Inleiding Kansrekening en Statistiek

S.J. de Lange

8-6-'93

Inleiding Kansrekening en Statistiek

CBC
91-01

WI
SV
Lang

Dit boek wordt ten behoeve van de Vereniging voor Studie- en Studentenbelangen te Delft (VSSD) uitgegeven door de Delftse Uitgevers Maatschappij. De VSSD is een vereniging van studenten aan de Technische Universiteit Delft, die zich ten doel stelt de belangen van studenten te behartigen.
Deze belangenbehartiging heeft vele, overigens samenhangende, kanten. De bevordering van de kwaliteit van het onderwijs, bezinning op de beroepspraktijk en het toegankelijker maken van het wetenschappleijk onderwijs voor alle lagen van de bevolking zijn de hoofdzaken van wat de 'ideële' kant van de belangenbehartiging genoemd kan worden.
De materiële kant van dit werk betreft het opkomen voor een aanvaardbaar inkomen voor de student, voor goede leefomstandigheden, goed en goedkoop studiemateriaal e.d.
Bij het verzorgen en doen uitgeven van boeken zoals deze zijn beide aspecten vertegenwoordigd: de beschikbaarheid van goede en handzame boeken vergroot de kwaliteit van het onderwijs, anderzijds worden ze zo goedkoop mogelijk in de handel gebracht.

Inleiding Kansrekening en Statistiek

ir. S.J. de Lange

Delftse Uitgevers Maatschappij

CIP-gegevens Koninklijke Bibliotheek, Den Haag

Lange, S.J. de

Inleiding kansrekening en statistiek / S.J. de Lange. — Delft : Delftsche Uitgevers Mij. — Ill., tab.
ISBN 90–6562–095–8
SISO 517.1 UDC 519.2 + 3.11(075.8)
Trefw.: kansrekening/statistiek.

Eerste druk 1989
Tweede druk 1991

Delftse Uitgevers Maatschappij b.v.
P.O. Box 2851, 2601 CW Delft, The Netherlands
Tel. 015–123725

ISBN 90 6562 095 8

Voorwoord

Dit boek is geschreven voor technische studenten die in één semester enige vaardigheid moeten krijgen in het hanteren van de begrippen en methoden van de kansrekening.
De zeer beperkte tijd die in de diverse studieprogramma's voor dit vak beschikbaar is heeft tot gevolg dat enige onderwerpen, die zelfs op dit niveau eigenlijk niet mogen ontbreken, niet of nauwelijks worden genoemd. Zo ontbreken voortbrengende en karakteristieke functies geheel en is de combinatoriek te vluchtig behandeld.
Daar statistiek een belangrijk toepassingsgebied is van de kansrekening worden ook de statistische standaardtechnieken behandeld.
De leer der verzamelingen is bekend verondersteld evenals de analyse die in het eerste jaar aan de TU Delft wordt onderwezen (in het bijzonder het berekenen van meervoudige integralen). De kansrekening zelf wordt echter, zij het beknopt, van de grond af aan opgebouwd.
Vooral in de latere hoofdstukken wordt een beroep op de intuïtie en de goedgelovigheid van de lezer gedaan. Het boek is dan ook niet bestemd voor (aspirant-) wiskundigen. Achterin is een aantal statistische tabellen opgenomen. Voor het toepassen van de besproken technieken zijn deze tabellen voldoende.
Vanzelfsprekend zijn op- en aanmerkingen, vooral voorstellen ter verbetering, welkom.
De vraagstukken zijn voor een deel overgenomen uit *Vraagstukken over waarschijnlijkheidsrekening* van dr. P.J.A. Kanters. Ik dank hem voor zijn toestemming daartoe. Mijn dank gaat voorts uit naar ir. Th.C.A. Mensch en dr. J.A.M. v.d. Weide voor hun bijdragen bij de totstandkoming van de inhoud. Zeer veel dank ben ik verschuldigd aan mevr. F.A. Zuidervaart-Murray en de medewerkers van de VSSD voor de prettige samenwerking bij de produktie van dit boek.

Delft, december 1988 S.J. de Lange

Inhoud

Notatie

Stochastische variabelen zijn aangeduid door onderstreping.

$\underline{x} \sim \ldots$: de stochastische variabele $\underline{x}$ heeft de … verdeling.

$a \approx b$: a is ongeveer gelijk aan b.

$P(\lambda)$: Poisson-verdeling met parameter λ.

$Exp(\lambda)$: Exponentiële verdeling met parameter λ, waarbij is gekozen voor die variant waarbij de verwachting $1/\lambda$ is.

$N(a;b)$: Normale verdeling met verwachting a en variantie b.

$B(n;p)$: Binomiale verdeling met parameters n en p.

$NB(r;p)$: Negatief-binomiale verdeling met parameters r en p.

Literatuur

Er zijn veel inleidende boeken over kansrekening en statistiek. Het merendeel is in het Engels. Uit die overvloed worden er slechts enkele genoemd.

Feller, W., *An Introduction to Probability Theory and its Applications*, Vol. I and Vol. II, John Wiley, New York.

Genugten, B.B. van der, *Inleiding tot de Waarschijnlijkheidsrekening en de Mathematische Statistiek*, deel I, Stenfert Kroese, Leiden.

Hogg, R.V. and A.T. Craig, *Introduction to Mathematical Statistics*, Collier MacMillan International Editions, London.

Kanters, dr. P.J.A., *Vraagstukken over waarschijnlijkheidsrekening*, DUM, Delft.

Larson, H.J., *Introduction to Probability Theory and Statistical Inference*, John Wiley, New York.

Meelen, A.J., J. van Soest en J.M.G. Vermeulen, *Aanvulling op Elementaire Statistiek*, DUM, Delft.

Mood, A.M., F.A. Graybill and D.C. Boes, *Introduction to the Theory of Statistics*, McGraw-Hill, New York.

Roes, P.B.M. en H.J.L. van Oorschot, *Kansrekening en Statistiek*, DUM, Delft.

Soest, ir. J. van, *Elementaire Statistiek*, DUM, Delft.

Stam, dr. A.J., *Inleiding tot de Waarschijnlijkheidsrekening*, Technische Uitgeverij H. Stam, Haarlem.

1

Kansrekening

1.1. Uitkomstenruimte en gebeurtenissen

In den beginne was er het experiment. Het resultaat dat optreedt na uitvoering van een experiment, de uitkomst, kan afhankelijk zijn van het toeval. Als dat het geval is schieten de gebruikelijke deterministische methoden tekort en moet men de kansrekening gebruiken om numerieke uitspraken te formuleren over de gevolgen van de uitvoering van het experiment.

Definities

- Een *experiment* is een handeling met één of meer mogelijke resultaten (uitkomsten).
- De *uitkomstenruimte*, Ω, is de verzameling van alle mogelijke uitkomsten van het experiment.
  Een element van Ω duiden we aan met ω.
- Een *gebeurtenis* A is een deelverzameling van de uitkomstenruimte Ω. De gebeurtenis A treedt op als het experiment eindigt in een van de uitkomsten die tot A behoren.
- Het *complement* $\overline{A}$ van de gebeurtenis A is de gebeurtenis dat A niet optreedt.
- Een *elementaire gebeurtenis* $\{\omega\}$ is een deelverzameling van Ω die slechts één element bevat.
- De *zekere gebeurtenis* is die deelverzameling van Ω die alle elementen van Ω bevat, m.a.w. Ω zelf.
- Een *onmogelijke gebeurtenis*, Φ, is een deelverzameling van Ω die geen enkel element bevat, m.a.w. een lege verzameling.
- Twee gebeurtenissen A en B heten *disjunct* of elkaar uitsluitend als $A \cap B = \Phi$.

Het aantal elementen van Ω kan eindig zijn of oneindig. In het laatste geval kan het aantal elementen aftelbaar zijn of meer dan aftelbaar. Is het aantal elementen eindig of aftelbaar dan noemt men Ω discreet.

Voorbeelden

a. Gooi een dobbelsteen éénmaal. Mogelijke uitkomsten: 1, 2, 3, 4, 5, 6;

$\Omega = \{1,2,3,4,5,6\}$. De gebeurtenis A: 'de worp is even' bevat de uitkomsten 2, 4 en 6; $A = \{2,4,6\}$. De gebeurtenis B: 'de worp is oneven' bevat de uitkomsten 1, 3 en 5; $B = \{1,3,5\}$.

$$A \cap B = \Phi; A \cup B = \Omega; \overline{A} = B.$$

b. Gooi een dobbelsteen net zo lang totdat de eerste 6 valt. De uitkomst van het experiment is het aantal malen dat de dobbelsteen moet worden gegooid.

$$\Omega = \{1,2,3,\ldots\}.$$

De uitkomstenruimte heeft oneindig veel elementen maar is aftelbaar.

c. Schiet op een schietschijf met een straal van 25 cm. De uitkomst van het experiment is de afstand in centimeters tussen het middelpunt van de schijf en het punt waar het schot de schijf heeft getroffen. Het is mogelijk dat de schijf niet wordt geraakt; ongeacht hoever het schot ernaast gaat, noemen we zo'n uitkomst 'mis'. De elementen van Ω hoeven dus niet gelijksoortig te zijn.

$$\Omega = \{\text{mis}\} \cup \{x | 0 \leq x \leq 25\}.$$

De uitkomstenruimte heeft meer dan aftelbaar veel elementen.
Zij A de gebeurtenis dat de schijf op minder dan 10 cm van het middelpunt wordt geraakt:

$$A = \{x | 0 \leq x < 10\}.$$

Zij B de gebeurtenis dat de schijf *niet* op minder dan 10 cm van het middelpunt wordt geraakt, m.a.w. $B = \overline{A}$:

$$B = \{\text{mis}\} \cup \{x | 10 \leq x \leq 25\}.$$

d. Gooi een rode en een groene dobbelsteen. Mogelijke uitkomsten: (1,1),(1,2), ..., (1,6),(2,1),...,(6,6) waarbij steeds het eerste getal het aantal ogen van de rode dobbelsteen aangeeft en het tweede getal dat van de groene.

$$\Omega = \{(i,j) \mid i = 1, 2, \ldots, 6; j = 1,2,\ldots,6\}.$$

De uitkomstenruimte bevat 36 elementen. Elk element bestaat uit een tweetal (i,j).
Zij A de gebeurtenis: 'het totale aantal ogen bij de worp is niet groter dan 4':

$$A = \{(1,1), (1,2), (1,3), (2,1), (2,2), (3,1)\}.$$

e. Laat de waterhoogte in cm ten opzichte van N.A.P., op een bepaald moment, op tien plaatsen langs de kust opmeten en registreer de resultaten:

$$\Omega = \{(x_1,x_2,\ldots,x_{10}) \mid a_i \leq x_i ; i = 1,2,\ldots,10\}.$$

Er is geen bovengrens aangegeven voor de x_i omdat er geen zekerheid is dat de waterhoogte een vooraf door ons gestelde grens niet zal overschrijden. De ondergrens kan echter gelijkgesteld worden aan het niveau van de zeebodem ter plaatse.

1.2. Axioma's van de kansrekening

Om de kansrekening goed te kunnen bedrijven hebben we een kansruimte nodig. Een kansruimte bestaat uit drie ingrediënten:

a. een uitkomstenruimte Ω;
b. een σ-algebra $\mathcal{A}$ van verzamelingen in Ω;
c. een kansmaat P die aan elk element A van $\mathcal{A}$ een getal, P(A), toekent.*

De uitkomstenruimte kwamen we al tegen in de vorige paragraaf. Het zou mooi zijn als we verder konden werken met de klasse van *alle* deelverzamelingen van Ω. Dat kan echter alleen als Ω discreet is. Als het aantal elementen van Ω meer dan aftelbaar is kunnen er problemen ontstaan zodat we ons enigszins moeten beperken, en wel tot een σ-algebra.

Definitie

Een σ-algebra $\mathcal{A}$ is een klasse van deelverzamelingen van Ω waarvoor het volgende geldt:

1. $\Omega \in \mathcal{A}$;
2. Als $A \in \mathcal{A}$ dan ook $\overline{A} \in \mathcal{A}$;
3. Als $A_i \in \mathcal{A}$ voor $i = 1, 2, 3, \ldots$ dan ook $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.

Het is gemakkelijk aan te tonen dat ook Φ en $\bigcap_{i=1}^{\infty} A_i$ tot $\mathcal{A}$ behoren. Kortom $\mathcal{A}$ is gesloten onder aftelbare verzamelingstheoretische operaties. Deelverzamelingen van Ω die niet tot $\mathcal{A}$ behoren zijn geen gebeurtenissen.

De kansmaat P is een functie van $\mathcal{A}$ naar [0,1] met de volgende eigenschappen:

1. $P(\Omega) = 1$;
2. $P(A) \geq 0$ voor alle $A \in \mathcal{A}$;
3. $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$ voor iedere rij disjuncte gebeurtenissen A_1, A_2, $A_3, \ldots$ (d.w.z. $A_i \cap A_j = \Phi$ als $i \neq j$).

Bij elkaar krijgt men zo de kansruimte $(\Omega, \mathcal{A}, P)$, ook een kans-triple genaamd. Er is een zeer grote vrijheid bij het bepalen van de functiewaarden van P. Het is echter gebruikelijk, en ook verstandig, om zodanige waarden te kiezen dat aan-

---

* Men komt ook andere schrijfwijzen tegen zoals P{A}, Pr(A), Pr{A} of P[A].

sluiting bij de praktijk wordt verkregen.
De drie eisen waaraan de kansmaat P moet voldoen worden de *axioma's* van de kansrekening genoemd. Uitgaande van deze axioma's kan een heel bouwwerk van stellingen worden opgetrokken.

Stelling 1.1. $P(\Phi) = 0$.
Bewijs: A, Φ, Φ, Φ, ... is een rij disjuncte gebeurtenissen, dus kan axioma 3 worden toegepast.

$$P(A \cup \Phi \cup \Phi \cup \ldots) = P(A) + P(\Phi) + P(\Phi) + \ldots \geq P(A) + P(\Phi)$$

$$P(A) \qquad\qquad \geq P(A) + P(\Phi) \quad \text{ofwel } P(\Phi) \leq 0.$$

Uit axioma 2 volgt echter dat $P(\Phi) \geq 0$ zodat $P(\Phi) = 0$. ❑

N.B.: Het omgekeerde geldt niet: als $P(A) = 0$ volgt daar niet uit dat $A = \Phi$.

Stelling 1.2. $P(A \cup B) = P(A) + P(B)$ als $A \cap B = \Phi$.
Bewijs: Als $A \cap B = \Phi$ dan is A, B, Φ, Φ, Φ, ... een rij disjuncte gebeurtenissen zodat axioma 3 weer kan worden gebruikt.

$$P(A \cup B \cup \Phi \cup \Phi \cup \Phi \cup \ldots) = P(A) + P(B) + P(\Phi) + P(\Phi) + P(\Phi) + \ldots$$

$$P(A \cup B) = P(A) + P(B) + 0 + 0 + 0 + \ldots$$

zodat $P(A \cup B) = P(A) + P(B)$. ❑

Deze stelling laat zich eenvoudig uitbreiden tot de volgende:

Stelling 1.3. Als $A_1, A_2, \ldots, A_n$ disjuncte gebeurtenissen zijn dan geldt

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i).$$ ❑

Stelling 1.4. $P(\overline{A}) = 1 - P(A)$.
Bewijs: $A \cap \overline{A} = \Phi$ dus $P(A \cup \overline{A}) = P(A) + P(\overline{A})$;
$A \cup \overline{A} = \Omega$ dus $1 = P(A) + P(\overline{A})$, waaruit het gestelde volgt. ❑

Stelling 1.5. $0 \leq P(A) \leq 1$.
Bewijs: Volgt direct uit axioma 2 en de voorgaande stelling. ❑

Stelling 1.6. $P(A \cap B) = P(B) - P(\overline{A} \cap B)$.
Bewijs: $B = B \cap \Omega = B \cap (A \cup \overline{A}) = (B \cap A) \cup (B \cap \overline{A})$.
$B \cap A$ en $B \cap \overline{A}$ zijn disjunct, dus
$P(B) = P\{(B \cap A) \cup (B \cap \overline{A})\} = P(B \cap A) + P(B \cap \overline{A})$ ofwel
$P(A \cap B) = P(B) - P(\overline{A} \cap B)$. ❑

Stelling 1.7. Als $A \in B$ dan geldt $P(A) \leq P(B)$.

Bewijs: Uit $A \in B$ volgt $A \cap B = A$. Dus geldt

$$P(B) = P(B \cap A) + P(B \cap \overline{A}) = P(A) + P(B \cap \overline{A}).$$

Volgens axioma 2 is $P(B \cap \overline{A}) \geq 0$ zodat

$$P(B) \geq P(A).$$

□

Stelling 1.8. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Bewijs: $A \cup B = (A \cap B) \cup (A \cap B) \cup (\overline{A} \cap B)$.

Het rechterlid is een vereniging van drie disjuncte gebeurtenissen. Dus

$$\begin{aligned} P(A \cup B) &= P(A \cap B) + P(A \cap B) + P(\overline{A} \cap B) = \\ &= P(A) - P(A \cap B) + P(A \cap B) + P(B) - P(A \cap B) = \\ &= P(A) + P(B) - P(A \cap B). \end{aligned}$$

Deze stelling, de *algemene somregel*, kan worden uitgebreid tot elk eindig aantal gebeurtenissen. De algemene vorm, bekend als de *regel van inclusie/exclusie*, is

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) - \ldots$$

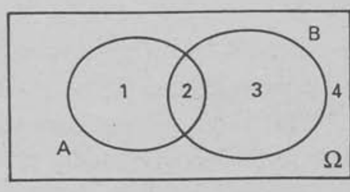
$$- (-1)^n P(A_1 \cap A_2 \cap \ldots \cap A_n).$$

In het geval dat $n = 3$ wordt het:

$$P(A \cup B \cup C) =$$

$$= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C).$$

Vaak is een Venn-diagram een goed hulpmiddel bij het bewijzen van stellingen. In een Venn-diagram duidt men grafisch de verschillende gebeurtenissen aan. Ter illustratie:



1: $A \cap \overline{B}$
2: $A \cap B$
3: $\overline{A} \cap B$
4: $\overline{A} \cap \overline{B}$

Figuur 1.1.

Voortaan zal, indien dat geen verwarring wekt, het doorsnijdingsteken '$\cap$' worden weggelaten en AB de betekenis $A \cap B$ hebben. Dit is vergelijkbaar met het weglaten van het maalteken in sommige gevallen: $ab = a \cdot b = a \times b$. Het verenigingsteken '$\cup$' zal echter (evenals het plusteken) steeds worden geplaatst.

1.3. Kansruimten

Om de kansruimte bruikbaar te maken moeten de functiewaarden van P worden vastgesteld. Iedere keuze die niet strijdig is met de axioma's is toelaatbaar. We willen echter resultaten krijgen die in praktische situaties betekenis hebben. Bij een experiment met een eindig aantal mogelijke uitkomsten kan dat worden bereikt door frequentie-quotiënten van de gebeurtenissen te bepalen. Men herhaalt het experiment n maal en telt het aantal malen n(A) dat daarbij de gebeurtenis A optreedt. Het frequentie-quotiënt $f_n(A)$ is dan gelijk aan

$$f_n(A) = \frac{n(A)}{n}.$$

Naarmate n groter wordt zal $f_n(A)$ minder fluctueren en in de buurt van een getal c blijven. Door nu P(A) = c te stellen verkrijgt men een bruikbare waarde voor de kans op de gebeurtenis A. Deze methode kan echter niet altijd worden toegepast.
Vaak is het mogelijk om bij een experiment een eindige uitkomstenruimte te bepalen, zodanig dat daarmede een symmetrische kansruimte kan worden geconstrueerd. In een symmetrische kansruimte heeft elk van de n elementen van Ω een even grote kans 1/n. De kans op een gebeurtenis A is dan gelijk aan het aantal uitkomsten van het experiment die tot de gebeurtenis A leiden, vermenigvuldigd met 1/n.
Iets anders geformuleerd verkrijgt men zo de klassieke kansdefinitie van Laplace:
Als bij een experiment alle mogelijke uitkomsten gelijkwaardig zijn dan geldt voor de kans dat de gebeurtenis A optreedt:

$$P(A) = \frac{\text{aantal uitkomsten dat gunstig is voor A}}{\text{totaal aantal mogelijke uitkomsten}}.$$

Voorbeelden

a. Gooi een zuivere dobbelsteen: $\Omega = \{1,2,3,4,5,6\}$. Omdat de dobbelsteen zuiver is zijn de uitkomsten gelijkwaardig en is $P(1) = P(2) = \ldots = P(6) = \frac{1}{6}$.

b. Gooi twee zuivere dobbelstenen. Als we alleen geïnteresseerd zijn in de som van de ogenaantallen is het mogelijk om te werken met $\Omega' = \{2,3,\ldots,12\}$. Hier zijn echter de elf uitkomsten niet gelijkwaardig. Een symmetrische kansruimte is mogelijk indien we onderscheid maken tussen de twee dobbelstenen zodat elke uitkomst door een getallenpaar wordt aangegeven. De uitkomsten (1,3) en (3,1) zijn verschillend omdat de drie in het eerste geval bij de tweede steen valt en in het laatste geval bij de eerste steen. Zo ontstaat een uitkomstenruimte met 36 elementen:

$$\Omega = \{(1,1),(1,2),\ldots,(1,6),(2,1),\ldots,(6,6)\},$$

waarin elk element een kans 1/36 heeft.

Iets meer moeite kost de constructie van een kansruimte als Ω oneindig veel elementen bevat, zoals uit de volgende voorbeelden blijkt.

c. Gooi een zuivere dobbelsteen net zo lang totdat de eerste 4 gevallen is. De uitkomst van het experiment is het aantal benodigde worpen, dus

$$\Omega = \{1,2,3,\ldots\}.$$

In voorbeeld b had net zo goed met een dobbelsteen twee keer achter elkaar gegooid kunnen worden. Dan blijkt bij zes van de 36 uitkomsten de eerste worp al een 4 op te leveren. Er zijn 5 uitkomsten waarbij pas bij de tweede worp een 4 verschijnt.
Een symmetrische kansruimte voor het drie maal werpen met een zuivere dobbelsteen heeft een uitkomstenruimte met 216 elementen van de vorm (i, j, k). Daarbij zijn er 36 met $i = 4$; 30 met $i \neq 4$ en $j = 4$; 25 met $i \neq 4$, $j \neq 4$ en $k = 4$. Dus

$$p_1 = P\{\text{de eerste 4 valt bij worp 1}\} = \frac{36}{216};$$

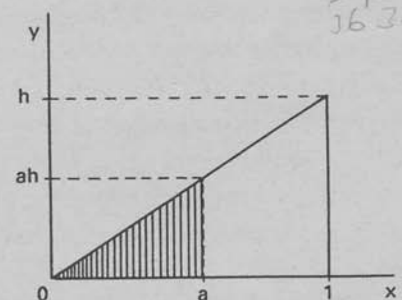
$$p_2 = P\{\text{de eerste 4 valt bij worp 2}\} = \frac{30}{216};$$

$$p_3 = P\{\text{de eerste 4 valt bij worp 3}\} = \frac{25}{216}.$$

Het valt op dat $p_2 = \frac{5}{6} p_1$ en $p_3 = (\frac{5}{6})^2 \cdot p_1$. Het vermoeden is dus dat bij het gooien net zolang totdat er een 4 valt zal gelden:
$P\{\text{de eerste 4 valt bij worp } n{+}1\} = \frac{5}{6} P\{\text{de eerste 4 valt bij worp } n\}$, $n = 1,2,3,\ldots$
waaruit volgt:

$$P\{\text{de eerste 4 valt bij worp } n\} = \frac{5^{n-1}}{6^n}, \; n = 1,2,3,\ldots .$$

Later zal blijken dat dit vermoeden gegrond is.
Merk op dat $p_n = \{1 - (p_1 + p_2 + \ldots + p_{n-1})\}/6$. $\frac{5}{36} + \frac{5}{36} +$



Figuur 1.2.

$5^0 = \frac{1}{6} + \frac{5}{36} + 25$

d. Prik blindelings in een driehoekig stuk karton als aangegeven in figuur 1.2. De uitkomst van het experiment is de x-coördinaat van het geprikte punt.

$$\Omega = \{x | 0 \leq x \leq 1\}.$$

'Blindelings' geeft aan dat elk punt in de driehoek een even grote kans heeft om gekozen te worden.
Uit het blindelings prikken volgt dat oppervlakten van gelijke grootte gelijke kans hebben om het geprikte punt te bevatten. Daaruit volgt weer dat de kans om in een deelgebied M van de driehoek te prikken gelijk is aan

$$\frac{\text{oppervlakte van gebied M}}{\text{oppervlakte van de hele driehoek}}.$$

Zij A de gebeurtenis dat $0 \leq x \leq a$ dan is dus

$$P(A) = \frac{1/2 \cdot a \cdot ah}{1/2 \cdot 1 \cdot h} = a^2.$$

Zij B de gebeurtenis dat $x = a$ dan volgt

$$P(B) = \frac{0}{1/2 \cdot 1 \cdot h} = 0$$

omdat het oppervlak dat gunstig is voor de gebeurtenis B (een lijn) de grootte nul heeft. Het is nu dus niet zinvol om, zoals bij een discrete uitkomstenruimte, voor elk element de kans aan te geven.
In dit geval hadden we ook een σ-algebra van toegelaten gebeurtenissen moeten kiezen. De σ-algebra van alle intervallen is ruim genoeg voor praktisch gebruik en beperkt genoeg om ons te behoeden voor problemen.

1.4. Combinatoriek

Symmetrische kansruimten worden al gauw erg groot. Het tellen van het aantal gunstige (en mogelijke) uitkomsten moet dan systematisch gebeuren. De mogelijkheid daartoe biedt de combinatoriek. Alle resultaten van de combinatoriek volgen uit de volgende twee grondregels.

Produktregel

Als handeling A op n_1 manieren kan worden uitgevoerd en handeling B op n_2 manieren, dan zijn er $n_1 \cdot n_2$ manieren om 'handeling A en handeling B' uit te voeren, mits de volgorde van A en B vastligt of er niet toe doet.

Somregel

Als handeling A op n_1 manieren kan worden uitgevoerd en handeling B op n_2 manieren, dan zijn er $n_1 + n_2$ manieren om 'handeling A of handeling B' uit te voeren, mits het niet mogelijk is met één handeling zowel A als B uit te voeren.

Reeds bekende en zeer bruikbare resultaten zijn:

- Het aantal permutaties van n voorwerpen is n!
- Het aantal permutaties van r voorwerpen uit n voorwerpen is $\frac{n!}{(n-r)!}$ voor $n \geq r$.
- Het aantal combinaties van r voorwerpen uit n voorwerpen is $\binom{n}{r}$ voor $n \geq r$.

Het aantal combinaties van r uit n is ook het aantal manieren waarop n voorwerpen over twee personen, A en B, verdeeld kunnen worden, zó dat A r voorwerpen krijgt en B n – r voorwerpen, waarbij A en B niet op de volgorde letten.

Dit kan worden uitgebreid tot het verdelen van n voorwerpen over k personen waarbij persoon i r_i voorwerpen ontvangt, $i = 1,2,\ldots, k$, en $\sum_{i=1}^{k} r_i = n$ en niet op de volgorde wordt gelet. Kies eerst r_1 voorwerpen uit voor persoon 1; dit kan op $\binom{n}{r_1}$ manieren. Kies uit de $n - r_1$ resterende de r_2 voorwerpen voor persoon 2; dat kan op $\binom{n-r_1}{r_2}$ manieren, enz. Het totaal aantal manieren wordt dan

$$\binom{n}{r_1}\cdot\binom{n-r_1}{r_2}\ldots\binom{n-r_1-r_2\ldots r_{k-1}}{r_k}$$

hetgeen na uitschrijven van de binominaalcoëfficiënten

$$\frac{n!}{r_1!r_2!\ldots r_k!}$$

oplevert.
Men krijgt hetzelfde resultaat voor het aantal permutaties van n voorwerpen als er daarbij r_i van soort i zijn, voor $i = 1,2,\ldots, k$, terwijl $\sum_{i=1}^{k} r_i = n$. Hierbij worden de voorwerpen van één soort niet van elkaar onderscheiden.

In combinatorische problemen komt men vaak de term ‘trekken met teruglegging’ of ‘trekken zonder teruglegging’ tegen. Bij trekken met teruglegging wordt elk getrokken voorwerp weer toegevoegd aan de verzameling waar men uit trekt, voordat een volgende trekking wordt gedaan. Het herhaald gooien met een dobbelsteen kan men beschouwen als trekken met teruglegging uit de verzameling $\{1,2,3,4,5,6\}$.
Bij trekken zonder teruglegging blijven de getrokken voorwerpen terzijde liggen. De verzameling waaruit men trekt wordt dus steeds kleiner. Als er wordt getrokken uit een vaas met n ballen dan kan men bij trekken zonder teruglegging hoogstens n maal een bal trekken. Bij trekken met teruglegging is er niet zo’n grens.
De toevoeging van de term ‘aselect’, ‘blindelings’ of ‘willekeurig’ duidt aan dat alle vergelijkbare uitkomsten even waarschijnlijk zijn. Dat betekent dus dat men een symmetrische kansruimte kan gebruiken.

1.5. Enige voorbeelden

a. In een doos zitten 7 bouten en 3 moeren. We pakken blindelings twee keer iets uit de doos zonder de getrokken voorwerpen terug te leggen. Hoe groot is de kans op de gebeurtenis A dat we een bout en een moer trekken.

Volgorde vast:

Om een symmetrische kansruimte te krijgen nummeren we de voorwerpen 1,2,...,10 (1-7 de bouten, 8-10 de moeren). Als uitkomstenruimte kiezen we de paren (i,j) met $i \neq j$ en $i = 1,2,\ldots,10$; $j = 1,2,\ldots,10$, waarbij i(j) het resultaat van de eerste (tweede) keer pakken weergeeft. Er zijn $\frac{10!}{(10-2)!} = 90$ permutaties van 2 uit 10 mogelijk, dus Ω heeft 90 elementen, elk met kans 1/90. Hoeveel van deze uitkomsten zijn er gunstig voor A? Dat zijn al de uitkomsten waarbij $i \leq 7$ en $j \geq 8$ is of waarbij $i \geq 8$ en $j \leq 7$ is. Volgens de produktregel, die mag worden toegepast omdat de volgorde vastligt zijn er $7 \cdot 3$ manieren om een $i \leq 7$ en een $j \geq 8$ te kiezen terwijl er $3 \cdot 7$ manieren zijn om een $i \geq 8$ en een $j \leq 7$ te kiezen. De somregel mag worden toegepast zodat er $21 + 21 = 42$ gunstige mogelijkheden zijn en dus volgt:

$$P(A) = \frac{42}{90} = \frac{7}{15}.$$

Volgorde doet er niet toe:

In dit geval was het echter helemaal niet nodig om de volgorde waarin de voorwerpen gepakt worden in de beschouwingen te betrekken. Alleen wat uiteindelijk naast de doos ligt is van belang, niet of een gekozen voorwerp als eerste of als tweede uit de doos kwam. Zo redenerend komt men tot een uitkomstenruimte met $\binom{10}{2} = 45$ elementen (i, j); $i = 1,2,\ldots,10$; $j = i+1,i+2,\ldots,10$.

Het aantal gunstige uitkomsten volgt weer m.b.v. de produktregel, die nu mag worden toegepast omdat de volgorde er niet toe doet, en is gelijk aan $7 \cdot 3 = 21$. Ook op deze wijze verkrijgt men

$$P(A) = \frac{21}{45} = \frac{7}{15}.$$

b. Kies een willekeurig getal van drie cijfers. Hoe groot is de kans op de gebeurtenis B dat minstens een van de drie cijfers een 6 is?

$\Omega = \{100,101,\ldots, 999\}$ en bevat 900 elementen. De kansruimte is symmetrisch. Het bepalen van het aantal gunstige uitkomsten kan door inspectie van de uitkomsten gebeuren. Handiger is het echter om P(B) te bepalen via $P(\overline{B})$, en dan gebruik te maken van $P(B) = 1 - P(\overline{B})$. $\overline{B}$ is de gebeurtenis dat het gekozen getal geen enkele 6 bevat. Hoeveel van zulke getallen zijn er in Ω? Voor het eerste cijfer zijn er 8 mogelijkheden (0 en 6 vallen uit). Voor het tweede zowel als voor het derde cijfer zijn er 9 mogelijkheden (alleen de 6 mag niet). Met behulp van de produktregel krijgen we dus 8·9·9 uitkomsten waar geen 6 in voorkomt.

Dus $\quad P(\overline{B}) = \frac{8\cdot 9\cdot 9}{900} = \frac{18}{25}$

en $\quad P(B) = 1 - \frac{18}{25} = \frac{7}{25}$.

c. De zes deelnemers aan een ronde-tafel conferentie kunnen op 5! verschillende manieren worden geplaatst. Omdat de tafel rond is maakt het niet uit op welke stoel de eerste deelnemer gaat zitten. De vijf anderen kunnen dan op 5! manieren over de resterende vijf stoelen worden verdeeld.

d. Iemand heeft een binair getal van vier cijfers opgeschreven, d.w.z. een rij ter lengte vier met op elke plaats een 0 of een 1. U mag raden welk getal het is. Er zijn $2\times2\times2\times2 = 16$ mogelijkheden en dus is de kans dat u het getal goed raadt 1/16. De kans dat u minstens 3 van de 4 symbolen goed raadt is 5/16 (ga dit na). Als u nu als extra informatie krijgt dat het getal uit 2 nullen en 2 enen bestaat zijn er maar 6 getallen mogelijk. De kans dat u de juiste volgorde van 2 nullen en twee enen raadt is 1/6. De kans dat u, zodoende, precies 3 van de 4 symbolen goed raadt is 0 (waarom?). U kunt echter toch de kans dat u minstens 3 van de vier symbolen goed raadt gelijk aan $\frac{1}{2}$ maken en wel door in een willekeurige volgorde 1 nul en 3 enen (of 3 nullen en 1 een) op te schrijven. U hebt dan met kans $\frac{1}{2}$ 1 symbool goed en eveneens met kans $\frac{1}{2}$ 3 symbolen goed.

1.6. Conditionele kans

Laat een experiment bestaan uit het tweemaal gooien met een zuivere dobbelsteen en de belangstelling uitgaan naar het totaal aantal ogen dat daarbij wordt geworpen.

$$\Omega = \{(i,j) | i = 1,2,\ldots,6;\ j = 1,2,\ldots,6\}.$$

Zij A de gebeurtenis dat $i + j = 10$, dan is $P(A) = 3/36$.
Zij B de gebeurtenis dat de eerste worp een drie is.
Met een drie bij de eerste worp is het onmogelijk om nog een totaal van tien te halen. Met andere woorden de kans op de gebeurtenis A, als gegeven is dat de gebeurtenis B is opgetreden is gelijk aan nul. Men spreekt in zo'n geval van de *conditionele* (of *voorwaardelijke*) kans op A gegeven B.
In het algemeen kan men, zodra bekend is dat B is opgetreden, een aangepaste kansmaat P_B bij het experiment construeren waarbij alle elementen van $\overline{B}$ de kans nul krijgen omdat zij niet meer als uitkomst van het experiment kunnen optreden. Het experiment kan alleen nog maar eindigen in één van de uitkomsten die bevat

zijn in B. Als de gebeurtenis A optreedt eindigt het experiment in een van de uitkomsten die bevat zijn in A. Dus A treedt op bij gegeven B als het experiment eindigt in een van de uitkomsten die behoren tot AB. Evenzo volgt dat $\overline{A}$ optreedt bij gegeven B als het experiment eindigt in een van de uitkomsten behorend tot $\overline{A}B$. Het is duidelijk dat $P_B(A) + P_B(\overline{A}) = 1$ moet zijn en dat

$$P_B(A) : P_B(\overline{A}) = P_B(AB) : P_B(\overline{A}B).$$

Hieruit volgt

$$P_B(A) : 1 = P(AB) : P(B), \text{ zodat } P_B(A) = \frac{P(AB)}{P(B)}.$$

2/3

Natuurlijk moet $P(B) \neq 0$ zijn. De gebruikelijke notatie voor de conditionele kans $P_B(A)$ is $P(A|B)$. Het is gemakkelijk aan te tonen dat de conditionele kansmaat aan de drie axioma's voldoet, m.a.w. dat voor $P(B) \neq 0$ geldt:

1. $P(\Omega|B) = 1$;
2. $P(A|B) \geq 0$ voor alle $A \in \mathcal{A}$;
3. $P(\bigcup_i A_i|B) = \sum_i P(A_i|B)$ mits $A_iA_j = \Phi$ voor $i \neq j$.

Dat betekent dat elke stelling die geldt voor gewone (of absolute) kansen ook geldt voor conditionele kansen.

Voorbeeld

Men trekt aselect en zonder teruglegging tweemaal een bal uit een vaas die 4 witte en 6 rode ballen bevat. Zij W_i de gebeurtenis dat de i^e getrokken bal wit is, $i = 1$ of 2.

$$P(W_1) = \frac{4}{10} = \frac{2}{5};\ P(W_1W_2) = \frac{4\cdot 3}{10\cdot 9} = \frac{2}{15};\ P(\overline{W}_1W_2) = \frac{6\cdot 4}{10\cdot 9} = \frac{4}{15};$$

$$P(W_2|W_1) = \frac{P(W_1W_2)}{P(W_1)} = \frac{2/15}{2/5} = \frac{1}{3};$$

$$P(W_2) = P(W_1W_2) + P((\overline{W}_1W_2) = \frac{2}{15} + \frac{4}{15} = \frac{2}{5};$$

$$P(W_1|W_2) = \frac{P(W_1W_2)}{P(W_2)} = \frac{2/15}{2/5} = \frac{1}{3}.$$

Voorbeeld

Men gooit tweemaal met een zuivere dobbelsteen. Hoe groot is de conditionele kans op 'tweemaal even' gegeven dat minstens één van de twee worpen een even resultaat geeft.

Zij A de gebeurtenis 'tweemaal even'; $P(A) = \frac{3\cdot 3}{6\cdot 6} = 1/4$.

Zij B de gebeurtenis 'minstens éénmaal even' dan is $\overline{B}$ 'tweemaal oneven';

$P(\overline{B}) = \frac{3 \cdot 3}{6 \cdot 6} = \frac{1}{4}$ zodat $P(B) = \frac{3}{4}$. Gevraagd wordt $P(A|B) = \dfrac{P(AB)}{P(B)}$.
Omdat $A \in B$ volgt $AB = A$, dus $P(A|B) = \dfrac{P(A)}{P(B)} = \dfrac{1/4}{3/4} = \dfrac{1}{3}$. □

Stelling 1.9.
$$P(AB) \begin{cases} = P(A|B) \cdot P(B), & \text{als } P(B) \neq 0; \\ = 0, & \text{als } P(B) = 0. \end{cases}$$
Bewijs: De eerste regel volgt onmiddellijk uit $P(A|B) = \dfrac{P(AB)}{P(B)}$; de tweede uit $0 \leq P(AB) \leq P(B)$. □

Stelling 1.10. $P(A_1A_2 \ldots A_n) =$
$$= P(A_n|A_1A_2 \ldots A_{n-1}) \cdot P(A_{n-1}|A_1A_2 \ldots A_{n-2}) \ldots P(A_2|A_1) \cdot P(A_1),$$
mits $P(A_1A_2 \ldots A_{n-1}) \neq 0$.
Dit heet de *algemene produktregel.*
Bewijs: $P(A_2|A_1) \cdot P(A_1) = P(A_1A_2)$ want uit $P(A_1A_2 \ldots A_{n-1}) \neq 0$ volgt $P(A_1) \neq 0$ omdat $A_1 \supset A_1A_2 \ldots A_{n-1}$.

Pas nu weer de vorige stelling toe waarbij A_1A_2 optreedt als conditie B dan volgt $P(A_3|A_1A_2) \cdot P(A_1A_2) = P(A_1A_2A_3)$, enz. □

Definitie

Een *partitie* $\mathcal{B}$ van Ω is een klasse van deelverzamelingen B_i, $i = 1,2,\ldots,n$, van Ω zodanig dat:

$$\bigcup_i B_i = \Omega;$$
$$B_iB_j = \Phi \text{ als } i \neq j;$$
$$P(B_i) \neq 0.$$

Stelling 1.11. Zij $\mathcal{B}$ een partitie van Ω dan geldt:

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i).$$

Dit heet de stelling van de *totale waarschijnlijkheid.*
Bewijs: $P(A|B_i) \cdot P(B_i) = P(AB_i)$ want $P(B_i) \neq 0$.
AB_i en AB_j zijn disjunct want B_i en B_j zijn disjunct, dus:

$$\sum_{i=1}^{n} P(A|B_i) \cdot P(B_i) = \sum_{i=1}^{n} P(AB_i) = P(\bigcup_{i=1}^{n} AB_i) = P(A \cap \bigcup_{i=1}^{n} B_i) =$$

$$= P(A\Omega) = P(A).$$ □

Stelling 1.12. Zij B een partitie van Ω en $P(A) \neq 0$ dan geldt:

$$P(B_k|A) = \frac{P(A|B_k)P(B_k)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)}.$$

Dit is de stelling van *Bayes*.

Bewijs: Volgens de stelling van de totale waarschijnlijkheid is

$$\sum_{i=1}^{n} P(A|B_i)P(B_i) = P(A).$$

Volgens de algemene produktregel is $P(A|B_k)P(B_k) = P(AB_k)$. Dit ingevuld levert $P(B_k|A) = \frac{P(AB_k)}{P(A)}$ hetgeen juist is omdat $P(A) \neq 0$. □

Voorbeeld

Een kast heeft drie laden. Lade A bevat 2 gouden munten, lade B een gouden en een zilveren terwijl lade C twee zilveren munten bevat. Men kiest aselect een lade en vervolgens aselect een munt uit die lade.

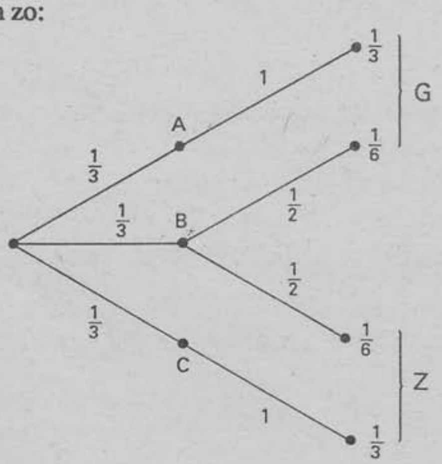
Zij A de gebeurtenis dat lade A wordt gekozen, idem voor B en C. Zij G de gebeurtenis dat een gouden munt wordt gekozen.

$$P(G) = P(G|A)P(A) + P(G|B)P(B) + P(G|C)P(C) = 1 \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} = \frac{1}{2}.$$

$$P(B|G) = \frac{P(G|B)P(B)}{P(G)} = \frac{1/2 \cdot 1/3}{1/2} = \frac{1}{3}.$$

Zo volgt ook $P(A|G) = \frac{2}{3}$ en $P(C|G) = 0$.

Men kan dit illustreren met een kansboom waarin alle mogelijkheden om het experiment uit te voeren grafisch worden weergegeven . Voor het bovenstaande voorbeeld krijgt men zo:



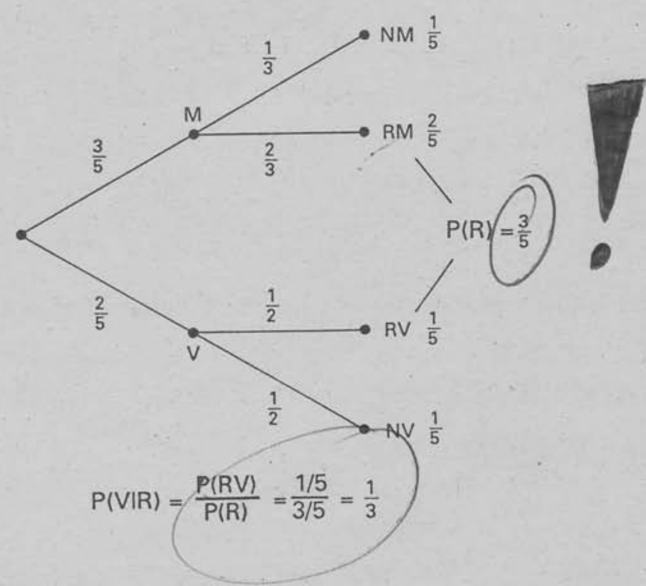
Figuur 1.3.

Hieruit valt af te lezen dat $P(G) = \frac{1}{2}$ en $P(AG) = \frac{1}{3}$.

Voorbeeld

In een trein zitten 60 mannen en 40 vrouwen. Van de mannen rookt 2/3, van de vrouwen de helft. De conducteur kiest aselect een reiziger uit waarvan hij het plaatsbewijs gaat controleren. Zij M(V) de gebeurtenis 'het is een man(vrouw)'. Zij R(N) de gebeurtenis 'de gekozen persoon rookt (rookt niet)'.

$$P(V|R) = \frac{P(R|V)\cdot P(V)}{P(R|V)P(V) + P(R|M)P(M)} = \frac{1/2\cdot 40/100}{1/2\cdot 40/100 + 2/3\cdot 60/100} = \frac{1/5}{3/5} = \frac{1}{3}.$$



Figuur 1.4.

1.7. Onafhankelijkheid

Definities

- Twee gebeurtenissen A en B zijn (stochastisch) onafhankelijk als geldt:

$$P(AB) = P(A)\cdot P(B).$$

- De gebeurtenissen $A_1, A_2, \ldots, A_n$ zijn (onderling) onafhankelijk als geldt:

$$P(A_iA_j) = P(A_i)\cdot P(A_j) \text{ voor alle } i \neq j$$

en

$$P(A_iA_jA_k) = P(A_i)\cdot P(A_j)\cdot P(A_k) \text{ voor alle } i \neq j \neq k$$

en

$$\vdots$$

en

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P(A_i).$$

– De gebeurtenissen $A_1, A_2, \ldots, A_n$ zijn paarsgewijs onafhankelijk als geldt:

$$P(A_iA_j) = P(A_i)\cdot P(A_j) \text{ voor alle } i \neq j.$$

Het is nodig om er met nadruk op te wijzen dat onafhankelijkheid en disjunctie van twee gebeurtenissen geheel verschillende begrippen zijn. Zodra de uitkomstenruimte van een experiment is bepaald kan worden nagegaan of A en B al dan niet disjunct zijn. Over onafhankelijkheid van A en B kan pas iets worden gezegd als de kansmaat P gekozen is. Door de keuze van P kan men desgewenst A en B onafhankelijk maken.

Stelling 1.13. Als A en B disjuncte gebeurtenissen zijn dan zijn zij alleen dan onafhankelijk als minstens één van de twee de kans nul heeft.
Bewijs: $AB = \Phi$ dus $P(AB) = 0$ volgt uit de disjunctie. Voor onafhankelijkheid moet gelden $P(AB) = P(A)\cdot P(B)$, dus hier $P(A)\cdot P(B) = 0$. □

Stelling 1.14. Als A en B onafhankelijke gebeurtenissen zijn en $0 < P(A) < 1$ dan geldt dat:

A^c

$$P(B|A) = P(B|\overline{A}) = P(B).$$

$A \cap B$

Bewijs: $$P(B|A) = \frac{P(AB)}{P(A)} = \frac{P(A)\cdot P(B)}{P(A)} = P(B).$$

$$\boxed{\overline{A}B} \cup \boxed{AB} = B, \text{ dus } P(\overline{A}B) + P(AB) = P(B) \text{ ofwel}$$

$$P(\overline{A}B) = P(B) - P(A)\cdot P(B) = P(B)\cdot P(\overline{A}),$$

zodat ook $P(B|\overline{A}) = P(B)$. □

Dit resultaat verklaart de naam onafhankelijkheid. Het wel of niet optreden van de gebeurtenis A heeft geen invloed op de kans dat gebeurtenis B optreedt.
Bij het construeren van een kansmaat zorgt men er dus voor dat gebeurtenissen waarvan men aanneemt dat ze elkaars kans van optreden niet beïnvloeden, onafhankelijk zijn.

Laat een experiment E bestaan uit het uitvoeren van n onafhankelijke proeven E_i, $i = 1,2,\ldots,n$. Bij elke proef E_i bepaalt men de uitkomstenruimte Ω_i met de bijbehorende kansmaat P_i. Een geschikte uitkomstenruimte Ω voor E is de produktruimte $\Omega_1 \times \Omega_2 \times \ldots \times \Omega_n$. Elk element ω van Ω is van de vorm $(\omega_1,\omega_2,\ldots,\omega_n)$ met $\omega_i \in \Omega_i$. Door nu te stellen

$$P(\{\omega\}) = P_1(\{\omega_1\}) \cdot P_2(\{\omega_2\}) \ldots P_n(\{\omega_n\})$$

ontstaat een kansmaat voor Ω waarin de gewenste onafhankelijkheid aanwezig is.

Voorbeeld

Werp tweemaal met een zuivere munt en eenmaal met een zuivere dobbelsteen.
$\Omega_1 = \{k, m\}$, $\Omega_2 = \{k, m\}$, $\Omega_3 = \{1,2,3,4,5,6\}$;
$P_1(\{k\}) = P_1(\{m\}) = 1/2$; $P_2(\{k\}) = P_2(\{m\}) = 1/2$; $P_3(\{i\}) = 1/6$ voor $i = 1,2,\ldots, 6$;
$P(\{(k, m, 5)\}) = P_1(\{k\}) \cdot P_2(\{m\}) \cdot P_3(\{5\}) = 1/2 \cdot 1/2 \cdot 1/6 = 1/24$.
Zij A de gebeurtenis dat 1× kruis, 1× munt en hoogstens 4 met de dobbelsteen wordt geworpen. Noem 'hoogstens 4' A_3 dan volgt dat

$$A = \bigcup_{i=1}^{4} \{(m, k, i)\} \cup \bigcup_{j=1}^{4} \{(k, m, j)\}$$

zodat $$P(A) = \sum_{i=1}^{4} P(\{(m, k, i)\}) + \sum_{j=1}^{4} P(\{(k, m, j)\}) =$$

$$= P_1(m) \cdot P_2(k) \cdot P_3(A_3) + P_1(k) \cdot P_2(m) \cdot P_3(A_3) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{2}{3} + \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{2}{3} = \frac{1}{3}.$$

Dezelfde constructie wordt ook toegepast bij een aftelbaar aantal onafhankelijke deel-experimenten maar is daar formeel iets gecompliceerder.

Definitie

Een aftelbare rij gebeurtenissen is onderling onafhankelijk als iedere eindige greep uit deze rij uit onafhankelijke gebeurtenissen bestaat.

Naar analogie van deze definitie kan men hier denken aan een experiment E dat bestaat uit de aftelbare rij onafhankelijke deel-experimenten $E_1, E_2, \ldots$ Als er voor elke eindige greep uit de rij deel-experimenten een uitkomstenruimte en kansmaat op de bovenstaande wijze wordt geconstrueerd dan gaat het wel goed.

Voorbeeld

Gooi met een zuivere dobbelsteen tot de eerste zes valt. Vooraf kunnen we geen grens aangeven waar het aantal benodigde worpen onder zal blijven. Beschouw het eenmaal werpen van de dobbelsteen als deel-experiment E_i met uitkomstenruimte $\Omega_i = \{s, m\}$ waarbij s staat voor succes (= 6 valt) en m voor mislukking (1,2,3,4 of 5 valt). Neem als P_i: $P_i(m) = 5/6$ en $P_i(s) = 1/6$.
Voor het berekenen van de kans op de gebeurtenis A_n dat bij de n^e worp $(n = 1,2,\ldots)$ de eerste zes valt hebben we alleen de eindige greep $E_1, E_2, \ldots, E_n$ nodig.

$B|A = B|A^c = P(B)$

$A_n = \{(\underbrace{m,m,m,\ldots,m}_{(n-1)x},s)\}$ zodat

$$P(A_n) = P_1(m)\cdot P_2(m)\ \ldots\ P_{n-1}(m)\cdot P_n(s) = (\tfrac{5}{6})^{n-1}.\tfrac{1}{6},\quad n = 1, 2, \ldots$$

Het gaat ook goed als men de kans op de gebeurtenis $B_n(n = 1,2,\ldots)$, dat de eerste zes valt bij of na de n^e worp, wil berekenen. Voor het optreden van B_n is nodig en voldoende dat de eerste $n-1$ worpen alle een mislukking opleveren. Men vindt zo $P(B_n) = (5/6)^{n-1}$, $n = 1,2,\ldots$ De kans dat het aantal worpen tot en met de eerste zes even is kan als volgt worden bepaald.

$$P(A_2 \cup A_4 \cup \ldots \cup A_{2m}) = \sum_{i=1}^{m} P(A_{2i}) = \frac{1}{6}\sum_{i=1}^{m} (\tfrac{5}{6})^{2i-1} =$$

$$= \frac{1}{5}\sum_{i=1}^{m} (\tfrac{25}{36})^{i} = \frac{1}{5}\,\frac{25}{36}\,\frac{1-(25/36)^m}{1-25/36} = \frac{5}{11}\{1-(\tfrac{25}{36})^m\}.$$

Maar voor de gebeurtenis 'even aantal worpen nodig' is het niet nodig m te begrenzen, zodat door het onbegrensd laten groeien van m volgt:

$$P(\text{even aantal worpen}) = \frac{5}{11}\,.$$

2

Stochastische variabelen

2.1. Kansfunctie

Veelal is van de uitkomst van een experiment alleen een bepaald aspect van belang. Als zich dit dan ook nog met behulp van een getal laat beschrijven kan met vrucht gebruik worden gemaakt van de technieken die zijn ontwikkeld voor stochastische variabelen.

Definitie
Een stochastische variabele is een functie van Ω naar $\mathbb{R}$.

Met andere woorden, een stochastische variabele kent aan elke $\omega \in \Omega$ een reëel getal toe. Heeft men meer getallen nodig om de eigenschappen van de uitkomst die van belang zijn vast te leggen dan kan men werken met een stochastische vector. □

Definitie
Een stochastische vector is een functie van Ω naar $\mathbb{R}^n$. □

Voorbeeld

1. Herhaal viermaal een experiment dat kan resulteren in s = succes of m = mislukking.
   De volgende vier elementen van Ω: (m, m, m, s), (m, m, s, m), (m, s, m, m) en (s, m, m, m) hebben gemeen dat er steeds 1 succes en 3 mislukkingen optreden. Als we alleen geïnteresseerd zijn in het aantal successen in de proevenreeks dan is een stochastische variabele toepasselijk die aan elke $\omega \in \Omega$ het getal toekent dat het bijbehorende aantal successen aangeeft. In dit geval dus 0, 1, 2, 3 of 4.

Een stochastische variabele wordt aangeduid door onderstreping, de waarde die een stochastische variabele $\underline{x}$ aan $\omega \in \Omega$ toekent door $\underline{x}(\omega)$. Na uitvoering van het experiment is bekend welke waarde $\underline{x}$ heeft aangenomen. Die waarde heet een realisatie van $\underline{x}$. De verzameling van alle mogelijke realisaties heet het waardenbereik W_x van $\underline{x}$.

Bijvoorbeeld in het voorgaande voorbeeld: $\underline{x}$ is het aantal successen bij het viermaal herhalen van het experiment. Het waardenbereik van $\underline{x}$ is $\{0, 1, 2, 3, 4\}$. Als $\omega_1 = (m, s, m, s)$ dan is $\underline{x}(\omega_1) = 2$.

Definities

- Een stochastische variabele heet discreet als het waardenbereik hoogstens aftelbaar veel getallen bevat.
- Een stochastische variabele heet continu als het waardenbreik uit één of meer intervallen bestaat.

❑

Voorbeeld

2. U vertrekt elke ochtend om 8.30 uur per fiets van huis naar school. De stochastische variabele $\underline{x}$ geeft de tijd in minuten aan die u nodig hebt om de school te bereiken. Dan is $\underline{x}$ een continue stochastische variabele met W_x bijvoorbeeld [10,20]. Als $\underline{y}$ aangeeft hoe vaak u daarbij voor een stoplicht hebt moeten wachten dan is $\underline{y}$ een discrete stochastische variabele met $W_y = \{0,1,2,\ldots\}$. Desgewenst kunt u $\underline{x}$ en $\underline{y}$ samen beschouwen als een stochastische vector $(\underline{x},\underline{y})$. De Ω kan hier van zeer simpel tot uiterst ingewikkeld gekozen worden.

Definitie

De kansfunctie van de discrete stochastische variabele $\underline{x}$ is een functie van $\mathbb{R}$ naar [0, 1] waarvan de waarde (notatie $P(\underline{x} = x)$ of $p_x(x)$) aangeeft hoe groot de kans is dat $\underline{x}$ de waarde x aanneemt.

❑

Zodra een stochastische variabele gedefinieerd is in een gegeven kansruimte ligt zijn kansfunctie vast. Immers de kans dat $\underline{x}$ de waarde x aanneemt is gelijk aan de kans op de gebeurtenis A(x) dat het experiment een uitkomst heeft waarvoor geldt $\underline{x}(\omega) = x$. Met andere woorden, $A(x) = \{\omega | \underline{x}(\omega) = x)\}$. Het is duidelijk dat A(x) en A(y) disjunct zijn als $x \neq y$. Ware dat niet het geval dan zou er minstens één $\omega \in A(x) \cap A(y)$ zijn, ofwel minstens één ω waarvoor $\underline{x}(\omega) = x$ en $\underline{x}(\omega) = y$. Dat kan niet.

Voorbeeld

3. Laat in voorbeeld 1 de vier experimenten onafhankelijk zijn en stel voorts $P_i(s) = p$ en $P_i(m) = q = 1 - p$, voor i = 1, 2, 3, 4.
$P(\underline{x} = 0) = p_x\,(0) = P((m, m, m, m)) = q^4$;
$P(\underline{x} = 1) = p_x(1) =$
$= P((m,m,m,s) \cup (m,m,s,m) \cup (m,s,m,m) \cup (s,m,m,m)) = 4pq^3$;
evenzo volgt $p_x(2) = 6p^2q^2$; $p_x(3) = 4p^3q$; $p_x(4) = p^4$; terwijl $p_x(x) = 0$, voor $x \neq 0,1,2,3,4$.
Hier is sprake van een binomiale kansfunctie.

❑

De kansfunctie van een discrete stochastische variabele $\underline{x}$ heeft twee eigenschappen:

$$p_x(x) \geq 0, \text{ voor alle } x \in \mathbb{R};$$

$$\Sigma\, p_x(x_i) = 1, \text{ als de som zich uitstrekt over alle } x_i \in W_x.$$

Het begrip kansfunctie heeft geen nut als de stochastische variabele continu is.

2.2. Verdelingsfunctie en kansdichtheid

Definitie

De verdelingsfunctie van de stochastische variabele $\underline{x}$ is een functie van $\mathbb{R}$ naar $[0,1]$ waarvan de waarde (notatie $P(\underline{x} \leq x)$ of $F_x(x)$) aangeeft hoe groot de kans is dat $\underline{x}$ een waarde $\leq x$ aanneemt.

Een van de voordelen van de verdelingsfunctie ten opzichte van de kansfunctie is dat het begrip verdelingsfunctie niet alleen voor discrete maar ook voor continue stochastische variabelen zinvol is.
Als van de discrete stochastische variabele $\underline{x}$ de kansfunctie $p_x(x)$ bekend is volgt $F_x(x)$ met behulp van

$$F_x(x) = \sum_{x_i \leq x} p_x(x_i).$$

Voor x in het waardenbereik W_x van $\underline{x}$ geldt:

$$F_x(x) = P(\underline{x} < x) + P(\underline{x} = x) = P(\underline{x} < x) + p_x(x)$$

met $p_x(x) > 0$.
Omdat er een gereduceerde omgeving van x bestaat die disjunct is met W_x geldt ook:

$$\lim_{h \downarrow 0} F_x(x-h) = \lim_{h \downarrow 0} P(\underline{x} \leq x - h) = P(\underline{x} < x)$$

en

$$\lim_{h \downarrow 0} F_x(x+h) = \lim_{h \downarrow 0} P(\underline{x} \leq x + h) = P(\underline{x} \leq x) = F_x(x).$$

F_x heeft dus sprongen in de punten $x \in W_x$ ter grootte van $p_x(x)$ en is aldaar continu van rechts.
Tussen twee opeenvolgende sprongpunten is F_x constant. Omdat de waarde van $F_x(x)$ gelijk is aan de kans op de gebeurtenis $\{\omega | \underline{x}(\omega) \leq x\}$ volgt voor F_x:

a. $F_x(x)$ is monotoon niet-dalend;

b. $\lim_{x \to -\infty} F_x(x) = 0$;

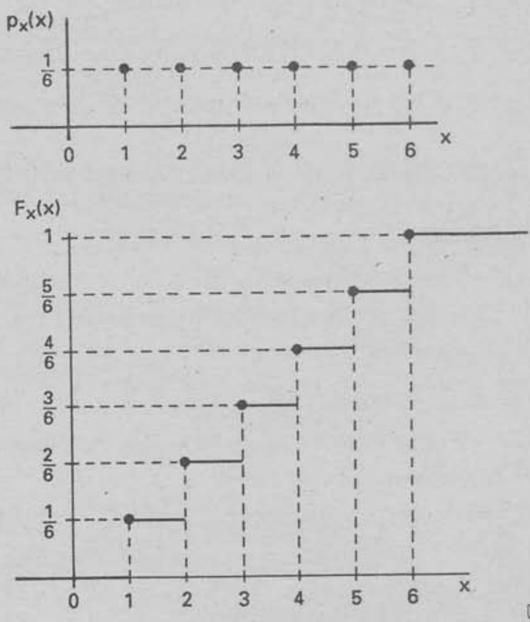
c. $\lim_{x \to \infty} F_x(x) = 1$.

Voorbeeld

De stochastische variabele $\underline{x}$ geeft het aantal ogen aan bij één worp met een zuivere dobbelsteen:

$$
\begin{aligned}
p_{\underline{x}}(x) &= 1/6 && x = 1, 2, \ldots, 6; \\
&= 0 && \text{overal elders.}
\end{aligned}
$$

$$
\begin{aligned}
F_{\underline{x}}(x) &= 0 && x < 1; \\
&= \tfrac{1}{6} && 1 \leq x < 2; \\
&= \tfrac{2}{6} && 2 \leq x < 3; \\
&= \tfrac{3}{6} && 3 \leq x < 4; \\
&= \tfrac{4}{6} && 4 \leq x < 5; \\
&= \tfrac{5}{6} && 5 \leq x < 6; \\
&= 1 && x \geq 6.
\end{aligned}
$$



Figuur 2.1

□

Bij een continue stochastische variabele $\underline{y}$ bestaat het waardenbreik W_y uit een of meer intervallen. De verdelingsfunctie $F_y(y) = P(\underline{y} \leq y)$ is in dit geval overal continu, ook als $y \in W_y$. Overigens heeft F_y dezelfde eigenschappen als de verdelingsfunctie van een discrete stochastische variabele, zoals gemakkelijk is na te gaan.

Voorbeeld

Kies aselect een punt in de cirkel met middelpunt (0, 0) en straal 1. De stochastische variabele $\underline{r}$ geeft de afstand van het middelpunt tot het gekozen punt aan.

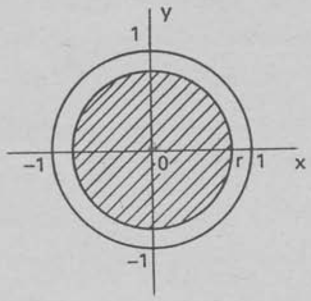
Voor $0 \leq r \leq 1$ geldt:

$$F_r(r) = P(\underline{r} \leq r) = \frac{\text{opp. cirkel met straal } r}{\text{opp. cirkel met straal } 1} = r^2,$$

terwijl $F_r(r) = 0$ voor $r < 0$ en $F_r(r) = 1$ voor $r > 1$.
De kans dat een stochastische variabele een waarde aanneemt in het interval $(y, y+h]$ wordt gegeven door

$$P(y < \underline{y} \leq y + h) = P(\underline{y} \leq y + h) - P(\underline{y} \leq y) = F_y(y + h) - F_y(y).$$

Voor een discrete stochastische variabele is dit gelijk aan nul, tenzij $(y, y + h]$ een of meer elementen van het waardenbereik bevat. Voor een continue stochastische



Figuur 2.2.

variabele kunnen we hiervoor schrijven, met behulp van de middelwaardestelling:

$$h \cdot \frac{dF_y(y+\theta h)}{dy}, 0 < \theta < 1,$$

mits F_y differentieerbaar is op $(y, y+h]$. Er zijn hoogstens aftelbaar veel punten waar F_y niet differentieerbaar is.
Blijkbaar geeft

$$f_y(y) = \lim_{h \downarrow 0} \frac{dF_y(y+\theta h)}{dy},$$

de hellingshoek van de raaklijn in het punt $(y, F_y(y))$ aan de grafiek van F_y, aan hoe snel de kans $P(\underline{y} \leq y)$ toeneemt als y toeneemt. In zekere zin is f_y dus vergelijkbaar met p_x, de kansfunctie van een discrete stochastische variabele, die ook de groei van de kans $P(\underline{x} \leq x)$ weergeeft.
De functie f_y heet de *kansdichtheid* van y. Uit de eigenschappen van $F_y(y)$ en omdat $f_y = F_y'$ volgt nu direct:

a. $\int_{-\infty}^{\infty} f_y(y)dy = 1$, want $F_y(-\infty) = 0$ en $F_y(+\infty) = 1$;

b. $\int_{-\infty}^{y} f_y(t)dt = F_y(y)$;

c. $f_y(y) \geq 0$ want $F_y(y)$ is monotoon niet-dalend;
$f_y(y) = 0$ als $y \notin W_y$.

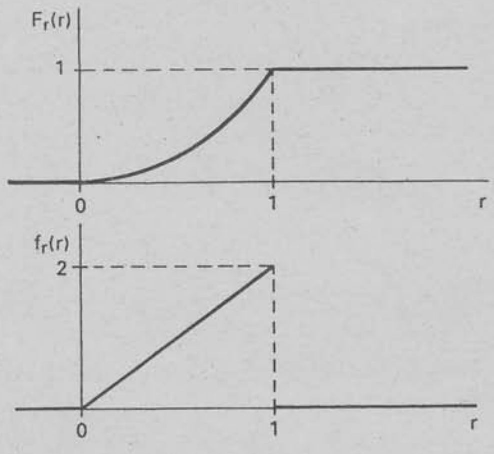
Voorbeeld

In het laatste voorbeeld was

$$\begin{aligned} F_r(r) &= 0, & r < 0; \\ &= r^2, & 0 \leq r \leq 1; \\ &= 1, & r > 1; \end{aligned}$$

zodat de kansdichtheid van $\underline{r}$ wordt:

$$
\begin{aligned}
f_r(r) &= 2r, && 0 < r < 1; \\
&= 0, && \text{elders.}
\end{aligned}
$$



Figuur 2.3.

In de punten waar F_y niet differentieerbaar is kan men f_y definiëren als de rechter- of linker-afgeleide van F_y.
Men kan de kans $P(a < \underline{y} \leq b)$ bepalen met behulp van F_y of met behulp van f_y:

$$P(a < \underline{y} \leq b) = P(\underline{y} \leq b) - P(\underline{y} \leq a) = F_y(b) - F_y(a),$$

of $$P(a < \underline{y} \leq b) = \int_{-\infty}^{b} f_y(y)dy - \int_{\infty}^{a} f_y(y)dy = \int_{a}^{b} f_y(y)dy.$$

De dichtheid wordt dus niet eenduidig bepaald door de verdelingsfunctie; omgekeerd is dat wel het geval.

Elke discrete functie $g(x)$ waarvoor geldt

$$g(x) \geq 0 \quad \text{en} \quad \sum_x g(x) = 1$$

kan als de kansfunctie van een discrete stochastische variabele worden beschouwd.

Elke continue functie $h(x)$ waarvoor geldt

$$h(x) \geq 0 \quad \text{en} \quad \int_{-\infty}^{+\infty} h(x)dx = 1$$

kan als de dichtheid van een continue stochastische variabele worden beschouwd.

Elke functie H(x) waarvoor geldt

a. $H(-\infty) = 0$
b. $H(+\infty) = 1$
c. H is monotoon niet-dalend

kan worden beschouwd als de verdelingsfunctie van een stochastische variabele. Is H continu dan is de bijbehorende stochastische variabele continu. Is H een stapfunctie dan is de bijbehorende stochastische variabele discreet.

2.3. Verwachting en variantie

Een stochastische variabele wordt volledig beschreven als zijn verdelingsfunctie gegeven is of zijn kansfunctie dan wel kansdichtheid. In veel gevallen is er echter behoefte aan een beknopte en daardoor ook meer overzichtelijke karakterisering van een stochastische variabele. Dat kan bijvoorbeeld met behulp van de verwachting $E(\underline{x})$ van de stochastische variabele $\underline{x}$.

Definitie

De verwachting $E(\underline{x})$ wordt gegeven door:

a. $E(\underline{x}) = \sum_{x_i \in W_x} x_i p_x(x_i)$ als $\underline{x}$ discreet is en de som absoluut convergent;

b. $E(\underline{x}) = \int_{-\infty}^{+\infty} x f_x(x) dx$ als $\underline{x}$ continu is en de integraal absoluut convergent.

Als de som of integraal niet absoluut convergent zijn zegt men dat de verwachting niet bestaat. □

Behalve in speciale gevallen zal voortaan worden aangenomen dat de beschouwde verwachtingen bestaan. De verwachting is vergelijkbaar met het gemiddelde. Als u alle loten in de Staatsloterij koopt en dus alle prijzen krijgt kunt u eenvoudig uitrekenen hoeveel u gemiddeld per lot hebt teruggekregen, zeg *f* 15,–. Als u één lot koopt en $\underline{x}$ is de stochastische variabele die aangeeft hoe groot uw prijs is dan is $E(\underline{x})$ ook *f* 15,–.

Voorbeeld

Voor de stochastische variabelen uit de voorbeelden in de vorige paragraaf geldt:

$$E(\underline{x}) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \ldots + 6 \cdot \frac{1}{6} = 3\frac{1}{2};$$

$$E(\underline{r}) = \int_{-\infty}^{\infty} r \cdot f_r(r) dr = \int_0^1 r \cdot 2r dr = \frac{2}{3}.$$

Als men de x-as als een staaf beschouwt die zelf geen gewicht heeft maar die

belast wordt met puntlasten (in het discrete geval) of met een continue belasting (in het continue geval), in beide gevallen de kansverdeling representerend, dan valt $E(\underline{x})$ samen met het zwaartepunt van deze opstelling. Er wordt ook wel over de *kansmassa* gesproken.
Een gebruikelijke notatie voor $E(\underline{x})$ is μ_x, of simpelweg μ als er geen verwarring met andere stochastische variabelen mogelijk is.
Als $\underline{x}$ een stochastische variabele is dan is ook een functie van $\underline{x}$ weer een stochastische variabele. Zij $y = g(x)$ dan zal de stochastische variabele $\underline{y} = g(\underline{x})$ de waarde y toekennen aan alle $\omega \in \Omega$ waaraan de stochastische variabele $\underline{x}$ de waarde x toekent. Natuurlijk is het mogelijk dat bij verschillende waarden x_i, $i = 1,2,\ldots$, dezelfde functiewaarde y behoort zodat

$$P(\underline{y} = y) = P(A)$$

waarin $A = \{\omega|\underline{x}(\omega) \in g^{-1}(y)\} = \bigcup_i \{\omega|\underline{x}(\omega) = x_i;\, g(x_i) = y\}$ een vereniging van disjuncte gebeurtenissen is.
Als $\underline{x}$ een discrete stochastische variabele is dan volgt voor $E(\underline{y}) = E\{g(\underline{x})\}$

$$E(\underline{y}) = \Sigma y_k \cdot P(\underline{y} = y_k)$$

waarbij gesommeerd wordt over alle $y_k \in W_y$. Bekijk één term uit deze som

$$y_k \cdot P(\underline{y} = y_k) = y_k\, \Sigma\, P(\underline{x} = x_i) \text{ gesommeerd over alle } x_i \in g^{-1}(y_k)$$
$$= g(x_i)\, \Sigma\, P(\underline{x} = x_i).$$

Daar $g^{-1}(y_k)$, $k = 1, 2, \ldots$, een partitie vormt van W_x volgt:

$$E(\underline{y}) = \Sigma y_k P(\underline{y} = y_k) = \Sigma g(x_j) P(\underline{x} = x_j),$$

waarbij de laatste sommatie zich uitstrekt voor alle $x_j \in W_x$.
In dit geval is dus

$$E(\underline{y}) = E\{g(\underline{x})\} = \sum_{x_i \in W_x} g(x_i) p_x(x_i).$$

Analoog geldt in het geval dat $\underline{x}$ een continue stochastische variabele is:

$$E(\underline{y}) = E\{g(\underline{x})\} = \int_{-\infty}^{+\infty} g(x) f_x(x)\, dx.$$

De verwachting is gedefinieerd als een som of integraal. Gebruikmakend daarvan is de volgende stelling eenvoudig te bewijzen.

Stelling 2.1. Zij $\underline{x}$ een stochastische variabele en zij c een constante, dan geldt:
Als $P(\underline{x} = c) = 1$ dan $E(\underline{x}) = E(c) = c$;
$E\{cg(\underline{x})\} = cE\{g(\underline{x})\}$;

$E\{g(\underline{x}) + h(\underline{x})\} = E\{g(\underline{x})\} + E\{h(\underline{x})\}$;

Als $a \leq \underline{x} \leq b$ dan $a \leq E(\underline{x}) \leq b$. □

Voorbeeld

Bepaal de verwachting van de grootte van het oppervlak van het vierkant waarvan de lengte in cm van de zijden bepaald wordt door het aantal ogen dat valt bij een worp met een zuivere dobbelsteen.

Zij $\underline{x}$ het aantal ogen; $\underline{y} = \underline{x}^2$ is de grootte in cm^2.

$$E(\underline{y}) = E(\underline{x}^2) = \sum_{i=1}^{6} i^2 \cdot P(\underline{x} = i) = \frac{1}{6}(1 + 4 + 9 + 16 + 25 + 36) = \frac{91}{6} \text{ cm}^2.$$

Idem als de lengte van de zijden gelijk is aan $1 + 5\underline{r}$ waarbij

$$\begin{aligned} f_r(r) &= 2r \quad 0 < r < 1, \\ &= 0 \quad \text{elders.} \end{aligned}$$

$$E(\underline{y}) = E\{(1 + 5\underline{r})^2\} = E(1 + 10\underline{r} + 25\underline{r}^2) = E(1) + 10E(\underline{r}) + 25E(\underline{r}^2) =$$

$$= 1 + 10\int_0^1 r \cdot 2r dr + 25\int_0^1 r^2 \cdot 2r dr = 1 + 10 \cdot \frac{2}{3} + 25 \cdot \frac{1}{2} = \frac{121}{6} \text{ cm}^2.$$

In beide gevallen moet $1 \leq E(\underline{y}) \leq 36$, hetgeen klopt.

Een ander veelgebruikt kental voor een stochastische variabele is de variantie. Dat is een maat voor de fluctuaties in de realisaties van een stochastische variabele.

Definities

- De variantie van de stochastische variabele $\underline{x}$, notatie $\sigma^2(\underline{x})$, is gedefinieerd als

$$\sigma^2(\underline{x}) = E\{(\underline{x} - \mu)^2\}.$$

- De standaard-afwijking, of spreiding, van de stochastische variabele $\underline{x}$, notatie $\sigma(\underline{x})$, is gedefinieerd als

$$\sigma(\underline{x}) = \sqrt{\sigma^2(\underline{x})}.$$

Stelling 2.2. $\sigma^2(\underline{x}) = E(\underline{x}^2) - \mu^2$.

Bewijs: $\sigma^2(\underline{x}) = E\{(\underline{x} - \mu)^2\} = E(\underline{x}^2) - 2\mu E(\underline{x}) + \mu^2 = E(\underline{x}^2) - \mu^2$. □

Voor de berekening is de laatste uitdrukking vaak handiger. Let wel, ook de variantie is een verwachting en hoeft dus niet te bestaan.

Stelling 2.3. Zij $\underline{x}$ een stochastische variabele en zij c een constante dan geldt:

$$\sigma^2(c\underline{x}) = c^2\sigma^2(\underline{x});$$
$$\sigma^2(\underline{x} + c) = \sigma^2(\underline{x});$$
$$\sigma^2(\underline{x}) \geq 0.$$

Als $P(\underline{x} = c) = 1$ dan $\sigma^2(\underline{x}) = \sigma^2(c) = 0.$

Bewijs: Volgt eenvoudig door uitschrijven. □

Voorbeeld

$\underline{x}$ is het aantal ogen dat valt bij een worp met een zuivere dobbelsteen.

$$\sigma^2(\underline{x}) = E(\underline{x}^2) - \{E(\underline{x})\}^2 = \frac{91}{6} - (\frac{6}{2})^2 = \frac{35}{12}.$$

$\underline{r}$ is de afstand tussen een punt dat aselect in de eenheidscirkel wordt gekozen en de oorsprong.

$$\sigma^2(\underline{r}) = \int_0^1 r^2 \cdot 2r dr - \{\int_0^1 r \cdot 2r dr\}^2 = \frac{1}{2} - (\frac{2}{3})^2 = \frac{1}{18}.$$

2.4. Momenten en andere kentallen

Verwachting en variantie zijn de meest gebruikte kentallen voor een stochastische variabele maar in sommige gevallen is er behoefte aan andere waarden.

Definities

- Het k-de moment van de stochastische variabele $\underline{x}$ is $E(\underline{x}^k)$ mits deze verwachting bestaat. Notatie $\mu_k = E(\underline{x}^k)$.
- Het k-de centrale moment van de stochastische variabele $\underline{x}$ is $E\{(\underline{x} - \mu)^k\}$ mits deze verwachting bestaat.

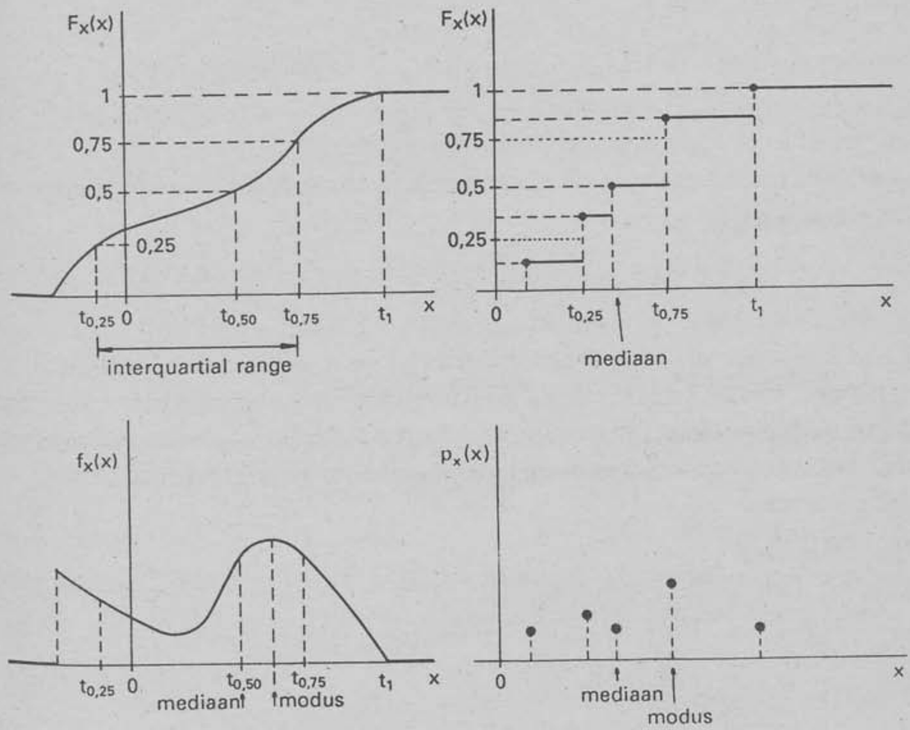
Verwachting en variantie van $\underline{x}$ zijn dus het eerste moment, respectievelijk het tweede centrale moment van $\underline{x}$. Het eerste centrale moment is altijd gelijk aan nul. Als μ_t bestaat dan bestaan alle momenten en centrale momenten voor $0 \leq k \leq t$.

- het 100·k-percentiel van $\underline{x}$ (notatie t_k) is het kleinste getal waarvoor geldt dat de kans dat $\underline{x}$ niet groter is dan dat getal gelijk is aan k, ofwel t_k is het kleinste getal zodanig dat $F_x(t_k) \geq k$.
- De mediaan van $\underline{x}$ is gelijk aan het 50-percentiel, $t_{0\cdot 50}$.

De mediaan is net als de verwachting een kental dat informatie geeft over de plaats van het centrum van de kansmassa.
De mediaan bestaat altijd.
In plaats van de variantie wordt soms $t_{0\cdot 75} - t_{0\cdot 25}$, de interquartiel range, als maat voor de variabiliteit gebruikt.

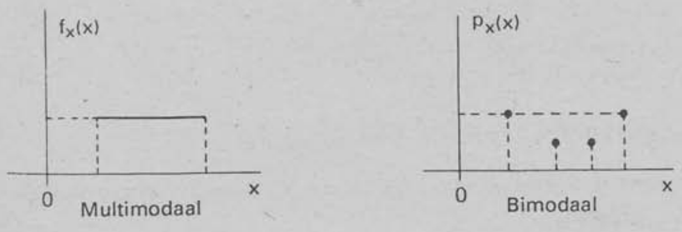
Een enkele keer komt men het begrip modus tegen.

– De modus van $\underline{x}$ is het getal x waarvoor
  $p_x(x)$ maximaal is als $\underline{x}$ een discrete stochastische variabele is;
  $f_x(x)$ maximaal is als $\underline{x}$ een continue stochastische variabele is.



Figuur 2.4.

De modus is niet altijd eenduidig bepaald zoals uit de volgende voorbeelden blijkt.



Figuur 2.5.

2.5. Functies van een stochastische variabele

In paragraaf 2.3 kwamen reeds functies van stochastische variabelen voor maar daar was alleen de verwachting van die functies nodig. Veelal zijn we ook geïnteresseerd in de verdeling van de zo ontstane stochastische variabelen. Zij $\underline{y} = g(\underline{x})$ en de verdelingsfunctie $F_x(x)$ van $\underline{x}$ gegeven. Dan geldt voor de verdelingsfunctie $F_y(y)$ van $\underline{y}$:

$$F_y(y) = P(\underline{y} \leq y) = P(g(\underline{x}) \leq y) = P(\underline{x} \in D(y))$$

waarbij $D(y) \in \mathbb{R}$ de verzameling is van alle $x \in W_x$ waarvoor geldt $g(x) \leq y$. Hiermee kan F_y bepaald worden. Als $\underline{x}$ een discrete stochastische variabele is kan men over het algemeen sneller het gewenste resultaat bereiken door $p_y(y)$ te bepalen uitgaande van $p_x(x)$

$$p_y(y) = P(\underline{y} = y) = P(g(\underline{x}) = y) = P(\underline{x} \in g^{-1}(y)) = \Sigma p_x(x_i)$$

waarbij de sommatie zich uitstrekt over alle $x_i \in g^{-1}(y)$.
Later wordt een methode behandeld waarbij voor een continue stochastische variabele $\underline{x}$ uit $f_x(x)$ rechtstreeks de dichtheid van $\underline{y} = g(\underline{x})$ bepaald kan worden als g een strikt monotone differentieerbare functie is. In het algemeen kan men de dichtheid van $\underline{y}$ bepalen door de op bovenbeschreven wijze gevonden F_y naar y te differentiëren.

Voorbeelden

1. $p_x(x) = 1/6, \quad x = 1, 2, \ldots, 6$
$\qquad\quad = 0, \quad$ elders;

$\underline{y} = -\underline{x}^2 + 4\underline{x} - 3$; gevraagd F_y.

x	1	2	3	4	5	6
y	0	1	0	–3	–8	–15

Dus $p_y(-15) = p_y(-8) = p_y(-3) = p_y(1) = \frac{1}{6}$;

$p_y(0) = p_x(1) + p_x(3) = \frac{1}{3}$;

$p_y(y) = 0$ voor $y \neq -15, -8, -3, 0, 1$.

$F_y(y)$ volgt op de gebruikelijke wijze.

2. $f_x(x) = 1/6, \quad 0 \leq x \leq 6$;
$\qquad\quad = 0, \quad$ elders.
Bepaal de dichtheid van $\underline{y} = -\underline{x}^2 + 4\underline{x} - 3$.
$F_y(y) = P(-\underline{x}^2 + 4\underline{x} - 3 \leq y) = P(\underline{x} \in D(y))$ met $D(y)$ de verzameling van alle $x \in W_x$ waarvoor $-x^2 + 4x - 3 \leq y$. Om $D(y)$ te bepalen lossen we eerst

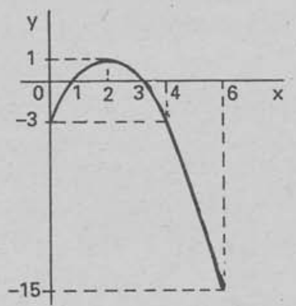
$-x^2 + 4x - 3 - y = 0$ op. Dat levert:

$$x_{1,2} = \frac{-4 \pm \sqrt{16 - 4\cdot(3+y)}}{-2} = 2 \pm \sqrt{1-y}, \quad \text{voor } y \le 1,$$

zodat $D(y) = [0; 2 - \sqrt{1-y}\,] \cup [2 + \sqrt{1-y}; 6]$ voor $-3 \le y \le 1$.
Voor $-15 \le y \le -3$ volgt:

$$D(y) = \Phi \cup [2 + \sqrt{1-y}\ ;\ 6],$$

terwijl voor $y \le -15$: $D(y) = \Phi \cup \Phi = \Phi$. Omdat $-x^2 + 4x - 3 \le 1$ voor alle x krijgen we tenslotte $D(y) = [0; 6]$ voor $y \ge 1$.



Figuur 2.6.

Zo vinden we:

$$\begin{aligned}
F_y(y) &= 0 & &= 0, & &y \le -15;\\
&= \int_{2+\sqrt{1-y}}^{6} f_x(x)dx & &= \tfrac{1}{6}(4 - \sqrt{1-y}), & &-15 < y \le -3;\\
&= \int_{0}^{2-\sqrt{1-y}} f_x(x)dx + \int_{2+\sqrt{1-y}}^{6} f_x(x)dx & &= \tfrac{1}{6}(6 - 2\sqrt{1-y}), & &-3 < y \le 1;\\
&= 1 & &= 1, & &y > 1.
\end{aligned}$$

De gevraagde dichtheid $f_y(y) = \frac{d}{dy} F_y(y)$:

$$\begin{aligned}
f_y(y) &= \frac{1}{12\sqrt{1-y}}, & &-15 < y < -3;\\
&= \frac{1}{6\sqrt{1-y}}, & &-3 < y < 1;\\
&= 0, & &\text{elders.}
\end{aligned}$$

3

Veel voorkomende verdelingen

3.1. Bernoulli- en binomiale verdeling

Definitie

Een Bernoulli-experiment is een experiment met twee mogelijke uitkomsten die in het algemeen 'succes' respectievelijk 'mislukking' worden genoemd.
De uitkomstenruimte Ω geven we aan met $\Omega = \{m, s\}$. Het is gebruikelijk om $P(s) = p$, $0 \leq p \leq 1$, en $P(m) = q = 1 - p$ te stellen.
De discrete stochastische variabele die aan het element s de waarde 1 toekent en aan m de waarde 0 heeft een Bernoulli-verdeling met parameter p.

$p_x(0) = q$; $p_x(1) = p$; $p_x(x) = 0$ als $x \neq 0, 1$.

$$
\begin{aligned}
F_x(x) &= 0, \quad x < 0; \\
&= q, \quad 0 \leq x < 1; \\
&= 1, \quad x \geq 1.
\end{aligned}
$$

$E(\underline{x}) = 0 \cdot q + 1 \cdot p = p$; $\quad \sigma^2(\underline{x}) = E(\underline{x}^2) - p^2 = 0^2 \cdot q + 1^2 \cdot p - p^2 = pq$.

Voer nu een vast aantal Bernoulli-experimenten uit, fysisch onafhankelijk van elkaar en alle met $P(s) = p$. De fysische onafhankelijkheid van de experimenten maakt het mogelijk om als kansruimte voor het totale experiment de produktkansruimte van de deel-experimenten te nemen:

$$\Omega = \Omega_1 \times \Omega_2 \times \ldots \times \Omega_n \text{ met } \Omega_i = \{s, m\} \text{ voor } i = 1, 2, \ldots n.$$

Een element $\omega \in \Omega$ heeft dus de vorm $\omega = (\omega_1, \omega_2, \ldots, \omega_n)$ met $\omega_i \in \Omega_i$ en $P(\omega) = P_1(\omega_1) \cdot P_2(\omega_2) \ldots P_n(\omega_n)$ waarin $P_i(\omega_i) = p$ als $\omega_i = s$ en $P_i(\omega_i) = q$ als $\omega_i = m$, $i = 1, 2, \ldots, n$.
Zo verkrijgt men een binomiale kansruimte. Als nu de discrete stochastische variabele $\underline{x}$ wordt gedefinieerd als

$$\underline{x}(\omega) = \text{aantal successen behorend bij } \omega$$

dan is $\underline{x}$ een binomiaal verdeelde stochastische variabele met parameters n en p.
Notatie: $\underline{x} \sim B(n; p)$.
De kansfunctie van $\underline{x}$ is eenvoudig te bepalen. Voor W_x, het waardenbereik van $\underline{x}$, geldt: $W_x = \{0, 1, 2, \ldots, n\}$. Zij $k \in W_x$ dan volgt

$$p_x(k) = \binom{n}{k} p^k q^{n-k}$$

want er zijn $\binom{n}{k}$ verschillende $\omega = (\omega_1, \omega_2, \ldots, \omega_n)$ met op k posities een s en op de overige positities een m, en elk van deze éénpuntsverzamelingen heeft een kans $p^k q^{n-k}$.

Dus $\quad p_x(k) = \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, 2, \ldots, n;$
$\qquad\qquad = 0, \quad$ elders.

De verdelingsfunctie volgt hieruit

$$F_x(x) = \begin{cases} 0 & x < 0; \\ \sum_{j=0}^{k} \binom{n}{j} p^j q^{n-j}, & k \le x < k+1, \quad k = 0, 1, 2, \ldots, n-1; \\ 1, & x \ge n. \end{cases}$$

Er zijn uitgebreide tabellen van de binomiale verdeling voor verschillende waarden van n en p, echter steeds met $p \le 0{,}5$. Als p, de kans op succes groter dan 0,5 is moet voor gebruik van de tabel de rol van 'succes' en 'mislukking' worden omgedraaid.

Voorbeeld

$\underline{x} \sim B(18; 0{,}6)$. Bepaal $P\{\underline{x} = 12\}$.

Gevraagd wordt de kans op 12 successen in 18 pogingen. Dat is hetzelfde als de kans op 6 mislukkingen in 18 pogingen. De stochastische variabele $\underline{y} = 18 - \underline{x}$ geeft het aantal mislukkingen aan en het is duidelijk dat $\underline{y} \sim B(18; 0{,}4)$. Met behulp van de tabel volgt nu:

$$P(\underline{x} = 12) = P(\underline{y} = 6) = P(\underline{y} \le 6) - P(\underline{y} \le 5) = 0{,}3743 - 0{,}2088 = 0{,}1655.$$

Stelling 3.1. Als $\underline{x} \sim B(n, p)$ dan is $E(\underline{x}) = np$ en $\sigma^2(\underline{x}) = npq$.

Bewijs:
$$E(\underline{x}) = \sum_{k=0}^{n} k \cdot \binom{n}{k} p^k q^{n-k} = \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} p^k q^{n-k} =$$
$$= np \sum_{k-1=0}^{n-1} \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} q^{n-k} = np(p+q)^{n-1} = np.$$

$$E\{\underline{x}(\underline{x}-1)\} = \sum_{k=0}^{n} k(k-1)\binom{n}{k} p^k q^{n-k} = \sum_{k=2}^{n} \frac{n!}{(k-2)!(n-k)!} p^k q^{n-k} =$$
$$= n(n-1)p^2 \sum_{k-2=0}^{n-2} \frac{(n-2)!}{(k-2)!(n-k)!} p^{k-2} q^{n-k} = n(n-1)p^2.$$

$$\sigma^2(\underline{x}) = E(\underline{x}^2) - \{E(\underline{x})\}^2 = E\{\underline{x}(\underline{x}-1)\} + E\underline{x} - \{E(\underline{x})\}^2 =$$
$$= n(n-1)p^2 + np - n^2p^2 = np(1-p) = npq. \qquad \square$$

Als van een stochastische variabele $\underline{x}$ bekend is dat hij binomiaal verdeeld is, en voorts de verwachting en de variantie van $\underline{x}$ gegeven zijn, ligt de verdeling van $\underline{x}$ geheel vast.

Voorbeeld

$\underline{x} \sim B(n; p)$ met $E(\underline{x}) = 10{,}80$ en $\sigma^2(\underline{x}) = 4{,}32$. Gevraagd n en p.

$$\frac{\sigma^2(\underline{x})}{E(\underline{x})} = \frac{npq}{np} = q \text{ dus } q = \frac{4{,}32}{10{,}80} = 0{,}4 \text{ ofwel } p = 0{,}6.$$

$$E(\underline{x}) = np \text{ dus } n = \frac{E(\underline{x})}{p} = \frac{10{,}80}{0{,}6} = 18.$$

3.2. Geometrische en hypergeometrische verdeling

Men voert onafhankelijke Bernoulli-experimenten uit, alle met kans p op succes, net zolang tot het eerste succes is opgetreden.

Zij $\underline{x}$ de discrete stochastische variabele die het aantal experimenten aangeeft, dan is $\underline{x}$ geometrisch verdeeld met parameter p. Men noemt $\underline{x}$ ook wel Pascal-verdeeld.

Het waardenbereik van $\underline{x}$ is: 1,2,3,... .

Als de eerste proef een succes oplevert is $\underline{x} = 1$. Dus $P(\underline{x} = 1) = p$. De stochastische variabele $\underline{x}$ neemt de waarde n aan, n = 2,3,..., als na n – 1 mislukkingen een succes optreedt, dus $P(\underline{x} = n) = q^{n-1}p$.

De kansverdeling van $\underline{x}$:

$$\begin{aligned} p_x(n) &= q^{n-1}p, && n = 1, 2, 3, \ldots; \\ &= 0, && \text{elders.} \end{aligned}$$

De verdelingsfunctie van $\underline{x}$:

$$\begin{aligned} F_x(x) &= \sum_{k=1}^{n} q^{k-1}p = 1 - q^n, && n \le x < n+1, n = 1, 2, \ldots \\ &= 0, && x < 1. \end{aligned}$$

De verwachting van $\underline{x}$:

$$E(\underline{x}) = \sum_{n=1}^{\infty} nq^{n-1}p = \frac{1}{p}.$$

De variantie van $\underline{x}$:

$$\sigma^2(\underline{x}) = \frac{q}{p^2}\,.$$

Voor het bepalen van $E(\underline{x})$ en $\sigma^2(\underline{x})$ is gebruik gemaakt van de volgende methode, die toelaatbaar is als $0 < z < 1$:

$$\sum_{n=1}^{\infty} nz^{n-1}(1-z) = (1-z)\sum_{n=0}^{\infty}\frac{d}{dz}z^n = (1-z)\frac{d}{dz}\sum_{n=0}^{\infty}z^n = (1-z)\frac{d}{dz}\left(\frac{1}{1-z}\right) =$$

$$= (1-z)\cdot\frac{1}{(1-z)^2} = \frac{1}{1-z}\,.$$

Substitutie van $z = q$ levert $E(\underline{x})$.

$$\sum_{n=1}^{\infty} n(n-1)z^{n-1}(1-z) = z(1-z)\sum_{n=0}^{\infty} n(n-1)z^{n-2} =$$

$$= z(1-z)\frac{d^2}{dz^2}\left(\frac{1}{1-z}\right) = \frac{2z(1-z)}{(1-z)^3} = \frac{2z}{(1-z)^2}\,.$$

$$\sigma^2(\underline{x}) = E(\underline{x})^2 - \{E(\underline{x})\}^2 = E\{\underline{x}(\underline{x}-1)\} + E(\underline{x}) - \{E(\underline{x})\}^2 =$$

$$= \sum_{n=1}^{\infty} n(n-1)q^{n-1}\cdot p + \frac{1}{p} - \frac{1}{p^2} = \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{q}{p^2}\,.$$

Voorbeeld

Jan en Piet gooien om de beurt met een zuivere dobbelsteen net zolang tot er een 6 valt. Jan begint, hoe groot is de kans dat hij ook als laatste gooit? Het aantal worpen $\underline{x}$ is geometrisch verdeeld met $p = \frac{1}{6}$. Jan gooit als laatste (gebeurtenis J) als $\underline{x} = 1, 3, 5, \ldots$

$$P(J) = \sum_{k=0}^{\infty} P(\underline{x} = 2k+1) = \frac{1}{6}\cdot\sum_{k=0}^{\infty}(5/6)^{2k} = \frac{1}{6}\cdot\frac{1}{1-25/36} = \frac{6}{11}\,.$$ □

Uit een vaas die N ballen bevat, waarvan er Z zwart zijn, worden aselect en zonder teruglegging n ballen getrokken. De discrete stochastische variabele $\underline{z}$, die het aantal getrokken zwarte ballen aangeeft, heeft een hypergeometrische verdeling.

Neem als uitkomstenruimte Ω de verzameling van alle $\binom{N}{n}$ combinaties van n ballen uit de N. Er wordt aselect getrokken, we mogen dus een symmetrische kansruimte gebruiken. De stochastische variabele $\underline{z}$ neemt de waarde k aan voor elke $\omega \in \Omega$ die behoort bij een trekking van k uit de Z zwarte en $n-k$ uit de $N-Z$ niet-zwarte ballen. Omdat niet op de volgorde van trekken gelet wordt is dit aantal gelijk aan $\binom{Z}{k}\binom{N-Z}{n-k}$.

Omdat we er niet meer uit kunnen trekken dan er in zit, moeten er grenzen worden gesteld aan k; dat gaat het eenvoudigst met de gebruikelijke conventie $\binom{a}{b} = 0$ als $b > a$. Zo krijgt men:

$$p_{\underline{z}}(k) = \frac{\binom{Z}{k}\binom{N-Z}{n-k}}{\binom{N}{n}}, \quad k = 0, 1, 2, \dots, n, \; n \leq N;$$

$$= 0, \qquad \text{elders.}$$

$$E(\underline{z}) = n\frac{Z}{N}; \quad \sigma^2(\underline{z}) = n\cdot\frac{Z}{N}\cdot\frac{N-Z}{N}\cdot\frac{N-n}{N-1}.$$

$E(\underline{z})$ vindt men door gebruik te maken van $\binom{a}{b} = \frac{a}{b}\cdot\binom{a-1}{b-1}$.

$$E(\underline{z}) = \sum_{k=0}^{n} k p_{\underline{z}}(k) = \sum_{k=1}^{n} k\frac{\binom{Z}{k}\binom{N-Z}{n-k}}{\binom{N}{n}} = \sum_{k=1}^{n} k\frac{\frac{Z}{k}\binom{Z-1}{k-1}\binom{N-Z}{n-k}}{\frac{N}{n}\binom{N-1}{n-1}} =$$

$$= n\cdot\frac{Z}{N}\sum_{k=1}^{n}\frac{\binom{Z-1}{k-1}\binom{N-Z}{n-k}}{\binom{N-1}{n-1}}$$

en deze laatste som is gelijk aan één.
Op overeenkomstige wijze vindt men de uitdrukking voor $\sigma^2(\underline{z})$.

Voorbeeld

In een doos zitten 10 lampen waarvan er 4 kapot zijn. Men pakt aselect 3 lampen uit de doos. Gevraagd de kans op de gebeurtenis A dat men zo minstens twee goede lampen krijgt.
Zij $\underline{z}$ het aantal goede lampen onder de 3 getrokken exemplaren, dan is $P(A) = p_{\underline{z}}(2) + p_{\underline{z}}(3)$.

$$p_{\underline{z}}(2) = \frac{\binom{6}{2}\binom{4}{1}}{\binom{10}{3}} = \frac{15\cdot4}{120} = \frac{1}{2},$$

$$p_{\underline{z}}(3) = \frac{\binom{6}{3}\binom{4}{0}}{\binom{10}{3}} = \frac{20\cdot1}{120} = \frac{1}{6},$$

zodat $\quad P(A) = \frac{2}{3}$. □

Als $N \gg n$ dan maakt het nauwelijks verschil of het trekken met of zonder teruglegging gebeurt. In dat geval kan de kansverdeling van $\underline{z}$ benaderd worden door een binomiale verdeling met parameters n en $p = \frac{Z}{N}$.

3.3. Poisson-verdeling

De discrete stochastische variabele $\underline{x}$ heeft een Poisson-verdeling met parameter $\lambda (> 0)$ als:

$$p_x(k) = \frac{\lambda^k}{k!}\, e^{-\lambda}, \quad k = 0, 1, 2, \ldots;$$

$$= 0; \qquad \text{elders.}$$

Hieruit volgt dat $E(\underline{x}) = \lambda$ en $\sigma^2(\underline{x}) = \lambda$, immers

$$E(\underline{x}) = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} \cdot e^{\lambda} = \lambda;$$

$$E(\underline{x}^2) = E\{\underline{x}(\underline{x}-1)\} + E(\underline{x}) = \sum_{k=0}^{\infty} k(k-1) \frac{\lambda^k}{k!} e^{-\lambda} + \lambda = \lambda^2 + \lambda$$

zodat

$$\sigma^2(\underline{x}) = E(\underline{x}^2) - \{E(\underline{x})\}^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

De Poisson-verdeling kan worden gebruikt als benadering van de binomiale verdeling $B(n; p)$ als n groot is en p klein (ruwweg $np < 5$ en $p < 0{,}1$).
Zij $\underline{x}_n \sim B(n; p)$, laat $n \to \infty$ en $p \to 0$, zó dat $np = \lambda$ constant is.

$$\lim_{n\to\infty} P(\underline{x}_n = k) = \lim_{n\to\infty} \frac{n!}{k!(n-k)!} p^k q^{n-k} = \frac{1}{k!} \lim_{n\to\infty} \frac{n!}{(n-k)!} \left(\frac{\lambda}{n}\right)^k \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{n-k} =$$

$$= \frac{\lambda^k}{k!} \cdot \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{\frac{n}{\lambda} \cdot \lambda} \cdot \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{-k} = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \ldots .$$

Voorbeeld

Op een kruispunt passeert gemiddeld één auto per 2 minuten. Veronderstel dat de auto's zich onafhankelijk van elkaar gedragen en dat de wegen zó smal zijn dat er per seconde hoogstens één auto het kruispunt kan passeren.
Om een indruk te krijgen van de grootte van de kans op de gebeurtenis A dat er in een bepaald interval van twee minuten géén auto passeert kunnen we dat interval verdelen in 120 opeenvolgende intervallen ter lengte van één seconde. Elk interval beschouwen we als een Bernoulli-experiment. Passeert er in die seconde een auto dan spreken we van een succes, zo niet dan van een mislukking. Zij $\underline{x}$ het aantal auto's dat passeert in de beschouwde twee minuten dan is $\underline{x} \sim B(120; \frac{1}{120})$, $E(\underline{x}) = 1$.

$$P(A) \approx P(\underline{x} = 0) = \binom{120}{0} \cdot \left(\frac{1}{120}\right)^0 \left(\frac{119}{120}\right)^{120} = 0{,}3663.$$

We hadden de 2 minuten ook in 1200 intervallen van 0,1 seconde kunnen verdelen en een bijbehorende stochastische variabele $\underline{y} \sim B(1200; 1/1200)$ introduceren.
Dan wordt

$$P(A) \approx P(\underline{y}=0) = \binom{1200}{0}\left(\frac{1}{1200}\right)^0 \left(\frac{1199}{1200}\right)^{1200} = 0{,}3677.$$

Gebruikmakend van een Poisson-verdeelde stochastische variabele $\underline{z}$ met $\lambda = 1$ krijgen we

$$P(A) \approx P(\underline{z}=0) = e^{-1} = 0{,}3679.$$

3.4. Uniforme en exponentiële verdeling

Een stochastische variabele $\underline{x}$ die de waarden $a + ih$, $i = 1, 2, 3, \ldots, n$, kan aannemen, elk met een kans $1/n$ heeft een discrete uniforme verdeling. Veelal is $a = 0$ en $h = 1$ en ook als dat niet zo is kan men de verdeling altijd terugbrengen tot een verdeling op $1, 2, 3, \ldots, n$ met behulp van $\underline{y} = (\underline{x} - a)/h$, ofwel $\underline{x} = \underline{y} \cdot h + a$.
Voor $\underline{y}$ geldt dan:

$$\begin{aligned} p_y(k) &= \frac{1}{n}, \quad && k = 1, 2, \ldots, n; \\ &= 0, && \text{elders.} \end{aligned}$$

$$E(\underline{y}) = \frac{1}{2}(n+1); \quad \sigma^2(y) = \frac{1}{12}(n^2 - 1).$$

Voorbeeld

Het aantal ogen bij een worp met een zuivere dobbelsteen, $\underline{z}$, is uniform verdeeld met $a = 0$, $h = 1$ en $n = 6$, dus

$$E(\underline{z}) = 3\tfrac{1}{2} \text{ en } \sigma^2(\underline{z}) = \frac{35}{12},$$

zoals reeds bekend was. Als er gesproken wordt van een uniform verdeelde stochastische variabele wordt echter meestal een continue stochastische variabele bedoeld.
De continue stochastische variabele $\underline{x}$ is uniform verdeeld op het interval (a, b) als (a, b) het waardenbereik is van $\underline{x}$ en de kansdichtheid f_x op (a, b) constant is.
Omdat $\int_{-\infty}^{+\infty} f_x(x)dx = 1$ en $f_x(x) = 0$ voor $x \notin W_x$ moet $\int_a^b c\, dx = 1$ ofwel $c = \frac{1}{b-a}$.

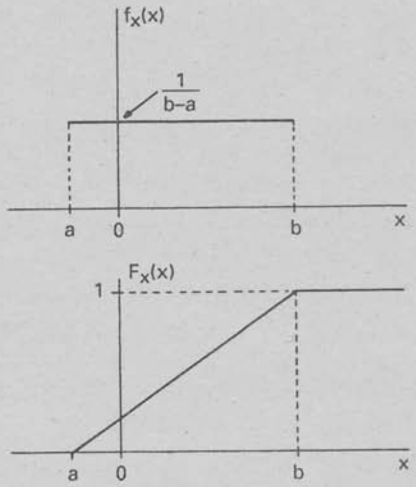
Daarmee is de verdeling van $\underline{x}$ geheel gespecificeerd:

$$\begin{aligned} f_x(x) &= \frac{1}{b-a}, \quad && a < x < b; \\ &= 0, && \text{elders,} \end{aligned}$$

of

$$
\begin{aligned}
F_x(x) &= 0, && x \le a;\\
&= \frac{x-a}{b-a}, && a < x < b;\\
&= 1, && x \ge b.
\end{aligned}
$$

$$E(\underline{x}) = \frac{1}{2}(a+b); \quad \sigma^2(\underline{x}) = \frac{1}{12}(b-a)^2.$$



Figuur 3.1.

De verwachting en de variantie volgen uit:

$$E(\underline{x}) = \int_a^b \frac{x}{b-a}\, dx = \frac{1}{2}\,\frac{b^2-a^2}{b-a} = \frac{1}{2}(a+b)$$

$$E(\underline{x}^2) = \int_a^b \frac{x^2}{b-a}\, dx = \frac{1}{3}\,\frac{b^3-a^3}{b-a} = \frac{1}{3}(b^2+ab+a^2)$$

$$\sigma^2(\underline{x}) = \frac{1}{3}(b^2+ab+a^2) - \frac{1}{4}(b^2+2ab+a^2) = \frac{1}{12}(b^2-2ab+a^2).$$

Als $a = 0$ en $b = 1$ spreekt men van de standaard-uniforme verdeling. Als $\underline{y}$ standaard uniform verdeeld is en $\underline{x}$ is uniform verdeeld op (a, b) dan geldt:

$$\underline{y} = \frac{\underline{x}-a}{b-a} \quad \text{of} \quad \underline{x} = (b-a)\underline{y} + a.$$

Voorbeeld

Men knipt een stuk touw met een lengte van 1 meter op een willekeurige plaats door. Gevraagd de verwachting van de lengte van het langste van de twee stukken touw die men zo verkrijgt.

Er is altijd één stuk korter dan 50 cm en één stuk dat langer dan 50 cm is. Zij $\underline{x}$ de lengte van het grootste stuk in cm, dan is W_x het interval (50, 100). Alle punten in dit interval zijn even waarschijnlijk, dus $\underline{x}$ is uniform verdeeld.

$$\begin{aligned} f_x(x) &= \frac{1}{50}, \quad 50 < x < 100; \\ &= 0, \quad \text{elders,} \end{aligned}$$

zodat $E(\underline{x}) = 75$. De gevraagde verwachting is 75 cm. □

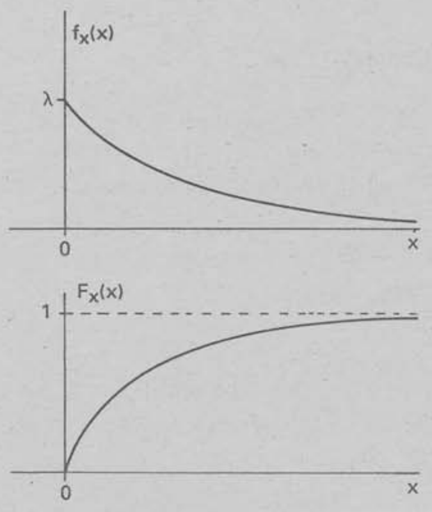
Een eveneens veel voorkomende continue verdeling is de (negatief) exponentiële verdeling.
De stochastische variabele $\underline{x}$ is *exponentieel verdeeld* met parameter λ (> 0) als

$$\begin{aligned} f_x(x) &= \lambda e^{-\lambda x}, \quad x \geq 0; \\ &= 0, \quad\quad x < 0. \end{aligned}$$

Hieruit volgt dat

$$\begin{aligned} F_x(x) &= 1 - e^{-\lambda x}, \quad x \geq 0; \\ &= 0, \quad\quad x < 0 \end{aligned}$$

en $$E(\underline{x}) = \frac{1}{\lambda}, \quad \sigma^2(\underline{x}) = \frac{1}{\lambda^2}.$$



Figuur 3.2.

Een exponentieel verdeelde stochastische variabele is bij uitstek geschikt om de levensduur aan te geven van niet aan mechanische slijtage onderhevige apparaten, of de lengte van een telefoongesprek.

Voorbeeld

Bij een opknapbeurt worden in een instructiezaal 100 nieuwe TL-buizen geïnstalleerd. Elke buis heeft een exponentieel verdeelde levensduur met verwachting 3 jaar. Neem aan dat de levensduren onafhankelijk van elkaar zijn. Bereken de kans op de gebeurtenis A dat na 1,5 jaar alle buizen nog werken.
Zij voor i = 1,2,3,...,100, $\underline{x}_i$ de levensduur van buis i, in jaren. $\underline{x}_i$ is exponentieel verdeeld met parameter $\lambda = 1/3$:

$$P(\underline{x}_i > 1\tfrac{1}{2}) = 1 - F_{\underline{x}_i}(1\tfrac{1}{2}) = e^{-\frac{1}{3}\cdot\frac{3}{2}} = e^{-\frac{1}{2}},\ i = 1, 2, 3, \ldots, 100.$$

Elke buis kan men nu beschouwen als een Bernoulli-experiment. De twee uitkomsten:

$$\text{succes} = \text{buis werkt na } 1\tfrac{1}{2} \text{ jaar: } P(s) = e^{-\frac{1}{2}} = 0{,}6065;$$

$$\text{mislukking} = \text{buis is kapot na } 1\tfrac{1}{2} \text{ jaar: } P(m) = 1 - e^{-\frac{1}{2}} = 0{,}3935.$$

Het aantal buizen dat na 1,5 jaar nog goed is, $\underline{z}$, is dus binomiaal verdeeld met parameters n = 100 en $p = e^{-\frac{1}{2}}$. Daarmee volgt:

$$P(A) = P(\underline{z} = 100) = p^{100} = e^{-50}.$$

3.5. Normale verdeling

Een continue stochastische variabele die Normaal verdeeld is heeft een op het eerste gezicht wat ingewikkelde dichtheid. Toch heeft deze verdeling veel goede eigenschappen. Statistiek bedrijven zonder de Normale verdeling te gebruiken is vrijwel onmogelijk.
De stochastische variabele $\underline{x}$ is Normaal verdeeld met parameters μ en σ^2, notatie $\underline{x} \sim N(\mu;\sigma^2)$, als voor de dichtheid geldt:

$$f_{\underline{x}}(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2},\ x \in \mathbb{R};\ \sigma > 0.$$

De verdelingsfunctie van $\underline{x}$:

$$F_{\underline{x}}(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt,\ x \in \mathbb{R},$$

kan alleen numeriek bepaald worden zodat we ons moeten behelpen met een tabel.
Voorts is

$$E(\underline{x}) = \mu \text{ en } \sigma^2(\underline{x}) = \sigma^2.$$

De gegeven functie f_x is zeker positief voor alle $x \in \mathbb{R}$. Dat ook voldaan is aan $\int_{-\infty}^{+\infty} f_x(x)dx = 1$ blijkt uit het volgende. Zij $I = \int_{-\infty}^{+\infty} e^{-\frac{1}{2}x^2}\,dx$ dan is

$$I^2 = \int_{-\infty}^{+\infty} e^{-\frac{1}{2}x^2}\,dx . \int_{-\infty}^{+\infty} e^{-\frac{1}{2}y^2}\,dy = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-\frac{1}{2}(x^2+y^2)}\,dxdy.$$

Overgaan op poolcoördinaten levert

$$I^2 = \int_{r=0}^{\infty}\int_{\varphi=0}^{2\pi} e^{-\frac{1}{2}r^2}\,rd\varphi dr = 2\pi \int_{0}^{+\infty} e^{-\frac{1}{2}r^2}\,d\tfrac{1}{2}r^2 = 2\pi.$$

Omdat $e^{-\frac{1}{2}x^2} \geq 0$ moet $I \geq 0$, dus $I = \sqrt{2\pi}$.

$$\int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}\,dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}\,d\frac{x-\mu}{\sigma} = \frac{1}{\sqrt{2\pi}}\cdot I = 1.$$

De verwachting volgt uit:

$$E(\underline{x}) = \int_{-\infty}^{+\infty} x f_x(x)dx = \int_{-\infty}^{+\infty} (x-\mu)f_x(x)dx + \mu \int_{-\infty}^{+\infty} f_x(x)dx =$$

$$= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \frac{x-\mu}{\sigma} \cdot e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}\,d\frac{x-\mu}{\sigma} + \mu\cdot 1 = \mu.$$

want $\int_{-\infty}^{+\infty} te^{-\frac{1}{2}t^2}\,dt = 0$

omdat de integrand een oneven functie is en de integraal absoluut convergent is.
Het bepalen van de variantie gaat als volgt:

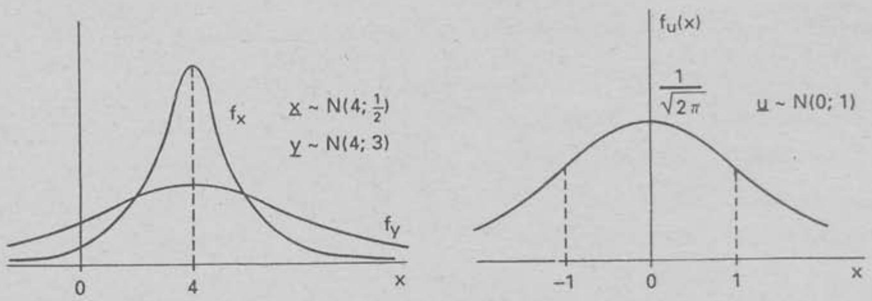
$$\sigma^2(\underline{x}) = \int_{-\infty}^{+\infty} (x-\mu)^2 \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}\,dx = \frac{2\sigma^2}{\sqrt{2\pi}} \int_{0}^{\infty} t^2 e^{-\frac{1}{2}t^2}\,dt =$$

$$= \frac{2\sigma^2}{\sqrt{\pi}} \int_{0}^{\infty} \frac{t}{\sqrt{2}}\, e^{-\frac{1}{2}t^2}\,d\frac{1}{2}t^2 = \frac{2\sigma^2}{\sqrt{\pi}}\,\Gamma(\tfrac{3}{2}) = \sigma^2,$$

want

$$\Gamma(\tfrac{3}{2}) = \tfrac{1}{2}\Gamma(\tfrac{1}{2}) = \tfrac{1}{2}\sqrt{\pi}.$$

In figuur 3.3 zijn enige dichtheden van Normaal-verdeelde stochastische variabelen geschetst. De buigpunten van de kromme liggen bij $\mu \pm \sigma$.
Een Normaal-verdeelde stochastische variabele met $\mu = 0$ en $\sigma^2 = 1$ heet *standaard-Normaal-verdeeld*. Alleen voor een dergelijke stochastische variabele,



Figuur 3.3.

$\underline{u} \sim N(0; 1)$, zijn de waarden van $1 - F_u(x)$ in de tabel opgenomen en dan nog alleen voor $x \geq 0$.

Voorbeeld

$\underline{u} \sim N(0; 1)$. Bepaal $P(\underline{u} > 2{,}25)$; $P(\underline{u} > -2{,}25)$; $P(-1{,}5 < \underline{u} < 0{,}5)$.

Zij $P(\underline{u} > a) = 0{,}6$, bepaal a.

$P(\underline{u} > 2{,}25) = 0{,}0122$.
$P(\underline{u} > -2{,}25) \overset{1}{=} P(\underline{u} < 2{,}25) = 1 - P(\underline{u} \geq 2{,}25) = 1 - P(\underline{u} > 2{,}25) = 0{,}9878$,
het gelijkteken 1 is een gevolg van de symmetrie ten opzichte van x = 0 van de dichtheid van $\underline{u}$.

$$\begin{aligned} P(-1{,}5 < \underline{u} < 0{,}5) &= P(\underline{u} < 0{,}5) - P(\underline{u} < -1{,}5) = 1 - P(\underline{u} \geq 0{,}5) - P(\underline{u} > 1{,}5) = \\ &= 1 - 0{,}3085 - 0{,}0668 = 0{,}6247. \end{aligned}$$

$P(\underline{u} > a) = 0{,}6$ dus $P(\underline{u} \leq a) = 1 - 0{,}6 = 0{,}4$. Met behulp van de symmetrie volgt dan $P(\underline{u} \geq -a) = 0{,}4$.
In de tabel vinden we: $P(\underline{u} > 0{,}25) = 0{,}4013$; $P(\underline{u} > 0{,}26) = 0{,}3974$ zodat lineaire interpolatie $P(\underline{u} > 0{,}253) = 0{,}4$ oplevert.
Dus $-a = 0{,}253$ ofwel $a = -0{,}253$. □

Met behulp van de volgende stelling is het mogelijk om de tabel van de standaard-Normale-verdeling ook te gebruiken voor het bepalen van kansen die horen bij andere Normaal-verdeelde stochastische variabelen.

Stelling 3.4. Zij $\underline{x} \sim N(\mu; \sigma^2)$ dan is $\frac{\underline{x}-\mu}{\sigma} \sim N(0;1)$.

Bewijs: $P(\frac{\underline{x}-\mu}{\sigma} \leq c) = P(\underline{x} \leq c\sigma + \mu) = \int_{-\infty}^{c\sigma+\mu} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx$. In de integraal passen we de substitutie $y = \frac{x-\mu}{\sigma}$ toe. De grens $x = -\infty$ wordt $y = -\infty$. De bovengrens $x = c\sigma + \mu$ wordt $y = c$, zodat

$$P(\frac{\underline{x}-\mu}{\sigma} \leq c) = \int_{-\infty}^{c} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy = P(\underline{u} \leq c) \text{ met } \underline{u} \sim N(0;1). \qquad \square$$

Op dezelfde manier kan worden bewezen dat als $\underline{x} \sim N(\mu;\sigma^2)$ volgt dat $\underline{y} = a\underline{x} + b$ Normaal-verdeeld is met als parameters

$$E(\underline{y}) = aE(\underline{x}) + b = a\mu + b$$

en

$$\sigma^2(\underline{y}) = a^2\sigma^2(\underline{x}) = a^2\sigma^2.$$

Voorbeeld

$\underline{x} \sim N(3;16)$. Bepaal $P(|\underline{x}| \leq 5)$.

$$\begin{aligned} P(|\underline{x}| \leq 5) &= P(-5 \leq \underline{x} \leq 5) = P(\underline{x} \leq 5) - P(\underline{x} \leq -5) = \\ &= P(\frac{\underline{x}-3}{4} \leq \frac{5-3}{4}) - P(\frac{\underline{x}-3}{4} \leq \frac{-5-3}{4}) = \\ &= P(\underline{u} \leq 0{,}5) - P(\underline{u} \leq -2) = \qquad \text{met } \underline{u} \sim N(0;1) \\ &= 1 - P(\underline{u} > 0{,}5) - P(\underline{u} \geq 2) = 1 - 0{,}3085 - 0{,}0228 = 0{,}6687. \end{aligned}$$

4

Simultane verdelingen

4.1. Twee-dimensionale verdelingen

Soms is het nodig om twee of meer stochastische variabelen op de uitkomstenruimte van een experiment te definiëren om voldoende informatie omtrent de afloop van dat experiment met behulp van stochastische variabelen te geven. Helaas is het dan niet voldoende om van elk van die variabelen apart een verdeling te geven zoals uit het volgende voorbeeld blijkt.

Voorbeeld

Twee heren, R en S, beiden even goed in het schaken, zullen door middel van een wedstrijd uitmaken wie zich kampioen mag noemen. Afgesproken wordt dat dat degene is die het eerst 2 partijen wint. Als remises buiten beschouwing blijven bestaat de uitkomstenruimte Ω uit 6 elementen (zie hieronder), waarbij r(s) correspondeert met een door R(S) gewonnen partij.
Zij $\underline{x}(\underline{y})$ het aantal door R(S) gewonnen partijen dan kan men het volgende schema maken:

Ω	(r,r)	(r,s,r)	(s,r,r)	(s,s)	(s,r,s)	(r,s,s)
$P(\{\omega\})$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$
$\underline{x}(\omega)$	2	2	2	0	1	1
$\underline{y}(\omega)$	0	1	1	2	2	2

$$
\begin{aligned}
\text{Zodat } p_x(k) &= \tfrac{1}{4}, \quad k=0, 1;\\
&= \tfrac{1}{2}, \quad k=2;\\
&= 0, \quad \text{elders;}\\
\text{en ook } p_y(k) &= \tfrac{1}{4}, \quad k=0, 1;\\
&= \tfrac{1}{2}, \quad k=2;\\
&= 0, \quad \text{elders.}
\end{aligned}
$$

Als de informatie omtrent het experiment alleen bestaat uit p_x en p_y dan is niet te

16

zien dat de combinaties $\underline{x} = 0$ en $\underline{y} = 0$ of 1, $\underline{x} = 1$ en $\underline{y} = 0$ of 1, $\underline{x} = 2$ en $\underline{y} = 2$ niet kunnen voorkomen.
Bij het bepalen van de kansmaat is onafhankelijkheid tussen de partijen verondersteld. □

Het is nodig om in één functie van Ω naar $\mathbb{R}^n$ de *samengestelde* of *simultane kansverdeling* vast te leggen. Voorlopig zullen we ons tot twee stochastische variabelen (een twee-dimensionale stochastische vector) beperken waarbij de kansverdeling dus wordt gegeven door een functie van Ω naar $\mathbb{R}^2$.

Definitie

De simultane verdelingsfunctie $F_{x,y}$ van de stochastische variabelen $\underline{x}$ en $\underline{y}$ is een functie van $\mathbb{R}^2$ naar [0,1] zodanig dat

$$F_{x,y}(x,y) = P(\underline{x} \le x \cap \underline{y} \le y). \qquad \square$$

Het is duidelijk dat $F_{x,y}$ monotoon niet-dalend is in x en in y. Voorts

$$F_{x,y}(-\infty,y) = P(\underline{x} \le -\infty \cap \underline{y} \le y) = P(\Phi \cap \underline{y} \le y) = 0$$

en natuurlijk ook $F_{x,y}(x,-\infty) = 0$.

Uit $\qquad F_{x,y}(\infty,y) = P(\underline{x} \le \infty \cap \underline{y} \le y) = P(\Omega \cap \underline{y} \le y) = P(\underline{y} \le y) = F_y(y)$

blijkt dat uit de simultane verdelingsfunctie van $\underline{x}$ en $\underline{y}$ de verdelingsfunctie van $\underline{y}$ alleen, en evenzo die van $\underline{x}$ alleen kan worden afgeleid. In deze meerdimensionale context heten F_x en F_y marginale verdelingsfuncties.

Definitie

De simultane kansfunctie $p_{x,y}$ van de discrete stochastische variabelen $\underline{x}$ en $\underline{y}$ is een functie van $\mathbb{R}^2$ naar [0,1] zodanig dat

$$p_{x,y}(x,y) = P(\underline{x} = x \cap \underline{y} = y).$$

Uit de simultane kansfunctie $p_{x,y}$ kan ook weer de marginale kansfunctie p_x van $\underline{x}$ of p_y van $\underline{y}$ worden bepaald.

$$\sum_y p_{x,y}(x,y) = \sum_y P(\underline{x} = x \cap \underline{y} = y) = P(\underline{x} = x \cap \bigcup_y \underline{y} = y) =$$

$$= P(\underline{x} = x \cap \Omega) = p_x(x),$$

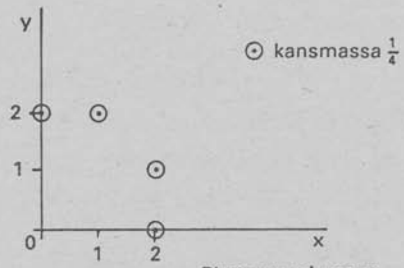
waarbij $\sum\limits_y$ en $\bigcup\limits_y$ zich uitstrekken over alle $y \in W_y$.
Er bestaat natuurlijk ook een verband tussen $F_{x,y}$ en $p_{x,y}$:

$$F_{x,y}(x,y) = \sum_{x_i \leq x} \sum_{y_j \leq y} p_{x,y}(x_i, y_j).$$

Voorbeeld

In het voorgaande voorbeeld wordt de simultane kansfunctie:

$$\begin{aligned} p_{x,y}(x,y) &= \tfrac{1}{4} \qquad (x = 0 \text{ of } 1 \text{ en } y = 2) \text{ of } (x = 2 \text{ en } y = 0 \text{ of } 1); \\ &= 0 \qquad \text{elders.} \end{aligned}$$



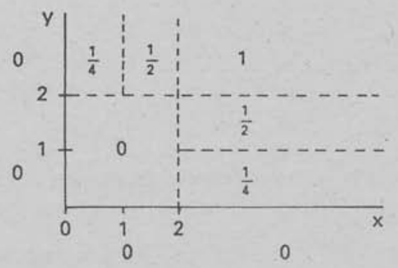
Figuur 4.1. Plattegrond van $p_{x,y}$.

$$F_{x,y}(1\tfrac{1}{2}, 2\tfrac{1}{2}) = P(\underline{x} \leq 1\tfrac{1}{2} \cap \underline{y} \leq 2\tfrac{1}{2}) = P(\underline{x} = 0 \cap \underline{y} = 2) +$$

$$+ P(\underline{x} = 1 \cap \underline{y} = 2) = \tfrac{1}{2};$$

$$F_{x,y}(x,y) = 0 \text{ voor } x < 0 \text{ of } y < 0.$$

De verdere waarden van $F_{x,y}$ staan in figuur 4.2.



Figuur 4.2.

Uit $p_{x,y}(x,y)$ kunnen ook p_x en p_y weer verkregen worden.

$$p_x(0) = \sum_y p_{x,y}(0,y) = p_{x,y}(0,2) = \tfrac{1}{4};$$

$$p_x(1) = \sum_y p_{x,y}(1,y) = p_{x,y}(1,2) = \tfrac{1}{4};$$

$$p_x(2) = \sum_y p_{x,y}(2,y) = p_{x,y}(2,0) + p_{x,y}(2,1) = \frac{1}{2};$$

$$p_x(x) = 0, \text{ voor } x \neq 0, 1, 2.$$

Op dezelfde manier volgt:

$$\begin{aligned} p_y(y) &= \tfrac{1}{4}, && y = 0, 1; \\ &= \tfrac{1}{2}, && y = 2; \\ &= 0, && \text{elders.} \end{aligned}$$

In meer-dimensionale situaties is het specificeren van de verdelingsfunctie $F_{x,y}$ meestal erg omslachtig zodat de voorkeur wordt gegeven aan het vastleggen van de kansverdeling door middel van de simultane kansfunctie (of kansdichtheid in het continue geval).
Als de componenten $\underline{x}$ en $\underline{y}$ van de stochastische vector $(\underline{x},\underline{y})$ continue stochastische variabelen zijn zal het meestal mogelijk zijn om $F_{x,y}$ te schrijven als de dubbelintegraal van een functie van x en y, de simultane dichtheid.

Definitie

De simultane dichtheid $f_{x,y}$ van de continue stochastische variabelen $\underline{x}$ en $\underline{y}$, zo hij bestaat, is een functie van $\mathbb{R}^2$ naar $[0,\infty]$ zodanig dat

$$F_{x,y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{x,y}(u,v)dvdu. \qquad \square$$

Dat er niet altijd een simultane dichtheid $f_{x,y}$ bestaat blijkt uit het volgende voorbeeld.

Voorbeeld

In het interval [0,1] wordt een willekeurig punt t gekozen. $\underline{x}$ is de afstand van 0 tot t, $\underline{y}$ is de afstand van t tot 1. $\underline{x}$ en $\underline{y}$ zijn beide continu en uniform verdeeld op [0,1]. $(\underline{x},\underline{y})$ als stochastische vector kan alleen waarden aannemen op de lijn $x + y = 1$ in het eerste kwadrant, de kansmassa is geheel geconcentreerd op dat lijnstuk. Er is geen simultane dichtheid. □

Een functie f(x,y) kan alleen dan een simultane dichtheid zijn als

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y)dxdy = 1$$

en

$$f(x,y) \geq 0 \text{ voor alle x en y.}$$

Als $f_{\underline{x},\underline{y}}$ de simultane dichtheid is van $\underline{x}$ en $\underline{y}$ dan is

$$f_{\underline{x},\underline{y}}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{\underline{x},\underline{y}}(x,y).$$

De marginale verdelingsfunctie $F_{\underline{x}}(x) = F_{\underline{x},\underline{y}}(x,\infty)$ dus

$$F_{\underline{x}}(x) = \int_{-\infty}^{x} \left\{ \int_{-\infty}^{+\infty} f_{\underline{x},\underline{y}}(u,y)dy \right\} du = \int_{-\infty}^{x} g(u)du,$$

waarin $g(u) = \int_{-\infty}^{+\infty} f_{\underline{x},\underline{y}}(u,y)dy.$

De (marginale) kansdichtheid van $\underline{x}$ volgt dus door de (marginale) verdelingsfunctie van $\underline{x}$ te differentiëren naar x of de simultane dichtheid te integreren naar y:

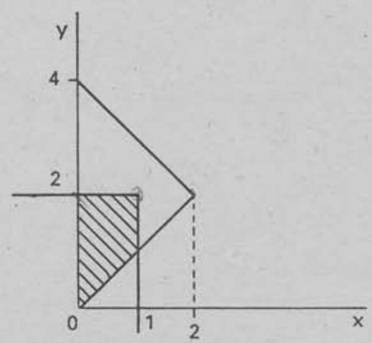
$$f_{\underline{x}}(x) = \frac{d}{dx} F_{\underline{x}}(x) \equiv \frac{d}{dx} \int_{-\infty}^{x} g(u)du = g(x) = \int_{-\infty}^{+\infty} f_{\underline{x},\underline{y}}(x,y)dy.$$

Op dezelfde manier volgt

$$f_{\underline{y}}(y) = \int_{-\infty}^{+\infty} f_{\underline{x},\underline{y}}(x,y)dx = \frac{d}{dy} F_{\underline{y}}(y).$$

Voorbeeld

Kies een willekeurig punt in de driehoek met hoekpunten (0,0), (2,2) en (0,4). De coördinaten van het punt geven we aan met de stochastische variabelen $\underline{x}$ en $\underline{y}$. Bepaal $f_{\underline{x},\underline{y}}$, $F_{\underline{x},\underline{y}}$, $f_{\underline{x}}$, $f_{\underline{y}}$, $F_{\underline{x}}$ en $F_{\underline{y}}$.



Figuur 4.3.

Buiten de driehoek is $f_{\underline{x},\underline{y}}(x,y)$ overal gelijk aan nul, de kansmassa strekt zich alleen uit over de driehoek. Binnen de driehoek is $f_{\underline{x},\underline{y}}(x,y) = c$ want alle punten daar zijn 'gelijkwaardig'. Omdat

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{x,y}(x,y)dxdy = 1,$$

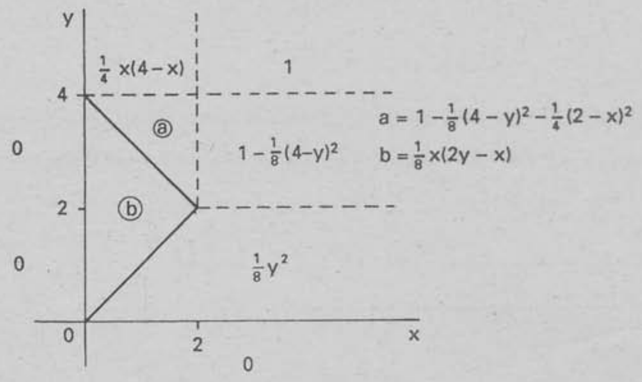
volgt: c^{-1} is gelijk aan het oppervlak van de driehoek, met andere woorden $c = \frac{1}{4}$.

$$\begin{aligned} f_{x,y}(x,y) &= \tfrac{1}{4}, \quad (x,y) \text{ in de driehoek;} \\ &= 0, \quad \text{elders.} \end{aligned}$$

$$F_{x,y}(1,2) = \int_{x=-\infty}^{1}\int_{y=-\infty}^{2} f_{x,y}(x,y)dydx = \int_{x=0}^{1}\int_{y=x}^{2} \frac{1}{4}dydx =$$

$$= \int_0^1 \frac{1}{4}(2-x)dx = \frac{1}{2} - \frac{1}{8} = \frac{3}{8}.$$

Omdat de dichtheid constant is binnen de driehoek is $F_{x,y}(1,2)$ natuurlijk gelijk aan het oppervlak van het gearceerde deel gedeeld door het oppervlak van de hele driehoek.
Zo vindt men ook $F_{x,y}(1,4) = \frac{3}{4}$, enz. De specificatie van $F_{x,y}$ staat in figuur 4.4.



Figuur 4.4.

De marginale dichtheid van $\underline{x}$ is

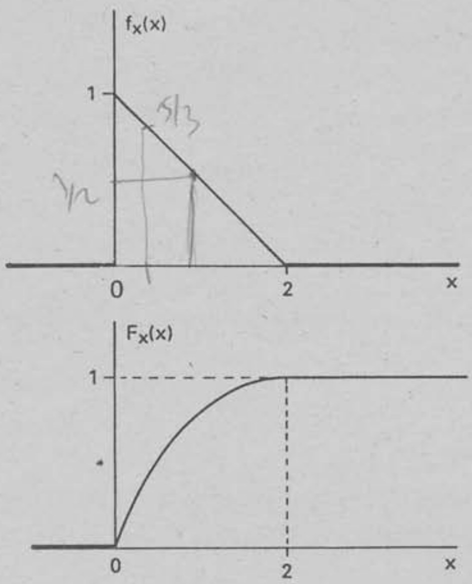
$$f_x(x) = \int_{y=-\infty}^{+\infty} f_{x,y}(x,y)dy,$$

dus voor $x < 0$ en voor $x > 2$ is $f_x(x)$ gelijk aan nul.

Voor $0 < x < 2$ volgt

$$f_x(x) = \int_{y=x}^{4-x} \frac{1}{4}dy = \frac{1}{2}(2-x);$$

zodat



Figuur 4.5.

$$f_x(x) = \frac{1}{2}(2 - x), \quad 0 < x < 2;$$

x = 1 f(x) = 1/2 x = 1/3

$$= 0, \quad \text{elders.}$$

De marginale verdelingsfunctie van $\underline{x}$ wordt dus

$$\begin{aligned} F_x(x) &= 0, & x < 0; \\ &= \frac{1}{4}x(4 - x), & 0 \le x \le 2; \\ &= 1, & x > 2. \end{aligned}$$

De marginale dichtheid van $\underline{y}$ is nul voor $y < 0$ en voor $y > 4$. Voor $0 < y \le 2$ volgt:

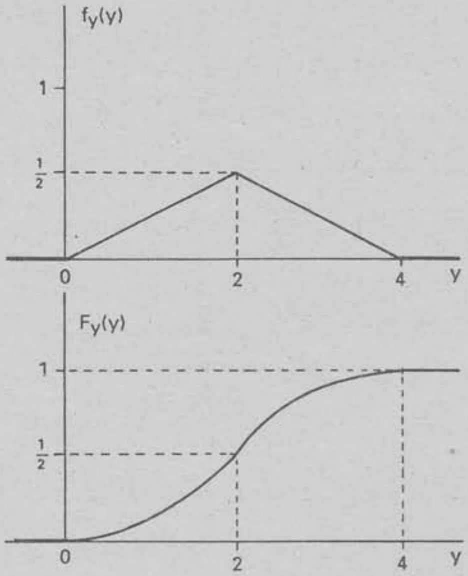
$$f_y(y) = \int_{x=0}^{y} \frac{1}{4}dx = \frac{1}{4}y.$$

Voor $2 < y < 4$:

$$f_y(y) = \int_{x=0}^{4-y} \frac{1}{4}dx = \frac{1}{4}(4 - y)$$

zodat

$$\begin{aligned} f_y(y) &= \frac{1}{4}y, & 0 < y \le 2; \\ &= \frac{1}{4}(4 - y), & 2 < y < 4; \\ &= 0, & \text{elders.} \end{aligned}$$



Figuur 4.6.

De marginale verdelingsfunctie van $\underline{y}$ wordt:

$$
\begin{aligned}
F_y(y) &= 0, && y < 0;\\
&= \tfrac{1}{8}\,y^2, && 0 \le y < 2;\\
&= 1 - \tfrac{1}{8}\,(4-y)^2, && 2 \le y < 4;\\
&= 1, && y \ge 4.
\end{aligned}
$$
□

Het is ook mogelijk dat van een stochastische vector component $\underline{x}$ discreet is, terwijl $\underline{y}$ continu is.

4.2. Verwachting, variantie, covariantie en correlatiecoëfficiënt

Meer nog dan in het ééndimensionale geval hebben we behoefte aan kentallen om op eenvoudige wijze het gedrag van de stochastische vector globaal te kunnen beschrijven. Die kentallen kunnen weer verkregen worden met behulp van het begrip verwachting.
Zij $g(\underline{x},\underline{y})$ een functie van de stochastische variabelen $\underline{x}$ en $\underline{y}$. Bij elke mogelijke realisatie (x,y) van $(\underline{x},\underline{y})$ hoort dan het getal $g(x,y)$. In het discrete geval treedt dat getal $g(x,y)$ op met kans $p_{x,y}(x,y)$. In het continue geval hoort bij dat getal de waarde $f_{x,y}(x,y)$ van de simultane dichtheid. De situatie verschilt dus niet wezenlijk van die in het ééndimensionale geval.

Definitie

De verwachting van de functie $g(\underline{x},\underline{y})$ van de stochastische variabelen $\underline{x}$ en $\underline{y}$ is gelijk aan

$$E\{g(\underline{x},\underline{y})\} = \sum_{x} \sum_{y} g(x,y)p_{x,y}(x,y)$$

als $\underline{x}$ en $\underline{y}$ een simultane kansfunctie hebben en

$$E\{g(\underline{x},\underline{y})\} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x,y)\ f_{x,y}(x,y)dxdy$$

als $\underline{x}$ en $\underline{y}$ een simultane dichtheid hebben.
Zij $g(\underline{x},\underline{y}) = \underline{x}$ dan volgt

$$\mu_x = E(\underline{x}) = \int_{x=-\infty}^{+\infty} \int_{y=-\infty}^{+\infty} x\ f_{x,y}(x,y)\ dydx = \int_{x=-\infty}^{+\infty} x\{\int_{y=-\infty}^{+\infty} f_{x,y}(x,y)\ dy\}dx =$$

$$= \int_{-\infty}^{+\infty} x\ f_x(x)\ dx$$

in het continue geval. Voor het discrete geval volgt op dezelfde manier

$$\mu_x = E(\underline{x}) = \sum_{x} x\ p_x(x).$$

Voor elke functie $g(\underline{x},\underline{y})$ die uitsluitend van $\underline{x}$ of uitsluitend van $\underline{y}$ afhangt kan de verwachting worden bepaald door gebruik te maken van de marginale kansverdeling. In het bijzonder volgt zo dat de variantie van $\underline{x}$

$$\sigma^2(\underline{x}) = E\{(\underline{x} - \mu_x)^2\} = E(\underline{x}^2) - \mu_x^2$$

uit de marginale kansverdeling van $\underline{x}$ te berekenen valt. Evenzo is voor het bepalen van de verwachting van functies, die te schrijven zijn als een lineaire combinatie van functies van één van de twee variabelen, alleen kennis van de marginale verdelingen nodig. Immers

$$E\{ag(\underline{x}) + bh(\underline{y})\} = \iint\{ag(x) + bh(y)\}\ f_{x,y}(x,y)dxdy =$$

$$= a\iint g(x)f_{x,y}(x,y)dxdy + b\iint h(y)f_{x,y}(x,y)dxdy =$$

$$= aE\{g(\underline{x})\} + bE\{h(\underline{y})\},$$

analoog in het discrete geval.
De verwachting van het produkt van $\underline{x}$ en $\underline{y}$ is echter in het algemeen alleen met behulp van de simultane verdeling te bepalen.

$$E(\underline{x}\underline{y}) = \sum_x \sum_y xy\, p_{x,y}(x,y) \qquad \text{(discreet)},$$

of
$$E(\underline{x}\underline{y}) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xy\, f_{x,y}(x,y)dxdy \qquad \text{(continu)}.$$

In het ééndimensionale geval wordt niet het tweede moment maar de variantie als kental gebruikt omdat bij de variantie de keuze van de oorsprong geen invloed heeft. In het twee-dimensionale geval wordt om dezelfde reden niet $E(\underline{x}\underline{y})$ maar $\text{cov}(\underline{x},\underline{y}) = E\{(\underline{x} - \mu_x)(\underline{y} - \mu_y)\}$, de covariantie van $\underline{x}$ en $\underline{y}$, als kental gebruikt. Voor het berekenen is de volgende uitdrukking meestal prettiger:

$$\begin{aligned}\text{cov}(\underline{x},\underline{y}) &= E(\underline{x}\underline{y} - \mu_x\underline{y} - \underline{x}\mu_y + \mu_x\mu_y) = E(\underline{x}\underline{y}) - \mu_x E(\underline{y}) - E(\underline{x})\mu_y \\ &\quad + \mu_x\mu_y = E(\underline{x}\underline{y}) - E(\underline{x})E(\underline{y}).\end{aligned}$$

Een veel gebruikte notatie voor variantie en covariantie is:

$$\sigma^2(\underline{x}) = \sigma_{x,x};$$

$$\text{cov}(\underline{x},\underline{y}) = \sigma_{x,y};$$

$$\sigma^2(\underline{y}) = \sigma_{y,y}.$$

De covariantie van $\underline{x}$ en $\underline{y}$ is een maat voor de samenhang tussen $\underline{x}$ en $\underline{y}$. Een nadeel is dat de waarde van de covariantie verandert bij een verandering van de schaal waarin de stochastische variabelen worden gemeten. Daarom geeft men er de voorkeur aan te werken met de correlatiecoëfficiënt $\rho(\underline{x},\underline{y})$,

$$\rho(\underline{x},\underline{y}) = \frac{\text{cov}(\underline{x},\underline{y})}{\sigma(\underline{x})\sigma(\underline{y})},$$

waarvan de waarde onafhankelijk is van de keuze van de oorsprong en de schaal. Voor de covariantie en de correlatiecoëfficënt gelden de volgende rekenregels die door uitschrijven eenvoudig zijn te controleren.

Stelling 4.1.
$$\text{cov}(\underline{x},\underline{x}) = \sigma^2(\underline{x}); \qquad \rho(\underline{x},\underline{x}) = 1;$$
$$\text{cov}(a\underline{x},b\underline{y}) = ab\,\text{cov}(\underline{x},\underline{y}); \qquad \rho(a\underline{x},b\underline{y}) = \rho(\underline{x},\underline{y});$$
$$\text{cov}(\underline{x},c) = 0; \qquad \rho(\underline{x},c) = 0.$$
$$\text{cov}(\sum_i a_i\underline{x}_i, \sum_j b_j\underline{y}_j) = \sum_i \sum_j a_i b_j\,\text{cov}(\underline{x}_i,\underline{y}_j). \qquad \square$$

Voorbeeld

$$p_{x,y}(x,y) = \frac{1}{4}, \quad (x = 0 \text{ of } 1, y = 2) \text{ of } (x = 2, y = 0 \text{ of } 1);$$

$$= 0, \quad \text{elders} \qquad \text{(zie figuur 4.1).}$$

$$E(\underline{x}) = \sum x p_x(x) = 0 \cdot \tfrac{1}{4} + 1 \cdot \tfrac{1}{4} + 2 \cdot \tfrac{2}{4} = \tfrac{5}{4} = E(\underline{y});$$

$$E(\underline{x}^2) = \sum x^2 p_x(x) = 0 \cdot \tfrac{1}{4} + 1 \cdot \tfrac{1}{4} + 4 \cdot \tfrac{2}{4} = \tfrac{9}{4} = E(\underline{y}^2);$$

$$\sigma^2(\underline{x}) = \tfrac{9}{4} - (\tfrac{5}{4})^2 = \tfrac{11}{16} = \sigma^2(y); \quad \sigma(\underline{x}) = \sigma(\underline{y}) = \tfrac{1}{4}\sqrt{11};$$

$$E(\underline{x}\underline{y}) = \sum\sum xy\, p_{x,y}(x,y) = 1\cdot 2 \cdot \tfrac{1}{4} + 2\cdot 1 \cdot \tfrac{1}{4} = 1;$$

$$\mathrm{cov}(\underline{x},\underline{y}) = 1 - \tfrac{5}{4} \cdot \tfrac{5}{4} = -\tfrac{9}{16};$$

$$\rho(\underline{x},\underline{y}) = \frac{-9/16}{(\tfrac{1}{4}\sqrt{11})^2} = -\frac{9}{11}.$$

Voorbeeld

$f_{x,y}(x,y) = \tfrac{1}{4}$, voor (x,y) in de driehoek met hoekpunten (0,0), (2,2) en (0,4);

$= 0$, elders (zie figuur 4.3).

$$E(\underline{x}) = \int_{x=0}^{2} x \cdot \int_{y=x}^{4-x} \tfrac{1}{4}\, dy\cdot dx = \int_0^2 x \cdot \tfrac{1}{2}(2-x)dx = 2 - \tfrac{4}{3} = \tfrac{2}{3};$$

$$E(\underline{y}) = \int_{y=0}^{2} y \int_{x=0}^{y} \tfrac{1}{4}\, dx\cdot dy + \int_{y=2}^{4} y \int_{x=0}^{4-y} \tfrac{1}{4}\, dx\cdot dy =$$

$$= \int_0^2 \tfrac{1}{4}\, y^2 dy + \int_2^4 \tfrac{1}{4}\, y(4-y)dy = = \tfrac{2}{3} + 6 - 4\tfrac{2}{3} = 2;$$

$$E(\underline{x}^2) = \int_0^2 x^2 \cdot \tfrac{1}{2}(2-x)dx = \tfrac{2}{3}; \quad \sigma^2(\underline{x}) = \tfrac{2}{3} - (\tfrac{2}{3})^2 = \tfrac{2}{9};$$

$$E(\underline{y}^2) = \int_0^2 y^2 \tfrac{1}{4}\, y\, dy + \int_2^4 y^2 \tfrac{1}{4}(4-y)dy = 1 + \tfrac{56}{3} - \tfrac{240}{16} = 4\tfrac{2}{3};$$

$$\sigma^2(\underline{y}) = 4\tfrac{2}{3} - 4 = \tfrac{2}{3};$$

$$E(\underline{x}\underline{y}) = \int_{x=0}^{2} \int_{y=x}^{4-x} xy \frac{1}{4}\, dydx = \int_{0}^{2} \frac{1}{4} x \cdot \frac{1}{2} \{(4-x)^2 - x^2\} dx =$$

$$= \int_{0}^{2} x(2-x)dx = 1\tfrac{1}{3};$$

$$\mathrm{cov}(\underline{x},\underline{y}) = 1\tfrac{1}{3} - \tfrac{2}{3} \cdot 2 = 0;$$

$$\rho(\underline{x},\underline{y}) = \frac{\mathrm{cov}(\underline{x},\underline{y})}{\sqrt{2/9 \cdot 2/3}} = 0. \qquad \square$$

Stelling 4.2. $\sigma^2(a\underline{x} + b\underline{y}) = a^2\sigma^2(\underline{x}) + 2ab\,\mathrm{cov}(\underline{x},\underline{y}) + b^2\sigma^2(\underline{y})$.

Bewijs: $\sigma^2(a\underline{x} + b\underline{y}) = E\{(a\underline{x} + b\underline{y})^2\} - \{E(a\underline{x} + b\underline{y})\}^2 =$

$$= a^2E(\underline{x}^2) + 2abE(\underline{x}\underline{y}) + b^2E(\underline{y})^2 - a^2\{E(\underline{x})\}^2 - 2abE(\underline{x})E(\underline{y}) - b^2\{E(\underline{y})\}^2 =$$

$$= a^2\sigma^2(\underline{x}) + 2ab\,\mathrm{cov}(\underline{x},\underline{y}) + b^2\sigma^2(\underline{y}). \qquad \square$$

Stelling 4.3. $|\rho(\underline{x},\underline{y})| \leq 1$.

Bewijs: $a\underline{x} + \underline{y}$ is een stochastische variabele dus $\sigma^2(a\underline{x} + \underline{y}) \geq 0$, voor alle $a \in \mathbb{R}$.

$$\sigma^2(a\underline{x} + \underline{y}) = a^2\sigma^2(\underline{x}) + 2a\mathrm{cov}(\underline{x},\underline{y}) + \sigma^2(\underline{y}) = g(a).$$

Omdat deze kwadratische vorm in a niet-negatief is mag $g(a) = 0$ hoogstens één reële wortel hebben. Dat wil zeggen, de discriminant $D \leq 0$, ofwel

$$\{2\,\mathrm{cov}(\underline{x},\underline{y})\}^2 \leq 4\sigma^2(\underline{x})\sigma^2(\underline{y}),$$

waaruit het gestelde volgt. $\square$

Stelling 4.4. Als $\sigma^2(\underline{z}) = 0$ dan is $P(\underline{z} = \mu_z) = 1$.

Bewijs: Zij $\underline{z}$ een discrete stochastische variabele dan is

$$\sigma^2(\underline{z}) = \sum_z (z - \mu_z)^2 P(\underline{z} = z)$$

een som van niet-negatieve termen. $\sigma^2(\underline{z})$ kan dus alleen nul zijn als elke term van de som gelijk aan nul is. $P(\underline{z} = z)$ moet dus nul zijn voor alle $z \neq \mu_z$. Omdat

$$\sum_z P(\underline{z} = z) = 1$$

volgt dat $P(\underline{z} = \mu_z) = 1$.
Zou $\underline{z}$ een continue stochastische variabele zijn dan volgt uit

$$\sigma^2(\underline{z}) = \int_{-\infty}^{+\infty} (z-\mu_z)^2 f_z(z)dz = 0$$

dat $f_z(z) = 0$ voor alle $z \neq \mu_z$. De hele kansmassa is dus weer geconcentreerd in $z = \mu_z$, m.a.w. $\underline{z}$ is geen continue stochastische variabele. □

Stelling 4.5. $\rho(\underline{x},\underline{y}) = \pm 1 \Leftrightarrow P(a\underline{x} + \underline{y} = c) = 1$.
Bewijs: $\sigma^2(\lambda\underline{x} + \underline{y}) = \lambda^2\sigma^2(\underline{x}) + 2\lambda\text{cov}(\underline{x},\underline{y}) + \sigma^2(\underline{y}) = g(\lambda)$.
Als $\rho(\underline{x},\underline{y}) = \pm 1$ dan is $\text{cov}(\underline{x},\underline{y}) = \pm\,\sigma(\underline{x})\sigma(\underline{y})$ zodat $g(\lambda)$ een nulpunt heeft voor $\lambda = a$, (met $|a| = \sigma(\underline{y})/\sigma(\underline{x})$) zodat:

$$\sigma^2(a\underline{x} + \underline{y}) = 0 \text{ dus } P(a\underline{x} + \underline{y} = a\mu_x + \mu_y) = 1.$$

Omgekeerd: bij het bepalen van kansverdelingen, en dus van verwachtingen hebben gebeurtenissen met de kans nul geen invloed zodat we kunnen doen alsof $a\underline{x} + \underline{y} = c$ voor alle $\omega \in \Omega$, ofwel $\underline{y} = c - a\underline{x}$.

$$\text{cov}(\underline{x},\underline{y}) = \text{cov}(\underline{x}, c - a\underline{x}) = -a\sigma^2(\underline{x});$$

$$\sigma^2(\underline{y}) = \sigma^2(c - a\underline{x}) = a^2\sigma^2(\underline{x});$$

$$\rho(\underline{x},\underline{y}) = \frac{-a\sigma^2(\underline{x})}{\sigma(\underline{x})\cdot|a|\sigma(\underline{x})} = \frac{-a}{|a|} = \pm 1.$$

Als $\rho(\underline{x},\underline{y}) = \pm 1$ is de hele kansmassa geconcentreerd op een lijn. De hellingshoek van die lijn is positief (negatief) als $\rho = 1$ ($\rho = -1$). □

4.3. Onafhankelijke stochastische variabelen; voorwaardelijke verdelingen

Twee stochastische variabelen $\underline{x}$ en $\underline{y}$ zijn onafhankelijk van elkaar als elke gebeurtenis A die kan worden beschreven met de stochastische variabele $\underline{x}$ onafhankelijk is van elke gebeurtenis B die kan worden beschreven met de stochastische variabele $\underline{y}$. De volgende definitie is equivalent met deze uitspraak.

Definitie

De stochastische variabelen $\underline{x}$ en $\underline{y}$ zijn onafhankelijk dan en slechts dan als $F_{x,y}(x,y) = F_x(x)\cdot F_y(y)$ voor alle x en y.

Stelling 4.6. De stochastische variabelen $\underline{x}$ en $\underline{y}$ zijn onafhankelijk dan en slechts dan als

dan als	$f_{x,y}(x,y) = f_x(x)\cdot f_y(y)$	(continu)	voor alle x en y waarvoor $f_{x,y}(x,y)$ continu is.
	$p_{x,y}(x,y) = p_x(x)\cdot p_y(y)$	(discreet)	voor alle x en y.

Bewijs: Stel $\underline{x}$ en $\underline{y}$ onafhankelijk (en continu) dan is

$$f_{x,y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{x,y}(x,y) = \frac{\partial}{\partial x} F_x(x) \cdot f_y(y) = f_x(x) \cdot f_y(y).$$

Voor het discrete geval volgt het bewijs in dezelfde trant.

Omgekeerd
Stel $p_{x,y}(x,y) = p_x(x) \cdot p_y(y)$.

$$F_{x,y}(x,y) = \sum_{x_i \le x} \sum_{y_j \le y} p_{x,y}(x_i,y_j) = \sum_{x_i \le x} \sum_{y_j \le y} p_x(x_i) p_y(y_j) =$$

$$= \sum_{x_i \le x} p_x(x_i) \cdot \sum_{y_j \le y} p_y(y_j) = F_x(x) \cdot F_y(y).$$

In het continue geval verloopt het bewijs analoog. □

Een noodzakelijke, maar zeker niet voldoende, voorwaarde voor onafhankelijkheid is dat het waardenbereik $W_{x,y}$ van het paar $(\underline{x},\underline{y})$ geschreven kan worden als het produkt $W_x \times W_y$ van de afzonderlijke waardenbereiken.
Als $\underline{x}$ en $\underline{y}$ onafhankelijk zijn dan zijn ook $\underline{z} = h(\underline{x})$ en $\underline{w} = g(\underline{y})$ onafhankelijk van elkaar.

Stelling 4.7. Als $\underline{x}$ en $\underline{y}$ onafhankelijk zijn dan geldt

$$E\{h(\underline{x}) \cdot g(\underline{y})\} = E\{h(\underline{x})\} \cdot E\{g(\underline{y})\}.$$

Bewijs: In het continue geval:

$$E\{h(\underline{x}) \cdot g(\underline{y})\} = \int_x \int_y h(x) g(y) f_{x,y}(x,y)\, dxdy =$$

$$= \int_x h(x) f_x(x) dx \cdot \int_y g(y) f_y(y) dy = E\{h(\underline{x})\} \cdot E\{g(\underline{y})\}.$$

Voor $\underline{x}$ en $\underline{y}$ discreet verloopt het bewijs analoog.
In het bijzonder is dus $E(\underline{x}\underline{y}) = E(\underline{x}) \cdot E(\underline{y})$ als $\underline{x}$ en $\underline{y}$ onafhankelijk zijn en dus de $\mathrm{cov}(\underline{x},\underline{y}) = 0$, waaruit volgt dat $\rho(\underline{x},\underline{y}) = 0$.
Ook hier geldt het omgekeerde in het algemeen niet. □

Voorbeelden

- $(\underline{x},\underline{y})$ is uniform verdeeld op $(0,1) \times (0, \frac{1}{2})$.

$$f_{x,y}(x,y) = 2, \quad 0 < x < 1, 0 < y < \tfrac{1}{2};$$
$$= 0, \quad \text{elders.}$$

$$f_x(x) = 1, \quad 0 < x < 1; \qquad f_y(y) = 2, \quad 0 < y < \tfrac{1}{2};$$
$$= 0, \quad \text{elders.} \qquad = 0, \quad \text{elders.}$$

$f_{x,y}(x,y) = f_x(x) \cdot f_y(y)$ voor alle x en y dus $\underline{x}$ en $\underline{y}$ zijn onafhankelijk.

• $(\underline{x},\underline{y})$ is uniform verdeeld op de driehoek (0,0), (1,0), (0,1).

$\underline{x}$ en $\underline{y}$ kunnen niet onafhankelijk zijn want $W_{x,y} \neq W_x \times W_y$.

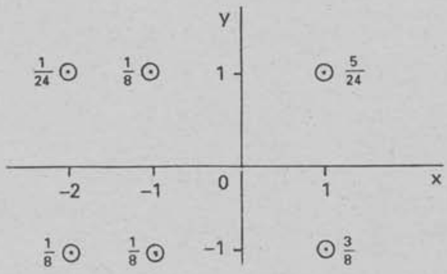
• $\underline{x}$ en $\underline{y}$ zijn onafhankelijk en beide exponentieel verdeeld met parameter $\lambda = 1$.

De simultane dichtheid wordt $f_{x,y}(x,y) = e^{-x-y}, \quad x > 0, y > 0;$

$= 0, \quad$ elders.

$$E(\underline{x}^2\underline{y}) = E(\underline{x}^2)E(\underline{y}) = \int_0^\infty x^2 e^{-x} dx \cdot \int_0^\infty y e^{-y} dy = 2.$$

• De waarden van $p_{x,y}(x,y) \neq 0$ zijn aangegeven in figuur 4.7.



Figuur 4.7.

$p_x(x) = 1/6,$	$x = -2;$	$p_y(y) = 3/8,$	$y = 1;$
$= 1/4,$	$x = -1;$	$= 5/8,$	$y = -1;$
$= 7/12,$	$x = 1;$	$= 0,$	elders.
$= 0,$	elders.		

$$E(\underline{x}\underline{y}) = (-2)\frac{1}{24} + (-1)\frac{1}{8} + 1 \cdot \frac{5}{24} + 1 \cdot \frac{1}{8} + 2 \cdot \frac{1}{8} + (-1)\frac{3}{8} = 0;$$

$$E(\underline{x}) = (-2) \cdot \frac{1}{6} + (-1) \cdot \frac{1}{4} + 1 \cdot \frac{7}{12} = 0 \quad \text{dus} \quad E(\underline{x}\underline{y}) = E(\underline{x}) \cdot E(\underline{y}) = 0.$$

Voorts is $W_{x,y} = W_x \times W_y$.

Toch zijn $\underline{x}$ en $\underline{y}$ niet onafhankelijk want bijvoorbeeld

$$p_{x,y}(-1;1) = \frac{1}{8} \neq p_x(-1) \cdot p_y(1) = \frac{1}{4} \cdot \frac{3}{8} = \frac{3}{32}.$$

Onafhankelijkheid maakt het mogelijk om in meer-dimensionale situaties bewerkingen terug te brengen tot bewerkingen met behulp van de een-dimensionale marginale verdelingen, hetgeen vaak veel eenvoudiger is.

Als $\underline{x}$ en $\underline{y}$ niet onafhankelijk zijn is het nodig om behalve naar simultane en marginale verdelingen ook naar voorwaardelijke verdelingen te kijken.
Het begrip voorwaardelijke of conditionele kans kan ook worden doorgetrokken naar stochastische variabelen. Zij $\underline{x}$ een stochastische variabele gedefinieerd op Ω,

met verdelingsfunctie F_x. Zij voorts gegeven dat bij uitvoering van het experiment gebeurtenis B is opgetreden. Gebruik makend van dit gegeven kan een aangepaste kansverdeling van $\underline{x}$ worden opgesteld. Immers $F_x(x) = P(A)$ met $A = \{\omega \mid \underline{x}(\omega) \le x\}$. Bij gegeven B echter wordt de kans dat $\underline{x}$ kleiner of gelijk x is gelijk aan $P(A \mid B)$. Zo onstaat de voorwaardelijke verdelingsfunctie $F_{x|B}(x)$, ook genoteerd als $F_x(x \mid B)$.
Soms is het mogelijk de gebeurtenis B te omschrijven met behulp van de stochastische variabele $\underline{x}$ zelf. Zo kunnen ook voorwaardelijke dichtheden en kansfuncties worden gevormd. De notatie is analoog.

• Zij B de gebeurtenis dat $\underline{x} > 2$ dan is, als $P(B) \neq 0$,

$$F_x(x \mid B) = P(\underline{x} \le x \mid \underline{x} > 2) = \frac{P(\underline{x} \le x \cap \underline{x} > 2)}{P(\underline{x} > 2)},$$

zodat

$$\begin{aligned} F_x(x \mid \underline{x} > 2) &= 0, && x \le 2; \\ &= \frac{F_x(x) - F_x(2)}{1 - F_x(2)}, && x > 2. \end{aligned}$$

In plaats van $P(B) \neq 0$ kan men de gebruikelijke eis ook formuleren als $F_x(2) \neq 1$.
Als $\underline{x}$ discreet is volgt

$$p_x(x \mid B) = P(\underline{x} = x \mid \underline{x} > 2) = \frac{P(\underline{x} = x \cap \underline{x} > 2)}{P(\underline{x} > 2)},$$

zodat

$$\begin{aligned} p_x(x \mid \underline{x} > 2) &= \frac{p_x(x)}{1 - F_x(2)}, && x \in (2, \infty) \cap W_x; \\ &= 0, && \text{elders.} \end{aligned}$$

Voor $\underline{x}$ continu geldt

$$f_x(x \mid B) = \frac{d}{dx} F_x(x \mid B),$$

dus

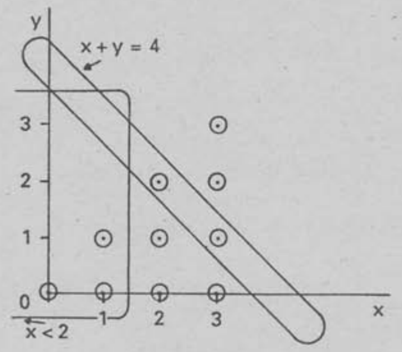
$$\begin{aligned} f_x(x \mid \underline{x} > 2) &= \frac{f_x(x)}{1 - F_x(2)}, && x > 2; \\ &= 0, && x \le 2. \end{aligned}$$

□

Bij simultane verdelingen kunnen ook voorwaardelijke verdelingen gevormd worden. In het discrete geval treden daarbij geen nieuwe moeilijkheden op, alleen is er een veel grotere variatie mogelijk bij het kiezen van de conditie B.
• De simultane kansfunctie van $\underline{x}$ en $\underline{y}$ is uniform over de in figuur 4.8 omcirkelde roosterpunten, met andere woorden

$$p_{\mathbf{x},\mathbf{y}}(x,y) = \frac{1}{10}, \quad x = 0,1,2,3;\ y = 0,1,2,3;\ y \le x;$$
$$= 0, \quad \text{elders.}$$

Zij de conditie: $\underline{x} < 2$.



Figuur 4.8.

$$p_{\mathbf{x},\mathbf{y}}(x,y \mid \underline{x} < 2) = \frac{P(\underline{x} = x \cap \underline{y} = y \cap \underline{x} < 2)}{P(\underline{x} < 2)} =$$
$$= \frac{1/10}{3/10} = \frac{1}{3}, \quad (x,y) = (0,0), (1,0), (1,1);$$
$$= 0, \quad \text{elders.}$$

$$p_{\mathbf{y}}(y \mid \underline{x} < 2) = \frac{P(\underline{y} = y \cap \underline{x} < 2)}{P(\underline{x} < 2)} =$$
$$= \frac{2}{3}, \quad y = 0;$$
$$= \frac{1}{3}, \quad y = 1;$$
$$= 0, \quad \text{elders.}$$

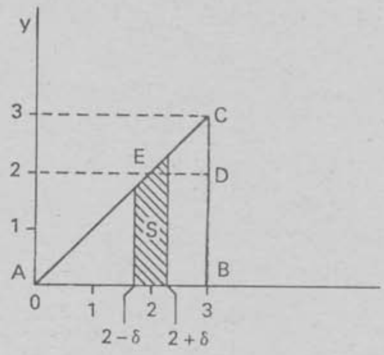
Zij de conditie $\underline{x} + \underline{y} = 4$.

$$p_{\mathbf{x},\mathbf{y}}(x,y \mid \underline{x} + \underline{y} = 4) = \frac{P(\underline{x} = x \cap \underline{y} = y \cap \underline{x} + \underline{y} = 4)}{P(\underline{x} + \underline{y} = 4)} =$$
$$= \frac{1/10}{2/10} = \frac{1}{2}, \quad (x,y) = (2,2), (3,1);$$
$$= 0, \quad \text{elders.} \qquad \square$$

Ook als $\underline{x}$ en $\underline{y}$ een continue simultane dichtheid hebben kan men op soortgelijke wijze te werk gaan.

- $\underline{x}$ en $\underline{y}$ zijn uniform verdeeld over de driehoek (0,0), (3,0), (3,3).

$$f_{\mathbf{x},\mathbf{y}}(x,y) = \frac{2}{9}, \quad 0 < y < x < 3;$$
$$= 0, \quad \text{elders.}$$



Figuur 4.9.

Zij de conditie $\underline{y} < 2$:

$$F_{x,y}(x,y \mid \underline{y}<2) = \frac{P(\underline{x} \le x \cap \underline{y} \le y \cap \underline{y} < 2)}{P(\underline{y} < 2)} =$$

$$= \frac{F_{x,y}(x,y)}{P(\underline{y} < 2)}, \qquad (x,y) \text{ in het trapezium ABDE;}$$

$$= \frac{F_{x,y}(x,2)}{P(\underline{y} < 2)}, \qquad (x,y) \text{ in de driehoek EDC.}$$

De voorwaardelijke simultane dichtheid volgt door differentiëren:

$$f_{x,y}(x,y \mid \underline{y}<2) = \frac{\partial^2}{\partial x \partial y} F_{x,y}(x,y \mid \underline{y}<2) =$$

$$= \frac{f_{x,y}(x,y)}{P(\underline{y} < 2)}, \quad (x,y) \text{ in ABDE;}$$

$$= 0, \qquad (x,y) \text{ in EDC.}$$

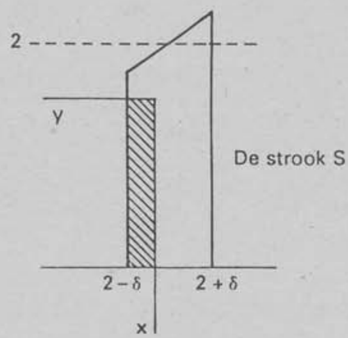
Voor (x,y) niet in de driehoek ABC is de voorwaardelijke dichtheid nul, omdat daar de onvoorwaardelijke dichtheid nul is en door het stellen van een conditie het waardenbereik van de stochastische variabelen nooit kan worden uitgebreid.
Resumerend krijgen we

$$f_{x,y}(x,y|\underline{y}<2) = \frac{2/9}{8/9} = \frac{1}{4}, \qquad (x,y) \text{ in ABDE;}$$

$$= 0, \qquad \text{elders.}$$

Zij nu B de gebeurtenis $2-\delta < \underline{x} \le 2+\delta$, dan is $P(B) = 2 \cdot 2\delta \cdot \frac{2}{9}$.
Voor $y < 2-\delta$ en $2-\delta < x < 2+\delta$ geldt

$$P(\underline{x} \le x \cap \underline{y} < y \cap 2-\delta < \underline{x} \le 2+\delta) = y(x-2+\delta)\cdot\frac{2}{9}$$



Figuur 4.10.

zodat

$$F_{\underline{x},\underline{y}}(x,y \mid 2-\delta < \underline{x} \leq 2+\delta) = \frac{y(x-2+\delta)\cdot 2/9}{2\cdot 2\delta\cdot 2/9}$$

waaruit voor de voorwaardelijke simultane dichtheid volgt

$$\begin{aligned} f_{\underline{x},\underline{y}}(x,y \mid 2-\delta < \underline{x} \leq 2+\delta) &= \tfrac{1}{4\delta}, \quad (x,y) \in S; \\ &= 0, \quad \text{elders.} \end{aligned}$$

Het is duidelijk dat de voorwaardelijke simultane dichtheid naar ∞ gaat als δ naar nul nadert.

Het is mogelijk dat bij een experiment, waarbij op de uitkomstenruimte twee continue stochastische variabelen $\underline{x}$ en $\underline{y}$ zijn gedefinieerd de realisatie van $\underline{x}$ al bekend is maar die van $\underline{y}$ nog niet. In zo'n geval is men geïnteresseerd in de voorwaardelijke dichtheid van $\underline{y}$, gegeven $\underline{x} = x$. Hier schiet de gebruikte methode tekort omdat de kans op de conditie nul is. De volgende definitie is echter voor de hand liggend.

Definitie

Zijn $\underline{x}$ en $\underline{y}$ stochastische variabelen met een continue simultane dichtheid dan is de voorwaardelijke dichtheid van $\underline{y}$, gegeven $\underline{x} = x$,

$$\begin{aligned} f_{\underline{y}}(y \mid \underline{x} = x) &= \frac{f_{\underline{x},\underline{y}}(x,y)}{f_{\underline{x}}(x)}, \text{ mits } f_{\underline{x}}(x) \neq 0; \\ &= \text{niet gedefinieerd voor die waarden van } x \text{ waarvoor } f_{\underline{x}}(x) = 0. \quad \square \end{aligned}$$

De gedefinieerde uitdrukking kan men zich ontstaan denken via de volgende benaderingen:

$$f_{\underline{x}}(x)dx \approx P(x < \underline{x} \leq x + dx);$$

$$f_{\underline{x},\underline{y}}(x,y)dxdy \approx P(x < \underline{x} \leq x + dx \cap y < \underline{y} \leq y + dy),$$

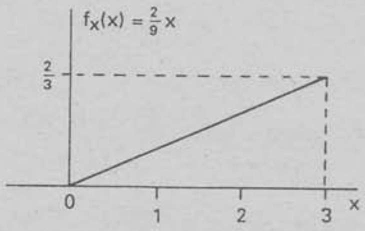
dus als $f_x(x) \neq 0$:

$$f_y(y \mid x < \underline{x} \leq x + dx)dy \approx P(y < \underline{y} \leq y + dy \mid x < \underline{x} \leq x + dx) =$$

$$= \frac{P(x < \underline{x} \leq x + dx \cap y < \underline{y} \leq y + dy)}{P(x < \underline{x} \leq x + dx)} \approx \frac{f_{x,y}(x,y)dxdy}{f_x(x)dx} = \frac{f_{x,y}(x,y)}{f_x(x)}\ dy,$$

daarna kan dx tot nul naderen zodat

$$f_y(y \mid \underline{x} = x) = \frac{f_{x,y}(x,y)}{f_x(x)}.$$



Figuur 4.11.

Voorbeeld

•Vervolg vorige voorbeeld.

$$f_y(y|\underline{x} = 2) = \frac{f_{x,y}(2,y)}{f_x(2)} = \frac{2/9}{4/9} = \frac{1}{2}, \qquad 0 < y < 2;$$

$$= 0, \qquad \text{elders.}$$

□

Ook deze voorwaardelijke dichtheid heeft de eigenschappen van een gewone dichtheid:

a. $f_y(y \mid \underline{x} = x)$ is een quotiënt van twee dichtheden en dus niet-negatief, als hij gedefinieerd is.

b. $\int_y \frac{f_{x,y}(x,y)}{f_x(x)}\, dy = \frac{f_x(x)}{f_x(x)} = 1.$

Voorwaardelijke dichtheden en kansfuncties kunnen op de gewone manier worden gebruikt om verwachtingen te bepalen die dan echter voorwaardelijke verwachtingen heten.

$$E(\underline{x}|B) = \sum_x x p_x(x|B) \text{ als } \underline{x} \text{ discreet is;}$$

$$E(\underline{x}|B) = \int_x x f_x(x|B)dx \text{ als } \underline{x} \text{ continu is.}$$

In het algemeen zal $E(\underline{y} \mid \underline{x} = x) = \int_y y\, f_y(y \mid \underline{x} = x)dy$ een functie van x zijn. Deze functie heet de regressie van y op x.

• Vervolg vorige voorbeeld.

$$E(\underline{y} \mid \underline{x} = x) = \int_0^x y \frac{2/9}{2x/9} dy = \frac{1}{x} \int_0^x y dy = \frac{1}{2} x, \qquad 0 < x < 3;$$

$$= \text{niet gedefinieerd}, \qquad \text{elders.}$$

□

Als $\underline{x}$ en $\underline{y}$ onafhankelijk zijn dan zijn alle voorwaardelijke verdelingen van $\underline{y}$ met een voorwaarde die geheel omschreven kan worden in termen van $\underline{x}$ gelijk aan de marginale verdeling van $\underline{y}$.

• De stochastische variabelen $\underline{x}$ en $\underline{y}$ zijn onafhankelijk. Dan geldt voor alle x:

$$f_x(x \mid \underline{y} = y) = f_x(x) \text{ voor alle } y \text{ waarvoor } f_y(y) \neq 0;$$

$$p_x(x \mid \underline{y} \leq y) = p_x(x) \text{ voor alle } y \text{ waarvoor } F_y(y) \neq 0.$$

In het algemeen zal niet gelden dat $f_x(x \mid \underline{x} + \underline{y} = z)$ gelijk is aan $f_x(x)$ omdat voor de omschrijving van de voorwaarde $\underline{x}$ en $\underline{y}$ gebruikt moeten worden.

4.4. Stochastische vectoren in n dimensies

Er is geen enkel bezwaar tegen om op de uitkomstenruimte Ω een n-tal stochastische variabelen te definiëren in plaats van slechts twee. Wil men alle n stochastische variabelen tegelijk beschouwen dan heeft men de n–dimensionale verdeling nodig. De stochastische vector $(\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_n)$ kent aan iedere $\omega \in \Omega$ een vector $(x_1,x_2,\ldots,x_n)$ in $\mathbb{R}^n$ toe. De verdeling van de stochastische vector$(\underline{x}_1,\underline{x}_2, \ldots,\underline{x}_n)$ is de simultane verdeling van de stochastische variabelen $\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_n$.
Uit de n-dimensionale kansfunctie

$$p_{x_1,\ldots,x_n}(x_1,\ldots,x_n) = P(\underline{x}_1 = x_1 \cap \underline{x}_2 = x_2 \cap \ldots \cap \underline{x}_n = x_n)$$

of de n-dimensionale dichtheid

$$f_{x_1,\ldots,x_n}(x_1,\ldots,x_n) = \frac{\partial^n}{\partial x_1 \ldots \partial x_n} F_{x_1,\ldots,x_n}(x_1,\ldots,x_n)$$

kan men nu tal van marginale verdelingen afleiden.
Zo wordt de (n – 1)-dimensionale marginale dichtheid van $\underline{x}_2,\ldots,\underline{x}_n$

$$\int_{x_1=-\infty}^{\infty} f_{x_1,\ldots,x_n}(x_1,\ldots,x_n)dx_1 = f_{x_2,\ldots,x_n}(x_2,\ldots,x_n).$$

In veel gevallen zullen de stochastische variabelen $\underline{x}_1,\ldots,\underline{x}_n$ onafhankelijk zijn. Een nodige en voldoende voorwaarde daarvoor is dat voor elke $(x_1,\ldots,x_n) \in \mathbb{R}^n$

$$p_{x_1,\ldots,x_n}(x_1,\ldots,x_n) = p_{x_1}(x_1)\ldots p_{x_n}(x_n)$$

dan wel

$$f_{x_1,\ldots,x_n}(x_1,\ldots,x_n) = f_{x_1}(x_1)\ldots f_{x_n}(x_n).$$

De stochastische variabelen $\underline{x}_1, \underline{x}_2, \ldots$ vormen een aftelbare rij onafhankelijke stochastische variabelen als elke eindige greep uit de rij een collectie onafhankelijke stochastische variabelen oplevert.
Als $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ een simultane verdeling hebben dan kan door uitschrijven (als som of integraal) gemakkelijk worden aangetoond dat

$$E\left(\sum_{i=1}^{n} a_i \underline{x}_i\right) = \sum_{i=1}^{n} a_i E(\underline{x}_i)$$

en

$$\sigma^2\left(\sum_{i=1}^{n} a_i \underline{x}_i\right) = \sum_{i=1}^{n} a_i^2 \sigma^2(\underline{x}_i) + \sum_{\substack{i=1 \\ i \neq j}}^{n} \sum_{j=1}^{n} a_i a_j \operatorname{cov}(\underline{x}_i, \underline{x}_j)$$

terwijl als de stochastische variabelen onafhankelijk zijn deze laatste uitdrukking wordt:

$$\sigma^2\left(\sum_{i=1}^{n} a_i \underline{x}_i\right) = \sum_{i=1}^{n} a_i^2 \sigma^2(\underline{x}_i).$$

Enkele vaak voorkomende meer-dimensionale verdelingen hebben een eigen naam gekregen.

A. De multinomiale verdeling

De *multinomiale verdeling* is een generalisatie van de binomiale verdeling. Men voert n onafhankelijke experimenten uit. Elk experiment heeft als mogelijke uitkomsten $s_1, s_2, \ldots, s_k$ en voor elk experiment geldt

$$P(s_i) = p_i,\ i = 1,2,\ldots,k;\ \sum p_i = 1.$$

Zij $\underline{x}_i$ het aantal malen dat uitkomst s_i, $i = 1,2,\ldots,k$ optreedt bij de n uitvoeringen van het experiment, dan heeft $(\underline{x}_1, \ldots, \underline{x}_k)$ een multinomiale verdeling. De bijbehorende kansfunctie is

$$p_{\underline{x}_1,\ldots,\underline{x}_k}(x_1,\ldots,x_k) = \frac{n!}{x_1!x_2!\ldots x_k!}\, p_1^{x_1} \cdot p_2^{x_2} \cdots p_k^{x_k}, \quad (x_1,\ldots,x_k) \in A$$

$$= 0, \qquad \text{elders}$$

waarbij $A \in \mathbb{N}^k$ met $\sum_{i=1}^{k} x_i = n$.

Immers een rij van n resultaten waarbij de eerste x_1 proeven s_1 opleveren, de volgende x_2 proeven s_2 opleveren, enz. heeft een kans

$$p_1^{x_1} \cdot p_2^{x_2} \cdots p_k^{x_k}.$$

Echter alleen het *aantal* resultaten per soort is van belang, niet wanneer een bepaald resultaat optreedt. Er moet dus nog worden vermenigvuldigd met het aantal verschillende volgordes van x_1 resultaten s_1, x_2 resultaten s_2, enzovoort.
De marginale verdeling van $\underline{x}_i$ is de binomiale verdeling met parameters n en p_i.

B. De multinormale verdeling

Zoals de naam al doet vermoeden, is de *multinormale verdeling* nauw verwant aan de *Normale verdeling*. Om niet meteen met een stadspoort in huis te vallen zullen we eerst de binormale (of 2-dimensionale Normale verdeling) beschouwen. De continue stochastische variabelen $\underline{x}$ en $\underline{y}$ zijn binormaal verdeeld als

$$f_{x,y}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\, e^{\frac{-1}{2(1-\rho^2)}\{x_1^2-2\rho x_1y_1+y_1^2\}}, \text{ voor alle } (x,y),$$

waarbij $x_1 = \dfrac{x-\mu_x}{\sigma_x}$ en $y_1 = \dfrac{y-\mu_y}{\sigma_y}$.

De dichtheid ligt geheel vast zodra de 5 parameters μ_x, σ_x, μ_y, σ_y en ρ bekend zijn. De voor deze parameters gebruikte symbolen hebben hun gebruikelijke betekenis:

$$\mu_x = E(\underline{x}),\ \ \sigma_x = \sqrt{\sigma^2(\underline{x})}\ \text{ en }\ \rho = \rho(\underline{x},\underline{y}).$$

De marginale verdeling van $\underline{x}$ is $N(\mu_x,\sigma_x^2)$, analoog voor $\underline{y}$.
De voorwaardelijke dichtheid van $\underline{x}$, gegeven $\underline{y} = y$ wordt

$$f_{x,y}(x,y)\cdot\frac{1}{f_y(y)} = f_{x,y}(x,y)\cdot\sigma_y\cdot\sqrt{2\pi}\cdot e^{\frac{1}{2}y_1^2}.$$

In de exponent van e komt dus

$$\frac{-1}{2(1-\rho^2)}\{x_1^2-2\rho x_1y_1+y_1^2-(1-\rho)^2y_1^2\} = -\frac{1}{2}\cdot\frac{(x_1-\rho y_1)^2}{1-\rho^2}$$

of $$-\frac{1}{2}\left\{\frac{x-\mu_x-(y-\mu_y)\rho\sigma_x/\sigma_y}{\sigma_x\sqrt{1-\rho^2}}\right\}^2.$$

Dus wordt $f_x(x \mid \underline{y} = y) = \dfrac{1}{a\sqrt{2\pi}}\cdot e^{-\frac{1}{2}(\frac{x-b}{a})^2}$

met $a = \sigma_x\sqrt{1-\rho^2}$ en $b = \mu_x + (y-\mu_y)\rho\dfrac{\sigma_x}{\sigma_y}$, zodat $\underline{x}$ bij gegeven $\underline{y} = y$ Normaal verdeeld is met verwachting b en variantie a^2. Als $\underline{x}$ en $\underline{y}$ onafhankelijk zijn dan is $\rho = 0$ zodat $b = \mu_x$ en $a = \sigma_x$, hetgeen te verwachten was.

Als $\rho(\underline{x},\underline{y}) = 0$ heten $\underline{x}$ en $\underline{y}$ ongecorreleerd. In het algemeen zijn ongecorreleerde variabelen niet onafhankelijk. Is echter bekend dat $\underline{x}$ en $\underline{y}$ binormaal verdeeld en ongecorreleerd zijn dan volgt dat zij onafhankelijk zijn. Immers voor $\rho = 0$ volgt

$$f_{x,y}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}(x_1^2+y_1^2)} = f_x(x)\cdot f_y(y), \text{ voor alle x en y.}$$

Als $\underline{x}$ en $\underline{y}$ afhankelijk zijn en beide een Normale marginale verdeling hebben is het toch mogelijk dat de simultane verdeling niet binormaal is.

De multinormale verdeling kan het beste beschreven worden met behulp van matrix-notatie.
Zij $\underline{X}'$ de stochastische vector $(\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_n)$. Zij $E(\underline{x}_i) = \mu_i$, $\sigma^2(\underline{x}_i) = \sigma_i^2$ en $\mathrm{cov}(\underline{x}_i,\underline{x}_j) = c_{ij}$.
De verwachtingen verzamelen we in de vector

$$\mu' = (\mu_1,\mu_2,\ldots,\mu_n) = E(\underline{X}').$$

De varianties en covarianties kunnen worden samengevat in de symmetrische variantie-covariantie-matrix V:

$$V = \begin{pmatrix} \sigma_1^2 & c_{12} & \ldots & c_{1n} \\ c_{21} & \sigma_2^2 & \ldots & c_{2n} \\ \vdots & & & \vdots \\ c_{n1} & \ldots & \ldots & \sigma_n^2 \end{pmatrix}.$$

Als we de verwachting van een matrix met stochastische elementen definiëren als de matrix van de verwachtingen dan is V te schrijven als

$$V = E\{(\underline{X}-\mu)(\underline{X}-\mu)'\}.$$

V is positief semidefiniet dus $|V| \geq 0$. Als $|V| > 0$ zodat V^{-1} bestaat kan de dichtheid van de n-dimensionale multinormale verdeling worden geschreven als

$$f(X) = \frac{1}{(2\pi)^{n/2}\sqrt{|V|}} e^{-\frac{1}{2}(X-\mu)'V^{-1}(X-\mu)}, \text{ voor alle } X \in \mathbb{R}^n.$$

Is V singulier, m.a.w. is $|V| = 0$ dan bestaat er een lineair verband tussen een aantal van de stochastische variabelen en hadden we beter in een ruimte met lagere dimensie kunnen werken.
Als $\underline{x}_1,\ldots,\underline{x}_n$ multinormaal verdeeld zijn dan heeft elke lineaire combinatie $\sum_{i=1}^{n} a_i\underline{x}_i$ een Normale verdeling. Door een van de a_i's gelijk één te kiezen en de overige nul volgt dat ook de $\underline{x}_i$'s allemaal Normaal verdeeld zijn.

4.5. De ongelijkheid van Chebychev; Wet van de grote aantallen

Stelling 4.8. Voor elke stochastische variabele $\underline{x}$ waarvan het tweede moment bestaat geldt de ongelijkheid van Chebychev, m.a.w.

$$P(|\underline{x}-\mu| \geq k\sigma) \leq \frac{1}{k^2} \text{ voor alle } k > 0.$$

Bewijs: Voor $\underline{x}$, een continue stochastische variabele met verwachting μ en variantie σ^2, geldt:

$$\sigma^2 = \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)dx =$$

$$= \int_{-\infty}^{\mu-k\sigma} \ldots + \int_{\mu-k\sigma}^{\mu+k\sigma} \ldots + \int_{\mu+k\sigma}^{\infty} (x-\mu)^2 f(x)dx = A + B + C.$$

In de integraal A is $x \leq \mu - k\sigma$ ofwel $x - \mu \leq -k\sigma$, dus $(x - \mu)^2 \geq k^2\sigma^2$. In integraal C is $x \geq \mu + k\sigma$, dus $(x - \mu)^2 \geq k^2\sigma^2$. Integraal B heeft een niet-negatieve integrand en voor $k > 0$ een positieve integratie-richting en is dus groter of gelijk aan nul. Met dit alles volgt:

$$\sigma^2 \geq k^2\sigma^2 \{ \int_{-\infty}^{\mu-k\sigma} f(x)dx + \int_{\mu+k\sigma}^{\infty} f(x)dx \}.$$

Tussen accolades wordt de dichtheid van $\underline{x}$ geïntegreerd over $x \leq \mu - k\sigma$ ofwel $x - \mu \leq -k\sigma$ en over $x \geq \mu + k\sigma$ ofwel over $x - \mu \geq k\sigma$. Samen is dat dus de kans dat $|\underline{x} - \mu| \geq k\sigma$. Dit levert

$$\sigma^2 \geq k^2\sigma^2 \cdot P(|\underline{x} - \mu| \geq k\sigma)$$

waaruit het gestelde volgt.
In het discrete geval verloopt het bewijs analoog. □

In de praktijk zijn de afschattingen die m.b.v. deze ongelijkheid kunnen worden verkregen meestal veel te ruw. Dat komt omdat er nauwelijks eisen aan de verdeling van de stochastische variabele $\underline{x}$ worden gesteld. Zodra meer informatie over de verdeling beschikbaar is zijn scherpere grenzen aan te geven.
Een aardige toepassing van Chebychev's ongelijkheid is het bewijs van de wet van de grote aantallen.

Stelling 4.9. Zwakke wet van de grote aantallen.
Zij $\underline{x}_i$, $i = 1,2,\ldots,n$ een rij onafhankelijke stochastische variabelen, alle met dezelfde verdeling met $E(\underline{x}_i) = \mu$ en $\sigma^2(\underline{x}_i) = \sigma^2$.

Zij $\underline{\bar{x}}_n = \frac{1}{n}\sum_{i=1}^{n} \underline{x}_i$ dan geldt dat

$$\lim_{n\to\infty} P(|\underline{\bar{x}}_n - \mu| > \varepsilon) = 0, \text{ voor alle } \varepsilon > 0.$$

Bewijs: $E(\underline{\bar{x}}_n) = \frac{1}{n}\sum_{i=1}^{n} E(\underline{x}_i) = \mu$

$$\sigma^2(\underline{x}_n) = \frac{1}{n^2}\sigma^2(\underline{x}_1+\underline{x}_2+\ldots+\underline{x}_n) = \frac{1}{n^2}\{\sigma^2(\underline{x}_1)+\sigma^2(\underline{x}_2)+\ldots+\sigma^2(\underline{x}_n)\} = \frac{1}{n}\cdot\sigma^2$$

omdat de stochastische variabelen onafhankelijk zijn.
Voor elke n geldt dus m.b.v. Chebychev:

$$P(|\underline{\bar{x}}_n - \mu| \geq k\frac{\sigma}{\sqrt{n}}) \leq \frac{1}{k^2} \text{ voor alle } k > 0.$$

Kies k zó dat $k\frac{\sigma}{\sqrt{n}} = \varepsilon$ ofwel $k = \frac{\varepsilon\sqrt{n}}{\sigma}$, dan

$$P(|\underline{\bar{x}}_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}.$$

Hierbij zijn σ en ε vaste getallen maar het rechterlid kan willekeurig dicht bij nul worden gebracht door n groot genoeg te kiezen. □

Het is mogelijk om deze stelling te bewijzen zonder te eisen dat de variantie van $\underline{x}_i$ bestaat. De zwakke wet rechtvaardigt het gebruik van frequentie-quotiënten als benadering van kansen.

Voorbeeld

Bij een experiment kan de gebeurtenis A optreden, P(A) = p. Zij $\underline{f}_n$ de relatieve frequentie van de gebeurtenis A bij n onafhankelijke uitvoeringen van het experiment. $n\underline{f}_n$ is het aantal malen dat A optreedt bij de n proeven, dus $n\underline{f}_n = \underline{x}_1+\underline{x}_2+\ldots+\underline{x}_n$ waarbij voor $i = 1,2,\ldots,n$ $\underline{x}_i = 1$ als A optreedt en $\underline{x}_i = 0$ als A niet optreedt bij proef i.
De stochastische variabelen $\underline{x}_i$ zijn onafhankelijk en hebben alle dezelfde verdeling. $\underline{f}_n = \underline{\bar{x}}_n$, $E(\underline{x}_i) = p$, $\sigma^2(\underline{x}_i) = p(1-p)$.
De zwakke wet levert

$$\lim_{n\to\infty} P(|\underline{f}_n - p| \geq \varepsilon) = 0$$

of

$$P(|\underline{f}_n - p| \geq \varepsilon) \leq \frac{p(1-p)}{n\varepsilon^2}.$$

Als nu P(A) = p onbekend is, hoe vaak moeten we dan het experiment herhalen voordat we met een kans van minstens 0,95 een waarde van de relatieve frequentie krijgen die hoogstens 0,01 verschilt van p?
$P(|\underline{f}_n - p| \geq 0{,}01)$ mag dus hoogstens 0,05 zijn, m.a.w.

$$\frac{p(1-p)}{n\cdot(0{,}01)^2} \leq 0{,}05 \text{ of } n \geq \frac{p(1-p)}{0{,}05(0{,}01)^2} = p(1-p)\cdot 2\cdot 10^5.$$

Als er geen enkel idee omtrent de waarde van p is moeten we het meest ongunstige geval nemen. $p(1-p)$ bereikt voor $p = \frac{1}{2}$ de maximale waarde $\frac{1}{4}$. Dus moet $n \geq \frac{1}{4}\cdot 2\cdot 10^5 = 5\cdot 10^4$ worden gekozen.

4.6. De Centrale Limietstelling; Benaderingen

De belangstelling voor de Normale verdeling is deels een gevolg van wat de Centrale Limietstelling (CLS) ons levert.

Stelling 4.10. Zij $\underline{x}_1, \underline{x}_2, \ldots$ een rij onafhankelijke, identiek verdeelde stochastische variabelen met $E(\underline{x}_i) = \mu$ en $\sigma^2(\underline{x}_i) = \sigma^2$, dan gaat de verdeling van

$$\underline{z}_n = \frac{\overline{\underline{x}}_n - \mu}{\sigma/\sqrt{n}}$$

voor $n \to \infty$ naar de standaard-Normale verdeling. □

We zullen deze stelling niet bewijzen, maar volstaan met er op te wijzen dat $E(\underline{z}_n) = 0$ en $\sigma^2(\underline{z}_n) = 1$ voor alle n.

Het resultaat kan ook worden omschreven als: de verdeling van

$$\frac{\sum_{i=1}^{n} \underline{x}_i - n\mu}{\sigma\sqrt{n}}$$

nadert tot de standaard-Normale verdeling. Zo kan ook de zwakke wet worden geschreven in de vorm

$$\frac{\sum_{i=1}^{n} \underline{x}_i - n\mu}{\sigma\cdot n} \to 0 \text{ met kans één.}$$

De factor $\sqrt{n}$ in de noemer is essentieel voor de CLS. Als er in de noemer n^α staat met $\alpha > \frac{1}{2}$ convergeert de rij stochastische variabelen met kans één naar 0, zoals bij de zwakke wet. Voor $\alpha < \frac{1}{2}$ wordt het gedrag van de rij te wild om prettig te beschrijven.

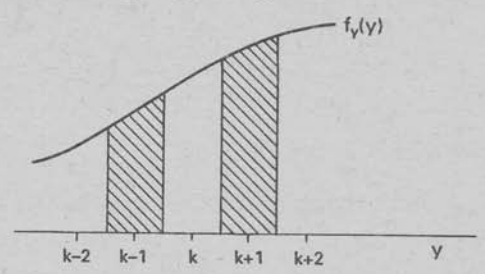
De CLS maakt het mogelijk om voor sommige stochastische variabelen de kansen te benaderen m.b.v. de tabel van de N(0;1)-verdeling. Zij $\underline{x} \sim B(n,p)$, dan kan $\underline{x}$ worden beschouwd als de som van n onafhankelijke stochastische variabelen $\underline{x}_i$ elk met de verdeling

$$P(\underline{x}_i = 0) = q;\ P(\underline{x}_i = 1) = p;$$

zodat $\mu_i = p$ en $\sigma_i = \sqrt{pq}$, $i = 1,2,\ldots,n$.
Hiermee volgt dat voor grote n ongeveer geldt

$$\frac{\underline{x}-np}{\sqrt{pqn}} \sim N(0;1) \text{ ofwel } \underline{x} \sim N(\mu_x;\sigma_x^2).$$

De Normale verdeling is continu zodat $P(\underline{x} = k)$ voor elke k met nul benaderd zou worden als er geen continuïteitscorrectie werd toegepast. Daartoe introduceren we de continue stochastische variabele $\underline{y} \sim N(\mu_x;\sigma_x^2)$. $P(\underline{x} = k)$ wordt nu benaderd door $P(k - \frac{1}{2} < \underline{y} \leq k + \frac{1}{2})$. Op die manier wordt de continu uitgesmeerde kansmassa geconcentreerd op de gehele getallen.



Figuur 4.12.

Natuurlijk krijgen zo ook de gehele getallen < 0 en > n een kans ongelijk aan nul maar die is meestal zo klein dat het er niet toe doet. De zo verkregen benadering geeft redelijke resultaten als min(np,nq) > 5. M.a.w. hoe verder p van $\frac{1}{2}$ af ligt hoe groter n moet zijn. Voor $p = \frac{1}{2}$ is de binomiale verdeling symmetrisch om zijn verwachting voor elke n en het is begrijpelijk dat de benadering met de symmetrische Normale verdeling dan sneller goede resultaten geeft.

Voorbeelden

$\underline{x} \sim B(50,\frac{1}{4})$. Stel $\underline{y} \sim N(\mu_x;\sigma_x^2)$ en $\underline{u} \sim N(0;1)$.
$P(\underline{x} \leq 14) = 0{,}7481$ volgens de tabel van de binomiale verdeling.

$$P(\underline{x} \leq 14) \approx P(\underline{y} \leq 14\tfrac{1}{2}) = P\left(\underline{u} \leq \frac{14\frac{1}{2} - 12\frac{1}{2}}{\sqrt{50\cdot\frac{1}{4}\cdot\frac{3}{4}}}\right) = P(\underline{u} \leq 0{,}653).$$

Uit de N(0;1)-tabel volgt

$P(\underline{u} \le 0{,}65) = 0{,}7422$;
$P(\underline{u} \le 0{,}66) = 0{,}7454$, zodat na lineaire interpolatie volgt $P(\underline{x} \le 14) \approx 0{,}7432$.

• $\underline{x} \sim B(20,\frac{1}{2})$. Stel $\underline{y} \sim N(10;5)$ en $\underline{u} \sim N(0;1)$.
$P(\underline{x} \ge 12) = 1 - P(\underline{x} \le 11) = 0{,}2517$, exact;

$$P(\underline{x} \ge 12) \approx P(\underline{y} > 11\tfrac{1}{2}) = P(\underline{u} > \frac{11\frac{1}{2} - 10}{\sqrt{5}}) = P(\underline{u} > 0{,}671) = 0{,}2511, \text{ benadering.}$$

4.7. De verdeling van functies van twee of meer stochastische variabelen

Soms is het voldoende als van een functie van stochastische variabelen de momenten bekend zijn maar het kan ook voorkomen dat de kansverdeling van zo'n functie bepaald moet worden. Daarvoor bestaan verschillende methoden waarvan we er twee zullen behandelen.

A. De verdelingsfunctie-methode

De methode is in wezen gelijk aan de in paragraaf 2.5 beschreven methode om de verdeling van een functie van één stochastische variabele te bepalen.
Zij $\underline{z} = g(\underline{x},\underline{y})$, dan is

$$F_z(z) = P(\underline{z} \le z) = P\{(\underline{x},\underline{y}) \in A\} \quad \text{met } A = \{(x,y) \mid g(x,y) \le z\}$$

zodat

$$F_z(z) = \iint_A f_{x,y}(x,y)\, dxdy, \qquad \text{continu;}$$

$$= \sum\sum_{(x_i,y_j) \in A} p_{x,y}(x_i,y_j), \qquad \text{discreet.}$$

Uit F_z volgt op de gebruikelijke wijze de dichtheid of kansfunctie van $\underline{z}$.

Voorbeeld

$\underline{x}$ en $\underline{y}$ zijn uniform verdeeld over het eenheidsvierkant; $\underline{z} = \underline{x} + \underline{y}$. Direct kan worden ingezien dat

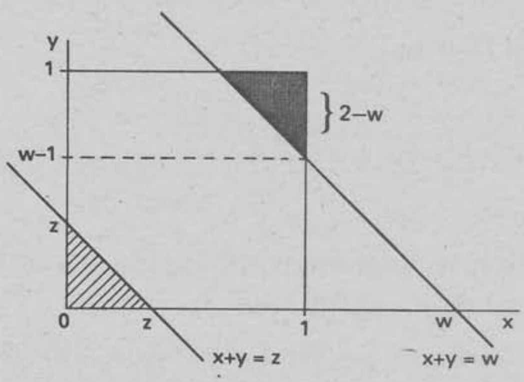
$$F_z(z) = 0, \quad z \le 0;$$
$$= 1, \quad z \ge 2.$$

Voor $0 < z < 1$ geldt dat $\{\underline{z} \le z\} = \{(\underline{x},\underline{y})$ in de gearceerde driehoek$\}$, dus

$$F_z(z) = \int_{y=0}^{z} \int_{x=0}^{z-y} f_{x,y}(x,y)\, dxdy = \tfrac{1}{2} z^2.$$

In dit simpele geval is het voldoende om het oppervlak van de driehoek te bepalen.

Voor $1 < w < 2$ geldt $\{\underline{z} \le w\} = \{(\underline{x},\underline{y})$ buiten de gestippelde driehoek$\}$, dus

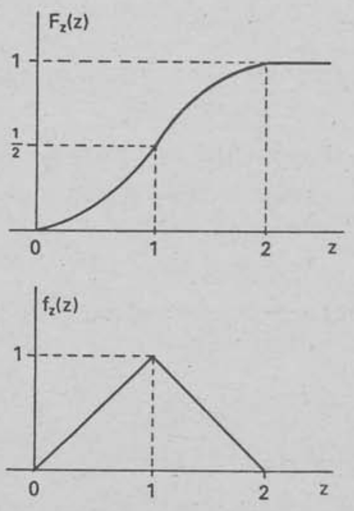


Figuur 4.13.

$$F_{\underline{z}}(w) = \int_{y=0}^{w-1}\int_{x=0}^{1} f_{\underline{x},\underline{y}}(x,y)\,dxdy + \int_{y=w-1}^{1}\int_{x=0}^{w-y} f_{\underline{x},\underline{y}}(x,y)\,dxdy = 1 - \frac{1}{2}(2-w)^2.$$

In dit geval is het handiger om $F_{\underline{z}}(w)$ via het complement te bepalen:

$$F_{\underline{z}}(w) = 1 - P(\underline{z} > w) = 1 - \int_{y=w-1}^{1}\int_{x=w-y}^{1} f_{\underline{x},\underline{y}}(x,y)\,dxdy.$$



Figuur 4.14.

Resumerend

$$
\begin{aligned}
F_{\underline{z}}(z) &= 0, && z \le 0;\\
&= \tfrac{1}{2} z^2, && 0 < z \le 1;\\
&= 1 - \tfrac{1}{2}(2-z)^2, && 1 < z \le 2;\\
&= 1, && z > 2.
\end{aligned}
$$

De dichtheid van $\underline{z}$ wordt dus:

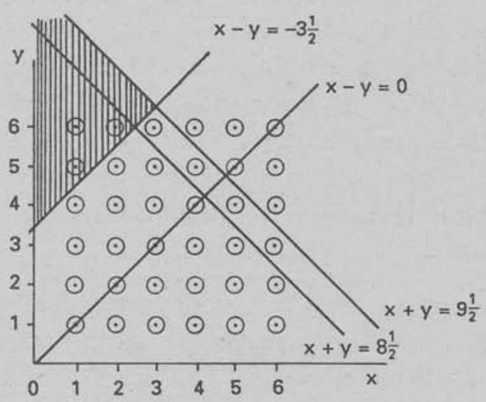
$$
\begin{aligned}
f_{\underline{z}}(z) &= z, && 0 < z < 1;\\
&= 2 - z, && 1 < z < 2;\\
&= 0, && \text{elders.}
\end{aligned}
$$

Op dezelfde manier kan ook de simultane verdeling van twee functies van stochastische variabelen worden bepaald.

Voorbeeld

$\underline{x}$ en $\underline{y}$ zijn onafhankelijk en beide uniform verdeeld op $\{1,2,\ldots,6\}$. Definieer de functies:

$\underline{u} = \underline{x} + \underline{y},\ \underline{v} = \underline{x} - \underline{y}.$



Figuur 4.15.

$$F_{\underline{u},\underline{v}}(u,v) = P(\underline{u} \le u, \underline{v} \le v) = P(\underline{x} + \underline{y} \le u, \underline{x} - \underline{y} \le v) =$$

$$= \sum_A\sum p_{\underline{x},\underline{y}}(x_i,y_j) \ \text{ met } \ A = \{(x_i,y_j) \mid (x_i + y_j) \le u;\ (x_i - y_j) \le v\}.$$

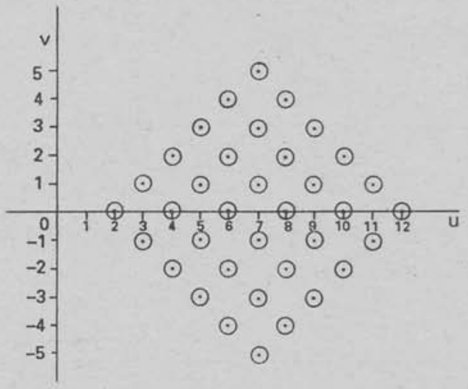
Neem $u = 9\frac{1}{2}$ en $v = -3\frac{1}{2}$ dan wordt $F_{\underline{u},\underline{v}}(9\frac{1}{2}, -3\frac{1}{2}) = \frac{3}{36}$ want A is nu het gearceerde gebied.

De volledige specificatie van $F_{\underline{u},\underline{v}}$ is eenvoudig maar omslachtig. Het is overzichtelijker om $p_{u,v}$ te bepalen, waaruit dan desgewenst $F_{u,v}$ door sommatie kan worden gevonden.

$$p_{u,v}(u,v) = P(\underline{x} + \underline{y} = u,\ \underline{x} - \underline{y} = v) = \frac{1}{36}, \quad \text{voor } (u,v) \in B;$$

$$= 0, \quad \text{elders.}$$

De verzameling B is in figuur 4.16 in beeld gebracht. □



Figuur 4.16.

Neem nu aan dat g en h continue functies zijn, en $\underline{x}$ en $\underline{y}$ continue stochastische variabelen. Het bepalen van de simultane verdelingsfunctie van $\underline{u} = g(\underline{x},\underline{y})$ en $\underline{v} = h(\underline{x},\underline{y})$ gaat op dezelfde manier. Ook daar is meestal de specificatie van $F_{u,v}$ omslachtig. Rechtstreeks uit de simultane dichtheid $f_{x,y}$ van $\underline{x}$ en $\underline{y}$ die van $\underline{u}$ en $\underline{v}$ afleiden is mogelijk met de, nog te behandelen, transformatiemethode. De verdelingsfunctie-methode is echter algemener bruikbaar. Het beschreven procédé is ook bruikbaar bij meer-dimensionale stochastische vectoren. De visuele steun die de grafiek in twee dimensies kan geven moeten we dan echter ontberen.

Voorbeeld

De stochastische variabelen $\underline{x}_i$, $i = 1,2,\ldots,n$ zijn continu, onafhankelijk en identiek verdeeld. Gevraagd de verdeling van de grootste van die n variabelen.
Zij $\underline{z} = \max(\underline{x}_i)$ en $F_x(x) = P(\underline{x}_i \le x)$; $i = 1,2,\ldots,n$.

$$P(\underline{z} \le z) = P(\underline{x}_1 \le z, \underline{x}_2 \le z, \ldots, \underline{x}_n \le z) =$$

$$= P(\underline{x}_1 \le z) \cdot P(\underline{x}_2 \le z) \ldots P(\underline{x}_n \le z) = \{F_x(z)\}^n$$

zodat $F_z(z) = \{F_x(z)\}^n$ en $f_z(z) = n\{F_x(z)\}^{n-1} f_x(z)$. □

B. De transformatie-methode

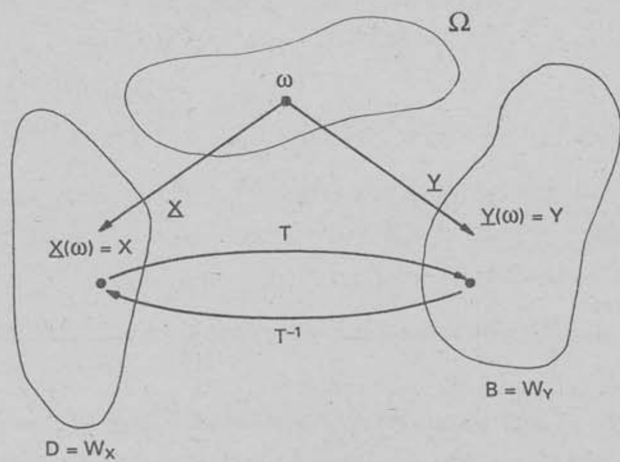
Stel de stochastische variabelen $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ hebben een simultane verdeling en de transformatie T van $D \subset \mathbb{R}^n$ naar $B \subset \mathbb{R}^n$ heeft een (éénduidige) inverse T^{-1} van

B naar D, m.a.w. er zijn 2n functies zó dat

$$y_i = g_i(X),\ i = 1,2,\ldots,n;\ \text{X een element van D;}$$

en $$x_i = h_i(Y),\ i = 1,2,\ldots,n;\ \text{Y een element van B.}$$

Verzamel de stochastische variabelen $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ in de stochastische vector $\underline{X}$, en laat de verdeling van $\underline{Y} = T(\underline{X})$ gevraagd zijn.
Kies D gelijk aan het waardenbereik W_X van de stochastische vector $\underline{X}$. Als $\underline{X}$



Figuur 4.17.

discreet is zal ook $\underline{Y}$ discreet zijn en omdat de transformatie één-éénduidig is correspondeert er met elk punt $X \in D = W_X$ precies één punt, $Y = T(X) \in B$ waar de kansfunctie van $\underline{Y} \neq 0$, en vice versa. Dus wordt B gelijk aan W_Y. Nu geldt: $P\{\underline{Y} = Y\} = P\{\underline{X} = T^{-1}(Y)\}$ en die laatste kans kan worden bepaald m.b.v. de kansfunctie van $\underline{X}$.

Voorbeeld

Vergelijk het voorbeeld op bladzijde 87.
$\underline{x}$ en $\underline{y}$ uniform op $\{1,2,\ldots,6\}$ en onafhankelijk.

$$\underline{u} = \underline{x} + \underline{y};\quad \underline{v} = \underline{x} - \underline{y}.$$

De transformatie T wordt gegeven door $\begin{cases} u = x + y; \\ v = x - y. \end{cases}$

De inverse transformatie T^{-1} is dus $\begin{cases} x = \frac{1}{2}(u + v); \\ y = \frac{1}{2}(u - v). \end{cases}$

$$P\{\underline{u}=u; \underline{v}=v\} = P\{\underline{x}=\tfrac{1}{2}(u+v); \underline{y}=\tfrac{1}{2}(u-v)\} =$$

$$= \frac{1}{36}, \qquad \text{als } (\tfrac{1}{2}(u+v), \tfrac{1}{2}(u-v)) \in \{1,2,\dots,6\}^2;$$

$$= 0, \qquad \text{elders}$$

□

Als $\underline{X}$ continu is en de transformatie één-éénduidig zal ook $\underline{Y}$ continu zijn. Zij A een gebied in B dan zal

$$P\{\underline{Y} \in A\} = \int \underset{A}{\cdots} \int f_Y(y_1,y_2,\dots,y_n)dy_1dy_2\dots dy_n$$

maar ook

$$P\{\underline{Y} \in A\} = P\{\underline{X} \in T^{-1}(A)\} = \int \underset{T^{-1}(A)}{\cdots} \int f_X(x_1,x_2,\dots,x_n)dx_1dx_2\dots dx_n$$

waarbij $T^{-1}(A) \in D$. Om de substitutie-methode voor het berekenen van meervoudige integralen (zie hoofdstuk 8 *Analyse*, J.H.J. Almering e.a., DUM 1991) te mogen toepassen moeten we nog eisen dat de partiële afgeleiden $\frac{\partial x_i}{\partial y_j}$, $i = 1,2,\dots,n$; $j = 1,2,\dots,n$ continu zijn binnen het beschouwde gebied. Als dat het geval is wordt

$$\int \underset{(x_1,\dots,x_n) \in T^{-1}(A)}{\cdots} \int f_X(x_1,x_2,\dots,x_n)dx_1dx_2\dots dx_n =$$

$$= \int \underset{(y_1,\dots,y_n) \in A}{\cdots} \int f_X(h_1(y_1,\dots,y_n),\dots,h_n(y_1,\dots,y_n))\cdot|J|\cdot dy_1\dots dy_n$$

waarbij

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_1} \\ \vdots & & \vdots \\ \frac{\partial x_1}{\partial y_n} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix},$$

de functionaal-determinant of determinant van Jacobi. De methode wordt daarom ook vaak aangeduid als de methode met de Jacobiaan.
Dus volgt dat de dichtheid van $\underline{Y}$ gelijk is aan $|J|$ maal de dichtheid van $\underline{X}$, waarin $x_1,x_2,\dots,x_n$ zijn vervangen door $h_i(y_1,y_2,\dots,y_n)$ voor $i = 1,2,\dots,n$; ofwel

$$f_Y(y_1,y_2,\dots,y_n) = f_X\{h_1(y_1,y_2,\dots,y_n),\dots,h_n(y_1,y_2,\dots,y_n))\}\cdot|J|.$$

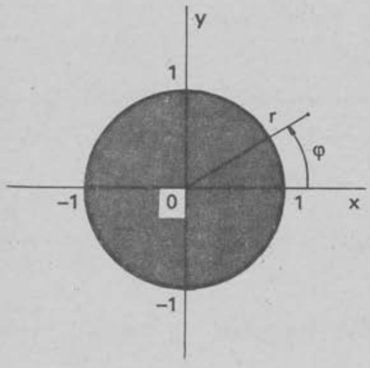
Voorbeeld

$\underline{x}$ en $\underline{y}$ zijn uniform verdeeld op de eenheidscirkel.
$\underline{r} = \sqrt{\underline{x}^2 + \underline{y}^2}$; $\underline{\varphi} = \arctan(\underline{y}/\underline{x})$.
Gevraagd de simultane dichtheid van $\underline{r}$ en $\underline{\varphi}$.

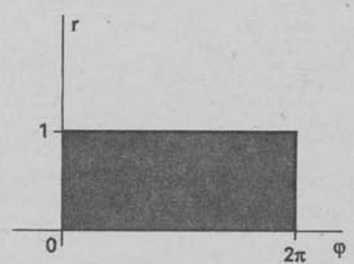
Dit is de bekende transformatie van rechthoekige naar poolcoördinaten. De inverse transformatie is

$$\left.\begin{matrix} x = r\cos\varphi \\ y = r\sin\varphi \end{matrix}\right\} \text{met } 0 \le r \le 1 \text{ en } 0 \le \varphi < 2\pi.$$

Afgezien van het punt x = 0, y = 0 ofwel r = 0, φ willekeurig is dit een één-éénduidige transformatie. Een dergelijk gebied met oppervlak nul kan echter geen roet in het eten gooien. (Analoog mag het in een drie-dimensionale context op een vlak misgaan.)



Figuur 4.18.



Figuur 4.19.

$$J = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \varphi} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \varphi} \end{vmatrix} = \begin{vmatrix} \cos\varphi & -r\sin\varphi \\ \sin\varphi & r\cos\varphi \end{vmatrix} = (r\cos^2\varphi + r\sin^2\varphi) = r.$$

Maar $r \ge 0$ dus $|J| = J = r$.

$$\begin{aligned} f_{x,y}(x,y) &= \tfrac{1}{\pi}, && (x,y) \text{ in de eenheidscirkel;} \\ &= 0, && \text{elders.} \\ f_{r,\varphi}(r,\varphi) &= \tfrac{1}{\pi}\cdot r, && 0 \le r \le 1;\ 0 \le \varphi < 2\pi; \\ &= 0, && \text{elders.} \end{aligned}$$

De marginale dichtheid van $\underline{r}$ wordt

$$f_r(r) = \int_{-\infty}^{+\infty} f_{r,\varphi}(r,\varphi)\,d\varphi = \int_0^{2\pi} \frac{r}{\pi}\,d\varphi = 2r, \quad 0 \le r \le 1;$$

$$= 0, \quad \text{elders.}$$

De marginale dichtheid van $\underline{\varphi}$:

$$f_\varphi(\varphi) = \int_0^1 \frac{r}{\pi}\,dr = \frac{1}{2\pi}, \qquad 0 \le \varphi < 2\pi;$$

$$= 0, \qquad \text{elders.}$$

Het blijkt dus dat $\underline{r}$ en $\underline{\varphi}$ onafhankelijk zijn.

Voor het gebruik van deze methode is het nodig dat het aantal functies $\underline{y}_i$ even groot is als het aantal stochastische variabelen $\underline{x}_i$ waarvan men uitgaat. Wil men de simultane dichtheid van m (< n) functies van $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ bepalen dan moeten er eerst n – m functies van $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ aan toegevoegd worden. Dan kan m.b.v. de transformatie-methode de simultane dichtheid van dat n-tal functies worden bepaald. Door weg-integreren van de overtollige variabelen ontstaat dan de gevraagde simultane dichtheid van de m functies als m-dimensionale marginale dichtheid. Het is raadzaam om voor de extra functies een zo eenvoudig mogelijke keuze te doen.

Voorbeeld

$\underline{x}, \underline{y}$ en $\underline{z}$ zijn onafhankelijk en exponentieel verdeeld met parameter λ. Gevraagd de dichtheid van $\underline{u} = \underline{x} + \underline{y} + \underline{z}$.

Er moeten twee functies worden verzonnen, neem $\underline{v} = \underline{x}$ en $\underline{w} = \underline{x} + \underline{y}$.

De inverse transformatie wordt dan:

$x = v;\ \ y = w - v;\ \ z = u - w,$

en de Jacobiaan:

$$J = \begin{vmatrix} 0 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 0 & -1 \end{vmatrix} = 1 \text{ zodat } |J| = 1.$$

De simultane dichtheid van $\underline{x}$, $\underline{y}$ en $\underline{z}$ is

$$f_X(x,y,z) = \lambda^3 e^{-\lambda(x+y+z)}, \quad x > 0;\ y > 0;\ z > 0;$$

$$= 0, \qquad \text{elders.}$$

zodat volgt

$$f_U(u,v,w) = \lambda^3 e^{-\lambda u}, \qquad u > w;\ w > v;\ v > 0;$$

$$= 0, \qquad \text{elders.}$$

N.B.: Let op dat ook het gebied voor $f_U \neq 0$ m.b.v. de nieuwe variabelen moet worden beschreven.

Nodig is nu de marginale dichtheid van $\underline{u}$, dus

$$\int\limits_{w}\int\limits_{v} f_U(u,v,w)\,dvdw = \int\limits_{w=0}^{u}\int\limits_{v=0}^{w} \lambda^3 e^{-\lambda u} dvdw = \int\limits_{w=0}^{u} w\lambda^3 e^{-\lambda u} dw =$$

$$= \frac{1}{2} u^2\lambda^3 e^{-\lambda u}, \quad u > 0;$$

$$= 0, \qquad \text{elders.}$$

Deze methode is natuurlijk ook bruikbaar om de dichtheid van een continue monotone functie $\underline{y}$ van een continue stochastische variabele $\underline{x}$ te bepalen. De Jacobiaan reduceert dan tot $\left|\frac{dx}{dy}\right|$.

Als de transformatie T geen éénduidige inverse heeft kan de methode nog gebruikt worden nadat het domein van T is gesplitst in delen voor elk waarvan T^{-1} wel eenduidig is. Het optellen van de zo verkregen dichtheid-achtige functies levert de gevraagde dichtheid.

Voorbeeld

$\underline{x} \sim N(0;1)$. Bepaal de dichtheid van $\underline{y} = \underline{x}^2$.
Bij $y = x^2$ horen $x = \sqrt{y}$ en $x = -\sqrt{y}$ als 'inverses'. Door het domein $\mathbb{R}$ te splitsen in $D_1 = \mathbb{R}^+$ en $D_2 = \mathbb{R}^-$ ontstaan twee één-éénduidige transformaties.
Op D_1 geldt $x = \sqrt{y}$ dus $|J| = \left|\frac{d\sqrt{y}}{dy}\right| = \frac{1}{2\sqrt{y}}$ en $f_x(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$ zodat daar

$$f_y^{o}(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y} \cdot \frac{1}{2\sqrt{y}}, \quad y > 0.$$

Op D_2 geldt $x = -\sqrt{y}$ dus $|J| = \left|\frac{d-\sqrt{y}}{dy}\right| = \frac{1}{2\sqrt{y}}$ en $f_x(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$

zodat ook hier volgt

$$f_y^{*}(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y} \cdot \frac{1}{2\sqrt{y}}, \quad y > 0.$$

De dichtheid van y volgt door het optellen van f^o en f^* zodat volgt:

$$f_y(y) = 0, \qquad y < 0;$$

$$= \frac{1}{\sqrt{2\pi y}} e^{-\frac{1}{2}y}, \qquad y > 0.$$

Dit resultaat kan echter eenvoudiger m.b.v. de verdelingsfunctie-methode verkregen worden.

$$F_y(y) = P(\underline{y} < y) = P(-\sqrt{y} < \underline{x} < \sqrt{y}) = F_x(\sqrt{y}) - F_x(-\sqrt{y});$$

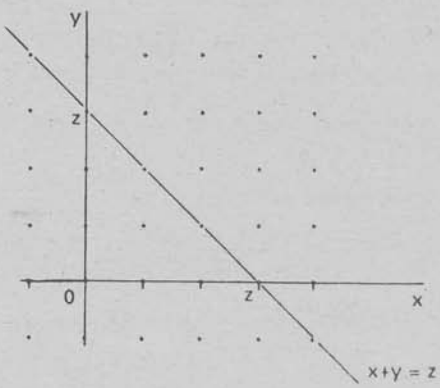
$$f_y(y) = \frac{d}{dy} F_y(y) = f_x(\sqrt{y})\cdot\frac{1}{2\sqrt{y}} - f_x(-\sqrt{y})\cdot\frac{-1}{2\sqrt{y}} = \frac{1}{\sqrt{y}} f_x(\sqrt{y}).$$

4.8. Convolutie

Zeer vaak is de verdeling van de som, $\underline{z}$, van twee (of meer) onafhankelijke stochastische variabelen, $\underline{x}$ en $\underline{y}$, nodig. De verdeling van $\underline{z}$ heet de *convolutie* van de verdeling van $\underline{x}$ en die van $\underline{y}$.
Zijn $\underline{x}$ en $\underline{y}$ discreet met kansfunctie p_x, respectievelijk p_y dan is ook $\underline{z}$ discreet.

$$P(\underline{z} = z) = P(\underline{x} + \underline{y} = z) = \sum_{x_i} P(\underline{x} = x_i;\ \underline{y} = z - x_i) = \sum_{x_i} p_x(x_i) \cdot p_y(z - x_i).$$

Men hoeft dus slechts de kansen op te tellen behorend bij punten die liggen op de lijn $x + y = z$.



Figuur 4.20.

Voorbeeld

$\underline{x}$ en $\underline{y}$ zijn onafhankelijk en beide Poisson-verdeeld, met parameter λ, resp. μ.
Gevraagd de verdeling van $\underline{z} = \underline{x} + \underline{y}$.

$$p_x(k) = \frac{\lambda^k}{k!} e^{-\lambda},\ k = 0,1,2,\ldots;$$

$$p_y(m) = \frac{\mu^m}{m!} e^{-\mu},\ m = 0,1,2,\ldots,$$

m.a.w. de kansmassa ligt geheel in het eerste kwadrant. Voor $n = 0,1,2,\ldots$ volgt:

$$p_z(n) = P\{\underline{x} + \underline{y} = n\} = \sum_{i=0}^{n} p_x(i)\, p_y(n-i) = \sum_{i=0}^{n} \frac{\lambda^i \mu^{n-i}}{i!(n-i)!} e^{-(\lambda+\mu)} =$$

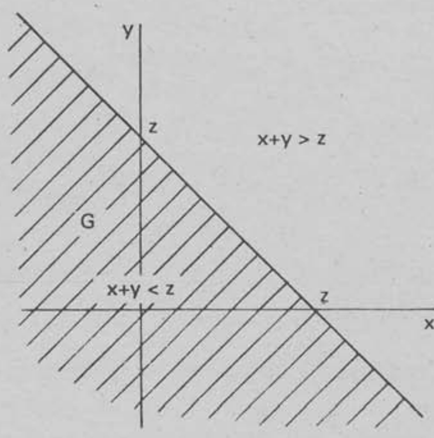
$$= \frac{1}{n!} e^{-(\lambda+\mu)} \sum_{i=0}^{n} \binom{n}{i} \lambda^i \mu^{n-i} = \frac{(\lambda+\mu)^n}{n!} e^{-(\lambda+\mu)}.$$

Dus $\underline{z}$ is Poisson-verdeeld met als parameter de som van de parameters.

In dit voorbeeld bleek dat de som van twee onafhankelijke Poisson-verdeelde stochastische variabelen weer Poisson-verdeeld is. Dat laat zich generaliseren tot:
De som van n onafhankelijke Poisson-verdeelde stochastische variabelen met parameters resp. $\lambda_1,\lambda_2,\ldots,\lambda_n$, is Poisson-verdeeld met parameter

$$\lambda = \sum_{i=1}^{n} \lambda_i.$$

Zijn $\underline{x}$ en $\underline{y}$ continu en onafhankelijk dan is ook $\underline{z} = \underline{x} + \underline{y}$ een continue stochastische variabele.



Figuur 4.2↑.

$$F_z(z) = P(\underline{z} \le z) = P\{\underline{x} + \underline{y} \le z\} = \iint_G f_{x,y}(x,y)\,dxdy =$$

$$= \int_{y=-\infty}^{\infty} \int_{x=-\infty}^{z-y} f_x(x)f_y(y)\,dxdy = \int_{y=-\infty}^{\infty} f_y(y)\cdot F_x(z-y)dy.$$

$$f_z(z) = \frac{d}{dz}F_z(z) = \frac{d}{dz}\int_{-\infty}^{\infty} f_y(y)\cdot F_x(z-y)dy = \int_{-\infty}^{\infty} f_y(y)f_x(z-y)dy.$$

Het resultaat kan natuurlijk ook als

$$f_z(z) = \int_{-\infty}^{\infty} f_x(x)f_y(z-x)dx$$

worden geschreven.
Weer wordt alleen de lijn x + y = z afgezocht. In elk punt van die lijn wordt het produkt van de marginale dichtheden bepaald en dan wordt alles bij elkaar geveegd door de integraal.

Als $\underline{x}$ en $\underline{y}$ niet onafhankelijk zijn hoeft $\underline{z} = \underline{x} + \underline{y}$ niet continu te zijn. Zij $\underline{x}$ een te meten continue grootheid en $\underline{y}$ de afrondfout bij meting. $\underline{x}$ en $\underline{y}$ zijn continu maar $\underline{z} = \underline{x} + \underline{y}$, het resultaat van de meting, is discreet.

Voorbeeld

$\underline{x}$ en $\underline{y}$ zijn onafhankelijk en beide N(0;1)-verdeeld. De som $\underline{z} = \underline{x} + \underline{y}$ is dan N(0;2)-verdeeld.

$$f_z(z) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-x)^2}\, dx =$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{4}z^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(2x^2-2zx+\frac{1}{2}z^2)}\, dx =$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{4}z^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x\sqrt{2}-z/\sqrt{2})^2} \cdot \frac{1}{\sqrt{2}}\, d(x\sqrt{2}) =$$

$$= \frac{1}{\sqrt{2}\sqrt{2\pi}} e^{-\frac{1}{2}(z/\sqrt{2})^2} \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x\sqrt{2}-z/\sqrt{2})^2}\, d(x\sqrt{2})$$

en de laatste integraal is gelijk aan één, want het is de integraal van een Normale dichtheid van $-\infty$ tot $+\infty$.
Het resultaat is de dichtheid van een N(0;2)-verdeelde stochastische variabele. □

Op soortgelijke wijze als in het voorbeeld kan worden aangetoond:
als $\underline{x}$ en $\underline{y}$ onafhankelijk zijn en $\underline{x} \sim N(\mu_1;\sigma_1^2)$ en $\underline{y} \sim N(\mu_2;\sigma_2^2)$ dan heeft $\underline{z} = \underline{x} + \underline{y}$ een Normale verdeling met verwachting $\mu_1 + \mu_2$ en variantie $\sigma_1^2 + \sigma_2^2$.
Als $\underline{y} \sim N(\mu_2;\sigma_2^2)$ dan is $-\underline{y} \sim N(-\mu_2;\sigma_2^2)$ zodat de verdeling van $\underline{w} = \underline{x} - \underline{y}$ ook Normaal is en wel $N(\mu_1-\mu_2;\ \sigma_1^2+\sigma_2^2)$.
Dit kan worden gegeneraliseerd tot:
Zijn $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ onafhankelijk en resp. $N(\mu_i;\sigma_i^2)$-verdeeld dan heeft $\underline{s} = \sum_{i=1}^{n} a_i \underline{x}_i$ een Normale verdeling met verwachting $\sum_{i=1}^{n} a_i\mu_i$ en variantie $\sum_{i=1}^{n} a_i^2\sigma_i^2$.

4.9. De negatief-binomiale verdeling

Soms biedt het voordelen om, ook als men maar in één stochastische variabele is geïnteresseerd, een omweg via een groter aantal, geschikt gekozen, stochastische variabelen te maken. Bijvoorbeeld bij een negatief-binomiaal verdeelde stochastische variabele.

Als we een Bernoulli-experiment, met kans p op succes, herhalen tot en met het optreden van het r-de succes dan is het aantal proeven, $\underline{n}$, negatief-binomiaal verdeeld met parameters r en p.

$$P\{\underline{n} = n\} = 0 \text{ voor } n \neq r, r+1, r+2, \dots .$$

Voor $n = r, r+1, r+2, \dots$ geldt:

$$\begin{aligned} P\{\underline{n} = n\} &= P\{r-1 \text{ successen in de eerste } n-1 \text{ proeven} \cap n\text{-de proef succes}\} \\ &= \binom{n-1}{r-1} \cdot p^{r-1} \cdot q^{n-r} \cdot p = \binom{n-1}{r-1} \cdot p^{r} \cdot q^{n-r}. \end{aligned}$$

Het rechtstreeks uitrekenen van $E(\underline{n})$ en $\sigma^2(\underline{n})$ is vervelend. Bekijk de reeks proeven (– = mislukking; + = succes):

$$\underbrace{---\overset{1}{+}}_{\underline{x}_1}\ \ \underbrace{--\overset{2}{+}}_{\underline{x}_2}\ \underset{\underline{x}_3}{\overset{3}{+}}\ \dots\ \overset{r-1}{+}\ \ \ \underbrace{----\overset{r}{+}}_{\underline{x}_r}$$

Stel $\underline{n} = \underline{x}_1 + \underline{x}_2 + \dots + \underline{x}_r$, waarbij $\underline{x}_i$ het aantal proeven na het i–1-de succes tot en met het i-de succes voorstelt.

De stochastische variabelen $\underline{x}_i$, $i = 1,2,\dots,r$ zijn onafhankelijk en elk geometrisch verdeeld met parameter p, dus

$$E(\underline{x}_i) = \frac{1}{p} \quad \text{en} \quad \sigma^2(\underline{x}_i) = \frac{q}{p^2} .$$

$$E(\underline{n}) = \sum_{i=1}^{r} E(\underline{x}_i) = \frac{r}{p};$$

$$\sigma^2(\underline{n}) = \sum_{i=1}^{r} \sigma^2(\underline{x}_i) = \frac{rq}{p^2} .$$

Ook verwachting en variantie van de binomiale verdeling kunnen langs deze weg eenvoudig worden afgeleid uit de verwachting en de variantie van de Bernoulli-verdeling.

5

S

5.1

Sta
uits
ver
Bij
Naa
van
de
pop
uits
stat
van
dan

5.2

Als
ove
die
maa
stee
sen

Vo
Een
pop

Vo
De
Lei
trei
die
bijv

5 Statistiek

5.1. Inleiding

Statistiek is ouder dan kansrekening. In het begin werd echter onder statistiek uitsluitend begrepen dat wat nu beschrijvende statistiek wordt genoemd: het vergaren en op overzichtelijke wijze presenteren van numerieke gegevens. Bijvoorbeeld het aantal paarden per district in een land op een bepaald tijdstip. Naarmate de te beschrijven populaties groter werden gingen niet alleen de kosten van een dergelijke beschrijving omhoog maar daalde ook de betrouwbaarheid van de resultaten. Daarom is men ertoe overgegaan slechts een gedeelte van de populatie, een steekproef, te onderzoeken en op grond van de resultaten daarvan uitspraken te doen over de hele populatie. Dat is de voorspellende, of inductieve, statistiek waarmee we ons nu verder bezig zullen houden. Voor het goed bedrijven van voorspellende statistiek is kansrekening nodig. De voorspellende statistiek is dan ook jonger dan de kansrekening.

5.2. Steekproef en populatie

Als men uit een populatie een steekproef trekt met de bedoeling om uitspraken over die populatie te baseren op dat wat men aantreft in de steekproef, dan moet die steekproef representatief zijn voor de populatie. Dit is voor de hand liggend maar in de praktijk vaak moeilijker te verwezenlijken dan het lijkt. Omdat bij elke steekproef ook wel een populatie te vinden is waarvoor die steekproef representatief is zullen we hier verder geen aandacht aan besteden.

Voorbeeld

Een worp met een dobbelsteen is een representatieve steekproef uit de oneindige populatie van alle worpen die met *die* dobbelsteen mogelijk zijn. □

Voorbeeld

De reizigers die op woensdag om 8.11 uur op Rotterdam CS in de trein richting Leiden stappen vormen geen representatieve steekproef voor de populatie van treinreizigers in Nederland, zelfs niet voor de veel kleinere populatie van reizigers die op die woensdag in Rotterdam in een trein richting Leiden stappen. (Denk bijvoorbeeld aan de 60-plussers die pas om negen uur goed op drift raken.) □

Een steekproef ter grootte n heet *aselect* als elke groep van n elementen uit de populatie dezelfde kans heeft om getrokken te worden. Als de populatie eindig is, ter grootte N, en het trekken gebeurt zonder teruglegging, kan men op $\binom{N}{n}$ manieren een steekproef ter grootte n daaruit trekken. Daarbij wordt dus niet op de volgorde binnen de steekproef gelet. Doet men dat wel, waarbij men spreekt van een geordende steekproef, dan zijn er $\frac{N!}{(N-n)!}$ manieren.

Trekt men met teruglegging n keer uit een eindige populatie N dan zijn er N^n geordende steekproeven mogelijk. Het aantal verschillende steekproeven als niet op de volgorde wordt gelet is nu gelijk aan $\binom{N+n-1}{n}$.

Bij het trekken uit een oneindige populatie zijn er natuurlijk oneindig veel verschillende steekproeven mogelijk.
Bij een oneindige, of een zeer grote eindige populatie maakt het geen verschil of men mèt dan wel zònder teruglegging trekt. In een populatie zo vol geladen mist men één, twee elementen niet.
Als het kenmerk van de elementen uit de populatie, waarin men geïnteresseerd is m.b.v. een getal kan worden vastgelegd dan zijn stochastische variabelen handig om de gang van zaken te beschrijven.
Zij $\underline{x}$ de stochastische variabele die aangeeft welke waarde het kenmerk heeft bij een willekeurig uit de populatie getrokken element. De verdelingsfunctie, F_x, van $\underline{x}$ volgt uit: $P(\underline{x} \le x) = v(x)$ = de fractie van het aantal elementen in de populatie waarvoor $\underline{x} \le x$. Helaas is daarmee eigenlijk alleen gezegd dat er een verdelingsfunctie van $\underline{x}$ bestaat, want de fractie $v(x)$ is onbekend. We werken precies andersom. Het trekken van de steekproef dient om voldoende informatie te verzamelen om aan $v(x)$ een redelijk betrouwbare waarde toe te kennen. Dat komt dus neer op het bepalen van F_x. Als bekend verondersteld wordt wat voor verdeling $\underline{x}$ heeft, hoeven alleen de parameters bepaald te worden. Is de stochastische variabele discreet (continu) dan spreekt men van een discrete (continue) populatie. Een continue populatie moet oneindig groot zijn.
Een aselecte steekproef uit een oneindige populatie, of uit een eindige maar dan met teruglegging, kan worden beschouwd als het n maal onafhankelijk herhalen van hetzelfde experiment. Met die n trekkingen associëren we de n stochastische variabelen $\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_n$. Deze n stochastische variabelen zijn onafhankelijk en hebben alle dezelfde verdelingsfunctie F_x. Na inspectie van de steekproef liggen de n realisaties $x_1,x_2,\ldots,x_n$ van $\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_n$ voor ons. Alleen uit die realisaties kunnen we de informatie verzamelen die we nodig hebben.

5.3. Gemiddelde en variantie van een steekproef

Van nu af aan zullen we spreken van een populatie F_x als de stochastische variabele $\underline{x}$ voor een willekeurig element van de populatie de geheel of gedeeltelijk onbekende verdelingsfunctie F_x heeft. Zo heeft $\underline{x}$ een Normale verdeling in een Normale populatie.
Onder een aselecte steekproef ter grootte n (of: met omvang n) zal, tenzij anders blijkt, een steekproef worden verstaan uit een oneindige populatie of uit een eindige met teruglegging. Zo'n steekproef is dus een n-tal onafhankelijke, identiek verdeelde stochastische variabelen $\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_n$ tot aan het moment waarop hij werkelijk getrokken is. Na het trekken hebben we de n getallen $x_1,x_2,\ldots,x_n$, de realisaties van de stochastische variabelen. Een steekproeffunctie (statistic), $\underline{z}$, is een functie van $\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_n$ waarvan we de realisatie kunnen bepalen als de steekproef is getrokken. $\underline{z}$ mag dus geen onbekende parameters bevatten.
De belangrijkste twee steekproeffuncties zijn het steekproefgemiddelde en de steekproefvariantie.

Definitie

Het gemiddelde $\overline{\underline{x}}_n$ van een steekproef $\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_n$ is de steekproeffunctie

$$\overline{\underline{x}}_n = \frac{1}{n}\sum_{i=1}^{n}\underline{x}_i. \qquad \square$$

Stelling 5.1. In een aselecte steekproef uit een populatie F_x geldt dat:

$$E(\overline{\underline{x}}_n) = E(\underline{x}) \quad \text{en} \quad \sigma^2(\overline{\underline{x}}_n) = \frac{1}{n}\sigma^2(\underline{x}).$$

Bewijs: $$E(\overline{\underline{x}}_n) = \frac{1}{n}\sum_{i=1}^{n}E(\underline{x}_i) = \frac{1}{n}\sum_{i=1}^{n}E(\underline{x}) = E(\underline{x});$$

$$\sigma^2(\overline{\underline{x}}_n) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2(\underline{x}_i) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2(\underline{x}) = \frac{1}{n}\sigma^2(\underline{x}). \qquad \square$$

Het gemiddelde van een steekproef is een voortreffelijke steekproeffunctie als het er om gaat informatie over de verwachting van $\underline{x}$ (in de populatie) te vergaren.

Definitie

De variantie $\underline{s}^2$ van een steekproef $\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_n$ is de steekproeffunctie

$$\underline{s}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\underline{x}_i - \overline{\underline{x}})^2. \qquad \square$$

Stelling 5.2. In een aselecte steekproef ter grootte n uit een populatie F_x geldt dat:

$$E(\underline{s}^2) = \sigma^2(\underline{x}).$$

Bewijs: Zij $E(\underline{x}_i) = E(\underline{x}) = \mu$ en $\sigma^2(\underline{x}_i) = \sigma^2(\underline{x}) = \sigma^2$.

$$\begin{aligned} E(\underline{s}^2) &= \frac{1}{n-1}\sum_{i=1}^{n} E\{((\underline{x}_i-\mu) - (\overline{\underline{x}}-\mu))^2\} = \\ &= \frac{1}{n-1}\sum_{i=1}^{n} [E\{(\underline{x}_i-\mu)^2\} + E\{(\overline{\underline{x}}-\mu)^2\} - 2E\{(\underline{x}_i-\mu)(\overline{\underline{x}}-\mu)\}] = \\ &= \frac{1}{n-1} [(n+1)\sigma^2 - 2E\{(\overline{\underline{x}}-\mu)\sum_{i=1}^{n}(\underline{x}_i-\mu)\}] = \\ &= \frac{1}{n-1} [(n+1)\sigma^2 - 2nE\{(\overline{\underline{x}}-\mu)^2\}] = \\ &= \frac{1}{n-1} \{(n+1)\sigma^2 - 2\sigma^2\} = \sigma^2(\underline{x}). \qquad \square \end{aligned}$$

Het bepalen van de variantie van $\underline{s}^2$ is mogelijk maar omslachtig. We zullen dat in de volgende paragraaf doen maar alleen voor een Normale populatie.
N.B. In sommige boeken wordt een afwijkende definitie van de steekproefvariantie gebruikt, namelijk $\frac{n-1}{n} \cdot \underline{s}^2$.

Definitie
De steekproefstandaardafwijking $\underline{s} = \sqrt{\underline{s}^2}$. ❑
N.B. De verwachting van $\underline{s}$ is niet gelijk aan $\sigma(\underline{x})$.
De naam van $\underline{s}^2$ en de functie die $\underline{s}^2$ vastlegt duiden aan dat de realisaties van $\underline{s}^2$ gebruikt worden om een idee te geven van de grootte van de populatievariantie $\sigma^2(\underline{x})$.

5.4. Andere steekproeffuncties: χ^2, t en F

A. De χ^2-verdeling

Definitie
Een stochastische variabele $\underline{z}$ heeft een χ^2-verdeling met k vrijheidsgraden als $\underline{z}$ te schrijven is als

$$\underline{z} = \underline{x}_1^2 + \underline{x}_2^2 + \ldots + \underline{x}_k^2,$$

waarbij de $\underline{x}_i$, i = 1,2,…,k onafhankelijk zijn en N(0;1)-verdeeld. ❑
Notatie: $\underline{z} \sim \chi^2(k)$.

Stelling 5.3. Als $\underline{z} \sim \chi^2(k)$ dan heeft $\underline{z}$ als dichtheid

$$f_z(z) = \frac{z^{\frac{1}{2}k-1} e^{-z/2}}{\Gamma(\frac{1}{2}k)2^{k/2}}, \qquad z > 0;$$
$$= 0, \qquad z < 0.$$

Bewijs: Met volledige inductie. Voor k = 1 is de bewering juist (zie het voorbeeld op blz. 93). Neem aan dat de bewering juist is voor k = n dus dat de dichtheid van

$\underline{w} = \sum_{i=1}^{n} \underline{x}_i^2$ gelijk is aan

$$f_w(w) = \frac{w^{\frac{1}{2}n-1} e^{-w/2}}{\Gamma(\frac{1}{2}n)2^{n/2}}, \qquad w > 0.$$

De dichtheid van $\underline{v} = \sum_{i=1}^{n+1} \underline{x}_i^2 = \underline{w} + \underline{x}_{n+1}^2$ volgt dan uit de convolutie van f_w met de dichtheid van $\underline{x}_{n+1}^2$, want $\underline{w}$ en $\underline{x}_{n+1}$ zijn onafhankelijk. Dus voor v > 0:

$$f_v(v) = \int_0^v \frac{w^{\frac{1}{2}n-1} e^{-w/2}}{\Gamma(\frac{1}{2}n)2^{n/2}} \cdot \frac{(v-w)^{-\frac{1}{2}} e^{-(v-w)/2}}{\sqrt{2\pi}} \, dw =$$

$$= \frac{e^{-v/2}}{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2}n)2^{(n+1)/2}} \int_0^v w^{\frac{1}{2}n-1} (v-w)^{\frac{1}{2}-1} dw.$$

$$\int_0^v w^{\alpha-1}(v-w)^{\beta-1} dw = v^{\alpha+\beta-1} \int_0^1 t^{\alpha-1}(1-t)^{\beta-1} dt =$$

$$= v^{\alpha+\beta-1} \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \text{ voor } \alpha > 0 \text{ en } \beta > 0.$$

Met dit resultaat volgt dat

$$f_v(v) = \frac{v^{\frac{1}{2}(n+1)-1} e^{-v/2}}{\Gamma(\frac{1}{2}(n+1))2^{(n+1)/2}}, \qquad v > 0$$
$$= 0, \qquad v < 0,$$

waarmee het gestelde is bewezen.

Stelling 5.4. Als $\underline{z} \sim \chi^2(k)$ dan is

$$E(\underline{z}) = k \quad \text{en} \quad \sigma^2(\underline{z}) = 2k.$$

Bewijs: $E(\underline{z}) = \sum_{i=1}^{k} E(\underline{x}_i^2) = kE(\underline{x}_1^2) = k$, want de $\underline{x}_i$ zijn N(0;1)-verdeeld zodat

$$E(\underline{x}_1^2) = \sigma^2(\underline{x}_1) = 1.$$

$$\sigma^2(\underline{z}) = \sum_{i=1}^{k} \sigma^2(\underline{x}_i^2) = k\sigma^2(\underline{x}_1^2).$$

$$\sigma^2(\underline{x}_1^2) = E\{(\underline{x}_1^2)^2\} - \{E(\underline{x}_1^2)\}^2 =$$

$$= \int_0^\infty y^2 \frac{1}{\sqrt{2\pi y}} e^{-y/2} dy - 1 = \frac{4}{\sqrt{\pi}} \int_0^\infty \left(\frac{y}{2}\right)^{\frac{5}{2}-1} e^{-y/2} d\frac{y}{2} - 1 =$$

$$= \frac{4}{\sqrt{\pi}} \Gamma\left(\frac{5}{2}\right) - 1 = \frac{4}{\sqrt{\pi}} \cdot \frac{3}{2} \cdot \frac{1}{2} \cdot \sqrt{\pi} - 1 = 2.$$

Dus $\sigma^2(\underline{z}) = 2k$. ❑

Door een slimme transformatie kan men aantonen dat bij een steekproef, met omvang n, uit een Normale populatie $\overline{\underline{x}}$ en $\underline{s}^2$ onafhankelijk zijn en dat $\frac{(n-1)}{\sigma^2}\underline{s}^2$ in dat geval $\chi^2(n-1)$ verdeeld is, waarbij σ^2 de populatie-variantie is.

Dus is de variantie van $(n-1)\underline{s}^2/\sigma^2$ dan gelijk aan $2(n-1)$, waaruit volgt dat de variantie van $\underline{s}^2$ gelijk is aan $\frac{2\sigma^4}{n-1}$.

Ook kan men aantonen dat $\overline{\underline{x}}$ en $\underline{s}^2$ alleen bij een steekproef uit een Normale populatie onafhankelijk zijn. In alle andere gevallen treedt afhankelijkheid op, wel kunnen $\overline{\underline{x}}$ en $\underline{s}^2$ dan ongecorreleerd zijn.

Er bestaat een merkwaardig verband tussen de $\chi^2(2)$-verdeling en de exponentiële verdeling. De dichtheid van $\chi^2(2)$, $\frac{1}{2}e^{-y/2}$, $y > 0$, is ook de dichtheid van een exponentieel verdeelde stochastische variabele met verwachting 2.

Dus is ook de dichtheid van $\chi^2(2n)$ dezelfde als de dichtheid van de som van n onafhankelijke exponentieel verdeelde stochastische variabelen elk met verwachting 2, dat wil zeggen een Erlang-n verdeling (met verwachting 2n).

B. De t-verdeling

Definitie

Een stochastische variabele $\underline{t}$ die kan worden geschreven als

$$\underline{t} = \frac{\underline{u}}{\sqrt{\underline{v}/n}}$$

waarbij $\underline{u}$ en $\underline{v}$ onafhankelijk zijn en $\underline{u} \sim N(0;1)$ en $\underline{v} \sim \chi^2(n)$ heeft een t-verdeling met n vrijheidsgraden. Notatie $\underline{t} \sim t(n)$. □

Stelling 5.5. De dichtheid van een t(n)-verdeelde stochastische variabele $\underline{t}$ is

$$f_t(t) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{\pi n}} \frac{1}{(1+t^2/n)^{(n+1)/2}}, \; -\infty < t < \infty; \;\; n = 1,2,3,\ldots .$$

Bewijs: Gebruik de transformatie $\left\{ \begin{array}{l} t = u/\sqrt{v/n} \\ s = v \end{array} \right.$ met inverse $\left\{ \begin{array}{l} u = t\sqrt{s/n} \\ v = s \end{array} \right.$.

De simultane dichtheid van $\underline{u}$ en $\underline{v}$ is het produkt van de marginale dichtheden want $\underline{u}$ en $\underline{v}$ zijn onafhankelijk. De absolute waarde van de Jacobiaan wordt

$$\left| \frac{\partial v}{\partial s} \cdot \frac{\partial u}{\partial t} \right| = \sqrt{s/n} .$$

Dus wordt de simultane dichtheid van $\underline{s}$ en $\underline{t}$:

$$\left. \begin{array}{ll} f(s,t) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2 \cdot s/n} \cdot \dfrac{s^{\frac{1}{2}n-1} e^{-s/2}}{\Gamma(n/2)2^{n/2}} \cdot \sqrt{\dfrac{s}{n}}, & s > 0; \\ \quad = 0, & s < 0; \end{array} \right\} \text{voor alle t.}$$

We moeten de marginale dichtheid van $\underline{t}$ hebben.

$$f_t(t) = \frac{1}{\sqrt{2\pi n}\,\Gamma(n/2)2^{n/2}} \int_{s=0}^{\infty} s^{\frac{1}{2}(n+1)-1} e^{-(s/2)(1+t^2/n)} ds, \qquad \text{(A)}$$

met de substitutie $y = \frac{s}{2}(1+t^2/n)$ kan men de integraal schrijven als

$$\frac{2^{(n+1)/2}}{(1+t^2/n)^{(n+1)/2}} \int_0^\infty y^{\frac{1}{2}(n+1)-1} e^{-y} dy = \frac{2^{(n+1)/2}}{(1+t^2/n)^{(n+1)/2}} \Gamma(\frac{n+1}{2}).$$

Dit resultaat ingevuld in (A) levert, na wegstrepen van de overtollige factoren twee, het gevraagde resultaat. □

De steekproeffunctie $\underline{t}$ is nuttig bij steekproeven uit een Normale populatie. In dat geval zijn $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ onafhankelijk en $N(\mu;\sigma^2)$ verdeeld.

$\underline{\bar{x}}$ is dus ook Normaal verdeeld en wel met verwachting μ en variantie σ^2/n.

D.w.z. dat $\underline{u} = \dfrac{\underline{\bar{x}}-\mu}{\sigma/\sqrt{n}} \sim N(0;1)$. De steekproefvariantie $\underline{s}^2$ is onafhankelijk van $\underline{\bar{x}}$ dus ook $\underline{v} = \dfrac{(n-1)\underline{s}^2}{\sigma^2}$ en $\underline{u}$ zijn onafhankelijk. Daar $\underline{v} \sim \chi^2(n-1)$ volgt dat

$$\frac{\underline{u}}{\sqrt{\underline{v}/(n-1)}} = \frac{\underline{\bar{x}}-\mu}{\sigma/\sqrt{n}} \cdot \frac{\sigma}{\underline{s}} = \frac{\underline{\bar{x}}-\mu}{\underline{s}/\sqrt{n}} \sim t(n-1).$$

Van dit resultaat zullen we later dankbaar gebruik maken.

Stelling 5.6. Als $\underline{x} \sim t(n)$ dan is $E(\underline{x}) = 0$ voor $n > 1$, terwijl voor $n = 1$ de verwachting niet bestaat.

Bewijs: De dichtheid van $\underline{x}$ is symmetrisch t.o.v. 0 voor alle n. Als $E(\underline{x})$ bestaat is hij dus gelijk aan nul.

Voor $n = 1$ moet voor het bestaan van $E(\underline{x})$ $\int_a^\infty \frac{t}{1+t^2}\,dt$ eindig zijn, maar in de buurt van ∞ gedraagt de integrand zich als $\frac{1}{t}$, dus dat gaat niet goed.

Voor $n > 1$ wordt de integrand $\dfrac{t}{(1+t^2/n)^{(n+1)/2}}$ en die vertoont voor $t \to \infty$ hetzelfde gedrag als $\frac{1}{t^n}$, waarmee de convergentie gewaarborgd is. □

Analoog kan men aantonen dat $E(\underline{x}^k)$ slechts bestaat als $n > k$. Dus zal $\sigma^2(\underline{x})$ alleen bestaan voor $n > 2$. In dat geval geldt $\sigma^2(\underline{x}) = \frac{n}{n-2}$.
Voor $n \to \infty$ gaat dus $\sigma^2(\underline{x})$ naar 1. Bovendien kan worden aangetoond dat de t-verdeling nadert tot de standaard-Normale verdeling als n, het aantal vrijheidsgraden, naar oneindig gaat. De getallen op de onderste regel in de t-tabel, corresponderend met ∞ veel vrijheidsgraden, zijn dus bruikbaar voor het bepalen van kansen behorend bij de N(0;1)-verdeling.
De t-verdeling wordt vaak de Student-verdeling genoemd.

C. De F-verdeling

Definitie

Een stochastische variabele $\underline{F}$ die kan worden geschreven als

$$\underline{F} = \frac{\underline{v}/m}{\underline{w}/n},$$

waarbij $\underline{v}$ en $\underline{w}$ onafhankelijk zijn terwijl $\underline{v} \sim \chi^2(m)$ en $\underline{w} \sim \chi^2(n)$, heeft een F-verdeling met m en n vrijheidsgraden. Notatie $\underline{F} \sim F(m;n)$. □

Stelling 5.7. Als $\underline{F} \sim F(m;n)$ dan heeft $\underline{F}$ de dichtheid

$$f_F(x) = \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})}(\frac{m}{n})^{m/2} \cdot \frac{x^{\frac{m}{2}-1}}{(1+\frac{m}{n}x)^{(m+n)/2}}, \quad x > 0;$$

$$= 0, \qquad x < 0.$$

Bewijs: Met de verdelingsfunctie-methode. Zij $x > 0$, dan:

$$F_F(x) = P(\underline{F} \le x) = P(\underline{v} \le \frac{xm}{n}\underline{w}) = \int_{w=0}^{\infty}\int_{v=0}^{\frac{xmw}{n}} f_v(v)\, f_w(w)\, dvdw.$$

$$f_F(x) = \frac{d}{dx}F_F(x) = \int_{w=0}^{\infty} f_w(w)\, f_v(\frac{xmw}{n})\cdot\frac{mw}{n}\, dw =$$

$$= \int_{w=0}^{\infty} \frac{w^{\frac{1}{2}n-1}\cdot e^{-w/2}}{\Gamma(\frac{n}{2})\cdot 2^{n/2}} \cdot \frac{(xmw/n)^{\frac{1}{2}m-1} e^{-xmw/2n}}{\Gamma(\frac{m}{2})2^{m/2}} \cdot \frac{mw}{n}\, dw =$$

$$= \frac{x^{\frac{m}{2}-1}(m/n)^{m/2}}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})}\int_0^{\infty}(\frac{w}{2})^{\frac{n+m}{2}-1} e^{-(1+xm/n)w/2}\, d\,\frac{w}{2}.$$

De integraal wordt $\Gamma(\frac{n+m}{2})\cdot(1+\frac{xm}{n})^{-\frac{n+m}{2}}$ waarmee de stelling is bewezen. □

Als $\underline{x} \sim F(m;n)$ en $n > 2$ dan is $E(\underline{x}) = \frac{n}{n-2}$.

Als $\underline{t} \sim t(k)$ dan is $\underline{t}^2 \sim F(1;k)$, want $\underline{t}^2 = \frac{\underline{u}^2/1}{\underline{v}/k}$ waarbij $\underline{u} \sim N(0;1)$, dus $\underline{u}^2 \sim \chi^2(1)$ en $\underline{v} \sim \chi^2(k)$. Bovendien zijn teller en noemer onafhankelijk.

Als $\underline{x} \sim F(m;n)$ dan is $\frac{1}{\underline{x}} \sim F(n;m)$. Deze relatie is nodig omdat de tabellen van de F-verdeling alleen de percentielen in de rechterstaart van de verdeling geven.

Voorbeeld

$\underline{x} \sim F(5;10)$. Bepaal a en b zó dat $P(\underline{x} > a) = 0{,}05$ en $P(\underline{x} < b) = 0{,}05$.
De tabel van de F-verdeling voor $\alpha = 0{,}05$ levert $P(\underline{x} > 3{,}33) = 0{,}05$ dus $a = 3{,}33$.
$P(\underline{x} < b) = P(\frac{1}{\underline{x}} > \frac{1}{b}) = 0{,}05$ en $\frac{1}{\underline{x}} \sim F(10;5)$ zodat de tabel levert $P(\frac{1}{\underline{x}} > 4{,}74) = 0{,}05$. Dus $\frac{1}{b} = 4{,}74$ ofwel $b = 0{,}21$. □

Als men twee onafhankelijke steekproeven uit Normale populaties heeft dan zijn de twee steekproefvarianties onafhankelijk en verwant aan de χ^2-verdeling. In die situatie is de F-verdeling nuttig.
De F-verdeling wordt ook wel Snedecor-verdeling genoemd.

5.5. Het rekenwerk

Voor het verkrijgen van resultaten in de statistiek is het niet alleen nodig om te werken met stochastische variabelen maar moeten ook vroeg of laat uit de getallen die door het trekken van de steekproef ter beschikking zijn gekomen de realisaties van de gebruikte steekproeffuncties worden bepaald. Dat komt vrijwel steeds neer op het bepalen van de realisaties $\overline{x}$ en s^2 van $\underline{\overline{x}}$, resp. $\underline{s}^2$. De meeste rekentuigjes hebben knoppen die, mits vakkundig bediend, na het invoeren van de getallen $x_1,x_2,\ldots,x_n$, het gemiddelde, de variantie of de standaardafwijking van die n getallen achter het venster laten verschijnen. Daarbij is het meestal zó dat bij de berekening van de variantie door n gedeeld wordt en bij de berekening van de standaardafwijking door n–1, of omgekeerd. U moet alleen de knop gebruiken die correspondeert met de deling door n–1 en de andere grootheid, door worteltrekken of kwadrateren, uit het verkregen resultaat afleiden (dus s via s^2, of s^2 via s).
Soms is het rekentuig kapot. Ook is het niet moeilijk zodanige getallen $x_1,x_2,\ldots,x_n$ te kiezen dat het geleverde resultaat onjuist is (de x_i groot en $|x_i-x_j|$ klein). Het is dus nuttig om toch enige vaardigheid te ontwikkelen in het met de hand uitrekenen van $\overline{x}$ en s^2.
Meestal is voor s^2 de berekening m.b.v. de volgende uitdrukking handiger:

$$s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2=\frac{1}{n-1}\sum_{i=1}^{n}x_i^2-\frac{2\overline{x}}{n-1}\sum_{i=1}^{n}x_i+\frac{n}{n-1}\overline{x}^2$$

dus

$$s^2=\frac{1}{n-1}\{\sum_{i=1}^{n}x_i^2-n\overline{x}^2\}=\frac{1}{n-1}\{\sum_{i=1}^{n}x_i^2-\frac{1}{n}(\sum_{i=1}^{n}x_i)^2\}.$$

Dit is $\frac{n}{n-1}(m_2-m_1^2)$, hetgeen doet denken aan de manier waarop $\sigma^2(\underline{x})$ meestal wordt bepaald*).
Als de x_i voor $i = 1,2,\ldots,n$ worden vervangen door

$$y_i=\frac{x_i-a}{b}$$

*) $\underline{m}_k$ is het k-de steekproefmoment en is gedefinieerd als $\underline{m}_k=\frac{1}{n}\sum_{i=1}^{n}\underline{x}_i^k$.

dan is $\quad \overline{y} = \frac{1}{n}\sum_{i=1}^{n}\frac{x_i-a}{b} = \frac{1}{nb}(n\overline{x}-na) = \frac{\overline{x}-a}{b}$

en $\quad s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\frac{x_i-a}{b} - \frac{\overline{x}-a}{b})^2 = \frac{1}{(n-1)b^2}\sum_{i=1}^{n}(x_i-x)^2 = \frac{1}{b^2}s_x^2.$

Door alle uitkomsten met een prettig gekozen constante te verminderen en de zo verkregen verschillen te delen door een geschikte constante kunnen we de zaak handelbaarder maken (ook als daarna het rekentuig wordt gebruikt).
Voor a kiezen we een getal dat op het oog in de buurt van $\overline{x}$ ligt. We kiezen b bij voorkeur zó dat breuken verdwijnen. Na de berekening van $\overline{y}$ en s_y^2 volgen $\overline{x}$ en s_x^2 uit $\overline{x} = b\overline{y} + a$ en $s_x^2 = b^2 s_y^2$.

Voorbeeld

Gegeven de steekproefresultaten

x:	56,3	54,7	55,8	52,1	59,7	58,2

Bepaal $\overline{x}$ en s_x^2.

Ga over op $y_i = \frac{x_i-56}{1/10}$, $i = 1,2,\ldots,6$; dus $a = 56$ en $b = 1/10$.

y:	3	–13	–2	–39	37	22
y^2:	9	169	4	1521	1369	484

$\overline{y} = \frac{1}{6}(3-13-2-39+37+22) = \frac{8}{6}$

$s_y^2 = \frac{1}{6-1}\{(9+169+4+1521+1369+484) - 6\overline{y}^2\} = \frac{1}{5}(3556 - 6\cdot(\frac{8}{6})^2) = \frac{10636}{15} = 709{,}07.$

Daaruit volgt $\overline{x} = 56 + \frac{1}{10}\cdot\frac{8}{6} = 56{,}13$ en $s_x^2 = 7{,}09$. □

Het is niet zinvol in dit voorbeeld meer decimalen te vermelden. Dat geeft een misplaatste indruk van de precisie. De getallen in de steekproef zijn vermoedelijk ook afgerond. Als de omvang van de steekproef groot is t.o.v. het aantal verschillende resultaten is het gemakkelijker om de x_i's niet elk apart te behandelen maar eerst een frequentie-tabel te maken:

waarde:	x_1	x_2	...	x_k
frequentie:	f_1	f_2	...	f_k

Als de steekproefomvang gelijk is aan n moet $\sum_{j=1}^{k} f_j = n$. Nu wordt $\overline{x} = \frac{1}{n}\sum_{j=1}^{k} f_j x_j$ en $\sum_{i=1}^{n} x_i^2 = \sum_{j=1}^{k} f_j x_j^2$. Ook hier kan natuurlijk de transformatie $y = \frac{x-a}{b}$ worden gebruikt.

Voorbeeld

De frequentie-tabel in een steekproef met n = 50 is:

x_i:	0,98	0,99	1,–	1,01	1,02
f_i:	2	8	20	13	7

Met $y = \frac{x-1}{1/100}$ krijgen we:

y_i:	–2	–1	0	1	2
y_i^2:	4	1	0	1	4

$\overline{y} = \frac{1}{50}\{2\cdot(-2) + 8\cdot(-1) + 20\cdot(0) + 13\cdot(1) + 7\cdot(2)\} = \frac{15}{50} = 0{,}3.$

$s_y^2 = \frac{1}{49}\{(2\cdot 4 + 8\cdot 1 + 20\cdot 0 + 13\cdot 1 + 7\cdot 4) - 50\cdot(0{,}3)^2\} = \frac{1}{49}(57 - \frac{9}{2}) = \frac{52{,}5}{49}.$

Dus $\overline{x} = 1 + \frac{1}{100}\cdot 0{,}3 = 1{,}003$ en $s_x^2 = \frac{1}{100^2}\cdot\frac{52{,}5}{49} = 0{,}0001.$ □

Bij grote steekproeven deelt men de waarnemingen in aansluitende, maar elkaar niet overlappende, klassen in en werkt verder met de klasse-middens i.p.v. de werkelijke waarden. Bij waarnemingen aan een continue variabele gebeurt dat, onbewust, natuurlijk ook doordat een resultaat 38,5 alleen maar aangeeft dat de waarneming lag tussen 38,45 en 38,55.

6 Schatten

6.1. Inleiding

De bedoeling is om op grond van een steekproef een zo goed mogelijke uitspraak te doen omtrent de waarde van één of meer parameters van de populatie-verdeling. Voor een onbekende parameter kan een puntschatting of een betrouwbaarheids-interval geconstrueerd worden. Een puntschatting is een getal dat we krijgen als realisatie van een geschikt gekozen steekproeffunctie. In de gebruikelijke situatie weten we vrijwel zeker dat dit getal niet gelijk zal zijn aan de te schatten parameter, maar we hopen dat de fout niet te groot is. Een betrouwbaarheidsinterval komt aan dit bezwaar tegemoet. Daarbij wordt, uitgaande van de kansverdeling van een functie van de stochastische variabelen in de steekproef èn de onbekende parameter, een interval geconstrueerd dat met een zekere betrouwbaarheid die onbekende parameter bevat.

6.2. Puntschatting

Een puntschatting, kortweg schatting, is de realisatie van een steekproeffunctie die schatter wordt genoemd. De schatter is het recept dat we moeten volgen om op grond van een t.z.t. te trekken steekproef $\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_n$ te komen tot een schatting $\widehat{\theta}$ van de onbekende parameter θ van de populatie-verdeling $F(x;\theta)$. Een schatter voor θ duiden we aan met $\widehat{\underline{\theta}}$.
Of andersom: er is een populatie met verdeling $F(x;\theta,\xi,\ldots,\nu)$. Afgezien van de onbekende parameters $\theta,\xi,\ldots,\nu$ is F geheel bepaald. Uit die populatie zal een steekproef $\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_n$ worden getrokken met het doel θ te schatten. Daartoe gebruiken we de schatter $\widehat{\underline{\theta}} = g(\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_n)$. Zodra de steekproef is getrokken kennen we de realisaties $x_1,x_2,\ldots,x_n$ en dus ook de realisatie $\widehat{\theta} = g(x_1,x_2,\ldots,x_n)$ van de schatter $\widehat{\underline{\theta}}$. Die realisatie is de schatting $\widehat{\theta}$ van θ. Het komt ook voor dat men een schatting wil hebben van een functie van één of meer van de parameters, bijvoorbeeld van het k-de moment van de populatie-verdeling of van de kans dat de stochastische variabele in de populatie $\leq t$ is. We willen natuurlijk een goede schatting krijgen. Helaas weten we in een speciaal geval nooit hoe goed of hoe slecht de schatting is. Wel kunnen we iets zeggen over de kwaliteit van het gebruikte recept, m.a.w. sommige schatters zijn 'beter' dan andere schatters. Er is

echter geen éénduidige maatstaf om de kwaliteit van een schatter te bepalen. Wel zijn direct een aantal eigenschappen te bedenken die wenselijk zijn voor een schatter.

A. Zuiverheid

Definitie

Een schatter $\underline{\hat{\theta}}$ voor θ is *zuiver* als $E(\underline{\hat{\theta}}) = \theta$, ongeacht de grootte van de steekproef en ongeacht de waarde van θ.

Voorbeeld

$\underline{\bar{x}}$ is een zuivere schatter voor de populatie-verwachting μ want $E(\underline{\bar{x}}) = E(\underline{x}) = \mu$.
$\underline{s}^2$ is een zuivere schatter voor de populatie-variantie σ^2 want $E(\underline{s}^2) = \sigma^2$. □

Een schatter die niet zuiver is, heet *onzuiver*.
Bij het gebruiken van een zuivere schatter is het resultaat 'gemiddeld' juist. We maken geen systematische fout.

Voorbeeld

Als bij onbekende verwachting, de variantie σ^2 geschat moet worden en we doen dat met de schatter $\underline{t} = \frac{1}{n}\sum_{i=1}^{n}(\underline{x}_i-\underline{\bar{x}})^2$ dan maken we een systematische fout want $E(\underline{t}) = \frac{n-1}{n}\sigma^2$. We geven dus 'gemiddeld' een te kleine schatting voor de variantie met deze schatter. Wel is het duidelijk dat voor grote n de fout verwaarloosbaar is.
Het verschil $E(\underline{\hat{\theta}}) - \theta$ heet de onzuiverheid van de schatter.

B. Nauwkeurigheid

Als men meer dan eens een schatting maakt voor θ m.b.v. de schatter $\underline{\hat{\theta}}$ dan is het niet prettig als de opeenvolgende schattingen grote verschillen vertonen. Ook al is de schatter zuiver dan hebben we in zo'n geval niet veel aan de schatting omdat het heel goed mogelijk is dat de gevonden schatting veel te groot of veel te klein is. Daarom is het gewenst dat de nauwkeurigheid van de schatter groot is.

Definitie

Een schatter $\underline{\hat{\theta}}_1$ is nauwkeuriger dan een schatter $\underline{\hat{\theta}}_2$ als bij gelijke steekproef-omvang

$$E\{(\underline{\hat{\theta}}_1 - \theta)^2\} \le E\{(\underline{\hat{\theta}}_2 - \theta)^2\}$$

voor elke waarde van θ, terwijl voor minstens één waarde van θ geldt:

$$E\{(\underline{\hat{\theta}}_1 - \theta)^2\} < E\{(\underline{\hat{\theta}}_2 - \theta)^2\}.$$

De grootheid $E\{(\hat{\underline{\theta}} - \theta)^2\}$ heet *gemiddelde kwadratische fout* (mean squared error), notatie $MSE(\hat{\underline{\theta}})$. Als $\hat{\underline{\theta}}$ een zuivere schatter is voor θ en dus $\theta = E(\hat{\underline{\theta}})$ is $MSE(\hat{\underline{\theta}}) = \sigma^2(\hat{\underline{\theta}})$. De schatter $\hat{\underline{\theta}}$ heet de beste (of nauwkeurigste) schatter voor θ als er geen schatter te vinden is die bij gelijke steekproefomvang nauwkeuriger is dan $\hat{\underline{\theta}}$. Daarbij wordt elke schatter die mogelijk is bij die steekproefomvang in de vergelijking betrokken.

C. Andere goede eigenschappen

Het is duidelijk dat we van een schatter verwachten dat hij beter wordt naarmate de steekproefomvang toeneemt. Voorts is het wenselijk dat een schatter alle relevante informatie die in de steekproef aanwezig is gebruikt. Tenslotte geniet een eenvoudige schatter de voorkeur boven een gecompliceerde. Dit laatste geeft aanleiding tot de introductie van de klasse van lineaire schatters.

Definitie

Een lineaire schatter $\hat{\underline{\theta}}_n$ is een lineaire functie van de steekproefvariabelen $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$.

Zo kan men de blik nog verder beperken en kijken naar zuivere, lineaire schatters. De schatter met de kleinste variantie binnen de klasse van zuivere, lineaire schatters heet de beste, lineaire, zuivere schatter (best linear unbiased estimator of *BLUE*).

Bij een vergelijking van alle mogelijke schatters stuit men vaak op het probleem hoe te kiezen tussen een zuivere schatter met grote variantie (kleine nauwkeurigheid) en een onzuivere schatter met grote nauwkeurigheid.

Stelling 6.1. Zij $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ een steekproef uit een populatie met verdeling F_x, met $E(\underline{x}) = \mu$ en $\sigma^2(\underline{x}) = \sigma^2$. Het steekproefgemiddelde $\bar{\underline{x}}$ is de beste, lineaire, zuivere schatter voor μ.

Bewijs: Voor elke lineaire, zuivere schatter $\hat{\underline{\mu}}$ voor μ moet gelden

$$\hat{\underline{\mu}} = \sum_{i=1}^{n} a_i \underline{x}_i + b \quad \text{en} \quad E(\hat{\underline{\mu}}) = \mu.$$

Daaruit volgt dat $b = 0$ en $\sum_{i=1}^{n} a_i = 1$.

Voor de beste, lineaire, zuivere schatter moet de variantie minimaal zijn.

Omdat $\sigma^2(\hat{\underline{\mu}}) = \sigma^2 \sum_{i=1}^{n} a_i^2$ moeten we dus $\sum_{i=1}^{n} a_i^2$ minimaliseren onder de nevenvoorwaarde $\sum_{i=1}^{n} a_i = 1$.

$$\sum_{i=1}^{n} a_i^2 = \sum_{i=1}^{n} (a_i - \frac{1}{n} + \frac{1}{n})^2 = \sum_{i=1}^{n} (a_i - \frac{1}{n})^2 + \frac{n}{n^2} + \frac{2}{n}\sum_{i=1}^{n} (a_i - \frac{1}{n})$$

zodat met de nevenvoorwaarde geldt:

$$\sum_{i=1}^{n} a_i^2 = \frac{1}{n} + \sum_{i=1}^{n} (a_i - \frac{1}{n})^2$$

en dit is minimaal gelijk aan $\frac{1}{n}$ als $a_i = \frac{1}{n}$ voor $i = 1,2,\ldots,n$.

De BLUE is in dit geval dus $\frac{1}{n}\sum \underline{x}_i = \underline{\bar{x}}$ met $\sigma^2(\underline{\bar{x}}) = \frac{1}{n}\sigma^2$. □

6.3. Constructie van schatters

We zullen twee veel gebruikte methoden bespreken om een schatter te vinden.

A. De momentenmethode

Zojuist is gebleken dat het steekproefgemiddelde, of eerste steekproefmoment, de BLUE voor het eerste moment van de populatie-verdeling is.

De steekproeffunctie $\underline{m}_k = \frac{1}{n}\sum_{i=1}^{n} \underline{x}_i^k$ heet het k-de steekproefmoment.

Stelling 6.2. Zij $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ een steekproef uit een populatie met verdelingsfunctie $F_{\underline{x}}$. Zij voorts $E(\underline{x}^r) = \mu_r$ dan is $E(\underline{m}_k) = \mu_k$ mits μ_k bestaat en $\sigma^2(\underline{m}_k) = \frac{1}{n}(\mu_{2k} - \mu_k^2)$ mits μ_{2k} bestaat.

Bewijs: $E(\underline{m}_k) = \frac{1}{n}\sum_{i=1}^{n} E(\underline{x}^k) = \frac{1}{n}\sum_{i=1}^{n} \mu_k = \mu_k$.

$$\sigma^2(\underline{m}_k) = \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2(\underline{x}^k) = \frac{1}{n}[E(\underline{x}^{2k}) - \{E(\underline{x}^k)\}^2] = \frac{1}{n}(\mu_{2k} - \mu_k^2).$$ □

De steekproefmomenten zijn dus zuivere schatters voor de overeenkomstige populatie-momenten. De populatie-momenten zijn in het algemeen functies van de parameters, zodat met een beetje geluk de parameters ook als functies van de populatie-momenten kunnen worden geschreven. Door in die laatste functies de populatie-momenten te vervangen door de steekproefmomenten verkrijgt men schatters voor de parameters. Zodra de steekproef getrokken is kunnen de realisaties m_k van de steekproefmomenten $\underline{m}_k$ worden berekend en daarmee de schattingen voor de parameters bepaald. Een schatter die op deze manier wordt gevonden heet een momentenschatter.

Bij het gebruik van deze methode heeft men voor het bepalen van schattingen van k parameters, k momenten nodig. Men kiest daartoe steeds zo laag mogelijke momenten, al zullen dat niet altijd het eerste tot en met het k-de moment kunnen zijn. Zo is bij een verdeling die symmetrisch is t.o.v. 0 het eerste moment altijd nul onafhankelijk van de waarde van de parameters. In dat geval is het eerste moment niet bruikbaar.

Voorbeeld

De populatie is Gamma-verdeeld met onbekende parameters α en β. Bepaal schatters voor α en β m.b.v. de momentenmethode, op grond van een steekproef $\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_n$ uit deze populatie.
De dichtheid van de Gamma-verdeling is

$$\begin{aligned} f_x(x) &= \frac{x^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha} e^{-x/\beta}, \quad x>0; \\ &= 0, \qquad\qquad \text{elders.} \end{aligned}$$

Dus $\quad E(\underline{x}) = \int_0^\infty \frac{x^\alpha}{\Gamma(\alpha)\beta^\alpha} e^{-x/\beta}dx = \frac{\beta}{\Gamma(\alpha)} \int_0^\infty (x/\beta)^\alpha e^{-x/\beta}dx/\beta = \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} \cdot \beta = \alpha\beta$

en $\quad E(\underline{x}^2) = \int_0^\infty \frac{x^{\alpha+1}}{\Gamma(\alpha)\beta^\alpha} e^{-x/\beta}dx = \frac{\beta^2}{\Gamma(\alpha)} \int_0^\infty (x/\beta)^{\alpha+1} e^{-x/\beta}dx/\beta =$

$$= \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)} \cdot \beta^2 = \alpha(\alpha+1)\beta^2.$$

$$\left.\begin{aligned} \mu_1 &= \alpha\beta \\ \mu_2 &= (\alpha\beta)^2 + \alpha\beta^2 \end{aligned}\right\} \rightarrow \left\{\begin{aligned} \alpha &= \frac{\mu_1^2}{\mu_2-\mu_1^2} \\ \beta &= \frac{\mu_2-\mu_1^2}{\mu_1} \end{aligned}\right.$$

De schatters voor α en β worden dus:

$$\underline{\widehat{\alpha}} = \frac{\underline{m}_1^2}{\underline{m}_2-\underline{m}_1^2} \text{ en } \underline{\widehat{\beta}} = \frac{\underline{m}_2-\underline{m}_1^2}{\underline{m}_1}.$$

Laat nu de steekproef als realisatie 2,5,1,4 hebben dan is

$$m_1 = \tfrac{1}{4}(2+5+1+4) = 3 \text{ en } m_2 = \tfrac{1}{4}(4+25+1+16) = 11{,}5$$

zodat de schattingen worden:

$$\widehat{\alpha} = \frac{9}{11{,}5-9} = 3{,}6 \text{ en } \widehat{\beta} = \frac{11{,}5-9}{3} = \frac{5}{6}.$$

Als de belangstelling was uitgegaan naar de variantie ($\alpha\beta^2$) van de populatieverdeling zou men via deze methode als schatting $3{,}6\cdot(\frac{5}{6})^2 = \frac{10}{4}$ gevonden hebben. De zuivere schatter voor de variantie $\underline{s}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\underline{x}_i - \overline{\underline{x}})^2$ levert echter de schatting $\frac{1}{3}(1^2+2^2+2^2+1^2) = \frac{10}{3}$, waar we meer vertrouwen in hebben. ❑

Het kan voorkomen dat men m.b.v. een momentenschatter een onmogelijke schatting verkrijgt. Men wijkt dan uit naar de 'dichtstbijzijnde' wel mogelijke waarde van de parameter.

Voorbeeld

$\underline{x}$ is uniform verdeeld op $(0;\theta)$. Bepaal een momentenschatter voor θ.
$E(\underline{x}) = \frac{1}{2}\theta$, dus $\theta = 2\mu_1$ en $\widehat{\underline{\theta}} = 2\underline{m}_1$.
Zij de realisatie van de steekproef 1,2,3,10, dan is $m_1 = 4$ dus $\widehat{\theta} = 8$. Maar een waarneming 10 kan niet uit een uniforme verdeling op (0;8) komen. Als schatting voor θ komt dan $\widehat{\theta} = 10$, de het dichtst bij 8 liggende wel mogelijke waarde in aanmerking. ❑

Een ander nadeel van momentenschatters is dat ze niet éénduidig bepaald zijn, het resultaat is afhankelijk van welke momenten men gebruikt.

Voorbeeld

Wordt in het vorige voorbeeld met μ_3 i.p.v. met μ gewerkt, dan krijgt men:

$$\mu_3 = \int_0^\theta x^3 \frac{1}{\theta}\,dx = \frac{1}{4}\theta^3 \text{ dus } \theta = \sqrt[3]{4\mu_3} \text{ en } \widehat{\underline{\theta}} = \sqrt[3]{4\underline{m}_3}$$

zodat $\widehat{\theta} = \sqrt[3]{(1+8+27+1000)} = 10{,}12.$

Dit is een waarde die wel mogelijk is. ❑

B. De methode van de grootste aannemelijkheid

Bij de methode van de grootste aannemelijkheid (maximum-likelihood), waarmee meest aannemelijke of ML-schatters worden gevonden, is de onderliggende gedachte dat het aannemelijker is dat de getrokken steekproef een grote kans van optreden heeft dan dat die kans klein is. De kans van optreden is een functie L van de onbekende parameters. Kies nu zodanige waarden voor de parameters dat L maximaal wordt. Die waarden noemt men de ML-schattingen voor de parameters. Men schat de parameters dus met die waarden waarvoor de kans van optreden van de voorliggende steekproefrealisatie zo groot mogelijk is.

Zo gesteld zou de methode alleen bruikbaar zijn bij een discrete stochastische variabele. In het continue geval kan men echter gebruik maken van de benadering $f(x)dx \approx P(x < \underline{x} \leq x + dx)$ en de simultane dichtheid van $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ maximaliseren door de juiste keuze van de parameters.
Daar in overeenkomstige gevallen steeds dezelfde functie van $x_1, x_2, \ldots, x_n$ de ML-schatting oplevert kan men door die functie als recept te nemen, dat moet worden toegepast op een nog te trekken steekproef $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$, komen tot een ML-schatter.

Voorbeeld

$\underline{x} \sim B(10;p)$, p is onbekend. Een steekproef aan $\underline{x}$ levert de realisaties $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$.

$$P(\underline{x}_i = x_i;\ i = 1,2,\ldots,n) = \prod_{i=1}^{n} \binom{10}{x_i} p^{x_i}(1-p)^{10-x_i}.$$

$$L(p) = p^{n\bar{x}}(1-p)^{n(10-\bar{x})}\prod_{i=1}^{n} \binom{10}{x_i} = c\cdot\{p^{\bar{x}}(1-p)^{10-\bar{x}}\}^n.$$

$$L'(p) = cn\{p^{\bar{x}}(1-p)^{10-\bar{x}}\}^{n-1}\cdot\{\bar{x}\cdot p^{\bar{x}-1}(1-p)^{10-\bar{x}} - (10-x)p^{\bar{x}}(1-p)^{10-\bar{x}-1}\}.$$

$L'(p) = 0$ als $\bar{x}\cdot p^{\bar{x}-1}(1-p)^{10-\bar{x}} = (10-\bar{x})p^{\bar{x}}(1-p)^{10-\bar{x}-1}$ ofwel $\bar{x}(1-p) = (10-\bar{x})p$ of $\bar{x} = 10p$ dus $p = \frac{\bar{x}}{10}$.
Nagegaan moet nu worden of voor $p = \frac{\bar{x}}{10}$ een maximum dan wel een minimum van $L(p)$ optreedt. Daar $L(p)$ voor $p = 0$ zowel als voor $p = 1$ gelijk nul is en $L(p)$ continu en voor $0 < p < 1$ positief is moet dit extreem wel een maximum zijn en is $p = \frac{\bar{x}}{10}$ de ML-schatting voor p. Dus is in dit geval $\hat{\underline{p}} = \frac{\bar{\underline{x}}}{10}$ de ML-schatter voor p.
❑

Veelal is het eenvoudiger om het maximum van ln L te bepalen. Dat valt samen met het maximum van L. Men kan dat doen omdat L als simultane dichtheid of kansfunctie nooit negatief is.

Voorbeeld

$\underline{x} \sim N(\mu;\sigma^2)$, μ en σ^2 zijn onbekend. Bepaal de ML-schatters voor μ en σ^2 op basis van een steekproef met omvang n.

$$L(\mu;\sigma^2) = \frac{1}{(\sigma\sqrt{2\pi})^n}\, e^{\frac{-1}{2\sigma^2}\sum(x_i-\mu)^2}$$

$$\ln L(\mu;\sigma^2) = c - \tfrac{1}{2} n \ln \sigma^2 - \frac{1}{2\sigma^2}\sum(x_i-\mu)^2.$$

Behalve voor $\sigma^2 = 0$ is ln L continu, dus voor een extreem is nodig

$$\frac{\partial \ln L}{\partial \mu} = 0 \quad \text{en} \quad \frac{\partial \ln L}{\partial \sigma^2} = 0.$$

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2}\sum(x_i-\mu); \quad \frac{\partial \ln L}{\partial \sigma^2} = \frac{-n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum(x_i-\mu)^2.$$

Stellen we de beide partiële afgeleiden gelijk aan nul dan volgt (met $\sigma^2 \neq 0$) dat voor $\mu_0 = \bar{x}$ en $\sigma_0^2 = \frac{1}{n}\sum(x_i-\bar{x})^2$ L een stationair punt heeft. Afgezien van de, evenals $\sigma^2 = 0$, niet interessante mogelijkheid $\sigma^2 = \infty$ is dit het enige stationaire punt. Controle levert dat het een maximum is zodat

$$\hat{\mu} = \bar{x} \quad \text{en} \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 \quad \text{de ML-schattingen}$$

en

$$\underline{\hat{\mu}} = \underline{\bar{x}} \quad \text{en} \quad \underline{\hat{\sigma}}^2 = \frac{1}{n}\sum_{i=1}^{n}(\underline{x}_i-\underline{\bar{x}})^2 \quad \text{de ML-schatters zijn voor } \mu \text{ en } \sigma^2. \qquad \square$$

Ook een ML-schatter is niet altijd zuiver. In het bovenstaande voorbeeld is $\underline{\hat{\mu}}$ zuiver maar $\underline{\hat{\sigma}}^2$ onzuiver.

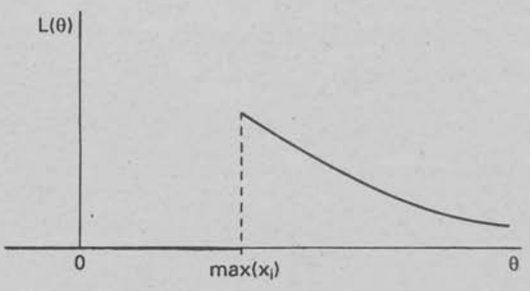
Het is mogelijk dat de aannemelijkheidsfunctie niet een éénduidig bepaald maximum heeft, maar dat komt zelden voor. Wel moet men erop bedacht zijn dat het gezochte maximum een randextreem kan zijn. In dat geval hoeft L geen stationair punt te hebben.

Voorbeeld

$\underline{x}$ is uniform verdeeld op $[0;\theta]$. Bepaal de ML-schatter voor θ, op basis van de steekproef $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$.

$$f(x_1,x_2,\ldots,x_n;\theta) = \frac{1}{\theta^n}, \quad 0 \leq x_i \leq \theta;\ i = 1,2,\ldots,n;$$

$$= 0, \qquad \text{elders.}$$



Figuur 6.1.

$$L(\theta) = \frac{1}{\theta^n}, \quad \theta \geq \max_i(x_i);$$
$$= 0, \quad \text{elders.}$$

Het interval $[0;\theta]$ moet alle waarnemingen bevatten dus θ is niet kleiner dan de grootste waarneming.
$L(\theta)$ is maximaal als $\theta = \max_i(x_i)$ en dus is $\underline{\widehat{\theta}} = \max_i(\underline{x}_i)$. ❑

Een ML-schatter produceert geen onmogelijke schattingen.
Zij z een functie $g(\theta_1,\theta_2,\ldots,\theta_n)$ van de onbekende parameters dan is de ML-schatter voor z gelijk aan $g(\underline{\widehat{\theta}}_1,\underline{\widehat{\theta}}_2,\ldots,\underline{\widehat{\theta}}_n)$, ofwel: de ML-schatter van de functie is de functie van de ML-schatters.

6.4. Betrouwbaarheidsintervallen

Voor een schatter is het niet nodig dat zijn verdeling bekend is. Zo kan $\underline{\bar{x}}$ altijd worden gebruikt om de populatie-verwachting te schatten.
Om een betrouwbaarheidsinterval voor een parameter θ te construeren moeten we echter uitgaan van een functie, $\underline{d}$ van $\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_n$ èn de onbekende θ, waarvan de verdelingsfunctie geheel bekend is.

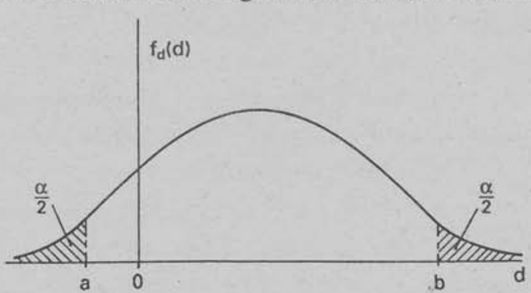
Voorbeeld

De steekproef $\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_{25}$ komt uit een $N(\mu;16)$-populatie.
Dan is $\underline{\bar{x}} \sim N(\mu,\frac{16}{25})$ en dus $\underline{d} = \frac{\underline{\bar{x}}-\mu}{0,8} \sim N(0;1)$.
Hier is $\underline{d}$ een functie van μ en de steekproefvariabelen $\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_n$. De verdeling van $\underline{d}$ is bekend. ❑

Omdat de verdeling van $\underline{d}$ bekend is kunnen we dus twee getallen, a en b, vinden zodanig dat

$$P(\underline{d} > b) = \frac{\alpha}{2} \quad \text{en} \quad P(\underline{d} < a) = \frac{\alpha}{2}$$

voor $0 < \alpha < 1$. We nemen voor het gemak aan dat $\underline{d}$ continu is.



Figuur 6.2.

Dan volgt dat

$$P(a < \underline{d} < b) = 1 - \alpha.$$

Hierin is $\underline{d}$ een functie van θ en de $\underline{x}_i$'s dus $\underline{d} = h(\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_n;\theta)$. Als het nu lukt om θ te isoleren kan de ongelijkheid $a < \underline{d} < b$ worden herschreven als $\underline{k} < \theta < \underline{g}$ waarin $\underline{k}$ en $\underline{g}$ functies zijn van $\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_n$ en a (of b) respectievelijk b (of a). We krijgen zo $P(\underline{k} < \theta < \underline{g}) = 1 - \alpha$.

Voorbeeld

Vervolg vorige voorbeeld:

$$P(a < \underline{d} < b) = P(a < \frac{\bar{\underline{x}}-\mu}{0{,}8} < b) = 1 - \alpha$$

$$a < \frac{\bar{\underline{x}}-\mu}{0{,}8} < b \rightarrow \bar{\underline{x}} - 0{,}8b < \mu < \bar{\underline{x}} - 0{,}8a.$$

Hier is dus $\underline{k} = \bar{\underline{x}} - 0{,}8b$ en $\underline{g} = \bar{\underline{x}} - 0{,}8a$ en $P(\bar{\underline{x}} - 0{,}8b < \mu < \bar{\underline{x}} - 0{,}8a) = 1 - \alpha$.❑

Dit is het recept voor de constructie van een betrouwbaarheidsinterval voor θ (vergelijkbaar met de schatter). Na het trekken van de steekproef kunnen de realisaties k en g van $\underline{k}$ respectievelijk $\underline{g}$ worden bepaald. Het betrouwbaarheidsinterval is dan

$$(k;g) \text{ met betrouwbaarheid } 100(1-\alpha)\ \%.$$

We kunnen niet meer spreken van een kans $1 - \alpha$ dat θ tussen k en g in ligt. Omdat de waarde van k en g afhangt van α moeten we aan het interval een aanduiding toevoegen waarin de waarde van α tot uitdrukking komt.

Voorbeeld

Vervolg vorige voorbeeld. Zij $\bar{x} = 22$ en $\alpha = 0{,}10$. Uit $\alpha = 0{,}10$ volgt $a = -1{,}645$ en $b = 1{,}645$; dus $k = 22 - 0{,}8 \cdot 1{,}645$; $g = 22 + 0{,}8 \cdot 1{,}645$.
Het betrouwbaarheidsinterval wordt (20,68; 23,32) met 90% betrouwbaarheid.
Zij $\bar{x} = 22$ en $\alpha = 0{,}05$. Nu wordt $k = 22 - 0{,}8 \cdot 1{,}96$ en $g = 22 + 0{,}8 \cdot 1{,}96$. Het betrouwbaarheidsinterval wordt (20,43; 23,57) met 95% betrouwbaarheid. ❑

Hoe kleiner α, m.a.w. hoe dichter de betrouwbaarheid bij de 100% komt, hoe langer het betrouwbaarheidsinterval wordt. Het geconstrueerde betrouwbaarheidsinterval (k;g) met betrouwbaarheid $100(1-\alpha)\%$ is een tweezijdig betrouwbaarheidsinterval. Eénzijdige betrouwbaarheidsintervallen $(-\infty,g)$ of (k,∞) met betrouwbaarheid $100(1-\frac{\alpha}{2})\%$ ontstaan door maar aan één zijde een eindige grens te stellen. $(-\infty,g)$ heet een rechtszijdig begrensd en (k,∞) een linkszijdig begrensd betrouwbaarheidsinterval. Bij de constructie van éénzijdige betrouwbaarheidsintervallen volgt a of b direct uit de eis $P(\underline{d} > b) = \alpha/2$ of $P(\underline{d} < a) = \alpha/2$, maar bij

een tweezijdig betrouwbaarheidsinterval is deze 'symmetrie' niet noodzakelijk. Als $\underline{d}$ een symmetrische dichtheid heeft, is het wel de beste keuze. Best in die zin dat het onder de gegeven omstandigheden het kortste betrouwbaarheidsinterval oplevert. Terwille van de eenvoud wordt echter steeds op de boven beschreven manier gewerkt.
De stochastische variabele $\underline{d}$ als functie van $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ en θ heet *spilfunctie* (pivotal-function).

Voorbeeld

Vervolg vorige voorbeeld.
Het tweezijdige 90%-betrouwbaarheidsinterval is (20,68; 23,32).
Het linkszijdig begrensd 95%-betrouwbaarheidsinterval wordt (20,68; ∞).
Het rechtszijdig begrensd 95%-betrouwbaarheidsinterval wordt (–∞, 23,32). ❑

De lengte $\underline{g} - \underline{k}$ van het betrouwbaarheidsinterval is een stijgende functie van b – a. De punten a en b volgen uit de verdeling van de spilfunctie $\underline{d}$. Het betrouwbaarheidsinterval zal dus korter zijn naarmate de variantie van $\underline{d}$ kleiner is. Als al gebruik gemaakt wordt van de beste spilfunctie en het betrouwbaarheidsinterval is te lang dan moet de steekproefomvang vergroot worden omdat meestal $\sigma^2(\underline{d})$ een dalende functie van n is.

Voorbeeld

De steekproef $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ komt uit een $N(\mu;\sigma^2)$-populatie; σ^2 is bekend. Een 90%-betrouwbaarheidsinterval voor μ mag hoogstens de lengte 1 hebben. Hoe groot moet n minimaal zijn?
De spilfunctie is $\underline{d} = \dfrac{\underline{\bar{x}}-\mu}{\sigma/\sqrt{n}}$; $\underline{d} \sim N(0;1)$.
Een 90%-betrouwbaarheidsinterval voor μ is een realisatie van

$$(\underline{\bar{x}} - 1{,}645 \frac{\sigma}{\sqrt{n}} ; \underline{\bar{x}} + 1{,}645 \frac{\sigma}{\sqrt{n}})$$

en heeft dus de lengte $3{,}290(\sigma/\sqrt{n})$ onafhankelijk van de waarde van μ en de realisatie van $\underline{\bar{x}}$. De eis is dus dat $3{,}29(\sigma/\sqrt{n}) \leq 1$ ofwel $n \geq (3{,}29)^2\sigma^2$. Natuurlijk moet n een geheel getal zijn. ❑

Met de volgende spilfuncties worden vaak betrouwbaarheidsintervallen geconstrueerd.

■ Bij een steekproef uit een Normale populatie:

Voor	Spilfunctie $\underline{d}$	Verdeling van $\underline{d}$
μ als σ^2 bekend is	$\dfrac{\bar{\underline{x}}-\mu}{\sigma/\sqrt{n}}$	$N(0;1)$
μ als σ^2 onbekend is	$\dfrac{\bar{\underline{x}}-\mu}{\underline{s}/\sqrt{n}}$	$t(n-1)$
σ^2 als μ bekend is	$\dfrac{1}{\sigma^2}\sum_{i=1}^{n}(\underline{x}_i-\mu)^2$	$\chi^2(n)$
σ^2 als μ onbekend is	$\dfrac{n-1}{\sigma^2}\underline{s}^2$	$\chi^2(n-1)$

■ Bij een steekproef uit een exponentiële populatie:

λ (d.w.z. $E\underline{x} = 1/\lambda$)	$2n\lambda\bar{\underline{x}}$	$\chi^2(2n)$

Als de spilfunctie $\underline{d}$ discreet is kunnen a en b over het algemeen niet zó worden gekozen dat $P(\underline{d} < a) = P(\underline{d} > b) = \alpha/2$. Men moet a en b dan vervangen door a_1, resp. b_1 waarvoor geldt:
a_1 is het grootste getal waarvoor geldt dat $P(\underline{d} < a_1) \leq \alpha/2$;
b_1 is het kleinste getal waarvoor geldt dat $P(\underline{d} > b_1) \leq \alpha/2$.
De betrouwbaarheid zal dus meestal groter zijn dan $(1-\alpha)\cdot 100\%$.

Het komt ook voor dat een betrouwbaarheidsinterval moet worden gemaakt op basis van twee steekproeven, uit elk van twee populaties één. Hierbij moeten twee fundamenteel verschillende situaties worden onderscheiden.

A. Gepaarde waarnemingen

Als van de twee steekproeven $\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_n$ en $\underline{y}_1,\underline{y}_2,\ldots,\underline{y}_n$ de waarnemingen met gelijke indices afhankelijk zijn van elkaar (dus $\underline{x}_i$ en $\underline{y}_i$, $i = 1,2,\ldots,n$) dan spreekt men van gepaarde waarnemingen. Men brengt dit terug tot de vorige situatie door één steekproef van verschillen $\underline{v}_i = \underline{x}_i - \underline{y}_i$, $i = 1,2,\ldots,n$ te bekijken. Als aangenomen mag worden dat de populatie van verschillen $\underline{v}$ Normaal verdeeld is kan dus op de gebruikelijke wijze een betrouwbaarheidsinterval voor $\mu_v = \mu_x - \mu_y$ geconstrueerd worden. Dit soort situaties komt bijvoorbeeld voor als men wil nagaan of een bepaalde behandeling effect heeft, en zo ja, wat voor effect.

Voorbeeld

25 studenten worden aselect gekozen en gewogen en daarna drie weken lang volgens een bepaald dieet gevoederd. Dan worden ze weer gewogen. Zij $\underline{x}_i$ het

gewicht in kg van student i vóór, en $\underline{y}_i$ ná de drie weken dan krijgt men resultaten van de volgende gedaante:

x_i:	65	72	77	62	...	80	(voor)
y_i:	66	70	80	63	...	75	(na)

Door over te gaan op $v_i = x_i - y_i$ ontstaat de steekproef:

v_i:	-1	2	-3	-1	...	5

Laat $\overline{v} = 1{,}4$ zijn. Als we mogen aannemen dat $\underline{v} \sim N(\mu_v;4)$ dan volgt als 90%-betrouwbaarheidsinterval voor μ_v:

$$(1{,}4 - 1{,}645 \cdot \tfrac{2}{5}\, ;\, 1{,}4 + 1{,}645 \cdot \tfrac{2}{5}) = (0{,}74;2{,}06).$$

Dit betrouwbaarheidsinterval bevat het punt 0 niet, dus we mogen met betrouwbaarheid 90% de uitspraak doen dat $\mu_v > 0$ of $\mu_x > \mu_y$ ofwel dat het dieet tot gewichtsverlies leidt.

B. Twee onafhankelijke steekproeven

Geheel anders wordt de situatie als $\underline{x}_i$, $i = 1,2,\ldots,n$ een steekproef is uit een $N(\mu_x;\sigma_x^2)$-populatie en $\underline{y}_j$, $j = 1,2,\ldots,m$ een daarvan onafhankelijke steekproef uit een $N(\mu_y;\sigma_y^2)$-populatie. Nu kunnen betrouwbaarheidsintervallen voor $\mu_x-\mu_y$ of voor σ_x^2/σ_y^2 worden geconstrueerd.
Immers $\overline{\underline{x}} \sim N(\mu_x;\sigma_x^2/n)$ en $\overline{\underline{y}} \sim N(\mu_y;\sigma_y^2/m)$ en $\overline{\underline{x}}$ en $\overline{\underline{y}}$ zijn onafhankelijk zodat

$$\overline{\underline{x}} - \overline{\underline{y}} \sim N(\mu_x-\mu_y; \frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m})$$

en

$$\underline{d} = \frac{(\overline{\underline{x}}-\overline{\underline{y}}) - (\mu_x-\mu_y)}{\sqrt{\sigma_x^2/n + \sigma_y^2/m}} \sim N(0;1).$$

Als de varianties bekend zijn kunnen hiermee betrouwbaarheidsintervallen voor $\mu_x-\mu_y$ worden geconstrueerd. Zijn σ_x^2 en σ_y^2 onbekend dan moeten die uit de spilfunctie $\underline{d}$ verdwijnen.

$$\frac{(n-1)}{\sigma_x^2}\underline{s}_x^2 \sim \chi^2(n-1) \text{ en } \frac{(m-1)}{\sigma_y^2}\underline{s}_y^2 \sim \chi^2(m-1).$$

$\underline{s}_x^2$ en $\underline{s}_y^2$ zijn onafhankelijk dus

$$\underline{w} = \frac{(n-1)}{\sigma_x^2}\underline{s}_x^2 + \frac{(m-1)}{\sigma_y^2}\underline{s}_y^2 \sim \chi^2(n+m-2).$$

De populaties waren Normaal dus $\underline{d}$ en $\underline{w}$ zijn onafhankelijk. Dat betekent dat

$$\frac{\underline{d}}{\sqrt{\underline{w}/(n+m-2)}} \sim t(n+m-2),$$

maar dat helpt in dit geval alleen als we kunnen aannemen dat de varianties, hoewel onbekend, aan elkaar gelijk zijn. Met $\sigma_x^2 = \sigma_y^2 = \sigma^2$ volgt

$$\underline{d} = \frac{1}{\sigma} \cdot \frac{(\underline{\bar{x}}-\underline{\bar{y}}) - (\mu_x - \mu_y)}{\sqrt{1/n + 1/m}}$$

en

$$\underline{w} = \frac{1}{\sigma^2} \{(n-1)\underline{s}_x^2 + (m-1)\underline{s}_y^2\}$$

zodat we als spilfunctie kunnen gebruiken:

$$\frac{\underline{d}}{\sqrt{\underline{w}/(n+m-2)}} = \frac{(\underline{\bar{x}}-\underline{\bar{y}}) - (\mu_x - \mu_y)}{\sqrt{\frac{1}{n} + \frac{1}{m}} \cdot \sqrt{\frac{(n-1)\underline{s}_x^2 + (m-1)\underline{s}_y^2}{n+m-2}}} \sim t(n+m-2).$$

Om te beoordelen of de aanname $\sigma_x^2 = \sigma_y^2$ redelijk is kunnen we een betrouwbaarheidsinterval voor σ_x^2/σ_y^2 construeren. Als dit betrouwbaarheidsinterval het punt 1 bevat is er geen duidelijke reden om de aanname dat het quotiënt gelijk aan één is te verwerpen. Als spilfunctie kunnen we hiervoor gebruiken

$$\underline{v} = \frac{(\underline{s}_x^2/\sigma_x^2)}{(\underline{s}_y^2/\sigma_y^2)}$$

want $(n-1) \times$ de teller is $\chi^2(n-1)$-verdeeld en $(m-1) \times$ de noemer is $\chi^2(m-1)$-verdeeld terwijl teller en noemer onafhankelijk zijn, zodat $\underline{v} \sim F(n-1;m-1)$.

Voorbeeld

Een steekproef met omvang 10 uit een Normale populatie resulteerde in $\bar{x} = 15{,}7$ en $s_x^2 = 12{,}5$. Onafhankelijk daarvan wordt een steekproef met omvang 16 uit een andere Normale populatie getrokken die $\bar{y} = 18{,}4$ en $s_y^2 = 10{,}3$ opleverde. Een 90%-betrouwbaarheidsinterval voor het quotiënt van de varianties ontstaat aldus:

$$\underline{v} = \frac{(\underline{s}_x^2/\sigma_x^2)}{(\underline{s}_y^2/\sigma_y^2)} \sim F(9;\, 15); \qquad P(a < \underline{v} < b) = 0{,}90$$

a en b volgen uit: $P(\underline{v} > b) = 0{,}05$ dus $b = 2{,}59$ en $P(\underline{v} < a) = 0{,}05$ dus $P(\frac{1}{\underline{v}} > \frac{1}{a}) = 0{,}05$ met $\frac{1}{\underline{v}} \sim F(15;\, 9)$ zodat $\frac{1}{a} = 3{,}01$ en $a = 0{,}33$ dus

$$P(0{,}33 < \frac{\sigma_y^2}{\sigma_x^2}\,\frac{\underline{s}_x^2}{\underline{s}_y^2} < 2{,}59) = 0{,}90$$

of

$$P(0{,}33\cdot\frac{\underline{s}_y^2}{\underline{s}_x^2} < \frac{\sigma_y^2}{\sigma_x^2} < 2{,}59\cdot\frac{\underline{s}_y^2}{\underline{s}_x^2}) = 0{,}90.$$

Nu wordt het tijd om naar de realisaties te kijken:

$$(0{,}33\cdot\frac{10{,}3}{12{,}5}\,;\ 2{,}59\cdot\frac{10{,}3}{12{,}5})$$

ofwel: (0,27; 2,13) is een 90%-betrouwbaarheidsinterval voor $\frac{\sigma_y^2}{\sigma_x^2}$.

Het betrouwbaarheidsinterval bevat het punt 1 dus kunnen we een betrouwbaarheidsinterval voor $\mu_x-\mu_y$ construeren en daarbij aannemen dat $\sigma_x^2 = \sigma_y^2$. De gebruikte spilfunctie is nu

$$\underline{t} = \frac{(\underline{\overline{x}}-\underline{\overline{y}}) - (\mu_x-\mu_y)}{\sqrt{\frac{1}{10}+\frac{1}{16}}\cdot\sqrt{\frac{9\underline{s}_x^2 + 15\underline{s}_y^2}{10+16-2}}} \sim t(10+16-2).$$

Een 90%-betrouwbaarheidsinterval voor $\mu_x-\mu_y$ volgt dus uit

$$P(-1{,}711 < \underline{t} < 1{,}711) = 0{,}90$$

ofwel

$$P\{(\underline{\overline{x}}-\underline{\overline{y}}) - 1{,}711\cdot\underline{n} < \mu_x-\mu_y < (\underline{\overline{x}}-\underline{\overline{y}}) + 1{,}711\cdot\underline{n}\} = 0{,}90$$

waarin $\underline{n}$ de noemer is van $\underline{t}$.
De realisaties leveren $\overline{x} - \overline{y} = -2{,}7$ en $n = 1{,}3445$, zodat het 90%-betrouwbaarheidsinterval wordt

$$(-2{,}7-2{,}3;\ -2{,}7+2{,}3) \text{ ofwel } (-5{,}0;\ -0{,}4).$$

Het punt nul valt buiten dit betrouwbaarheidsinterval.
Een 95%-betrouwbaarheidsinterval voor $\mu_x-\mu_y$ wordt

$$(-2{,}7 - \frac{2{,}064}{1{,}711}\cdot 2{,}3;\ -2{,}7 + \frac{2{,}064}{1{,}711}\cdot 2{,}3) \text{ of } (-5{,}47;\ +0{,}07).$$

We kunnen dus de uitspraak $\mu_x \neq \mu_y$ wel doen met een betrouwbaarheid van 90%, maar niet met een betrouwbaarheid van 95%. □

Men kan de F-verdeling ook gebruiken om een betrouwbaarheidsinterval voor het quotiënt van σ_x^2 en σ_y^2 te construeren, zonder dat men de bedoeling heeft om eventueel daarna een betrouwbaarheidsinterval voor $\mu_x-\mu_y$ te construeren. Als μ_x en μ_y onbekend zijn gaat men daarbij uit van $\underline{s}_x^2$ en $\underline{s}_y^2$. Als er een populatieverwachting bekend is wordt er geen gebruik gemaakt van de corresponderende $\underline{s}^2$ maar van $\underline{r}^2 = \frac{1}{n}\sum(\underline{z}_i-\mu_z)^2$ waarvoor geldt dat

$$\frac{n\underline{r}^2}{\sigma_z^2} \sim \chi^2(n).$$

Men komt dus weer bij dezelfde soort spilfuncties terecht maar op een of twee plaatsen heeft men een vrijheidsgraad méér.

Men mag in het algemeen niet door combinatie van twee betrouwbaarheidsintervallen, bijvoorbeeld één voor μ en één voor σ^2, komen tot een betrouwbaarheidsgebied voor het parameterpaar. Dat mag alleen als de gebruikte spilfuncties onafhankelijk zijn. Omdat bij een steekproef uit een Normale populatie $\underline{x}$ en $\underline{s}^2$ inderdaad onafhankelijk zijn kan in dat geval een betrouwbaarheidsgebied voor $(\mu;\sigma^2)$ worden geconstrueerd.

7

Toetsen van hypothesen

7.1. Inleiding

Een hypothese is een veronderstelling omtrent de verdeling van een stochastische variabele. Een toets is een procedure om op grond van waarnemingen aan (realisaties van) die stochastische variabele te komen tot een uitspraak over de houdbaarheid van de hypothese. Voor het toetsen van een hypothese omtrent de verdeling van $\underline{x}$ heeft men dus een steekproef nodig uit een populatie met verdelingsfunctie $F_{\underline{x}}$.

Voorbeeld

Van een munt beweert Jan dat P(kruis) = 0,6 terwijl Piet meent dat P(kruis) = 0,4.

Hypothese van Jan H_0: P(kruis) = p = 0,6;
Hypothese van Piet H_1: P(kruis) = p = 0,4.

Om de strijd te beslechten gooien ze de munt 20 keer en tellen het aantal malen $\underline{x}$ dat daarbij kruis valt. Het is duidelijk dat bij een grote waarde van x, de realisatie van $\underline{x}$, Jan gelijk zal krijgen, terwijl bij een kleine waarde van x Piet gelijk krijgt. Alleen de grens tussen 'groot' en 'klein' moet nog worden bepaald.
Er zijn vier mogelijkheden:

	H_0 is waar	H_1 is waar
H_0 verwerpen	fout van de eerste soort	goed
H_0 aanvaarden	goed	fout van de tweede soort

Stel dat de grens zó gekozen wordt dat bij meer dan tien keer kruis Jan gelijk krijgt (H_0 aanvaard wordt) en bij tien of minder keren kruis Piet gelijk krijgt (H_1 aanvaard wordt).
Als Jan gelijk heeft, dus H_0 is waar, dan geldt dat $\underline{x} \sim B(20; 0{,}6)$ zodat:

$$P(\text{Jan krijgt gelijk}) = P(\underline{x} > 10) = 0{,}7553;$$
$$P(\text{Piet krijgt gelijk}) = P(\underline{x} \leq 10) = 0{,}2447 = \alpha.$$

Als Piet gelijk heeft, dus H_1 is waar, dan geldt dat $\underline{x} \sim B(20; 0{,}4)$ zodat:

$$P(\text{Jan krijgt gelijk}) = P(\underline{x} > 10) = 0{,}1275 = \beta;$$
$$P(\text{Piet krijgt gelijk}) = P(\underline{x} \leq 10) = 0{,}8725.$$

Hierbij is dus α de kans dat er een fout van de eerste soort wordt gemaakt en β de kans op een fout van de tweede soort. De vier bovenstaande kansen zijn afhankelijk van de gekozen grenswaarde.
Het liefst zouden we zowel α als β klein maken maar een wijziging van de grenswaarde heeft tot gevolg dat ofwel α kleiner en β groter wordt ofwel β kleiner maar α groter.
Als in dit voorbeeld de grens wordt gelegd bij z, zó dat
bij $\underline{x} > z$ Jan gelijk krijgt en bij $\underline{x} \leq z$ Piet gelijk krijgt dan volgt dat

$$\alpha_z = P(\underline{x} \leq z) \text{ met } \underline{x} \sim B(20;\, 0{,}6) \text{ en}$$
$$\beta_z = P(\underline{x} > z) \text{ met } \underline{x} \sim B(20;\, 0{,}4).$$

We krijgen dan:

z	α_z	β_z
7	0,0210	0,5841
8	0,0565	0,4044
9	0,1275	0,2447
10	0,2447	0,1275
11	0,4044	0,0565
12	0,5841	0,0210

Het zo ontstane dilemma wordt opgelost door één van de twee fouten als 'ernstig' aan te merken en de grens zo te kiezen dat de kans op het maken van die ernstige fout een vooraf vastgestelde waarde niet overschrijdt. Het is nu gebruikelijk om de namen H_0 en H_1, respectievelijk nulhypothese en alternatieve hypothese, zo aan de veronderstellingen toe te voegen dat de fout van de eerste soort de ernstige fout is. M.a.w. het ten onrechte verwerpen van H_0 heeft onaangename consequenties. De kans daarop (α) wordt onbetrouwbaarheidsdrempel genoemd. Als de maximaal toelaatbare waarde van α gegeven is kan de grens tussen de gebieden A en K worden bepaald. A is het aanvaardingsgebied en bestaat uit alle realisaties van $\underline{x}$ die leiden tot het aanvaarden van H_0. K is het kritieke gebied en bestaat uit alle realisaties van $\underline{x}$ die leiden tot het verwerpen van H_0. Nodig is dat $A \cup K = W_x$ en $A \cap K = \Phi$ zodat bij elke realisatie van $\underline{x}$ eenduidig vastligt wat de conclusie is.
Als hier gegeven is dat $\alpha \leq 6\%$ moet zijn dan volgt dat z = 8, want $\alpha_8 = 0{,}0565$ en $\alpha_9 = 0{,}1275$. Het waardenbereik van $\underline{x}$ wordt dus gesplitst in $K = \{0,1,2,\ldots,8\}$: het kritieke gebied en $A = \{9,10,11,\ldots,20\}$: het aanvaardingsgebied.
Natuurlijk zou met $K = \{0,1,2,\ldots,7\}$ en $A = \{8,9,10,\ldots,20\}$ ook voldaan zijn aan de eis dat α = P(fout van de eerste soort) $\leq 0{,}06$ maar dan wordt β = P(fout van de tweede soort) groter dan nodig is. Het kritieke gebied wordt altijd zo groot mogelijk gekozen. ❑

In het voorbeeld bestond alleen verschil van mening omtrent de waarde van de parameter p. Dat $\underline{x}$ binomiaal verdeeld is ligt vast. Men spreekt in zo'n geval van parametrisch toetsen. Daarnaast bestaan er non-parametrische toetsen, bijvoorbeeld om de hypothese dat een stochastische variabele Normaal verdeeld is te toetsen.

7.2. Parametrische toetsen

Voor het toetsen van een hypothese is een toetsingsgrootheid nodig. Een toetsingsgrootheid is een functie van de stochastische variabelen in de steekproef waarvan de verdeling volledig bekend is als H_0, de nulhypothese, waar is. Bij parametrische toetsen is de toetsingsgrootheid dus een functie van de steekproefvariabelen met een verdeling die, afgezien van de waarde van de omstreden parameter θ, volledig bekend is. Om een toetsingsgrootheid te vinden, kan men meestal uitgaan van de meest-aannemelijke schatter voor de parameter waarover de toets gaat. In het geval dat de nulhypothese enkelvoudig is, d.w.z. van de vorm H_0: $\theta = c$, is er geen probleem. In het geval dat de nulhypothese samengesteld is, d.w.z van de vorm H_0: $\theta \leq c$ of H_0: $\theta \geq c$, ligt de waarde van θ en daarmee de verdeling van de toetsingsgrootheid, ook als we aannemen dat H_0 waar is, niet vast. Zo'n samengestelde nulhypothese wordt dan vervangen door de meest ongunstige enkelvoudige nulhypothese. 'Meest ongunstig' betekent daarbij 'het dichtst bij H_1'. Dus

H_0: $\theta \leq c$ $\quad$ H_1: $\theta > c$ wordt vervangen door

H_0: $\theta = c$ $\quad$ H_1: $\theta > c$.

H_0 en H_1 moeten zó gekozen worden dat er geen overlapping optreedt, m.a.w. er mag geen waarde voor θ zijn waarvoor zowel H_0 als H_1 waar is.

Vervolgens wordt het kritieke gebied bepaald. Daartoe worden intuïtief geschikte waarden uit het waardenbereik van de toetsingsgrootheid verzameld. Dat zijn dus waarden waarvan de kans van optreden als H_1 waar is groot is t.o.v. de kans van optreden als H_0 waar is. Het kritieke gebied K wordt zo groot mogelijk gemaakt. Als de toetsingsgrootheid $\underline{tg}$ een continue stochastische variabele is geldt dus $P(\underline{tg} \in K) = \alpha$ als H_0 waar is.

Het kritieke gebied bestaat uit één stuk bij eenzijdig toetsen zoals H_0: $\theta = c$ tegen H_1: $\theta > c$ maar het bestaat uit twee stukken bij tweezijdig toetsen, bijvoorbeeld H_0: $\theta = c$ tegen H_1: $\theta \neq c$.

Als de toetsingsgrootheid $\underline{tg}$, de onbetrouwbaarheidsdrempel α en het kritieke gebied vastliggen wordt het tijd om uit de waarnemingen (de steekproef) de realisatie tg van $\underline{tg}$ te berekenen. Als tg $\in$ K wordt H_0 verworpen, zo niet dan wordt H_0 aanvaard.

Het aanvaarden van H_0 houdt niet in dat H_0 waar is. Er kan een fout van de tweede soort zijn gemaakt. De kans daarop is afhankelijk van de werkelijke waarde van de parameter; als $\theta = \theta_0$ dan is de kans op een fout van de tweede soort: $\beta(\theta_0) = P(\underline{tg} \in A$; aangenomen dat $\theta = \theta_0)$.

Voorbeeld

Een fabrikant produceert lampen met een levensduur die Normaal verdeeld is met verwachting $\mu = 2200$ uur en standaardafwijking $\sigma = 250$ uur. Een slimme werknemer dient een verbeteringsvoorstel in dat volgens hem zal resulteren in een hogere waarde van μ bij gelijkblijvende σ. Het bedrijf besluit om 100 lampen volgens het nieuwe procédé te maken en die te toetsen met $\alpha = 0{,}05$. Als nulhypothese kiest men H_0: $\mu \le 2200$, met daarbij H_1: $\mu > 2200$.

Concluderen dat de nieuwe lamp beter is dan de gebruikelijke, terwijl dat in werkelijkheid niet zo is, is een ernstige fout omdat dan ten onrechte kosten gemaakt worden voor de wijziging van het produktieproces. Bij deze keuze van H_0, H_1 en α is dus de kans dat het produktieproces ten onrechte wordt gewijzigd hoogstens 0,05.

Het gaat dus om een stochastische variabele $\underline{x} \sim N(\mu; 250^2)$, de levensduur. Op grond van een steekproef met omvang $n = 100$ moet met een onbetrouwbaarheidsdrempel $\alpha = 0{,}05$ H_0: $\mu \le 2200$ tegen H_1: $\mu > 2200$ worden getoetst. De meestaannemelijke schatter voor μ is $\underline{\bar{x}}$. We gebruiken dus als toetsingsgrootheid $\underline{\bar{x}}$.

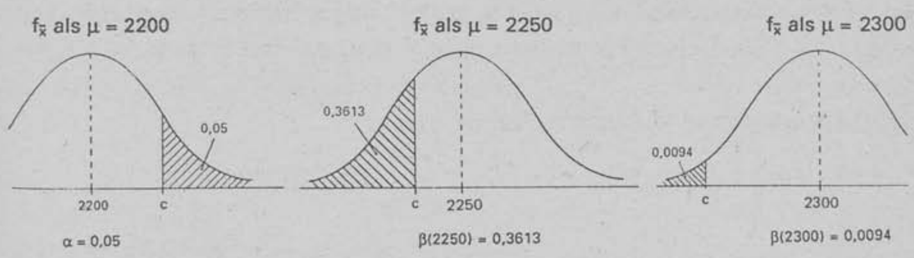
De nulhypothese is samengesteld, voor het bepalen van het kritieke gebied kiezen we de ongunstigste waarde van μ, dus $\mu = 2200$. Als we uitgaan van $\mu = 2200$ volgt dat $\underline{\bar{x}} \sim N(2200; 25^2)$. Grote waarden van $\bar{x}$ maken H_1 geloofwaardiger dan H_0 dus K, het kritieke gebied, is van de vorm $(c;\infty)$. De grens c wordt bepaald uit $P(\underline{\bar{x}} > c) \le 0{,}05$ als $\mu = 2200$.

Maar het kritieke gebied moet zo groot mogelijk zijn en omdat $\underline{\bar{x}}$ een continue stochastische variabele is kan c zó worden gekozen dat $P(\underline{\bar{x}} > c) = 0{,}05$ bij $\mu = 2200$. Dit leidt tot

$$P\left(\frac{\underline{\bar{x}}-2200}{25} > \frac{c-2200}{25}\right) = 0{,}05 \text{ ofwel } \frac{c-2200}{25} = 1{,}645$$

want als we aannemen dat $\mu = 2200$ volgt dat $\dfrac{\underline{\bar{x}}-2200}{25} \sim N(0;1)$. Hieruit volgt dat $c = 2241{,}125$. Het kritieke gebied is dus $(2241{,}125; \infty)$.

Als de realisatie $\bar{x} > 2241{,}125$ wordt H_0 verworpen. Is de realisatie $\bar{x} < 2241{,}125$ dan wordt H_0 niet verworpen, toch kan ook dan H_1 waar zijn. Het is mogelijk dat er een fout van de tweede soort is gemaakt. De kans daarop is afhankelijk van de werkelijke waarde van μ.



Figuur 7.1.

$\beta(\mu) = P(\underline{\bar{x}} < c)$ als $E(\underline{\bar{x}}) = E(\underline{x}) = \mu$ dus $\underline{\bar{x}} \sim N(\mu; 25^2)$.

Zo wordt $\beta(2250) \quad = P\left\{\frac{\underline{\bar{x}}-2250}{25} < \frac{2241{,}125-2250}{25}\right\} =$

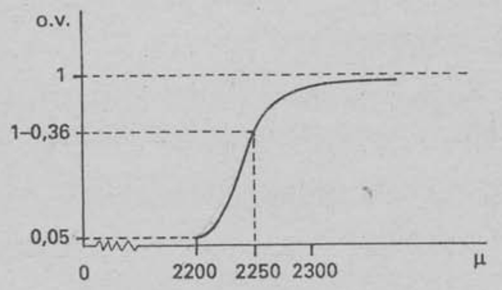
$= P(\underline{u} > 0{,}355) = 0{,}3613.$ ❑

Veelal wordt niet β maar $1-\beta$ bepaald, het *onderscheidingsvermogen* (power) van de toets geheten.

$1-\beta(\theta)$ = P(het verwerpen van de nulhypothese als de omstreden parameter de waarde θ heeft), waarbij θ een zodanige waarde moet hebben dat H_1 waar is.

Voorbeeld

Vervolg van het vorige voorbeeld. Het is duidelijk dat als μ groter wordt de kans dat H_0 wordt verworpen stijgt, dus $1-\beta$ is hier een stijgende functie van μ die gedefinieerd is voor $\mu > 2200$.



Figuur 7.2. ❑

Natuurlijk kan men eisen dat, bij vaste α, voor een zekere waarde van de parameter de kans op een fout van de tweede soort beneden een vooraf gekozen grens moet blijven. Dan moet men echter de mogelijkheid hebben om de steekproefomvang te vergroten

Voorbeeld

Vervolg van het vorige voorbeeld. Stel dat geëist wordt dat voor μ = 2250 het onderscheidingsvermogen van de toets 0,90 is, d.w.z. dat $\beta(2250) = 0{,}10$.

Het zojuist gevonden kritieke gebied is nu niet langer bruikbaar, daarbij hoorde $\beta(2250) = 0{,}36$. De steekproefomvang, stel deze m, moet opnieuw worden bepaald.
Zij het kritieke gebied $(d;\infty)$, dan volgt:

$$P(\underline{\bar{x}} > d) = 0{,}05 \text{ uit de eis } \alpha = 0{,}05 \ \{\text{nu } \underline{\bar{x}} \sim N(2200; \frac{250^2}{m})\}$$

en

$$P(\underline{\bar{x}} < d) = 0{,}10 \text{ uit de eis } \beta(2250) = 0{,}10 \ \{\text{nu } \underline{\bar{x}} \sim N(2250; \frac{250^2}{m})\}.$$

Overgang op de standaard-Normale verdeling levert:

$$\frac{d-2200}{250/\sqrt{m}} = 1{,}645 \quad \text{en} \quad \frac{d-2250}{250/\sqrt{m}} = -1{,}282$$

waaruit volgt

$$50 = \frac{250}{\sqrt{m}} \cdot 2{,}927 \quad \text{ofwel} \quad m = 214{,}18.$$

Natuurlijk moet de steekproefomvang een geheel getal zijn. Er moet altijd naar boven worden afgerond. Voor m = 214 voldoen we nog net niet aan de eisen en de steekproef moet dus uit 215 lampen bestaan. Het nieuwe kritieke gebied heeft als grens

$$d = 2200 + \frac{250}{\sqrt{215}} \cdot 1{,}645 = 2228{,}05.$$

❑

7.3. Enige standaardtoetsen

In alle volgende gevallen wordt uitgegaan van een steekproef met omvang n uit een Normale populatie $(\mu;\sigma^2)$ en een onbetrouwbaarheidsdrempel $\alpha\%$.

7.3.1. Toetsen voor de verwachting bij bekende variantie

Toetsingsgrootheid $\underline{\bar{x}} \sim N(\mu;\sigma^2/n)$. Het kritieke gebied wordt bepaald m.b.v.

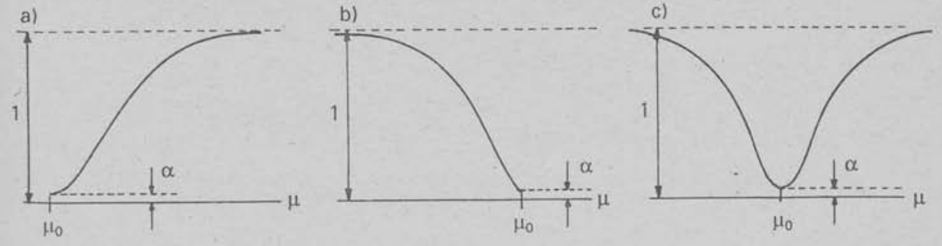
$$\underline{u} = \frac{\underline{\bar{x}}-\mu}{\sigma/\sqrt{n}} \sim N(0;1).$$

a. $H_0: \mu = \mu_0$ (of $\mu \le \mu_0$) $\quad H_1: \mu > \mu_0$.
Kritieke gebied $u \ge u_\alpha$ $\quad$ d.w.z. $\bar{x} \ge \mu_0 + u_\alpha \cdot \sigma/\sqrt{n}$;

b. $H_0: \mu = \mu_0$ (of $\mu \ge \mu_0$) $\quad H_1: \mu < \mu_0$.
Kritieke gebied $u \le -u_\alpha$ $\quad$ d.w.z. $\bar{x} \le \mu_0 - u_\alpha \cdot \sigma/\sqrt{n}$;

c. $H_0: \mu = \mu_0$ $\quad H_1: \mu \ne \mu_0$.
Kritieke gebied $u \le -u_{\alpha/2} \cup u \ge u_{\alpha/2}$ d.w.z.
$\bar{x} \le \mu_0 - u_{\alpha/2}\sigma/\sqrt{n} \cup \bar{x} \ge \mu_0 + u_{\alpha/2}\sigma/\sqrt{n}$.

Het onderscheidingsvermogen heeft in deze drie gevallen het volgende verloop:



Figuur 7.3.

7.3.2. Toetsen voor de verwachting bij onbekende variantie

Toetsingsgrootheid $\underline{\bar{x}}$. M.b.v. $\underline{t} = \dfrac{\underline{\bar{x}}-\mu}{\underline{s}/\sqrt{n}} \sim t(n-1)$ wordt het kritieke gebied bepaald.

Het enige verschil met de toetsen in 7.3.1 is dat overal σ^2 (onbekend) is vervangen door $\underline{s}^2$ (zuivere schatter voor σ^2) dan wel s^2 (de realisatie), en dat de t-verdeling in plaats van de N(0;1)-verdeling wordt gebruikt.

a. H_0: $\mu = \mu_0$ (of $\mu \le \mu_0$) H_1: $\mu > \mu_0$.

Kritieke gebied $t \ge t_{n-1;\alpha}$ d.w.z. $\bar{x} \ge \mu_0 + t_{n-1;\alpha} \cdot s/\sqrt{n}$.

b en **c** gaan analoog.

N.B. De verdeling van $\underline{t}$ is symmetrisch t.o.v. 0. Voor $n \to \infty$ nadert de verdeling van $\underline{t}$ naar de standaard-Normale verdeling.

7.3.3. Toetsen voor de variantie bij bekende verwachting

Toetsingsgrootheid $\underline{v} = \sum_{i=1}^{n} \left(\dfrac{\underline{x}_i-\mu}{\sigma}\right)^2 \sim \chi^2(n)$ als $\mathrm{Var}(\underline{x}_i) = \sigma^2$.

Dus onder H_0: $\sigma^2 = \sigma_0^2$ geldt dat $\underline{v} = \sum_{i=1}^{n} \dfrac{(\underline{x}_i-\mu)^2}{\sigma_0^2} \sim \chi^2(n)$ zodat het kritieke gebied kan worden bepaald.

a. H_0: $\sigma^2 = \sigma_0^2$ (of $\sigma^2 \le \sigma_0^2$) H_1: $\sigma^2 > \sigma_0^2$.

Kritieke gebied $v \ge \chi^2_{n;\alpha}$;

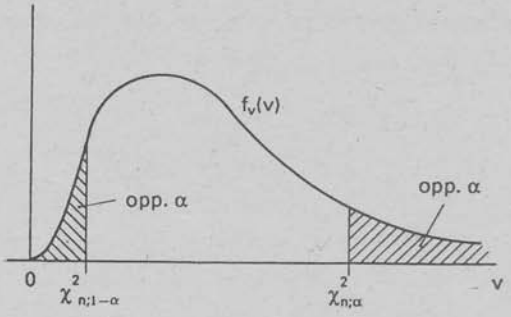
b. H_0: $\sigma^2 = \sigma_0^2$ (of $\sigma^2 \ge \sigma_0^2$) H_1: $\sigma^2 < \sigma_0^2$.

Kritieke gebied $v \le \chi^2_{n;1-\alpha}$;

c. H_0: $\sigma^2 = \sigma_0^2$ H_1: $\sigma^2 \ne \sigma_0^2$.

Kritieke gebied $v \le \chi^2_{n;1-\alpha/2} \cup v \ge \chi^2_{n;\alpha/2}$.

N.B. $\underline{v}$ en v zijn altijd niet-negatief.



Figuur 7.4.

7.3.4. Toetsen voor de variantie bij onbekende verwachting

Toetsingsgrootheid $\underline{v} = \sum_{i=1}^{n} \left\{\frac{\underline{x}_i - \overline{\underline{x}}}{\sigma}\right\}^2 = (n-1)\underline{s}^2/\sigma^2 \sim \chi^2(n-1)$ als $\mathrm{Var}(\underline{x}_i) = \sigma^2$.

Dus onder H_0: $\sigma^2 = \sigma_0^2$ geldt dat $\underline{v} = (n-1)\underline{s}^2/\sigma_0^2 \sim \chi^2(n-1)$ zodat het kritieke gebied kan worden bepaald. Het verschil met 7.3.3 is dat μ (nu onbekend) is vervangen door $\overline{\underline{x}}$, de zuivere schatter voor μ, of door de realisatie $\overline{x}$. We moeten daarbij één vrijheidsgraad inleveren.

a. H_0: $\sigma^2 = \sigma_0^2$ (of $\sigma^2 \leq \sigma_0^2$) H_1: $\sigma^2 > \sigma_0^2$.

Kritieke gebied $v \geq \chi^2_{n-1;\alpha}$;

b en **c** gaan analoog.

7.4. Twee steekproeven

Net als bij de constructie van betrouwbaarheidsintervallen moet hier onderscheid worden gemaakt tussen het geval van gepaarde waarnemingen en het geval dat er twee onafhankelijke steekproeven getrokken zijn. Bij gepaarde waarnemingen heeft men in feite één steekproef van paren $(\underline{x}_i, \underline{y}_i)$, $i = 1,2,\dots,n$. Daaruit bepaalt men een steekproef van verschillen $\underline{z}_i = \underline{x}_i - \underline{y}_i$, $i = 1,2,\dots,n$. Als aangenomen mag worden dat $\underline{z}_i \sim N(\mu;\sigma^2)$, dan kunnen de toetsen uit de vorige paragraaf worden gebruikt. De hypothese $\mu_z = 0$ is hetzelfde als de hypothese $\mu_x = \mu_y$. Omdat $\sigma^2(\underline{z}) = \sigma^2(\underline{x}) + \sigma^2(\underline{y}) - 2\,\mathrm{cov}(\underline{x},\underline{y})$ kan een hypothese omtrent $\sigma^2(\underline{z})$ niet eenvoudig worden uitgedrukt in termen van $\sigma^2(\underline{x})$ en $\sigma^2(\underline{y})$.

N.B. Het is niet nodig dat $\underline{x}_i$, resp. $\underline{y}_i$, uit een Normale populatie komt. Voldoende is dat $\underline{z}_i$ Normaal verdeeld is.

Bij twee onafhankelijke steekproeven

$\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ uit een $N(\mu_x;\sigma_x^2)$-populatie, en

$\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_m$ uit een $N(\mu_y;\sigma_y^2)$-populatie

zijn andere toetsen nodig. De toetsingsgrootheden bij de nu volgende toetsen komen overeen met de spilfuncties, als beschreven in paragraaf 6.4, die worden gebruikt bij de constructie van betrouwbaarheidsintervallen gebaseerd op twee onafhankelijke steekproeven.

7.4.1. Het verschil van de verwachtingen bij bekende varianties

$$\text{Toetsingsgrootheid } \underline{u} = \frac{\overline{\underline{x}}-\overline{\underline{y}}-(\mu_x-\mu_y)}{\sqrt{\frac{1}{n}\sigma_x^2 + \frac{1}{m}\sigma_y^2}} \sim N(0;1).$$

a. H_0: $\mu_x - \mu_y = k$ $\quad H_1$: $\mu_x - \mu_y > k$.

Kritieke gebied $u \geq u_\alpha$ $\quad$ d.w.z. $\overline{x} - \overline{y} \geq k + u_\alpha\sqrt{\frac{1}{n}\sigma_x^2 + \frac{1}{m}\sigma_y^2}$;

b en **c** gaan analoog.

7.4.2. Het verschil van de verwachtingen bij onbekende maar gelijke varianties

$$\text{Toetsingsgrootheid } \underline{t} = \frac{\overline{\underline{x}}-\overline{\underline{y}}-(\mu_x-\mu_y)}{\sqrt{\left(\frac{1}{n} + \frac{1}{m}\right)\frac{(n-1)\underline{s}_x^2 + (m-1)\underline{s}_y^2}{n+m-2}}} \sim t(n+m-2).$$

a. H_0: $\mu_x - \mu_y = k$ $\quad H_1$: $\mu_x - \mu_y > k$.

Kritieke gebied $t \geq t_{n+m-2;\alpha}$ d.w.z. $\overline{x} - \overline{y} \geq k + t_{n+m-2;\alpha}\sqrt{\ldots\ldots}$.

b en **c** gaan analoog.

7.4.3. Het quotiënt van de varianties bij bekende verwachtingen

$$\text{Toetsingsgrootheid } \underline{f} = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\underline{x}_i-\mu_x}{\sigma_x}\right)^2}{\frac{1}{m}\sum_{j=1}^{m}\left(\frac{\underline{y}_j-\mu_y}{\sigma_y}\right)^2} \sim F(n;m).$$

Hiermee kan H_0: $\sigma_y^2/\sigma_x^2 = k$ worden getoetst tegen de gebruikelijke alternatieven. Dit geval komt echter vrijwel niet voor. Meestal wordt getoetst of de varianties gelijk zijn bij onbekende verwachtingen teneinde daarna (als de hypothese aanvaard is) een toets als in 7.4.2 uit te kunnen voeren.

7.4.4. Het quotiënt van de varianties bij onbekende verwachtingen

Toetsingsgrootheid $\underline{f} = \dfrac{\underline{s}_x^2/\sigma_x^2}{\underline{s}_y^2/\sigma_y^2} = \dfrac{\sigma_y^2}{\sigma_x^2} \cdot \dfrac{\frac{1}{n-1}\sum_{i=1}^{n}(\underline{x}_i-\overline{\underline{x}})^2}{\frac{1}{m-1}\sum_{j=1}^{m}(\underline{y}_j-\overline{\underline{y}})^2} \sim F(n-1;m-1).$

Omdat van de F-verdeling alleen de rechterstaart is getabelleerd zorgt men er steeds voor dat het kritieke gebied van de vorm $(c;\infty)$ is. Dit geldt ook voor de toetsen in 7.4.3 vermeld.

a. $H_0: \dfrac{\sigma_x^2}{\sigma_y^2} = k \qquad H_1: \dfrac{\sigma_x^2}{\sigma_y^2} > k.$

Onder H_0 is $\underline{f}_1 = \dfrac{1}{k}\dfrac{\underline{s}_x^2}{\underline{s}_y^2} \sim F(n-1;m-1)$ en zijn grote waarden van de realisatie verdacht, immers een grote σ_x^2 zal de kans op grote waarden van $\underline{s}_x^2$ doen toenemen.

Kritieke gebied: $f_1 \geq F_{n-1;m-1;\alpha}$ d.w.z. $\dfrac{s_x^2}{s_y^2} \geq k \cdot F_{n-1;m-1;\alpha}$.

b. $H_0: \dfrac{\sigma_x^2}{\sigma_y^2} = k \qquad H_1: \dfrac{\sigma_x^2}{\sigma_y^2} < k$ vervangen we door het gelijkwaardige:

$H_0: \dfrac{\sigma_y^2}{\sigma_x^2} = \dfrac{1}{k} \qquad H_1: \dfrac{\sigma_y^2}{\sigma_x^2} > \dfrac{1}{k}$ met $\underline{f}_2 = \dfrac{\underline{s}_y^2/\sigma_y^2}{\underline{s}_x^2/\sigma_x^2} \sim F(m-1;n-1),$

zodat onder H_0 $\underline{f}_2 = k\dfrac{\underline{s}_y^2}{\underline{s}_x^2} \sim F(m-1;n-1).$

Kritieke gebied: $f_2 \geq F_{m-1;n-1;\alpha}$ d.w.z. $\dfrac{s_y^2}{s_x^2} \geq \dfrac{1}{k} \cdot F_{m-1;n-1;\alpha}$.

c. $H_0: \dfrac{\sigma_x^2}{\sigma_y^2} = k \qquad H_1: \dfrac{\sigma_x^2}{\sigma_y^2} \neq k.$

Als toetsingsgrootheid kan zowel $\underline{f}_1$ als $\underline{f}_2$ gebruikt worden.
$\underline{f}_1$ geeft als kritieke gebied $f_1 \leq F_{n-1;m-1;1-\alpha/2} \cup f_1 \geq F_{n-1;m-1;\alpha/2}$;
$\underline{f}_2$ geeft als kritieke gebied $f_2 \leq F_{m-1;n-1;1-\alpha/2} \cup f_2 \geq F_{m-1;n-1;\alpha/2}$.

Nu is $f_1 = \dfrac{1}{f_2} = \dfrac{1}{k}\dfrac{s_x^2}{s_y^2}$, dus of f_1 of f_2 is groter dan één.

Als $f_1 > 1$ dan nemen we als toetsingsgrootheid $\underline{f}_1$ en hoeven dan alleen het rechterstuk van het bijbehorende kritieke gebied te bepalen omdat $F_{n-1;m-1;1-\alpha/2} < 1$. Dit leidt tot

$$\frac{s_x^2}{s_y^2} \geq k\, F_{n-1;m-1;\alpha/2}.$$

Is $f_1 < 1$ dan nemen we als toetsingsgrootheid $\underline{f}_2$ en omdat dan $f_2 > 1$ is volgt dat we weer alleen met het rechterdeel van het kritieke gebied te maken hebben. Dit levert

$$\frac{s_y^2}{s_x^2} \geq \frac{1}{k}\, F_{m-1;n-1;\alpha/2}.$$

Het is dus in dit geval toegestaan om naar de realisaties te kijken voordat de procedure geheel is vastgelegd.

Voorbeeld

Een vertegenwoordiger komt bij een jamfabriek om een vulmachine te verkopen. Hij beweert dat de variantie van de hoeveelheid jam per pot bij de nieuwe machine minder dan de helft is van die bij de oude. Om geen geld in het water te gooien besluit men eerst de bewering van de vertegenwoordiger te onderzoeken d.m.v. een statistische toets met $\alpha = 0{,}05$. Aangenomen wordt dat het vulgewicht in de potjes bij de oude zowel als bij de nieuwe machine Normaal verdeeld is. Een steekproef $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_{10}$ van 10 potjes gevuld met de oude machine en een steekproef $\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_8$ van 8 potjes gevuld met de nieuwe machine zullen daarbij worden gebruikt.

Het ten onrechte nog een tijdje doormodderen met de oude machine is een minder ernstige fout dan het ten onrechte kopen van een dure nieuwe machine dus α (de kans op de ernstige fout) is de kans dat we de vertegenwoordiger geloven terwijl zijn bewering, $\sigma_x^2/\sigma_y^2 > 2$ niet waar is. α is de kans op het verwerpen van H_0, d.w.z. het aanvaarden van H_1, terwijl H_0 waar is. Hier wordt dus H_0: $\sigma_x^2/\sigma_y^2 \leq 2$; H_1: $\sigma_x^2/\sigma_y^2 > 2$. Dit vervangen we door H_0: $\sigma_x^2/\sigma_y^2 = 2$; H_1: $\sigma_x^2/\sigma_y^2 > 2$. Van beide steekproeven wordt de steekproefvariantie s^2 bepaald.

De toetsingsgrootheid is $\underline{f}_1 = \dfrac{\sigma_y^2}{\sigma_x^2}\dfrac{\underline{s}_x^2}{\underline{s}_y^2} \sim F(9;7)$ zodat onder H_0 $\dfrac{1}{2}\dfrac{\underline{s}_x^2}{\underline{s}_y^2} \sim F(9;7)$.

Het kritieke gebied wordt dus

$$\frac{s_x^2}{s_y^2} \geq 2 \cdot F_{9;7;0{,}05} = 2 \cdot 3{,}68 = 7{,}36.$$

Als $s_x^2 = 12{,}34$ en $s_y^2 = 1{,}59$ volgt dat $\dfrac{s_x^2}{s_y^2} = 7{,}76$. Dit ligt in het kritieke gebied, dus H_0 wordt verworpen.

De toets heeft als resultaat dat het vertrouwen in de juistheid van de bewering van de vertegenwoordiger is toegenomen. ❑

7.5. Twee andere toetsen

7.5.1. Voor de parameter van een exponentiële verdeling

Zij $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ een steekproef uit een exponentiële populatie met onbekende parameter λ. Omdat uit $2\lambda\underline{x} \sim \chi^2(2)$ volgt dat $\sum_{i=1}^{n} 2\lambda\underline{x}_i = 2n\lambda\underline{\bar{x}} \sim \chi^2(2n)$, kunnen op de gebruikelijke wijze toetsen voor λ worden geconstrueerd.

$$H_0: \lambda = \lambda_0; \qquad\qquad H_1: \lambda > \lambda_0.$$

Onder H_0 is $2\lambda_0 n\underline{\bar{x}} \sim \chi^2(2n)$. Als H_1 waar is, dus $\lambda > \lambda_0$, zal $E(\underline{\bar{x}}) = E(\underline{x}) = \frac{1}{\lambda}$ kleiner zijn dan $\frac{1}{\lambda_0}$. Kleine waarden van $2\lambda_0 n\bar{x}$ zijn dus verdacht, m.a.w. het kritieke gebied is van de vorm $(0,c)$. De waarde van c volgt uit α m.b.v. de tabel van de χ^2-verdeling.

7.5.2. Voor de parameter p van een alternatief verdeelde populatie

Het komt vaak voor dat men de populatie slechts in twee delen wil splitsen: de goeden en de slechten. Men is dan geïnteresseerd in de grootte van de fractie goeden. Een stochastische variabele die dit weergeeft is de alternatief verdeelde stochastische variabele $\underline{x}$ waarvoor geldt: $P(\underline{x} = 0) = 1 - p = q$ en $P(\underline{x} = 1) = p$. Laat $\underline{x} = 0$ corresponderen met slecht en $\underline{x} = 1$ met goed, dan is p de fractie goeden in de populatie. Voor het toetsen van hypothesen omtrent p kan een aselecte steekproef $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ worden gebruikt. De toetsingsgrootheid $\underline{b} = \underline{x}_1 + \underline{x}_2 + \ldots + \underline{x}_n$ ligt voor de hand. Immers $\underline{b}$ is binomiaal verdeeld met parameters n en p. Onder een enkelvoudige (of daartoe teruggebrachte samengestelde) nulhypothese ligt de verdeling van $\underline{b}$ geheel vast. M.b.v. de tabel van de binomiale verdeling kan bij gegeven α het kritieke gebied worden bepaald, mits n en p waarden hebben die in de tabel voorkomen.

Voorbeeld

De dienst Marketing-reizigersvervoer van de NS denkt dat minstens 80% van de treinreizigers tevreden is over de nieuwe opzet van het spoorboekje. De stichting Rover gelooft dat niet. Alvorens een boze brief naar de NS te schrijven voert men een toets uit. Zij p de fractie reizigers die tevreden is over de nieuwe opzet.
Men trekt een steekproef met n = 100 uit de populatie van treinreizigers, stelt $\alpha = 0{,}05$ en toetst H_0: $p = 0{,}8$ tegen H_1: $p < 0{,}8$. Het kritieke gebied wordt dus $\{0,1,2,\ldots,c\}$ waarbij c het grootste gehele getal is dat voldoet aan $P(\underline{b} \leq c) \leq 0{,}05$ met $\underline{b} \sim B(100; 0{,}8)$. De tabel is nu niet direct bruikbaar. We moeten overgaan op

$\underline{b}' = 100 - \underline{b}$, met $\underline{b}' \sim B(100; 0{,}2)$. Het kritieke gebied voor $\underline{b}'$ wordt dan $\{100-c,\ldots,98,99,100\}$. $P(\underline{b}' \geq 100-c) \leq 0{,}05$ ofwel $P(\underline{b}' \leq 99-c) \geq 0{,}95$ moet nu de waarde van c leveren.

Uit de tabel volgt: $P(\underline{b}' \leq 26) = 0{,}9442$ (voldoet niet)

$P(\underline{b}' \leq 27) = 0{,}9658$ (voldoet wel).

Dus $99-c = 27$ m.a.w. $c = 72$.

De nulhypothese wordt dus verworpen als er in de steekproef 72 of minder tevreden reizigers zijn.

Het onderscheidingsvermogen van de toets kan ook m.b.v. de tabel worden bepaald. Bijvoorbeeld voor $p = 0{,}7$. Het onderscheidingsvermogen is nu gelijk aan $P(\underline{y} \leq 72)$ met $\underline{y} \sim B(100; 0{,}7)$.

Dat is $P(\underline{y}' \geq 28) = 1 - P(\underline{y}' \leq 27)$ met $\underline{y}' \sim B(100; 0{,}3)$.

Uit de tabel blijkt dit $1 - 0{,}2964 = 0{,}7036$ te zijn. ❑

Helaas is de tabel niet altijd bruikbaar. Als n groot is en p wijkt niet al te veel af van 0,5 kan zonder bezwaar de Normale benadering van de binomiale verdeling worden gebruikt.

Voorbeeld

De steekproef van het vorige voorbeeld werd uitgevoerd via een schriftelijke enquête. Helaas stuurden slechts 81 van de 100 uitverkorenen hun formulier correct ingevuld terug. Als we de andere 19 geheel buiten beschouwing laten dan kunnen we werken met de toetsingsgrootheid $\underline{z}$ die onder H_0 $B(81; 0{,}8)$ verdeeld is. De verdeling van $\underline{z}$ wordt benaderd door de $N(64{,}8; 12{,}96)$ verdeling.

De toets wordt H_0: $\mu = 64{,}8$; H_1: $\mu < 64{,}8$ en met $\alpha = 0{,}05$ wordt de bovengrens van het kritieke gebied $64{,}8 - 1{,}645 \cdot 3{,}6 = 58{,}878$ in termen van de continue variabele. Bij de toetsingsgrootheid $\underline{z}$ hoort dus het kritieke gebied $\{0,1,\ldots,58\}$.

N.B. Men heeft ook hier te maken met de continuïteitscorrectie. Als het kritieke gebied voor een binomiaal verdeelde stochastische variabele $\underline{x}$ $\{0,1,\ldots,c\}$ is dan moet in termen van de continue stochastische variabele $\underline{y}$, die als benadering wordt gebruikt, c worden bepaald als het grootste gehele getal waarvoor geldt:

$P(\underline{y} \leq c + \frac{1}{2}) \leq \alpha$. ❑

8

Verdelingsvrije toetsen

8.1. Inleiding

Naast het parametrisch toetsen, waarbij men aanneemt dat de populatie-verdeling, op een of meer onbekende parameters na, bekend is, zijn er ook toetsen ontwikkeld voor die gevallen waarin men zelfs omtrent het type verdelingsfunctie in het duister tast.
Slechts enkele verdelingsvrije toetsen zullen hier worden besproken. Allereerst een toets voor de hypothese dat een steekproef afkomstig is uit een populatie met verdelingsfunctie F. Vervolgens twee toetsen om na te gaan of twee steekproeven geacht kunnen worden afkomstig te zijn uit populaties die, in zeker opzicht, gelijk zijn.

8.2. De χ^2-toets voor aanpassing

Als men het idee heeft, of hoopt, dat de stochastische variabele $\underline{x}$ een bepaalde verdelingsfunctie F_x heeft en men beschikt over een voldoend grote steekproef aan $\underline{x}$ dan kan men op de gegevens een χ^2-toets toepassen om de redelijkheid van het idee te onderzoeken. Deze toets wordt meestal de χ^2-goodness of fit-toets genoemd, in het Nederlands: de χ^2-toets voor aanpassing.
In de eenvoudigste vorm gaat het als volgt:
Verdeel het waardenbereik (a,b) van $\underline{x}$ in twee stukken (a,m] en (m,b). De verdelingsfunctie waarvan we hopen dat het de goede is geeft $P(\underline{x} \leq m) = p_1$, $P(\underline{x} > m) = p_2 = 1 - p_1$. Dit is de nulhypothese. De alternatieve hypothese is $P(\underline{x} \leq m) \neq p_1$. Onder de nulhypothese is iedere waarneming x_i een realisatie van een alternatief verdeelde stochastische variabele met kans p_1 op succes.

Bij een grote aselecte steekproef is $\sum_{i=1}^{n} \underline{x}_i$, het aantal successen, na standaardisatie, bij benadering N(0;1)-verdeeld (centrale limietstelling), d.w.z.

$$\frac{\sum_{i=1}^{n} \underline{x}_i - np_1}{\sqrt{np_1(1-p_1)}} \sim N(0;1).$$

Dan is $\underline{z} = \dfrac{\left\{\sum\limits_{i=1}^{n} \underline{x}_i - np_1\right\}^2}{np_1(1-p_1)}$ ongeveer $\chi^2(1)$-verdeeld.

$$\underline{z} = \frac{1}{np_1}\left\{\sum_1^n \underline{x}_i - np_1\right\}^2 + \frac{1}{n(1-p_1)}\left\{\sum_1^n \underline{x}_i - np_1\right\}^2,$$

en met $\underline{y}_i = 1 - \underline{x}_i$ dus $\sum\limits_i^n \underline{x}_i = n - \sum\limits_i^n \underline{y}_i$ kunnen we daarvoor schrijven:

$$\underline{z} = \frac{\left\{\sum\limits_1^n \underline{x}_i - np_1\right\}^2}{np_1} + \frac{\left\{\sum\limits_1^n \underline{y}_i - np_2\right\}^2}{np_2}.$$

Hierin is $\sum\limits_1^n \underline{x}_i$ het aantal waarnemingen $\leq m$ bij de n in de steekproef en np_1 is de verwachting daarvan. Het aantal waarnemingen in de steekproef dat groter is dan m is $\sum\limits_1^n \underline{y}_i$ en de verwachting van dat aantal is np_2.

Onder de nulhypothese zal $\underline{z}$ meestal niet al te groot zijn, het kritieke gebied bestaat dus uit het interval (c,∞) waarbij c volgt uit $P\{\underline{z} > c\} = \alpha$ met $\underline{z} \sim \chi^2(1)$ en α de gekozen onbetrouwbaarheidsdrempel van de toets.

Soms wordt echter tweezijdig getoetst. Heel kleine waarden voor $\underline{z}$ kunnen verdacht zijn als de mogelijkheid bestaat dat er met de waarnemingen is geknoeid.

Meestal verdeelt men het waardenbereik van de stochastische variabele in een groter aantal stukken, bijvoorbeeld k: $(a = m_0, m_1]$, $(m_1, m_2], \ldots, (m_{k-1}, m_k = b)$.

Elke waarneming is nu de realisatie van een multinomiaal verdeelde stochastische variabele die met kans p_i een waarde aanneemt in het i-de deelinterval. De kansen p_i volgen uit de verdelingsfunctie als $F_x(m_i) - F_x(m_{i-1})$. Onder H_0: de verdelingsfunctie F_x is de juiste, zijn deze bekend.

De variabele

$$\underline{z} = \sum_{j=1}^{k} \frac{(\underline{O}_j - np_j)^2}{np_j},$$

waarin $\underline{O}_j$ het aantal waarnemingen is dat in de j-de klasse = het j-de deelinterval terecht komt, is nu bij benadering $\chi^2(k-1)$ verdeeld. Verder verloopt de toets als in het voorgaande.

Er zit nogal wat benadering in. Pas voor $n \to \infty$ gaat alles echt goed. Meestal wordt aangegeven dat men, om te grote afwijkingen te voorkomen, ervoor moet zorgen dat np_i, voor alle i, groter is dan vijf.

Als H_0 alleen het type verdelingsfunctie specificeert maar één of meer parameters nog onbekend zijn, is het mogelijk om deze parameters m.b.v. de waarnemingen te schatten. Dat maakt $\underline{z}$ natuurlijk kleiner. Het gaat dan ook ten koste van één vrijheidsgraad per parameter die uit de waarnemingen wordt geschat.

Voorbeeld

Een NS-conducteur met belangstelling voor statistiek heeft het idee dat het aantal plaatsbewijzen dat hij per werkdag moet uitschrijven Poisson-verdeeld is. Om zijn vermoeden te toetsen noteert hij de desbetreffende aantallen gedurende 50 werkdagen en onderwerpt de resultaten $\underline{x}_1, \underline{x}_2, \dots, \underline{x}_{50}$ aan een χ^2-toets met $\alpha = 0{,}05$. H_0: $\underline{x}_i \sim$ Poisson; H_1: de verdeling van $\underline{x}_i$ is anders. Neem aan dat de resultaten als volgt kunnen worden samengevat.

x_i:	5	6	7	8	9	10	11	12	13	14	15
f_i:	1	1	5	8	9	5	7	6	4	2	2

Als eerste moet nu de parameter λ worden geschat. $\lambda = E(\underline{x})$ dus we schatten λ met $\bar{x} = \frac{1}{50}\sum x_i f_i = 10$. Het waardenbereik van een Poisson-verdeelde variabele is $\{0,1,2,\dots\}$. Dit moet worden verdeeld in klassen waarvoor geldt dat per klasse het verwachte aantal waarnemingen minstens vijf is. Uit de tabel van de cumulatieve Poisson-verdeling volgt voor $\lambda = 10$:

$$P(\underline{x} \le 5) = 0{,}0671;$$
$$P(\underline{x} = 6) = P(\underline{x} \le 6) - P(\underline{x} \le 5) = 0{,}1301 - 0{,}0671 = 0{,}0630,$$

enz. Daaruit volgt dat bij n = 50 het verwachte aantal waarnemingen ≤ 5 gelijk is aan 50·0,0671 = 3,35 enz. Deze gegevens kunnen het beste meteen in een voor de verdere berekening handige tabel worden geplaatst.

klasse j	p_j	e_j	e_i	O_i	$\lvert o_i - e_i \rvert$	$(o_i - e_i)^2/e_i$
≤ 5	0,0671	3,35	6,50	2	4,50	3,115
6	0,0630	3,15				
7	0,0901	4,50	10,13	13	2,87	0,813
8	0,1126	5,63				
9	0,1251		6,26	9	2,74	1,199
10	0,1251		6,26	5	1,26	0,254
11	0,1138		5,69	7	1,31	0,302
12	0,0948	4,74	8,38	10	1,62	0,313
13	0,0729	3,64				
14	0,0520	2,60	6,78	4	2,78	1,140
≥ 15	0,0835	4,18				
	———— +		———— +	—— +		———— +
	1,0000		50,00	50		7,136

Het aantal klassen, na het samenvoegen, is zeven. Het aantal vrijheidsgraden van de toetsingsgrootheid zou dus 7 − 1 = 6 zijn geweest als we niet de parameter λ geschat hadden uit dezelfde waarnemingen. Nu moet er nog een vrijheidsgraad worden afgetrokken.

Onder H_0 is dus $\underline{z} = \sum_{i=1}^{7} \frac{(\underline{o}_i - e_i)^2}{e_i}$ $\chi^2(5)$ verdeeld.

Het kritieke gebied, bij $\alpha = 0{,}05$, is $(11{,}1; \infty)$. De realisatie van $\underline{z}$ ligt niet in het kritieke gebied zodat de nulhypothese niet verworpen kan worden op grond van deze steekproef. ❑

De hier beschreven toets is bruikbaar voor elke verdelingsfunctie. Er zijn echter ook diverse toetsen ontwikkeld die alleen kunnen worden gebruikt in combinatie met een speciaal type populatie-verdeling, bijvoorbeeld de Normale verdeling of de exponentiële verdeling.

8.3. De tekentoets

De tekentoets wordt gebruikt om de hypothese dat de mediaan M van een verdeling de waarde m_0 heeft tegen het alternatief $M > m_0$ (of $M < m_0$, of $M \neq m_0$) te toetsen. Het is dus gewoon een toets voor de parameter p van een alternatief verdeelde populatie in het speciale geval p = 0,5. Immers als M de mediaan is van de verdeling F_x van $\underline{x}$ dan geldt $P(\underline{x} < M) = P(\underline{x} > M) = 0{,}5$ waarbij we $\underline{x}$ gemakshalve continu veronderstellen zodat $P(\underline{x} = M) = 0$. Voor de uitvoering van de toets kan de steekproef $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ worden vervangen door $\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_n$ waarbij $\underline{y}_i = 1$ als $\underline{x}_i > m_0$ en $\underline{y}_i = 0$ als $\underline{x}_i < m_0$. Onder de nulhypothese, $M = m_0$, is $\sum \underline{y}_i \sim B(n; 0{,}5)$. Het is echter gebruikelijk, en daaraan ontleent de toets zijn naam, om i.p.v. de symbolen '1' en '0' respectievelijk '+' en '−' te noteren, zodat als toetsingsgrootheid $\underline{z}$ het aantal plustekens (of het aantal mintekens) onder de n symbolen wordt gebruikt. M.a.w. als $x_i - m_0 > 0$ schrijft men een plus en als $x_i - m_0 < 0$ een min. Het kan natuurlijk voorkomen dat $x_i = m_0$. Dergelijke waarnemingen worden uit de steekproef verwijderd. De steekproefomvang n wordt daardoor kleiner. Het kritieke gebied kan op de gebruikelijke manier worden bepaald. Er bestaan ook tabellen voor de meest voorkomende gevallen.

De tekentoets wordt vaak toegepast bij gepaarde waarnemingen $(\underline{x}_i, \underline{y}_i)$, $i = 1,2,\ldots,n$. Men vormt dan een steekproef van verschillen $\underline{v}_i = \underline{x}_i - \underline{y}_i$, $i = 1,2,\ldots,n$ en toetst de nulhypothese dat de mediaan van $\underline{v}$ gelijk is aan nul. Die nulhypothese laat zich vertalen als: er is geen systematisch verschil in ligging tussen de verdeling van $\underline{x}$ en de verdeling van $\underline{y}$.

Voorbeeld

Bij een onderzoek naar de krimp veroorzaakt door de wasmiddelen Bam en Dof worden 16 lakens in tweeën geknipt. Van elk laken wordt een helft met Bam en de andere helft met Dof gewassen, onder overigens gelijke omstandigheden.
Het onderzoek wordt uitgevoerd door de fabrikant van Dof in de hoop bevestiging te krijgen van het idee dat Bam meer krimp veroorzaakt dan Dof. Na het wassen bleek in 12 gevallen de Bam-helft meer te zijn gekrompen dan de Dof-helft, in drie gevallen was het omgekeerde het geval terwijl van één laken beide helften evenveel waren gekrompen. We toetsen H_0: 'gelijke krimp' tegen 'Bam veroorzaakt een grotere krimp dan Dof' met onbetrouwbaarheidsdrempel $\alpha = 0{,}05$.
De steekproefomvang wordt teruggebracht tot $n = 15$ omdat het laken waarvan de beide helften even sterk gekrompen waren buiten beschouwing blijft. Onder H_0 is dan $\underline{z}$ (het aantal malen dat Bam meer krimp veroorzaakt dan Dof) binomiaal verdeeld met parameters $n = 15$ en $p = 0{,}5$. Het kritieke gebied wordt $\{12,13,14,15\}$ want voor $\underline{z} \sim B(15; 0{,}5)$ geldt

$$P(\underline{z} \geq 11) = 1 - P(\underline{z} \leq 10) = 0{,}0592 \quad \text{(te groot)}$$
$$P(\underline{z} \geq 12) = 1 - P(\underline{z} \leq 11) = 0{,}0176.$$

De realisatie van $\underline{z}$ $(z = 12)$ ligt in het kritieke gebied dus de nulhypothese wordt verworpen en de fabrikant van Dof kan concluderen dat Bam meer krimp veroorzaakt dan Dof.

8.4. De toets van Wilcoxon

Met de toets van Wilcoxon kan men bij twee onafhankelijke steekproeven de hypothese toetsen dat de beide populatie-verdelingen dezelfde ligging hebben tegen de alternatieve hypothese dat de verdelingen t.o.v. elkaar verschoven zijn. Eigenlijk mag men de toets alleen toepassen als de beide populatie-verdelingen continu zijn, maar daarmee neemt men het meestal niet al te nauw. De toets verloopt als volgt.
Gegeven zijn twee onafhankelijke steekproeven:

$$\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n \quad \text{uit een populatie met } F_x$$
$$\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_m \quad \text{uit een populatie met } F_y.$$

Als toetsingsgrootheid kiest men $\underline{w} = \sum_{i=1}^{n} \sum_{j=1}^{m} \underline{w}_{ij}$, waarin de 'scores' $\underline{w}_{ij}$ voor $i = 1,2,\ldots,n$; $j = 1,2,\ldots,m$ zijn gedefinieerd door

$$\underline{w}_{ij} = 2 \text{ als } \underline{x}_i > \underline{y}_j;$$
$$= 1 \text{ als } \underline{x}_i = \underline{y}_j;$$
$$= 0 \text{ als } \underline{x}_i < \underline{y}_j.$$

Als F_x en F_y continu zijn is $P(\underline{w}_{ij} = 1) = 0$, zodat dan onder H_0 geldt $P(\underline{w}_{ij} = 2) = P(\underline{w}_{ij} = 0) = 0{,}5$.
De realisatie van de toetsingsgrootheid $\underline{w}$ wordt bepaald door elke waarneming in de ene steekproef te vergelijken met elke waarneming uit de andere steekproef en de zo verkregen scores te sommeren.

Voorbeeld

x_i: 5, 9 (n = 2)

y_j: 3, 4, 5, 5, 7, 9, 10, 10 (m = 8)

(Terwille van de overzichtelijkheid zijn de waarnemingen naar grootte geordend.)
$x_1(5)$ is groter dan y_1 en y_2(3, resp. 4) en gelijk aan y_3 en y_4 maar kleiner dan y_5 t/m y_8 zodat

$$\sum_{j=1}^{8} w_{1j} = 2 + 2 + 1 + 1 = 6$$

de bijdrage is van x_1 in de totale score. Zo volgt als bijdrage van x_2:

$$\sum_{j=1}^{8} w_{2j} = 2 + 2 + 2 + 2 + 2 + 1 = 11.$$

De totale score w = 17. ❑

De maximale waarde van w wordt bereikt als alle waarnemingen in de eerste steekproef groter zijn dan de grootste waarneming in de tweede steekproef. w_{max} is dus 2nm. Analoog volgt dat $w_{min} = 0$. Onder H_0 is de verdeling van $\underline{w}$ symmetrisch en dus de verwachting van $\underline{w}$ gelijk aan nm. Bij tweezijdig toetsen zal het kritieke gebied de volgende punten uit het waardenbereik van $\underline{w}$ bevatten:
0,1,...,c en 2nm–c, 2nm–c+1,...,2nm
waarbij c bepaald wordt m.b.v.

$$\left.\begin{array}{l} P(\underline{w} \leq c) \leq \frac{1}{2}\alpha \\ P(\underline{w} \leq c+1) > \frac{1}{2}\alpha \end{array}\right\} \text{ als } H_0 \text{ waar is.}$$

Bij een eenzijdige toets ligt het kritieke gebied geheel in de buurt van 0 of van 2nm en wordt op overeenkomstige wijze bepaald.
De grenswaarden c van de linker-kritieke gebieden zijn getabelleerd voor n en m niet groter dan 20. De grenswaarde van een overeenkomstig rechter-kritiek gebied volgt als 2nm – c op grond van symmetrie.
Daar $\underline{w}$ onder H_0 een symmetrische verdeling heeft is het niet verbazingwekkend dat voor grotere waarden van m en n de verdeling van $\underline{w}$ wordt benaderd met een Normale verdeling met verwachting $E(\underline{w})$ = nm en variantie

$V(\underline{w}) = \frac{1}{3}nm(n+m+1)$. Het is niet moeilijk maar wel omslachtig om de exacte verdeling van $\underline{w}$, onder H_0, te bepalen. Dat gaat als volgt, waarbij we F_x en F_y continu veronderstellen zodat er geen gelijke waarnemingen voorkomen.
Voor de waarde die $\underline{w}$ aanneemt is het niet van belang in welke volgorde de waarnemingen in de steekproef $x_1,x_2,\ldots,x_n$ (en in $y_1,y_2,\ldots,y_m$) voorkomen.
Voorts kunnen we ook de werkelijke waarde van x_i (of y_j) vervangen door het rangnummer dat x_i (of y_j) zou krijgen als alle n+m waarnemingen tijdelijk bij elkaar worden gedaan en, van klein naar groot, worden geordend.

Voorbeeld

Voor de steekproeven

x_i:	5,23	3,17	
y_j:	4,78	1,14	9,56

volgt $w = (2+2)+2 = 6$.
Vervang x_i en y_j door hun rangnummers in de gecombineerde steekproef:

x_i':	4	2	
y_j':	3	1	5

Rangschik dit op een overzichtelijke manier:

x_i':		2		4	
y_j':	1		3		5

Ook nu volgt $w = 2+(2+2) = 6$. ❑

Het is zelfs voldoende als bekend is welke van de n+m rangnummers horen bij waarnemingen uit de steekproef $x_1,x_2,\ldots,x_n$. Er zijn $\binom{n+m}{n}$ manieren om n rangnummers uit de n+m te kiezen die bij de x-steekproef horen en onder H_0 heeft elk van die manieren een even grote kans. Er is maar één volgorde die $w = 0$ oplevert dus $P(\underline{w} = 0) = 1/\binom{n+m}{n}$. Ook is er maar één volgorde met $w = 2$, namelijk:

x:	1	2	...	n–2		n			
y:					n–1		n+1	...	n+m

zodat ook $P(\underline{w} = 2) = 1/\binom{n+m}{n}$. Op grond van de symmetrie volgt

$$P(\underline{w} = 2nm) = P(\underline{w} = 0) \text{ en } P(\underline{w} = 2mn-2) = P(\underline{w} = 2).$$

Voorbeeld

Voor $n = 2$ en $m = 3$ zijn de volgende opstellingen mogelijk. Alleen de rangnummers behorend bij de twee waarnemingen uit de x-steekproef zijn vermeld.

1,2	$w = 0$
1,3	$w = 2$

1,4; 2,3	w = 4
1,5; 2,4	w = 6
2,5; 3,4	w = 8
3,5	w = 10
4,5	w = 12

Dus is onder H_0 $P(\underline{w} = 0) = P(\underline{w} = 2) = P(\underline{w} = 10) = P(\underline{w} = 12) = 0{,}1$ en $P(\underline{w} = 4) = P(\underline{w} = 6) = P(\underline{w} = 8) = 0{,}2$. ❑

Men maakt natuurlijk geen gebruik van de toets van Wilcoxon als bekend is dat de populatie-verdelingen Normaal zijn. De in hoofdstuk 7 genoemde toetsen zijn dan veel beter.

Voorbeeld

De NS wenst na te gaan of remblokken van materiaal B gemiddeld een grotere wielbandslijtage veroorzaken dan remblokken van materiaal A. Men kiest 20 wielen en plaatst bij tien daarvan remblokken van materiaal A, bij de andere helft remblokken van materiaal B. Na 10 000 km meet men de slijtage van de wielbanden. Dit gaf de volgende resultaten (in een of andere lengte-eenheid):

A:	14	14	16	14	15	16	15	14	13	15
B:	15	14	17	16	14	17	16	15	16	15,5

of na ordening:

A:	13	4 × 14	3 × 15		2 × 16	
B:		2 × 14	2 × 15	15,5	3 × 16	2 × 17

Er wordt getoetst H_0: 'Het materiaal heeft geen invloed' tegen H_1: 'Materiaal B geeft een grotere wielbandslijtage dan materiaal A' met onbetrouwbaarheidsdrempel $\alpha = 0{,}05$.

Als $w_{ij} = 2$ voor $A_i > B_j$ dan hebben kleine waarden van de toetsingsgrootheid $\underline{w}$ een grotere kans als H_1 waar is dan onder H_0. Het kritieke gebied is dan van de vorm $\{0,1,\ldots,c\}$. De tabel voor $n = m = 10$ en $\alpha = 0{,}05$ geeft $c = 54$. H_0 wordt dus verworpen als $\underline{w} \leq 54$.

Uit de waarnemingen volgt

$$w = 0 + 4\cdot2\cdot1 + (3\cdot2\cdot2 + 3\cdot2\cdot1) + (2\cdot5\cdot2 + 2\cdot3\cdot1) = 8 + 18 + 26 = 52.$$

De waarneming ligt in het kritieke gebied zodat H_0 wordt verworpen. Materiaal B geeft meer slijtage dan materiaal A.

Zeker als $n = m$ is het zaak om goed na te gaan of men niet met gepaarde waarnemingen te maken heeft zodat de tekentoets in aanmerking komt. De toets van Wilcoxon mag alleen worden gebruikt als de beide steekproeven onafhankelijk van elkaar zijn.

9

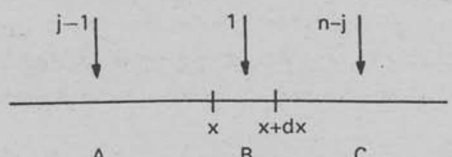
Geordende steekproeven

Als men uit een continue populatie F_x een steekproef trekt met omvang n teneinde bijvoorbeeld een schatting te maken van de verwachting kan zich een tot nu toe niet besproken moeilijkheid voordoen. Laat $\underline{x}$ de levensduur zijn van een onderdeel. Kies uit een grote partij van die onderdelen n exemplaren, de steekproef. Daarna breekt een tijd van wachten aan tot alle exemplaren het hebben begeven. Immers x_i is de realisatie van de levensduur $\underline{x}_i$ en die is pas bekend op het moment waarop het onderdeel i kapot gaat $(i = 1,2,\ldots,n)$. De nummering van de onderdelen is hierbij geheel willekeurig.
Een natuurlijker nummering kent het rangnummer 1 toe aan het onderdeel dat als eerste kapot gaat, 2 aan het onderdeel dat daarna als eerste kapot gaat; en zo verder tot aan rangnummer n voor het onderdeel dat als laatste kapot gaat. Men krijgt zo een stijgende rij realisaties.
Zij $\underline{x}_{(i)}$ de stochastische variabele die de levensduur aangeeft van het onderdeel dat als i-de kapot gaat $(i = 1,2,\ldots,n)$. De notatie $\underline{x}_{(1)},\underline{x}_{(2)},\ldots,\underline{x}_{(n)}$ is nodig ter onderscheiding van de oorspronkelijke willekeurige nummering $\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_n$. De variabelen $\underline{x}_i$ en $\underline{x}_j$ zijn onafhankelijk en hebben dezelfde verdeling. De variabelen $\underline{x}_{(i)}$ en $\underline{x}_{(j)}$ zijn niet onafhankelijk en hebben ook niet dezelfde verdeling, immers voor $j > i$ is het zeker dat $\underline{x}_{(j)} > \underline{x}_{(i)}$. De variabelen $\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_n$ horen bij een (gewone) aselecte steekproef. De variabelen $\underline{x}_{(1)},\underline{x}_{(2)},\ldots,\underline{x}_{(n)}$ associëren we met een (naar opklimmende grootte) geordende steekproef.
Het is duidelijk dat als $\underline{x}$ een continue stochastische variabele is dat dan ook $\underline{x}_{(j)}$ continue stochastische variabelen zijn. Voor f_j, de dichtheid van $\underline{x}_{(j)}$, kan op de volgende manier een uitdrukking in termen van f_x, de dichtheid van $\underline{x}$, en F_x worden gevonden.

$$f_j(x)dx \approx P\{x < \underline{x}_{(j)} \leq x+dx\} =$$

$$= P\{\underline{x}_{(1)},\underline{x}_{(2)},\ldots,\underline{x}_{(j-1)} \leq x \cap x < \underline{x}_{(j)} \leq x+dx \cap \underline{x}_{(j+1)},\underline{x}_{(j+2)},\ldots,\underline{x}_{(n)} > x+dx\} =$$

$$= P\{j-1 \text{ van de } n \leq x \cap 1 \text{ van de } n \in (x,x+dx] \cap n-j \text{ van de } n > x+dx\}.$$



Figuur 9.1.

Deze laatste kans kan worden bepaald met behulp van de multinomiale verdeling. Er worden n onafhankelijke proeven gedaan. Voor elke proef geldt:

$$p_1 = P\{\text{resultaat A}\} = F_x(x) = \int_{-\infty}^{x} f_x(t)\,dt;$$

$$p_2 = P\{\text{resultaat B}\} \approx f_x(x)\,dx;$$

$$p_3 = P\{\text{resultaat C}\} = 1 - F_x(x+dx) = \int_{x+dx}^{\infty} f_x(t)\,dt.$$

Daarmee volgt:

$$\begin{aligned} f_j(x)\,dx &\approx P\{j-1 \text{ maal A} \cap 1 \text{ maal B} \cap n-j \text{ maal C}\} = \\ &= \frac{n!}{(j-1)!\,1!\,(n-j)!}\, p_1^{j-1} p_2^{1} p_3^{n-j} = \\ &= j\binom{n}{j}\cdot\{F_x(x)\}^{j-1}\cdot f_x(x)\,dx\cdot\{1-F_x(x+dx)\}^{n-j} \end{aligned}$$

zodat na deling door dx en het vervolgens tot nul laten naderen van dx als resultaat overblijft

$$f_j(x) = j\binom{n}{j} f_x(x)\cdot\{F_x(x)\}^{j-1}\cdot\{1-F_x(x)\}^{n-j} \quad \text{voor} \quad j = 1,2,\ldots,n.$$

Voor $j = 1$ krijgt men de dichtheid f_1 van het minimum van $\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_n$ en voor $j = n$ de dichtheid f_n van het maximum.
De bovengenoemde moeilijkheid bestaat nu daaruit dat er ongelukkigerwijze wel eens één of meer bijzonder goede onderdelen in de steekproef kunnen zitten. Goed betekent in dit verband: met een lange levensduur. In dat geval duurt het heel lang voor de realisaties van $\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_n$ allemaal bekend zijn. Er zijn twee manieren om hier iets aan te doen.

a. Men kiest een vaste tijd T en noteert de levensduur van de onderdelen die vóór T kapot gaan. Men verkrijgt zo de realisaties $x_{(1)},x_{(2)},\ldots,x_{(i)}$. Hierbij is $\underline{i}$ een stochastische variabele omdat vooraf niet bekend is hoeveel onderdelen voor T kapot zullen gaan. Het waardenbereik van $\underline{i}$ is $0,1,2,\ldots,n$. Van de overige onderdelen is alleen bekend dat de levensduur groter is dan T.
b. Men kiest een natuurlijk getal $j < n$ en noteert de realisaties $x_{(1)},x_{(2)},\ldots,x_{(j)}$. Van de overige $n-j$ onderdelen is alleen bekend dat de levensduur groter is dan $x_{(j)}$. Nu is de tijd nodig voor het verzamelen van de informatie een stochastische variabele $\underline{x}_{(j)}$. Het kan nog heel lang duren maar wel is zeker dat $x_{(j)}$ kleiner is dan $x_{(n)}$, het tijdstip waarop het laatste onderdeel uit de steekproef kapot gaat.

Het is, in beide gevallen, niet eerlijk om verder te werken met de verkregen realisaties en net te doen alsof de overige onderdelen (met de lange levensduren) niet in de steekproef zaten. Voor een correcte verwerking van de verkregen resultaten is het nodig om gebruik te maken van de verdeling van $\underline{x}_{(i)}$.
Zij $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ een aselecte steekproef uit een standaard-uniforme populatie, met andere woorden:

$$\begin{aligned} F_x(x) &= 0, \quad && x < 0; \\ &= x, && 0 \leq x \leq 1; \\ &= 1, && x > 1; \end{aligned}$$

dan vinden we voor de dichtheid van $\underline{x}_{(i)}$, voor $i = 1,2,\ldots,n$

$$\begin{aligned} f_i(x) &= i\binom{n}{i} x^{i-1}(1-x)^{n-i}, \qquad && 0 < x < 1; \\ &= 0, && \text{elders.} \end{aligned}$$

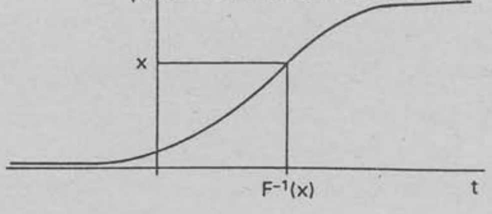
Daaruit volgt dat de verwachting van $\underline{x}_{(i)}$ gelijk is aan

$$\begin{aligned} E(\underline{x}_{(i)}) &= i\binom{n}{i}\int_0^1 x^i(1-x)^{n-i}dx = i\binom{n}{i} B(i+1;n-i+1) = \\ &= i\cdot\frac{n!}{i!(n-i)!}\cdot\frac{i!(n-i)!}{(n+1)!} = \frac{i}{n+1}\,. \end{aligned}$$

Ook de modus van $\underline{x}_{(i)}$ is eenvoudig te bepalen. Het blijkt dat $f_i(x)$ een maximale waarde aanneemt voor $x = \dfrac{i-1}{n-1}$.

Het is niet toevallig dat hierboven wat extra aandacht is gegeven aan de standaard-uniforme verdeling. Zij namelijk $\underline{x}$ een continue stochastische variabele met verdelingsfunctie $F(t)$ dan is $F(\underline{x})$ standaard-uniform verdeeld.
Immers voor $0 \leq x \leq 1$ (zie figuur 9.2):

$$P\{F(\underline{x}) \leq x\} = P\{\underline{x} \leq F^{-1}(x)\} = F\{F^{-1}(x)\} = x.$$



Figuur 9.2.

Van deze relatie wordt ook gebruik gemaakt bij simulatie. Het is voldoende om een generator te hebben voor getallen die uniform verdeeld zijn over het interval [0,1]. Van de zo verkregen uitkomsten $u_1, u_2, \ldots, u_n$ maakt men met behulp van de inverse F^{-1} van de verdelingsfunctie F een rij gesimuleerde realisaties $x_1 = F^{-1}(u_1)$, $x_2 = F^{-1}(u_2), \ldots, x_n = F^{-1}(u_n)$ van de stochastische variabele $\underline{x}$ met verdelingsfunctie F.

Voorbeeld

Simulatie van realisaties van een exponentieel verdeelde stochastische variabele met parameter λ.

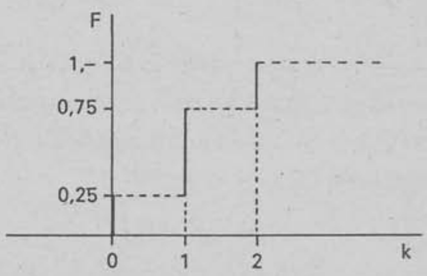
$$F(x) = 1 - e^{-\lambda x} \quad \text{zodat} \quad x = \frac{-1}{\lambda} \ln\{1 - F(x)\} \qquad x > 0.$$

Het resultaat u_i van de generator leidt tot $x_i = \dfrac{-1}{\lambda} \ln(1 - u_i)$, waarbij u_i uniform op [0,1] verdeeld is. De berekening kan in dit geval nog iets worden vereenvoudigd omdat ook $1-u_i$ uniform op [0,1] verdeeld is zodat we kunnen werken met $x_i = \dfrac{-1}{\lambda} \ln u_i$. ❑

Een analoge werkwijze is mogelijk bij een discrete stochastische variabele.

Voorbeeld

Simulatie van realisaties van een $B(2;\frac{1}{2})$-verdeelde stochastische variabele $\underline{x}$ (zie figuur 9.3).



Figuur 9.3.

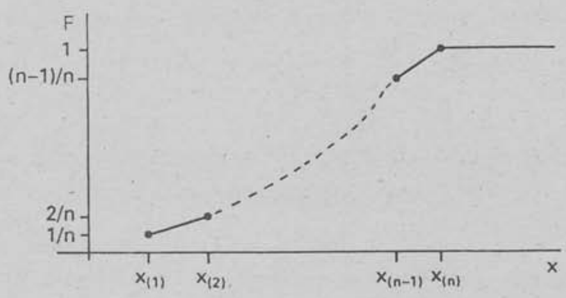
$P\{\underline{x} = 0\} = \frac{1}{4}$; $P\{\underline{x} = 1\} = \frac{1}{2}$; $P(\underline{x} = 2\} = \frac{1}{4}$.

Verdeel het interval [0,1] op de verticale as in de stukken [0;0,25], (0,25;0,75] en (0,75;1]. Genereer een getal u_i, uniform op [0,1]. De gegenereerde realisatie van $\underline{x}$ wordt dan:

$$
\begin{array}{lll}
x_i = 0 & \text{als} & 0 \le u_i \le 0{,}25; \\
x_i = 1 & \text{als} & 0{,}25 < u_i \le 0{,}75; \\
x_i = 2 & \text{als} & 0{,}75 < u_i \le 1.
\end{array}
$$
❑

Het gebruik van een generator kan worden vervangen door het werken met een tabel van toevalsgetallen.

Als men een indruk wenst te krijgen van de onbekende verdelingsfunctie van een stochastische variabele kan men als volgt te werk gaan. Trek een steekproef ter grootte n, met als uitkomsten $x_1, x_2, \ldots, x_n$. Orden de resultaten tot $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$. In de steekproef zien we dat i van de n uitkomsten een waarde hebben kleiner of gelijk aan $x_{(i)}$. Voor deze steekproef geldt dus dat de fractie uitkomsten, kleiner of gelijk aan $x_{(i)}$ gelijk is aan i/n. Op deze manier ontstaat een (slechte) schatting voor een aantal punten van de verdelingsfunctie F, zie figuur 9.4. Door deze punten te verbinden krijgt men inderdaad iets dat op een verdelingsfunctie lijkt.



Figuur 9.4.

Echter, de suggestie dat de stochastische variabele $\underline{x}$ nooit groter dan $x_{(n)}$ kan zijn, is natuurlijk misleidend. Een volgende steekproef aan dezelfde variabele kan best een grotere waarde bevatten.
De in figuur 9.4 geschetste grafiek wordt wel de empirische verdelingsfunctie genoemd. Een betere schatting van de verdelingsfunctie krijgt men door te werk te gaan op een manier analoog aan de bepaling van momentenschatters of meest-aannemelijke schatters zoals beschreven in hoofdstuk 6.

A. Momenten-methode

Zij F(x) de onbekende verdelingsfunctie van $\underline{x}$ dan is $F(\underline{x})$ standaard-uniform verdeeld.
De n geordende waarnemingen $\underline{x}_{(1)}, \underline{x}_{(2)}, \ldots, \underline{x}_{(n)}$ corresponderen met n geordende waarnemingen $F(\underline{x}_{(1)}), F(\underline{x}_{(2)}), \ldots, F(\underline{x}_{(n)})$ aan de standaard-uniform verdeelde variabele $F(\underline{x})$. Daaruit volgt dat $E\{F(\underline{x}_{(i)})\} = \frac{i}{n+1}$. Als (momenten-)schatting gebruiken we dus:

$$F(x_{(i)}) = \frac{i}{n+1}\,.$$

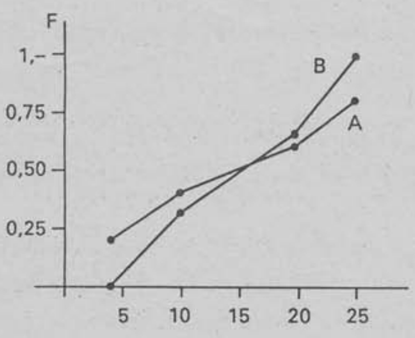
Daarmee kan het grootste deel van de grafiek van de geschatte verdelingsfunctie worden getekend. Slechts begin en eind (voor $0 \leq F(x) < \frac{1}{n+1}$, respectievelijk $\frac{n}{n+1} < F(x) \leq 1$) moeten nog worden toegevoegd.

B. Methode der grootste aannemelijkheid

Als in het voorgaande gaan we over op $F(\underline{x}_{(1)}), F(\underline{x}_{(2)}), \ldots, F(\underline{x}_{(n)})$. We kiezen nu F zó dat het aangetroffen resultaat in de steekproef de grootst mogelijke kans van optreden heeft (in dit continue geval: een zo groot mogelijke waarde van de simultane dichtheid). Die wordt bereikt door in elke één-dimensionale marginale dichtheid de modus te kiezen dus aan $x_{(i)}$ als functiewaarde van F de modus van $\underline{x}_{(i)}$, ofwel $\frac{i-1}{n-1}$ toe te voegen. In dit geval krijgt men een grafiek voor F die bij nul begint en tot één doorgroeit.

Voorbeeld

De geordende waarnemingen 4, 10, 20, 25 zijn gegeven. Men construeert hiermee de volgende schattingen voor de verdelingsfunctie (zie figuur 9.5):



A:	$(4;\frac{1}{5})$	B:	$(4;0)$
	$(10;\frac{2}{5})$		$(10;\frac{1}{3})$
	$(20;\frac{3}{5})$		$(20;\frac{2}{3})$
	$(25;\frac{4}{5})$		$(25;1)$

Figuur 9.5.

Het is duidelijk dat noch methode A noch methode B geheel bevredigend is. Er zijn dan ook nog diverse andere manieren om aan een waarneming $x_{(i)}$ een waarde van F toe te kennen. Men spreekt in deze gevallen van het bepalen van *plotting-positions.*

Slechts één andere regel zij nog kort vermeld. Daarbij kiest men voor de mediane waarde van $F(\underline{x}_{(i)})$, dat wil zeggen men kiest die waarde die met kans $\frac{1}{2}$ te klein en met kans $\frac{1}{2}$ te groot is. Dit wordt benaderd door aan $x_{(i)}$ als F-waarde $\frac{i-0,3}{n+0,4}$ toe te voegen.

Opgaven

1. Van een grote partij draadeinden (M24 Staal 32) wordt een duizendtal beproefd met 14, 16 en 18 ton. Het blijkt dat van deze duizend er 30 reeds breken bij een last van 14 ton, 670 bij een last van 16 ton en een verdere 210 bij 18 ton. Construeer met behulp van deze gegevens een kansruimte voor de experimentele bepaling van de sterkteklasse van een draadeind uit dezelfde partij. Onderscheid 4 uitkomsten.

2. Bepaal de onderling gelijkwaardige uitkomsten bij een worp met twee zuivere dobbelstenen. Welke van deze uitkomsten leiden tot een totaal aantal van zes geworpen ogen?

3. Een oude hifi-installatie bestaat uit een tuner T, een platenspeler P, een versterker V en twee luidsprekerboxen R en L. Ieder van deze onderdelen kan wel of niet functioneren. De gebeurtenis dat muziek ten gehore kan worden gebracht duidt men aan met $(T \cup P) \cap V \cap (R \cup L)$, een van de toestanden van de installatie waarbij deze gebeurtenis optreedt is $T\bar{P}VR\bar{L}$.
Geef alle toestanden die tot de bovenbeschreven gebeurtenis leiden. Hoeveel toestanden kunnen er in totaal bij het experiment worden onderscheiden?

4. Een apparaat bestaat uit drie componenten A, B en C. Van een grote partij van deze apparaten is bekend, dat bij 76% component A functioneert, bij 64% component B en bij 52% zowel A als B. Bij 76% van de partij werkt component C.
Verder is bij 8% van de partij component A en component C defect, bij 16% component B èn component C defect en bij 4% werkt geen der componenten.
Men kiest een willekeurig apparaat uit de partij. Bereken de kans dat
a. minstens één van de componenten A en B werkt;
b. component A wel werkt en component B niet;
c. alle drie de componenten werken.

5. Bij het invullen van het formulier van de voetbaltoto kiest iemand voor ieder van de 13 wedstrijden lukraak één van de cijfers 1,2,3. Bereken de kans dat hij
a. alle 13 uitslagen goed heeft geraden;
b. in elk geval de eerste 10 wedstrijden goed heeft;
c. precies 10 wedstrijden goed heeft;
d. minstens 10 wedstrijden goed heeft.

6. Een parkeerterrein bestaat uit 12 parkeerplaatsen op een rij, waarop 8 auto's geparkeerd staan. Als iedere parkeerder willekeurig heeft gekozen uit de resterende lege plaatsen, hoe groot is dan de kans dat de 4 nog lege plaatsen naast elkaar liggen? Is de gevonden kans groter of kleiner dan u had verwacht?

7. Van 10 personen is bekend dat ze allen in november jarig zijn. Hoe groot is de kans dat er minstens 2 personen op dezelfde dag jarig zijn?

8. Een lineaal van 30 cm lengte breekt op een willekeurige plaats. Bereken de kans dat het langste stuk minstens twee keer zo lang is als het kortste stuk.

9. Iedere 25ste bezoeker van het Spoorwegmuseum krijgt een oud treinkaartje.
a. Als u dit niet weet en nietsvermoedend naar binnen gaat, hoe groot is dan de kans dat u zo'n kaartje krijgt?
b. Als u het wel weet en daarom eerst 15 anderen naar binnen laat gaan, die niets krijgen, hoe groot is dan de kans dat u wel zo'n kaartje krijgt?

10. *a.* Bereken $P(A \cup B)$ als is gegeven $P(A) = \frac{1}{3}$ en $P(B \mid \bar{A}) = \frac{1}{4}$.
b. Bereken $P(B)$ als is gegeven $P(A \cup B) = \frac{2}{3}$ en $P(\bar{A} \mid \bar{B}) = \frac{1}{2}$.
c. Bereken $P(A)$ als is gegeven dat er van de gebeurtenissen A, B en C minstens één optreedt maar dat ze niet alle drie tegelijk kunnen optreden en voorts dat $P(A \cup (BC) \mid B) = \frac{1}{2}$, $P(A \mid \bar{C}) = \frac{1}{2}$ en $P(A) = P(B) = P(C)$.

11. De machines A, B en C produceren respectievelijk 60%, 30% en 10% van de door een fabriek afgeleverde werkstukken. De percentages ondeugdelijke werkstukken bedragen voor de drie machines respectievelijk 2%, 3% en 4% van de produktie van de betreffende machine.
Een willekeurig gekozen werkstuk blijkt ondeugdelijk te zijn. Bepaal de kans dat het werkstuk is gemaakt op machine C.

12. Van een meetapparaat voor chips is bekend dat $P(A \mid B) = P(\bar{A} \mid \bar{B}) = p$, waarbij A de gebeurtenis is dat het meetapparaat de chip als ondeugdelijk classificeert terwijl B de gebeurtenis is dat de chip ondeugdelijk is. Dit meetapparaat wordt gebruikt om de slechte exemplaren in een grote partij chips aan te wijzen. In die partij is 5% van de chips ondeugdelijk.
a. Omschrijf de betekenis van $P(\bar{B} \mid A)$.
b. Bereken $P(\bar{B} \mid A)$ als $p = 0{,}95$.
c. Hoe groot moet p zijn opdat $P(B \mid A) = 0{,}9$?

13. Er zijn 10 apparaten in voorraad; 2 van de apparaten zijn defect. Er worden 9 apparaten verkocht.
a. Bereken de kans dat het derde apparaat dat wordt verkocht defect is.
b. Bereken de kans dat het resterende apparaat defect is.

14. Uit een doos met 1000 houtschroeven waarvan er 50 onvoldoende gegalvaniseerd zijn worden achtereenvolgens voor de bouw van een fietsenstalling schroeven genomen en gebruikt. Hoe groot is de kans dat de 1e, 2e, 3e,...,k-de gebruikte schroef binnen afzienbare tijd zal roesten? Hoe groot is de kans dat er van de k-de en de (k+1)-de gebruikte schroef juist één zal roesten?

15. De kans dat een transistor langer dan t uren functioneert is $10^{-t/100.000}$. Na tien jaar in bedrijf te zijn geweest wordt besloten deze transistor te vernieuwen om de onaangename gevolgen van falen in belaste toestand te voorkomen (preventief onderhoud). Als de gegeven tijdsafhankelijke kans juist is, is dit preventief onderhoud dan zinvol?
Bereken eens de conditionele kans op een levensduur langer dan 10 jaar + t uur, gegeven dat de levensduur van de transistor groter dan tien jaar is.

16. Een schakeling wordt opgebouwd uit twee weerstanden en twee condensatoren. Er zijn acht weerstanden beschikbaar. Het is bekend dat er daarvan slechts vijf goed functioneren. Van de vijf beschikbare condensatoren functioneren er slechts drie goed.
Bij de bouw van de schakeling kiest men willekeurig uit de beschikbare componenten. Bereken de kans dat:
a. de schakeling goed werkt;
b. precies twee van de gekozen componenten goed werken;
c. beide weerstanden goed werken als bekend is dat precies twee van de gekozen componenten goed werken.

17. Bij de stormramp van 1953 deed zich een extreem hoge waterstand voor van 6 meter boven N.A.P. We noemen een kalenderjaar waarin een dergelijke hoge waterstand wordt bereikt of overschreden een rampjaar. We stellen de kans dat een jaar een rampjaar is gelijk aan p. Rampjaren komen onafhankelijk van elkaar voor. Hoe groot is de kans dat zich in de 21e eeuw minstens één zo'n ramp zal voordoen?

18. Noem A de gebeurtenis 'in een gezin zijn zowel jongens als meisjes'. Noem B de gebeurtenis 'er is hoogstens één jongen'. Neem aan dat P(jongen) = P(meisje) = $\frac{1}{2}$.
a. Laat zien dat A en B onafhankelijk zijn wanneer er in het gezin drie kinderen zijn
b. Ga na of onafhanklijkeheid van A en B geldt voor een gezin met twee kinderen.

19. Uit ervaring is bekend dat een cafébezoeker bij het verlaten van een café met gelijke kans (nl. $\frac{1}{2}$) linksaf of rechtsaf gaat. Het gedrag van de man wordt waargenomen bij twee achtereenvolgende cafébezoeken. De volgende gebeurtenissen worden onderscheiden:

A = {na het eerste cafébezoek gaat de man rechtsaf};
B = {na het tweede cafébezoek gaat de man rechtsaf};
C = {de man gaat eenmaal rechtsaf en eenmaal linksaf}.

a. Toon aan dat de gebeurtenissen A, B en C paarsgewijs onafhankelijk zijn.
b. Onderzoek of de gebeurtenissen A, B en C onderling onafhankelijk zijn.

20. Drie schutters schieten ieder eenmaal. De kans dat in de roos wordt geschoten is respectievelijk $\frac{1}{6}$, $\frac{1}{4}$ en $\frac{1}{3}$.
a. Bereken de kans dat precies één van de drie in de roos schiet.
b. Als slechts één van de drie in de roos schiet, hoe groot is dan de kans dat dit de slechtste schutter is?

21. Als gegeven is dat

$$F(x) = \begin{cases} 0, & x < 0; \\ x^2, & 0 \le x < \frac{1}{2}; \\ 1 - 3(1-x)^2, & \frac{1}{2} \le x < 1; \\ 1, & x \ge 1; \end{cases}$$

kan F(x) dan de verdelingsfunctie van een stochastische variabele zijn? Zo ja, bepaal dan de bijbehorende kansdichtheid.

22. De continue stochastische variabele $\underline{x}$ heeft als dichtheid

$$f_x(x) = \begin{cases} \frac{1}{10} e^{-x/10}, & x > 0; \\ 0, & \text{elders.} \end{cases}$$

a. Bepaal $F_x(x)$.
b. Bereken $P(5 < \underline{x} \le 15)$.
c. Bereken $E(\underline{x})$.

23. Een aannemer weet uit ervaring dat het laagste bedrag waarvoor door de andere aannemers op een werk wordt ingeschreven uniform verdeeld is over het interval $(\frac{3}{4}C, 2C)$, waarbij C zijn eigen schatting is van de kosten (zonder winst of verlies) van het werk. Definieer de winst van de aannemer als 0 als het werk hem niet wordt gegund en als het verschil tussen zijn inschrijfbedrag en C als het werk hem wel wordt gegund. Als het werk naar de laagste inschrijver gaat, voor welk bedrag moet hij dan inschrijven om de verwachting van zijn winst zo groot mogelijk te maken?

24. In de maand maart is $\underline{w}$, de temperatuur in °C 's avonds om 18.00 uur, uniform verdeeld op [–1,10].
a. Geef $f_w(w)$, $E(\underline{w})$ en $Var(\underline{w})$;

b. $\underline{z}$ is dezelfde temperatuur in °F. Geef de dichtheid, de verwachting en de variantie van $\underline{z}$.

25. Bij een tentamen waaraan 50 studenten deelnamen werden de volgende resultaten behaald:

Cijfer:	1	2	3	4	5	6	7	8	9	10
Aantal:	0	1	2	4	11	12	10	6	3	1

Zij $\underline{x}$ het cijfer van de student die tijdens het tentamen aan tafel 37 zat. Geef $p_x(k)$, $E(\underline{x})$ en $Var(\underline{x})$.

26. Iemand schiet op een ronde schijf. De schijf heeft een straal van 5 cm en is door vier concentrische cirkels, met straal resp. 1, 2, 3 en 4 cm, verdeeld in vijf gebieden. De kans dat de schijf wordt geraakt is $\frac{1}{2}$; de plaats waar de schijf wordt geraakt is willekeurig.

a. Bereken de kans dat gebied i (i = 1,2,3,4,5, genummerd van binnen naar buiten) wordt getroffen.

De schutter krijgt 6–i punten wanneer hij gebied i raakt en 0 punten als hij naast de schijf schiet. Zij $\underline{x}$ het aantal punten bij eenmaal schieten.

b. Geef de kansverdeling, de verwachting en de variantie van $\underline{x}$.

c. Dezelfde schutter schiet nu 27 keer. $\underline{z}$ is het aantal keren dat gebied 1 wordt getroffen. Geef $p_z(k)$, $E(\underline{z})$ en $Var(\underline{z})$.

d. Dezelfde schutter schiet nu totdat gebied 1 voor de eerste maal wordt geraakt. $\underline{w}$ is het aantal malen dat wordt geschoten. Geef $p_w(k)$, $E(\underline{w})$ en $Var(\underline{w})$.

27. Iemand besluit net zolang deel te nemen aan de Staatsloterij tot hij een keer een prijs wint. Voor iedere trekking koopt hij precies één lot. $\underline{x}$ is het aantal keren dat hij een lot moet kopen. De kans op een prijswinnend lot is 0,6.

a. Bepaal de verdelingsfunctie en de kansfunctie van $\underline{x}$.

b. Hoe groot is de kans dat hij minstens vier keer een lot moet kopen?

c. Bereken de kans dat hij hoogstens *f* 150,– aan loten moet besteden als de prijs per lot *f* 25,– bedraagt.

d. Bereken $E(\underline{x})$.

28. U koopt één lot in een loterij met 1000 loten. Er wordt, zodra alle loten zijn verkocht, één prijs uitgekeerd van *f* 100,–.

a. De stochastische variabele $\underline{x}$ is het bedrag dat aan u als prijs wordt uitgekeerd. Bepaal de kansverdeling van $\underline{x}$. Bereken $E(\underline{x})$ en $Var(\underline{x})$.

b. De loten kosten *f* 0,25 per stuk. De stochastische variabele $\underline{z}$ is uw netto-opbrengst. Bepaal de kansverdeling van $\underline{z}$. Bereken $E(\underline{z})$ en $Var(\underline{z})$.

29. De politie weet uit ervaring dat 15% van de rondrijdende automobilisten geen geldig rijbewijs heeft. Bij een verkeerscontrole zullen 50 auto's worden aangehouden.

a. Zij $\underline{x}$ het aantal zonder geldig rijbewijs rijdende chauffeurs bij die 50. Bepaal $p_x(k)$ en bereken $E(\underline{x})$ en $Var(\underline{x})$.

b. Zij $\underline{z}$ het aantal aangehouden chauffeurs met een geldig rijbewijs. Bepaal $p_z(k)$ en bereken $E(\underline{z})$ en $Var(\underline{z})$.

c. Bereken m.b.v. de tabel: $P\{\underline{z} \leq 40\}$, $P\{\underline{z} \geq 36\}$ en $P\{35 < \underline{z} < 45\}$.

30. U probeert een tentamen te halen. Hoewel u iedere keer dat u het tentamen doet de stof goed beheerst, hebt u toch, door de gemene manier waarop de vragen worden gesteld, elke keer maar een kans van 0,83 om het tentamen te halen. De stochastische variabele $\underline{x}$ is gelijk aan het nummer van de poging waarbij u het tentamen haalt.

a. Bepaal $p_x(k)$ en bereken $E(\underline{x})$ en $Var(\underline{x})$.

b. Bereken $P\{\underline{x} \geq 2\}$ en $P\{\underline{x} > 2\}$.

31. Een schip is uitgerust met 12 brandblus-apparaten, waarvan er vier ondeugdelijk zijn. De brandweer komt het schip inspecteren en beproeft vijf willekeurig gekozen brandblussers.

a. Bereken de kans dat er precies twee van die vijf ondeugdelijk zijn.

b. Bereken de kans dat er minder dan twee van die vijf ondeugdelijk zijn.

32. Van 25 schakelaars is bekend dat er vijf tussen zitten die kortsluiting veroorzaken. Iemand heeft voor een toestel vijf van die schakelaars nodig en kiest die willekeurig uit de hele partij.

a. Bereken de kans dat het toestel goed functioneert.

b. Als de man twee van die toestellen tegelijk bouwt hoe groot is dan de kans dat minstens één van die twee toestellen goed functioneert.

33. Van een zeer grote partij boutjes is bekend dat 20% rechtse draad heeft. Iemand grijpt een handvol boutjes (en dat blijken er 50 te zijn).

a. Bepaal de kans dat 10 van de 50 boutjes rechtse draad hebben.

b. Bepaal de kans dat meer dan 10 van de 50 boutjes rechtse draad hebben.

c. Bepaal c zodanig dat de kans dat meer dan c van de 50 boutjes rechtse draad hebben kleiner is dan 0,05.

34. Van een grote partij vaten olie is 0,2% lek. Een schip laadt 500 van die vaten olie.

a. Bepaal de kans p_0 dat géén van de vaten lekt.

b. Bepaal de kans p_i dat i van die vaten lekken voor i = 1,2,3.

c. Bepaal de kans dat minstens vier vaten lekken.

35. Een speelautomaat keert met kans 1/1000 een prijs uit. Iemand speelt 1000 keer; $\underline{y}$ is het aantal malen dat daarbij een prijs wordt uitgekeerd.
a. Bepaal $p_y(k)$ en $F_y(k)$.
b. Bereken $E(\underline{y})$ en $Var(\underline{y})$.

36. De stochastische variabele $\underline{x}$ is uniform verdeeld op (–1,+1). Bepaal de variantie van $\underline{x}$.

37. De levensduur in jaren, $\underline{x}$, van een koelkast is exponentieel verdeeld met verwachting 5.
a. Bereken $P(\underline{x} > 4)$.
De fabrikant garandeert een levensduur van t jaar, en kiest t natuurlijk zo groot mogelijk. Hij wil echter een risico van hoogstens 5% lopen dat een klant zich terecht beklaagt.
b. Bepaal t.

38. De stochastische variabele $\underline{x}$ is Normaal verdeeld met verwachting 1 en variantie 4.
a. Bepaal $P(\underline{x} > 0)$.
b. Bepaal $P(|\underline{x}| > 3)$.
c. Als $P(\underline{x} < b) = 0{,}10$, wat is dan de waarde van b?

39. De stochastische variabele $\underline{x}$ is binomiaal verdeeld met parameters n = 50 en p = 2/3.
a. Bepaal $P(\underline{x} \leq 40)$ en $P(\underline{x} = 40)$.
De stochastische variabele $\underline{y}$ is Normaal verdeeld. $E(\underline{y}) = E(\underline{x})$ en $\sigma^2(\underline{y}) = \sigma^2(\underline{x})$.
b. Bepaal $P(\underline{y} \leq 40)$, $P(\underline{y} \leq 40{,}5)$ en $P(39{,}5 < \underline{y} < 40{,}5)$.

40. De stochastische variabele $\underline{x}$ is binomiaal verdeeld met parameters n = 100 en p = 0,05. De stochastische variabele $\underline{y}$ is Poisson-verdeeld met parameter $\lambda = 5$.
a. Bepaal $P(\underline{x} = 4)$ en $P(\underline{y} = 4)$.
De stochastische variabele $\underline{v}$ is Normaal verdeeld met $E(\underline{v}) = E(\underline{x})$ en $\sigma^2(\underline{v}) = \sigma^2(\underline{x})$; $\underline{w}$ is Normaal verdeeld met $E(\underline{w}) = E(\underline{y})$ en $\sigma^2(\underline{w}) = \sigma^2(\underline{y})$.
b. Bepaal $P(3{,}5 < \underline{v} < 4{,}5)$ en $P(3{,}5 < \underline{w} < 4{,}5)$.

41. F is de simultane verdelingsfunctie van $\underline{x}$ en $\underline{y}$. Toon aan dat voor alle a,b,c en d met a < b en c < d geldt:

$$P(a < \underline{x} \leq b;\ c < \underline{y} \leq d) = F(b,d) - F(b,c) - F(a,d) + F(a,c).$$

42. De simultane dichtheid van $\underline{u}$ en $\underline{v}$ is

$$f_{u,v}(u,v) = cuv, \qquad 0 < u < 1; 0 < v < 1;$$
$$= 0, \qquad \text{elders.}$$

a. Bereken c.
b. Bepaal $f_u(u)$.
c. Bereken $P(0 < \underline{u} \le \frac{1}{2}; \frac{1}{4} < \underline{v} \le 1)$.
d. Bereken $P(\underline{u} < \underline{v})$.

43. $\underline{x}$ en $\underline{y}$ zijn onafhankelijk en beide Poisson-verdeeld met parameter λ, resp. μ.
a. Bepaal $P(\underline{x} = j; \underline{y} = k - j)$ voor $j \in \mathbb{N}; k \in \mathbb{N}; j \le k$.
b. Bepaal $P(\underline{x} + \underline{y} = k)$ voor $k \in \mathbb{N}$.

44. De stochastische variabelen $\underline{x}$ en $\underline{y}$ zijn onafhankelijk, $\underline{x} \sim B(n; p)$ en $\underline{y} \sim B(m; p)$. Bepaal de verdeling van $\underline{x} + \underline{y}$.

45. De stochastische variabelen $\underline{x}$ en $\underline{y}$ hebben de simultane dichtheid

$$f_{x,y}(x,y) = 6e^{-2x-3y}, \qquad x > 0; y > 0;$$
$$= 0, \qquad \text{elders.}$$

a. Bepaal de marginale dichtheid van $\underline{x}$ en van $\underline{y}$.
b. Zijn $\underline{x}$ en $\underline{y}$ onafhankelijk?
c. Bereken $P(\underline{x} < \underline{y})$ en $P(\underline{x} > \underline{y})$.

46. De simultane dichtheid van $\underline{x}$ en $\underline{y}$ is:

$$f_{x,y}(x,y) = 21x^2y^3, \qquad 0 < x < y < 1;$$
$$= 0, \qquad \text{elders.}$$

a. Bepaal de marginale dichtheid van $\underline{x}$ en van $\underline{y}$.
b. Bereken de covariantie van $\underline{x}$ en $\underline{y}$.
c. Zijn $\underline{x}$ en $\underline{y}$onafhankelijk?

47. Er wordt 15 maal geworpen met een dobbelsteen. Het aantal malen dat 1 resp. 2 wordt gegooid wordt aangegeven met $\underline{x}$ resp. $\underline{y}$.
a. Is $\text{cov}(\underline{x},\underline{y})$ negatief, positief of nul?
b. Bereken $\text{cov}(\underline{x},\underline{y})$ en $\rho(\underline{x},\underline{y})$.

48. De stochastische variabelen $\underline{x}$ en $\underline{y}$ hebben de simultane kansdichtheid

$$f_{x,y}(x,y) = \frac{12}{(1+x+y)^5}, \qquad x > 0; y > 0;$$
$$= 0, \qquad \text{elders.}$$

a. Onderzoek of $\underline{x}$ en $\underline{y}$ onafhankelijk zijn.
b. Als $\underline{u} = \underline{x} + \underline{y}$, bereken dan $\sigma^2(\underline{u})$.

49. Een vaas bevat 1 witte, 2 rode en 2 zwarte ballen. Aselect en zonder teruglegging worden 2 ballen uit deze vaas getrokken. De stochastische variabele $\underline{x}$ ($\underline{y}$) stelt het aantal getrokken witte (rode) ballen voor.
a. Bepaal de simultane kansfunctie van $\underline{x}$ en $\underline{y}$.
b. Bereken $P(\underline{x} > \underline{y})$ en $P(\underline{x} < \underline{y})$.
c. Bepaal de marginale kansfunctie van $\underline{y}$.
d. Bereken $E(\underline{y})$, $\sigma^2(\underline{y})$ en $E(\underline{x}\,\underline{y})$.
e. Bereken $\rho(\underline{x},\underline{y})$.

50. De stochastische variabelen $\underline{x}$ en $\underline{y}$ zijn onafhankelijk en hebben dezelfde verdeling. Zij $\underline{u} = \underline{x} + \underline{y}$ en $\underline{v} = \underline{x} - \underline{y}$.
a. Bereken $\rho(\underline{u},\underline{x})$.
b. Bereken $\rho(\underline{u},\underline{v})$.

51. De stochastische variabele $\underline{x}$ is geometrisch verdeeld met parameter $p = \frac{1}{3}$.
a. Bepaal $P(\underline{x} = k \mid \underline{x} \geq 3)$ voor $k = 3,4,\ldots$.
b. Bereken $E(\underline{x} \mid \underline{x} \geq 3)$.
Zij $\underline{z} = \underline{x} - 2$.
c. Bepaal de verdeling van $\underline{z}$, gegeven $\underline{x} \geq 3$.

52. De stochastische variabele $\underline{v}$ heeft als dichtheid

$$
\begin{aligned}
f_v(v) &= 0{,}08v, \qquad && 0 < v < 5; \\
&= 0, && \text{elders.}
\end{aligned}
$$

a. Bepaal $f_v(v \mid \underline{v} > 2)$.
b. Bepaal $F_v(v \mid \underline{v} > 2)$.
c. Bereken $\sigma^2(\underline{v} \mid \underline{v} > 2)$.

53. De stochastische variabelen $\underline{x}$ en $\underline{y}$ hebben de simultane dichtheid

$$
\begin{aligned}
f_{x,y}(x,y) &= x + y, \qquad && 0 < x < 1; 0 < y < 1; \\
&= 0, && \text{elders.}
\end{aligned}
$$

a. Bepaal de voorwaardelijke dichtheid van $\underline{x}$ als gegeven is dat $\underline{y} < \frac{1}{2}$.
b. Bepaal de voorwaardelijke dichtheid van $\underline{x}$ als gegeven is dat $\underline{y} = y$.

54. Bij de uitvoering van een rij Bernoulli-experimenten, elk met kans p op succes, stelt $\underline{x}$ het aantal successen voor bij de eerste m experimenten en $\underline{n}$ het rangnummer van het experiment waarbij het derde succes optreedt.
Bepaal $E(\underline{x} \mid \underline{n} = k)$ voor $k = 3,4,\ldots$.

55. Zij $\underline{x}$ het aantal uitgaande telefoongesprekken in de TU-centrale dat in een uur begint. Neem aan dat $\underline{x}$ Poisson-verdeeld is met parameter λ. Elk gesprek heeft een kans p om langer dan 3 minuten te duren. Zij $\underline{y}$ het aantal lange gesprekken in dat uur. Bepaal de verdeling van $\underline{y}$.

56. De simultane verdeling van $\underline{x}$ en $\underline{y}$ is uniform over het oppervlak van de cirkel met middelpunt 0 en straal 1.
a. Bepaal de voorwaardelijke dichtheid van $\underline{x}$ bij gegeven $\underline{y} = y$.
b. Bereken de covariantie van $\underline{x}$ en $\underline{y}$.
c. Zijn $\underline{x}$ en $\underline{y}$ onafhankelijk?

57. De stochastische variabele $\underline{x} \sim B(75; \frac{1}{3})$. Geef een benadering m.b.v. de CLS van de volgende kansen:
a. $P(\underline{x} < 31)$.
b. $P(\underline{x} \leq 22)$.
c. $P(21 < \underline{x} < 31)$.

58. Van de stochastische variabele $\underline{x}$ is gegeven dat $E(\underline{x}) = 3$ en $E(\underline{x}^2) = 13$. Bepaal een ondergrens voor $P(-2 < \underline{x} < 9)$ m.b.v. de ongelijkheid van Chebychev.

59. *a.* Bereken de variantie van $\underline{x}$, als $\underline{x}$ exponentieel verdeeld is met parameter λ. Stel nu $\lambda = 1$.
b. Bepaal de kleinste waarde van t waarvoor $P(|\underline{x} - 1| \geq t) \leq 0{,}35$.
c. Wat zegt de ongelijkheid van Chebychev over de waarde van t, als $\underline{x}$ een onbekende verdeling heeft met dezelfde verwachting en variantie?

60. $\underline{x}$ en $\underline{y}$ zijn binormaal verdeeld met als parameters $\mu_x = 3$, $\mu_y = 1$, $\sigma_x^2 = 16$, $\sigma_y^2 = 25$ en $\rho = \frac{3}{5}$. Bereken de volgende kansen:
a. $P(3 < \underline{y} < 8)$.
b. $P(3 < \underline{y} < 8 \mid \underline{x} = 7)$.
c. $P(-3 < \underline{x} < 3)$.
d. $P(-3 < \underline{x} < 3 \mid \underline{y} = -4)$.

61. De stochastische variabelen $\underline{x}$ en $\underline{y}$ zijn uniform verdeeld over het eenheidsvierkant. Zij $\underline{u} = \underline{x} + \underline{y}$ en $\underline{v} = \underline{x} - \underline{y}$. Bepaal de simultane dichtheid van $\underline{u}$ en $\underline{v}$.

62. De stochastische variabelen $\underline{x}$ en $\underline{y}$ zijn onafhankelijk en beide exponentieel verdeeld met parameter λ. Bepaal de dichtheid van $\underline{z} = \underline{x} + \underline{y}$.

63. De stochastische variabelen $\underline{x}$ en $\underline{y}$ zijn onafhankelijk en beide N(0;1)-verdeeld.
a. Zijn $\underline{u} = \underline{x} + \underline{y}$ en $\underline{v} = \underline{x} - \underline{y}$ onafhankelijk?
b. Wat is de simultane dichtheid van $\frac{1}{2}\underline{u}\sqrt{2}$ en $\frac{1}{2}\underline{v}\sqrt{2}$?
c. Toon aan dat $2\underline{x}\,\underline{y}$ en $\underline{x}^2 - \underline{y}^2$ dezelfde verdeling hebben.

64. De stochastische variabelen $\underline{x}$ en $\underline{y}$ hebben de simultane kansfunctie

$$\begin{aligned} p_{x,y}(x,y) &= \frac{xy}{36}, & x = 1,2,3;\ y = 1,2,3; \\ &= 0, & \text{elders.} \end{aligned}$$

a. Bepaal de simultane kansfunctie van $\underline{u} = \underline{x}\,\underline{y}$ en $\underline{v} = \underline{x}$.

b. Bepaal de marginale kansfunctie van $\underline{v}$.

65. De stochastische variabelen $\underline{u}$, $\underline{v}$ en $\underline{w}$ zijn onafhankelijk en alle standaard-Normaal verdeeld. De stochastische variabelen $\underline{x}$, $\underline{y}$ en $\underline{z}$ zijn gedefinieerd door:

$$\begin{aligned} \underline{u} &= \underline{x} \cos \underline{y} \sin \underline{z}; \\ \underline{v} &= \underline{x} \sin \underline{y} \sin \underline{z}; \\ \underline{w} &= \underline{x} \cos \underline{z}; \end{aligned}$$

terwijl $0 \leq \underline{x} < \infty$; $0 \leq \underline{y} < 2\pi$; $0 \leq \underline{z} < \pi$.

Toon aan dat $\underline{x}$, $\underline{y}$ en $\underline{z}$ onafhankelijk zijn.

66. Uit een volledig kaartspel wordt aselect en met teruglegging herhaaldelijk een kaart getrokken. De stochastische variabele $\underline{x}$ is gedefinieerd, als het rangnummer van de trekking, waarbij de derde schoppenkaart verschijnt.

a. Bepaal de kansverdeling van $\underline{x}$; bereken $E(\underline{x})$.

b. Bepaal de conditionele kansverdeling van $\underline{x}$, als gegeven is dat onder de eerste vier getrokken kaarten geen schoppenkaart voorkomt. Bereken eveneens de met dit gegeven corresponderende conditionele verwachting van $\underline{x}$.

67. A en B werpen om de beurt een onzuivere munt, net zolang tot iemand een worp doet die hetzelfde resultaat heeft als de direct daaraan voorafgaande worp. Zij $\underline{x}$ het totaal aantal worpen en p de kans op munt ($q = 1 - p$ de kans op kruis) per worp.

a. Bepaal de kansfunctie van $\underline{x}$.

b. Bepaal $E(\underline{x})$.

c. Bepaal de kans dat de laatste worp wordt gedaan door A als gegeven is dat A met het gooien begint.

68. Zij de verdeling van $\underline{x}$ binomiaal met parameters n en p en die van $\underline{y}$ negatief binomiaal met parameters r en p. Toon het volgende verband aan tussen de verdelingsfuncties van $\underline{x}$ en $\underline{y}$: $F_x(r-1) = 1 - F_y(n)$.

69. De stochastische variabelen $\underline{x}_1, \underline{x}_2, \ldots$ zijn onafhankelijk en hebben alle dezelfde dichtheid

$$\begin{aligned} f(x) &= \lambda e^{-\lambda x}, & x > 0; \\ &= 0, & \text{elders.} \end{aligned}$$

Toon m.b.v. volledige inductie aan dat voor $n = 1,2,\ldots$ de dichtheid van

$\underline{z}_n = \sum_{i=1}^{n} \underline{x}_i$ gelijk is aan:

$$f_{z_n}(y) = \frac{\lambda^n y^{n-1}}{(n-1)!} e^{-\lambda y}, \quad y > 0;$$
$$= 0, \quad \text{elders.}$$

70. Apparaat A bestaat uit drie in serie geschakelde componenten. De levensduren van de componenten zijn onafhankelijk en exponentieel verdeeld met dezelfde parameter λ.

a. Bepaal de verdeling van de tijd dat het apparaat kan werken.

b. Dezelfde vraag als de componenten parallel geschakeld zijn.

71. Een steekproef leverde het volgende resultaat:

440	591	590	424	564	582	412	443
532	505	417	557	533	394	507	

Bepaal het steekproefgemiddelde, de steekproefvariantie en de steekproefstandaardafwijking.

72. Een steekproef met omvang 100 gaf de volgende resultaten

waarde x_i:	1,10	2,30	4,45	6,22	8,47	9,01
aantal f_i:	3	19	29	37	8	4

Bepaal de steekproefstandaardafwijking.

73. De stochastische variabele $\underline{x}$ is exponentieel verdeeld met $E(\underline{x}) = 1/\lambda$.

a. Bepaal de verdeling van $\underline{y} = 2\lambda\underline{x}$.

b. Bepaal m.b.v. de χ^2-tabel de waarde van a als gegeven is dat $P(\underline{x} < a) = 0{,}75$ en $\lambda = \frac{1}{3}$.

74. De stochastische variabele $\underline{x}$ is uniform verdeeld op het interval $(a-b;\ a+b)$. Bepaal a en b zodanig dat $E(\underline{x})$ en $\sigma^2(\underline{x})$ gelijk zijn aan de verwachting, resp. de variantie van de $\chi^2(8)$-verdeling.

75. Zij $\underline{x}_1, \underline{x}_2, \underline{x}_3, \underline{x}_4$ een aselecte steekproef uit een $N(\mu;\sigma^2)$-verdeelde populatie. $\underline{z} = c\{(\underline{x}_1 - \underline{x}_2)^2 + (\underline{x}_3 - \underline{x}_4)^2\}$.

a. Bepaal c zodanig dat $E(\underline{z}) = \sigma^2$.

b. Bepaal voor de zojuist berekende waarde van c de variantie van $\underline{z}$.

c. Vergelijk deze resultaten met $E(\underline{s}^2)$ en $Var(\underline{s}^2)$.

76. Zijn $\underline{x}_1, \underline{x}_2, \ldots$ onderling onafhankelijke stochastische variabelen, ieder met een standaard-Normale dichtheid en zij $\underline{s}_k = \underline{x}_1 + \ldots + \underline{x}_k$.

a. Bepaal de simultane dichtheid van $\underline{s}_m$ en $\underline{s}_n$ en de conditionele dichtheid van $\underline{s}_m$ gegeven dat $\underline{s}_n = t$.

b. Bepaal de simultane dichtheid van $\underline{u} = \underline{x}_1^2 + \ldots + \underline{x}_m^2$ en $\underline{v} = \underline{x}_1^2 + \ldots + \underline{x}_n^2$ en de conditionele dichtheid van $\underline{u}$, gegeven $\underline{v} = v$.

77. De stochastische variabele $\underline{t} \sim t(14)$.
a. Bepaal a zó dat $P(\underline{t} > a) = 0{,}10$.
b. Bepaal b zó dat $P(|\underline{t}| > b) = 0{,}10$.
c. Bepaal c zó dat $P(|\underline{t}| < c) = 0{,}98$.

78. De stochastische variabele $\underline{w} \sim F(6;\, 10)$.
a. Bepaal a zó dat $P(\underline{w} \geq a) = 0{,}95$.
b. Bepaal b zó dat $P(\underline{w} \leq b) = 0{,}95$.
c. Bereken $P(a \leq \underline{w} \leq b)$.

79. Uit een Normale populatie met bekende verwachting $\mu = \mu_0$ en onbekende variantie σ^2 wordt een steekproef $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ getrokken.
Toon aan dat $\underline{z} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (\underline{x}_i - \mu_0)^2$ een $\chi^2(n)$-verdeling heeft.
Bepaal $E(\underline{z})$ en $Var(\underline{z})$.

80. Uit een Normale populatie met verwachting 5 en variantie 16 wordt een aselecte steekproef $\underline{x}_i$, $i = 1,2,\ldots,25$ getrokken.
a. Bepaal $P(4 < \underline{\bar{x}} < 6)$.
b. Bepaal a zó dat $P(\underline{s}^2 > a) = 0{,}05$.
c. Bepaal b zó dat $P(\underline{z} > b) = 0{,}05$ waarbij $\underline{z} = \frac{1}{25} \sum_{i=1}^{25} (\underline{x}_i - 5)^2$.

81. $\underline{x}$ is een waarneming aan een Bernoulli-verdeelde stochastische variabele met onbekende parameter p. Welke van de volgende drie functies van $\underline{x}$ zijn zuivere schatters van p?
$\underline{t}_1 = \underline{x}$, $\underline{t}_2 = \frac{1}{2}$, $\underline{t}_3 = 2p - \underline{x}$.

82. Een vaas bevat z zwarte en w witte ballen. z en w zijn onbekend. Bepaal de ML-schatter voor de verhouding $r = z/w$ op basis van een steekproef verkregen door:
a. Aselect en met teruglegging n ballen te trekken;
b. Aselect en met teruglegging ballen te trekken totdat er een zwarte bal wordt getrokken en deze procedure n maal te herhalen. Daarbij zijn $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ de aantallen getrokken ballen (inclusief de zwarte).

83. Aan de stochastische variabele $\underline{x}$ met dichtheid

$$f(z) = \begin{cases} K(\theta) z e^{-z/\theta}, & z > 0; \\ 0, & \text{elders.} \end{cases}$$

worden n onafhankelijke waarnemingen $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ gedaan.

a. Welke waarde heeft de constante $K(\theta)$?
b. Bepaal verwachting en variantie van $\underline{x}$.
c. Bepaal een momentenschatter voor θ.
d. Bepaal de meest aannemelijke schatter voor θ.
e. Bepaal de meest aannemelijke schatter voor var $\underline{x}$.
f. Bepaal een momentenschatter voor var $\underline{x}$ door het tweede steekproefmoment te relateren aan het tweede moment van $\underline{x}$.
g. Onderzoek de zuiverheid van de onder e. en f. gevonden schatters voor var $\underline{x}$.

84. De levensduur van buizen is exponentieel verdeeld met parameter λ. Gegeven zijn twee buizen met levensduren $\underline{x}_1$, $\underline{x}_2$.
a. Een schatter voor de verwachte levensduur is $\overline{\underline{x}}$. Gevraagd verwachting, variantie en gemiddelde kwadratische fout.
b. Een andere schatter voor de verwachte levensduur is het meetkundig gemiddelde: $\overline{\underline{x}}_g = \sqrt{\underline{x}_1\underline{x}_2}$. Gevraagd verwachting, variantie en gemiddelde kwadratische fout.
c. Maak de onder b. gegeven schatter zuiver. Bereken de variantie van deze zuivere schatter.
d. Aan welke van de schatters bedoeld onder a. en c. geeft u de voorkeur en waarom?

85. De stochastische variabele $\underline{x}$ is uniform verdeeld op $(0,\theta)$. Aan $\underline{x}$ worden n onafhankelijke waarnemingen $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ gedaan. Op basis van deze steekproef kunnen we de volgende schatters voor θ introduceren:

$\underline{t}_1 = 2\overline{\underline{x}}$ (een momentenschatter);

$\underline{t}_2 = \max\limits_i(\underline{x}_i)$ (de ML-schatter);

$\underline{t}_3 = (n+1)\min\limits_i(\underline{x}_i)$.

a. Welke van deze schatters zijn zuiver?
b. Maak de onzuivere schatters zuiver.
c. Bepaal van de zo verkregen zuivere schatters de variantie.
d. Welke schatter heeft uw voorkeur?

86. Zij $\overline{\underline{x}}$ het steekproefgemiddelde van n onafhankelijke waarnemingen uit een $N(\mu;\sigma^2)$-verdeling, $\underline{s}^2$ de steekproefvariantie.
a. Als $n = 20$, $\sigma^2 = 80$, $\overline{x} = 81{,}2$, geef dan een 95%-betrouwbaarheidsinterval voor μ.
b. Als $\sigma^2 = 9$, hoe groot moet n dan minstens zijn opdat $(\overline{x}-1, \overline{x}+1)$ een 90%-betrouwbaarheidsinterval voor μ is?

c. Als $\overline{x} = 4{,}7$, $s^2 = 5{,}76$, $n = 16$ bepaal dan een 90%-betrouwbaarheidsinterval voor μ.

d. Als $n = 15$, $\overline{x} = 3{,}2$ en $s^2 = 4{,}24$, geef dan een 90%-betrouwbaarheidsinterval voor σ^2.

e. Zij $n_1 = 16$, $n_2 = 10$, $\overline{x}_1 = 3{,}6$, $\overline{x}_2 = 13{,}6$, $s_1^2 = 4{,}14$ en $s_2^2 = 7{,}26$. Geef een 90%-betrouwbaarheidsinterval voor de verhouding σ_2^2/σ_1^2 van de varianties van de populaties waaruit de steekproeven 1 en 2 getrokken zijn. De populatiegemiddelden μ_1 en μ_2 zijn onbekend.

f. De volgende negen getallen vormen een aselecte steekproef uit een $N(8; \sigma^2)$-verdeling. Bepaal hieruit een 90%-betrouwbaarheidsinterval voor σ^2. (8,6; 7,9; 8,3; 6,4; 8,4; 9,8; 7,2; 7,8; 7,5).

87. Bij een duurzaamheidsproef werden 12 apparaten belast tot het moment dat zij niet meer functioneerden. De gemiddelde levensduur was 330 uren. Geef een 95%-betrouwbaarheidsinterval voor de verwachte levensduur van een apparaat, aannemende dat de levensduurverdeling exponentieel is.

88. Gegeven zijn de steekproeven

$$75,\ 80,\ 85,\ 84,\ 82,\ 68,\ 78$$

en

$$40,\ 39,\ 45,\ 44,\ 38$$

afkomstig uit Normale verdelingen.
Ga na met behulp van een betrouwbaarheidsinterval (betrouwbaarheid 0,90) of de varianties verschillend zijn.

89. De druksterkten in kg/cm^2 van vijf monsters pressup-beton bleken

350, 360, 400, 370 en 380.

De druksterkten van vijf monsters readymix-beton, op analoge wijze beproefd waren:

340, 350, 320, 330 en 350.

Aannemende dat de waarnemingen afkomstig zijn uit twee Normale verdelingen $N(\mu_1; \sigma^2)$ en $N(\mu_2; \sigma^2)$, wordt gevraagd een tweezijdig begrensd betrouwbaarheidsinterval voor $\mu_1 - \mu_2$ te construeren (betrouwbaarheid 90%).
Als men wenst na te gaan of pressup-beton een grotere druksterkte heeft dan readymix-beton, is een tweezijdig betrouwbaarheidsinterval dan een geschikte keuze?

90. $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ is een aselecte steekproef uit een populatie die uniform verdeeld is op $(\theta - \frac{1}{2}; \theta + \frac{1}{2})$. Zij $\underline{v} = \min_i \underline{x}_i$ en $\underline{w} = \max_i \underline{x}_i$. Als betrouwbaarheidsinterval voor θ gebruikt men (v,w) waarbij v en w de realisaties zijn van $\underline{v}$, resp. $\underline{w}$.

a. Hoe groot is de betrouwbaarheid van (v,w)?

b. Dezelfde vraag als de populatie $N(\theta; 1)$-verdeeld is en $n = 2$.

c. Geef een op $\overline{x}$ gebaseerd betrouwbaarheidsinterval voor θ in dit laatste geval met gelijke betrouwbaarheid als onder b.

d. Bereken de verwachte lengte van de drie betrouwbaarheidsintervallen.

91. Uit een Normale populatie met variantie $\sigma^2 = 16$ wordt een steekproef getrokken ter grootte $n = 25$. Toets met $\alpha = 0{,}05$ de hypothese H_0: $\mu = 5$ tegen H_1: $\mu > 5$, als gegeven is dat het steekproefgemiddelde gelijk is aan 5,9.

92. $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_{16}$ is een steekproef uit een Normale populatie met onbekende variantie.Toets de hypothese H_0: $\mu = 10$ tegen H_1: $\mu \neq 10$, als gegeven is dat $\overline{x} = 13$ en $s_x^2 = 22$. Neem als onbetrouwbaarheidsdrempel 10% en 1%. Geef in beide gevallen het kritieke gebied.

93. Een steekproef $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_{10}$ uit een Normale populatie wordt getrokken om de hypothese H_0: $\sigma^2 \leq 5$ tegen $\sigma^2 > 5$ te toetsen, met $\alpha = 0{,}05$.
De steekproef heeft als resultaat:

2,9; 4,1; 3,2; 8,7; 7,5; 9,9; 10,2; 4,2; 7,5; 10,8.

a. Hoe luidt de conclusie als de populatie-verwachting onbekend is?

b. Hoe luidt de conclusie als de populatie-verwachting $\mu = 6$?

94. Een toets H_0: $\mu \leq \mu_0$ tegen H_1: $\mu > \mu_0$ wordt uitgevoerd op basis van een steekproef $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_{16}$ uit een Normale populatie met variantie $\sigma^2 = 9$. Het kritieke gebied is $(15, \infty)$. Neem aan dat $\mu_0 = 13$.

a. Bepaal de kans op een fout van de tweede soort voor $\mu = 13, 14, 15, 17\frac{1}{2}, 20$.

b. Schets het onderscheidingsvermogen van de toets.

95. Om de zuiverheid van een munt te onderzoeken wordt hij 100 maal geworpen. Als men toetst met $\alpha = 0{,}05$, welke uitkomsten liggen dan in het kritieke gebied?

96. Een nieuw geneesmiddel B dient ter vervanging van een middel A dat in 70% van de gevallen succesvol is. Geneesmiddel B bleek bij toepassing op 50 patiënten in 38 gevallen te helpen.

a. Kan men hieruit concluderen (onbetrouwbaarheid 0,05) dat geneesmiddel B beter is dan A?

b. Hoe groot is de kans dat men bij toepassing op 50 patiënten tot de conclusie komt dat B beter is dan A, indien geneesmiddel B in 80% van de gevallen werkzaam is?

c. Dezelfde vraag voor 90% en voor 85%.
d. Geef een schets van het onderscheidingsvermogen.

97. Een Engelse dame beweert dat zij kan proeven of bij het inschenken van thee de melk vóór dan wel ná de thee in de kopjes werd gedaan. Ter controle van deze bewering worden tien kopjes ingeschonken. Afhankelijk van de uitkomst van een loting met een zuivere munt wordt eerst de melk dan wel eerst de thee in een kopje gedaan. Dit wordt haar medegedeeld, zodat zij niet onnodig haar smaakoordeel aanpast in de mogelijke veronderstelling dat precies vijf van de tien kopjes op de ene, en vijf op de andere manier zouden zijn ingeschonken. Als de dame na proeven in acht van de tien gevallen de juiste diagnose stelt, is haar bewering daarmede dan gestaafd bij onbetrouwbaarheidsdrempel 0,05?

98. (Zie som 71) De treksterkte van 15 stukken henneptouw werd bepaald. Men vond (in kg):

440	591	590	424	564	582	412	443
532	505	417	557	533	394	507	

Neem aan dat de treksterkte Normaal verdeeld is.
a. Toets de hypothese dat de waarnemingen afkomstig zijn uit een N(500; σ^2)-verdeling tegen het alternatief dat de verwachting groter is dan 500 (onbetrouwbaarheidsdrempel 0,05).
b. Teken het onderscheidingsvermogen.

99. De levensduurverdeling van gloeilampen, die door een zeker bedrijf gemaakt worden, is Normaal met verwachting 1500 uur en standaardafwijking 100 uur. Iemand stelt een wijziging voor in het produktieproces, waardoor bewerkstelligd zou worden dat de verwachting van de levensduur groter wordt dan 1500 uur. De overige karakteristieken van de levensduurverdeling blijven ongewijzigd.
a. Er worden 100 lampen gemaakt volgens het gewijzigde procédé. Indien de gemiddelde levensduur van de 100 onderzochte lampen minstens gelijk is aan 1525 uur zal men op het gewijzigde produktieproces overschakelen. Bereken de kans dan men ten onrechte overschakelt.
b. Indien de onder a. berekende kans hoogstens gelijk mag zijn aan 0,001, hoeveel lampen dient men dan te onderzoeken?
c. Beschouw H_0: $\mu = 1500$ tegen H_1: $\mu > 1500$. Men wenst slechts 100 lampen te onderzoeken en het kritieke gebied zodanig te kiezen dat de kans op een fout van de eerste soort hoogstens gelijk is aan 0,001. Geef het kritieke gebied van de toets.
d. Bepaal het kritieke gebied en het aantal waarnemingen zodanig dat voldaan is aan de volgende twee eisen:

1. de kans op een fout van de eerste soort is hoogstens gelijk aan 0,001;
2. het onderscheidingsvermogen indien $\mu = 1525$ is minstens gelijk aan 0,99.

100. Een machine produceert kogeltjes met een nominale diameter van 1,550 mm. Uit de produktie neemt men een steekproef van 10 stuks. Deze kogeltjes hebben de volgende diameters (in mm):

1,551	1,551	1,549	1,553	1,553
1,559	1,555	1,553	1,551	1,547

Gegeven is dat de populatie van diameters van de door de machine geproduceerde kogeltjes Normaal verdeeld is.
Onderzoek door middel van een statistische toets met onbetrouwbaarheidsdrempel 5% of de verwachting van de diameter van de geproduceerde kogeltjes afwijkt van de bovengenoemde nominale waarde.

101. Een sigarettenfabrikant maakt reclame met de bewering dat de verwachting van het nicotine-gehalte van de door hem vervaardigde sigaretten hoogstens 23 mg is. Bij onderzoek van 5 sigaretten blijken deze resp. 26, 28, 22, 23 en 29 mg nicotine te bevatten. Kunt u op grond van de steekproefuitkomsten de bewering van de fabrikant bestrijden aangenomen dat de sigaretten een aselecte steekproef uit de produktie vormen en dat in die produktie het nicotine-gehalte Normaal verdeeld is? Onbetrouwbaarheidsdrempel 0,10.

102. Stalen pijp wordt door een machine van beschermende coating voorzien. Een plek waar de coating te dun is aangebracht wordt een 'kaaltje' genoemd. Indien de machine normaal functioneert is het aantal kaaltjes per 100 m pijp een Poisson-verdeelde stochastische variabele met verwachting $\lambda = 0{,}9$.

a. Op gezette tijden wordt 1000 m pijp onderzocht. De machine wordt opnieuw afgeregeld indien het aantal kaaltjes 10 of meer bedraagt.
Bereken de kans dat men ten onrechte de machine opnieuw afregelt.
Bereken ook de kans dat men de instelling van de machine ongewijzigd laat, terwijl in werkelijkheid de verwachting van het aantal kaaltjes per 100 m gelijk is aan 1.

b. Toets H_0: $\lambda = 0{,}9$ tegen H_1: $\lambda > 0{,}9$. De nulhypothese wordt verworpen ten gunste van de alternatieve hypothese als het aantal kaaltjes op 1000 m pijp tenminste de waarde c heeft. Bepaal c zodanig dat de kans op een fout van de eerste soort hoogstens gelijk is aan 0,05.

c. Bepaal het onderscheidingsvermogen voor de toets onder b. voor het geval $\lambda = 1$.

d. Schets een grafiek van het onderscheidingsvermogen.

103. Van twee onafhankelijke steekproeven $\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_{10}$ en $\underline{y}_1,\underline{y}_2,\ldots,\underline{y}_{10}$ is bekend dat ze komen uit Normaal-verdeelde populaties met verwachting μ_x resp. μ_y, maar beide met variantie 12.

a. Toets de hypothese dat $\mu_x = \mu_y$ tegen het alternatief dat μ_x groter is dan μ_y als gegeven is dat $\overline{x} = 20$ en $\overline{y} = 18$. Neem $\alpha = 5\%$.

b. Iemand anders heeft op dezelfde waarnemingen dezelfde toets toegepast, alleen met een andere waarde van de onbetrouwbaarheidsdrempel α. Als het resultaat dat de ander krijgt tegengesteld is aan het uwe, wat kunt u dan zeggen over de waarde van α die de ander heeft gebruikt?

104. De steekproef $\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_8$ komt uit een $N(\mu_x;\ 32)$-verdeelde populatie. Onafhankelijk daarvan wordt een steekproef $\underline{y}_1,\underline{y}_2,\ldots,\underline{y}_{12}$ getrokken uit een $N(\mu_y;\ 27)$-verdeelde populatie. Dit alles werd gedaan omdat u hoopt d.m.v. een statistische toets bevestiging te krijgen van het vermoeden dat μ_x kleiner is dan μ_y. Men kiest $\alpha = 2{,}5\%$.

Een ijverige rekenaar bepaalt voor u de volgende grootheden:

$$\overline{x} = 23{,}6;\ \overline{y} = 32{,}2;\ s_x = 5{,}92;\ s_y = 5{,}01.$$

a. Bent u tevreden over de tijdspassering van de rekenaar?

b. Wat is het resultaat van de toets?

105. De steekproeven $\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_{12}$ en $\underline{y}_1,\underline{y}_2,\ldots,\underline{y}_{12}$ zijn onafhankelijk en komen beide uit Normale populaties waarvan aangenomen kan worden dat de varianties gelijk zijn.

Toets met onbetrouwbaarheidsdrempel $\alpha = 10\%$ H_0: $\mu_x = \mu_y$ tegen H_1: $\mu_x \neq \mu_y$.
Uit de steekproeven werd berekend: $\overline{x} = 15$; $\overline{y} = 17$; $s_x^2 = 5{,}6$; $s_y^2 = 5{,}1$.

106. Om te onderzoeken of de treinen die uit Delft vertrekken in de richting Den Haag systematisch meer vertraging hebben dan de treinen die in de richting Rotterdam vertrekken past men een statistische toets toe op twee onafhankelijke steekproeven uit de desbetreffende populaties. De resultaten (in secondes) waren:

richting Den Haag:	45;	85;	50;	30;	90;	
Rotterdam:	20;	35;	90;	75;	85;	40.

a. Neem aan dat de populaties Normaal zijn en dezelfde variantie hebben en voer een toets uit met $\alpha = 5\%$.

b. Hoe oordeelt u over de aanname dat de varianties gelijk zijn?

c. Hoe oordeelt u over de aanname dat de populaties Normaal zijn?

d. Wat voor waarde hecht u aan het resultaat van de toets?

107. Aan twee groepen, de een bestaande uit 50 jongens, de ander uit 50 meisjes, werd de taak gegeven om een aantal blokjes tot een bepaalde figuur te rangschikken. Voor iedere jongen en ieder meisje werd de tijd genoteerd die voor het vervullen van de taak nodig was. De resultaten zijn in de twee onderstaande frequentieverdelingen gegeven.

Tijd (in sec.)	40	41	42	43	44	45	46	47	48	49	50	51
meisjes	1	4	3	6	12	9	9	3	1	1	1	0
jongens	0	0	3	2	6	10	12	7	3	4	2	1

Aangenomen mag worden dat de twee steekproeven uit Normaal verdeelde populaties met gelijke, bekende variantie 4,29 afkomstig zijn.
Toets met $\alpha = 5\%$ de hypothese dat er geen verschil is tussen jongens en meisjes (althans op dit punt) tegen het alternatief dat er wel verschil is.

108. Een elektrotechnisch bedrijf vervaardigt weerstanden, waarvoor als kwaliteitseis geldt dat de standaardafwijking niet meer dan 1% van de nominale waarde mag bedragen. Er worden weerstanden gemaakt met een nominale waarde van 500 Ohm, van twee verschillende soorten materiaal, A en B te noemen. Metingen aan een aantal weerstanden van beide soorten, leverden de volgende uitkomsten op.

A		B	
495	496	511	511
496	504	497	493
499	495	501	507
504	500	505	509
503	495	493	
504	504	501	
495	503	499	
504	495	497	

Deze waarnemingen mogen worden beschouwd als onafhankelijke steekproeven uit Normaal verdeelde populaties.

a. Ga door middel van een statistische toets na of de weerstanden van elk der beide soorten aan de kwaliteitseis voldoen.
b. Ga daarna door toetsing na of er een verschil in kwaliteit is tussen de beide soorten weerstanden.

N.B. De onbetrouwbaarheidsdrempel voor de toetsen wordt op 0,10 gesteld.

109. In een proef is een aantal metingen verricht om na te gaan of synthetische vezels sterker krimpen naarmate ze op hogere temperatuur worden gebracht. Daartoe zijn 12 metingen verricht bij 120 °C en, onafhankelijk daarvan, 10 metingen bij 140 °C.
De resultaten (krimp in procenten van de oorspronkelijke vezellengte) zijn:

krimp bij	
120 °C	140 °C
3,45	3,72
3,64	4,03
3,57	3,60
3,62	4,01
3,56	3,40
3,44	3,76
3,60	3,54
3,52	3,96
3,56	3,91
3,49	3,67
3,53	
3,43	

De waarnemingen kunnen worden beschouwd als steekproeven uit Normaal verdeelde populaties. Men onderzoekt nu met de t-toets voor het verschil van verwachtingen of de krimp bij 140 °C hoger is dan bij 120 °C.
Bent u het met de keuze van deze toets eens?

110. Iemand gooit een munt tot en met het voor de eerste keer bovenkomen van kruis. Dit experiment wordt tweemaal uitgevoerd waarbij $\underline{x}_1$ en $\underline{x}_2$ maal de munt wordt geworpen.
a. Bepaal de meest aannemelijke schatter $\hat{\underline{p}}$ voor p de kans op kruis.
Toets de hypothese H_0: $p = \frac{1}{2}$ tegen de alternatieve hypothese H_1: $p < \frac{1}{2}$. Kies daarbij onbetrouwbaarheidsdrempel 0,05 en neem eenvoudigheidshalve als toetsingsgrootheid niet $\hat{\underline{p}}$ maar $\underline{x} = \underline{x}_1 + \underline{x}_2$.
b. Bepaal het kritieke gebied voor $\underline{x}$.
c. Wordt H_0 verworpen indien 2 en 6 worpen worden gedaan?

111. Het aantal malen per week dat de koffie-automaat defect raakt is gedurende de afgelopen twee jaar genoteerd omdat de kantinedienst het idee dat dit aantal storingen Poisson-verdeeld is wil onderzoeken. Het resultaat over deze 100 weken is als volgt:

aantal storingen	x_i:	1	2	3	4	5	6	7	8	9	10
frequentie	$f(x_i)$:	7	14	12	20	21	10	8	5	2	1

Is het idee gegrond? **N.B.** Gebruik als onbetrouwbaarheidsdrempel $\alpha = 5\%$.

112. Om te onderzoeken of de levensduur, in weken, van TL-buizen inderdaad exponentieel verdeeld is met verwachting 10, zoals de fabrikant beweert, worden

100 van die TL-buizen aangestoken. Om de drie weken gaat men kijken hoeveel er nog branden, met het volgende resultaat. Bij de inspectie aan het einde van week i treft men nog n_i brandende buizen aan.

i:	3	6	9	12	15	18	21	24
n_i:	65	47	33	23	14	8	5	3

Na 24 weken staakt men het waarnemen en gaat over tot het uitvoeren van een χ^2-toets met $\alpha = 10\%$. Ondersteunen de resultaten van die toets de bewering van de fabrikant?

113. De postzegelhandel Philador verkoopt postzegels in pakketten van 100 gram. Een klant heeft het idee dat in het door hem gekochte pakket erg veel beschadigde zegels voorkomen. Hij gaat zich in de winkel beklagen.
De baas van Philador beweert dat het percentage beschadigde zegels per pakket Normaal verdeeld is met verwachting 10 en variantie 16. De klant gelooft dat niet. Zij besluiten om een steekproef te nemen van 50 pakketten en die pakketten te onderzoeken. Het resultaat van de steekproef is als volgt

% beschadigd	0–5	5–10	10–15	15–20	20–25	> 25
aantal pakketten	6	10	14	9	8	3

Pas hierop een goodness-of-fit-toets toe met $\alpha = 0{,}10$ om uit te maken wie gelijk heeft.

114. Over het aantal verkeersongevallen met dodelijke afloop in twee steden A en B zijn de gegevens in de volgende tabel beschikbaar. Onderzoek of er in één van beide steden systematisch meer van dergelijke ongevallen optreden dan in de andere.

Jaar	Aantal ongelukken in A	Aantal ongelukken in B
1960	99	75
1961	87	83
1962	105	81
1963	97	101
1964	109	93
1965	105	97
1966	116	87
1967	110	113
1968	125	103
1969	117	93
1970	130	105
1971	137	107
1972	121	124
1973	126	112
1974	131	121
1975	145	130
1976	142	121

115. Twee soorten staaldraad worden in sterkte met elkaar vergeleken door van beide soorten aselect 6 stukken draad te nemen en ieder stuk aan een trekproef te onderwerpen. De draden worden in een machine belast totdat zij breken of totdat de voor deze machine maximaal mogelijke belasting van 1000 kg is bereikt. De resultaten zijn als volgt:

Belasting bij breken in kg draadsoort X	Y
790	750
750	820
630	510
840	980
700	960
niet gebroken	840

Toets met een onbetrouwbaarheid van 10% of beide soorten draad even sterk zijn.

116. Bij een onderzoek naar de slijtage van rubberbanden wordt onder andere de diepte van de groeven in het profiel van de band gemeten. De twee onderzoekers (X en Y) die zich hiermee bezighouden, krijgen nogal uiteenlopende resultaten. Men kiest nu op één band tien punten uit en laat zowel X als Y op die punten een meting doen. De resultaten zijn:

Punt	1	2	3	4	5	6	7	8	9	10
X	126	128	167	131	142	159	152	138	138	142
Y	125	120	163	118	129	152	150	136	140	136

Toets met onbetrouwbaarheidsdrempel 0,05 de hypothese dat er geen systematisch verschil is tussen X en Y tegen de alternatieve hypothese dat X systematisch hoger meet dan Y.

N.B. Men mag niet veronderstellen dat de waarnemingen afkomstig zijn uit Normaal verdeelde populaties.

117. De legerleiding overweegt de schoenen van militairen van een nieuw type zolen te voorzien. Dit type B is slijtvaster dan het oude type A, maar het is niet onmogelijk dat het aanleiding geeft tot meer voetklachten.
Een afdeling van 60 man die gewend is geweest type A te dragen, werd gedurende een zekere tijd en onder soortgelijke omstandigheden als vroeger, met type B uitgerust. Het bleek dat er 36 man meer last kreeg van voetklachten dan vroeger, bij 14 man was het minder geworden en bij 10 man was het aantal voetklachten onveranderd gebleven.
Onderzoek in de vorm van een toets met onbetrouwbaarheidsdrempel 0,05 of het aantal voetklachten systematisch is toegenomen.

118. Een bepaalde hoeveelheid van een voedingsmiddel wordt gesuspendeerd in steriel water en met vloeibare voedingsbodem gemengd. Na stolling hiervan wordt bij 20 °C gedurende drie dagen gebroed. Daarna wordt het aantal bacteriekolonies, dat zich op de voedingsbodem heeft ontwikkeld, geteld.
De proef is met vijf verschillende monsters uitgevoerd en hierbij werden de volgende aantallen kolonies geteld:

31 200 267 417 149

Daarna is de proef uitgevoerd met nieuwe monsters, waaraan een conserveringsmiddel is toegevoegd. Men vermoedt dat de toevoeging van het conserveringsmiddel het aantal bacteriekolonies doet afnemen. De resultaten met zes monsters zijn nu:

96 68 239 16 26 6

Onderzoek met betrouwbaarheidsdrempel 0,05 of dat vermoeden door de bovenstaande gegevens bevestigd wordt. Hierbij mag niet worden aangenomen dat de beide steekproeven uit Normaal verdeelde populaties afkomstig zijn.

119. Bij een onderzoek van een consumentenvereniging wil men de krimp vergelijken die door de wasmiddelen Druft en Remax wordt veroorzaakt. Daartoe heeft men 15 wollen lappen aselect uit een voorraad genomen en van iedere lap wordt de ene helft 10 keer met Druft en de andere helft 10 keer met Remax gewassen. De resultaten (in procenten krimp) zijn:

lap nr.	Druft	Remax
1	6	7
2	8	10
3	5	7
4	6	6
5	9	7
6	7	8
7	5	6
8	5	7
9	7	6
10	6	8
11	5	7
12	5	7
13	7	8
14	9	5
15	7	7

Ga door middel van een statistische toets met onbetrouwbaarheidsdrempel 0,05 na of Remax meer krimp veroorzaakt dat Druft. Hierbij mag men geen veronderstelling maken over de verdeling van de populaties waaruit beide reeksen waarnemingen steekproeven vormen.

120. Er worden titraties in duplo uitgevoerd (I en II) op monsters van een bepaalde vloeistof. Men vermoedt dat de bepalingen II die na de bepalingen I zijn uitgevoerd systematisch hoger liggen. Ga op grond van de volgende uitkomsten na of dit vermoeden juist is.

	Bepaling	
Monster	I	II
1	76,3	77,0
2	77,2	77,3
3	73,7	74,9
4	75,8	75,2
5	77,4	77,7
6	74,5	75,0
7	78,2	78,5
8	73,8	74,1
9	75,7	75,4
10	76,1	76,8

a. Aannemende dat de waarnemingen Normaal verdeeld zijn.

b. Aannemende dat de waarnemingen niet Normaal verdeeld zijn.

In beide gevallen dient een onbetrouwbaarheidsdrempel van 0,05 te worden aangehouden.

Antwoorden

1. $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$; ω_1: $b \leq 14$; ω_2: $14 < b \leq 16$; ω_3: $16 < b \leq 18$; ω_4: $b > 18$; waarbij b de belasting is die tot breuk leidt; $\mathcal{A}$ is de klasse van alle deelverzamelingen van Ω. Als kansmaat P ligt voor de hand:
$P(\omega_1) = 0{,}03$; $P(\omega_2) = 0{,}67$; $P(\omega_3) = 0{,}21$; $P(\omega_4) = 0{,}09$.

2. $\Omega = \{(i;j) \mid i = 1,2,\ldots,6;\ j = 1,2,\ldots,6\}$ met i resultaat steen één en j resultaat steen twee; $\{(1;5), (2;4), (3;3), (4;2), (5;1)\}$.

3. $TPVRL$, $TPVR\bar{L}$, $TPV\bar{R}L$, $T\bar{P}VRL$, $\bar{T}PVRL$, $T\bar{P}VR\bar{L}$, $T\bar{P}V\bar{R}L$, $\bar{T}PVR\bar{L}$, $\bar{T}PV\bar{R}L$; 32.

4. $P(A \cup B) = 0{,}88$; $P(A\bar{B}) = 0{,}24$; $P(ABC) = 0{,}48$.

5. *a.* 0,000.000.6; *b.* 0,000.017; *c.* 0,001.435; *d.* 0,0016.

6. 1/55. Dat hangt van u af.

7. 0,8154. **8.** $\frac{2}{3}$. **9.** *a.* 0,04; *b.* 0,1.

10. *a.* $\frac{1}{2}$; *b.* $\frac{1}{3}$; *c.* $\frac{1}{2}$. **11.** 0,16.

12. *a.* De kans dat een willekeurige chip, die uit de stapel als ondeugdelijk geclassificeerde chips wordt gepakt, goed blijkt te zijn; *b.* 0,5; *c.* 0,9942.

13. *a.* 0,2; *b.* 0,2.

14. P(i-de schroef zal gaan roesten) = 0,05 voor $i = 1,2,\ldots,1000$; 0,095.

15. Nee; $P\{\text{leeft} > 10 \text{ jaar} + t \text{ uur} \mid \text{leeft} > 10 \text{ jaar}\} = 10^{-t/100.000}$ voor $t > 0$.

16. *a.* 0,1071; *b.* 0,3893; *c.* 0,0917. **17.** $1 - (1-p)^{100}$.

18. *a.* Onafhankelijk als P(J) gelijk is aan 0; 0,5 of 1; *b.* Alleen onafhankelijk als P(J) gelijk is aan 0 of 1 (dus niet leuk).

19. *b.* Niet onderling onafhankelijk.

20. *a.* 0,4306; *b.* 0,1935.

21. Ja;
$$f(x) = \begin{cases} 2x, & 0 < x < \frac{1}{2}; \\ 6(1-x), & \frac{1}{2} < x < 1; \\ 0, & \text{elders.} \end{cases}$$

22. *a.*
$$F(x) = \begin{cases} 0, & x \leq 0; \\ 1 - e^{-x/10}, & x > 0; \end{cases}$$

b. 0,3834; *c*. 10.

23. $3C/2$.

24. *a*. $f(w) = 1/11, \quad -1 < w < 10$; $E(\underline{w}) = 4{,}5$; $\sigma^2(\underline{w}) = 10{,}08$;
$\qquad = 0, \quad$ elders;

b. $\underline{z} = 1{,}8\underline{w} + 32$; $f(z) = 5/99, \quad 30{,}2 < z < 50$; $E(\underline{z}) = 40{,}1$; $\sigma^2(\underline{z}) = 32{,}67$.
$\qquad = 0, \quad$ elders.

25.

k:	2	3	4	5	6	7	8	9	10	$k \neq 2,\ldots,10$
$p_x(k)$:	0,02	0,04	0,08	0,22	0,24	0,20	0,12	0,06	0,02	0

$E(\underline{x}) = 6{,}12$ en $\sigma^2(\underline{x}) = 2{,}7456$.

26. *a*. $p(1) = 0{,}02$; $p(2) = 0{,}06$; $p(3) = 0{,}10$; $p(4) = 0{,}14$; $p(5) = 0{,}18$; $p(\text{mis}) = 0{,}50$;
b. $p_x(0) = 0{,}50$; $p_x(1) = 0{,}18$; $p_x(2) = 0{,}14$; $p_x(3) = 0{,}10$; $p_x(4) = 0{,}06$; $p_x(5) = 0{,}02$; $E(\underline{x}) = 1{,}10$ en $\sigma^2(\underline{x}) = 1{,}89$;
c. $p_z(k) = \binom{27}{k}(0{,}02)^k(0{,}98)^{27-k}, k = 0,1,\ldots,27$; $E(\underline{z}) = 0{,}54$ en $\sigma^2(\underline{z}) = 0{,}5292$;
$\qquad = 0, \quad$ elders;

d. $p_w(k) = (0{,}98)^{k-1} \cdot 0{,}02, \quad k = 1,2,3,\ldots$; $E(\underline{w}) = 50$ en $\sigma^2(\underline{w}) = 2450$;
$\qquad = 0, \quad$ elders.

27. *a*. $F(x) = 0, \quad x < 1$;
$\qquad = 1 - (0{,}4)^k, \quad k \leq x < k+1$ met $k = 1,2,3,\ldots$;
$p(k) = (0{,}4)^{k-1} \cdot 0{,}6, \quad k = 1,2,3,\ldots$;
$\qquad = 0, \quad$ elders;

b. 0,064; *c*. 0,9959; *d*. 5/3.

28. *a*. $p_x(0) = 0{,}999$; $p_x(100) = 0{,}001$; $p_x(x) = 0$ voor $x \neq 0$ of 100;
$E(\underline{x}) = 10$ cent en $\sigma^2(\underline{x}) = 9{,}99$ gulden^2.
b. $p_z(-0{,}25) = 0{,}999$; $p_z(99{,}75) = 0{,}001$; $p_z(z) = 0$ voor $z \neq -0{,}25$ of $99{,}75$;
$E(\underline{z}) = 15$ cent verlies en $\sigma^2(\underline{z}) = 9{,}99$ gulden^2.

29. *a*. $p_x(k) = \binom{50}{k}(0{,}15)^k(0{,}85)^{50-k}, k = 0,\ldots,50$; $E(\underline{x}) = 7{,}5$ en $\sigma^2(\underline{x}) = 6{,}375$.
$\qquad = 0, \quad$ elders;

b. $p_z(k) = \binom{50}{k}(0{,}85)^k(0{,}15)^{50-k}, k = 0,1,\ldots,50$; $E(\underline{z}) = 42{,}5$ en $\sigma^2(\underline{z}) = 6{,}375$.
$\qquad = 0, \quad$ elders;

c. 0,2089; 0,9947; 0,7753.

30. *a.* $p(k) = (0{,}17)^{k-1} \cdot 0{,}83$, $k = 1,2,3,\ldots$; $E(\underline{x}) = 1{,}20$ en $\sigma^2(\underline{x}) = 0{,}2468$;
$\quad = 0$, elders;
b. 0,17; 0,0289.

31. *a.* 0,424; *b.* 0,424. **32.** *a.* 0,292; *b.* 0,527.

33. *a.* 0,140; *b.* 0,416; *c.* $c \geq 15$.

34. *a.* 0,3675; *b.* $p_1 = 0{,}3682$; $p_2 = 0{,}1841$; $p_3 = 0{,}0613$ (Binomiaal);
a. 0,3679; *b.* $p_1 = 0{,}3679$; $p_2 = 0{,}1839$; $p_3 = 0{,}0613$ (Poisson);
c. 0,0189 (Binomiaal); 0,0190 (Poisson).

35. *a.* Binomiaal $n = 1000$, $p = 0{,}001$ te benaderen met Poisson $\lambda = 1$.
b. $E(\underline{y}) = 1$ en $\sigma^2(\underline{y}) = 0{,}999$ (Binomiaal); $E(\underline{y}) = 1$ en $\sigma^2(\underline{y}) = 1$ (Poisson).

36. $E(\underline{x}) = 0$ en $\sigma^2(\underline{x}) = \frac{1}{3}$.

37. *a.* 0,449; *b.* $t = 0{,}256$;

38. *a.* 0,6915; *b.* 0,1815; *c.* $-1{,}564$.

39. *a.* 0,9873 en 0,0157; *b.* 0,9772; 0,9842; 0,0164.

40. *a.* 0,1782 en 0,1755; *b.* 0,1639 en 0,1602.

42. *a.* 4; *b.* $f_u(u) = 2u$, $0 < u < 1$; *c.* 15/64; *d.* 0,5.
$\quad = 0$, elders.

43. *a.* $\frac{\lambda^j}{j!} \frac{\mu^{k-j}}{(k-j)!} e^{-(\lambda+\mu)}$; *b.* $\frac{(\lambda+\mu)^k}{k!} e^{-(\lambda+\mu)}$, $k \in \mathbb{N}$, m.a.w. Poisson met parameter $(\lambda + \mu)$.

44. Binomiaal met parameters n+m en p.

45. *a.* $f_x(x) = 2e^{-2x}$, $x > 0$; $f_y(y) = 3e^{-3y}$, $y > 0$;
$\quad = 0$, elders; $\quad = 0$, elders;
b. ja; *c.* 0,4 en 0,6.

46. *a.* $f_x(x) = \frac{21}{4} x^2(1-x^4)$, $0 < x < 1$; $f_y(y) = 7y^6$, $0 < y < 1$;
$\quad = 0$, elders; $\quad = 0$, elders;
b. 0,009; *c.* nee.

47. *a.* negatief; *b.* $-5/12$ en $-0{,}2$.

48. *a.* niet onafhankelijk; *b.* $E(\underline{u}) = 1$, $\sigma^2(\underline{u}) = 2$.

49. *a.* $p(0;0) = 0{,}1$; $p(0;1) = 0{,}4$; $p(0;2) = 0{,}1$; $p(1;0) = 0{,}2$;
$p(1;1) = 0{,}2$; $p(x,y) = 0$ elders; *b.* 0,2 en 0,5;

c. $p(0) = 0{,}3$; $p(1) = 0{,}6$; $p(2) = 0{,}1$; $p(x) = 0$ voor $x \neq 0{,}1{,}2$;
d. 0,8; 0,36; 0,2; *e*. −0,408.

50. *a*. 0,707; *b*. 0.

51. *a*. $\frac{1}{3}\cdot(\frac{2}{3})^{k-3}$, $k = 3{,}4{,}\ldots$;
0, elders.

b. 5; *c*. $P(\underline{z} = k \mid \underline{x} \geq 3) = \frac{1}{3}\cdot(\frac{2}{3})^{k-1}$, $k = 1{,}2{,}\ldots$;
$= 0$, elders.

52. *a*. $2v/21$, $2 < v < 5$, *b*. 0, $v < 2$, *c*. 0,704.
0, elders; $(v^2-4)/21$, $2 \leq v \leq 5$,
1, $v > 5$;

53. *a*. $(4x+1)/3$, $0 < x < 1$, *b*. voor $y \in (0,1)$: $(x+y)/(y+0{,}5)$, $0 < x < 1$,
0, elders; 0, elders;
voor $y \notin (0,1)$: niet gedefinieerd.

54. $m \geq k$: $3 + (m-k)p$; $m < k$: $2m/(k-1)$.

55. Poisson met parameter λp.

56. *a*. Uniform op $(-\sqrt{1-y^2}\,;\sqrt{1-y^2})$ voor $|y| < 1$; Niet gedefinieerd voor $|y| \geq 1$;
b. $\rho(\underline{x},\underline{y}) = 0$; *c*. $\underline{x}$ en $\underline{y}$ niet onafhankelijk.

57. *a*. 0,9110; *b*. 0,2702; *c*. 0,7154. **58.** 0,84.

59. *a*. $1/\lambda^2$; *b*. 0,7965; *c*. 1,6903.

60. *a*. 0,2638; *b*. 0,4400; *c*. 0,4332; *d*. 0,6431.

61. Uniform op vierkant: (0,0); (1,1); (2,0); (1,−1).

62. $\lambda^2 z\cdot e^{-\lambda z}$, $z > 0$; 0, elders.

63. *a*. Ja; *b*. Binormaal met parameters (0; 0; 1; 1; 0).

64. *a*. $u/36$, $v = 1{,}2{,}3$ en $u = v, 2v, 3v$; 0, elders.
b. $v/6$, $v = 1{,}2{,}3$; 0, elders.

65. simultane dichtheid van $\underline{x}$, $\underline{y}$ en $\underline{z}$:

$$\frac{1}{2\pi\sqrt{2\pi}}e^{-x^2/2}\cdot x^2\cdot \sin z, \quad x > 0, 0 \leq y < 2\pi, 0 \leq z < \pi.$$
0, elders.

66. *a*. Negatief binomiaal met $p = \frac{1}{4}$ en $k = 3$; 12;
b. $\binom{k-5}{2}(\frac{1}{4})^3(\frac{3}{4})^{k-7}$, $k = 7{,}8{,}\ldots$; 0, elders; 16.

67. *a.* $P(\underline{x} = 2k) = (p^2+q^2)(pq)^{k-1}$; $P(\underline{x} = 2k+1) = (pq)^k$, $k = 1,2,\ldots$;
$P(\underline{x} = x) = 0$, $x \neq 2,3,\ldots$; *b.* $(2+pq)/(1-pq)$; *c.* $pq/(1-pq)$.

70. *a.* Exp(3λ); *b.* dichtheid: $3\lambda(1-e^{-\lambda y})^2 e^{-\lambda y}$, $y > 0$;
0, elders.

71. $\overline{x} = 499{,}4$; $s_x^2 = 5080{,}40$; $s_x = 71{,}28$.

72. $s_x = 2{,}03$. **73.** *a.* $\chi^2(2)$; *b.* 4,155. **74.** $a = 8$; $b = 6{,}93$.

75. *a.* $c = 0{,}25$; *b.* σ^4; *c.* $E(\underline{s}^2) = \sigma^2$ en $\sigma^2(\underline{s}^2) = 2\sigma^4/3$, dus beide schatters zijn zuiver maar $\underline{s}^2$ is beter omdat hij een kleinere variantie heeft.

76. *a.* $m < n$: $\underline{s}_m$ en $\underline{s}_n$ binormaal met parameters $(0; 0; m; n; \sqrt{m/n})$;
gegeven $\underline{s}_n = t$ is $\underline{s}_m \sim N(tm/n; m(1-m/n))$;
$m > n$: $\underline{s}_m$ en $\underline{s}_n$ binormaal met parameters $(0; 0; m; n; \sqrt{n/m})$;
gegeven $\underline{s}_n = t$ is $\underline{s}_m \sim N(t; m-n)$;
b. f_k is de $\chi^2(k)$-dichtheid.
$m < n$: $f(u,v) = f_m(u) \cdot f_{n-m}(v-u)$, $u > 0$, $v > u$;
$f(u \mid \underline{v} = v) = f_m(u) \cdot f_{n-m}(v-u)/f_n(v)$, $0 < u < v$ mits $v > 0$;
niet gedefinieerd voor $v < 0$;
$m > n$: $f(u,v) = f_n(v) \cdot f_{m-n}(u-v)$, $v > 0$, $u > v$;
$f(u \mid \underline{v} = v) = f_{m-n}(u-v)$, $u > v$ mits $v > 0$; niet gedef. voor $v < 0$.

77. *a.* 1,345; *b.* 1,761; *c.* 2,624.

78. *a.* 0,246; *b.* 3,22; *c.* 0,90.

79. $E(\underline{z}) = n$ en $\sigma^2(\underline{z}) = 2n$.

80. *a.* 0,7888; *b.* 24,267; *c.* 24,128.

81. $\underline{t}_1$ is een zuivere schatter; $\underline{t}_2$ is niet-zuivere schatter; $\underline{t}_3$ is geen schatter dus zeker geen zuivere schatter.

82. *a.* $\hat{\underline{r}} = \underline{b}/(n-\underline{b})$ als $\underline{b}$ het aantal zwarte ballen is dat wordt getrokken;
b. $\hat{\underline{r}} = 1/(\overline{\underline{x}}-1)$.

83. *a.* $1/\theta^2$; *b.* $E(\underline{x}) = 2\theta$ en $\sigma^2(\underline{x}) = 2\theta^2$; *c.* $\hat{\underline{\theta}} = \underline{x}/2$; *d.* $\hat{\underline{\theta}} = \overline{\underline{x}}/2$;
e. $\overline{\underline{x}}^2/2$; *f.* $\underline{m}_2/3$; *g.* $E(\overline{\underline{x}}^2/2) = (2+1/n)\theta^2$, deze schatter is niet zuiver;
$E(\underline{m}_2/3) = 2\theta^2$, deze schatter is zuiver.

84. *a.* $E(\overline{\underline{x}}) = 1/\lambda$; $\sigma^2(\overline{\underline{x}}) = 1/(2\lambda^2) = MSE(\overline{\underline{x}})$ want $\overline{\underline{x}}$ is een zuivere schatter;
b. $E(\overline{\underline{x}}_g) = \pi/(4\lambda)$; $\sigma^2(\overline{\underline{x}}_g) = (16-\pi^2)/(16\lambda^2)$; $MSE(\overline{\underline{x}}_g) = (4-\pi)/(2\lambda^2)$;
c. $4\overline{\underline{x}}_g/\pi$; $\sigma^2(4\overline{\underline{x}}_g/\pi) = (16-\pi^2)/(\pi\lambda)^2 = 0{,}62/\lambda^2$;
d. Aan $\overline{\underline{x}}$ omdat die een kleinere variantie heeft, terwijl beide schatters zuiver zijn.

85. *a.* $\underline{t}_1$ en $\underline{t}_3$; *b.* $\underline{t}_4 = (n+1)\underline{t}_2/n$ is zuiver; *c.* $\sigma^2(\underline{t}_1) = \theta^2/(3n)$; $\sigma^2(\underline{t}_3) = n\theta^2/(n+2)$; $\sigma^2(\underline{t}_4) = \theta^2/(2n+n^2)$; *d.* $\underline{t}_4$ (kleinste variantie).

86. *a.* (77,3; 85,1); *b.* $n \geq 25$; *c.* (3,65; 5,75); *d.* (2,50; 9,04); *e.* (0,68; 5,29); *f.* (0,43; 2,21).

87. (201; 639).

88. (0,58; 16,7) of (0,06; 1,72) dus de varianties kunnen best gelijk zijn.

89. (14,7; 53,3) met 90% betrouwbaarheid; Nee, dan kiest men een linkszijdig begrensd betrouwbaarheidsinterval.

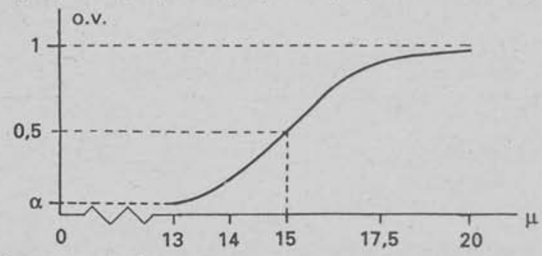
90. *a.* $\{1-(\frac{1}{2})^{n-1}\}\cdot 100\%$; *b.* 50%; *c.* $(\overline{x}-0{,}477;\ \overline{x}+0{,}477)$; *d.* $(n-1)/(n+1)$; 1,13; 0,955.

91. Kritiek gebied voor $\overline{x}$: $(6{,}32;\ \infty)$; H_0 wordt niet verworpen.

92. Kritiek gebied voor $\overline{x}$ als $s^2 = 22$:
$\alpha = 10\%$: $(-\infty;\ 7{,}9) \cup (12{,}1;\ \infty)$; H_0 wordt verworpen;
$\alpha = 1\%$: $(-\infty;\ 6{,}54) \cup (13{,}46;\ \infty)$; H_0 wordt niet verworpen.

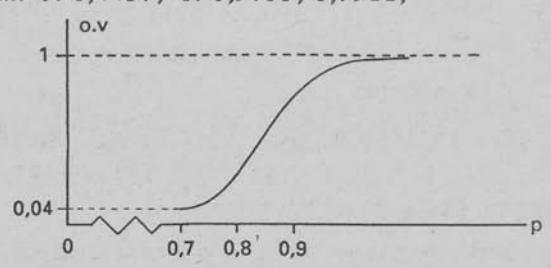
93. *a.* Kritiek gebied voor s^2: $(9{,}4;\ \infty)$; H_0 wordt niet verworpen;
b. Kritiek gebied voor $\Sigma(x_i-6)^2$: $(91{,}5;\ \infty)$; H_0 wordt verworpen.

94. *a.* 0,9962; 0,9092; 0,5000; 0,0004; 0;
b.



95. H_0: $p = \frac{1}{2}$ H_1: $p \neq \frac{1}{2}$; Kritiek gebied: $\{0,1,\ldots,39\} \cup \{61,62,\ldots,100\}$.

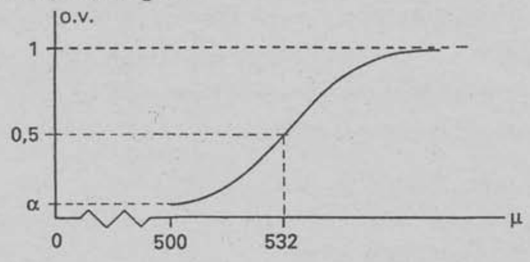
96. *a.* H_0: $p = 0{,}7$ H_1: $p > 0{,}7$; Kritiek gebied: $\{41,42,\ldots,50\}$; H_0 wordt niet verworpen, m.a.w. de conclusie dat B beter is dan A wordt niet gestaafd door dit resultaat. *b.* 0,4437; *c.* 0,9755; 0,7911;
d.



N.B. Voor $p \downarrow 0{,}7$ wordt het o.v. hier niet gelijk aan α (= 0,05) maar aan 0,04.

97. H_0: $p = \frac{1}{2}$ H_1: $p > \frac{1}{2}$; Kritiek gebied: $\{9,10\}$; H_0 wordt niet verworpen, er is geen reden om aan te nemen dat ze niet raadt;
N.B. Als H_1: $p \neq \frac{1}{2}$ als alternatieve hypothese wordt genomen volgt dezelfde conclusie.

98. *a.* H_0: $\mu = 500$ H_1: $\mu > 500$; Kritiek gebied als $s^2 = 5080{,}40$: $(532{,}4; \infty)$; H_0 wordt niet verworpen.
b.

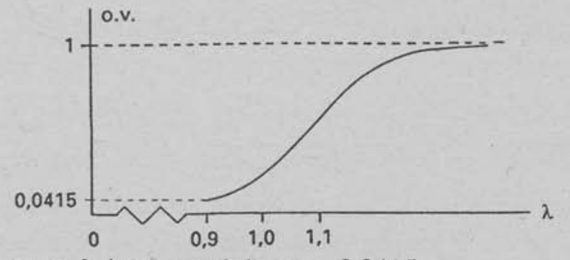


99. *a.* 0,0062; *b.* 153; *c.* $(1530{,}9; \infty)$; *d.* $(1514{,}25; \infty)$ en $n \geq 470$.

100. Vervang de waarnemingen x_i door $y_i = 1000x_i - 1550$.
H_0: $\mu_y = 0$ H_1: $\mu_y \neq 0$;
Kritiek gebied voor $\overline{y}$ met $s_y^2 = 10{,}84$: $(-\infty; -2{,}35) \cup (2{,}35; \infty)$; H_0 wordt niet verworpen.

101. H_0: $\mu = 23$ H_1: $\mu > 23$; Kritiek gebied voor $\overline{x}$ met $s^2 = 9{,}3$: $(25{,}09; \infty)$; H_0 wordt verworpen.

102. *a.* 0,4126 en 0,4579; *b.* $c = 15$; *c.* 0,0835;
d.



N.B. Voor $\lambda \downarrow 0{,}9$ wordt het o.v. 0,0415.

103. *a.* H_0: $\mu_x - \mu_y = 0$ H_1: $\mu_x - \mu_y > 0$;
Kritiek gebied voor $\overline{x} - \overline{y}$: $(2{,}55; \infty)$; H_0 wordt niet verworpen.
b. $\alpha > 10\%$.

104. *a.* Nee, de berekening van s_x en s_y is overbodig.

b. H_0: $\mu_x - \mu_y = 0$ H_1: $\mu_x - \mu_y < 0$; Kritiek gebied voor $\overline{x} - \overline{y}$: $(-\infty; -4{,}9)$; H_0 wordt verworpen.

105. Kritiek gebied voor $\overline{x} - \overline{y}$: $(-\infty; -1{,}62) \cup (1{,}62; \infty)$; H_0 wordt verworpen.

106. *a.* Naar Den Haag: $\underline{g}_i$; naar Rotterdam $\underline{r}_j$.
H_0: $\mu_g - \mu_r = 0$ H_1: $\mu_g - \mu_r > 0$; Kritiek gebied voor $\overline{g} - \overline{r}$: $(31{,}15; \infty)$;
H_0 wordt niet verworpen; er is geen reden om een systematisch verschil in de vertraging per richting te veronderstellen;
b. H_0: $\sigma_g^2 = \sigma_r^2$ H_1: $\sigma_g^2 \neq \sigma_r^2$; Kritiek gebied voor s_r^2/s_g^2 met $\alpha = 10\%$: $(0; a) \cup (6{,}26; \infty)$; met $0 < a < 1$. H_0 wordt niet verworpen. De aanname is niet onredelijk;
c. De aanname is onredelijk want veel treinen vertrekken op tijd; de populatie heeft dus geen continue verdeling; er is een discontinuïteit in 0;
d. Geen enkele.

107. H_0: $\mu_j = \mu_m$ H_1: $\mu_j \neq \mu_m$; Kritiek gebied voor $\overline{m} - \overline{j}$: $(-\infty; -0{,}812) \cup (0{,}812; \infty)$; H_0, geen verschil, wordt verworpen.

108. N.B. $\mu = 500$ dus niet met s^2 werken.
a. H_0: $\sigma_A^2 = 25$ H_1: $\sigma_A^2 > 25$; Kritiek gebied voor tg: $(23{,}5; \infty)$; tg = 10,2; dus H_0 niet verwerpen.
H_0: $\sigma_B^2 = 25$ H_1: $\sigma_B^2 > 25$; Kritiek gebied voor tg: $(18{,}5; \infty)$; tg = 20,6; dus H_0 wordt verworpen.
b. H_0: $\sigma_A^2 = \sigma_B^2$ H_1: $\sigma_A^2 \neq \sigma_B^2$; Kritiek gebied voor tg: $(0; a) \cup (2{,}42; \infty)$ met $0 < a < 1$; tg = 2,69; dus H_0 wordt verworpen.

109. De t-toets mag alleen worden toegepast als $\sigma_{120}^2 = \sigma_{140}^2$;
H_0: $\sigma_{120}^2 = \sigma_{140}^2$ H_1: $\sigma_{120}^2 \neq \sigma_{140}^2$: zelfs met $\alpha = 1\%$ wordt H_0 verworpen want dan is het kritieke gebied voor s_{140}^2/s_{120}^2: $(5{,}54; \infty)$ en de realisatie is 9,22; De toets mag niet worden toegepast.

110. *a.* $\widehat{p} = 2/(\underline{x}_1 + \underline{x}_2)$; *b.* Als $p = \frac{1}{2}$ volgt $P(\underline{x} = k) = (k-1)(\frac{1}{2})^k$, $k = 2,3,\ldots$; Kritiek gebied voor x: $\{9,10,\ldots\}$; *c.* Nee.

111. $\widehat{\lambda} = \overline{x} = 4{,}4$; 8 klassen: $\leq 1, 2, \ldots, 7, \geq 8$; 1 parameter geschat: 6 vrijheidsgraden; Kritiek gebied bij $\alpha = 5\%$: $(12{,}6; \infty)$, realisatie 3,61; de steekproefresultaten geven geen aanleiding om het idee ongegrond te verklaren.

112. 8 klassen: [0; 3), [3; 6) … [18; 24), [24; ∞); 7 vrijheidsgraden; Kritiek gebied bij $\alpha = 10\%$: $(12{,}0; \infty)$; realisatie 8,35; de resultaten ondersteunen de bewering van de fabrikant.

113. 4 klassen; 3 vrijheidsgraden; Kritiek gebied bij $\alpha = 10\%$: $(6{,}25; \infty)$; realisatie 47,6; het idee van de klant wordt door dit resultaat ondersteund.

114. Tekentoets, tweezijdig; H_0: $p = \frac{1}{2}$ H_1: $p \neq \frac{1}{2}$; $n = 17$, $\underline{x}$ = aantal malen $A_i > B_i$; $x = 14$; Kritiek gebied als $\alpha = 1\%$: $\{0,1,2\} \cup \{15,16,17\}$ dus H_0 wordt niet verworpen; Kritiek gebied als $\alpha = 2\%$: $\{0,1,2,3\} \cup \{14,15,16,17\}$ dus H_0 wordt wel verworpen, ook voor alle waarden van α die groter zijn dan 2% (zelfs $\geq 1{,}28\%$).

115. Wilcoxon, tweezijdig; H_0: ligging X = ligging Y H_1: ligging X $\neq$ ligging Y; $n = m = 6$, $\alpha = 10\%$ dus kritiek gebied is $\{0,1,\ldots,14\} \cup \{58,59,\ldots,72\}$; realisatie, met $w_{ij} = 2$ als $x_i > y_j$: 30; H_0 wordt niet verworpen.

116. Tekentoets, eenzijdig; $\underline{z}_i = \underline{x}_i - \underline{y}_i$; H_0: $p = P(\underline{z}_i > 0) = \frac{1}{2}$ H_1: $p > \frac{1}{2}$; $n = 10$ en $\alpha = 5\%$; Kritiek gebied: $\{9,10\}$; realisatie 9; H_0 wordt verworpen.

117. Tekentoets, eenzijdig; H_0: aantal klachten gelijk H_1: aantal klachten neemt toe; $n = 60 - 10 = 50$; kritiek gebied bij $\alpha = 5\%$: $\{32,33,\ldots,50\}$; realisatie 36; H_0 wordt verworpen dus type B veroorzaakt meer voetklachten.

118. Wilcoxon, eenzijdig; $\underline{z}$ is zonder, $\underline{m}$ is met; H_0: ligging $\underline{z}$ = ligging $\underline{m}$ H_1: $\underline{z}$ "ligt rechts van" $\underline{m}$; $n = 5$, $m = 6$ en $\alpha = 5\%$; met $w_{ij} = 2$ als $z_i > m_j$ wordt het kritieke gebied voor w: $\{50,51,\ldots,60\}$; realisatie 50, H_0 wordt verworpen en het vermoeden is bevestigd.

119. Tekentoets, eenzijdig; $\underline{z}_i = \underline{R}_i - \underline{D}_i$; H_0: $p = P(\underline{z}_i > 0) = \frac{1}{2}$ H_1: $p > \frac{1}{2}$; $\alpha = 5\%$, na het verwijderen van het vierde en vijftiende waarnemingspaar (gelijk resultaat) wordt $n = 13$; Kritiek gebied: $\{10,11,12,13\}$; realisatie 10, H_0 wordt verworpen.

120. *a.* Dit is één steekproef van verschillen; zij $\underline{v}_i = \underline{II}_i - \underline{I}_i$, m.a.w. $v_1 = 0{,}7$, enz.; H_0: $\mu_v = 0$ H_1: $\mu_v > 0$; $\alpha = 5\%$ en $n = 10$; kritiek gebied voor $\overline{v}$ als $s_v^2 = 0{,}264$: $(0{,}30; \infty)$; realisatie $\overline{v} = 0{,}32$, H_0 wordt verworpen;
b. Tekentoets, eenzijdig; H_0: $p = P(\underline{v}_i > 0) = \frac{1}{2}$ H_1: $p > \frac{1}{2}$; $\alpha = 5\%$ en $n = 10$; kritiek gebied: $\{9,10\}$; realisatie 8, H_0 kan niet worden verworpen.

Trefwoordenlijst

Appendix: formules en tabellen

Klein repertorium

Een verzamelingsoperatie

De Morgan regels

$\overline{(A \cup B)} = \bar{A} \cap \bar{B}$,

$\overline{(A \cap B)} = \bar{A} \cup \bar{B}$.

Axioma's

a. $P(A) \geq 0$ voor alle A.

b. $P(\Omega) = 1$.

c. $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$, indien $A_iA_j = \Phi$ voor alle $i \neq j$.

Enkele elementaire stellingen

1. *Verenigingsformules.* De kans dat minstens één van de gebeurtenissen A en B optreedt is

$$P(A \cup B) = P(A) + P(B) - P(AB)$$

voor willekeurige gebeurtenissen A en B.
Voor iedere drie willekeurige gebeurtenissen geldt dat

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC).$$

2. *Ongelijkheid van Boole.* Voor de kans dat minstens één van de gebeurtenissen $A_1, A_2, \ldots$, optreedt geldt dat

$$P(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i).$$

Combinatoriek

Het aantal verschillende steekproeven zonder teruglegging ter grootte r uit een populatie van n verschillende elementen is

$$(n)_r = n(n-1) \ldots (n-r+1) = \frac{n!}{(n-r)!}.$$

Het aantal deelverzamelingen ter grootte r uit een populatie van n verschillende elementen is

$$\binom{n}{r} = \frac{(n)_r}{r!} = \frac{n(n-1)\ldots(n-r+1)}{r!} = \frac{n!}{(n-r)!r!}, \quad r \leq n.$$

Voorwaardelijke kans

Algemene produktregel. Voor ieder n-tal gebeurtenissen $A_1, A_2, \ldots, A_n$ met

$$P(\bigcap_{i=1}^{n-1} A_i) > 0,$$

geldt
$$P(\bigcap_{i=1}^{n} A_i) = P(A_1)P(A_2 \mid A_1)\ldots P(A_n \mid \bigcap_{i=1}^{n-1} A_i).$$

Theorema van de totale waarschijnlijkheid. Zij $A_1, A_2, \ldots$ een volledig stelsel elkaar uitsluitende gebeurtenissen (dat wil zeggen $A_iA_j = \Phi$ voor $i \neq j$, $\bigcup_{i=1}^{\infty} A_i = \Omega$; ook wel partitie van Ω genaamd) met $P(A_i) > 0$, $i = 1,2,\ldots$, dan is voor iedere gebeurtenis A:

$$P(A) = P(A \mid A_1)P(A_1) + P(A \mid A_2)P(A_2) + \ldots .$$

Theorema van Bayes. Zij $A_1, A_2, \ldots$ een partitie van Ω met $P(A_i) > 0$, $i = 1,2,\ldots$ en is A een gebeurtenis met positieve kans van optreden, dan geldt

$$P(A_k|A) = \frac{P(A|A_k)P(A_k)}{\sum_{i=1}^{\infty} P(A|A_i)P(A_i)}.$$

Stochastische variabelen

Verwachting. Laat $\underline{x}$ en $\underline{y}$ twee stochastische variabelen zijn met gezamenlijke kansfunctie of dichtheid $f_{\underline{x},\underline{y}}$, dan geldt:

a. Als $\underline{x} = a$ constant is, dan is ook $E\underline{x} = a$.
b. Als a en b constanten zijn, dan is $E(a\underline{x} + b\underline{y}) = aE\underline{x} + bE\underline{y}$.
c. Als $P(\underline{x} \leq \underline{y}) = 1$, dan is $E\underline{x} \leq E\underline{y}$.
d. Zijn $\underline{x}$ en $\underline{y}$ onafhankelijke stochastische variabelen en φ en ψ twee reële functies, dan is $E(\varphi(\underline{x})\psi(\underline{y})) = E\varphi(\underline{x})E\psi(\underline{y})$.

Variantie, covariantie en correlatiecoëfficiënt

Zijn a, b en c reële getallen en $\underline{x}$, $\underline{y}$ en $\underline{z}$ stochastische variabelen, dan geldt

a. $\text{var}(\underline{x}) = E\underline{x}^2 - (E\underline{x})^2$.

b. $\text{var}(a\underline{x} + b) = a^2\text{var}(\underline{x})$.

c. $\text{var}(\underline{x}) \geq 0$.

d. $\text{var}(\underline{x}) = 0 \Leftrightarrow P(\underline{x} = E\underline{x}) = 1$.

e. $\text{cov}(\underline{x},\underline{y}) = E\underline{x}\underline{y} - (E\underline{x})(E\underline{y})$.

f. Als $\underline{x}$ en $\underline{y}$ onafhankelijk zijn is $\text{cov}(\underline{x},\underline{y}) = 0$.

g. $\text{var}(a\underline{x} + b\underline{y}) = a^2\text{var}(\underline{x}) + b^2\text{var}(\underline{y}) + 2ab\,\text{cov}(\underline{x},\underline{y})$.

h. $\text{cov}(a\underline{x}+b\underline{y},c\underline{z}) = ac\,\text{cov}(\underline{x},\underline{z}) + bc\,\text{cov}(\underline{y},\underline{z})$.

i. $|\text{cov}\,(\underline{x},\underline{y})| \leq \sqrt{\text{var}(\underline{x})\cdot\text{var}(\underline{y})}$.

j. $|\rho_{x,y}| \leq 1$, waar $\rho_{x,y} = \dfrac{\text{cov}(\underline{x},y)}{\sqrt{\text{var}(\underline{x})\cdot\text{var}(\underline{y})}}$.

k. In beide ongelijkheden geldt het gelijkteken dan en slechts dan als er een lineaire relatie tussen $\underline{x}$ en $\underline{y}$ bestaat, d.w.z. als er reële getallen a en b (niet beide 0) en c bestaan, zodat $P(a\underline{x} + b\underline{y} + c = 0) = 1$.

Ongelijkheid van Chebyshev. Voor iedere $t > 0$ en iedere stochastische variabele $\underline{x}$ waarvan $\text{var}(\underline{x})$ bestaat is

$$P(|\underline{x} - \mu| \geq t) \leq t^{-2}\text{var}(\underline{x}), \text{ waar } \mu = E\underline{x}.$$

Relaties tussen verdelingen

1. De verdeling van de som van n onafhankelijke identiek alternatief verdeelde stochastische variabelen is binomiaal.
2. De verdeling van de som van n onafhankelijke identiek Pascal (geometrisch) verdeelde stochastische variabelen is negatief binomiaal.
3. De som van n onafhankelijke identiek exponentieel verdeelde stochastische variabelen heeft een Erlang verdeling.
4. Het quotiënt van twee onafhankelijke normaal verdeelde stochastische variabelen, beide met verwachting nul, heeft een Cauchy verdeling.
5. Het verschil van twee onafhankelijk identiek exponentieel verdeelde stochastische variabelen heeft een Laplace verdeling.

6. De verdeling van de som van de kwadraten van n onafhankelijke N(0; 1) verdeelde stochastische variabelen is $\chi^2(n)$.
7. Zij $\underline{u}$ een standaard normaal verdeelde stochastische variabele en $\underline{v}$ een $\chi^2(k)$-verdeelde variabele, terwijl $\underline{u}$ en $\underline{v}$ onafhankelijk zijn, dan heeft $\underline{t} = \underline{u}/\sqrt{\underline{v}/k}$ een t-verdeling met k vrijheidsgraden.
8. Zijn $\underline{v}$ en $\underline{w}$ twee onafhankelijke stochastische variabelen met respectievelijk $\chi^2(k_1)$- en $\chi^2(k_2)$-verdeling, dan heeft
$$\underline{F} = \frac{\underline{v}/k_1}{\underline{w}/k_2}$$
een F-verdeling met k_1, k_2 vrijheidsgraden.
9. *Centrale limietstelling.* $\underline{x}_1, \underline{x}_2, \ldots$ zijn onafhankelijke identiek verdeelde stochastische variabelen met verwachting μ en variantie σ^2; $\underline{S}_n = \sum_{k=1}^{n} \underline{x}_k$, dan geldt, voor alle a en b $\in \mathbb{R}$:
$$\lim_{n\to\infty} P[a < \frac{\underline{S}_n - n\mu}{\sigma\sqrt{n}} \leq b] = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}u^2} du.$$

Steekproeven uit normale populaties

Laat $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_n$ een aselecte steekproef zijn uit een populatie met verdeling $N(\mu; \sigma^2)$. Zij
$$\underline{\bar{x}} = \frac{1}{n}\sum_{i=1}^{n} \underline{x}_i \quad \text{en} \quad \underline{s}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\underline{x}_i - \underline{\bar{x}})^2 = \frac{1}{n-1} (\sum_{i=1}^{n} \underline{x}_i^2 - n\underline{\bar{x}}^2),$$
dan zijn $\underline{\bar{x}}$ en $\underline{s}^2$ onafhankelijke stochastische variabelen. De verdeling van $\underline{\bar{x}}$ is $N(\mu; \sigma^2/n)$, die van $(n-1)\underline{s}^2/\sigma^2$ is $\chi^2(n-1)$.

Toets van Wilcoxon

Toetsingsgrootheid is $\underline{W} = \sum_i \sum_j \underline{w}_{ij}$, waarin de 'scores' w_{ij} gedefinieerd zijn door $w_{ij} = 2$ als $x_i > y_j$, $w_{ij} = 1$ als $x_i = y_j$ en $w_{ij} = 0$ als $x_i < y_j$. Hierbij zijn $x_1, x_2, \ldots, x_n$ de waarnemingen uit de eerste steekproef, $y_1, y_2, \ldots, y_m$ de waarnemingen uit de tweede steekproef.

Overzicht verdelingen

restricties parameters	naam	parameters	waardenbereik	kansfunctie of dichtheid	verwachting	variantie
$0 < p < 1$	Alternatief	p	$\{0,1\}$	$p^k(1-p)^{1-k}$	p	$p(1-p)$
$0 < p < 1, n \in \mathbb{N}^+$	Binomiaal	n, p	$\{0,1,\ldots,n\}$	$\binom{n}{k}p^k(1-p)^{n-k}$	np	$np(1-p)$
$0 < p < 1$	Geometrisch	p	$\{1,2,\ldots\}$	$p(1-p)^{k-1}$	p^{-1}	$(1-p)/p^2$
$0 < p < 1, r \in \mathbb{N}^+$	Negatief binomiaal	r, p	$\{r,r+1,\ldots\}$	$\binom{k-1}{k-r}p^r(1-p)^{k-r}$	rp^{-1}	$r(1-p)p^{-2}$
$\mu > 0$	Poisson	μ	$\{0,1,2,\ldots\}$	$(\mu^k/k!)e^{-\mu}$	μ	μ
$a < b$	Uniform	a, b	(a,b)	$(b-a)^{-1}$	$\frac{1}{2}(a+b)$	$\frac{1}{12}(b-a)^2$
$\lambda > 0$	Exponentieel	λ	$(0,\infty)$	$\lambda e^{-\lambda x}$	λ^{-1}	λ^{-2}
$\lambda > 0, n \in \mathbb{N}^+$	Erlang	n, λ	$(0,\infty)$	$\lambda^n x^{n-1}e^{-\lambda x}/(n-1)!$	$n\lambda^{-1}$	$n\lambda^{-2}$
$\alpha > 0, \beta > 0$	Gamma	α, β	$(0,\infty)$	$x^{\alpha-1}e^{-x/\beta}/(\Gamma(\alpha)\beta^\alpha)$	$\alpha\beta$	$\alpha\beta^2$
$\sigma > 0$	Normaal	μ, σ	$(-\infty,\infty)$	$\frac{1}{\sigma\sqrt{2\pi}}e^{-(x-\mu)^2/(2\sigma^2)}$	μ	σ^2
$a > 0$	Cauchy	a	$(-\infty,\infty)$	$\frac{a}{\pi(a^2+x^2)}$	—	—
$\lambda > 0$	Laplace	λ	$(-\infty,\infty)$	$\frac{1}{2}\lambda e^{-\lambda\|x\|}$	0	$2\lambda^{-2}$
$k \in \mathbb{N}^+$	Chi-kwadraat	k	$(0,\infty)$	$\frac{z^{\frac{1}{2}k-1}e^{-z/2}}{\Gamma(\frac{k}{2})2^{k/2}}$	k	$2k$

Cumulatieve binomiale verdeling

$$P(\underline{x} \leq x) = \sum_{k=0}^{x} \binom{n}{k} p^k (1-p)^{n-k}$$

		p											
n	x	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,50	1/6	1/3
2	0	0,9025	0,8100	0,7725	0,6400	0,5625	0,4900	0,4225	0,3600	0,3025	0,2500	0,6944	0,4444
	1	0,9975	0,9900	0,9775	0,9600	0,9375	0,9100	0,8775	0,8400	0,7975	0,7500	0,9722	0,8889
	2	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
3	0	0,8574	0,7290	0,6141	0,5120	0,4219	0,3430	0,2746	0,2160	0,1664	0,1250	0,5787	0,2963
	1	0,9928	0,9720	0,9392	0,8960	0,8438	0,7840	0,7182	0,6480	0,5748	0,5000	0,9259	0,7407
	2	0,9999	0,9990	0,9966	0,9920	0,9844	0,9730	0,9571	0,9360	0,9089	0,8750	0,9954	0,9630
	3	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
4	0	0,8145	0,6561	0,5220	0,4096	0,3164	0,2401	0,1785	0,1296	0,0915	0,0625	0,4822	0,1975
	1	0,9860	0,9477	0,8905	0,8192	0,7383	0,6517	0,5630	0,4752	0,3910	0,3125	0,8681	0,5926
	2	0,9995	0,9963	0,9880	0,9728	0,9492	0,9163	0,8735	0,8202	0,7585	0,6875	0,9838	0,8889
	3	1,000	0,9999	0,9995	0,9984	0,9961	0,9919	0,9850	0,9744	0,9590	0,9375	0,9992	0,9877
	4		1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
5	0	0,7738	0,5905	0,4437	0,3277	0,2373	0,1681	0,1160	0,0778	0,0503	0,0312	0,4019	0,1317
	1	0,9774	0,9185	0,8352	0,7373	0,6328	0,5282	0,4284	0,3370	0,2562	0,1875	0,8038	0,4609
	2	0,9988	0,9914	0,9734	0,9421	0,8965	0,8369	0,7648	0,6826	0,5931	0,5000	0,9645	0,7901
	3	1,000	0,9995	0,9978	0,9933	0,9844	0,9692	0,9460	0,9130	0,8688	0,8125	0,9967	0,9547
	4		0,9999	0,9999	0,9997	0,9990	0,9976	0,9948	0,9898	0,9816	0,9688	0,9999	0,9959
	5		1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
6	0	0,7351	0,5314	0,3771	0,2621	0,1780	0,1176	0,0754	0,0467	0,0277	0,0156	0,3349	0,0878
	1	0,9672	0,8857	0,7765	0,6554	0,5339	0,4202	0,3191	0,2333	0,1636	0,1094	0,7368	0,3512
	2	0,9978	0,9842	0,9527	0,9011	0,8306	0,7443	0,6471	0,5443	0,4415	0,3438	0,9377	0,6804
	3	0,9999	0,9987	0,9941	0,9830	0,9624	0,9295	0,8726	0,8208	0,7447	0,6562	0,9913	0,8999
	4	1,000	0,9999	0,9996	0,9984	0,9954	0,9891	0,9777	0,9590	0,9308	0,8906	0,9993	0,9822
	5		1,000	1,000	0,9999	0,9998	0,9993	0,9982	0,9959	0,9917	0,9844	1,000	0,9986
	6				1,000	1,000	1,000	1,000	1,000	1,000	1,000		1,000
7	0	0,6983	0,4783	0,3206	0,2097	0,1335	0,0824	0,0490	0,0280	0,0152	0,0078	0,2791	0,0585
	1	0,9556	0,8503	0,7166	0,5767	0,4450	0,3294	0,2338	0,1586	0,1024	0,0625	0,6698	0,2634
	2	0,9962	0,9743	0,9262	0,8520	0,7564	0,6471	0,5323	0,4199	0,3164	0,2266	0,9042	0,5706
	3	0,9998	0,9973	0,9879	0,9667	0,9294	0,8740	0,8002	0,7102	0,6083	0,5000	0,9824	0,8267
	4	1,000	0,9998	0,9988	0,9953	0,9871	0,9712	0,9444	0,9037	0,8471	0,7734	0,9980	0,9547
	5		1,000	0,9999	0,9996	0,9987	0,9962	0,9910	0,9812	0,9643	0,9375	0,9999	0,9931
	6			1,000	1,000	0,9999	0,9998	0,9994	0,9984	0,9963	0,9922	1,000	0,9995
	7				1,000	1,000	1,000	1,000	1,000	1,000	1,000		1,000
8	0	0,6634	0,4305	0,2725	0,1678	0,1001	0,0576	0,0319	0,0168	0,0084	0,0039	0,2326	0,0391
	1	0,9428	0,8131	0,6572	0,5033	0,3671	0,2553	0,1691	0,1064	0,0632	0,0352	0,6047	0,1951
	2	0,9942	0,9619	0,8948	0,7969	0,6786	0,5518	0,4278	0,3154	0,2201	0,1445	0,8652	0,4682
	3	0,9996	0,9950	0,9786	0,9437	0,8862	0,8059	0,7064	0,5941	0,4770	0,3633	0,9693	0,7413
	4	1,000	0,9996	0,9972	0,9896	0,9727	0,9420	0,8939	0,8263	0,7396	0,6367	0,9954	0,9121
	5		1,000	0,9998	0,9988	0,9958	0,9887	0,9747	0,9502	0,9115	0,8555	0,9996	0,9803
	6			1,000	0,9999	0,9996	0,9987	0,9964	0,9915	0,9819	0,9648	1,000	0,9974
	7				1,000	1,000	0,9999	0,9998	0,9993	0,9983	0,9961		0,9998
	8						1,000	1,000	1,000	1,000	1,000		1,000

Cumulatieve binomiale verdeling (vervolg)

		p											
n	*x*	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,50	1/6	1/3
9	0	0,6302	0,3874	0,2316	0,1342	0,0751	0,0404	0,0207	0,0101	0,0046	0,0020	0,2038	0,0260
	1	0,9288	0,7748	0,5995	0,4362	0,3004	0,1960	0,1211	0,0705	0,0385	0,0195	0,5427	0,1431
	2	0,9916	0,9470	0,8592	0,7382	0,6007	0,4628	0,3373	0,2318	0,1495	0,0898,	0,8217	0,3772
	3	0,9994	0,9917	0,9661	0,9144	0,8343	0,7297	0,6089	0,4826	0,3614	0,2539	0,9520	0,6503
	4	1,000	0,9991	0,9944	0,9804	0,9511	0,9012	0,8283	0,7334	0,6214	0,5000	0,9910	0,8552
	5		0,9999	0,9994	0,9969	0,9900	0,9747	0,9464	0,9007	0,8342	0,7461	0,9989	0,9576
	6		1,000	1,000	0,9997	0,9987	0,9957	0,9888	0,9750	0,9502	0,9102	0,9999	0,9917
	7				1,000	0,9999	0,9996	0,9986	0,9962	0,9909	0,9805	1,000	0,9990
	8					1,000	1,000	0,9999	0,9997	0,9992	0,9980		1,000
	9							1,000	1,000	1,000	1,000		
10	0	0,5987	0,3487	0,1969	0,1074	0,0563	0,0282	0,0135	0,0060	0,0025	0,0010	0,1615	0,0173
	1	0,9139	0,7361	0,5443	0,3758	0,2440	0,1493	0,0860	0,0464	0,0233	0,0107	0,4845	0,1040
	2	0,9885	0,9298	0,8202	0,6778	0,5256	0,3828	0,2616	0,1673	0,0996	0,0547	0,7752	0,2991
	3	0,9990	0,9872	0,9500	0,8791	0,7759	0,6496	0,5138	0,3823	0,2670	0,1719	0,9303	0,5593
	4	0,9999	0,9984	0,9901	0,9672	0,9219	0,8497	0,7515	0,6331	0,5044	0,3770	0,9845	0,7969
	5	1,000	0,9998	0,9986	0,9936	0,9803	0,9526	0,9051	0,8338	0,7384	0,6230	0,9976	0,9234
	6		1,000	0,9999	0,9991	0,9965	0,9894	0,9740	0,9452	0,8980	0,8281	0,9997	0,9803
	7			1,000	0,9999	0,9996	0,9984	0,9952	0,9877	0,9726	0,9453	1,000	0,9966
	8				1,000	1,000	0,9999	0,9995	0,9983	0,9955	0,9893		0,9996
	9						1,000	1,000	0,9999	0,9997	0,9990		1,000
	10								1,000	1,000	1,000		
11	0	0,5688	0,3138	0,1673	0,0859	0,0422	0,0198	0,0088	0,0036	0,0014	0,0005	0,1346	0,0116
	1	0,8981	0,6974	0,4922	0,3221	0,1971	0,1130	0,0606	0,0302	0,0139	0,0059	0,4307	0,0751
	2	0,9848	0,9104	0,7788	0,6174	0,4552	0,3127	0,2001	0,1189	0,0652	0,0327	0,7268	0,2341
	3	0,9984	0,9815	0,9306	0,8389	0,7133	0,5696	0,4256	0,2963	0,1911	0,1133	0,9044	0,4726
	4	0,9999	0,9972	0,9841	0,9496	0,8854	0,7897	0,6683	0,5328	0,3971	0,2744	0,9755	0,7110
	5	1,0000	0,9997	0,9973	0,9883	0,9657	0,9218	0,8513	0,7535	0,6331	0,5000	0,9954	0,8779
	6		1,0000	0,9997	0,9980	0,9924	0,9784	0,9499	0,9006	0,8262	0,7256	0,9994	0,9614
	7			1,0000	0,9998	0,9988	0,9957	0,9878	0,9707	0,9390	0,8867	0,9999	0,9912
	8				1,0000	0,9999	0,9994	0,9980	0,9941	0,9852	0,9673	1,0000	0,9986
	9					1,0000	1,0000	0,9998	0,9993	0,9978	0,9941		0,9999
	10							1,0000	1,0000	0,9998	0,9995		1,0000
	11									1,0000	1,0000		
12	0	0,5404	0,2824	0,1422	0,0687	0,0317	0,0138	0,0057	0,0022	0,0008	0,0002	0,1122	0,0077
	1	0,8816	0,6590	0,4435	0,2749	0,1584	0,0850	0,0424	0,0196	0,0083	0,0032	0,3813	0,0540
	2	0,9804	0,8891	0,7358	0,5583	0,3907	0,2528	0,1513	0,0834	0,0421	0,0193	0,6774	0,1811
	3	0,9978	0,9744	0,9078	0,7946	0,6488	0,4925	0,3467	0,2253	0,1345	0,0730	0,8748	0,3931
	4	0,9998	0,9957	0,9761	0,9274	0,8424	0,7237	0,5833	0,4382	0,3044	0,1938	0,9636	0,6315
	5	1,0000	0,9995	0,9954	0,9806	0,9456	0,8822	0,7873	0,6652	0,5269	0,3872	0,9921	0,8223
	6		0,9999	0,9993	0,9961	0,9857	0,9614	0,9154	0,8418	0,7393	0,6128	0,9987	0,9336
	7		1,0000	0,9999	0,9994	0,9972	0,9905	0,9745	0,9427	0,8883	0,8062	0,9998	0,9812
	8			1,0000	0,9999	0,9996	0,9983	0,9944	0,9847	0,9644	0,9270	1,0000	0,9961
	9				1,0000	1,0000	0,9998	0,9992	0,9972	0,9921	0,9807		0,9995
	10						1,0000	0,9999	0,9997	0,9989	0,9968		1,0000
	11							1,0000	1,0000	0,9999	0,9998		
	12									1,0000	1,0000		

Cumulatieve binomiale verdeling (vervolg)

		p											
n	x	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,50	1/6	1/3
13	0	0,5133	0,2542	0,1209	0,0550	0,0238	0,0097	0,0037	0,0013	0,0004	0,0001	0,0935	0,0051
	1	0,8646	0,6213	0,3983	0,2336	0,1267	0,0637	0,0296	0,0126	0,0049	0,0017	0,3365	0,0385
	2	0,9755	0,8661	0,6920	0,5017	0,3326	0,2025	0,1132	0,0579	0,0269	0,0112	0,6281	0,1387
	3	0,9959	0,9658	0,8820	0,7473	0,5843	0,4206	0,2783	0,1686	0,0929	0,0461	0,8419	0,3224
	4	0,9997	0,9935	0,9658	0,9009	0,7940	0,6543	0,5005	0,3530	0,2279	0,1334	0,9488	0,5520
	5	1,0000	0,9991	0,9925	0,9700	0,9198	0,8346	0,7159	0,5744	0,4268	0,2905	0,9873	0,7587
	6		0,9999	0,9987	0,9930	0,9757	0,9376	0,8705	0,7712	0,6437	0,5000	0,9976	0,8965
	7		1,0000	0,9998	0,9988	0,9944	0,9818	0,9538	0,9023	0,8212	0,7095	0,9997	0,9653
	8			1,0000	0,9998	0,9990	0,9960	0,9874	0,9679	0,9302	0,8666	1,0000	0,9912
	9				1,0000	0,9999	0,9993	0,9975	0,9922	0,9797	0,9539		0,9984
	10					1,0000	0,9999	0,9997	0,9987	0,9959	0,9888		0,9998
	11						1,0000	1,0000	0,9999	0,9995	0,9983		1,0000
	12								1,0000	1,0000	0,9999		
	13										1,0000		
14	0	0,4877	0,2288	0,1028	0,0440	0,0178	0,0068	0,0024	0,0008	0,0002	0,0001	0,0779	0,0034
	1	0,8470	0,5846	0,3567	0,1979	0,1010	0,0475	0,0205	0,0081	0,0029	0,0009	0,2960	0,0274
	2	0,9699	0,8416	0,6479	0,4481	0,2811	0,1608	0,0839	0,0398	0,0170	0,0065	0,5795	0,1053
	3	0,9958	0,9559	0,8535	0,6982	0,5213	0,3552	0,2205	0,1243	0,0632	0,0287	0,8063	0,2612
	4	0,9996	0,9908	0,9533	0,8702	0,7415	0,5842	0,4227	0,2793	0,1672	0,0898	0,9310	0,4755
	5	1,0000	0,9985	0,9885	0,9561	0,8883	0,7805	0,6405	0,4859	0,3373	0,2120	0,9809	0,6898
	6		0,9998	0,9978	0,9884	0,9617	0,9067	0,8164	0,6925	0,5461	0,3953	0,9959	0,8505
	7		1,0000	0,9997	0,9976	0,9897	0,9685	0,9247	0,8499	0,7414	0,6047	0,9993	0,9424
	8			1,0000	0,9996	0,9978	0,9917	0,9757	0,9417	0,8811	0,7880	0,9999	0,9826
	9				1,0000	0,9997	0,9983	0,9940	0,9825	0,9574	0,9102	1,0000	0,9960
	10					1,0000	0,9998	0,9989	0,9961	0,9886	0,9713		0,9993
	11						1,0000	0,9999	0,9994	0,9978	0,9935		0,9999
	12							1,0000	0,9999	0,9997	0,9991		1,0000
	13								1,0000	1,0000	0,9999		
	14										1,0000		
15	0	0,4633	0,2059	0,0874	0,0352	0,0134	0,0047	0,0016	0,0005	0,0001	0,0000	0,0649	0,0023
	1	0,8290	0,5490	0,3186	0,1671	0,0802	0,0353	0,0142	0,0052	0,0017	0,0005	0,2596	0,0194
	2	0,9638	0,8159	0,6042	0,3980	0,2361	0,1268	0,0617	0,0271	0,0107	0,0037	0,5322	0,0794
	3	0,9945	0,9444	0,8227	0,6482	0,4613	0,2969	0,1727	0,0905	0,0424	0,0176	0,7685	0,2092
	4	0,9994	0,9873	0,9383	0,8358	0,6865	0,5155	0,3519	0,2173	0,1204	0,0592	0,9102	0,4041
	5	0,9999	0,9978	0,9832	0,9389	0,8516	0,7216	0,5643	0,4032	0,2608	0,1509	0,9726	0,6184
	6	1,0000	0,9997	0,9964	0,9819	0,9434	0,8689	0,7548	0,6098	0,4522	0,3036	0,9934	0,7970
	7		1,0000	0,9994	0,9958	0,9827	0,9500	0,8868	0,7869	0,6535	0,5000	0,9987	0,9118
	8			0,9999	0,9992	0,9958	0,9848	0,9578	0,9050	0,8182	0,6964	0,9998	0,9692
	9			1,0000	0,9999	0,9992	0,9963	0,9876	0,9662	0,9231	0,8491	1,0000	0,9915
	10				1,0000	0,9999	0,9993	0,9972	0,9907	0,9745	0,9408		0,9982
	11					1,0000	0,9999	0,9995	0,9981	0,9937	0,9824		0,9997
	12						1,0000	0,9999	0,9997	0,9989	0,9963		1,0000
	13							1,0000	1,0000	0,9999	0,9995		
	14									1,0000	1,0000		
16	0	0,4401	0,1853	0,0743	0,0281	0,0100	0,0033	0,0010	0,0003	0,0001	0,0000	0,0541	0,0015
	1	0,8108	0,5147	0,2839	0,1407	0,0635	0,0261	0,0098	0,0033	0,0010	0,0003	0,2272	0,0137
	2	0,9571	0,7892	0,5614	0,3518	0,1971	0,0994	0,0451	0,0183	0,0066	0,0021	0,4868	0,0594
	3	0,9930	0,9316	0,7899	0,5981	0,4050	0,2459	0,1339	0,0651	0,0281	0,0106	0,7291	0,1659
	4	0,9991	0,9830	0,9209	0,7982	0,6302	0,4499	0,2892	0,1666	0,0853	0,0384	0,8866	0,3391
	5	0,9999	0,9967	0,9765	0,9183	0,8103	0,6598	0,4900	0,3288	0,1976	0,1051	0,9622	0,5469
	6	1,0000	0,9995	0,9944	0,9733	0,9204	0,8247	0,6881	0,5272	0,3660	0,2272	0,9899	0,7374
	7		0,9999	0,9989	0,9930	0,9729	0,9256	0,8406	0,7161	0,5629	0,4018	0,9979	0,8735
	8		1,0000	0,9998	0,9985	0,9925	0,9743	0,9329	0,8577	0,7441	0,5982	0,9996	0,9500
	9			1,0000	0,9998	0,9984	0,9929	0,9771	0,9417	0,8759	0,7728	1,0000	0,9841
	10				1,0000	0,9997	0,9984	0,9938	0,9809	0,9514	0,8949		0,9960
	11					1,0000	0,9997	0,9987	0,9951	0,9851	0,9616		0,9992
	12						1,0000	0,9998	0,9991	0,9965	0,9894		0,9999
	13							1,0000	0,9999	0,9994	0,9979		1,0000
	14								1,0000	0,9999	0,9997		
	15									1,0000	1,0000		

Cumulatieve binomiale verdeling (vervolg)

		p											
n	*x*	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,50	1/6	1/3
17	0	0,4181	0,1668	0,0631	0,0225	0,0075	0,0023	0,0007	0,0002	0,0000	0,0000	0,0451	0,0010
	1	0,7922	0,4818	0,2525	0,1182	0,0501	0,0193	0,0067	0,0021	0,0006	0,0001	0,1983	0,0096
	2	0,9497	0,7618	0,5198	0,3096	0,1637	0,0774	0,0327	0,0123	0,0041	0,0012	0,4435	0,0442
	3	0,9912	0,9174	0,7556	0,5489	0,3530	0,2019	0,1028	0,0464	0,0184	0,0064	0,6887	0,1304
	4	0,9988	0,9779	0,9013	0,7582	0,5739	0,3887	0,2348	0,1260	0,0596	0,0245	0,8604	0,2814
	5	0,9999	0,9953	0,9681	0,8943	0,7653	0,5968	0,4197	0,2639	0,1471	0,0717	0,9496	0,4777
	6	1,0000	0,9992	0,9917	0,9623	0,8929	0,7752	0,6188	0,4478	0,2902	0,1662	0,9853	0,6739
	7		0,9999	0,9983	0,9891	0,9598	0,8954	0,7872	0,6405	0,4743	0,3145	0,9965	0,8281
	8		1,0000	0,9997	0,9974	0,9876	0,9597	0,9006	0,8011	0,6626	0,5000	0,9993	0,9245
	9			1,0000	0,9995	0,9969	0,9873	0,9617	0,9081	0,8166	0,6855	0,9999	0,9727
	10				0,9999	0,9994	0,9968	0,9880	0,9652	0,9174	0,8338	1,0000	0,9920
	11				1,0000	0,9999	0,9993	0,9970	0,9894	0,9699	0,9283		0,9981
	12					1,0000	0,9999	0,9994	0,9975	0,9914	0,9755		0,9997
	13						1,0000	0,9999	0,9995	0,9981	0,9936		1,0000
	14							1,0000	0,9999	0,9997	0,9988		
	15								1,0000	1,0000	0,9999		
	16										1,0000		
18	0	0,3972	0,1501	0,0536	0,0180	0,0056	0,0016	0,0004	0,0001	0,0000	0,0000	0,0376	0,0007
	1	0,7735	0,4503	0,2241	0,0991	0,0395	0,0142	0,0046	0,0013	0,0003	0,0001	0,1728	0,0068
	2	0,9419	0,7338	0,4797	0,2713	0,1353	0,0600	0,0236	0,0082	0,0025	0,0007	0,4027	0,0326
	3	0,9891	0,9018	0,7202	0,5010	0,3057	0,1646	0,0783	0,0328	0,0120	0,0038	0,6479	0,1017
	4	0,9985	0,9718	0,8794	0,7164	0,5187	0,3327	0,1886	0,0942	0,0411	0,0154	0,8318	0,2311
	5	0,9998	0,9936	0,9581	0,8671	0,7175	0,5344	0,3550	0,2088	0,1077	0,0481	0,9347	0,4122
	6	1,0000	0,9988	0,9882	0,9487	0,8610	0,7217	0,5491	0,3743	0,2258	0,1189	0,9794	0,6085
	7		0,9998	0,9973	0,9837	0,9431	0,8593	0,7283	0,5634	0,3915	0,2403	0,9947	0,7767
	8		1,0000	0,9995	0,9957	0,9807	0,9404	0,8609	0,7368	0,5778	0,4073	0,9989	0,8924
	9			0,9999	0,9991	0,9946	0,9790	0,9403	0,8653	0,7473	0,5927	0,9998	0,9567
	10			1,0000	0,9998	0,9988	0,9939	0,9788	0,9424	0,8720	0,7597	1,0000	0,9856
	11				1,0000	0,9998	0,9986	0,9938	0,9797	0,9463	0,8811		0,9961
	12					1,0000	0,9997	0,9986	0,9942	0,9817	0,9519		0,9991
	13						1,0000	0,9997	0,9987	0,9951	0,9846		0,9999
	14							1,0000	0,9998	0,9990	0,9962		1,0000
	15								1,0000	0,9999	0,9993		
	16									1,0000	0,9999		
	17										1,0000		
19	0	0,3774	0,1351	0,0456	0,0144	0,0042	0,0011	0,0003	0,0001	0,0000	0,0000	0,0313	0,0005
	1	0,7547	0,4203	0,1985	0,0829	0,0310	0,0104	0,0031	0,0008	0,0002	0,0000	0,1502	0,0047
	2	0,9335	0,7054	0,4413	0,2369	0,1113	0,0462	0,0170	0,0055	0,0015	0,0004	0,3643	0,0240
	3	0,9858	0,8850	0,6841	0,4551	0,2631	0,1332	0,0591	0,0230	0,0077	0,0022	0,6070	0,0787
	4	0,9980	0,9648	0,8556	0,6733	0,4654	0,2822	0,1500	0,0696	0,0280	0,0096	0,8011	0,1879
	5	0,9998	0,9914	0,9463	0,8369	0,6678	0,4739	0,2968	0,1629	0,0777	0,0318	0,9176	0,3519
	6	1,0000	0,9983	0,9837	0,9324	0,8251	0,6655	0,4812	0,3081	0,1727	0,0835	0,9719	0,5431
	7		0,9997	0,9959	0,9767	0,9225	0,8180	0,6656	0,4878	0,3169	0,1796	0,9921	0,7207
	8		1,0000	0,9992	0,9933	0,9713	0,9161	0,8145	0,6675	0,4940	0,3238	0,9982	0,8538
	9			0,9999	0,9984	0,9911	0,9674	0,9125	0,8139	0,6710	0,5000	0,9996	0,9352
	10			1,0000	0,9997	0,9977	0,9895	0,9653	0,9115	0,8159	0,6762	0,9999	0,9759
	11				1,0000	0,9995	0,9972	0,9886	0,9648	0,9129	0,8204	1,0000	0,9926
	12					0,9999	0,9994	0,9969	0,9884	0,9658	0,9165		0,9981
	13					1,0000	0,9999	0,9993	0,9969	0,9891	0,9682		0,9996
	14						1,0000	0,9999	0,9994	0,9972	0,9904		0,9999
	15							1,0000	0,9999	0,9995	0,9978		1,0000
	16								1,0000	0,9999	0,9996		
	17									1,0000	1,0000		

Cumulatieve binomiale verdeling (vervolg)

		p												
n	*x*	0.01	0,05	0.10	0.15	0,20	0,25	0.30	0,35	0,40	0,45	0,50	1/6	1/3
20	0	0,8179	0.3585	0,1216	0,0388	0,0115	0,0032	0,0008	0,0002	0,0000	0,0000	0,0000	0,0261	0,0003
	1	0,9831	0.7358	0,3917	0,1756	0,0692	0,0243	0,0076	0,0021	0,0005	0,0001	0,0000	0,1304	0,0033
	2	0,9990	0,9245	0,6769	0,4049	0,2061	0,0913	0,0355	0,0121	0,0036	0,0009	0,0002	0,3287	0,0176
	3	1,0000	0,9841	0,8670	0,6477	0,4114	0,2252	0,1071	0,0444	0,0160	0,0049	0,0013	0,5665	0,0604
	4		0,9974	0,9568	0,8298	0,6296	0,4148	0,2375	0,1182	0,0510	0,0189	0,0059	0,7687	0,1515
	5		0,9997	0,9887	0,9327	0,8042	0,6172	0,4164	0,2454	0,1256	0,0553	0,0207	0,8982	0,2972
	6		1,0000	0,9976	0,9781	0,9133	0,7858	0,6080	0,4166	0,2500	0,1299	0,0577	0,9629	0,4793
	7			0,9996	0,9941	0,9679	0,8982	0,7723	0,6010	0,4159	0,2520	0,1316	0,9887	0,6615
	8			0,9999	0,9987	0,9900	0,9591	0,8867	0,7624	0,5956	0,4143	0,2517	0,9972	0,8095
	9			1,0000	0,9998	0,9974	0,9861	0,9520	0,8782	0,7553	0,5914	0,4119	0,9994	0,9081
	10				1,0000	0,9994	0,9961	0,9829	0,9468	0,8725	0,7507	0,5881	0,9999	0,9624
	11					0,9999	0,9991	0,9949	0,9804	0,9435	0,8692	0,7483	1,0000	0,9870
	12					1,0000	0,9998	0,9987	0,9940	0,9790	0,9420	0,8684		0,9963
	13						1,0000	0,9997	0,9985	0,9935	0,9786	0,9423		0,9991
	14							1,0000	0,9997	0,9984	0,9936	0,9793		0,9998
	15								1,0000	0,9997	0,9985	0,9941		1,0000
	16									1,0000	0,9997	0,9987		
	17										1,0000	0,9998		
	18											1,0000		
50	0	0,6050	0,0769	0,0052	0,0003	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0000
	1	0,9106	0,2794	0,0338	0,0029	0,0002	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0012	0,0000
	2	0,9852	0,5405	0,1117	0,0142	0,0013	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000	0,0066	0,0000
	3	0,9984	0,7604	0,2503	0,0460	0,0057	0,0005	0,0000	0,0000	0,0000	0,0000	0,0000	0,0238	0,0000
	4	0,9999	0,8964	0,4312	0,1121	0,0185	0,0021	0,0002	0,0000	0,0000	0,0000	0,0000	0,0643	0,0000
	5	1,0000	0,9622	0,6161	0,2194	0,0480	0,0070	0,0007	0,0001	0,0000	0,0000	0,0000	0,1388	0,0001
	6		0,9882	0,7702	0,3613	0,1034	0,0194	0,0025	0,0002	0,0000	0,0000	0,0000	0,2506	0,0005
	7		0,9968	0,8779	0,5188	0,1904	0,0453	0,0073	0,0008	0,0001	0,0000	0,0000	0,3911	0,0017
	8		0,9992	0,9421	0,6681	0,3073	0,0916	0,0183	0,0025	0,0002	0,0000	0,0000	0,5421	0,0050
	9		0,9998	0,9755	0,7911	0,4437	0,1637	0,0402	0,0067	0,0008	0,0001	0,0000	0,6830	0,0127
	10		1,0000	0,9906	0,8801	0,5836	0,2622	0,0789	0,0160	0,0022	0,0002	0,0000	0,7986	0,0284
	11			0,9968	0,9372	0,7107	0,3816	0,1390	0,0342	0,0057	0,0006	0,0000	0,8827	0,0570
	12			0,9990	0,9699	0,8139	0,5110	0,2229	0,0661	0,0133	0,0018	0,0002	0,9373	0,1035
	13			0,9997	0,9868	0,8894	0,6370	0,3279	0,1163	0,0280	0,0045	0,0005	0,9693	0,1715
	14			0,9999	0,9947	0,9393	0,7481	0,4468	0,1878	0,0540	0,0104	0,0013	0,9862	0,2612
	15			1,0000	0,9981	0,9692	0,8369	0,5692	0,2801	0,0955	0,0220	0,0033	0,9943	0,3690
	16				0,9993	0,9856	0,9017	0,6839	0,3889	0,1561	0,0427	0,0077	0,9978	0,4868
	17				0,9998	0,9937	0,9449	0,7822	0,5060	0,2369	0,0765	0,0164	0,9992	0,6046
	18				0,9999	0,9975	0,9713	0,8594	0,6216	0,3356	0,1273	0,0325	0,9997	0,7126
	19				1,0000	0,9991	0,9861	0,9152	0,7264	0,4465	0,1974	0,0595	0,9999	0,8036
	20					0,9997	0,9937	0,9522	0,8139	0,5610	0,2862	0,1013	1,0000	0,8741
	21					0,9999	0,9974	0,9749	0,8813	0,6701	0,3900	0,1611		0,9244
	22					1,0000	0,9990	0,9877	0,9290	0,7660	0,5019	0,2399		0,9576
	23						0,9996	0,9944	0,9604	0,8438	0,6134	0,3359		0,9778
	24						0,9999	0,9976	0,9793	0,9022	0,7160	0,4439		0,9892
	25						1,0000	0,9991	0,9900	0,9427	0,8034	0,5561		0,9951
	26							0,9997	0,9955	0,9686	0,8721	0,6641		0,9979
	27							0,9999	0,9981	0,9840	0,9220	0,7601		0,9992
	28							1,0000	0,9993	0,9924	0,9556	0,8389		0,9997
	29								0,9997	0,9966	0,9765	0,8987		0,9999
	30								0,9999	0,9986	0,9884	0,9405		1,0000
	31								1,0000	0,9995	0,9947	0,9675		
	32									0,9998	0,9978	0,9836		
	33									0,9999	0,9991	0,9923		
	34									1,0000	0,9997	0,9967		
	35										0,9999	0,9987		
	36										1,0000	0,9995		
	37											0,9998		
	38											1,0000		

Cumulatieve binomiale verdeling (slot)

		p												
n	*x*	0,01	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,50	1/6	1/3
100	0	0,3650	0,0059	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	1	0,7358	0,0371	0,0003	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	2	0,9206	0,1183	0,0019	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	3	0,9816	0,2578	0,0078	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
	4	0,9966	0,4360	0,0237	0,0004	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0000
	5	0,9995	0,6160	0,0576	0,0016	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0004	0,0000
	6	0,9999	0,7660	0,1172	0,0047	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0013	0,0000
	7	1,0000	0,8720	0,2061	0,0122	0,0003	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0038	0,0000
	8		0,9369	0,3209	0,0275	0,0009	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0095	0,0000
	9		0,9718	0,4513	0,0551	0,0023	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0213	0,0000
	10		0,9885	0,5832	0,0994	0,0057	0,0001	0,0000	0,0000	0,0000	0,0000	0,0000	0,0427	0,0000
	11		0,9957	0,7030	0,1635	0,0126	0,0004	0,0000	0,0000	0,0000	0,0000	0,0000	0,0777	0,0000
	12		0,9985	0,8018	0,2473	0,0253	0,0010	0,0000	0,0000	0,0000	0,0000	0,0000	0,1297	0,0000
	13		0,9995	0,8761	0,3474	0,0469	0,0025	0,0001	0,0000	0,0000	0,0000	0,0000	0,2000	0,0000
	14		0,9999	0,9274	0,4572	0,0804	0,0054	0,0002	0,0000	0,0000	0,0000	0,0000	0,2874	0,0000
	15		1,0000	0,9601	0,5683	0,1285	0,0111	0,0004	0,0000	0,0000	0,0000	0,0000	0,3877	0,0000
	16			0,9794	0,6725	0,1923	0,0211	0,0010	0,0000	0,0000	0,0000	0,0000	0,4942	0,0001
	17			0,9900	0,7633	0,2712	0,0376	0,0022	0,0001	0,0000	0,0000	0,0000	0,5994	0,0002
	18			0,9954	0,8372	0,3621	0,0630	0,0045	0,0001	0,0000	0,0000	0,0000	0,6965	0,0005
	19			0,9980	0,8935	0,4602	0,0995	0,0089	0,0003	0,0000	0,0000	0,0000	0,7803	0,0011
	20			0,9992	0,9337	0,5595	0,1488	0,0165	0,0008	0,0000	0,0000	0,0000	0,8481	0,0024
	21			0,9997	0,9607	0,6540	0,2114	0,0288	0,0017	0,0000	0,0000	0,0000	0,8998	0,0048
	22			0,9999	0,9779	0,7389	0,2864	0,0479	0,0034	0,0001	0,0000	0,0000	0,9369	0,0091
	23			1,0000	0,9881	0,8109	0,3711	0,0755	0,0066	0,0003	0,0000	0,0000	0,9621	0,0164
	24				0,9939	0,8686	0,4617	0,1136	0,0121	0,0006	0,0000	0,0000	0,9783	0,0281
	25				0,9970	0,9125	0,5535	0,1631	0,0211	0,0012	0,0000	0,0000	0,9881	0,0458
	26				0,9986	0,9442	0,6417	0,2244	0,0351	0,0024	0,0001	0,0000	0,9938	0,0715
	27				0,9994	0,9658	0,7224	0,2964	0,0558	0,0046	0,0002	0,0000	0,9969	0,1066
	28				0,9997	0,9800	0,7925	0,3768	0,0848	0,0084	0,0004	0,0000	0,9985	0,1524
	29				0,9999	0,9888	0,8505	0,4623	0,1236	0,0148	0,0008	0,0000	0,9993	0,2093
	30				1,0000	0,9939	0,8962	0,5491	0,1730	0,0248	0,0015	0,0000	0,9997	0,2766
	31					0,9969	0,9307	0,6331	0,2331	0,0398	0,0030	0,0001	0,9999	0,3525
	32					0,9984	0,9554	0,7107	0,3029	0,0615	0,0055	0,0002	1,0000	0,4344
	33					0,9993	0,9724	0,7793	0,3803	0,0913	0,0098	0,0004		0,5188
	34					0,9997	0,9836	0,8371	0,4624	0,1303	0,0166	0,0009		0,6019
	35					0,9999	0,9906	0,8839	0,5458	0,1795	0,0272	0,0018		0,6803
	36					0,9999	0,9948	0,9201	0,6269	0,2386	0,0429	0,0033		0,7511
	37					1,0000	0,9973	0,9470	0,7024	0,3068	0,0651	0,0060		0,8123
	38						0,9986	0,9660	0,7699	0,3822	0,0951	0,0105		0,8630
	39						0,9993	0,9790	0,8276	0,4621	0,1343	0,0176		0,9034
	40						0,9997	0,9875	0,8750	0,5433	0,1831	0,0284		0,9341
	41						0,9999	0,9928	0,9123	0,6225	0,2415	0,0443		0,9566
	42						0,9999	0,9960	0,9406	0,6967	0,3087	0,0666		0,9724
	43						1,0000	0,9979	0,9611	0,7635	0,3828	0,0967		0,9831
	44							0,9989	0,9754	0,8211	0,4613	0,1356		0,9900
	45							0,9995	0,9850	0,8689	0,5413	0,1841		0,9943
	46							0,9997	0,9912	0,9070	0,6196	0,2421		0,9969
	47							0,9999	0,9950	0,9362	0,6931	0,3086		0,9983
	48							0,9999	0,9973	0,9577	0,7596	0,3822		0,9991
	49							1,0000	0,9985	0,9729	0,8173	0,4602		0,9996
	50								0,9993	0,9832	0,8654	0,5398		0,9998
	51								0,9996	0,9900	0,9040	0,6178		0,9999
	52								0,9998	0,9942	0,9338	0,6914		1,0000
	53								0,9999	0,9968	0,9559	0,7579		
	54								1,0000	0,9983	0,9716	0,8159		
	55									0,9991	0,9824	0,8644		
	56									0,9996	0,9894	0,9033		
	57									0,9998	0,9939	0,9334		
	58									0,9999	0,9966	0,9557		
	59									1,0000	0,9982	0,9716		
	60										0,9991	0,9824		
	61										0,9995	0,9895		
	62										0,9998	0,9940		
	63										0,9999	0,9967		
	64										1,0000	0,9982		
	65											0,9991		
	66											0,9996		
	67											0,9998		
	68											0,9999		
	69											1,0000		

Cumulatieve Poissonverdeling

$$P(\underline{x} \leqslant x) = \sum_{k=0}^{x} \frac{e^{-\lambda}\,\lambda^{k}}{k!}$$

	λ									
x	0,05	0,10	0,15	0,20	0,25	0,30	0,35	0,40	0,45	0,50
0	0,9512	0,9048	0,8607	0,8187	0,7788	0,7408	0,7047	0,6703	0,6376	0,6065
1	0,9988	0,9953	0,9898	0,9825	0,9735	0,9631	0,9513	0,9384	0,9246	0,9098
2	1,000	0,9998	0,9995	0,9988	0,9978	0,9964	0,9945	0,9921	0,9891	0,9856
3		1,000	1,000	0,9999	0,9999	0,9997	0,9995	0,9992	0,9988	0,9982
4				1,000	1,000	1,000	1,000	0,9999	0,9999	0,9998
5								1,000	1,000	1,000
	0,55	0,60	0,65	0,70	0,75	0,80	0,85	0,90	0,95	1,00
0	0,5770	0,5488	0,5220	0,4966	0,4724	0,4493	0,4274	0,4066	0,3867	0,3679
1	0,8943	0,8781	0,8614	0,8442	0,8266	0,8088	0,7907	0,7725	0,7541	0,7358
2	0,9815	0,9769	0,9717	0,9659	0,9595	0,9526	0,9451	0,9371	0,9287	0,9197
3	0,9975	0,9966	0,9956	0,9942	0,9927	0,9909	0,9889	0,9865	0,9839	0,9810
4	0,9997	0,9996	0,9994	0,9992	0,9989	0,9986	0,9982	0,9977	0,9971	0,9963
5	1,000	1,000	0,9999	0,9999	0,9999	0,9998	0,9998	0,9997	0,9995	0,9994
6			1,000	1,000	1,000	1,000	1,000	1,000	0,9999	0,9999
7									1,000	1,000
	1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	2,0
0	0,3329	0,3012	0,2725	0,2466	0,2231	0,2019	0,1827	0,1653	0,1496	0,1353
1	0,6990	0,6626	0,6268	0,5918	0,5578	0,5249	0,4932	0,4628	0,4337	0,4060
2	0,9004	0,8795	0,8571	0,8335	0,8088	0,7834	0,7572	0,7306	0,7037	0,6767
3	0,9743	0,9662	0,9569	0,9463	0,9344	0,9212	0,9068	0,8913	0,8747	0,8571
4	0,9946	0,9922	0,9893	0,9857	0,9814	0,9763	0,9704	0,9636	0,9559	0,9473
5	0,9990	0,9985	0,9978	0,9968	0,9955	0,9940	0,9920	0,9896	0,9868	0,9834
6	0,9999	0,9998	0,9996	0,9994	0,9991	0,9987	0,9981	0,9974	0,9966	0,9955
7	1,000	1,000	0,9999	0,9999	0,9998	0,9997	0,9996	0,9994	0,9992	0,9989
8			1,000	1,000	1,000	1,000	0,9999	0,9999	0,9998	0,9998
9							1,000	1,000	1,000	1,000
	2,1	2,2	2,3	2,4	2,5	2,6	2,7	2,8	2,9	3,0
0	0,1225	0,1108	0,1003	0,0907	0,0821	0,0743	0,0672	0,0608	0,0550	0,0498
1	0,3796	0,3546	0,3309	0,3084	0,2873	0,2674	0,2487	0,2311	0,2146	0,1991
2	0,6496	0,6227	0,5960	0,5697	0,5438	0,5184	0,4936	0,4695	0,4460	0,4232
3	0,8386	0,8194	0,7993	0,7787	0,7576	0,7360	0,7141	0,6919	0,6696	0,6472
4	0,9379	0,9275	0,9162	0,9041	0,8912	0,8774	0,8629	0,8477	0,8318	0,8153
5	0,9796	0,9751	0,9700	0,9643	0,9580	0,9510	0,9433	0,9349	0,9258	0,9161
6	0,9941	0,9925	0,9906	0,9884	0,9858	0,9828	0,9794	0,9756	0,9713	0,9665
7	0,9985	0,9980	0,9974	0,9967	0,9958	0,9947	0,9934	0,9919	0,9901	0,9881
8	0,9997	0,9995	0,9994	0,9991	0,9989	0,9985	0,9981	0,9976	0,9969	0,9962
9	0,9999	0,9999	0,9999	0,9998	0,9997	0,9996	0,9995	0,9993	0,9991	0,9989
10	1,000	1,000	1,000	1,000	0,9999	0,9999	0,9999	0,9998	0,9998	0,9997
11					1,000	1,000	1,000	1,000	0,9999	0,9999
12									1,000	1,000

Cumulatieve Poissonverdeling (slot)

	λ									
x	3,2	3,4	3,6	3,8	4,0	4,2	4,4	4,6	4,8	5,0
0	0,0408	0,0334	0,0273	0,0224	0,0183	0,0150	0,0123	0,0101	0,0082	0,0067
1	0,1712	0,1468	0,1257	0,1074	0,0916	0,0780	0,0663	0,0563	0,0477	0,0404
2	0,3799	0,3397	0,3027	0,2689	0,2381	0,2102	0,1851	0,1626	0,1425	0,1247
3	0,6025	0,5584	0,5152	0,4735	0,4335	0,3954	0,3595	0,3257	0,2942	0,2650
4	0,7806	0,7442	0,7064	0,6678	0,6288	0,5898	0,5512	0,5132	0,4763	0,4405
5	0,8946	0,8705	0,8441	0,8156	0,7851	0,7531	0,7199	0,6858	0,6510	0,6160
6	0,9554	0,9421	0,9267	0,9091	0,8893	0,8675	0,8437	0,8180	0,7908	0,7622
7	0,9832	0,9769	0,9692	0,9599	0,9489	0,9361	0,9214	0,9050	0,8867	0,8666
8	0,9943	0,9917	0,9883	0,9840	0,9786	0,9721	0,9642	0,9549	0,9442	0,9319
9	0,9982	0,9973	0,9960	0,9942	0,9919	0,9889	0,9851	0,9805	0,9749	0,9682
10	0,9995	0,9992	0,9987	0,9981	0,9972	0,9959	0,9943	0,9922	0,9896	0,9863
11	0,9999	0,9998	0,9996	0,9994	0,9991	0,9986	0,9980	0,9971	0,9960	0,9946
12	1,000	0,9999	0,9999	0,9998	0,9997	0,9996	0,9993	0,9990	0,9986	0,9980
13		1,000	1,000	1,000	0,9999	0,9999	0,9998	0,9997	0,9995	0,9993
14					1,000	1,000	0,9999	0,9999	0,9999	0,9998
15							1,000	1,000	1,000	0,9999
16										1,000

x	5,5	6,0	6,5	7,0	7,5	8,0	8,5	9,0	9,5	10,0
0	0,0041	0,0025	0,0015	0,0009	0,0006	0,0003	0,0002	0,0001	0,0001	0,0000
1	0,0266	0,0174	0,0113	0,0073	0,0047	0,0030	0,0019	0,0012	0,0008	0,0005
2	0,0884	0,0620	0,0430	0,0296	0,0203	0,0138	0,0093	0,0062	0,0042	0,0028
3	0,2017	0,1512	0,1119	0,0818	0,0592	0,0424	0,0301	0,0212	0,0149	0,0103
4	0,3575	0,2851	0,2237	0,1730	0,1321	0,0996	0,0744	0,0550	0,0403	0,0293
5	0,5289	0,4457	0,3690	0,3007	0,2414	0,1912	0,1496	0,1157	0,0885	0,0671
6	0,6860	0,6063	0,5265	0,4497	0,3782	0,3134	0,2562	0,2068	0,1650	0,1301
7	0,8095	0,7440	0,6728	0,5987	0,5246	0,4530	0,3856	0,3239	0,2687	0,2202
8	0,8944	0,8472	0,7916	0,7291	0,6620	0,5926	0,5231	0,4557	0,3918	0,3328
9	0,9462	0,9161	0,8774	0,8305	0,7764	0,7166	0,6530	0,5874	0,5218	0,4579
10	0,9748	0,9574	0,9332	0,9015	0,8622	0,8159	0,7634	0,7060	0,6453	0,5830
11	0,9890	0,9799	0,9661	0,9467	0,9208	0,8881	0,8487	0,8030	0,7520	0,6968
12	0,9956	0,9912	0,9840	0,9730	0,9573	0,9362	0,9091	0,8758	0,8364	0,7916
13	0,9983	0,9964	0,9929	0,9872	0,9784	0,9658	0,9486	0,9262	0,8981	0,8645
14	0,9994	0,9986	0,9970	0,9943	0,9897	0,9827	0,9726	0,9585	0,9400	0,9165
15	0,9998	0,9995	0,9988	0,9976	0,9954	0,9918	0,9862	0,9780	0,9665	0,9513
16	0,9999	0,9998	0,9996	0,9990	0,9980	0,9963	0,9934	0,9889	0,9823	0,9730
17	1,000	0,9999	0,9999	0,9996	0,9992	0,9984	0,9970	0,9947	0,9911	0,9857
18		1,000	1,000	0,9999	0,9997	0,9994	0,9987	0,9976	0,9957	0,9928
19				1,000	0,9999	0,9998	0,9995	0,9989	0,9980	0,9965
20					1,000	0,9999	0,9998	0,9996	0,9991	0,9984
21						1,000	0,9999	0,9998	0,9996	0,9993
22							1,000	0,9999	0,9999	0,9997
23								1,000	0,9999	0,9999
24									1,000	1,000

Linker-kritieke waarden $K_{1-\alpha}(n)$ van de tekentoets

$n \backslash \alpha$	.005	.010	.025	.050	.125	$n \backslash \alpha$	.005	.010	.025	.050	.125
1	—	—	—	—	—	51	15	16	18	19	20
2	—	—	—	—	—	52	16	17	18	19	21
3	—	—	—	—	0	53	16	17	18	20	21
4	—	—	—	—	0	54	17	18	19	20	22
5	—	—	—	0	0	55	17	18	19	20	22
6	—	—	0	0	1	56	17	18	20	21	23
7	—	0	0	0	1	57	18	19	20	21	23
8	0	0	0	1	1	58	18	19	21	22	24
9	0	0	1	1	2	59	19	20	21	22	24
10	0	0	1	1	2	60	19	20	21	23	25
11	0	1	1	2	3	61	20	20	22	23	25
12	1	1	2	2	3	62	20	21	22	24	25
13	1	1	2	3	3	63	20	21	23	24	26
14	1	2	2	3	4	64	21	22	23	24	26
15	2	2	3	3	4	65	21	22	24	25	27
16	2	2	3	4	5	66	22	23	24	25	27
17	2	3	4	4	5	67	22	23	25	26	28
18	3	3	4	5	6	68	22	23	25	26	28
19	3	4	4	5	6	69	23	24	25	27	29
20	3	4	5	5	6	70	23	24	26	27	29
21	4	4	5	6	7	71	24	25	26	28	30
22	4	5	5	6	7	72	24	25	27	28	30
23	4	5	6	7	8	73	25	26	27	28	31
24	5	5	6	7	8	74	25	26	28	29	31
25	5	6	7	7	9	75	25	26	28	29	32
26	6	6	7	8	9	76	26	27	28	30	32
27	6	7	7	8	10	77	26	27	29	30	32
28	6	7	8	9	10	78	27	28	29	31	33
29	7	7	8	9	10	79	27	28	30	31	33
30	7	8	9	10	11	80	28	29	30	32	34
31	7	8	9	10	11	81	28	29	31	32	34
32	8	8	9	10	12	82	28	30	31	33	35
33	8	9	10	11	12	83	29	30	32	33	35
34	9	9	10	11	13	84	29	30	32	33	36
35	9	10	11	12	13	85	30	31	32	34	36
36	9	10	11	12	14	86	30	31	33	34	37
37	10	10	12	13	14	87	31	32	33	35	37
38	10	11	12	13	14	88	31	32	34	35	38
39	11	11	12	13	15	89	31	33	34	36	38
40	11	12	13	14	15	90	32	33	35	36	39
41	11	12	13	14	16	91	32	33	35	37	39
42	12	13	14	15	16	92	33	34	36	37	39
43	12	13	14	15	17	93	33	34	36	38	40
44	13	13	15	16	17	94	34	35	37	38	40
45	13	14	15	16	18	95	34	35	37	38	41
46	13	14	15	16	18	96	34	36	37	39	41
47	14	15	16	17	19	97	35	36	38	39	42
48	14	15	16	17	19	98	35	37	38	40	42
49	15	15	17	18	19	99	36	37	39	40	43
50	15	16	17	18	20	100	36	37	39	41	43

$\Pr\{k \leqslant K_{1-\alpha}(n)\} \leqslant \alpha < \Pr\{k \leqslant K_{1-\alpha}(n) + 1\}$ met $k \sim B(n; 0{,}5)$.

Linker-kritieke waarden $W_{1-\alpha}(n_1,n_2)$ van de toets van Wilcoxon

$\alpha = 0{,}005$

n_1 \ n_2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0
3	-	-	-	-	-	-	-	-	0	0	0	2	2	2	4	4	4	4	6	6
4	-	-	-	-	-	0	0	2	2	4	4	6	6	8	10	10	12	12	14	16
5	-	-	-	-	0	2	2	4	6	8	10	12	14	14	16	18	20	22	24	26
6	-	-	-	0	2	4	6	8	10	12	14	18	20	22	24	26	30	32	34	36
7	-	-	-	0	2	6	8	12	14	18	20	24	26	30	32	36	38	42	44	48
8	-	-	-	2	4	8	12	14	18	22	26	30	34	36	40	44	48	52	56	60
9	-	-	0	2	6	10	14	18	22	26	32	36	40	44	48	54	58	62	66	72
10	-	-	0	4	8	12	18	22	26	32	36	42	48	52	58	62	68	74	78	84
11	-	-	0	4	10	14	20	26	32	36	42	48	54	60	66	72	78	84	90	95
12	-	-	2	6	12	18	24	30	36	42	48	54	62	68	74	82	88	94	102	108
13	-	-	2	6	14	20	26	34	40	48	54	62	68	76	84	90	98	106	114	120
14	-	-	2	8	14	22	30	36	44	52	60	68	76	84	92	100	108	116	126	134
15	-	-	4	10	16	24	32	40	48	58	66	74	84	92	102	110	120	128	138	146
16	-	-	4	10	18	26	36	44	54	62	72	82	90	100	110	120	130	140	148	158
17	-	-	4	12	20	30	38	48	58	68	78	88	98	108	120	130	140	150	162	172
18	-	-	4	12	22	32	42	52	62	74	84	94	106	116	128	140	150	162	174	184
19	-	0	6	14	24	34	44	56	66	78	90	102	114	126	138	148	162	174	186	198
20	-	0	6	16	26	36	48	60	72	84	96	108	120	134	146	158	172	184	198	210

$\alpha = 0{,}01$

n_1 \ n_2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2	-	-	-	-	-	-	-	-	-	-	-	-	0	0	0	0	0	0	2	2
3	-	-	-	-	-	-	0	0	2	2	2	4	4	4	6	6	8	8	8	10
4	-	-	-	-	0	2	2	4	6	6	8	10	10	12	14	14	16	18	18	20
5	-	-	-	0	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32
6	-	-	-	2	4	6	8	12	14	16	18	22	24	26	30	32	36	38	40	44
7	-	-	0	2	6	8	12	14	18	22	24	28	32	34	38	42	46	48	52	56
8	-	-	0	4	8	12	14	18	22	26	30	34	40	44	48	52	56	60	64	68
9	-	-	2	6	10	14	18	22	28	32	36	42	46	52	56	62	66	72	76	80
10	-	-	2	6	12	16	22	26	32	38	44	48	54	60	66	72	76	82	88	94
11	-	-	2	8	14	18	24	30	36	44	50	56	62	68	74	82	88	94	100	106
12	-	-	4	10	16	22	28	34	42	48	56	62	70	76	84	92	98	106	112	120
13	-	0	4	10	18	24	32	40	46	54	62	70	78	86	94	102	110	118	126	134
14	-	0	4	12	20	26	34	44	52	60	68	76	86	94	102	112	120	130	138	146
15	-	0	6	14	22	30	38	48	56	66	74	84	94	102	112	122	132	140	150	160
16	-	0	6	14	24	32	42	52	62	72	82	92	102	112	122	132	142	152	164	174
17	-	0	8	16	26	36	46	56	66	76	88	98	110	120	132	142	154	164	176	186
18	-	0	8	18	28	38	48	60	72	82	94	106	118	130	140	152	164	176	188	200
19	-	2	8	18	30	40	52	64	76	88	100	112	126	138	150	164	176	188	202	214
20	-	2	10	20	32	44	56	68	80	94	106	120	134	146	160	174	186	200	214	228

Linker-kritieke waarden $W_{1-\alpha}(n_1,n_2)$ van de toets van Wilcoxon (slot)

$\alpha = 0{,}025$

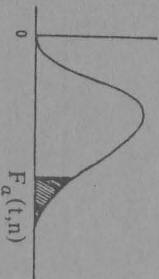
$n_1 \backslash n_2$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2	-	-	-	-	-	-	-	0	0	0	0	2	2	2	2	2	4	4	4	4
3	-	-	-	-	0	2	2	4	4	6	6	8	8	10	10	12	12	14	14	16
4	-	-	-	0	2	4	6	8	8	10	12	14	16	18	20	22	22	24	26	26
5	-	-	0	2	4	6	10	12	14	16	18	22	24	26	28	30	34	36	38	40
6	-	-	2	4	6	10	12	16	20	22	26	28	32	34	38	42	44	48	50	54
7	-	-	2	6	10	12	16	20	24	28	32	36	40	44	48	52	56	60	64	68
8	-	0	4	8	12	16	20	26	30	34	38	44	48	52	58	62	68	72	76	82
9	-	0	4	8	14	20	24	30	34	40	46	52	56	62	68	74	78	84	90	96
10	-	0	6	10	16	22	28	34	40	46	52	58	66	72	78	84	90	96	104	110
11	-	0	6	12	18	26	32	38	46	52	60	66	74	80	88	94	102	110	116	124
12	-	2	8	14	22	28	36	44	52	58	66	74	82	90	98	106	114	122	130	138
13	-	2	8	16	24	32	40	48	56	66	74	82	90	100	108	118	126	134	144	152
14	-	2	10	18	26	34	44	52	62	72	80	90	100	110	118	128	134	148	156	166
15	-	2	10	20	28	38	48	58	68	78	88	98	108	118	128	140	150	160	170	180
16	-	2	12	22	30	42	52	62	74	84	94	106	118	128	140	150	162	172	184	196
17	-	4	12	22	34	44	56	68	78	90	102	114	126	134	150	162	174	186	198	210
18	-	4	14	24	36	48	60	72	84	96	110	122	134	148	160	172	186	198	212	224
19	-	4	14	26	38	50	64	76	90	104	116	130	144	156	170	184	198	212	226	238
20	-	4	16	26	40	54	68	82	96	110	124	138	152	166	180	196	210	224	238	254

$\alpha = 0{,}05$

$n_1 \backslash n_2$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0
2	-	-	-	-	0	0	0	2	2	2	2	4	4	4	6	6	6	8	8	8
3	-	-	0	0	2	4	4	6	6	8	10	10	12	14	14	16	18	18	20	22
4	-	-	0	2	4	6	8	10	12	14	16	18	20	22	24	28	30	32	34	36
5	-	0	2	4	8	10	12	16	18	22	24	26	30	32	36	38	40	44	46	50
6	-	0	4	6	10	14	16	20	24	28	32	34	38	42	46	50	52	56	60	64
7	-	0	4	8	12	16	22	26	30	34	38	42	48	52	56	60	66	70	74	78
8	-	2	6	10	16	20	26	30	36	40	46	52	56	62	66	72	78	82	88	94
9	-	2	6	12	18	24	30	36	42	48	54	60	66	72	78	84	90	96	102	108
10	-	2	8	14	22	28	34	40	48	54	62	68	74	82	88	96	102	110	116	124
11	-	2	10	16	24	32	38	46	54	62	68	76	84	92	100	108	114	122	130	138
12	-	4	10	18	26	34	42	52	60	68	76	84	94	102	110	120	128	136	144	154
13	-	4	12	20	30	38	48	56	66	74	84	94	102	112	122	130	140	150	160	168
14	-	4	14	22	32	42	52	62	72	82	92	102	112	122	132	142	154	164	174	184
15	-	6	14	24	36	46	56	66	78	88	100	110	122	132	144	154	166	176	188	200
16	-	6	16	28	38	50	60	72	84	96	108	120	130	142	154	166	178	190	202	214
17	-	6	18	30	40	52	66	78	90	102	114	128	140	154	166	178	192	204	218	230
18	-	8	18	32	44	56	70	82	96	110	122	136	150	164	176	190	204	218	232	246
19	0	8	20	34	46	60	74	88	102	116	130	144	160	174	188	202	218	232	246	260
20	0	8	22	36	50	64	78	94	108	124	138	154	168	184	200	214	230	246	260	276

F-verdeling

$\alpha = 0{,}05$



vrijheidsgraden van de teller

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161·4	199·5	215·7	224·6	230·2	234·0	236·8	238·9	240·5	241·9	243·9	245·9	248·0	249·1	250·1	251·1	252·2	253·3	254·3
2	18·51	19·00	19·16	19·25	19·30	19·33	19·35	19·37	19·38	19·40	19·41	19·43	19·45	19·45	19·46	19·47	19·48	19·49	19·50
3	10·13	9·55	9·28	9·12	9·01	8·94	8·89	8·85	8·81	8·79	8·74	8·70	8·66	8·64	8·62	8·59	8·57	8·55	8·53
4	7·71	6·94	6·59	6·39	6·26	6·16	6·09	6·04	6·00	5·96	5·91	5·86	5·80	5·77	5·75	5·72	5·69	5·66	5·63
5	6·61	5·79	5·41	5·19	5·05	4·95	4·88	4·82	4·77	4·74	4·68	4·62	4·56	4·53	4·50	4·46	4·43	4·40	4·36
6	5·99	5·14	4·76	4·53	4·39	4·28	4·21	4·15	4·10	4·06	4·00	3·94	3·87	3·84	3·81	3·77	3·74	3·70	3·67
7	5·59	4·74	4·35	4·12	3·97	3·87	3·79	3·73	3·68	3·64	3·57	3·51	3·44	3·41	3·38	3·34	3·30	3·27	3·23
8	5·32	4·46	4·07	3·84	3·69	3·58	3·50	3·44	3·39	3·35	3·28	3·22	3·15	3·12	3·08	3·04	3·01	2·97	2·93
9	5·12	4·26	3·86	3·63	3·48	3·37	3·29	3·23	3·18	3·14	3·07	3·01	2·94	2·90	2·86	2·83	2·79	2·75	2·71
10	4·96	4·10	3·71	3·48	3·33	3·22	3·14	3·07	3·02	2·98	2·91	2·85	2·77	2·74	2·70	2·66	2·62	2·58	2·54
11	4·84	3·98	3·59	3·36	3·20	3·09	3·01	2·95	2·90	2·85	2·79	2·72	2·65	2·61	2·57	2·53	2·49	2·45	2·40
12	4·75	3·89	3·49	3·26	3·11	3·00	2·91	2·85	2·80	2·75	2·69	2·62	2·54	2·51	2·47	2·43	2·38	2·34	2·30
13	4·67	3·81	3·41	3·18	3·03	2·92	2·83	2·77	2·71	2·67	2·60	2·53	2·46	2·42	2·38	2·34	2·30	2·25	2·21
14	4·60	3·74	3·34	3·11	2·96	2·85	2·76	2·70	2·65	2·60	2·53	2·46	2·39	2·35	2·31	2·27	2·22	2·18	2·13
15	4·54	3·68	3·29	3·06	2·90	2·79	2·71	2·64	2·59	2·54	2·48	2·40	2·33	2·29	2·25	2·20	2·16	2·11	2·07
16	4·49	3·63	3·24	3·01	2·85	2·74	2·66	2·59	2·54	2·49	2·42	2·35	2·28	2·24	2·19	2·15	2·11	2·06	2·01
17	4·45	3·59	3·20	2·96	2·81	2·70	2·61	2·55	2·49	2·45	2·38	2·31	2·23	2·19	2·15	2·10	2·06	2·01	1·96
18	4·41	3·55	3·16	2·93	2·77	2·66	2·58	2·51	2·46	2·41	2·34	2·27	2·19	2·15	2·11	2·06	2·02	1·97	1·92
19	4·38	3·52	3·13	2·90	2·74	2·63	2·54	2·48	2·42	2·38	2·31	2·23	2·16	2·11	2·07	2·03	1·98	1·93	1·88
20	4·35	3·49	3·10	2·87	2·71	2·60	2·51	2·45	2·39	2·35	2·28	2·20	2·12	2·08	2·04	1·99	1·95	1·90	1·84
21	4·32	3·47	3·07	2·84	2·68	2·57	2·49	2·42	2·37	2·32	2·25	2·18	2·10	2·05	2·01	1·96	1·92	1·87	1·81
22	4·30	3·44	3·05	2·82	2·66	2·55	2·46	2·40	2·34	2·30	2·23	2·15	2·07	2·03	1·98	1·94	1·89	1·84	1·78
23	4·28	3·42	3·03	2·80	2·64	2·53	2·44	2·37	2·32	2·27	2·20	2·13	2·05	2·01	1·96	1·91	1·86	1·81	1·76
24	4·26	3·40	3·01	2·78	2·62	2·51	2·42	2·36	2·30	2·25	2·18	2·11	2·03	1·98	1·94	1·89	1·84	1·79	1·73
25	4·24	3·39	2·99	2·76	2·60	2·49	2·40	2·34	2·28	2·24	2·16	2·09	2·01	1·96	1·92	1·87	1·82	1·77	1·71
26	4·23	3·37	2·98	2·74	2·59	2·47	2·39	2·32	2·27	2·22	2·15	2·07	1·99	1·95	1·90	1·85	1·80	1·75	1·69
27	4·21	3·35	2·96	2·73	2·57	2·46	2·37	2·31	2·25	2·20	2·13	2·06	1·97	1·93	1·88	1·84	1·79	1·73	1·67
28	4·20	3·34	2·95	2·71	2·56	2·45	2·36	2·29	2·24	2·19	2·12	2·04	1·96	1·91	1·87	1·82	1·77	1·71	1·65
29	4·18	3·33	2·93	2·70	2·55	2·43	2·35	2·28	2·22	2·18	2·10	2·03	1·94	1·90	1·85	1·81	1·75	1·70	1·64
30	4·17	3·32	2·92	2·69	2·53	2·42	2·33	2·27	2·21	2·16	2·09	2·01	1·93	1·89	1·84	1·79	1·74	1·68	1·62
40	4·08	3·23	2·84	2·61	2·45	2·34	2·25	2·18	2·12	2·08	2·00	1·92	1·84	1·79	1·74	1·69	1·64	1·58	1·51
60	4·00	3·15	2·76	2·53	2·37	2·25	2·17	2·10	2·04	1·99	1·92	1·84	1·75	1·70	1·65	1·59	1·53	1·47	1·39
120	3·92	3·07	2·68	2·45	2·29	2·17	2·09	2·02	1·96	1·91	1·83	1·75	1·66	1·61	1·55	1·50	1·43	1·35	1·25
∞	3·84	3·00	2·60	2·37	2·21	2·10	2·01	1·94	1·88	1·83	1·75	1·67	1·57	1·52	1·46	1·39	1·32	1·22	1·00

vrijheidsgraden van de noemer

F-verdeling (vervolg)

$\alpha = 0{,}025$

vrijheidsgraden van de teller

vrijheidsgraden van de noemer

$\nu_2 \backslash \nu_1$	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	647·8	799·5	864·2	899·6	921·8	937·1	948·2	956·7	963·3	968·6	976·7	984·9	993·1	997·2	1001	1006	1010	1014	1018
2	38·51	39·00	39·17	39·25	39·30	39·33	39·36	39·37	39·39	39·40	39·41	39·43	39·45	39·46	39·46	39·47	39·48	39·49	39·50
3	17·44	16·04	15·44	15·10	14·88	14·73	14·62	14·54	14·47	14·42	14·34	14·25	14·17	14·12	14·08	14·04	13·99	13·95	13·90
4	12·22	10·65	9·98	9·60	9·36	9·20	9·07	8·98	8·90	8·84	8·75	8·66	8·56	8·51	8·46	8·41	8·36	8·31	8·26
5	10·01	8·43	7·76	7·39	7·15	6·98	6·85	6·76	6·68	6·62	6·52	6·43	6·33	6·28	6·23	6·18	6·12	6·07	6·02
6	8·81	7·26	6·60	6·23	5·99	5·82	5·70	5·60	5·52	5·46	5·37	5·27	5·17	5·12	5·07	5·01	4·96	4·90	4·85
7	8·07	6·54	5·89	5·52	5·29	5·12	4·99	4·90	4·82	4·76	4·67	4·57	4·47	4·42	4·36	4·31	4·25	4·20	4·14
8	7·57	6·06	5·42	5·05	4·82	4·65	4·53	4·43	4·36	4·30	4·20	4·10	4·00	3·95	3·89	3·84	3·78	3·73	3·67
9	7·21	5·71	5·08	4·72	4·48	4·32	4·20	4·10	4·03	3·96	3·87	3·77	3·67	3·61	3·56	3·51	3·45	3·39	3·33
10	6·94	5·46	4·83	4·47	4·24	4·07	3·95	3·85	3·78	3·72	3·62	3·52	3·42	3·37	3·31	3·26	3·20	3·14	3·08
11	6·72	5·26	4·63	4·28	4·04	3·88	3·76	3·66	3·59	3·53	3·43	3·33	3·23	3·17	3·12	3·06	3·00	2·94	2·88
12	6·55	5·10	4·47	4·12	3·89	3·73	3·61	3·51	3·44	3·37	3·28	3·18	3·07	3·02	2·96	2·91	2·85	2·79	2·72
13	6·41	4·97	4·35	4·00	3·77	3·60	3·48	3·39	3·31	3·25	3·15	3·05	2·95	2·89	2·84	2·78	2·72	2·66	2·60
14	6·30	4·86	4·24	3·89	3·66	3·50	3·38	3·29	3·21	3·15	3·05	2·95	2·84	2·79	2·73	2·67	2·61	2·55	2·49
15	6·20	4·77	4·15	3·80	3·58	3·41	3·29	3·20	3·12	3·06	2·96	2·86	2·76	2·70	2·64	2·59	2·52	2·46	2·40
16	6·12	4·69	4·08	3·73	3·50	3·34	3·22	3·12	3·05	2·99	2·89	2·79	2·68	2·63	2·57	2·51	2·45	2·38	2·32
17	6·04	4·62	4·01	3·66	3·44	3·28	3·16	3·06	2·98	2·92	2·82	2·72	2·62	2·56	2·50	2·44	2·38	2·32	2·25
18	5·98	4·56	3·95	3·61	3·38	3·22	3·10	3·01	2·93	2·87	2·77	2·67	2·56	2·50	2·44	2·38	2·32	2·26	2·19
19	5·92	4·51	3·90	3·56	3·33	3·17	3·05	2·96	2·88	2·82	2·72	2·62	2·51	2·45	2·39	2·33	2·27	2·20	2·13
20	5·87	4·46	3·86	3·51	3·29	3·13	3·01	2·91	2·84	2·77	2·68	2·57	2·46	2·41	2·35	2·29	2·22	2·16	2·09
21	5·83	4·42	3·82	3·48	3·25	3·09	2·97	2·87	2·80	2·73	2·64	2·53	2·42	2·37	2·31	2·25	2·18	2·11	2·04
22	5·79	4·38	3·78	3·44	3·22	3·05	2·93	2·84	2·76	2·70	2·60	2·50	2·39	2·33	2·27	2·21	2·14	2·08	2·00
23	5·75	4·35	3·75	3·41	3·18	3·02	2·90	2·81	2·73	2·67	2·57	2·47	2·36	2·30	2·24	2·18	2·11	2·04	1·97
24	5·72	4·32	3·72	3·38	3·15	2·99	2·87	2·78	2·70	2·64	2·54	2·44	2·33	2·27	2·21	2·15	2·08	2·01	1·94
25	5·69	4·29	3·69	3·35	3·13	2·97	2·85	2·75	2·68	2·61	2·51	2·41	2·30	2·24	2·18	2·12	2·05	1·98	1·91
26	5·66	4·27	3·67	3·33	3·10	2·94	2·82	2·73	2·65	2·59	2·49	2·39	2·28	2·22	2·16	2·09	2·03	1·95	1·88
27	5·63	4·24	3·65	3·31	3·08	2·92	2·80	2·71	2·63	2·57	2·47	2·36	2·25	2·19	2·13	2·07	2·00	1·93	1·85
28	5·61	4·22	3·63	3·29	3·06	2·90	2·78	2·69	2·61	2·55	2·45	2·34	2·23	2·17	2·11	2·05	1·98	1·91	1·83
29	5·59	4·20	3·61	3·27	3·04	2·88	2·76	2·67	2·59	2·53	2·43	2·32	2·21	2·15	2·09	2·03	1·96	1·89	1·81
30	5·57	4·18	3·59	3·25	3·03	2·87	2·75	2·65	2·57	2·51	2·41	2·31	2·20	2·14	2·07	2·01	1·94	1·87	1·79
40	5·42	4·05	3·46	3·13	2·90	2·74	2·62	2·53	2·45	2·39	2·29	2·18	2·07	2·01	1·94	1·88	1·80	1·72	1·64
60	5·29	3·93	3·34	3·01	2·79	2·63	2·51	2·41	2·33	2·27	2·17	2·06	1·94	1·88	1·82	1·74	1·67	1·58	1·48
120	5·15	3·80	3·23	2·89	2·67	2·52	2·39	2·30	2·22	2·16	2·05	1·94	1·82	1·76	1·69	1·61	1·53	1·43	1·31
∞	5·02	3·69	3·12	2·79	2·57	2·41	2·29	2·19	2·11	2·05	1·94	1·83	1·71	1·64	1·57	1·48	1·39	1·27	1·00

F-verdeling (vervolg)

$\alpha = 0{,}01$

vrijheidsgraden van de teller

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4052	4999·5	5403	5625	5764	5859	5928	5981	6022	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	98·50	99·00	99·17	99·25	99·30	99·33	99·36	99·37	99·39	99·40	99·42	99·43	99·45	99·46	99·47	99·47	99·48	99·49	99·50
3	34·12	30·82	29·46	28·71	28·24	27·91	27·67	27·49	27·35	27·23	27·05	26·87	26·69	26·60	26·50	26·41	26·32	26·22	26·13
4	21·20	18·00	16·69	15·98	15·52	15·21	14·98	14·80	14·66	14·55	14·37	14·20	14·02	13·93	13·84	13·75	13·65	13·56	13·46
5	16·26	13·27	12·06	11·39	10·97	10·67	10·46	10·29	10·16	10·05	9·89	9·72	9·55	9·47	9·38	9·29	9·20	9·11	9·02
6	13·75	10·92	9·78	9·15	8·75	8·47	8·26	8·10	7·98	7·87	7·72	7·56	7·40	7·31	7·23	7·14	7·06	6·97	6·88
7	12·25	9·55	8·45	7·85	7·46	7·19	6·99	6·84	6·72	6·62	6·47	6·31	6·16	6·07	5·99	5·91	5·82	5·74	5·65
8	11·26	8·65	7·59	7·01	6·63	6·37	6·18	6·03	5·91	5·81	5·67	5·52	5·36	5·28	5·20	5·12	5·03	4·95	4·86
9	10·56	8·02	6·99	6·42	6·06	5·80	5·61	5·47	5·35	5·26	5·11	4·96	4·81	4·73	4·65	4·57	4·48	4·40	4·31
10	10·04	7·56	6·55	5·99	5·64	5·39	5·20	5·06	4·94	4·85	4·71	4·56	4·41	4·33	4·25	4·17	4·08	4·00	3·91
11	9·65	7·21	6·22	5·67	5·32	5·07	4·89	4·74	4·63	4·54	4·40	4·25	4·10	4·02	3·94	3·86	3·78	3·69	3·60
12	9·33	6·93	5·95	5·41	5·06	4·82	4·64	4·50	4·39	4·30	4·16	4·01	3·86	3·78	3·70	3·62	3·54	3·45	3·36
13	9·07	6·70	5·74	5·21	4·86	4·62	4·44	4·30	4·19	4·10	3·96	3·82	3·66	3·59	3·51	3·43	3·34	3·25	3·17
14	8·86	6·51	5·56	5·04	4·69	4·46	4·28	4·14	4·03	3·94	3·80	3·66	3·51	3·43	3·35	3·27	3·18	3·09	3·00
15	8·68	6·36	5·42	4·89	4·56	4·32	4·14	4·00	3·89	3·80	3·67	3·52	3·37	3·29	3·21	3·13	3·05	2·96	2·87
16	8·53	6·23	5·29	4·77	4·44	4·20	4·03	3·89	3·78	3·69	3·55	3·41	3·26	3·18	3·10	3·02	2·93	2·84	2·75
17	8·40	6·11	5·18	4·67	4·34	4·10	3·93	3·79	3·68	3·59	3·46	3·31	3·16	3·08	3·00	2·92	2·83	2·75	2·65
18	8·29	6·01	5·09	4·58	4·25	4·01	3·84	3·71	3·60	3·51	3·37	3·23	3·08	3·00	2·92	2·84	2·75	2·66	2·57
19	8·18	5·93	5·01	4·50	4·17	3·94	3·77	3·63	3·52	3·43	3·30	3·15	3·00	2·92	2·84	2·76	2·67	2·58	2·49
20	8·10	5·85	4·94	4·43	4·10	3·87	3·70	3·56	3·46	3·37	3·23	3·09	2·94	2·86	2·78	2·69	2·61	2·52	2·42
21	8·02	5·78	4·87	4·37	4·04	3·81	3·64	3·51	3·40	3·31	3·17	3·03	2·88	2·80	2·72	2·64	2·55	2·46	2·36
22	7·95	5·72	4·82	4·31	3·99	3·76	3·59	3·45	3·35	3·26	3·12	2·98	2·83	2·75	2·67	2·58	2·50	2·40	2·31
23	7·88	5·66	4·76	4·26	3·94	3·71	3·54	3·41	3·30	3·21	3·07	2·93	2·78	2·70	2·62	2·54	2·45	2·35	2·26
24	7·82	5·61	4·72	4·22	3·90	3·67	3·50	3·36	3·26	3·17	3·03	2·89	2·74	2·66	2·58	2·49	2·40	2·31	2·21
25	7·77	5·57	4·68	4·18	3·85	3·63	3·46	3·32	3·22	3·13	2·99	2·85	2·70	2·62	2·54	2·45	2·36	2·27	2·17
26	7·72	5·53	4·64	4·14	3·82	3·59	3·42	3·29	3·18	3·09	2·96	2·81	2·66	2·58	2·50	2·42	2·33	2·23	2·13
27	7·68	5·49	4·60	4·11	3·78	3·56	3·39	3·26	3·15	3·06	2·93	2·78	2·63	2·55	2·47	2·38	2·29	2·20	2·10
28	7·64	5·45	4·57	4·07	3·75	3·53	3·36	3·23	3·12	3·03	2·90	2·75	2·60	2·52	2·44	2·35	2·26	2·17	2·06
29	7·60	5·42	4·54	4·04	3·73	3·50	3·33	3·20	3·09	3·00	2·87	2·73	2·57	2·49	2·41	2·33	2·23	2·14	2·03
30	7·56	5·39	4·51	4·02	3·70	3·47	3·30	3·17	3·07	2·98	2·84	2·70	2·55	2·47	2·39	2·30	2·21	2·11	2·01
40	7·31	5·18	4·31	3·83	3·51	3·29	3·12	2·99	2·89	2·80	2·66	2·52	2·37	2·29	2·20	2·11	2·02	1·92	1·80
60	7·08	4·98	4·13	3·65	3·34	3·12	2·95	2·82	2·72	2·63	2·50	2·35	2·20	2·12	2·03	1·94	1·84	1·73	1·60
120	6·85	4·79	3·95	3·48	3·17	2·96	2·79	2·66	2·56	2·47	2·34	2·19	2·03	1·95	1·86	1·76	1·66	1·53	1·38
∞	6·63	4·61	3·78	3·32	3·02	2·80	2·64	2·51	2·41	2·32	2·18	2·04	1·88	1·79	1·70	1·59	1·47	1·32	1·00

vrijheidsgraden van de noemer

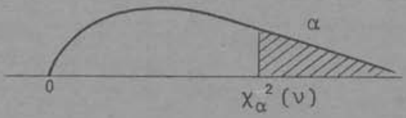
F-verdeling (slot) $\alpha = 0{,}005$

vrijheidsgraden van de teller

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	16211	20000	21615	22500	23056	23437	23715	23925	24091	24224	24426	24630	24836	24940	25044	25148	25253	25359	25465
2	198·5	199·0	199·2	199·2	199·3	199·3	199·4	199·4	199·4	199·4	199·4	199·4	199·4	199·5	199·5	199·5	199·5	199·5	199·5
3	55·55	49·80	47·47	46·19	45·39	44·84	44·43	44·13	43·88	43·69	43·39	43·08	42·78	42·62	42·47	42·31	42·15	41·99	41·83
4	31·33	26·28	24·26	23·15	22·46	21·97	21·62	21·35	21·14	20·97	20·70	20·44	20·17	20·03	19·89	19·75	19·61	19·47	19·32
5	22·78	18·31	16·53	15·56	14·94	14·51	14·20	13·96	13·77	13·62	13·38	13·15	12·90	12·78	12·66	12·53	12·40	12·27	12·14
6	18·63	14·54	12·92	12·03	11·46	11·07	10·79	10·57	10·39	10·25	10·03	9·81	9·59	9·47	9·36	9·24	9·12	9·00	8·88
7	16·24	12·40	10·88	10·05	9·52	9·16	8·89	8·68	8·51	8·38	8·18	7·97	7·75	7·65	7·53	7·42	7·31	7·19	7·08
8	14·69	11·04	9·60	8·81	8·30	7·95	7·69	7·50	7·34	7·21	7·01	6·81	6·61	6·50	6·40	6·29	6·18	6·06	5·95
9	13·61	10·11	8·72	7·96	7·47	7·13	6·88	6·69	6·54	6·42	6·23	6·03	5·83	5·73	5·62	5·52	5·41	5·30	5·19
10	12·83	9·43	8·08	7·34	6·87	6·54	6·30	6·12	5·97	5·85	5·66	5·47	5·27	5·17	5·07	4·97	4·86	4·75	4·64
11	12·23	8·91	7·60	6·88	6·42	6·10	5·86	5·68	5·54	5·42	5·24	5·05	4·86	4·76	4·65	4·55	4·44	4·34	4·23
12	11·75	8·51	7·23	6·52	6·07	5·76	5·52	5·35	5·20	5·09	4·91	4·72	4·53	4·43	4·33	4·23	4·12	4·01	3·90
13	11·37	8·19	6·93	6·23	5·79	5·48	5·25	5·08	4·94	4·82	4·64	4·46	4·27	4·17	4·07	3·97	3·87	3·76	3·65
14	11·06	7·92	6·68	6·00	5·56	5·26	5·03	4·86	4·72	4·60	4·43	4·25	4·06	3·96	3·86	3·76	3·66	3·55	3·44
15	10·80	7·70	6·48	5·80	5·37	5·07	4·85	4·67	4·54	4·42	4·25	4·07	3·88	3·79	3·69	3·58	3·48	3·37	3·26
16	10·58	7·51	6·30	5·64	5·21	4·91	4·69	4·52	4·38	4·27	4·10	3·92	3·73	3·64	3·54	3·44	3·33	3·22	3·11
17	10·38	7·35	6·16	5·50	5·07	4·78	4·56	4·39	4·25	4·14	3·97	3·79	3·61	3·51	3·41	3·31	3·21	3·10	2·98
18	10·22	7·21	6·03	5·37	4·96	4·66	4·44	4·28	4·14	4·03	3·86	3·68	3·50	3·40	3·30	3·20	3·10	2·99	2·87
19	10·07	7·09	5·92	5·27	4·85	4·56	4·34	4·18	4·04	3·93	3·76	3·59	3·40	3·31	3·21	3·11	3·00	2·89	2·78
20	9·94	6·99	5·82	5·17	4·76	4·47	4·26	4·09	3·96	3·85	3·68	3·50	3·32	3·22	3·12	3·02	2·92	2·81	2·69
21	9·83	6·89	5·73	5·09	4·68	4·39	4·18	4·01	3·88	3·77	3·60	3·43	3·24	3·15	3·05	2·95	2·84	2·73	2·61
22	9·73	6·81	5·65	5·02	4·61	4·32	4·11	3·94	3·81	3·70	3·54	3·36	3·18	3·08	2·98	2·88	2·77	2·66	2·55
23	9·63	6·73	5·58	4·95	4·54	4·26	4·05	3·88	3·75	3·64	3·47	3·30	3·12	3·02	2·92	2·82	2·71	2·60	2·48
24	9·55	6·66	5·52	4·89	4·49	4·20	3·99	3·83	3·69	3·59	3·42	3·25	3·06	2·97	2·87	2·77	2·66	2·55	2·43
25	9·48	6·60	5·46	4·84	4·43	4·15	3·94	3·78	3·64	3·54	3·37	3·20	3·01	2·92	2·82	2·72	2·61	2·50	2·38
26	9·41	6·54	5·41	4·79	4·38	4·10	3·89	3·73	3·60	3·49	3·33	3·15	2·97	2·87	2·77	2·67	2·56	2·45	2·33
27	9·34	6·49	5·36	4·74	4·34	4·06	3·85	3·69	3·56	3·45	3·28	3·11	2·93	2·83	2·73	2·63	2·52	2·41	2·29
28	9·28	6·44	5·32	4·70	4·30	4·02	3·81	3·65	3·52	3·41	3·25	3·07	2·89	2·79	2·69	2·59	2·48	2·37	2·25
29	9·23	6·40	5·28	4·66	4·26	3·98	3·77	3·61	3·48	3·38	3·21	3·04	2·86	2·76	2·66	2·56	2·45	2·33	2·21
30	9·18	6·35	5·24	4·62	4·23	3·95	3·74	3·58	3·45	3·34	3·18	3·01	2·82	2·73	2·63	2·52	2·42	2·30	2·18
40	8·83	6·07	4·98	4·37	3·99	3·71	3·51	3·35	3·22	3·12	2·95	2·78	2·60	2·50	2·40	2·30	2·18	2·06	1·93
60	8·49	5·79	4·73	4·14	3·76	3·49	3·29	3·13	3·01	2·90	2·74	2·57	2·39	2·29	2·19	2·08	1·96	1·83	1·69
120	8·18	5·54	4·50	3·92	3·55	3·28	3·09	2·93	2·81	2·71	2·54	2·37	2·19	2·09	1·98	1·87	1·75	1·61	1·43
∞	7·88	5·30	4·28	3·72	3·35	3·09	2·90	2·74	2·62	2·52	2·36	2·19	2·00	1·90	1·79	1·67	1·53	1·36	1·00

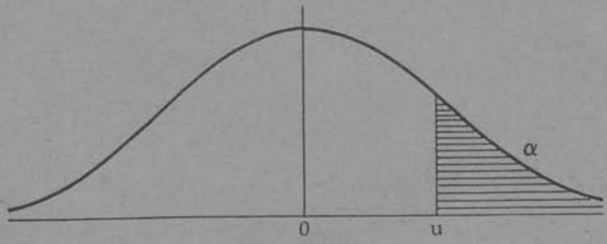
vrijheidsgraden van de noemer

Chi-kwadraat-verdeling



ν \ α	.995	.99	.975	.95	.90	.75	.50	.25	.10	.05	.025	.010	.005	.001
1	-	-	.001	.004	.016	.102	.445	1.32	2.71	3.84	5.02	6.63	7.88	10.8
2	.010	.020	.051	.103	.211	.575	1.39	2.77	4.61	5.99	7.38	9.21	10.6	13.8
3	.072	.115	.216	.352	.584	1.21	2.37	4.11	6.25	7.81	9.35	11.3	12.8	16.3
4	.207	.297	.484	.711	1.06	1.92	3.36	5.39	7.78	9.49	11.1	13.3	14.9	18.5
5	.412	.554	.831	1.15	1.61	2.67	4.35	6.63	9.24	11.1	12.8	15.1	16.7	20.5
6	.676	.872	1.24	1.64	2.20	3.45	5.35	7.84	10.6	12.6	14.4	16.8	18.5	22.5
7	.989	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.0	14.1	16.0	18.5	20.3	24.3
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.2	13.4	15.5	17.5	20.1	22.0	26.1
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.4	14.7	16.9	19.0	21.7	23.6	27.9
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.5	16.0	18.3	20.5	23.2	25.2	29.6
11	2.60	3.05	3.82	4.57	5.58	7.58	10.3	13.7	17.3	19.7	21.9	24.7	26.8	31.3
12	3.07	3.57	4.40	5.23	6.30	8.44	11.3	14.8	18.5	21.0	23.3	26.2	28.3	32.9
13	3.57	4.11	5.01	5.89	7.04	9.30	12.3	16.0	19.8	22.4	24.7	27.7	29.8	34.5
14	4.07	4.66	5.63	6.57	7.79	10.2	13.3	17.1	21.1	23.7	26.1	29.1	31.3	36.1
15	4.60	5.23	6.26	7.26	8.55	11.0	14.3	18.2	22.3	25.0	27.5	30.6	32.8	37.7
16	5.14	5.81	6.91	7.96	9.31	11.9	15.3	19.4	23.5	26.3	28.8	32.0	34.3	39.3
17	5.70	6.41	7.56	8.67	10.1	12.8	16.3	20.5	24.8	27.6	30.2	33.4	35.7	40.8
18	6.26	7.01	8.23	9.39	10.9	13.7	17.3	21.6	26.0	28.9	31.5	34.8	37.2	42.3
19	6.84	7.63	8.91	10.1	11.7	14.6	18.3	22.7	27.2	30.1	32.9	36.2	38.6	43.8
20	7.43	8.26	9.59	10.9	12.4	15.5	19.3	23.8	28.4	31.4	34.2	37.6	40.0	45.3
21	8.03	8.90	10.3	11.6	13.2	16.3	20.3	24.9	29.6	32.7	35.5	38.9	41.4	46.8
22	8.64	9.54	11.0	12.3	14.0	17.2	21.3	26.0	30.8	33.9	36.8	40.3	42.8	48.3
23	9.26	10.2	11.7	13.1	14.8	18.1	22.3	27.1	32.0	35.2	38.1	41.6	44.2	49.7
24	9.89	10.9	12.4	13.8	15.7	19.0	23.3	28.2	33.2	36.4	39.4	43.0	45.6	51.2
25	10.5	11.5	13.1	14.6	16.5	19.9	24.3	29.3	34.4	37.7	40.6	44.3	46.9	52.6
26	11.2	12.2	13.8	15.4	17.3	20.8	25.3	30.4	35.6	38.9	41.9	45.6	48.3	54.1
27	11.8	12.9	14.6	16.2	18.1	21.7	26.3	31.5	36.7	40.1	43.2	47.0	49.6	55.5
28	12.5	13.6	15.3	16.9	18.9	22.7	27.3	32.6	37.9	41.3	44.5	48.3	51.0	56.9
29	13.1	14.3	16.0	17.7	19.8	23.6	28.3	33.7	39.1	42.6	45.7	49.6	52.3	58.3
30	13.8	15.0	16.8	18.5	20.6	24.5	29.3	34.8	40.3	43.8	47.0	50.9	53.7	59.7
40	20.7	22.2	24.4	26.5	29.1	33.7	39.3	45.6	51.8	55.8	59.3	63.7	66.8	73.4
50	28.0	29.7	32.4	34.8	37.7	42.9	49.3	56.3	63.2	67.5	71.4	76.2	79.5	86.7
60	35.5	37.5	40.5	43.2	46.5	52.3	59.3	67.0	74.4	79.1	83.3	88.4	91.6	99.6
70	43.3	45.4	48.8	51.7	55.3	61.7	69.3	77.6	85.5	90.5	95.0	100.4	104.2	112.3
80	51.2	53.5	57.2	60.4	64.3	71.1	79.3	88.1	96.6	101.9	106.6	112.3	116.3	124.8
90	59.2	61.8	65.6	69.1	73.3	80.6	89.3	98.6	107.6	113.1	118.1	124.1	128.3	137.2
100	67.3	70.1	74.2	77.9	82.4	90.1	99.3	109.1	118.5	124.3	129.6	135.8	140.2	149.4

Standaard-normale verdeling

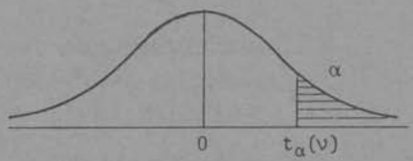


overschrijdingskansen:
waarden van $\alpha \cdot 10^4$ voor
$u_\alpha = 0{,}00(0{,}01)3{,}49$

met $\int_{u_\alpha}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx = \alpha$

u	0	1	2	3	4	5	6	7	8	9
0,0	5000	4960	4920	4880	4840	4801	4761	4721	4681	4641
0,1	4602	4562	4522	4483	4443	4404	4364	4325	4286	4247
0,2	4207	4168	4129	4090	4052	4013	3974	3936	3897	3859
0,3	3821	3783	3745	3707	3669	3632	3594	3557	3520	3483
0,4	3446	3409	3372	3336	3300	3264	3228	3192	3156	3121
0,5	3085	3050	3015	2981	2946	2912	2877	2843	2810	2776
0,6	2743	2709	2676	2643	2611	2578	2546	2514	2483	2451
0,7	2420	2389	2358	2327	2296	2266	2236	2206	2177	2148
0,8	2119	2090	2061	2033	2005	1977	1949	1922	1894	1867
0,9	1841	1814	1788	1762	1736	1711	1685	1660	1635	1611
1,0	1587	1562	1539	1515	1492	1469	1446	1423	1401	1379
1,1	1357	1335	1314	1292	1271	1251	1230	1210	1190	1170
1,2	1151	1131	1112	1093	1075	1056	1038	1020	1003	0985
1,3	0968	0951	0934	0918	0901	0885	0869	0853	0838	0823
1,4	0808	0793	0778	0764	0749	0735	0721	0708	0694	0681
1,5	0668	0655	0643	0630	0618	0606	0594	0582	0571	0559
1,6	0548	0537	0526	0516	0505	0495	0485	0475	0465	0455
1,7	0446	0436	0427	0418	0409	0401	0392	0384	0375	0367
1,8	0359	0351	0344	0336	0329	0322	0314	0307	0301	0294
1,9	0287	0281	0274	0268	0262	0256	0250	0244	0239	0233
2,0	0228	0222	0217	0212	0207	0202	0197	0192	0188	0183
2,1	0179	0174	0170	0166	0162	0158	0154	0150	0146	0143
2,2	0139	0136	0132	0129	0125	0122	0119	0116	0113	0110
2,3	0107	0104	0102	0099	0096	0094	0091	0089	0087	0084
2,4	0082	0080	0078	0075	0073	0071	0069	0068	0066	0064
2,5	0062	0060	0059	0057	0055	0054	0052	0051	0049	0048
2,6	0047	0045	0044	0043	0041	0040	0039	0038	0037	0036
2,7	0035	0034	0033	0032	0031	0030	0029	0028	0027	0026
2,8	0026	0025	0024	0023	0023	0022	0021	0021	0020	0019
2,9	0019	0018	0018	0017	0016	0016	0015	0015	0014	0014
3,0	0013	0013	0013	0012	0012	0011	0011	0011	0010	0010
3,1	0010	0009	0009	0009	0008	0008	0008	0008	0007	0007
3,2	0007	0007	0006	0006	0006	0006	0006	0005	0005	0005
3,3	0005	0005	0005	0004	0004	0004	0004	0004	0004	0003
3,4	0003	0003	0003	0003	0003	0003	0003	0003	0003	0002

Student-verdeling



ν \ α	0.40	0.25	0.10	0.05	0.025	0.010	0.005	0.0025	0.0010	0.0005
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	127.321	318.309	636.619
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	22.327	31.598
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.214	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.223	3.646	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.768
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797	3.090	3.467	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	0.254	0.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
120	0.254	0.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	0.253	0.674	1.282	1.645	1.960	2.326	2.576	2.807	3.090	3.290