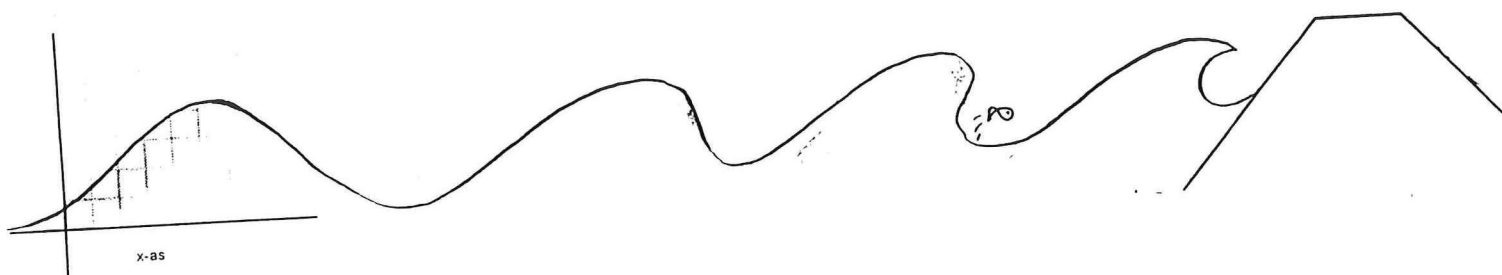




De invloed van statistische onzekerheid op de bepaling van dijkhoogtes



Karen A.H. Slikhuis

TU Delft
Faculteit der Civiele Techniek
Vakgroep Waterbouwkunde

Prof. drs. ir. J.K. Vrijling, TU Delft
ir. R.E. Jorissen, DWW
ir. P. v. Gelder

Prof. dr. ir. A.C.W.M. Vrouwenvelder, TU Delft
ir. W. Meermans

Delft, Oktober 1996

De invloed van statistische onzekerheid op de bepaling van dijkhoogtes

Karen A.H. Slijkhuis

TU Delft
Faculteit der Civiele Techniek
Vakgroep Waterbouwkunde

Prof. drs. ir. J.K. Vrijling, TU Delft
ir. R.E. Jorissen, DWW
ir. P. v. Gelder

Prof. dr. ir. A.C.W.M. Vrouwenvelder, TU Delft
ir. W. Meermans

Delft, Oktober 1996

Inhoudsopgave

Voorwoord	7
Samenvatting	9
Lijst van de afbeeldingen	11
Inleiding	13
De waterstanden	15
1.1 Inleiding	15
1.1.1 Een absoluut maximum	15
1.1.2 Overschrijdingslijnen	15
1.2 Selectie van de waterstanden	19
1.2.1 Inleiding	19
1.2.3 Bewerken van de waterstanden	21
1.2.4 Selectie op basis van jaarmaxima	21
1.2.5 Selectie op basis van de Peak-over-Threshold-methode	23
Schatters en verdelingen	29
2.1 Inleiding	33
2.2 Verdelingen	33
2.2.1 Inleiding	33
2.2.2 Verdelingen van de waterstanden	37
2.3 Methode der momenten	37
2.3.1 Inleiding	37
2.3.2 Bepaling van de momentenschatters	39
2.3.3 Onzekerheid van de momentenschatters	47
2.4 De Methode van de Maximum-Likelihood	47
2.4.1 Inleiding	47
2.4.2 Bepaling van een Maximum Likelihood-schatter	51
2.4.3 Onzekerheid van een Maximum Likelihood-schatter	55
2.5 Kleinste kwadraten methode	55
2.5.1 Inleiding	55
2.5.2 Bepaling van een kleinste-kwadraten-schatter	59
2.5.3 Onzekerheid van een kleinste-kwadraten-schatter	61
2.6 Vergelijken van de verschillende methoden	61
2.6.1 Inleiding	61
2.6.2 Zuiverheid en efficiëntie	63
2.6.3 Toepasbaarheid	63
Overschrijdingslijnen van waterstanden	65
3.1 Inleiding	65
3.1.1 Bepalen van een overschrijdingslijn	65
3.1.2 De invloed van statistische onzekerheid op een overschrijdingslijn	69
3.2 De overschrijdingslijnen van de jaarmaxima	69
3.2.1 Inleiding	69
3.2.2 Resultaten	75
3.3 De overschrijdingslijn van de reeks verkregen met de POT-methode	75
3.3.1 Inleiding	77
3.3.2 Resultaten	81
3.4 Vergelijking van de overschrijdingslijnen	81
3.4.1 Inleiding	81
3.4.2 Resultaten	81

De optimale dijkhoogte	
4.1 Inleiding	85
4.2 Het model	87
4.2.1 Inleiding	87
4.2.2 De investeringen	87
4.2.3 De schadeverwachting	89
4.2.4 De optimale dijkhoogte	91
4.3 De invloed van de onzekerheid van de verdelingsparameter	93
4.3.1 Inleiding	93
4.3.2 De invloed van de onzekerheid op de totale kosten	93
4.3.3 Resultaten	93
4.4 De invloed van de onzekerheid van de kosten van dijkverhoging en de schade	97
4.4.1 Inleiding	97
4.4.2 De invloed van de onzekerheid op de totale kosten	97
4.5 Kansdichtheid van de totale kosten	105
4.5.1 Inleiding	105
4.5.2 Kansdichtheidsfunctie van de totale kosten	105
Conclusies	
5.1 Conclusies	109
5.1.1 Inleiding	109
5.1.2 Hoofdconclusies	109
5.1.3 Aanvullende conclusies	109
5.2 Aanbevelingen	115
5.2.1 Inleiding	115
5.2.2 Aanbevelingen	115
Literatuurlijst	117
Bijlagen	
1 Jaarmaxima in Hoek van Holland in de periode 1887 - 1994	
2 Waterstanden in Hoek van Holland in de periode 1887-1994, geselecteerd met de POT-methode	
3 Bepaling van de parameters van een exponentiële verdeling met de Methode der Momenten	
4 Bepaling van de parameters van een exponentiële en een lognormale verdeling met de Methode van de Maximum Likelihood	
5 Pascal-programma om numeriek de Maximum-Likelihood-schatter te bepalen	
6 Bepaling van de standaardafwijking van de Maximum-Likelihood-schatter	
7 Bewijs dat de kleinste-kwadraten-schatter een onzuivere schatter is	
8 Resultaten van het numeriek fitten van een exponentiële verdeling op de 14 datareeksen van Ron Beem	
9 Afleiding van de invloed van statistische onzekerheid op de schatter van parameter A van de exponentiële verdeling	
10 Bepaling van de verwachtingswaarde van de contante waarde van de schade	

Voorwoord

Dit verslag is geschreven in het kader van mijn afstuderen aan de Technische Universiteit Delft, Faculteit der Civiele Techniek, vakgroep Waterbouwkunde. Het onderzoek is uitgevoerd bij de Dienst Weg- en Waterbouw in Delft.

Bij het onderzoek ben ik begeleid door Prof. drs. J.K. Vrijling, ir. P.H.A.J.M. van Gelder, ir. W. Meermans en prof. dr. ir. A.C.W.M. Vrouwenvelder. Ik wil hen hiervoor heel erg hartelijk bedanken.

Verder wil ik al die mensen op de achtste verdieping van wie ik de computer mocht gebruiken bedanken. Bijzondere dank gaat uit naar de "dames" voor alle ondersteuning, voor de zoethoutthee en de koekjes. Aan dit eindverslag gingen verscheidene concepten vooraf. Ter compensatie van al dat papiergebruik heb ik bij de DWW een boompje geplant.

Om mensen naar de presentatie van mijn afstuderen te lokken, had ik de eigenlijke titel van mijn afstuderen, "De invloed van statistische onzekerheid op de bepaling van dijkhoogtes" verandert in "Waar de blanke top der duinen...", naar het niet zo bekende tweede couplet van dat Oud-Hollandse lied.

"Waar de blanke top der duinen
Schittert in de zonnegloed,
Zijn de winden fel en hevig,
Telkens dreigt de stormvloed.

Want hoe hoog de dijk ook leit
Altijd blijft, onzekerheid.
Ach, hoe hoog wordt toch die piek ?
Lang leve de statistiek !"

Karen A.H. Slijkhuis,
Delft, Oktober 1996.

Samenvatting

Om waterstanden statistisch te kunnen analyseren moet gewerkt worden met een reeks homogene en onafhankelijke waarnemingen. Er zijn in dit onderzoek twee reeksen met waterstanden geselecteerd uit de waarnemingen die gedaan zijn in de periode van 1887-1994 in Hoek van Holland. De ene reeks is een reeks met jaarmaxima. Deze reeks waarnemingen is Gumbel- of lognormaal verdeeld. Voor de andere reeks zijn de waterstanden geselecteerd met de POT-methode. Aangenomen wordt dat deze waarnemingen exponentieel verdeeld zijn.

Uit de reeksen kunnen de overschrijdingslijnen van de waterstanden bepaald worden. Om de parameters van de verdelingsfunctie te schatten, worden in dit verslag drie schattingsmethoden beschreven. De Methode der Momenten, de Methode van de Maximum Likelihood en de Kleinste-Kwadraten-methode. Een schatter is een functie van willekeurige trekkingen uit een stochastische variabele en zal daarom altijd stochastisch zijn. Dit is statistische onzekerheid. Inzicht in de kansdichtheidsfunctie van de schatter is (soms) op analytische en (meestal) op numerieke wijze te verkrijgen.

De invloed van de statistische onzekerheid op de ligging van het 10^{-4} -kwantiel is op twee manieren te bepalen. Bij de ene methode wordt uit het gemiddelde en de standaardafwijking van de schatters het gemiddelde en de standaardafwijking van het 10^{-4} -kwantiel vastgesteld. Bij de andere methode wordt de onzekerheid van de schatter op Bayesiaanse wijze uit de verdelingsfunctie van de waterstanden geïntegreerd.

Met behulp van de overschrijdingslijnen kan de economisch optimale dijkhoogte bepaald worden als aangenomen wordt dat overstroom het enige faalmechanisme van de dijk is. Bij iedere dijkhoogte is er een kans op falen met als gevolg een ramp. In het econometrisch model is de dijkhoogte optimaal als de totale kosten minimaal zijn. De totale kosten zijn een som van de kosten van dijkverbetering en de gekapitaliseerde schadeverwachting.

De parameters in het econometrisch model zijn stochasten. Eerst wordt beschouwd wat de invloed van de statistische onzekerheid van de parameters van de verdelingsfunctie op de optimale dijkhoogte. Vervolgens wordt bekeken hoe de optimale dijkhoogte verandert als de investeringen en de schade stochastisch zijn.

Lijst van de afbeeldingen

- 1.1 Overschrijdingslijnen van de stormvloedstanden te Hoek van Holland, Deltaciel [2]
- 1.2 Definitie van opzet
- 1.3 Jaargemiddelden in Hoek van Holland
- 1.4 Voorbeeld van een DS-i-selectie met $i=4$.
- 1.5 Invloed van de keuze van de drempel van de hoogwateropzet op de overschrijdingslijn
- 1.6 Invloed van de keuze van de drempel van de hoogwaterstand op de overschrijdingslijn

- 2.1 Kansdichtheidsfunctie van de schatter van het gemiddelde gebaseerd op verschillende aantallen waarnemingen
- 2.2 Kansdichtheidsfunctie van de schatter van de standaardafwijking gebaseerd op verschillende aantallen waarnemingen
- 2.3 Bootstrap van de momentenschatters
- 2.4 Bootstrap van de Maximum-Likelihood-schatters
- 2.5 Gelineariseerde data en verdelingen
- 2.6 Niet-gelineariseerde data en verdelingen
- 2.7 Bootstrap van de gelineariseerde Kleinste-Kwadraten-schatters

- 3.1 Jaarmaxima met Gumbelverdeling
- 3.2 Jaarmaxima met lognormale verdeling
- 3.3 Waterstanden na selectie met de POT-methode met exponentiële verdeling
- 3.4 Invloed van onzekerheid rond parameter A van de exponentiële verdeling
- 3.5 Invloed van onzekerheid rond parameter B van de exponentiële verdeling
- 3.6 De overschrijdingslijnen
- 3.7 De gevoeligheid van de schattingsmethoden voor de keuze van de drempel van de hoogwaterstand

- 4.1 Dijkkringgebied Centraal Holland
- 4.2 Definitie van dijkverhoging
- 4.3 Kansdichtheidsfunctie van de totale schade
- 4.4 Economisch optimale dijkhoogte
- 4.5 De totale kosten met statistische onzekerheid
- 4.6 Kansdichtheidsfunctie van de totale schade, met normaal verdeelde te beschermen waarde W
- 4.7 Gemiddelde en standaardafwijking van de totale kosten
- 4.8 De totale kosten. Gemiddelde met $\frac{1}{2}$, 1 en $1\frac{1}{2}$ keer de standaardafwijking
- 4.9 Totale kosten. Gemiddelde en standaardafwijking voor verschillende varianties van de investeringen
- 4.10 Totale kosten. Gemiddelde en standaardafwijking voor verschillende varianties van de te beschermen waarde
- 4.11 Invloed van de variantie van de investeringen op de variantie van de totale schade
- 4.12 Invloed van de variantie van de te beschermen waarde op de variantie van de totale schade
- 4.13 De kansdichtheid van de kosten zonder schade, bij een dijkhoogte van NAP +5m.
- 4.14 De kansdichtheid van de schade bij een dijkhoogte van NAP +5m.
- 4.15 Totale kosten in geval van schade
- 4.16 Totale kosten in geval van schade met stochastische rente

Inleiding

Nederland is als land uniek in het feit dat zij voor haar voortbestaan afhankelijk is van dijken. Zonder die waterkeringen zou $\frac{2}{3}$ van Nederland regelmatig onder water staan en zou bijna de gehele bevolking moeten emigreren.

In het verleden werd bij de vaststelling van de hoogte van de zeedijken altijd uitgegaan van de tot dan toe bekende hoogste waterstanden. Het water bleek toch altijd weer hoger te komen. Verscheidene malen in haar geschiedenis werd Nederland getroffen door een watersnoodramp.

Na de ramp in 1953 werd de Delta commissie ingesteld. Alle dijken moesten verbeterd opdat een tweede ramp van dergelijke omvang voorkomen zou worden. Voortaan werd de hoogte van de dijken zoveel mogelijk wetenschappelijk vastgesteld. Aan de hand van de te verwachten waterstanden in de toekomst werden dijkverbeteringen vastgesteld, waarbij voor elk gebied een aannemelijke, economisch en sociaal verantwoorde mate van beveiliging kon worden verzekerd.

Het uitgangspunt hierbij was niet dat de hoogte van de nieuwe dijken nooit meer overschreden zou mogen worden, maar dat die hoogte een zeer kleine kans van overschrijding zou mogen hebben. De Delta commissie stelde als basispeil de waterstand met een overschrijdingsfrequentie van 1 maal per 10.000 jaar.

Al eerder had men ontdekt dat de hoogste waterstanden bij benadering een rechte lijn vormen als ze op logaritmisch papier uitgezet zijn tegen het aantal malen dat ze overschreden worden. Om het basispeil vast te stellen worden deze overschrijdingslijnen gebruikt. Een probleem van deze statistische benadering is dat het niet volkomen duidelijk is welke verdelingsfunctie de waterstanden hebben. Zelfs als dat wel bekend zou zijn, dan zijn de bijbehorende parameters om die functie weer te geven nog niet met zekerheid te geven.

Het is lastig om met behulp van zo'n honderd jaar waarnemingen een waterstand vast te stellen die maar 1 maal in de 10.000 jaar voorkomt. Het basispeil met een overschrijdingsfrequentie van 10^{-4} per jaar is niet met zekerheid te bepalen.

De Delta commissie bepaalde op mathematisch-economische wijze het ramppeil. Het ramppeil is het peil waarop de kosten van dijkverbetering aan de ene kant en de kosten van de schade bij inundatie aan de andere kant minimaal zijn. Aan de hand van het basis- en het ramppeil werd een ontwerppeil vastgesteld. Het ontwerppeil komt in de meeste gevallen overeen met het basispeil. Alleen in geval van economisch belangrijke of dichtbevolkte gebieden werd het basispeil naar boven bijgesteld.

Bij een waterstand gelijk aan het ontwerppeil moet er nog volledige veiligheid tegen doorbreken zijn. Als het water toch hoger komt, dan moet er voldoende reserve zijn om ook die stormvloedstand te kunnen keren. Als het water daarentegen hoger komt dan het ramppeil dan zal het hele achtergelegen gebied geïnundeerd worden en zal alles verloren gaan.

Het doel van dit afstudeeronderzoek is om meer inzicht te krijgen in de invloed van onzekerheden op de bepaling van het basis- en het ramppeil. Om te beginnen worden de statistische onzekerheden in een overschrijdingslijn beschouwd. Daarbij wordt gekeken hoe uit alle bekende waterstanden een statistisch te analyseren reeks verkregen kan worden. Gedetailleerd wordt ingegaan op het schatten van de parameters van een verdelingsfunctie. Uiteindelijk wordt op econometrische wijze de optimale dijkhoogte bepaald. Daarbij wordt niet alleen de onzekerheid van de overschrijdingskans, maar ook van de verschillende kosten meegenomen. We zullen zien wat de invloed is van de onzekerheid op de bepaling van dijkhoogtes.

De waterstanden

1.1 Inleiding

1.1.1 Een absoluut maximum

Door de eeuwen heen is men er altijd van uitgegaan dat hoge waterstanden een maximum hebben. Van de hoogste waterstand die ooit voorgekomen was, werd aangenomen dat het de maximale stand was. De dijken werden vervolgens net iets hoger gebouwd dan die hoogste waterstand. Als bij een storm het water toch over de dijk heen kwam, dan kon daaruit alleen geconcludeerd worden dat het maximum blijkbaar toch hoger lag.

Ook in deze eeuw is het niet gelukt om met een meer wetenschappelijke benadering de absoluut maximale waterstand te vinden. Een hoge waterstand bestaat voor een deel uit een astronomisch getij en voor een deel uit een windopzet. De hoogwaterstanden die veroorzaakt worden door enkel het getij zijn begrensd. Een extreem hoge waterstand ontstaat pas als bovendien sprake is van windopzet. Omdat de windopzet geen aantoonbare bovengrens heeft, is de maximale waterstand niet te bepalen.

De Deltacommissie en later het RIKZ hebben onderzoek gedaan naar de maximale hoogwaterstand. Om toch een fysische bovengrens voor de stormvloedstanden te vinden, hebben zij het gedrag van de waterstanden langs de Nederlandse kust onder zeer extreme meteorologische condities onderzocht. Een opzet wordt veroorzaakt door een storm en een storm door een depressie. Als de zwaarst mogelijke depressie bepaald zou kunnen worden, dan zou daaruit de hoogste stormopzet en de maximale waterstand volgen.

De Deltacommissie heeft in de jaren '50 bij haar onderzoek naar de bovenste grens voor de hevigheid van stormen elders op de oceaan of in het Noordzeegebied opgetreden depressies op zodanige wijze verplaatst, dat zij in de meest ongunstige positie kwamen te liggen ([2b]). Zij concludeerden dat het fysisch gezien niet ondenkbeeldig was dat zwaardere stormen dan de storm op 1 februari 1953 op de Noordzee zouden voorkomen.

Het RIKZ heeft zo'n 30 jaar later de depressies op vergelijkbare wijze in tijd en ruimte verschoven en daarbij de windsnelheden vergroot. ([1b]) Er was nogal wat kritiek op deze 'manipulaties', omdat ze niet duidelijk te beargumenteren waren. De conclusie was dat een representatieve superstorm voor de Nederlandse kust niet te bepalen is. Een storm wordt bepaald door verscheidene meteorologische factoren, die stochastisch van aard zijn, maar waarvan de statistiek onvoldoende bekend is.

1.1.2 Overschrijdingslijnen

Al in 1939 had Wemelsfelder een wetmatigheid in het optreden van hoogwaterstanden ontdekt. De hoogste waterstanden op logaritmisch papier uitgezet tegen het aantal malen dat ze overschreden worden, vormen een rechte lijn. Zie ook afbeelding 1.1. Als aangenomen wordt dat die trend doorzet, is een absoluut maximum voor een stormvloedstand niet aan te geven. Iedere waterstand heeft dan een kans op voorkomen, hoe klein ook.

De Deltacommissie baseerde de ontwerpen van de dijken op deze zogenaamde overschrijdingslijnen. Als norm stelde men een overschrijdingsfrequentie van 1 maal in de 10.000 jaar voor de dijken van Centraal Holland. Om dergelijke lage frequenties van voorkomen te bepalen dient men de overschrijdingsfrequentielijn te extrapoleren. Op hoe meer waarnemingen de overschrijdingslijn gebaseerd is, des te nauwkeuriger die extrapolatie kan plaatsvinden.

Halverwege de jaren '80 was de beschikbare reeks waarnemingen aanzienlijk langer en bovendien de kennis inzake overschrijdingslijnen groter. In het rapport Basispeilen staat het hernieuwd onderzoek naar de overschrijdingslijnen beschreven. In het rapport is de norm, zoals gesteld door de Deltacommissie overgenomen.

1.2 Selectie van de waterstanden

1.2.1 Inleiding

Bij dit afstudeeronderzoek zal de overschrijdingslijn van de waterstanden opnieuw vastgesteld worden op basis van de waterstanden die totnogtoe in Hoek van Holland zijn waargenomen. Bekeken zal worden of de nauwkeurigheid waarmee die overschrijdingslijn vastgesteld kan worden van invloed is op de maatgevende hoogwaterstand.

Men heeft sinds 1887 bijna zonder onderbreking de hoogwaterstanden gemeten en geregistreerd. Dat betekent dat in de beschikbare periode, van 1887 tot 1994, een kleine 80.000 hoogwaterstanden waargenomen zijn. Dat enorme aantal waarnemingen doet vermoeden dat het niet moeilijk zal zijn om daarmee een nauwkeurige overschrijdingslijn van de waterstanden te bepalen.

Aan het op statistische wijze analyseren van de hoge waterstanden zijn echter twee voorwaarden verbonden. De te gebruiken gegevens dienen *homogeen* en *onafhankelijk* van elkaar te zijn. Als waarnemingen uit een zelfde kansverdeling komen, zijn ze homogeen. Ze zijn onafhankelijk te noemen als er geen relatie is tussen de opvolgende waarden. Als een statistische analyse, die bedoeld is voor onafhankelijke waarden, toegepast wordt op afhankelijke waarden, dan worden te lage extreme waarden gevonden (zie ook [1c], blz.33).

Om aan de voorwaarden van homogeniteit en onafhankelijkheid te voldoen, worden de beschikbare waterstanden geselecteerd. De selectie zal echter geen volkomen zekerheid kunnen geven dat aan de voorwaarden zal worden voldaan. In dit hoofdstuk zullen twee mogelijkheden om de waterstanden te selecteren beschouwd worden.

1.2.2 De waterstand als de te verwerken grootheid

Voordat overgegaan kan worden tot het homogeen en onafhankelijk maken van de waarnemingen, moet beslist worden welke grootheid geanalyseerd gaat worden. Een hoge waterstand bestaat uit twee delen. Het astronomisch getij, een deterministisch gedeelte, en de windopzet, een stochastisch gedeelte. Opzet is het verschil tussen de waargenomen waterstand en het astronomisch getij op een zelfde tijdstip. De hoogwateropzet (HW-opzet) is het verschil tussen een waargenomen hoogwaterstand en het bijbehorende astronomisch hoogwater. De definitie van opzet en HW-opzet is weergegeven in afbeelding 1.2. Door het RIKZ is gekeken naar de HW-opzet, omdat die opzet niet gevoelig is voor vervroeging of verlating van het getij door weersinvloeden.

Voor de statistische analyse van waterstanden is het een nadeel dat ze voor een groot gedeelte uit het deterministische astronomisch getij bestaan. De HW-opzetten zijn wel puur stochastisch en voldoen, als ze geselecteerd zijn, aan de eis van onafhankelijkheid en homogeniteit.

De voorkeur gaat er echter naar uit om de waterstanden en niet de HW-opzetten statistisch te analyseren. De waterstand heeft rechtstreeks betrekking op het vraagstuk. Het is de waargenomen en de gezochte grootheid en dus neemt de kans op fouten niet onnodig toe als de waterstand verwerkt wordt.

Verder bestaan juist hoge waterstanden voor een relatief groot deel uit de stochastische windopzet. Indien de analyse gebaseerd zal worden op uitsluitend hoge waterstanden zullen die waterstanden als stochastisch beschouwd mogen worden.

1.2.3 Bewerken van de waterstanden

Om de waterstanden te kunnen vergelijken zonder invloed van zeespiegelstijging, grote waterbouwkundige ingrepen, klimatologische veranderingen e.d., zijn alle waterstanden herleid naar de toestand van 1994.

De trendlijn van de waterstanden is door het RIKZ bepaald door alle jaargemiddelden uit te zetten tegen het jaar van voorkomen. [1c, §4.5] In afbeelding 1.3 wordt het verloop van het jaargemiddelde van de hoogwaterstand weergegeven met een lichte lijn. Rond 1965 is een sprong in de trendlijn te zien. De oorzaak daarvan is waarschijnlijk de aanleg van de Maasvlakte. Daarom heeft het RIKZ twee trendlijnen bepaald. Eén voor en één na de trendbreuk.

In een zeker jaar J kan de gecorrigeerde waterstand W_c gevonden worden door de volgende bewerking van de werkelijke waterstand W_j :

$$\begin{array}{ll} \text{Van 1887-1964} & : \quad W_c = 0.12(1994-J)/100 + 0.158 + W_j \\ \text{In 1965} & : \quad W_c = 0.14 + W_j \\ \text{Van 1966-1994} & : \quad W_c = 0.53(1994-J)/100 + W_j \end{array}$$

J : jaar waarvoor de correctie geldt
 W_j : oorspronkelijke waterstand in m +NAP
 W_c : gecorrigeerde waterstand

In afbeelding 1.3 zijn met een donkere lijn de gecorrigeerde jaargemiddelden weergegeven.

De stijging van de gemiddelde waterstand door de jaren heen, wordt niet meegenomen in de overschrijdingslijn, maar als afzonderlijke grootheid bij het ontwerp van de dijk. Na de trendcorrectie van de waarnemingen wordt in dit verslag de invloed van zeespiegelstijging e.d. buiten beschouwing gelaten.

1.2.4 Selectie op basis van jaarmaxima

Om op statistische wijze de maatgevende waterstand te mogen bepalen, dienen de waarnemingen onafhankelijk en homogeen te zijn. Er zijn twee manieren om de waterstanden te selecteren.

De eerste methode selecteert de hoogste waterstand van ieder jaar. Jaarmaxima kunnen beschouwd worden als onderling onafhankelijke waarnemingen, omdat elke waarneming voortkomt uit één meteorologische jaarcyclus. Daar kan de extreme waarde theorie op toegepast worden. Bij selectie op basis van jaarmaxima wordt volkomen objectief een datareeks verkregen.

In Hoek van Holland zijn waterstanden beschikbaar vanaf 1887 tot aan 1994. Selectie op basis van jaarmaxima geeft een reeks met 108 onafhankelijke waarnemingen met een extreme waarde verdeling. In bijlage 1 staan de 108 geselecteerde jaarmaxima.

1.2.5 Selectie op basis van de Peak-over-Threshold-methode

Meer waarnemingen geven een nauwkeuriger schatting van de overschrijdingslijn. In 108 jaar zijn er slechts 108 jaarmaxima, maar het is een kleine 80.000 keer hoogwater geweest. Het lijkt dus dat, door gebruik te maken van alle hoogwaterstanden, een nauwkeuriger overschrijdingslijn vastgesteld zou kunnen worden.

Een tweede mogelijke selectie om aan de voorwaarden van homogeniteit en onafhankelijkheid te voldoen, is de selectie van waterstanden met de Peak-Over-Threshold-methode. Dat betekent letterlijk dat enkel waarnemingen boven een nader te bepalen drempel geselecteerd worden. De methode houdt daarnaast echter in dat de waterstanden 'gedund' worden om onafhankelijkheid te bewerkstelligen. Verder wordt de reeks beperkt tot waarnemingen in een stormseizoen om te zorgen dat de waarnemingen uit een zelfde verdeling komen. De POT-methode zal geen volkomen zekerheid kunnen geven dat aan de voorwaarden van homogeniteit en onafhankelijkheid zal worden voldaan.

Zowel door de Deltacommissie [2] als in het rapport Basispeilen [1] wordt met op dergelijke wijze geselecteerde waterstanden gewerkt. De selectiecriteria die door het RIKZ in [1] besproken zijn, zullen ook hier toegepast worden. Uiteindelijk zal de invloed van de verschillende criteria op de overschrijdingslijn en de extreme waarden besproken worden. In bijlage 2 staan alle waterstanden die met de Peak-Over-Threshold-methode zijn geselecteerd. De selectiecriteria van de toegepaste Peak-over-Threshold-methode zullen hier achtereenvolgend besproken worden.

• Stormseizoen

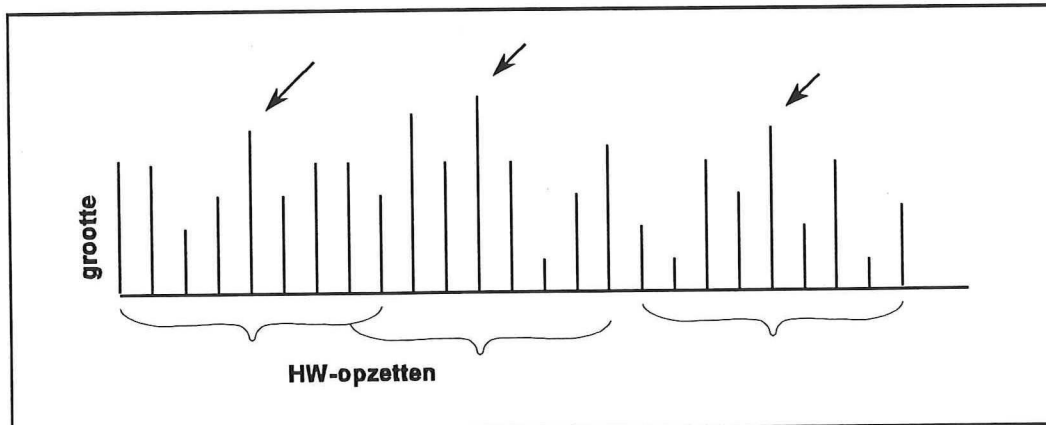
Gedurende een jaar verschilt de mate van voorkomen en de hoogte van de HW-opzetten. Indien aangenomen wordt dat opzetten uit eenzelfde verdeling bij benadering eenzelfde frequentie en gemiddelde grootte zullen hebben, dient een zogenaamd stormseizoen gekozen te worden. In het gekozen seizoen is zowel het aantal opzetten als de gemiddelde grootte constant. Gekozen is voor een stormseizoen van 1 oktober tot en met 15 maart. Alleen waterstanden uit die periode worden in de analyse meegenomen.

• De dunningsselectie

Iedere opzet ontstaat door een storm. De storm ontstaat op zijn beurt weer door een depressie. Eén depressie kan meer stormen veroorzaken. Eén storm kan meer HW-opzetten teweegbrengen. HW-opzetten kunnen daardoor gecorreleerd kunnen zijn. Opzetten zullen bij benadering onafhankelijk zijn als ze in de tijd ver genoeg uit elkaar liggen. In [1] is gebruik gemaakt van de zogenaamde "DS-i selectie".

Bij de DS-i-selectie moet iedere geselecteerde HW-opzet groter zijn dan i voorafgaande en i volgende waarnemingen. Hoe groter i , hoe kleiner de afhankelijkheid van de geselecteerde waarnemingen, maar ook hoe minder waarnemingen. Bij de bepaling van i zal een compromis gevonden moeten worden tussen de wens van onafhankelijkheid aan de ene kant en aan de andere kant de wens om zo veel mogelijk waarnemingen behouden.

Door het RIKZ is gekozen voor $i=4$. Afbeelding 1.4 geeft ter verduidelijking aan welke HW-opzetten uit een reeks geselecteerd worden bij een dunningsselectie met $i=4$. Voor de bepaling van i zie [1c].



Afbeelding 1.4 Voorbeeld van een DS-i-selectie, met $i=4$.

• Drempel voor de HW-opzet

Tijdreeksen van HW-opzetten hebben een zekere ruis. Door onnauwkeurigheden in de grootheden die de grootte van de opzetten bepalen, zou het kunnen gebeuren dat er een opzet lijkt te zijn, die er op dat moment niet is. Om ervoor te zorgen dat alleen opzetten meegenomen worden, waarbij sprake is van opwaaiing, wordt een drempel voor de HW-opzet ingevoerd.

De drempel voor HW-opzet wordt gekozen op ongeveer twee maal de standaardafwijking die geschat wordt uit de standaardafwijkingen van de meting van het hoogwater, ($\sigma_{HW} \approx 25$ mm), de luchtdruk, ($\sigma_l \approx 100$ mm), en het astronomisch getij, ($\sigma_a \approx 100$ mm). Als aangenomen wordt dat de grootheden onafhankelijk zijn, dan geldt:

$$\rightarrow \sigma_o = \sqrt{(\sigma_{HW}^2 + \sigma_l^2 + \sigma_a^2)} = 0.144 \text{ m}$$

Afgerond wordt de minimale drempel dan : $2\sigma_o = 0.30\text{m}$.

• Drempel voor de waterstand

Verhoudingsgewijs bestaat een lage waterstand voor een groter deel uit het deterministische astronomisch getij dan een hoge waterstand. De verdelingsfunctie van de lage waterstanden is daardoor anders dan die van de hoge waterstanden. Om te bewerkstelligen dat de te analyseren hoge waterstanden uit dezelfde verdelingsfunctie komen, wordt een drempel voor de waterstanden gekozen.

De gekozen ondergrens is de som van het hoogste astronomische getij uit de geselecteerde reeks (NAP +1.52m) en de ondergrens van de HW-opzet (0.30 m). Afgerond komt hiermee de drempel voor de hoge waterstanden in Hoek van Holland op NAP +1.85m.

Invloed van de verschillende selectiecriteria van de POT-methode

Na selectie met de POT-methode zijn van de 80.000 hoogwaterstanden er slechts 530 overgebleven die voldeden aan de volgende criteria:

- Waarnemingen uit het stormseizoen van 1 Oktober tot 15 Maart.
- Dunningselectie (DS-4) van de HW-opzetten.
- HW-opzet ≥ 0.30 m.
- Hoogwaterstanden $\geq$ NAP +1.85m.

Deze criteria zijn vrij arbitrair gekozen, daarom worden hier de gevolgen van de verschillende keuzes besproken.

De beperking tot een **stormseizoen** vermindert het aantal waargenomen hoogwaterstanden. Een extreem hoge stormvloed die in een ander deel van het jaar valt, wordt niet geselecteerd. In de 108 jaar viel het jaarmaximum 6 keer buiten het stormseizoen.

De Deltacommissie definieerde ook een stormseizoen. Zij kozen voor de beperking van de waarnemingen tot de waterstanden uit de wintermaanden. Het werd een stormseizoen van 1 november tot 1 februari. Als de grens op 31 januari gelegd zou worden, dan zou de hoogste waterstand tot dan toe, van 1 februari 1953, buiten de selectie vallen.

De **DS-i-selectie** dunt de waarnemingen uit. Hoe groter i is, hoe minder waarnemingen geselecteerd zullen worden. De gevolgen van een dunningselectie zijn vergelijkbaar met die van een drempel voor hoogwateropzetten. Daarom wordt voor uitleg verwezen naar de volgende alinea. In het kort kan gesteld worden dat na een dunningselectie een waterstand van bijvoorbeeld NAP +1.85m een lagere frequentie van voorkomen heeft en de bijbehorende overschrijdingslijn steiler zal lopen.

De keuze van de drempels is een moeilijk te onderbouwen selectiecriteria. Als eerste wordt de keuze van de **drempel van de hoogwateropzet** beschouwd. De drempel voor de HW-opzetten werkt als de dunningselectie. Als de drempel hoger gekozen wordt dan zullen, vooral bij de lagere hoogwaterstanden, minder waarnemingen geselecteerd worden.

In afbeelding 1.5 zijn drie verschillende overschrijdingsfrequentielijnen weergegeven. Iedere lijn is gebaseerd op waterstanden met een andere drempel van de HW-opzet. Alle waarnemingen hebben dezelfde drempel voor de HW-stand, NAP +1.85m. Het is niet duidelijk te zien, maar als de drempel voor de HW-opzet hoger gekozen wordt, dan loopt de overschrijdingslijn net iets steiler. Een hogere drempel geeft een grotere waarde voor het 10^{-4} -kwantiel.

Het omgekeerde geldt voor de **drempel van de hoogwaterstanden**. Als de drempel hoger gekozen wordt, dan zullen alleen de lagere waterstanden tot aan de drempel uit de reeks verdwijnen. Een waterstand van bijvoorbeeld NAP +2.05m zal nog steeds dezelfde overschrijdingsfrequentie hebben. In afbeelding 1.6 zijn drie overschrijdingsfrequentielijnen weergegeven van waterstanden, die allemaal dezelfde drempel voor de HW-opzet hebben, maar verschillende drempels voor de HW-standen. Hier is duidelijk te zien dat bij een hogere drempel voor de waterstand een flauwer lopende overschrijdingsfrequentielijn hoort. Een hogere drempel geeft dus een lagere waarde voor het 10^{-4} -kwantiel.

Schatters en verdelingen

2.1 Inleiding

In het eerste hoofdstuk is besproken dat hoogwaterstanden niet deterministisch zijn, maar dat het voorkomen ervan onvoorspelbaar en willekeurig is. Deze onvoorspelbaarheid van de waterstanden wordt intrinsieke onzekerheid genoemd.

Een maatgevende waterstand dient dus op statistische wijze bepaald te worden. In hoofdstuk 1 is beschreven hoe uit de waargenomen waterstanden bij Hoek van Holland, twee reeksen met onafhankelijke en homogene waarnemingen zijn verkregen. De ene reeks is geselecteerd op basis van de jaarmaxima. De andere reeks is verkregen met de Peak-Over-Threshold-methode. Om de maatgevende hoogwaterstand te kunnen vaststellen, zal eerst van beide reeksen de overschrijdingslijn bepaald moeten worden.

In dit onderzoek wordt aangenomen dat de verdelingsfuncties van de twee reeksen waterstanden bekend zijn. Het voorkomen van hoogwaterstanden is een fysisch proces waarvan niet bekend is hoe de kansverdeling er uit ziet. Deze onzekerheid heet modelonzekerheid. De modelonzekerheid wordt in dit onderzoek buiten beschouwing gelaten.

De Gumbel- en de lognormale verdeling zijn gekozen als verdelingsfuncties van de reeks jaarmaxima. De exponentiële verdeling wordt beschouwd als de verdeling van de waterstanden die geselecteerd zijn met de POT-methode. Verdelingsfuncties zullen in § 2.2 meer in detail besproken worden.

2.1.1 Bepalen van de schatters

Wanneer de verdelingsfunctie bekend is, moeten de parameters van de verdeling geschat worden. Daarvoor zijn verscheidene methoden beschikbaar. De methoden die in dit hoofdstuk behandeld zullen worden zijn:

- Methode der Momenten (§ 2.3)
- Methode van de Maximum Likelihood (§ 2.4)
- Kleinste kwadraten methode (§ 2.5)

Van de laatste methode zal zowel lineaire, als niet-lineaire regressie besproken worden. Na de bespreking zullen in § 2.6 de afzonderlijke methoden vergeleken worden.

2.1.2 Onzekerheid van de schatters

Een parameter van een verdeling wordt geschat uit een reeks met waarnemingen. De waarnemingen worden opgevat als onafhankelijke trekkingen uit de stochastische variabele 'hoogwaterstanden'. Een schatter is een functie van deze willekeurige trekkingen uit de stochastische variabele. Een schatter zal daarom niet deterministisch zijn, maar ook stochastisch. De verdeling van de schatter volgt uit het functievoorschrift waarmee de schatter uit de waarnemingen kan worden bepaald, het aantal waarnemingen en de verdeling van de waarnemingen. Het niet exact kunnen schatten van de verdelingsparameters wordt statistische onzekerheid genoemd. De statistische onzekerheid wordt met de kansdichtheid van de schatters gekarakteriseerd.

Ter onderscheiding van de werkelijke waarde van een parameter θ , wordt een schatter van een parameter weergegeven met een dakje. Zo is bijvoorbeeld $\hat{\theta}$ de schatter van θ . Schatters zullen zelden of nooit gelijk zijn aan de werkelijke, gezochte parameter van de verdeling. Pas bij een groot aantal waarnemingen nadert de schatter de gezochte waarde van de parameter.

Een goede schatter is een zuivere en efficiënte schatter. De maatstaf voor zuiverheid is de verwachtingswaarde van het verschil tussen de schatter en de werkelijke parameter.

$$\text{Zuiverheid} = E[\hat{\theta} - \theta].$$

Bij een zuivere schatter zal die verwachtingswaarde gelijk aan 0 zijn.

Efficiëntie is gedefinieerd als de variantie van de schatter ten opzichte van de werkelijke parameter.

$$\text{Efficiëntie} = E[(\hat{\theta} - \theta)^2]$$

Als de standaardafwijking kleiner is, dan is de spreiding van de parameter geringer en de efficiëntie groter.

Er zijn twee mogelijkheden om inzicht te krijgen in de statistische onzekerheid.

Bij de ene methode wordt inzicht verkregen door analytisch de kansdichtheidsfunctie van de schatter te bepalen. Deze methode is alleen toepasbaar als het functievoorschrift en de kansdichtheidsfunctie van de waarnemingen een algebraïsche oplossing toelaten. Vaak is dit niet het geval.

De tweede methode om inzicht te krijgen in statistische onzekerheid is een numerieke methode die wél op alle schatters en alle verdelingen toepasbaar is. Deze methode is het 'bootstrappen'. Bij het bootstrappen worden uit de beschikbare reeks met waarnemingen nieuwe datasets gegenereerd. Uit een oorspronkelijke reeks met waarnemingen worden willekeurig, met teruglegging, waarden getrokken, totdat een nieuwe dataset is ontstaan met precies hetzelfde aantal waarnemingen. Door voor iedere nieuwe dataset de parameters van de verdeling te schatten, is de kansdichtheidsfunctie van de afzonderlijke schatters vast te stellen.

In dit hoofdstuk zal waar mogelijk met beide methoden naar de onzekerheid van de schatters gekeken worden.

2.2 Verdelingen

2.2.1 Inleiding

De verdelingsfunctie $F(x)$ geeft de kans weer dat een stochastische variabele een waarde x onderschrijdt. Aangezien een kans op onderschrijden nooit groter dan 1 of kleiner dan 0 kan worden, zal een verdelingsfunctie altijd een tussen 0 en 1 liggen.

De verdelingsfunctie gedifferentieerd naar x geeft de kansdichtheidsfunctie $f(x)$. De oppervlakte van de kansdichtheidsfunctie over een bandbreedte Δx geeft de kans weer dat de variabele een waarde in die marge aan zal nemen. De oppervlakte onder de gehele kansdichtheidsfunctie is altijd gelijk aan 1.

In het geval van de waterstanden in Hoek van Holland zijn we geïnteresseerd in de kans dat een bepaalde hoogte x overschreden zal worden. Daarbij zal dus gebruik gemaakt worden van $1-F(x)$.

2.2.2 Verdelingen van de waterstanden

De meest eenvoudige en meest toegepaste verdeling van extreme waarden is de exponentiële verdeling. In navolging van Van Dantzig en de Deltacommissie worden de hoogwaterstanden die geselecteerd zijn met de Peak-Over-Threshold-methode in dit onderzoek als exponentieel verdeeld beschouwd.

De verdelingsfunctie van de exponentiële verdeling met twee parameters, A en B wordt gegeven door :

$$F(x) = 1 - \exp^{-\frac{(x-A)}{B}} \quad \text{met } x > A \text{ en } B > 0$$

Door de verdelingsfunctie naar x te differentiëren wordt de kansdichtheidsfunctie verkregen:

$$f(x) = \frac{1}{B} \exp^{-\frac{(x-A)}{B}}$$

Omdat in dit verslag ter illustratie regelmatig gebruik wordt gemaakt van een nog eenvoudiger variant, wordt hier voor de volledigheid ook de kansdichtheids- en de verdelingsfunctie van de exponentiële verdeling met één parameter λ gegeven :

$$F(x) = 1 - \exp^{-\lambda x}$$

$$f(x) = \lambda \cdot \exp^{-\lambda x}$$

De waterstanden die geselecteerd zijn op basis van de jaarmaxima worden verondersteld Gumbel- of lognormaal verdeeld te zijn.

De Gumbelverdeling heeft de volgende verdelingsfunctie:

$$F(x) = e^{-e^{-\frac{(x-A)}{B}}}$$

De verdelingsfunctie naar x gedifferentieerd is de kansdichtheidsfunctie:

$$f(x) = \frac{1}{B} e^{-\left[-\frac{(x-A)}{B}\right] - e^{-\frac{(x-A)}{B}}}$$

Als een stochast x lognormaal verdeeld is, dan is de stochast $y = \ln(x)$ normaal verdeeld. De lognormale verdeling kan met een transformatie uit de, nog te bespreken, normale verdeling verkregen worden. De kansdichtheidsfunctie van de lognormale verdeling met twee parameters, A en B is als volgt gedefinieerd :

$$f(x) = \frac{1}{\sqrt{2\pi}Bx} \exp \frac{-(\ln(x)-A)^2}{2B^2}$$

De verdelingsfunctie die bij deze kansdichtheidsfunctie hoort is:

$$F(x) = \int_0^x \frac{1}{\sqrt{2\pi}Bx} \exp \frac{-(\ln(x)-A)^2}{2B^2} dx$$

Deze integraal kan niet analytisch opgelost worden.

De totnogtoe beschreven verdelingen kunnen gebruikt worden om de intrinsieke onzekerheid van de hoogwaterstanden te beschrijven. Een andere vorm van onzekerheid is de onzekerheid van de schatters (zie §2.1.2) oftewel statistische onzekerheid. Zoals later in dit verslag besproken zal worden, is voor beschrijving van die onzekerheid bij benadering de normale verdeling geschikt.

De normale verdeling of Gaussische verdeling heeft de volgende kansdichtheidsfunctie :

$$f(x) = \frac{1}{B\sqrt{2\pi}} \exp \frac{-(x-A)^2}{2B^2}$$

Het gemiddelde van de verdeling is A en de standaardafwijking is B .

De bijbehorende verdelingsfunctie is :

$$F(x) = \int_{-\infty}^x \frac{1}{B\sqrt{2\pi}} \exp \frac{-(x-A)^2}{2B^2} dx$$

Net zoals de lognormale verdeling, die van de normale is afgeleid, kan ook deze integraal niet analytisch opgelost worden.

2.3 Methode der momenten

2.3.1 Inleiding

Bij de methode der momenten wordt er van uitgegaan dat de momenten van een dataset gelijk zullen zijn aan de momenten van de bijbehorende verdelingsfunctie. De parameters van de verdeling kunnen dus geschat worden door de momenten van zowel de dataset als de verdelingsfunctie te bepalen en aan elkaar gelijk te stellen. Het aantal momenten dat bepaald dient te worden hangt samen met het aantal te schatten parameters.

2.3.2 Bepaling van de momentenschatters

Eerst zal beschreven worden hoe de momenten van een dataset bepaald kunnen worden en vervolgens wordt de bepaling van de momenten van een verdeling besproken.

- **Momenten van een dataset**

De definitie van het k^e moment (m_k) van de waarnemingen is:

$$m_k = E(x_i - \bar{x})^k = \frac{\sum_{i=1}^N (x_i - \bar{x})^k}{N}$$

Waarin $\bar{x}$ het gemiddelde van de waarnemingen is.

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

Het gemiddelde $\bar{x}$ is gelijk aan het eerste moment, m_1 . Het tweede moment van de dataset, m_2 , is de variantie of de gekwadrateerde standaardafwijking van de waarnemingen, S^2 .

$$S^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$$

Dit is te vereenvoudigen tot

$$S^2 = \frac{1}{N} \left[\sum_{i=1}^N x_i^2 - 2\bar{x} \cdot \sum_{i=1}^N x_i + N \cdot \bar{x} \cdot \bar{x} \right] \rightarrow S^2 = \frac{1}{N} \left[\sum_{i=1}^N x_i^2 \right] - \bar{x}^2$$

Aangezien de meest gebruikte verdelingen slechts twee parameters hebben, voldoen dit eerste en tweede moment in het merendeel van de gevallen.

Mochten meer gegevens nodig zijn om de parameters van de verdeling te kunnen schatten, dan kan het derde moment, m_3 , de scheefheid, of zelfs het vierde moment, m_4 , de gepiektetheid, alsnog bepaald worden. Ook die momenten zijn vrij eenvoudig te bepalen volgens bovenstaande definitie.

- **Momenten van een verdelingsfunctie**

De verwachtingswaarde van een variabele x , gedefinieerd door een continue verdeling $f(x)$, is:

$$E[X] = \int_{-\infty}^{\infty} x \cdot f(x) dx$$

Analoog hieraan geldt voor het k^e moment $E[X^k]$:

$$E[X^k] = \int_{-\infty}^{\infty} x^k \cdot f(x) dx$$

Het k^e centrale moment, m_k , van x wordt gedefinieerd door:

$$m_k = E(x - m_x)^k = \int_{-\infty}^{\infty} (x - m_x)^k \cdot f(x) dx$$

Met $m_x = E[x]$. Het eerste centrale moment is dus altijd gelijk aan 0.

Het tweede centrale moment, m_2 , is de variantie, geschreven als $\text{Var}[x]$ of σ_x^2 . De variantie geeft de spreiding van de stochast X rond de verwachtingswaarde $E[X]$ weer. Eenvoudig is aan te tonen dat de variantie ook geschreven kan worden als:

$$\text{Var}[x] = E[(x)^2] - (E[x])^2$$

Om de werking te illustreren, worden in bijlage 3 de parameters van de exponentiële verdeling met één en met twee parameters bepaald volgens de methode der momenten.

De uitkomst voor de verdelingen die in dit verslag toegepast worden, is in tabel 2.1 weergegeven.

Schatter	Methode der Momenten	
exponentieel $\hat{\lambda}$	$\hat{\lambda} = 1/\bar{x}$	
exponentieel $\hat{A}, \hat{B}$	$\hat{A} = \bar{x} - \hat{B}$	$\hat{B} = \sqrt{\frac{1}{N} \sum x_i^2 - \bar{x}^2}$
Gumbel, $\hat{A}, \hat{B}$	$\hat{A} = (\bar{x} - \gamma \hat{B})$	$\hat{B} = \frac{\sqrt{6} \sqrt{\frac{1}{N} \sum x_i^2 - \bar{x}^2}}{\pi}$
lognormaal, $\hat{A}, \hat{B}$	$\hat{A} = \ln(\bar{x}) - \frac{1}{2} \ln\left(\frac{\sum x_i^2}{N \bar{x}^2}\right)$	$\hat{B} = \sqrt{\ln\left[\frac{\sum x_i^2}{N \bar{x}^2}\right]}$
normaal, $\hat{A}, \hat{B}$	$\hat{A} = \bar{x}$	$\hat{B} = \sqrt{\frac{1}{N} \sum x_i^2 - \bar{x}^2}$

Tabel 2.1.

In deze tabel is $\bar{x}$ geschreven, waar het gemiddelde van de waarnemingen wordt bedoeld.
 $\gamma = 0.5772$, het getal van Euler.

2.3.3 Onzekerheid van de momentenschatters

Hoewel de werkelijke, gezochte parameters van een verdelingsfunctie constanten zijn, zijn de schatters van de parameters stochasten, zoals in §2.1.2 werd uiteengezet. De schatter zal zelden gelijk zijn aan de werkelijke waarde van de parameter.

Iedere momentenschatter bestaat uit een functie van de schatting van het gemiddelde en/of de standaardafwijking van de waarnemingen. Om inzicht te krijgen in de kansdichtheidsfunctie van de schatters, wordt gekeken naar de verdeling van de schatter van het gemiddelde en van de standaardafwijking.

Voor de bepaling van zowel het gemiddelde als de standaardafwijking van de waarnemingen worden de willekeurige en onafhankelijke waarden, X_i , van de stochast X gesommeerd. De som van een groot aantal onafhankelijke waarden is normaal verdeeld. Dit wordt de centrale limiet stelling genoemd. De schatters van het gemiddelde en de standaardafwijking van een verdeling zullen daarom bij benadering normaal verdeeld zijn. Bij een klein aantal waarnemingen zal de schatter minder nauwkeurig te bepalen zijn dan bij een groot aantal waarnemingen.

Om inzicht te krijgen in de onzekerheid, wordt het gemiddelde en de standaardafwijking van de schatters vastgesteld. (Zie ook [7], blz. 377e.v.). Om te beginnen zal de verwachtingswaarde en de variantie van de schatter van het gemiddelde besproken worden.

De verwachtingswaarde van het gemiddelde $\bar{x}$ van N onafhankelijke waarnemingen met dezelfde verdeling wordt gegeven door

$$E[\bar{X}] = E\left[\frac{1}{N} \sum X_i\right] = \frac{1}{N} \sum E[X_i] = \frac{1}{N} \sum \mu_x = \frac{N \cdot \mu_x}{N} = \mu_x$$

De verwachtingswaarde van het gemiddelde van de waarnemingen blijkt gelijk aan het, exacte maar onbekende, gemiddelde van de verdeling, μ_x . Met andere woorden, het gemiddelde van de waarnemingen, $\bar{x}$, is een zuivere schatter voor het gemiddelde van de verdeling, μ .

Het gemiddelde van de waarnemingen, $\bar{x}$, zal een afwijking hebben ten opzichte van het exacte gemiddelde van de verdeling. De gemaakte fout bij het schatten is $\bar{x} - \mu_x$. Die fout kan gezien worden als een stochast met een gemiddelde gelijk aan 0. Om inzicht te krijgen in de grootte van de afwijking van de schatter wordt gekeken naar de verwachtingswaarde van de gekwadrateerde gemaakte fout. De "Mean Square Error" wordt vastgesteld.

De Mean Square Error van de schatter $\bar{x}$ is te vereenvoudigen tot :

$$E[\bar{X} - \mu_x]^2 = E[(\bar{X} - E[\bar{X}])^2] = \text{Var}[\bar{X}]$$

De samenhang tussen de bovengenoemde variantie van het gemiddelde, $\text{Var}[\bar{X}]$, en de variantie van de verdeling is als volgt te vinden:

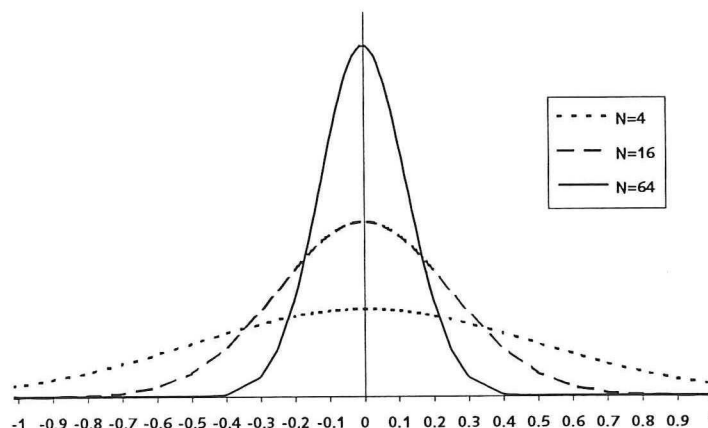
$$\text{Var}[\bar{X}] = \text{Var}\left[\frac{1}{N} \sum X_i\right] = \frac{1}{N^2} \text{Var}[\sum X_i]$$

Als X_i onafhankelijke waarnemingen zijn, dan geldt :

$$\text{Var}[\sum X_i] = N\sigma_x^2 \rightarrow \sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{N}$$

Zoals verwacht, is de standaardafwijking van het geschatte gemiddelde $\bar{x}$ ten opzichte van het werkelijke gemiddelde afhankelijk van het aantal waarnemingen.

In afbeelding 2.1 is de kansdichtheidsfunctie van het gemiddelde van de waarnemingen, $\bar{x}$, weergegeven voor verschillende aantallen waarnemingen, N . Duidelijk is te zien dat de spreiding van de schatter kleiner wordt als de schatter gebaseerd is op een groot aantal waarnemingen.



Afbeelding 2.1

De schatter van de variantie is eigenlijk, in het geval van een normale verdeling, χ^2 -verdeeld. Bij een groot aantal waarnemingen kan de schatter als normaal verdeeld beschouwd worden. In dit onderzoek wordt aangenomen dat de schatter van de variantie in ieder geval normaal verdeeld is. Als de verwachtingswaarde van de variantie bepaald wordt, blijkt dat:

$$E[\text{Var}(X_i)] = \frac{N-1}{N} \text{Var}[X]$$

De verwachtingswaarde van de schatter van de variantie zal voor iedere N kleiner zijn dan de eigenlijke variantie van de verdeling. Naarmate het beschikbare aantal waarnemingen, N , groter is, zal het gemiddelde dichter bij de werkelijke variantie komen te liggen. Schatters, zoals deze, waarvan de verwachtingswaarde niet gelijk is aan de parameter, noemen we een onzuivere schatter. De vaak gebruikte Engelse term hiervoor is een 'biased estimator'.

Stel nu dat de variantie van de waarnemingen gedefinieerd zou worden door

$$\text{Var}^*(X_i) = \frac{N}{N-1} \text{Var}[X_i] = \frac{1}{N-1} \sum (X_i - \bar{X})^2$$

De verwachtingswaarde van deze nieuwe schatter van de variantie, $\text{Var}^*(X)$, is gelijk aan:

$$E[\text{Var}^*(X_i)] = \text{Var}[X]$$

Dit is dus een zuivere schatter, of 'unbiased estimator'. De verwachtingswaarde van de schatter van de variantie is gelijk aan de variantie van de verdeling. In veel gevallen wordt voor deze schatter gekozen.

Hoewel de schatter zuiverder is, hoeft ze niet altijd 'beter' te zijn. Om de twee schatters van de variantie te vergelijken wordt ook hier de Mean Square Error bepaald. De Mean Square Error geeft een maat voor de afwijking tussen de geschatte variantie, S^2 of S^{*2} , en de variantie van de verdeling, σ_x^2 .

Voor de Mean Square Error van de onzuivere schatter S^2 geldt:

$$E[(S^2 - \sigma^2)^2] = \text{Var}[S^2] + \left(\frac{N-1}{N} \sigma^2 - \sigma^2\right)^2 = \text{Var}[S^2] + \frac{1}{N^2} \sigma^4$$

voor de zuivere schatter S^{*2} geldt:

$$E[(S^{*2} - \sigma^2)^2] = \text{Var}[S^{*2}] = \frac{N^2}{(N-1)^2} \text{Var}[S^2]$$

Deze fouten zijn nog steeds niet goed te vergelijken. In [7], blz.381 wordt meegedeeld, dat voor een normaal verdeelde stochast 'some simple algebra will show that':

$$\text{Var}[S^2] = \frac{2(N-1)}{N^2} \sigma^4$$

Hiermee is het vergelijken van de twee schatters mogelijk geworden. De bovenstaande vergelijking geldt bij benadering ook voor andere verdelingen. Met die vergelijking kan dan de standaardafwijking van de zuivere schatter, $\text{Var}[S^{*2}]$, als functie van de te schatten parameter σ afgeleid worden :

$$\text{Var}[S^{*2}] = \frac{2}{N-1} \sigma^4$$

Het blijkt dat bij een klein aantal waarnemingen de standaardafwijking van de onzuivere schatter van de variantie, $\text{Var}[S^2]$, kleiner is dan de standaardafwijking van de zuivere schatter, $\text{Var}[S^{*2}]$

In afbeelding 2.2 is de kansdichtheid van een zuivere en een onzuivere schatter van de variantie bij verschillende aantallen waarnemingen weergegeven. Goed is te zien dat naarmate het aantal waarnemingen groter wordt, de verwachtingswaarde van de onzuivere schatter steeds dichter bij de werkelijke variantie komt te liggen. $E[S^2 - \sigma^2] \rightarrow 0$ als het aantal waarnemingen, N , groter wordt. Het is ook duidelijk dat de benadering van de kansdichtheidsfunctie van de variantie door een normale verdeling niet voldoet. Bij een klein aantal waarnemingen ligt met die benadering een deel

van de kansdichtheid in het negatieve gebied.

In afbeelding 2.2 op de vorige bladzijde is tevens te zien dat de spreiding van de zuivere schatter eerst groter is dan die van de onzuivere schatter, maar dat dit verschil verdwijnt naarmate N groter wordt. De standaardafwijking van de variantie is omgekeerd evenredig met de wortel uit het aantal waarnemingen, N .

Met de standaardafwijking van de schatter van de variantie is het nog niet mogelijk om inzicht te krijgen in de standaardafwijking van de schatters van de parameters van de verdelingen. Die schatters worden namelijk niet bepaald door de variantie, maar door de standaardafwijking. Het is daarom nodig om $\text{Var}[S]$ in plaats van $\text{Var}[S^2]$ te bepalen.

Om verwarring te voorkomen wordt " V " in plaats van " S^2 " geschreven.

We stellen $y=V$ en $x=S$. $V=S^2$, dus $y=x^2$.

Op blz.2-15, [5], is met behulp van de Mean-Value Approach benaderd dat als $y=x^2$ voor de standaardafwijkingen geldt (voor een groot aantal waarnemingen N):

$$\sigma(y) = |2\mu(x)| \cdot \sigma(x) \rightarrow \sigma(V) = |2\mu(S)| \cdot \sigma(S)$$

$$\rightarrow \sigma(S) = \frac{\sigma(V)}{2\mu(S)} \rightarrow \text{Var}[S] = \frac{\text{Var}[V]}{4 \cdot (\mu(S))^2} = \frac{2}{N-1} \cdot \sigma^4 \cdot \frac{1}{4\sigma^2} = \frac{1}{2(N-1)} \sigma^2$$

Het is nu mogelijk om met de variantie van de zuivere schatter van de variantie, $\text{Var}[S^2]$ (of $\text{Var}[V]$), het gemiddelde en de variantie van enkele schatters analytisch te bepalen. In tabel 2.1 wordt de verwachtingswaarde van de schatters gegeven. De variantie is te bepalen met de volgende rekenregel: $\text{Var}[a+bX+cY]=b^2\text{Var}[X]+c^2\text{Var}[Y]$, waarbij X en Y onafhankelijke stochasten zijn en a, b en c deterministische waarden. Als aangenomen wordt dat de schatter van het gemiddelde onafhankelijk is van de schatter van de standaardafwijking dan kunnen de varianties van de schatters van bepaald worden.

Voor de schatters van de parameters van de exponentiële verdeling geldt:

$$\text{Var}[\hat{A}] = \text{Var}[\bar{X}] + \text{Var}[S] = \frac{\sigma^2}{N} + \sigma^2 \cdot \frac{1}{2(N-1)}$$

$$\text{Var}[\hat{B}] = \text{Var}[S] = \frac{1}{2(N-1)} \sigma^2$$

En voor de schatters van de parameters van de Gumbel verdeling geldt:

$$\text{Var}[\hat{A}] = \text{Var}[\bar{X}] + \gamma^2 \text{Var}[S] = \frac{\sigma^2}{N} + \gamma^2 \sigma^2 \cdot \frac{1}{2(N-1)}$$

$$\text{Var}[\hat{B}] = \frac{6}{\pi^2} \text{Var}[S] = \sigma^2 \cdot \frac{3}{\pi^2(N-1)}$$

Voor de schatters van de parameters van de lognormale verdeling is de variantie niet op deze wijze te bepalen, omdat de momentenschatters nog een niet-lineaire transformatie ondergaan.

Een tweede methode om de onzekerheid van schatters te bepalen is het bootstrappen. In dit geval is inzicht verkregen in de verdeling van de momentenschatters van de verschillende verdelingen door 500 keer de datareeks behorend bij de verdeling te bootstrappen en de parameters te bepalen. Voor de parameters van de exponentiële verdeling is de datareeks die verkregen is met behulp van de POT-methode, bestaande uit 530 observaties, gebootstrapt. In het geval van de Gumbel- en de lognormale verdeling is de reeks met 108 jaarmaxima gebruikt.

De resultaten staan in afbeelding 2.3. In iedere grafiek staan de uitkomsten van de bootstrap voor een schatter met een dunne lijn en de daaruit bepaalde normale verdeling met een dikke getrokken lijn. De normale verdeling die op analytische wijze gevonden is, is met een gestippelde lijn aangegeven. De twee methoden komen in de meeste gevallen goed overeen. Opvallend is het verschil tussen de twee methoden in geval van de standaardafwijking van de kansdichtheidsfunctie van de schatter van B van de Gumbel-verdeling; Die verschilt een factor 2.

2.4 De Methode van de Maximum-Likelihood

2.4.1 Inleiding

Stel dat van een zekere kansdichtheidsfunctie $f(x)$ de parameter θ bekend is. Van een reeks onafhankelijke waarnemingen is dan de gezamenlijke kansdichtheidsfunctie

$$f(x_1, x_2, \dots, x_N | \theta) = f(x_1) \cdot f(x_2) \cdot \dots \cdot f(x_N) = \prod f(x_i | \theta)$$

Ook de omgekeerde veronderstelling is mogelijk. Een reeks waarnemingen met een zekere kansverdeling, maar de parameters van de verdeling zijn onbekend. Nu wordt gezocht naar een parameter θ , waarbij juist de waargenomen reeks data het meest waarschijnlijk is. De kans op de data is

$$f(x_1) \cdot f(x_2) \cdot \dots \cdot f(x_N)$$

Dit is de functie van de grootste aannemelijkheid van de waarnemingen, of in de Engelse term, van de Maximum Likelihood.

$$L(\theta | x_1, x_2, \dots, x_N) = \prod f(x_i | \theta)$$

De meest aannemelijke schatter van de parameter θ is die waarde waarvoor de bovenstaande functie maximaal is.

2.4.2 Bepaling van een Maximum Likelihood-schatter

Om het maximum van de functie te vinden zou de afgeleide van de Likelihood gelijk aan 0 gesteld kunnen worden. Meestal is het echter beduidend eenvoudiger om de logaritme van de functie te differentiëren en gelijk aan 0 te stellen. De logaritme van een functie heeft bij precies dezelfde waarde een maximum. De log-likelihoodfunctie heeft het grote voordeel dat er geen sprake meer is van een product, maar van een som van componenten.

$$\log\left[\prod_{i=1}^N f(x_i | \theta)\right] = \sum_{i=1}^N \log[f(x_i | \theta)]$$

Iedere component kan vervolgens afzonderlijk gedifferentieerd worden.

Ter illustratie wordt de Maximum-Likelihood-schatter bepaald van de parameter λ van een exponentiële verdeling met de volgende kansdichtheidsfunctie

$$f(x | \lambda) = \lambda e^{-\lambda x}$$

De likelihoodfunctie is dan

$$L(\lambda | x_1, x_2, \dots, x_N) = \prod f(x_i | \lambda) = \lambda_1 \exp^{-\lambda x_1} \cdot \lambda_2 \exp^{-\lambda x_2} \cdot \dots \cdot \lambda_N \exp^{-\lambda x_N} = \lambda^N \exp^{-\lambda \sum x_i}$$

Als daar de natuurlijk logaritme van genomen wordt, volgt

$$\ln L(\lambda | x_i) = N \ln(\lambda) - \lambda \sum_{i=1}^N x_i$$

Om deze log-likelihood functie te maximaliseren wordt de afgeleide gelijk aan 0 gesteld:

$$\frac{d \ln L(\lambda | x_i)}{d \lambda} = \frac{N}{\lambda} - \sum x_i = 0$$

De maximum-likelihood-schatter $\hat{\lambda}$, waarvoor de functie maximaal is:

$$\hat{\lambda} = \frac{N}{\sum x_i} = \frac{1}{\bar{x}}$$

Toevallig is deze schatter gelijk aan de schatter die met de methode der momenten bepaald is.

Als van een kansdichtheidsfunctie niet één, maar meer parameters bepaald dienen te worden, dan wordt de log-likelihoodfunctie naar de afzonderlijke parameters gedifferentieerd. In het algemeen kan gesteld worden dat de volgende functies gemaximaliseerd moet worden

$$\frac{d \sum_{i=1}^N \log[f(x_i | \theta)]}{d \theta_j} = 0$$

Voor iedere $j=1..n$, het aantal parameters van de kansdichtheidsfunctie.

Opgemerkt dient te worden dat bij het zoeken naar de Maximum Likelihood wel rekening gehouden moet worden met de grenzen van de parameters. Zo zal een schatter voor een parameter A van een exponentiële verdeling altijd kleiner moeten zijn dan de laagste waarde uit de dataset, omdat de verdelingsfunctie anders groter dan 1 zou kunnen worden. Het kan voorkomen dat het Maximum van de Likelihood gevonden wordt bij een onder- of bovengrens van een parameter. Een voorbeeld hiervan is de schatter van A van een exponentiële verdeling. Voor illustratie wordt verwezen naar bijlage 4.

In bijlage 4 zijn voor de exponentiële en de lognormale verdeling de schatters van de parameters bepaald. De resultaten voor enkele verdelingen zijn in tabel 2.2 weergegeven.

Schatter	Methode van de Maximum Likelihood	
exponentieel $\hat{\lambda}$	$\hat{\lambda} = 1/\bar{x}$	
exponentieel $\hat{A}, \hat{B}$	$\hat{A} = x_{(1)}$	$\hat{B} = \bar{x} - x_{(1)}$
lognormaal, $\hat{A}, \hat{B}$	$\hat{A} = \frac{1}{N} \sum \ln x_i$	$\hat{B} = \sqrt{\frac{1}{N} \sum (\ln x_i)^2 - (\frac{1}{N} \sum \ln x_i)^2}$

Tabel 2.2

In de tabel staat voor het gemiddelde van de waarnemingen $\bar{x}$ geschreven. $x_{(1)}$ is de kleinste waarde van de waarnemingen.

Voor niet alle verdelingen, bijvoorbeeld de Gumbel-verdeling, is de ML-schatter eenvoudig te bepalen. Het is dan mogelijk om numeriek de Likelihood-functie te maximaliseren. In bijlage 5 is een Pascal programma weergegeven waarmee numeriek de parameters van een Gumbel-verdeling met behulp van de Maximum Likelihood bepaald kunnen worden. In het geval van de Gumbel-verdeling is de Likelihood een functie van twee parameters A en B en wordt gezocht naar die combinatie waarbij de Likelihoodfunctie maximaal is.

2.4.3 Onzekerheid van een Maximum Likelihood-schatter

Net zoals bij de schatters van de methode der momenten het geval was, is ook nu de geschatte parameter geen deterministische grootheid. De geschatte parameter $\hat{\theta}$ is normaal verdeeld. Het gemiddelde van $\hat{\theta}$ is asymptotisch gelijk aan θ voor $N \rightarrow \infty$. Een Maximum Likelihood schatter wordt zuiverder naarmate het aantal waarnemingen N groter wordt.

Voor de Mean Square Error, de gekwadeerde fout tussen de geschatte parameter $\hat{\theta}$ en de werkelijke waarde van de parameter θ geldt:

$$E[(\hat{\theta} - \theta)^2] = \text{Var}[\hat{\theta}] + (E[\hat{\theta}] - \theta)^2$$

Aangezien $E[\hat{\theta}] - \theta = 0$ voor grote N , is de Mean Square Error gelijk aan $\text{Var}[\hat{\theta}]$. In [7] op blz.400 wordt de volgende benadering gegeven van $\text{Var}[\hat{\theta}]$, als $N \rightarrow \infty$:

$$\text{Var}[\hat{\theta}] = \frac{-1}{N \cdot E\left[\frac{d^2 \ln f(x|\theta)}{d\theta^2}\right]}$$

Ter illustratie wordt de variantie van de in de vorige paragraaf geschatte $\hat{\lambda}$ bepaald.

Om te beginnen wordt de logaritme van de kansdichtheidsfunctie genomen.

$$f(x|\lambda) = \lambda e^{-\lambda x} \rightarrow \ln[f(x|\lambda)] = \ln(\lambda) - \lambda x$$

Dan volgt voor de tweede afgeleide:

$$\rightarrow \frac{d^2 \ln[f(x|\lambda)]}{d\lambda^2} = -\frac{1}{\lambda^2}$$

In dit geval is het resultaat een constante die niet afhankelijk is van de waarnemingen.

$$\rightarrow E\left[\frac{d^2 \ln[f(x|\lambda)]}{d\lambda^2}\right] = -\frac{1}{\lambda^2}$$

$$\rightarrow \text{Var}[\hat{\lambda}] \approx \frac{\lambda^2}{N} \rightarrow \sigma[\hat{\lambda}] \approx \frac{\lambda}{\sqrt{N}}$$

Net zoals bij de schatters bepaald met de Methode der momenten is de standaardafwijking van de parameter omgekeerd evenredig met de wortel uit het aantal waarnemingen.

Op vergelijkbare wijze kan voor andere verdelingen de standaardafwijking bepaald worden. In bijlage 6 wordt de standaardafwijking van de schatters voor de exponentiële verdeling en de lognormale verdeling bepaald. De schatter van parameter A van de exponentiële verdeling gedefinieerd is door het minimum van de waarnemingen. De schatter is daarom niet normaal verdeeld, maar heeft een exponentiële verdeling. Daarvoor geldt:

$$P(\text{Min}(x_1, \dots, x_N) < a) = 1 - P(\text{Min}(x_1, \dots, x_N) > a) = 1 - \prod (P(x_i > a)) = 1 - \exp^{-N(\frac{a-A}{B})}$$

Zodat voor de schatter van A geldt:

$$F_{\hat{A}}(a) = 1 - \exp^{-N(\frac{a-x_{(1)}}{\bar{x}-x_{(1)}})}$$

Naast de analytische methode is het ook mogelijk om inzicht in de onzekerheid van schatters te krijgen door het bootstrappen van de beschikbare waarnemingen. Voor iedere onderzochte verdeling is 500 keer de datareeks behorend bij de verdeling gebootstrapt en zijn de ML-schatters bepaald.

De resultaten staan in afbeelding 2.4. In iedere grafiek staan de resultaten van de bootstrap en zo mogelijk de daaruit bepaalde normale verdeling. De ML-schatter van parameter A van de exponentiële verdeling was iedere keer gelijk aan de drempelwaarde voor de waterstanden bij de POT-methode. Alle andere onderzochte schatters kunnen als normaal verdeeld beschouwd worden.

In de grafieken met de schatters van de parameters van de lognormale verdeling staat ook analytisch bepaalde normale verdeling van de schatter. De twee normale verdelingen komen bijzonder goed overeen. Het blijkt dat bootstrappen een goed inzicht in de onzekerheid van een schatter geeft.

2.5 Kleinste kwadraten methode

2.5.1 Inleiding

Deze methode om de parameters van een verdeling te schatten is van oorsprong een grafische methode voor lineaire functies. Het principe is dat er een lijn door de data getrokken kan worden, indien deze op juiste wijze in een grafiek zijn uitgezet. Uiteindelijk zijn de schatters van de parameters van de verdeling te bepalen uit de positie en de helling van de gevonden lijn.

Voor verscheidene eenvoudige kansdichtheidsfuncties is het mogelijk om de verdelingsfunctie $F(x)$ voor verschillende waarden van de stochast X als een rechte lijn te plotten. Daartoe kan de functie getransformeerd worden of waarschijnlijkheidspapier gebruikt worden.

We beschikken over een op grootte geordende dataset met N waarnemingen, $X_1, \dots, X_N$, uit dezelfde verdeling met onbekende parameters. Voor iedere waarde X_i wordt een overschrijdingsfrequentie $F(X_i)$ geschat. Voor N waarnemingen ligt het voor de hand om $F(X_i) = i/N$ te kiezen. Het blijkt echter dat bij een geschatte overschrijdingsfrequentie van $F(X_i) = i/(N+1)$ zuiverder schatters gevonden worden. De overschrijdingslijn van de waarnemingen wordt dus gevonden door elke waarneming X_i uit te zetten tegen haar overschrijdingsfrequentie $F(X_i)$.

Als de verdeling geen lineaire functie is, dan dient eerst de plotpositie getransformeerd te worden, zodat de waarnemingen uitgezet worden op een rechte lijn.

De bepaling van de regressielijn vindt plaats met de methode der kleinste kwadraten. Het principe daarvan is dat de schatters van de parameters zo gekozen worden, dat de som van de gekwadraterde verschillen tussen de waarden y_i en de geschatte lijn minimaal is. A en B moeten dus zo gekozen worden, dat de volgende vergelijking minimaal is :

$$\text{Min}_{\theta, \theta} \sum_{i=1}^N \left[\frac{i}{N+1} - F(x, \theta) \right]^2$$

2.5.2 Bepaling van een kleinste-kwadraten-schatter

Om te beginnen zal de bepaling van een schatter met behulp van lineaire regressie besproken worden. Bij lineaire regressie wordt de verdelingsfunctie gelineariseerd. Een voordeel daarvan is dat na de transformatie van een niet-lineaire verdeling de schatters redelijk eenvoudig met behulp van de kleinste kwadraten methode te bepalen zijn.

Om de te volgen werkwijze bij lineaire regressie te verduidelijken worden van een exponentiële functie de schatters van de twee parameters A en B bepaald. De verdelingsfunctie is:

$$F(x) = 1 - e^{-\frac{x-A}{B}}$$

Om te beginnen wordt de functie getransformeerd :

$$\Leftrightarrow 1 - F(x) = e^{-\frac{x-A}{B}} \quad \Leftrightarrow \ln(1 - F(x)) = -\left(\frac{x-A}{B}\right)$$

De plotpositie, $y_i = F(x_i)$, gesubstitueerd in de getransformeerde vergelijking geeft :

$$F(x_i) = \frac{i}{N+1} \quad \rightarrow \quad y_i^* = \ln\left(1 - \frac{i}{N+1}\right) = -\left(\frac{x_i - A}{B}\right)$$

Hierin is y^* de getransformeerde y . y^* is lineair afhankelijk van x . Nu is het dus mogelijk om de schatters voor A en B te bepalen. Om de som van de gekwadraterde afwijkingen te minimaliseren, moet de afgeleide naar respectievelijk A en B gelijk aan 0 moet zijn.

$$\frac{d \left(\sum_{i=1}^N \left[y_i^* - \frac{(x_i - A)}{B} \right]^2 \right)}{dA} = 0 \rightarrow -\frac{2}{B} \sum_{i=1}^N \left(y_i^* - \frac{(x_i - A)}{B} \right) = 0$$

$$\frac{d \left(\sum_{i=1}^N \left[y_i^* - \frac{(x_i - A)}{B} \right]^2 \right)}{dB} = 0 \rightarrow -\frac{2}{B^2} \sum_{i=1}^N (x_i - A) \cdot \left(y_i^* - \frac{(x_i - A)}{B} \right) = 0$$

Uit deze twee vergelijkingen met twee onbekenden kan de schatter voor A en B opgelost worden.

Een voordeel van het lineariseren van de verdeling is dat het redelijk eenvoudig is om de parameters te bepalen. De nadelen van het lineariseren zijn aan de hand van een exponentiële verdeling duidelijk uit te leggen.

Een nadeel van het lineariseren is dat een functie F andere zwaartepunten heeft dan zijn logaritme, $\log(F)$. Bovendien wordt een fout gemaakt, doordat de gekwadrateerde afstand tussen de gelineariseerde y_i en de gelineariseerde verdelingsfunctie wordt geminimaliseerd. Dit is niet de gezochte minimale fout. De som van de gekwadrateerde afwijking moet niet geminimaliseerd worden in de zogenaamde $\log(y)$ -ruimte, maar in de y -ruimte.

Om dit probleem te ondervangen, kan met behulp van een computer de gekwadrateerde fout numeriek geminimaliseerd worden voor de werkelijke verdeling en de werkelijke plotpositie. Het grote voordeel van deze numerieke methode is dat er zuiverder schatters gevonden worden. Een bijkomend voordeel is echter dat de methode op een verscheidenheid van verdelingen toepasbaar is, aangezien ze niet lineariseerbaar hoeven te zijn. De meeste verdelingen zijn dat echter wel.

Om numeriek een oplossing te vinden moet wel een beginwaarde gegeven worden. Aangezien het erg aantrekkelijk is voor de programmeur om de parameters die met de gelineariseerde verdeling gevonden zijn hiervoor te gebruiken, vervalt het voordeel van niet-lineariseerbaarheid.

Ter illustratie van het verschil tussen lineaire en niet-lineaire regressie zijn afbeelding 2.5 en 2.6 geconstrueerd. Uit een bekende exponentiële verdeling zijn 10 onafhankelijke waarnemingen getrokken. Met lineaire regressie is de gestippelde lijn bepaald. De gestreepte lijn is met niet-lineaire regressie bepaald.

In de bovenste afbeelding zijn de waarnemingen en de geschatte verdelingen gelineariseerd weergegeven; in de onderste niet-lineair. In de bovenste afbeelding is de gekwadrateerde afstand tussen de data en de gestippelde lijn (lin.regr.) minimaal. Dit is de minimale gekwadrateerde afstand in de $\log(y)$ -ruimte. Bij de methode van de kleinste kwadraten moet echter de minimale gekwadrateerde afstand in de y -ruimte gevonden worden. In de onderste afbeelding is te zien dat de gekwadrateerde afstand tussen de waarnemingen en de gestreepte lijn (niet-lin.regr.) minimaal is.

2.5.3 Onzekerheid van een kleinste-kwadraten-schatter

Net zoals bij de andere schattersmethoden, is ook de kleinste-kwadraten-schatter geen deterministische grootheid, maar een stochast.

Bij deze methode is het niet mogelijk om, zoals bij de Methode van de Maximum Likelihood, de verwachtingswaarde en de standaardafwijking van de schatter analytisch te bepalen.

In de praktijk leek het alsof de kleinste-kwadraten-schatter van parameter B van een exponentiële verdeling altijd groter was dan de zuivere momentenschatter. In Bijlage 7 is dit bewezen. De kleinste-kwadraten-schatter nadert de zuivere momentenschatter naarmate het aantal waarnemingen waarop de schatter gebaseerd is toeneemt.

De verdeling van de kleinste-kwadraten-schatter die gebruik maakt van lineaire regressie is wel numeriek met behulp van een bootstrap te bepalen voor lineariseerbare verdelingsfuncties. De resultaten zijn weergegeven in afbeelding 2.7. De gebootstrapte waarden komen goed overeen met een normale verdeling.

De standaardafwijking van de kleinste-kwadratenschatter met *niet*-lineaire regressie is niet te bepalen. Ook het bootstrappen biedt praktisch gesproken geen uitkomst, vanwege de bij dit onderzoek gebruikte programmatuur. Om 500 keer nieuwe kleinste-kwadratenschatters te bepalen, zou 500 keer een nieuwe datareeks in het gebruikte fit-programma 'Superfit' handmatig ingevoerd moeten worden en de routine van schatters bepalen herhaald.

2.6 Vergelijken van de verschillende methoden

2.6.1 Inleiding

In de vorige paragrafen zijn drie verschillende methoden om parameters van een kansverdeling te schatten behandeld.

- Methode der momenten
- Methode van de Maximum Likelihood
- Kleinste kwadraten methode

In de nevenstaande tabel 2.3 zijn de schatters die bepaald zijn met de eerste twee methoden weergegeven. In deze paragraaf zal een vergelijking van de verschillende methoden plaatsvinden, waarmee in het volgende hoofdstuk een keuze gemaakt kan worden. De methoden worden vergeleken op zuiverheid en efficiëntie van de schatter en toepasbaarheid van de methode.

2.6.2 Zuiverheid en efficiëntie

Een eerste vergelijking vindt plaats met de maatstaven van zuiverheid en efficiëntie van de schatter. Zuiverheid is de verwachtingswaarde van het verschil tussen de parameter en de schatter.

Zuiverheid = $E[\hat{\theta} - \theta]$.

Efficiëntie is gedefinieerd als de standaardafwijking van de schatter ten opzichte van de parameter.

Efficiëntie = $E[(\hat{\theta} - \theta)^2]$

De enige schatter waarvan de verwachtingswaarde gelijk is aan de gezochte parameter is de momentenschatter. Dit is de enige zuivere schatter. Alle andere schatters zijn niet zuiver, maar de verwachtingswaarde van de schatter benadert de parameter als het aantal waarnemingen erg groot wordt.

De efficiëntie van de schattingsmethoden is overeenkomstig. Als aangenomen wordt dat bij een groot aantal waarnemingen niet alleen de momentenschatter, maar ook de Maximum Likelihood-schatter zuiver is, dan komt de efficiëntie bij die twee methoden vaak overeen met σ^2/N .

In [4] is de zuiverheid en de efficiëntie van de drie methoden numeriek vergeleken. In het onderzoek zijn volgens de drie methoden de parameters A en B van een exponentiële verdeling geschat voor 14 datasets met ieder 10 waarnemingen. De datasets zijn gegenereerd uit een exponentiële verdeling met $A=1.96$ en $B=0.33$. Het onderzoek is uitgebreid door van precies dezelfde datasets de parameters A en B te bepalen, met de kleinste kwadraten methode met niet-lineaire regressie. In Bijlage 8 zijn de uitkomsten weergegeven. Met de resultaten is numeriek de zuiverheid en de efficiëntie van de verschillende methoden bepaald. Zie tabel 2.4. Een lage waarde bij de zuiverheid of de efficiëntie betekent een zuivere, respectievelijk efficiënte schatter. De laagste waarde is steeds dikgedrukt.

De momentenschatter is qua zuiverheid een goede schatter, zoals ook analytisch reeds bepaald was. De Maximum Likelihoodschatter is niet alleen redelijk zuiver, (al bij 10 waarnemingen vergelijkbaar met de methode der momenten !), maar ook een erg efficiënte schatter in vergelijking met de andere schattingsmethoden. De kleinste kwadratenschatter met lineaire regressie geeft de slechtste schattingen van parameters.

2.6.3 Toepasbaarheid

Een tweede vergelijking van de schattingsmethoden betreft de toepasbaarheid van de methoden. Niet alle methoden zijn toepasbaar op alle verdelingen.

Bij de *methode der momenten* is het voor verscheidene functies lastig om de momenten zelf te bepalen. In de literatuur zijn de uitdrukkingen meestal wel te vinden.

Bij de *methode van de Maximum Likelihood* is het vaak mogelijk om met stug doorrekenen de schatter analytisch te bepalen. Het is ook mogelijk om numeriek een oplossing te vinden.

De *kleinste-kwadraten-methode* heeft in welke vorm dan ook het nadeel dat een plotpositie aan de waarnemingen toegevoegd moet worden. Een verdeling moet een lineariseerbare verdelingsfunctie hebben, wil een bepaling van een lineaire regressie mogelijk zijn. De Kleinste-Kwadraten-schatter met niet-lineaire regressie heeft in principe geen beperkingen wat betreft de verdelingsfunctie.

Overschrijdingslijnen van waterstanden

3.1 Inleiding

3.1.1 Bepalen van een overschrijdingslijn

De waterstand wordt als een stochastische variabele beschouwd. Om die variabele statistisch te beschrijven wordt de bijbehorende verdeling gezocht. In Hoofdstuk 1 is beschreven hoe uit de waterstanden die, in Hoek van Holland in de periode van 1887 tot 1994, gemeten zijn, twee reeksen met onafhankelijke en homogene waarnemingen zijn geselecteerd. Met die reeksen is het mogelijk om de parameters van de verdelingsfunctie te schatten.

Verondersteld is dat de reeks jaarmaxima lognormaal of Gumbel verdeeld zijn. De waterstanden die geselecteerd zijn met de Peak-Over-Threshold-methode worden als exponentieel verdeeld beschouwd. In hoofdstuk 2 is beschreven hoe de parameters van een verdeling en hun onzekerheid geschat kunnen worden. In dit hoofdstuk worden de resultaten van de schattingen gegeven. Van iedere schatter wordt de waarde van het gemiddelde en de standaardafwijking gegeven. Waar mogelijk is gekozen voor de analytische oplossing. In de tabel met de uitkomsten is aangeduid of de schatter op analytische (An.) of numerieke wijze (Nu.) is bepaald.

Als de verdelingsfunctie en de parameters bekend zijn, is de overschrijdingslijn vastgesteld. Het is nu mogelijk om de waterstand met een overschrijdingsfrequentie van 1 keer in 10.000 jaar te bepalen. Die waterstand wordt het 10^{-4} -kwantiel genoemd.

3.1.2 De invloed van statistische onzekerheid op een overschrijdingslijn

Zoals eerder besproken, zijn schatters ook stochasten. In dit hoofdstuk wordt bekeken wat de invloed van die statistische onzekerheid op het 10^{-4} -kwantiel is. Er zijn twee methoden om de invloed van de onzekere schatters mee te nemen.

De eerste methode noemen we de klassieke of frequentistische methode. Een overschrijdingskans, p , van een waterstand, x_p , wordt beschreven door de verdelingsfunctie en haar parameters.

$$p = F(x_p, A, B)$$

Om de statistische onzekerheid mee te nemen wordt die vergelijking zo herschreven dat de waterstand x_p een functie is van de overschrijdingskans p . Bij een gegeven overschrijdingskans wordt de bijbehorende waterstand dan een functie van de stochasten A en B .

$$x_p = G(p, A, B)$$

Met behulp van het gemiddelde en de standaardafwijking van de schatters is vervolgens het gemiddelde en de standaardafwijking van het kwantiel te bepalen.

De tweede methode om de invloed van statistische onzekerheid op een overschrijdingslijn mee te nemen is de zogenaamde Bayesiaanse methode. Bij die methode wordt de verdeling van de schatter uit de verdelingsfunctie van de waterstand geïntegreerd. Een schatter zal daarbij weergegeven worden door de werkelijke, gezochte parameter met een normaal verdeelde spreiding, ϵ . In het vorige hoofdstuk is gebleken dat niet alle schatters als normaal verdeeld beschouwd kunnen worden. De uitzonderingen zullen besproken worden op het moment dat ze aan de orde komen.

De kansdichtheidsfunctie van de spreiding ϵ is :

$$f(\epsilon) = \frac{1}{\sigma_{\epsilon} \sqrt{2\pi}} \cdot e^{-\frac{(\epsilon - \mu_{\epsilon})^2}{2\sigma_{\epsilon}^2}}$$

Het gemiddelde is gedefinieerd door: $\mu_{\epsilon} = E[\hat{\theta}] - \theta$. Als het een zuivere schatter betreft zal de verwachtingswaarde van de geschatte parameter gelijk zijn aan de werkelijke parameter van de verdeling. Dit betekent dat voor een zuivere schatter geldt: $\mu_{\epsilon} = 0$.

De standaardafwijking van de spreiding ϵ hangt samen met het type verdelingsfunctie en met het aantal beschikbare waarnemingen. Bij een groot aantal waarnemingen zal de normaal verdeelde spreiding een kleinere standaardafwijking hebben, dan bij een klein aantal waarnemingen. De geschatte parameter zal dan dichter bij de werkelijke parameter liggen.

De verdelingsfunctie van x , waarin een strooing ϵ rond een parameter θ meegenomen is, is als volgt te bepalen:

$$F_{\theta}(x) = \int F_{\theta}(x|\epsilon) \cdot f(\epsilon) d\epsilon$$

De voorwaarde is wel dat de grenzen waartussen geïntegreerd gaat worden, zo gekozen zijn, dat de verdelingsfunctie geen waarden kleiner dan 0 of groter dan 1 kan aannemen.

Voor het overzicht is in eerste instantie de invloed van de statistische onzekerheid voor iedere parameter, A en B, afzonderlijk meegenomen. De invloed van de verschuivingsparameter A bleek in alle gevallen verwaarloosbaar te zijn. Het is dus niet nodig geweest om de invloed van de onzekerheid van beide parameters vast te stellen.

3.2 De overschrijdingslijnen van de jaarmaxima

3.2.1 Inleiding

We hebben de beschikking over 108 jaarmaxima. Met die waarnemingen kunnen de parameters van de lognormale en de Gumbelverdeling geschat worden. In het vorige hoofdstuk zijn de parameters van de Gumbelverdeling met drie verschillende schattingsmethoden bepaald. De parameters van een lognormale verdeling zijn alleen geschat met behulp van de methode der momenten en de methode van de Maximum Likelihood.

Bij de methode der momenten en de methode van de Maximum Likelihood worden uitsluitend de waarnemingen gebruikt om de schatters te bepalen. Bij de methode van de kleinste kwadraten is wel een plotpositie benodigd. Iedere waarneming i heeft een plotpositie $y_i = i/(N+1)$ gekregen. De waarnemingen zijn van groot naar klein geordend. i is het rangnummer. N is het totaal aantal jaren.

3.2.2 Resultaten

De schatters van de **Gumbel-verdeling** zijn alleen met de methode der Momenten (MdM) analytisch vastgesteld. Met de methode van de Maximum Likelihood (ML) zijn de parameters geschat door de Likelihoodfunctie numeriek te maximaliseren (§ 2.4.2). De Kleinste Kwadratenschatter is bepaald met lineaire regressie. De schatters van parameter A en van B blijken weinig afhankelijk van de toegepaste methode. Zie tabel 3.1.

In afbeelding 3.1 is goed te zien dat de overschrijdingslijn die bepaald is met de methode der Momenten bijna over de lijn heen valt die bepaald is met de methode van de Maximum Likelihood. De lijn die bepaald is met de kleinste-kwadraten-methode ligt hoger.

De standaardafwijking van de schatters (tabel 3.1) is bepaald met behulp van een bootstrap van de jaarmaxima.

Gumbel, N=108	$\mu(A)$		$\sigma(A)$		$\mu(B)$		$\sigma(B)$	
MdM	An.	2.361	Nu.	0.027	An.	0.2634	Nu.	0.026
ML	Nu.	2.361	Nu.	0.028	Nu.	0.2664	An.	0.028
KKW	An.	2.355	Nu.	0.027	An.	0.2781	Nu.	0.028

Tabel 3.1

In de tabel staat tevens aangegeven of de waarde analytisch (An.) of Numeriek (Nu.) bepaald is.

Met de bepaalde overschrijdingslijnen kan de waterstand met een overschrijdingskans van 10^{-4} berekend worden. Meestal wordt de verdeling van een schatter verwaarloosd en uitsluitend de gemiddelde waarde gebruikt.

Om (het gemiddelde van) het 10^{-4} -kwantiel vast te stellen, wordt de verdelingsfunctie zo herschreven dat de waterstand, x_p , een functie is van de overschrijdingskans, p . Voor de Gumbel-verdeling is dat: $x_p = A - B \cdot \ln(-\ln(p))$.

Door de gemiddelden van de schatters, $\mu(\hat{A})$ en $\mu(\hat{B})$, te substitueren voor A en B wordt het gemiddelde van het kwantiel, $\mu(X_{10^{-4}})$, vastgesteld. De uitkomsten zijn vermeld in Tabel 3.2 op de volgende bladzijde.

De variantie is op vergelijkbare wijze te bepalen. Opgemerkt dient te worden dat voor variantie de volgende rekenregel geldt: $\text{Var}(a+bX+cY)=b^2\text{Var}(X)+c^2\text{Var}(Y)$. Waarbij X en Y onafhankelijke stochasten zijn en a , b en c deterministisch. Als aangenomen wordt dat de parameters A en B onafhankelijk zijn, dan geldt voor de variantie van een kwantiel:

$$\text{Var}(x_p) = \text{Var}(A^2) + \text{Var}(B^2) \cdot (\ln(-\ln(p)))^2$$

Met de Bayesiaanse methode is de statistische onzekerheid van afzonderlijk A en B uit de verdelingsfunctie geïntegreerd. Als sprake is van een spreiding ϵ rond parameter A , respectievelijk rond parameter B , dan wordt de verdelingsfunctie:

$$F_A(x|\epsilon) = \exp\left(-\exp\left(-\frac{x-(A+\epsilon)}{B}\right)\right) \quad \text{voor } A \text{ en voor parameter } B : \quad F_B(x|\epsilon) = \exp\left(-\exp\left(-\frac{x-A}{(B+\epsilon)}\right)\right)$$

De onzekerheid ϵ kan uit de verdelingsfuncties geïntegreerd worden door de functie te vermenigvuldigen met de kansdichtheidsfunctie van ϵ en vervolgens te integreren over ϵ .

$$F_A(x) = \int F_A(x|\epsilon) \cdot f(\epsilon) d\epsilon \quad \text{en} \quad F_B(x) = \int F_B(x|\epsilon) \cdot f(\epsilon) d\epsilon$$

Van de 'nieuwe' kansverdeling kan vervolgens de waterstand met een overschrijdingskans van 10^{-4} vastgesteld worden. De berekeningen zijn uitgevoerd met het rekenprogramma 'Maple'. De resultaten zijn weergegeven in Tabel 3.2.

Gumbel, N=108	$\mu(X_{10^{-4}})$	$\sigma(X_{10^{-4}})$	$F_A(X_{10^{-4}})$	$F_B(X_{10^{-4}})$
MdM	NAP +4.787 m	0.24 m	NAP +4.787m	NAP +4.871 m
ML	NAP +4.814 m	0.24 m	NAP +4.815 m	NAP +4.902 m
KKW	NAP +4.916 m	0.26 m	NAP +4.916 m	NAP +5.001 m

Tabel 3.2

De invloed van de onzekerheid van de schatter van A is verwaarloosbaar. ($F_A(X_{10^{-4}}) \approx \mu(X_{10^{-4}})$). De onzekerheid van de schatter van B is wel terug te zien. ($F_B(X_{10^{-4}}) > \mu(X_{10^{-4}})$). Bij 108 waarnemingen is de onzekerheid van de schatter nog zo groot dat de standaardafwijking van het 10^{-4} -kwantiel $\pm 0.25\text{m}$ is en met de methode van uitintegreren $\pm 0.09\text{m}$ hoger. Opvallend is dat het 10^{-4} -kwantiel van de overschrijdingslijn die vastgesteld is met de methode van de kleinste kwadraten hoger uitvalt dan het 10^{-4} -kwantiel van de overschrijdingslijn die bepaald is met de andere twee methoden en waarbij ook de onzekerheid is meegenomen.

Ook bij de **lognormale verdeling** zijn de momentenschatters en de Maximum-Likelihood-schatters praktisch aan elkaar gelijk. Zie Tabel 3.3. Zelfs de standaardafwijking van de schatters komt goed overeen. De standaardafwijking van de momentenschatter is bepaald met een bootstrap van de jaarmaxima. De standaardafwijking van de Maximum-Likelihood-schatter is op analytische wijze bepaald.

lognormaal, N=108	$\mu(A)$		$\sigma(A)$		$\mu(B)$		$\sigma(B)$	
MdM	An.	0.9125	Nu.	0.0124	An.	0.1345	Nu.	0.0116
ML	An.	0.9130	An.	0.0124	An.	0.1284	An.	0.0097

Tabel 3.3

In afbeelding 3.2 zijn de overschrijdingslijnen weergegeven. Opvallend is dat de lognormale verdeling vergeleken met de Gumbel verdeling in het gebied van de kleine kansen naar beneden buigt. Het 10^{-4} -kwantiel zal een lage waarde hebben. Omdat de lognormale verdeling geen analytische verdelingsfunctie heeft, is het niet mogelijk om de waterstand x_p als functie van de kans p te herschrijven.

Het gemiddelde van het 10^{-4} -kwantiel is iteratief uit het gemiddelde van de schatters van A en B bepaald. De variantie van het 10^{-4} -kwantiel is op numerieke wijze bepaald. Daarbij zijn de schatters van A en B als normaal verdeeld beschouwd. Door een groot aantal malen uit die twee verdelingen te trekken en daarbij het 10^{-4} -kwantiel te bepalen is inzicht verkregen in de verdeling van het kwantiel.

Bij de tweede methode om de invloed van statistische onzekerheid op de overschrijdingslijn te beschouwen wordt de onzekerheid uit de verdelingsfunctie geïntegreerd. Als parameter A of B een spreiding ϵ heeft, dan kan de verdelingsfunctie geschreven worden als:

$$F_A(x|\epsilon) = \int_0^{\infty} \frac{1}{B\sqrt{2\pi x}} \exp \left(-\frac{(\ln(x) - (A+\epsilon))^2}{2B^2} \right) dx \quad \text{voor A en voor B:} \quad F_B(x|\epsilon) = \int_0^{\infty} \frac{1}{(B+\epsilon)\sqrt{2\pi x}} \exp \left(-\frac{(\ln(x) - A)^2}{2(B+\epsilon)^2} \right) dx$$

Om uit deze verdelingsfunctie de onzekerheid ϵ te integreren, zou de functie vermenigvuldigd moeten worden met de kansdichtheidsfunctie van de onzekerheid, $f(\epsilon)$, en geïntegreerd over ϵ . De statistische onzekerheid kan door numerieke integratie van deze dubbelintegraal bepaald worden, maar dit is in dit rapport niet meer aan de orde gekomen.

lognormaal, N=108	$\mu(X_{10^{-4}})$	$\sigma(X_{10^{-4}})$
MdM	NAP +4.107m	0.176 m
ML	NAP +3.996m	0.140 m

Tabel 3.4

3.3 De overschrijdingslijn van de reeks verkregen met de POT-methode

3.3.1 Inleiding

Door de selectie van de waterstanden met de POT-methode mag aangenomen worden, dat de overgebleven waarnemingen dezelfde kansverdeling hebben en onafhankelijk van elkaar zijn. Met behulp van de overschrijdingsfrequentie van de waterstanden kan de kansverdeling gevonden worden.

Met 'overschrijdingsfrequentie' wordt het aantal malen in een jaar bedoeld, dat de waterstand hoger is dan een zekere hoogte x . Om die frequentie te vinden worden de waarnemingen eerst geordend van hoog naar laag. Vervolgens krijgt iedere waarneming een rangnummer i . De plotpositie van de overschrijdingsfrequentie is dan: $F=i/T$

F is de overschrijdingsfrequentie van de i^e waarneming. T is het totaal aantal jaren waarin waarnemingen zijn gedaan. Om van de frequentie een kans te maken moet nog de volgende vertaalslag gemaakt worden:

$$P = 1 - \left(1 - \frac{F}{n_{\text{gem}}}\right)^{n_{\text{gem}}}$$

Hierin is P de overschrijdingskans. n_{gem} is het gemiddeld aantal malen in een jaar dat een waarneming boven een zekere drempel komt. In dit geval zijn er 530 waarnemingen in 108 jaar gedaan boven een vastgestelde drempel van NAP +1.85m, dus $n_{\text{gem}} = 530/108 = 4.912$

Bij de interpretatie van overschrijdingslijnen worden de begrippen overschrijdingsfrequentie en -kans vaak met elkaar verward. Voor alle, ook niet geselecteerde, waarnemingen mag je wel een frequentie geven, maar geen kans. Frequentie is een aantal per tijdseenheid en een kans is een getal tussen 0 en 1. Voor frequenties groter dan 1 maal per jaar is het snel duidelijk dat het geen kans is. Voor lagere waarden geldt, zoals ook in de afbeelding te zien is, dat bij waarden kleiner dan ± 0.1 frequenties en kansen aan elkaar gelijk zijn. In [1c] blz. 97 wordt hiervoor een bijzonder charmante afleiding gegeven.

Het probleem van de methode der momenten en de methode van de Maximum-Likelihood, is dat de parameters geschat worden voor een overschrijdingslijn van kansen en niet van frequenties. De Deltacommissie stelde reeds dat de asymptoot van de overschrijdings-frequentielijn gelijk is aan de kansverdeling voor de kleine kansen. Om het 10^{-4} -kwantiel te bepalen wordt dus een overschrijdingslijn van frequenties, $G(x)$, gezocht, met als randvoorwaarde dat alleen naar frequenties kleiner dan 0.1 gekeken mag worden.

De frequentieverdeling $G(x)$ is een kansverdeling $F(x)$ vermenigvuldigd met n_{gem} .

$$G(x) = n_{\text{gem}} \cdot F(x).$$

$G(x)$ is het gemiddeld aantal keren dat een waterstand groter is dan een zekere drempel in een jaar. $F(x)$ is de kansverdeling waarvan de parameters geschat kunnen worden met de MdM en ML. Om de overschrijdingslijn van de waterstanden te vinden dient nog een bewerking plaats te vinden.

Bij een exponentiële verdeling geldt $F(x) = \exp\left(-\frac{x-A}{B}\right)$. De frequentieverdeling volgt dan uit:

$$G(x) = n_{\text{gem}} \cdot \exp\left(-\frac{x-A}{B}\right) = \exp^{\ln n_{\text{gem}}} \cdot \exp\left(-\frac{x-A}{B}\right) = \exp\left(-\frac{x-A-B \cdot \ln n_{\text{gem}}}{B}\right) = \exp\left(-\frac{x-A'}{B}\right)$$

Uiteindelijk ziet de functie er dan uit als een gewone kansverdeling, maar dat is het dus niet. De maximale waarde die bereikt kan worden is namelijk niet 1, maar n_{gem} .

De Deltacommissie heeft deze bewerking ook toegepast. Zij heeft waterstanden boven een drempel van NAP +1.70m geselecteerd. $\rightarrow A=1.70$. De waarde die ze bij de overschrijdingslijn heeft toegepast is $A'=1.96$. Met $B=0.33$ betekent dit dat in het Deltarapport $n_{\text{gem}} = \exp[(A'-A)/B] = 2.199$.

In navolging daarvan zal in dit verslag uitsluitend naar de waarden van de parameters na bewerking gekeken worden. Dit is toegestaan mits alleen kleine frequenties of kansen worden beschouwd.

3.3.2 Resultaten

De schatters van de parameters van de kansverdeling van de waarnemingen, geselecteerd met de POT-methode, zijn met de methode der momenten en de methode van de Maximum Likelihood bepaald. De resultaten zijn voor de exponentiële verdeling weergegeven in tabel 3.5.

De Maximum-Likelihoodschatte van parameter A is niet normaal verdeeld. Het maximum van de Likelihood functie werd namelijk aan de rand gevonden. De schatter heeft daarom een extreme-waardeverdeling voor minima. Aangezien bij de bootstrap uitsluitend de drempelwaarde als schatter werd gevonden, is de standaardafwijking van de ML-schatte gelijk aan 0.

Exponentieel, N=530	$\mu(A)$		$\sigma(A)$		$\mu(B)$		$\sigma(B)$	
MdM	An.	2.316	Nu.	0.011	An.	0.279	Nu.	0.011
ML	An.	2.329	Nu.	0	An.	0.301	An.	0.013
KKW	An.	2.323	Nu.	0.012	An.	0.286	Nu.	0.017

Tabel 3.5

Er is een redelijk verschil tussen de momentenschatte en de ML-schatte van parameter B. Of dit komt doordat de ML-schatte van B gedefinieerd wordt door een randminimum is niet onderzocht. In afbeelding 3.3. is te zien dat de lijn bepaald met de methode der momenten veel flauwer loopt. Dit is van invloed op de bepaling van het kwantiel.

Om (het gemiddelde van) het 10^{-4} -kwantiel vast te stellen, wordt de verdelingsfunctie zo herschreven dat de waterstand, x_p , een functie is van de overschrijdingskans, p . Voor de exponentiële verdeling is dat:

$$x_p = A - B \cdot \ln(p).$$

Door de gemiddelden van de schatters, $\mu(\hat{A})$ en $\mu(\hat{B})$, te substitueren voor A en B wordt het gemiddelde van het kwantiel, $\mu(X_{10^{-4}})$, vastgesteld. De uitkomsten zijn vermeld in Tabel 3.6.

Voor de variantie van een kwantiel geldt:

$$\text{Var}(x_p) = \text{Var}(A^2) + \text{Var}(B^2) \cdot (\ln(p))^2$$

Hiermee is de standaardafwijking van het 10^{-4} -kwantiel te bepalen. Het is niet van grote invloed op die standaardafwijking dat voor de ML-schatte van A geen onzekerheid wordt meegenomen.

Exponentieel, N=530	$\mu(X_{10^{-4}})$	$\sigma(X_{10^{-4}})$	$F_A(X_{10^{-4}})$	$F_B(X_{10^{-4}})$
MdM	NAP +4.885 m	0.111 m	NAP +4.885 m	NAP +4.902 m
ML	NAP +5.10 m	0.10 m	NAP +5.10 m	NAP +5.119 m
KKW	NAP +4.957 m	0.157 m	NAP +4.957 m	NAP +4.989 m

Tabel 3.6

De tweede manier om statistische onzekerheid mee te nemen is de Bayesiaanse methode. De onzekerheid wordt daarbij uit de verdelingsfunctie geïntegreerd. Een verdelingsfunctie heeft een statistisch onzekere parameter A. Zoals ook in paragraaf 3.2 wordt de onzekerheid weergegeven door ϵ . De verdelingsfunctie wordt vermenigvuldigd met de kansdichtheidsfunctie van ϵ en vervolgens geïntegreerd over ϵ . In dit geval is de daar uit volgende integraal vrij eenvoudig op te lossen. De werkwijze wordt verduidelijkt in Bijlage 9. Als grenzen voor integratie wordt gekozen $\epsilon = -\infty$ en $\epsilon = +\infty$.

$$F_A(x) = \int_{\epsilon=-\infty}^{\epsilon=+\infty} e^{-\frac{(x-A-\epsilon)}{B}} \cdot \frac{1}{\sigma_\epsilon \sqrt{2\pi}} \cdot e^{-\frac{1}{2} \left(\frac{\epsilon}{\sigma_\epsilon}\right)^2} d\epsilon \rightarrow F_A(x) = e^{-\frac{(x-A)}{B}} \cdot e^{-\frac{\sigma_\epsilon^2}{2B^2}}$$

Voor de standaardafwijking van de schatter van A is bepaald: $\sigma_A = \frac{B}{\sqrt{N}}$
 Als dat gesubstitueerd wordt volgt:

$$F_A(x) = e^{-\frac{(x-A)}{B}} \cdot e^{-\frac{1}{2N}}$$

De gevolgen van een onzekerheid rond A zijn terug te brengen tot uitsluitend de invloed van het aantal waarnemingen, N. In afbeelding 3.4 is de 'nieuwe' verdelingsfunctie weergegeven voor verschillende waarden van N. De overschrijdingslijn verschuift bij een klein aantal waarnemingen een klein stukje naar boven parallel aan de oorspronkelijke overschrijdingslijn. Gesteld kan worden dat de invloed van een onzekere parameter A verwaarloosbaar is. Zie ook Tabel 3.6.

De gevolgde rekenwijze wordt ook toegepast op een normaal verdeelde parameter B. De verdelingsfunctie wordt nu:

$$F_B(x) = \int e^{-\frac{(x-A)}{(B+\epsilon)}} \cdot f(\epsilon) d\epsilon$$

Aangezien de ϵ in de noemer van de e-macht staat, kan deze integraal slechts numeriek opgelost worden.

De overschrijdingslijn met onzekerheid rond B, wordt niet alleen beïnvloed door het aantal waarnemingen, N, maar ook door de waterstand, x. Dit is te zien in grafiek 3.5. Hoe kleiner het aantal waarnemingen, hoe meer de lijn afwijkt van de rechte. Bij een constant aantal waarnemingen zal de lijn meer afwijken van de rechte naarmate de overschrijdingsfrequentie kleiner wordt. Bij de bepaling van het 10^{-4} -kwantiel is echter gebruik gemaakt van 530 waarnemingen. De invloed van statistische onzekerheid is nagenoeg verdwenen. Zie Tabel 3.6. $F_B(X_{10^{-4}}) \approx \mu(X_{10^{-4}})$ Het verschil tussen het 10^{-4} -kwantiel met en zonder onzekerheid bedraagt nog slechts 0,03m.

Opvallend is dat de invloed van de keuze van de schattingsmethode van groter invloed is op het kwantiel dan de invloed van de statistische onzekerheid. Een mogelijke oorzaak daarvoor is een niet-homogene reeks met waarnemingen. De waterstanden zullen niet exponentieel verdeeld zijn. De onzekerheid van de parameters is weliswaar niet groot is door het grote aantal waarnemingen dat geselecteerd is, maar de onzekerheid van het model, die in dit onderzoek verwaarloosd is, speelt daarnaast nog een rol.

3.4 Vergelijking van de overschrijdingslijnen

3.4.1 Inleiding

Met behulp van twee reeksen onafhankelijke waarnemingen zijn de parameters van de exponentiële, de lognormale en de Gumbelverdeling bepaald met behulp van drie schattingmethoden. Van iedere verdeling zijn nu drie sets met parameters beschikbaar. Om te bepalen met welke verdeling en welke set parameters in het volgende hoofdstuk verder gerekend zal worden, worden in deze paragraaf de verschillende overschrijdingslijnen met elkaar vergeleken.

3.4.2 Resultaten

In afbeelding 3.6 zijn de verschillende overschrijdingslijnen weergegeven. De lognormale verdeling in een lagere overschrijdingslijn resulteert dan de andere twee verdelingen. Hetzelfde kwantiel heeft een beduidend lagere waarde. Bij de econometrische analyse van de optimale dijkhoogte wordt bij voorkeur gemaakt van een overschrijdingslijn die hoog ligt, omdat dit veiliger is. Daarom wordt gekozen voor de exponentiële en de Gumbel-verdeling.

De exponentiële verdeling en de Gumbel-verdeling geven kwantielschattingen in de zelfde orde van grootte. De Gumbel-verdeling, gebaseerd op de geselecteerde jaarmaxima geeft iets lagere kwantielschattingen dan de exponentiële verdeling van de waterstanden die geselecteerd zijn met de POT-methode. (Zie afbeelding 3.6) Op grond hiervan heeft de exponentiële verdeling een lichte voorkeur.

Een voordeel van de exponentiële verdeling boven de Gumbel verdeling is een eenvoudig, inzichtelijk functievoorschrift.

Een tweede voordeel is dat de parameters van de exponentiële verdeling geschat zijn met behulp van 530 waterstanden die geselecteerd zijn met de POT-methode in plaats van 108 jaarmaxima. Schatters die bepaald zijn met behulp van 108 waarnemingen hebben een grotere statistische onzekerheid dan schatters die bepaald zijn met 530 waarnemingen. De selectie op basis van jaarmaxima is echter een fysisch beter te onderbouwen selectie. De modelonzekerheid van de jaarmaxima is kleiner dan van de waterstanden die geselecteerd zijn met de POT-methode.

Omdat de exponentiële en de Gumbel-verdeling een eenvoudig functievoorschrift, een 'veilige' overschrijdingslijn hebben, zal in dit verslag verder gewerkt worden met deze twee verdelingen.

Een tweede keuze dient gemaakt te worden wat betreft de te gebruiken schattingsmethoden. Bij de exponentiële verdeling viel op dat de invloed van de keuze van de schattingsmethode groter was dan de invloed van de statistische onzekerheid.

De statistische onzekerheid van parameter A van de exponentiële verdeling op de overschrijdingslijn bleek verwaarloosbaar. In dit afstudeeronderzoek zal verder uitsluitend naar de onzekerheid van parameter B gekeken worden. De kansdichtheidsfunctie van de schatter van B kon uitsluitend in geval van de methode van de Maximum Likelihood op analytische wijze benaderd worden.

De meest hoge kwantiel-schattingen en daardoor de meest 'veilige' overschrijdingslijn geeft de Maximum Likelihood. Normaliter geeft een overschrijdingslijn die geschat is met de methode van de Kleinste Kwadraten de hoogste schatting van het 10^{-4} -kwantiel. Dat blijkt hier niet het geval. Zoals in afbeelding 3.7 te zien is, is de Maximum Likelihood-schatter veel gevoeliger voor de keuze van de drempelwaarde dan de andere schattingsmethoden. Bij iedere drempel hoger dan NAP +1.65m is de overschrijdingslijn bepaald met de Maximum Likelihood de hoogste overschrijdinglijn. De methode der Momenten geeft een lage overschrijdingslijn, wat nadelig is.

Omdat de methode van de Maximum Likelihood niet alleen een 'veilige' overschrijdingslijn geeft, maar ook een analytische afleiding van de onzekerheid van parameter B heeft, wordt in dit verslag verder gewerkt met deze methode.

De optimale dijkhoogte

4.1 Inleiding

Een absoluut maximum voor de waterstanden is niet aan te geven. Welke dijkhoogte ook gekozen wordt, altijd blijft er een kans, hoe klein ook, dat een hogere waterstand zal voorkomen. Een absolute veiligheid tegen overstromen is dus niet te garanderen, welke stormvloedstand ook als uitgangspunt genomen wordt. Desondanks zal tot een zekere dijkhoogte besloten moeten worden.

In 1956 was Van Dantzig [3] de eerste om de keuze voor de maatgevende hoogwaterstand theoretisch te motiveren. Hij koos niet de waterstand, maar een zeker veiligheidsniveau als uitgangspunt. Overschrijdingslijnen van waterstanden werden gebruikt om te bepalen met welk risico rekening gehouden zou moeten worden.

In zijn model werd de bepaling van het aanvaardbare risiconiveau als een mathematisch-economisch beslissingsprobleem beschouwd. Dat wil zeggen dat het risico van een gebeurtenis in geld uitgedrukt wordt en afgewogen wordt tegen de kosten van het voorkomen van die gebeurtenis. Het hier gebruikte model is gebaseerd op dat econometrische beslissingsmodel van Van Dantzig. [3]

De econometrische analyse van het optimale veiligheidsniveau werd uitgevoerd voor Centraal Holland. Wat betreft economie en inwoneraantal is dat het meest belangrijke gebied van Nederland te noemen. Centraal Holland is het gebied dat begrensd wordt door de Noordzee, de Rotterdamse Waterweg, de Nieuwe Maas, de Lek, Het Amsterdam-Rijnkanaal, Het IJ en Het Noordzeekanaal. Zie afbeelding 4.1.

Bij een econometrische analyse van een dijk wordt de dijkhoogte zo gekozen, dat de som van de kosten voor dijkverbetering en de gekapitaliseerde schadeverwachting minimaal is. Het aldus berekende peil is een ramppeil. Overschrijding van het peil leidt tot een ramp. Overschrijding van het in Hoofdstuk 3 bepaalde ontwerppeil hoeft niet meteen tot een ramp te leiden.

Om de optimale dijkhoogte te bepalen wordt uitgegaan van enkele vooronderstellingen.

- De beschouwing wordt beperkt tot één enkele polder, die door dijken tegen de zee beschermd is.
- Het overstromen van de dijken bij een waterstand hoger dan de dijkhoogte is het enige bezwijkmechanisme.
- Bij inundatie van de polder gaan alle goederen volledig verloren.
- Het verlies van mensenlevens bij een overstroming valt buiten deze beschouwing.

Aangenomen is dat de waterstanden beschreven kunnen worden door een exponentiële verdeling met twee parameters. De waarden van de parameters zijn in hoofdstuk 3 met de methode van de Maximum-Likelihoodschatters bepaald.

In dit hoofdstuk zal eerst het optimalisatiemodel besproken worden. Vervolgens zal nagegaan worden wat de invloed is van de onzekerheid van schatters op de econometrische dijkhoogtebepaling. De overige variabelen in het model zijn ook geen deterministische grootheden. Er zal tevens onderzocht worden wat de invloed is op de dijkhoogtebepaling als die variabelen als stochastisch beschouwd worden.

4.2 Het model

4.2.1 Inleiding

In deze paragraaf wordt het economisch-mathematisch model besproken dat door de Deltacommissie beschreven is. [2c]. De investeringen in een veiliger systeem worden vergeleken met de daling van de contante waarde van het jaarlijkse risico.

De optimale situatie wordt gevonden door de som van de kosten van de dijkverhoging en de contante waarde van de schadeverwachting te minimaliseren.

De te minimaliseren functie is dus:

$$K = I + R$$

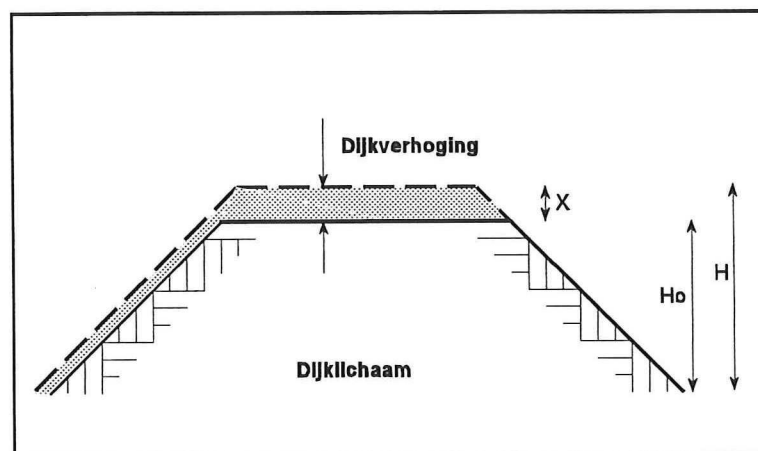
K = Totale kosten ; I = Investerings, kosten van de dijkverhoging ; R = Risico, schadeverwachting.

4.2.2 De investeringen

De hoogte van de bestaande dijk is een deterministische grootte. Verondersteld wordt dat de gehele dijk overal de hoogte H_0 heeft. De dijk zal over de gehele lengte met X meter verhoogd worden tot een nader te bepalen niveau H .

$$\Rightarrow X = H - H_0$$

X = Ophoging ; H = Nieuwe dijkhoogte ; H_0 = Huidige dijkhoogte.



Afbeelding 4.2

Het verhogen van de dijken brengt kosten, I , met zich mee, die een functie zullen zijn van de ophoging X . Als de dijkverhoging klein is, dan zijn de kosten I een lineaire functie van de verhoging, X , en onafhankelijk van het huidige niveau van de dijk, H_0 . De mobilisatiekosten voor de dijkverbetering, I_0 , zijn kosten die hoe dan ook gemaakt zullen worden als de dijken verhoogd moeten worden.

$$I = I_0 + I' X$$

I_0 = Mobilisatiekosten; I' = Kosten per meter dijkverhoging.

4.2.3 De schadeverwachting

In dit econometrisch model wordt aangenomen dat overstromen het enige bezwijkmechanisme van de dijk is. Als de waterstand hoger is dan de dijkhoogte vindt een ramp plaats waarbij alle goederen in de polder verloren gaan.

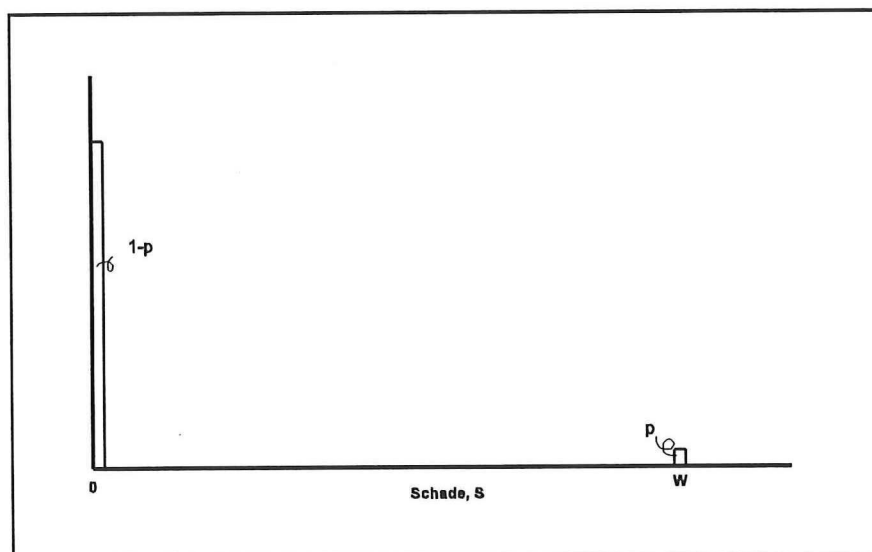
De kans p op een ramp is afhankelijk van de overschrijdingslijn van de waterstanden en de dijkhoogte. De overschrijdingslijn van de waterstanden, h , wordt in dit model gegeven door een exponentiële verdeling of een Gumbel-verdeling. Om het model te verduidelijken zal gebruik gemaakt worden van de exponentiële verdeling:

$$F(h) = e^{-\frac{h-A}{B}}$$

De verdelingsfunctie geeft de kans weer dat een waterstand groter wordt dan het peil h . Als een waterstand groter wordt dan de dijkhoogte, H , dan is er sprake van een ramp. Als in plaats van h H geschreven wordt, geeft deze functie de kans, p , op een ramp bij een zekere dijkhoogte H weer.

De door de dijken te beschermen waarde, W , bestaat uit de totale reële waarde van alle goederen in de polder vermeerderd met de kosten van dijkherstel en droogleggen van de polder na een overstroming.

De kansdichtheidsfunctie van de schade is in afbeelding 4.3 weergegeven. Er is een kans $1-p$ dat de schade gelijk is aan 0 en een hele kleine kans p op een ramp waarbij W verloren gaat. Opgemerkt dient te worden dat de kans gelijk is aan de oppervlakte onder de grafiek en niet aan de hoogte.



Afbeelding 4.3

In bijlage 10 wordt de wiskundige verwachting van de totale schade S afgeleid. Hier zal verder gerekend worden met het resultaat.

De schadeverwachting per jaar is $E[S]=F(h).W$; De kans op een ramp vermenigvuldigd met de gevolgen. Hierbij wordt ervan uitgegaan dat de schade binnen een jaar hersteld kan worden. Stel nu dat het bedrag $F(h).W$ gereserveerd moet worden om de verliezen van het komende jaar te dekken. Als de gereduceerde rentevoet in procenten per jaar r is (zie bijlage 10), dan wordt, met $h=H=H_0+X$, de totale verdisconteerde schadeverwachting R (met exponentiële verdeling):

$$R = F(h) \cdot \frac{W}{r} = \frac{W \cdot e^{-\frac{(H_0+X)-A}{B}}}{r}$$

R kan gezien worden als het bedrag dat nu gereserveerd dient te worden om de schade in de toekomst te kunnen betalen.

4.2.4 De optimale dijkhoogte

In § 4.2.2 zijn de investeringen gedefinieerd en in §4.2.3 de schadeverwachting. Opgeteld leveren ze de totale kosten, K.

$$K(X) = I + R = I_0 + I'X + \frac{W \cdot F(H_0 + X)}{r}$$

De economisch optimale dijkhoogte wordt gevonden door de kostenfunctie te minimaliseren. Het minimum wordt gevonden door de functie naar X te differentiëren en het differentiecoëfficiënt gelijk aan 0 te stellen.

$$K'(X) = I' + \frac{W \cdot e^{-\frac{((H_0 + X) - A)}{B}}}{Br} = 0$$

Hiermee kan de optimale dijkhoogte, $H_0 + X$, vastgesteld worden.

$$H_0 + X = A - B \cdot \ln\left(\frac{I'Br}{W}\right)$$

Dit geldt voor waterstanden die als exponentieel verdeeld beschouwd worden. In het geval van een Gumbel-verdeling geldt voor de optimale dijkhoogte:

$$H_0 + X = A - B \cdot \ln\left(-\ln\left(\frac{I'Br}{W}\right)\right)$$

De Delta commissie heeft voor de bepaling van de economisch optimale dijkhoogte van Centraal Holland de volgende waarden gebruikt:

Mobilisatiekosten: $I_0 = 110 \cdot 10^6$ gld.
 Kosten per meter dijkverhoging: $I' = 40 \cdot 10^6$ gld/m.
 Huidige dijkhoogte: $H_0 = \text{NAP} + 5.0\text{m}$.
 Gemiddelde, gereduceerde rentevoet : $r = 1.5 \%$
 Te beschermen waarde : $W = 24.2 \cdot 10^9$ gld.

Voor de schatters van de parameters worden de waarden gebruikt die in het vorige hoofdstuk zijn bepaald met de Methode van de Maximum Likelihood.

Exponentiële verdeling: $A=2.329$; $B=0.301$;
 Gumbel-verdeling: $A=2.361$; $B=0.2664$;

In afbeelding 4.3 zijn de investeringen, I, de contante waarde van het risico, R, en de som van die twee, de totale kosten, uitgezet tegen de dijkhoogte.

Als de overschrijdingslijn een exponentiële verdeling is, dan is de economisch optimale dijkhoogte: $h_{opt} = \text{NAP} + 5,88\text{m}$. De functie van de totale kosten heeft daar een minimum. De bijbehorende overschrijdingskans: $F(h_{opt}) = 7,52 \cdot 10^{-6}$ per jaar. Voor de Gumbel-verdeling geldt dat de economisch optimale dijkhoogte $h_{opt} = \text{NAP} + 5,51\text{m}$ en de bijbehorende overschrijdingskans: $F(h_{opt}) = 6,43 \cdot 10^{-6}$ per jaar.

4.3 De invloed van de onzekerheid van de verdelingsparameter

4.3.1 Inleiding

De optimale dijkhoogte is bepaald met behulp van een exponentiële en een Gumbel-verdeling met twee constante parameters. Zoals in hoofdstuk 2 besproken is zijn de schatters van parameters ook stochasten.

De economisch optimale dijkhoogte wordt nogmaals bepaald, maar nu wordt parameter B als een normaal verdeelde stochast beschouwd. Dit gebeurt, net zoals in § 3.3, door een normaal verdeelde onzekerheid ϵ aan de parameter toe te voegen. Die onzekerheid kan vervolgens met de Bayesiaanse methode uit de verdelingsfunctie geïntegreerd worden. De onzekerheid van de verschuivingsparameter A bleek bij beide verdelingen verwaarloosbaar (§ 3.3) en daarom zal in dit hoofdstuk uitsluitend naar de invloed van de onzekerheid van de schatter van parameter B gekeken worden.

4.3.2 De invloed van de onzekerheid op de totale kosten

De totale kosten als functie van de dijkverhoging, X, en de waterstanden exponentieel verdeeld:

$$K(X) = I + R = I_0 + I'X + F(h) \cdot \frac{W}{r} \quad \text{met} \quad F(h) = e^{-\frac{(h-A)}{B}}$$

Aangenomen wordt dat de schatter $\hat{B}$ normaal verdeeld en zuiver is. De schatter wordt weergegeven als een sommatie van B en een normaal verdeelde onzekerheid ϵ . ($\mu_\epsilon = 0$, $\sigma_\epsilon = B/\sqrt{N}$) De onzekerheid ϵ kan op Bayesiaanse wijze uit de verdelingsfunctie geïntegreerd worden door de $F(h)$ te vermenigvuldigen met de kansdichtheidsfunctie van ϵ , $f(\epsilon)$, en te integreren over ϵ . Zie § 3.3.

$$F(h) = \int_{-3\sigma}^{3\sigma} e^{-\frac{(h-A)}{(B+\epsilon)}} \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{\epsilon}{\sigma}\right)^2} d\epsilon$$

Met deze verdelingsfunctie worden de totale kosten opnieuw bepaald. De verdelingsfunctie $F(h)$, waarin de onzekerheid van B is meegenomen, is slechts numeriek te bepalen. De totale kosten, $K(X)$, kunnen dus ook uitsluitend numeriek bepaald worden.

Op dezelfde wijze kunnen de totale kosten, $K(X)$, bepaald worden als de waterstanden Gumbel-verdeeld zijn. De schatter $\hat{B}$ is wederom een sommatie van B en een normaal verdeelde onzekerheid ϵ met $\mu_\epsilon = 0$, $\sigma_\epsilon = B/\sqrt{N}$. De verdelingsfunctie waar de onzekerheid uit geïntegreerd is, wordt gedefinieerd door:

$$F(h) = \int_{-3\sigma}^{3\sigma} e^{-e^{-\frac{(h-A)}{(B+\epsilon)}}} \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{\epsilon}{\sigma}\right)^2} d\epsilon$$

Op numerieke wijze kunnen hieruit de totale kosten, $K(X)$, bepaald worden.

4.3.3 Resultaten

In afbeelding 4.4 zijn de totale kosten uitgezet tegen de dijkhoogte. De waterstanden zijn als exponentieel verdeeld beschouwd. Een Gumbel-verdeling heeft een gelijkvormige, maar iets lager gelegen grafiek tot gevolg.

De hooggelegen gestippelde lijn geeft de totale kosten weer als de parameters geschat zijn uit 10 waarnemingen. De getrokken lijn geeft de totale kosten weer, als de parameters van als constanten gezien worden. Naarmate de onzekerheid van de schatters kleiner wordt komen de totale kosten dichter bij de totale kosten zonder onzekerheid te liggen.

Als parameter B van de exponentiële verdeling onzeker is, dan verschillen de kosten in het gebied met een dijkhoogte lager dan $\pm$ NAP +6.5m aanzienlijk met de kosten waarbij geen rekening is gehouden met onzekerheid. Als de schatter van B gebaseerd zou zijn op 10 waarnemingen dan zijn de kosten minimaal bij een dijkhoogte van NAP +6.77m. De optimale dijkhoogte bij 100 waarnemingen is NAP +6.0m. De invloed van de statistische onzekerheid is praktisch verwaarloosbaar geworden als de overschrijdingslijn geschat is uit 530 waarnemingen. De econometrisch optimale dijkhoogte is dan NAP +5.91m. Dat verschilt slechts 0.03 meter met de dijkhoogtebepaling waarbij statistische onzekerheid werd verwaarloosd.

In het geval van een Gumbel-verdeling verschilt de optimale dijkhoogte, waarbij wel rekening is gehouden met statistische onzekerheid 0.16 meter met de optimale dijkhoogte, waarbij de onzekerheid wordt verwaarloosd. Als parameter B geschat zou worden uit slechts 10 waarnemingen dan ligt de optimale dijkhoogte 0.82meter hoger.

In tabel 4.1 staan voor het overzicht de minimale kosten, de optimale dijkhoogtes met de bijbehorende overschrijdingskans weergegeven voor zowel de exponentiële als de Gumbel-verdeling. Hoe groter de onzekerheid van de parameters, dus hoe minder waarnemingen, hoe hoger de optimale dijkhoogte komt te liggen.

Het volgende dient hierbij opgemerkt. De invloed van de onzekerheid van de parameters is niet groot bij de exponentiële verdeling, omdat met de POT-methode 530 waarnemingen zijn geselecteerd. De modelonzekerheid is echter verwaarloosd bij dit onderzoek. In hoofdstuk 3 bleek reeds dat de invloed van die onzekerheid aanzienlijk is. De reeks waarnemingen geselecteerd met de POT-methode is geen homogene reeks. De modelonzekerheid is bij de jaarmaxima veel kleiner.

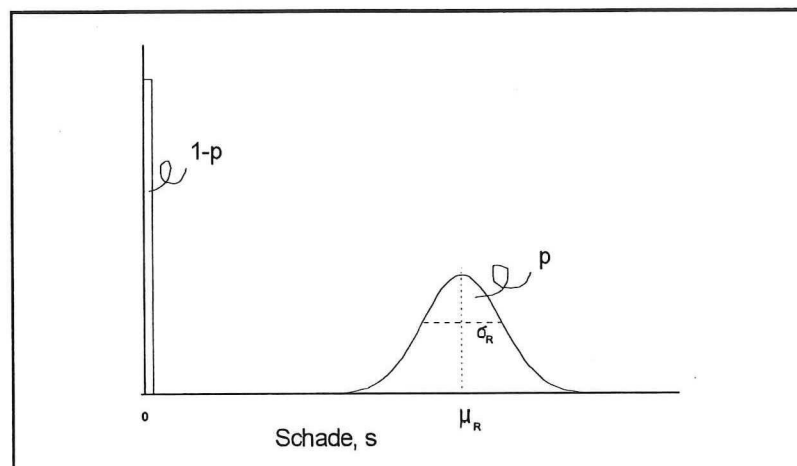
4.4 De invloed van de onzekerheid van de kosten van dijkverhoging en de schade

4.4.1 Inleiding

Bij de berekening van de totale kosten is totnogtoe alleen in beschouwing genomen dat de parameters van een verdelingsfunctie onzeker zijn. De investeringen en de schade zijn daarbij als deterministische waarden beschouwd. In de werkelijkheid zullen die waarden ook niet exact te bepalen zijn.

In deze paragraaf worden de rollen omgedraaid. De kosten van dijkverhoging (I_0 , I') en van de schade in het geval van een ramp (W) worden als stochastisch beschouwd, terwijl de parameters van de verdelingsfunctie constant gehouden worden. De onzekerheid van de parameters wordt buiten beschouwing gelaten om de som niet te compliceren. Als verdelingsfunctie wordt gekozen voor de exponentiële verdeling.

Aangenomen wordt dat I_0 , I en W normaal verdeeld zijn. Dit is een vereenvoudigde weergave van de werkelijkheid. De kansdichtheidsfunctie van de totale schade is weergegeven in afbeelding 4.6.



Afbeelding 4.6

4.4.2 De invloed van de onzekerheid op de totale kosten

Met I_0 , I' en S als stochastisch, wordt de definitie van de totale kosten K :

$$K(X) = I + R = I_0 + I'X + \frac{S}{r}$$

Het gemiddelde van de totale kosten is de som van de verwachtingswaarden van de stochastische variabelen. De verwachtingswaarde van de schade is de kans vermenigvuldigd met het gemiddelde van het gevolg: $p \cdot \mu(W)$. (Zie ook bijlage 10)

$$\rightarrow \mu(K) = \mu(I_0) + \mu(I') \cdot X + \frac{p \cdot \mu(W)}{r}$$

Om de standaardafwijking van de totale kosten te bepalen kunnen de standaardafwijkingen van de verschillende stochasten niet zondermeer opgeteld worden. De afleiding van de variantie van de schade is beschreven in Bijlage 10. In die bijlage staat ook beschreven hoe de contante waarde van de variantie van de schade gevonden kan worden. Hier wordt verder gerekend met de uitkomst van de variantie van de totale schade:

$$\sigma^2(K) = \sigma^2(I_0) + \sigma^2(I') \cdot X^2 + \frac{p \cdot [\sigma^2(W) + (1-p)\mu^2(W)]}{(1+r)^2 - 1}$$

De gemiddelde waarden van de stochastische variabelen zijn gelijk aan de waarden die door de Deltacommissie gebruikt zijn. Zie Tabel 4.2

De kosten van dijkverhoging zijn goed van tevoren vast te stellen. Voor de variatiecoëfficiënt ($V = \sigma/\mu$) van I_0 en I' is daarom een waarde van 0,1 gekozen. De schade bij een overstroming is veel minder nauwkeurig bekend. Voor de variatiecoëfficiënt van de schade is gekozen voor een waarde van 0.3. Hiermee zijn de standaardafwijkingen bepaald. (Zie tabel 4.2)

	gemiddelde, μ	variatiecoëfficiënt	standaardafwijking, σ
I_0	$110 \cdot 10^6$ gld	$V_{I_0} = 0.1$	$11 \cdot 10^6$ gld
I'	$40 \cdot 10^6$ gld/m	$V_{I'} = 0.1$	$4 \cdot 10^6$ gld/m
S	$24,2 \cdot 10^9$ gld	$V_S = 0.3$	$7,26 \cdot 10^9$ gld

Tabel 4.2

De overige parameters zijn constant gehouden: $r=1.5\%$; $A=2.329$; $B=0.301$.

In Afbeelding 4.7 zijn het gemiddelde en de standaardafwijking van de totale kosten weergegeven. Het is opvallend dat de standaardafwijking van de totale kosten voor dijkhoogtes kleiner dan $\pm \text{NAP} + 6.5\text{m}$ enorm groot is ten opzichte van het gemiddelde. Dit komt doordat in de variantie van de schadeverwachting het gekwadrateerde gemiddelde meetelt. (Zie ook de betreffende vergelijking.) Bij grotere dijkhoogtes valt de hele schade-term weg, omdat de overschrijdingskans dan erg klein is.

Het minimum van de totale kosten kan nu niet meer zondermeer bepaald worden. De optimale dijkhoogte behorende bij het gemiddelde van de totale kosten is natuurlijk gelijk aan de dijkhoogte die bepaald is door de Delta commissie. Voordat het minimum vastgesteld kan worden, moet beslist worden hoe vaak de standaardafwijking meegeteld gaat worden, met andere woorden, hoe groot de risico-aversie is. Mensen nemen bij voorkeur het zekere voor het onzekere. Ze investeren liever iets meer in een hogere dijk dan dat ze geld reserveren voor de onzekere gevolgen van een overstroming.

De totale kosten bij een zekere dijkhoogte kunnen dan weergegeven worden als een functie van het gemiddelde en de standaardafwijking op die hoogte:

$$K(x) = \mu(K(x)) + k \cdot \sigma(K(x))$$

De vraag wordt dus : Hoe groot moet k zijn ?

In afbeelding 4.8 staan de totale kosten en de bijbehorende optimale dijkhoogte weergegeven voor $k=1/2, k=1$ en $k=1 1/2$. Hoe meer de standaardafwijking meetelt, des te hoger de optimale dijkhoogte is. Aan de ene kant heeft dit een grotere veiligheid tot gevolg, aan de andere kant worden de kosten ook hoger. Als de standaardafwijking één keer meegeteld wordt, dan vallen de kosten al meer dan $1 1/2$ keer zo hoog uit. De overschrijdingskans wordt daarbij 20 keer zo klein. Zie ook tabel 4.3.

	Minimale kosten, $K_{\min}$ in 10^6 gld.	Optimale dijkhoogte, h_{opt} in m +NAP	Overschrijdingskans, $F(h)$
$k=0$	157	5,88	$7,52 \cdot 10^{-6}$
$k=1/2$	221	7,13	$1,18 \cdot 10^{-7}$
$k=1$	241	7,48	$3,70 \cdot 10^{-7}$
$k=1 1/2$	255	7,65	$2,10 \cdot 10^{-8}$

Tabel 4.3

Een andere keuze moet gemaakt worden wat betreft de grootte van de variatiecoëfficiënt van de normaal verdeelde variabelen. Als de waarde van de coëfficiënt lager is, is de spreiding van de stochastische variabele kleiner. Een redelijke waarde van de variatiecoëfficiënt ligt tussen 0,1 en 0,5. Voor de coëfficiënt van de investeringen is een waarde van 0,1 gekozen en voor de schade een waarde van 0,3.

Om de invloed van de keuze te beschouwen is eerst de variatiecoëfficiënt van de investeringen gevarieerd (afbeelding 4.9) en vervolgens de coëfficiënt van de schade (afbeelding 4.10). Bij de bepaling van de totale kosten is één keer de standaardafwijking meegenomen ($k=1$).

De invloed van een variërende coëfficiënt van de investeringen is aanzienlijk. Hoe kleiner de standaardafwijking van de investeringen, hoe lager ook de totale kosten, met één standaardafwijking, worden, terwijl de optimale dijkhoogte hoger komt te liggen. Een lage variatiecoëfficiënt van de investeringen geeft dus een veiliger, goedkoper systeem. Ter verduidelijking zijn de grafieken in afbeelding 4.11 gecreëerd (Vervolg volgende bladzijde).

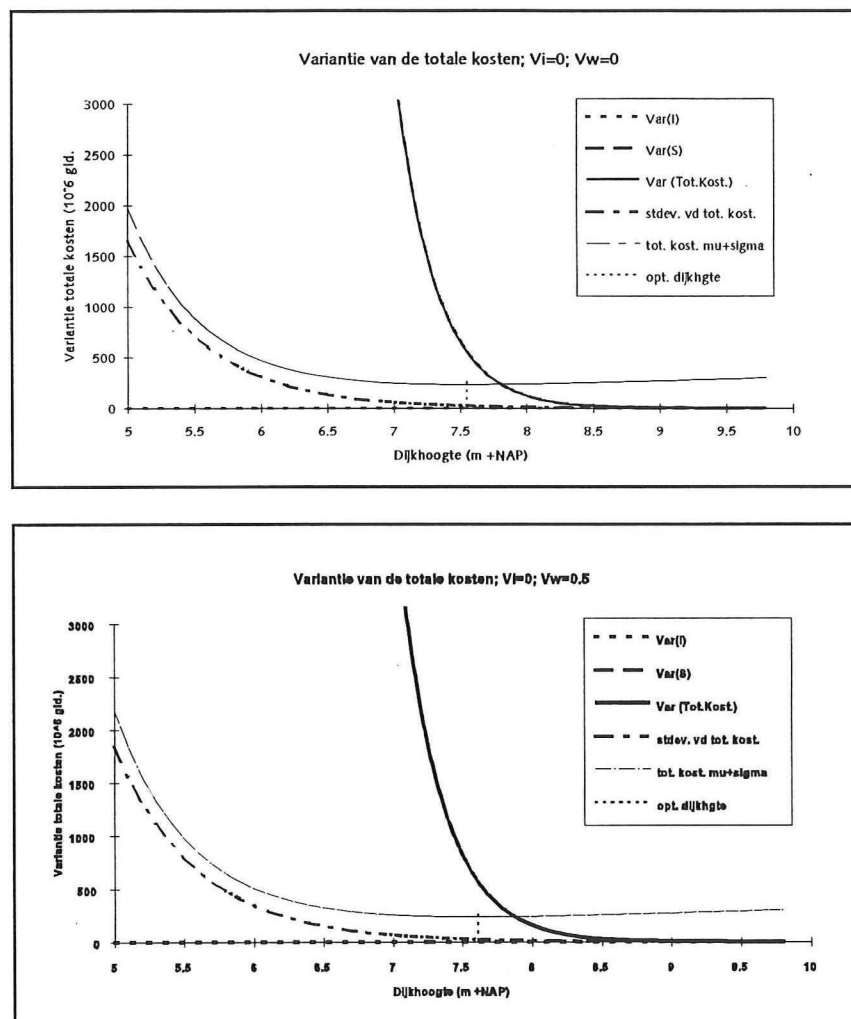
In de afbeelding staan de variantie van de investeringen, de variantie van de totale schade en de som van die twee, de variantie van de totale kosten afgebeeld. Om inzicht te krijgen in de orde van grootte zijn ook de standaardafwijking van de totale kosten en het gemiddelde plus één standaardafwijking van de totale kosten weergegeven.

De standaardafwijking van de totale kosten wordt weergegeven door de onderste dunne lijn. Een grotere variantie van de investeringen 'tilt' als het ware de staart van de standaardafwijking van de totale kosten op. Door het 'optillen' van de staart wordt de dijkhoogte met de minimale kosten lager en gaan de bijbehorende minimale kosten omhoog.

Zoals te zien in afbeelding 4.10, is de invloed van de variantie van de schade in geval van een ramp, W , verwaarloosbaar. Dit is een gevolg van de definitie van de variantie van de totale schade.

$$\sigma^2(S) = \frac{p \cdot [\sigma^2(W) + (1-p)\mu^2(W)]}{(1+r)^2 - 1}$$

Deze variantie bestaat voornamelijk uit het gekwadrateerde gemiddelde van de te beschermen waarde W . Een variërende $\sigma^2(W)$ heeft daar nauwelijks invloed op, zoals ook te zien is in afbeelding 4.12. Bij de linker grafiek is de variatiecoëfficiënt van de te beschermen waarde W gelijk aan 0. Bij de rechter grafiek is de variatiecoëfficiënt gelijk aan 0.5. De grafieken zijn ondanks dat grote verschil praktisch identiek. De standaardafwijking van de totale kosten wordt iets groter bij een dijkhoogte van NAP +5m, maar dat heeft geen invloed op de optimale dijkhoogte. De optimale dijkhoogte is bij een variatiecoëfficiënt van 0.5 slechts 0.06 meter hoger dan bij een variatiecoëfficiënt van 0.



Afbeelding 4.12

4.5 Kansdichtheid van de totale kosten

4.5.1 Inleiding

Om niet alleen inzicht te krijgen in de minimale kosten als functie van het gemiddelde, de standaardafwijking en de variatiecoëfficiënt, maar ook meer te weten over de 'kansdichtheidsfunctie' van de kosten bij een zekere dijkhoogte, is een Monte Carlo-simulatie uitgevoerd.

Het doel van een Monte Carlo simulatie is om meer inzicht te krijgen in een stochast die een functie is van verschillende stochastische variabelen. Dit inzicht wordt verkregen door een groot aantal malen voor alle stochastische variabelen een waarde uit de bijbehorende verdeling te 'trekken' en vervolgens de waarde van de gezochte stochast te bepalen. Door uiteindelijk het aantal malen dat een zekere uitkomst is voorgekomen te sommeren en te delen door het totale aantal uitkomsten wordt een benadering gevonden van de kansdichtheid van die uitkomst.

4.5.2 Kansdichtheidsfunctie van de totale kosten

In dit geval is er voor gekozen om de kansdichtheid van de kosten te bepalen bij een dijkhoogte van NAP +5m. Er is gekozen voor een zo laag mogelijke dijkhoogte, omdat de overschrijdingskans dan groot is en de kans op schade dus groot is. Dijkhoogtes lager dan NAP +5m zouden geen reële waarden voor de kosten van de dijkverhoging geven.

Opnieuw zijn de variabelen I_0 , I' en W normaal verdeeld. Het is mogelijk om, gebruikmakend van een randomgenerator, waarden uit een normale verdeling $N(\mu, \sigma)$ te trekken.

Om te beginnen wordt er een waarde uit een standaardnormale verdeling getrokken:

$$N(0,1) = \sqrt{-2 \cdot \ln(u)} \cdot \cos(2 \cdot \pi \cdot v)$$

Waarin u en v volkomen willekeurige getallen tussen 0 en 1 zijn.

Hiermee is vervolgens een willekeurig getal uit een normale verdeling met μ en σ te bepalen:

$$\rightarrow N(\mu, \sigma) = \sigma \cdot N(0,1) + \mu$$

De vergelijking van de totale kosten gegeven:

$$K(X) = I + R = I_0 + I'X + \frac{p \cdot W}{r}$$

p is een random getal tussen 0 en 1. Een stormvloedramp 'vindt plaats' als $p < 0.0001$. In de meeste jaren zal de schade dus gelijk aan 0 zijn.

Als in het k^e jaar de dijk overstroomt, zal de contante waarde van de schade gelijk zijn aan $\frac{W}{(1+r)^k}$.

De totale kosten zullen bepaald worden voor een levensduur van de dijk van 100 jaar. Als een langere levensduur gekozen zou worden, dan wordt de contante waarde van een schade die pas na een groot aantal jaren k op zal treden erg klein door de factor $1/(1+r)^k$.

Het is nu mogelijk om inzicht te krijgen in de kosten bij een dijkhoogte van NAP +5m. Als K keer de totale schade met een Monte-Carlo-simulatie wordt bepaald zullen naar verwachting $K \cdot [1 - (1-p)^N]$ gevallen met schade optreden. Waarin p de overschrijdingskans en N de levensduur van de waterkering is. Bij deze dijkhoogte van NAP +5m en een levensduur van 100 jaar is bij 1000 keer trekken een kleine 10 maal schade te verwachten. Om de verdeling van de gevallen met schade te bepalen moet de simulatie bijzonder vaak uitgevoerd worden.

Afbeelding 4.13 toont de uitkomst. De kansdichtheid wordt gegeven voor iedere bandbreedte van $1 \cdot 10^6$ gld. De uitkomst komt goed overeen met de kansdichtheidsfunctie van I_0 . De contant gemaakte schadegevallen en hun kansdichtheid, afbeelding 4.14, zijn weergegeven in bandbreedtes van $1 \cdot 10^9$ gld. Zelfs met logaritmische schalen is het niet mogelijk om de twee figuren in één grafiek weer te geven.

Omdat de puntjes in afbeelding 4.14 geen duidelijk beeld geven van het deel van de kansdichtheidsfunctie waar schade optreedt, wordt dat gedeelte in detail beschouwd. Daarom is de analytische kansdichtheid van de gevallen met schade bepaald. Als vele malen voor een levensduur N alleen naar de gevallen met schade gekeken zou worden, dan zou de kans dat de schade in jaar j optreedt gelijk zijn aan de kans dat hij in jaar k optreedt. In ieder jaar, $(1..N)$, kan de contante waarde van het gemiddelde van de schade bepaald worden. Als vervolgens gekeken wordt naar het aantal schadegevallen in een bandbreedte van $500 \cdot 10^9$ gld. dan geeft dat een inzicht in de verdeling van de contant gemaakte schade.

Een andere mogelijkheid om naar de kansdichtheidsfunctie in het deel van de schade te kijken is een numerieke methode. De totale kosten voor een levensduur van 100 jaar worden met een Monte-Carlo-simulatie bepaald, totdat zich 1000 gevallen met schade hebben voorgedaan. Alle contante waarden zijn gesorteerd en geteld in een bandbreedte van $500 \cdot 10^9$. De uitkomsten zijn weergegeven in afbeelding 4.15. De analytische benadering komt inderdaad overeen met de simulatie. Welbeschouwd is afbeelding 4.15 een close-up van afbeelding 4.14. De waarden op de y-as moeten nog wel met de overschrijdingskans p vermenigvuldigd worden.

Als de dijkhoogte groter wordt, zal de overschrijdingskans kleiner worden. De totale kosten zullen dus steeds minder gevallen van schade inhouden. De vorm van de kansdichtheid van de kosten zal gelijk blijven. Het gedeelte waarin tijdens de levensduur geen overstroming heeft plaatsgevonden, zal relatief groter van oppervlak worden. Het andere gedeelte, waarbij wel een ramp heeft plaatsgevonden, zal steeds kleiner worden. De vorm zal wel gelijk blijven.

Om te kijken of het constant houden van de rente nog enige invloed heeft op de verdeling van de contante waarde, is nogmaals een Monte-Carlo-simulatie uitgevoerd. De reële rente is als normaal verdeeld beschouwd, met $\mu=1.5\%$ en $\sigma=0.15\%$. Deze kleine marge is gekozen om te voorkomen dat ze kleiner dan 0 zou worden. Dit probleem had verholpen kunnen worden door in plaats van een normale een lognormale verdeling te kiezen.

Bij een schade in het k^e jaar wordt de contante waarde gevonden door de schade te delen door: $(1+r_1) \cdot (1+r_2) \dots (1+r_k)$, met $r_j = N(0.015, 0.0015)$.

Wederom zijn bij een dijk met een hoogte van NAP +5m en een levensduur van 100 jaar de totale kosten bepaald voor 1000 gevallen waarin een overstroming heeft plaatsgevonden. De resultaten zijn, gesorteerd in eerder genoemde bandbreedtes, weergegeven in afbeelding 4.16. Het blijkt dat de rente niet van grote invloed is op de kansdichtheidsfunctie van de schade. Naar verwachting is de invloed van de rente groter als de rente in opeenvolgende jaren niet als onafhankelijk worden gemodelleerd.

Conclusies

5.1 Conclusies

5.1.1 Inleiding

In dit onderzoek is de invloed van statistische onzekerheid op de bepaling van de ontwerpwaterstand en de econometrische dijkhoogtebepaling nader onder de loep genomen. Uit het onderzoek zijn een paar belangrijke conclusies te trekken. In §5.1.2 zullen de hoofdconclusies besproken worden. Gaandeweg zijn verscheidene kleinere zaken aan het licht gekomen. In §5.1.3 worden de aanvullende conclusies per hoofdstuk besproken.

5.1.2 Hoofdconclusies

- Voor het schatten van de parameters van kansverdelingen zijn drie methoden onderzocht. Methode der Momenten, de Methode van de Maximum Likelihood en de methode van de kleinste kwadraten met lineaire regressie. De methode van de Maximum Likelihood is erg efficiënt in vergelijking met de andere methoden. (§2.6.2)
- De geschatte parameters van verdelingen zijn met een statistische onzekerheid behept. Deze onzekerheid kan zowel analytisch (vaak) als numeriek bepaald worden. (§2.3 ..2.5)
- Om de invloed van de statistische onzekerheid op een overschrijdingslijn te beschouwen zijn twee methoden mogelijk. De ene methode geeft een gemiddelde en een standaardafwijking van een kwantiel. De tweede methode integreert de onzekerheid uit de verdelingsfunctie en resulteert in een toeslag op het kwantiel. (§3.1.2)
- De twee methoden zijn toegepast op de hoogwaterstanden in Hoek van Holland. De resultaten zijn weergegeven in tabel 5.1. De standaardafwijking van het 10^{-4} -kwantiel varieert tussen 0.11 en 0.25 meter voor de verschillende methoden en verdelingen. Als de onzekerheid uit de verdelingsfunctie geïntegreerd wordt, dan blijkt dat de invloed van de statistische onzekerheid van parameter A verwaarloosbaar is. De invloed van de statistische onzekerheid van parameter B is een verhoging van het 10^{-4} -kwantiel met ± 0.08 meter in het geval van de Gumbel-verdeling en een verhoging van ± 0.02 meter in het geval van de exponentiële verdeling. (§3.2,3.3)
- De selectie van de waterstanden met de POT-methode levert geen homogene reeks met waarnemingen op. De invloed van de keuze van de schattingsmethode is een belangrijke factor voor de bepaling van het 10^{-4} -kwantiel. (§3.2,3.3)
- De invloed van de statistische onzekerheid van de waterstanden op de bepaling van de economisch optimale dijkhoogte is verwaarloosbaar als de waterstanden met de POT-methode geselecteerd worden en als exponentieel verdeeld beschouwd worden. De optimale dijkhoogte bepaald met invloed van statistische onzekerheid ligt 0.03 meter hoger dan de optimale dijkhoogte zonder die onzekerheid. Als de waterstanden geselecteerd worden op basis van jaarmaxima is de modelonzekerheid kleiner, maar de invloed van de statistische onzekerheid groter. De optimale dijkhoogte ligt dan ten gevolge van statistische onzekerheid 0.16 meter hoger. (§ 4.3)

5.1.3 Aanvullende conclusies

- Conclusies van Hoofdstuk 1

Er zijn twee mogelijkheden behandeld om die onafhankelijke en homogene reeks te verkrijgen:

- Selectie op basis van de jaarmaxima (§ 1.2.4)
- Selectie met de Peak-Over-Threshold-methode. (§ 1.2.5)

De selectie op basis van jaarmaxima is een fysisch goed te onderbouwen selectie. Iedere waarneming komt voort uit één meteorologische jaarcyclus. Uit de beschikbare reeks waarnemingen van waterstanden worden 108 jaarmaxima geselecteerd (§ 1.2.4).

Het voordeel van de selectie met de POT-methode is dat 530 waarnemingen overblijven (§ 1.2.5).

Een nadeel is dat de afzonderlijke selectiecriteria vrij arbitrair gekozen zijn. (§ 1.2.5).

Een nadeel is tevens dat nog steeds geen homogene reeks met waarnemingen verkregen is.

• Conclusies van Hoofdstuk 2

Een schatter is een functie van willekeurige trekkingen uit een stochastische variabele en zal daarom altijd stochastisch zijn. Het functievoorschrift van de kansdichtheid van de schatter volgt uit het aantal en de verdeling van de waarnemingen en de schattingsmethode. Inzicht in de kansdichtheidsfunctie van de schatter is op analytische en numerieke wijze te verkrijgen (§ 2.1.2).

De momentenschatter van het gemiddelde is een zuivere schatter. De standaardafwijking van de schatter van het gemiddelde is omgekeerd evenredig met de wortel uit het aantal waarnemingen. $\sigma_{\bar{x}} = \sigma_x \sqrt{n}$. De momentenschatter van de variantie is ook een zuivere schatter. De standaardafwijking van de schatter van de variantie wordt benaderd door de standaardafwijking van de schatter van de variantie van een normale verdeling. (§ 2.3.3).

De verwachtingswaarde van een Maximum-Likelihood-schatter is asymptotisch gelijk aan de werkelijke waarde van de parameter. De variantie van een Maximum-Likelihood-schatter kan met een benaderingsformule vastgesteld worden. (§ 2.4.3)

Bewezen kan worden dat de kleinste-kwadraten-schatter van parameter B van de exponentiële en van de Gumbel-verdeling altijd groter is dan de momentenschatter. (Bijlage 7).

• Conclusies van Hoofdstuk 3

De verschillende schattingsmethoden geven overeenkomstige schattingen van de parameters van een Gumbel verdeling. De schattingen van het hoogste en het laagste 10^{-4} -kwantiel liggen toch 0.13 meter uit elkaar. (kwantielschatting van MdM en ML scheelt 0.02 meter). (§ 3.2.2)

Bij de Gumbel-verdeling is de invloed van de schattingsmethode ongeveer zo groot als de invloed van de statistische onzekerheid. (§ 3.2.2)

Bij de lognormale verdeling schelen MdM en ML 0.11 meter in de kwantielschattingen.

De hoogste en de laagste waarde van het 10^{-4} -kwantiel verschillen bij een exponentiële verdeling ongeveer 0.22 meter. (§ 3.3.2) Bij de exponentiële verdeling is de invloed van de keuze van de schattingsmethode groter (in dit geval 10x zo groot) dan de invloed van de statistische onzekerheid. (§ 3.3.2) Hieruit blijkt dat een selectie met de POT-methode geen homogene reeks oplevert.

De ML-schatter is erg gevoelig voor de drempelwaarde van de waterstand bij de POT-methode. Daardoor ligt de overschrijdingslijn die bepaald is met de Maximum Likelihood hoger dan de lijn bepaald met de Kleinste-kwadratenschatter. De laatstgenoemde lijn zou naar verwachting hoger liggen. (§ 3.4.2)

De invloed van de statistische onzekerheid is groter is bij de Gumbel-verdeelde jaarmaxima, omdat bij die selectie van de waterstanden slechts 108 waarnemingen geselecteerd worden. De reeks is echter homogener dan de reeks waterstanden geselecteerd met de POT-methode. Omdat bij die selectie 530 waarnemingen overblijven, is de onzekerheid van de parameters kleiner, maar de onzekerheid

5.2 Aanbevelingen

5.2.1 Inleiding

Bij dit afstudeeronderzoek is niet alleen inzicht gekregen in de invloed van statistische onzekerheid op de bepaling van dijkhoogtes, maar ook in de problemen die met dat onderzoek samenhangen. Om een vervolg te vereenvoudigen worden enkele aanbevelingen gedaan.

5.2.2 Aanbevelingen

Om de invloed van de modelonzekerheid met recht te kunnen verwaarlozen, zoals in dit verslag heeft plaatsgevonden, dienen de reeksen met waarnemingen homogeen te zijn. Daartoe dienen de selectieprocedures veranderd of verbeterd. Het is niet zinvol om de drempels die in de POT-methode toegepast worden zonder meer te verhogen of te verlagen.

Het is mogelijk om het onderzoek uit te breiden door het aantal toegepaste schattingsmethoden te vergroten. Dit geeft niet noodzakelijkerwijs meer inzicht in de invloed van statistische onzekerheid. Het is wel mogelijk dat daardoor zuiverder en efficiëntere schatters toegepast zullen worden.

Voor de compleetheid van het onderzoek is het raadzaam om de variantie van de momentenschatter van de variantie voor andere verdelingen dan een normale verdeling bepalen.

Het is mogelijk om voor de volledigheid van dit onderzoek de overschrijdingslijn van de waterstanden op Bayesiaanse wijze te bepalen met zowel parameter A als B onzeker. De dubbele integraal die dan opgelost dient te worden ziet er als volgt uit:

$$\int \int F_x(x|A,B) f_A(a) f_B(b) da db$$

Het is raadzaam om het econometrisch model te verbeteren. Hierbij denk ik aan de volgende verbeteringen:

- Meer typen belasting op de dijk, bijvoorbeeld ook overslag aan het model toevoegen
- Niet alleen de belasting op, maar ook de sterkte van de dijk op probabilistische wijze in het model opnemen.
- Onderzoek naar de verdelingsfunctie van de totale schade.
Het is bijvoorbeeld denkbaar dat de te beschermen waarde beter op een andere wijze gemodelleerd kan worden.
- De werkelijke, huidige kosten van de investeringen en de schade toepassen.

In dit onderzoek is inzicht verkregen in de invloed van statistische onzekerheid op de econometrische dijkhoogtebepaling met behulp van een Nivo-III-som. Om meer inzicht te krijgen in de invloed van de verschillende stochasten dient de aanbeveling om ook de Nivo-II-som te maken.

Literatuurlijst

- [1] De basispeilen langs de nederlandse kust (1993),
 - [a] Eindverslag
 - [b] Fysisch Onderzoek
 - [c] Statistisch Onderzoek, Tekst & Bijlagen.Rijksinstituut voor Kust en Zee, 's Gravenhage.

- [2] Rapport Deltacommissie (1960),
 - [a] Deel 1: Eindverslag en Interimadviezen.
 - [b] Deel 3: Beschouwingen over stormvloed en getijbeweging.Deltacommissie.
Staatsdrukkerij- en uitgeverijbedrijf, 's Gravenhage.

- [3] Economic decision problems for flood prevention (1956),
D. van Dantzig.
Econometrica nr. 24, blz. 276 t/m 287, New Haven.

- [4] Onderzoek naar methoden ter bepaling van de parameters van een verdelingsfunctie (1992),
R. Beem.
Bouwdienst Rijkswaterstaat.

- [5] Probabilistisch Ontwerpen, b3 (1984),
A.C.W.M. Vrouwenvelder, J.K. Vrijling.
Faculteit der Civiele Techniek, TU Delft.

- [6] Probabilistisch ontwerpen in de waterbouwkunde, f30 (1994),
J.K. Vrijling.
Faculteit der Civiele Techniek, TU Delft.

- [7] Probability, Statistics and Decision for Civil Engineers (1970),
J.R. Benjamin, C.A. Cornell.
McGraw-Hill Publishing Company, U.S.A.

- [8] Statistiek en operationele analyse, a106 (1992),
P. Groenenboom, et al.
Faculteit der Technische Wiskunde en informatica, TU Delft.

- [9] Structural reliability theory and its applications (1982),
P. Thoft-Christensen, M.J. Baker.
Springer-Verlag, Berlin.

Bijlage 1

De jaarmaxima in Hoek van Holland in de periode van 1887-1994. De eenheid is meters boven NAP. (m +NAP).

3.95	2.49	2.14
3.54	2.49	2.13
3.2	2.48	2.13
3.19	2.46	2.12
3.13	2.45	2.12
3.13	2.45	2.09
3.1	2.44	2.08
3.03	2.43	2.02
3.02	2.43	2.02
2.94	2.43	1.94
2.93	2.4	1.86
2.88	2.39	
2.88	2.39	
2.87	2.39	
2.87	2.39	
2.85	2.37	
2.84	2.37	
2.83	2.37	
2.82	2.35	
2.81	2.35	
2.81	2.33	
2.8	2.32	
2.8	2.32	
2.8	2.32	
2.79	2.31	
2.74	2.31	
2.73	2.3	
2.71	2.3	
2.71	2.29	
2.7	2.28	
2.69	2.28	
2.68	2.27	
2.66	2.27	
2.65	2.25	
2.62	2.24	
2.61	2.24	
2.6	2.24	
2.58	2.23	
2.58	2.23	
2.57	2.23	
2.54	2.22	
2.54	2.22	
2.53	2.22	
2.52	2.22	
2.51	2.2	
2.51	2.18	
2.51	2.17	
2.5	2.16	
2.49		

Bijlage 2

De waterstanden in Hoek van Holland die geselecteerd zijn met de Peak-over-Threshold- methode. De waarnemingen zijn gedaan in de periode van 1887-1994. Alle waterstanden zijn groter dan NAP + 1.65 meter, met een opzet die groter is dan 0.3 meter. Deze waterstanden zijn geselecteerd na een DS-4-selectie. (Bron: RIKZ)

3,95	2,5	2,33	2,24	2,16	2,09	2,03	1,98
3,54	2,5	2,33	2,24	2,16	2,09	2,03	1,98
3,2	2,49	2,33	2,24	2,16	2,09	2,03	1,98
3,19	2,49	2,32	2,24	2,16	2,08	2,03	1,98
3,13	2,48	2,32	2,24	2,16	2,08	2,03	1,98
3,13	2,48	2,32	2,24	2,15	2,08	2,03	1,98
3,1	2,48	2,31	2,24	2,15	2,08	2,03	1,98
3,03	2,47	2,31	2,24	2,15	2,08	2,03	1,98
3,02	2,46	2,31	2,23	2,15	2,07	2,03	1,98
2,94	2,45	2,31	2,23	2,15	2,07	2,03	1,98
2,93	2,45	2,31	2,23	2,15	2,07	2,03	1,98
2,88	2,45	2,3	2,23	2,14	2,07	2,02	1,97
2,88	2,44	2,3	2,23	2,14	2,07	2,02	1,97
2,88	2,44	2,3	2,22	2,14	2,07	2,02	1,97
2,87	2,44	2,3	2,22	2,13	2,07	2,02	1,97
2,87	2,43	2,3	2,22	2,13	2,07	2,02	1,97
2,84	2,43	2,29	2,22	2,13	2,07	2,02	1,97
2,83	2,43	2,29	2,22	2,13	2,07	2,02	1,97
2,82	2,43	2,29	2,22	2,13	2,07	2,02	1,97
2,81	2,43	2,28	2,21	2,13	2,07	2,02	1,96
2,8	2,42	2,28	2,21	2,13	2,06	2,02	1,96
2,8	2,41	2,28	2,21	2,13	2,06	2,01	1,96
2,8	2,41	2,28	2,21	2,13	2,06	2,01	1,96
2,79	2,41	2,28	2,21	2,12	2,06	2,01	1,96
2,74	2,4	2,28	2,21	2,12	2,06	2,01	1,96
2,73	2,4	2,28	2,21	2,12	2,06	2,01	1,96
2,71	2,4	2,28	2,21	2,12	2,06	2,01	1,96
2,7	2,39	2,28	2,21	2,12	2,06	2,01	1,96
2,69	2,39	2,28	2,2	2,12	2,05	2,01	1,96
2,68	2,39	2,28	2,2	2,12	2,05	2,01	1,96
2,66	2,39	2,27	2,2	2,12	2,05	2,01	1,95
2,65	2,39	2,27	2,2	2,12	2,05	2,01	1,95
2,65	2,39	2,27	2,2	2,12	2,05	2	1,95
2,65	2,39	2,27	2,2	2,12	2,05	2	1,95
2,64	2,38	2,27	2,19	2,11	2,05	2	1,95
2,61	2,38	2,27	2,19	2,11	2,05	2	1,95
2,61	2,37	2,27	2,19	2,1	2,05	2	1,95
2,6	2,37	2,27	2,19	2,1	2,05	2	1,95
2,58	2,37	2,26	2,18	2,1	2,04	1,99	1,95
2,58	2,37	2,26	2,18	2,1	2,04	1,99	1,94
2,57	2,37	2,26	2,18	2,1	2,04	1,99	1,94
2,54	2,36	2,26	2,18	2,1	2,04	1,99	1,94
2,53	2,36	2,26	2,18	2,09	2,04	1,99	1,94
2,52	2,35	2,26	2,18	2,09	2,04	1,99	1,94
2,52	2,35	2,26	2,18	2,09	2,04	1,99	1,94
2,51	2,35	2,25	2,17	2,09	2,04	1,99	1,94
2,51	2,35	2,25	2,17	2,09	2,04	1,99	1,94
2,51	2,35	2,25	2,17	2,09	2,04	1,99	1,94
2,51	2,34	2,25	2,17	2,09	2,04	1,98	1,94
2,51	2,34	2,25	2,17	2,09	2,04	1,98	1,94
2,51	2,34	2,25	2,16	2,09	2,04	1,98	1,93

[illegible]

Bijlage 3

In deze bijlage worden ter verduidelijking van § 2.3.2 de parameters van een exponentiële verdeling met één en met twee parameters bepaald met behulp van de methode der momenten.

De exponentiële verdeling met één parameter

De kansdichtheidsfunctie van de exponentiële verdeling wordt gegeven door

$$f_X(x) = \lambda e^{-\lambda x}$$

Het gemiddelde van de verdeling wordt bepaald met

$$E[X] = \int_0^{\infty} x f_X(x) dx$$

$$\Rightarrow E[X] = \int_0^{\infty} x \lambda e^{-\lambda x} dx$$

$$= \left[-x e^{-\lambda x} - \frac{1}{\lambda} e^{-\lambda x} \right]_0^{\infty} = 0 - \left(0 - \frac{1}{\lambda} \right) = \frac{1}{\lambda}$$

Zo is dus eenvoudig aan te tonen dat voor deze verdeling moet gelden:

$$\Rightarrow \mu_x = \frac{1}{\lambda}$$

Aangezien ook van een dataset eenvoudig een gemiddelde bepaald kan worden, is de schatter van λ :

$$\hat{\lambda} = \frac{N}{\sum x_i}$$

De variantie van deze exponentiële verdeling is eenvoudig te bepalen. De variantie van een verdeling wordt gedefinieerd door

$$\text{Var}[X] = \int_0^{\infty} (x - \mu_x)^2 f_X(x) dx$$

Voor de exponentiële verdeling met één parameter geeft dit

$$\Rightarrow \text{Var}[X] = \int_0^{\infty} \left(x - \frac{1}{\lambda}\right)^2 \lambda e^{-\lambda x} dx$$

$$= \int_0^{\infty} \left(x^2 \lambda - 2x + \frac{1}{\lambda}\right) e^{-\lambda x} dx = \left[-x^2 e^{-\lambda x} - \frac{1}{\lambda^2} e^{-\lambda x} \right]_0^{\infty}$$

Door de grenzen hierin in te vullen leidt dat tot

$$\Rightarrow \sigma_x^2(x) = \frac{1}{\lambda^2} \Rightarrow \sigma_x = \frac{1}{\lambda}$$

De exponentiële verdeling met twee parameters

Voor de exponentiële verdeling met de kansdichtheidsfunctie

$$f(x) = \frac{1}{B} e^{-\frac{(x-A)}{B}}$$

geldt in principe hetzelfde. Alleen wordt er dit maal niet geïntegreerd vanaf 0 tot oneindig, maar van A tot oneindig. Als x immers kleiner wordt dan A, dan wordt de bovenstaande vergelijking groter dan 1 en is daarmee geen kansdichtheidsfunctie meer.

Voor het gemiddelde geldt:

$$E[X] = \int_A^{\infty} \frac{x}{B} e^{-\frac{(x-A)}{B}} dx = \left[-x e^{-\frac{(x-A)}{B}} - e^{-\frac{(x-A)}{B}} \right]_A^{\infty} = 0 - (-A - B) = A + B$$

$$\Rightarrow \mu_x = A + B$$

En voor de variantie:

$$\text{Var}[X] = \int_A^{\infty} (x - \mu_x)^2 e^{-\frac{(x-A)}{B}} dx = \int_A^{\infty} (x - (A+B))^2 \frac{1}{B} e^{-\frac{(x-A)}{B}} dx$$

$$= \int_A^{\infty} \left(\frac{x^2}{B} - \frac{2Ax}{B} + 2x + \frac{A^2}{B} + 2A + B \right) e^{-\frac{(x-A)}{B}} dx$$

$$= \left[\left(-x^2 + 2Ax - A^2 - B^2 \right) e^{-\frac{(x-A)}{B}} \right]_A^{\infty} = 0 - (-A^2 + 2A^2 - A^2 - B^2) = B^2$$

$$\text{Var}[X] = B^2 \quad \Rightarrow \quad \sigma_x = B$$

Bijlage 4

In deze bijlage worden ter verduidelijking van paragraaf 2.4.2 de parameters van de exponentiële en de lognormale verdeling bepaald met behulp van de Maximum Likelihood methode.

De exponentiële verdeling

De kansdichtheidsfunctie van een exponentiële verdeling met twee parameters A en B is:

$$f(x_i|A, B) = \frac{1}{B} \cdot e^{-\frac{(x_i-A)}{B}}$$

Voor iedere x_i moet gelden: $x_i \geq A$. Dan en slechts dan is $0 < f(x) \leq 1$.

De likelihood-functie is het product van de kansdichtheden van alle x_i

$$L = \prod f(x_i|A, B) = \frac{1}{B} \cdot e^{-\frac{(x_1-A)}{B}} \cdot \frac{1}{B} \cdot e^{-\frac{(x_2-A)}{B}} \cdot \dots \cdot \frac{1}{B} \cdot e^{-\frac{(x_N-A)}{B}} = \frac{1}{B^N} \cdot e^{\frac{(-\sum x_i + NA)}{B}}$$

De beste schatter van zowel A als B wordt gevonden door de likelihoodfunctie te maximaliseren.

Voor parameter A geldt dat de likelihood-functie toeneemt als A toeneemt. Om de functie te maximaliseren moet dus A maximaal zijn. De beste schatter voor A zal de grootst mogelijke waarde van A zijn. Aangezien $A \leq \min(x_{(1)} \dots x_{(N)})$ geldt:

$$\hat{A} = x_{(1)}$$

Voor de schatter van B wordt het de likelihood-functie gemaximaliseerd, door de afgeleide van log-likelihood-functie gelijk aan 0 te stellen.

$$\ln(L) = -N \ln(B) - \frac{\sum x_i}{B} + \frac{NA}{B}$$

$$\frac{d \ln(L)}{dB} = \frac{-N}{B} + \frac{\sum x_i}{B^2} - \frac{NA}{B^2} = 0$$

$$\Rightarrow \frac{-NB - NA + \sum x_i}{B^2} = 0 \Rightarrow -NB - NA + \sum x_i = 0$$

$$\Rightarrow A + B = \frac{\sum x_i}{N} = \bar{x}$$

Met $\hat{A} = x_{(1)}$ volgt: $\hat{B} = \bar{x} - x_{(1)}$.

De lognormale verdeling

Voor de lognormale verdeling geldt de volgende kansdichtheidsfunctie:

$$f(x) = \frac{1}{\sqrt{2\pi Bx}} \exp^{-\frac{(\ln(x)-A)^2}{2B^2}}$$

De likelihood functie is het product van de kansdichtheidsfunctie voor alle x_i . Dit geeft de volgende onaantrekkelijke vergelijking

$$L = \prod_{i=1}^N (f(x_i)) = \frac{1}{(\sqrt{2\pi B})^N} \cdot \frac{1}{\prod (x_i)} \cdot \exp^{\frac{\sum [-(\ln x_i)^2 + 2\ln x_i \cdot A]}{2B^2}} \cdot \exp^{\frac{-NA^2}{2B^2}}$$

Ook hier wordt het maximum van de functie bepaald door het maximum van het logaritme van de functie vast te stellen.

Het logaritme van de likelihood functie is

$$\ln(L) = -N \ln(\sqrt{2\pi B}) - \sum \ln x_i + \frac{\sum [-(\ln x_i)^2 + 2\ln x_i \cdot A]}{2B^2} - \frac{NA^2}{2B^2}$$

Om het maximum te bepalen wordt dit gedifferentieerd naar respectievelijk de parameters A en B en vervolgens gelijk aan 0 gesteld.

Voor parameter A volgt dan

$$\frac{d \ln(L)}{dA} = \frac{2 \sum \ln x_i}{2B^2} - \frac{2NA}{2B^2} = 0$$

Hiermee kan de beste schatter van parameter A gevonden worden:

$$\Rightarrow \hat{A} = \frac{1}{N} \cdot \sum \ln x_i$$

Voor parameter B wordt de zelfde werkwijze gevolgd

$$\frac{d \ln(L)}{dB} = -\frac{N}{B} - \frac{\sum [-(\ln x_i)^2 + 2\ln x_i \cdot A]}{B^3} + \frac{NA^2}{B^3} = 0$$

Aangezien B niet gelijk aan 0 is, kan dit herschreven worden als:

$$NB^2 = \sum (\ln x_i)^2 - 2A \sum \ln x_i + NA^2$$

Als hierin de uitdrukking voor A wordt gesubstitueerd volgt:

$$NB^2 = \sum (\ln x_i)^2 - \frac{2(\sum \ln x_i)^2}{N} + \frac{N(\sum \ln x_i)^2}{N^2} = \sum (\ln x_i)^2 - \frac{(\sum \ln x_i)^2}{N}$$

Waaruit de schatter voor B gevonden kan worden:

$$\Rightarrow \hat{B} = \sqrt{\frac{1}{N} \sum (\ln x_i)^2 - \left(\frac{1}{N} \sum \ln x_i\right)^2}$$

Bijlage 5

In deze bijlage wordt het in Pascal geschreven programma gegeven, waarmee de Maximum-Likelihoodschatters vande Gumbelverdeling van de jaarmaxima bepaald zijn.

De opzet van het programma, waarmee de Maximum-Likelihoodschatter van de Gumbelverdeling bepaald kan worden, is als volgt.

De likelihoodfunctie is een functie van parameter A en B. Voor een nog onbekende A en B is de likelihoodfunctie maximaal. De zoektocht begint bij de waarden van A en B die met de methode der momenten zijn bepaald. Om die waarden wordt een raster met 9 punten opgezet. $(A-\delta, A, A+\delta)$, $(B-\delta, B, B+\delta)$.

Vervolgens wordt in ieder punt de Maximum Likelihood bepaald. Het nieuwe middelpunt van het raster wordt dat punt waar de likelihood maximaal is. Is de likelihood maximaal in het middelste punt dan wordt de stapgrootte δ van het raster verkleind. Deze handelingen van verschuiven en verkleinen van het raster worden herhaald totdat de stapgrootte δ een zekere gewenste kleine waarde heeft.

PROGRAM MLGUMBEL;

uses dos, crt;

type
plotpos = record
 no1 : real;
 no2 : real;
end;

var
A,B : real; {parameters van de verdeling}
i,j,c,tel : integer; {tellers}
outfile : text;
dtafile : text;
positie : array[1..9] of plotpos; {de positie in het raster}
delta : real; {stapgrootte van het raster}
L : array[1..9] of real; {De likelihoodfunctie van positie i}
jaarmax : array[1..108] of real; {de jaarmaxima}
ML : real; {De maximum Likelihood}
MLpos : integer; {De positie met de Maximum Likelihood}

begin
clrscr;
assign(dtafile,'max.dat');
reset(dtafile);
assign(outfile,'res.dat');
rewrite(outfile);

A:=2.4; {startwaarden van de variabelen}
B:=0.3;
delta := 0.1;

```

repeat                                {herhaal dit programma tot delta een zekere kleine waarde heeft}
begin

for i:=1 to 9 do                      {positie i krijgt een waarde A en een waarde B}
begin
    positie[i].no1 :=A-delta;
    if (i+1) MOD 3 < > 0 then c:=0 else positie[i].no1 :=A;
    if i MOD 3 < > 0 then c:=0 else positie[i].no1 :=A+delta;
    positie[i].no2 := B-delta;
    if i > 3 then positie[i].no2 := B;
    if i > 6 then positie[i].no2 := B+delta;
end;

for i:=1 to 9 do                      {Voor iedere positie i wordt de Likelihoodfunctie bepaald}
begin
    A:=positie[i].no1;
    B:=positie[i].no2;
    L[i]:=1;
    reset(dtatfile);
    for j:=1 to 108 do
    begin
        readln(dtatfile,jaarmax[j]);
        L[i]:=L[i]*(1/B)*exp(-(jaarmax[j]-A)/B)-exp(-(jaarmax[j]-A)/B)); {bepalen van de likelihood}
    end;
end;

ML:=0;
for i:=1 to 9 do {De positie i met de maximale likelihood wordt gevonden}
begin
    if L[i] > ML then
    begin
        ML:=L[i];
        MLpos:=i;
    end;
end;

if MLpos < > 5 then c:=0 else delta:=delta/2;          {raster wordt verkleind of verschoven}
A:=positie[MLpos].no1;
B:=positie[MLpos].no2;

end;
until delta < 0.001;                    {eindig de lus als delta kleiner dan bv. 0.001}

writeln(outfile,'De Likelihood is maximaal voor');
writeln(outfile,'A = ',positie[MLpos].no1,', ' ; B = ',positie[MLpos].no2);

close(dtatfile);
close(outfile);
end.

```

Bijlage 6

In deze bijlage wordt de standaardafwijking van de schatters van de parameters van de lognormale en de exponentiële verdeling bepaald met behulp van een benaderingsformule. De schatters zijn bepaald met de Maximum Likelihood methode.

Inleiding

Als een schatter $\hat{\theta}$ bepaald is met behulp van de methode van de Maximum Likelihood, is de variantie van de schatter asymptotisch gelijk aan [7, blz. 400]:

$$\text{Var}[\hat{\theta}] = - \frac{1}{N E\left[\frac{\partial^2 \ln f(x|\theta)}{\partial \theta^2} \right]}$$

De exponentiële verdeling

Voor de schatters van de parameters van een exponentiële verdeling met twee parameters wordt de variantie als volgt gevonden:

$$f(x|A, B) = \frac{1}{B} e^{-\frac{(x-A)}{B}}$$

$$\ln f(x|A, B) = -\ln B - \frac{(x-A)}{B}$$

Als deze vergelijking twee keer naar A gedifferentieerd wordt, dan blijft er niets over. Met deze asymptotische benadering van de variantie van schatters is de variantie van parameter A niet te bepalen.

Om de variantie van de schatter van B te bepalen wordt er twee keer naar B gedifferentieerd.

$$\frac{\partial \ln [f(x|A, B)]}{\partial B} = -\frac{1}{B} + \frac{(x-A)}{B^2}$$

$$\frac{\partial^2 \ln [f(x|A, B)]}{\partial B^2} = -\frac{1}{B^2} + \frac{2A}{B^3} - \frac{2x}{B^3}$$

$$E\left[\frac{\partial^2 \ln [f(x|A, B)]}{\partial B^2} \right] = E\left[\frac{B}{B^3} + \frac{2A}{B^3} - \frac{2x}{B^3} \right]$$

Bij de bepaling van de schatters van de parameters van de verdeling met de methode van de Maximum Likelihood is gebleken dat $E[x] = \bar{x} = A + B$. Hiermee wordt het volgende gevonden:

$$E\left[\frac{B}{B^3} + \frac{2A}{B^3} - \frac{2x}{B^3} \right] = \frac{B}{B^3} + \frac{2A}{B^3} - \frac{2(A+B)}{B^3} = \frac{-1}{B^2}$$

$$\Rightarrow \text{Var}[\hat{B}] \approx \frac{B^2}{N} \Rightarrow \sigma[\hat{B}] \approx \frac{B}{\sqrt{N}}$$

De standaardafwijking van de schatter van parameter B is omgekeerd evenredig met de wortel uit het aantal beschikbare waarnemingen N.

De lognormale verdeling

Voor de lognormale verdeling geldt de volgende kansdichtheidsfunctie:

$$f(x|A, B) = \frac{1}{\sqrt{2\pi} Bx} \exp \left\{ -\frac{(\ln(x) - A)^2}{2B^2} \right\}$$

We bepalen nu achtereenvolgens voor de schatter van parameter A en B de standaardafwijking volgens de eerder gegeven definitie. Daartoe wordt eerst het natuurlijk logaritme van de functie genomen.

$$\ln[f(x|A, B)] = -\ln B - \ln \sqrt{2\pi} x - \frac{(\ln(x) - A)^2}{2B^2}$$

De bovenstaande vergelijking wordt twee maal naar A gedifferentieerd om de variantie van de schatter van A te bepalen.

$$\frac{\partial \ln[f(x|A, B)]}{\partial A} = \frac{2 \ln x}{2B^2} - \frac{2A}{2B^2}$$

$$\frac{\partial^2 \ln[f(x|A, B)]}{\partial A^2} = -\frac{1}{B^2}$$

Voor de variantie van de geschatte parameter A geldt nu

$$\text{Var}[\hat{A}] = -\frac{1}{N \cdot E\left[\frac{\partial^2 \ln f(x|A, B)}{\partial A^2}\right]} = \frac{B^2}{N}$$

$$\Rightarrow \sigma_{\hat{A}} = \frac{B}{\sqrt{N}}$$

De standaardafwijking van de schatter van A is wederom omgekeerd evenredig met de wortel uit het aantal waarnemingen. De variantie hangt echter niet met A samen, maar met de werkelijke waarde van B!

Voor de variantie van de schatter van B geldt in principe hetzelfde. Ook hier wordt het logaritme van de kansdichtheidsfunctie twee keer naar B gedifferentieerd.

$$\frac{\partial \ln[f(x|A, B)]}{\partial B} = \frac{-1}{B} + \frac{(\ln(x) - A)^2}{B^3}$$

$$\frac{\partial^2 \ln[f(x|A, B)]}{\partial B^2} = \frac{1}{B^2} - \frac{3(\ln(x) - A)^2}{B^4}$$

$$E\left[\frac{\partial^2 \ln[f(x|A, B)]}{\partial B^2}\right] = \frac{1}{B^2} - \frac{3 \cdot E[(\ln(x) - A)^2]}{B^4}$$

Met behulp van de Maximum Likelihood methode is bepaald dat $\hat{A} = \frac{1}{N} \sum \ln(x_i)$, zodat de verwachtings waarde in het rechterlid herschreven kan worden als:

$$E[(\ln(x) - A)^2] \Rightarrow E\left[\left(\ln(x) - \frac{1}{N} \sum \ln(x)\right)^2\right]$$

Stellen we nu dat $\ln(x) = y$, met y normaal verdeeld ([5], blz. 2-18), dan kan bovenstaande verwachtingswaarde herschreven worden als

$$E\left(y - \frac{1}{N} \sum y\right)^2 = E[(y - \bar{y})^2] = \text{Var}(y)$$

Volgens de definitie is ([5]) :

Bijlage 7

In deze bijlage wordt bewezen dat de schatter van parameter B van een exponentiële verdeling geschat met behulp van de methode der momenten altijd kleiner is dan dezelfde schatter als die bepaald wordt met behulp van de methode van de kleinste kwadraten.

We beschouwen een gelineariseerde exponentiële verdeling (zie ook paragraaf 2.5.2):

$$1 - F(x) = e^{-\frac{x-A}{B}} \Rightarrow \ln(1 - F(x)) = -\left(\frac{x-A}{B}\right)$$

De plotpositie van de i^e waarneming, $y_i = F(x_i) = \frac{i}{(N+1)}$, gesubstitueerd in de getransformeerde vergelijking geeft:

$$y_i^* = \ln\left(1 - \frac{i}{N+1}\right) = -\left(\frac{x_i - A}{B}\right)$$

Hierin is y^* de getransformeerde y .

Met de methode van de kleinste kwadraten wordt de regressielijn bepaald met de vergelijking $y = \alpha + \beta x$. De vergelijking van β kan gevonden worden door de correlatie-coëfficiënt te vermenigvuldigen met de standaardafwijking van y gedeeld door de standaardafwijking van x . ([7], blz.422)

$$\beta = \rho \frac{\sigma_y}{\sigma_x}$$

ρ , de correlatie-coëfficiënt wordt gedefinieerd door: $\rho = \frac{\text{cov}(xy)}{\sigma_x \sigma_y}$

Parameter B uit de exponentiële verdeling kan na transformatie met de methode der kleinste kwadraten geschat worden:

$$-\frac{1}{\hat{B}_k} = \beta = \rho \frac{\sigma_y}{\sigma_x} \Rightarrow \hat{B}_k = -\frac{\sigma_x}{\rho \cdot \sigma_{y^*}}$$

Het subscript k verwijst naar de methode van de Kleinste Kwadraten.

In paragraaf 2.3.2 is beschreven dat parameter B in het geval van de methode der momenten geschat kon worden door de standaardafwijking van de waarnemingen x .

$$\hat{B}_m = \sigma_x.$$

Het subscript m verwijst in dit geval naar de methode der Momenten.

In de praktijk bleek dat altijd $\hat{B}_m < \hat{B}_k$ en dus dat $\sigma_x < -\frac{\sigma_x}{\rho \cdot \sigma_{y^*}}$.

Deze ongelijkheid geldt als te bewijzen is dat $-1 < \rho \cdot \sigma_{y^*} < 0$.

De correlatie-coëfficiënt $\rho \leq 0$, want y^* is een dalende functie van x . Als een rangorde i groter wordt, dan wordt de waarde van x hoger. Als i groter wordt dan wordt $1 - \frac{i}{(N+1)}$ steeds kleiner. Zoals in de

grafiek op de nevenstaande bladzijde te zien is, wordt de waarde van $y_i^* = \ln\left(1 - \frac{i}{(N+1)}\right)$ steeds kleiner.

De ongelijkheid is dus bewezen als $\sigma_{y^*} < 1$.

De standaardafwijking van y is af te leiden uit de uitdrukking van y^* : $y_i^* = \ln\left(1 - \frac{i}{N+1}\right)$

Om schrijfwerk te beperken stellen we nu dat $1 - \frac{i}{N+1} = z$, waarin z uniform verdeeld is tussen 0 en 1.

$y = \ln(z)$. NB! y is dus een functie van een stochast met een uniforme verdeling.

De verwachtingswaarde van y^*

$$E[y^*] = \int_{\delta}^{1-\delta} \ln(z) \cdot f(z) dz = -1 \quad \text{voor } \delta \rightarrow 0 \quad \text{Met } \delta \text{ is } 1/(N+1)$$

De variantie van y^* is

$$\begin{aligned} \text{Var}[y^*] &= \int_{\delta}^{1-\delta} \ln^2(z) f(z) dz - (-1)^2 = 1 \quad \text{voor } \delta \rightarrow 0 \\ &= \sigma_y = 1 \quad \text{voor } \delta \rightarrow 0 \end{aligned}$$

Voor datasets van eindige grootte zal δ nooit gelijk aan 0 worden. Voor iedere δ groter dan 0 geldt dat $\sigma_{y^*} < 1$

Omdat $\rho \in (-1, 0)$ en $\sigma_{y^*} < 1$ is het bewijs geleverd dat

$$\sigma_x < -\frac{\sigma_x}{\rho \cdot \sigma_{y^*}} \quad \text{oftewel dat } \hat{B}_m < \hat{B}_k.$$

Als $\sigma_x = \frac{1}{N-1} \sum (x_i - \bar{x})^2$, met andere woorden als σ_x een zuivere schatter is, dan is met dit bewijs

tegelijkertijd bewezen dat de kleinste-kwadratenschatter van parameter B met lineaire regressie geen zuivere schatter oplevert.

Bijlage 8

Deze tabel geeft de uitkomsten weer van het fitten van de datasets uit [4] met het programma 'Superfit'. Ter vergelijking zijn ook de uitkomsten voor de Maximum Likelihood methode weergegeven. Een mooie toevoeging voor het 'groene boekje'.

	Maximum Likelihood		Numeriek fitten	
	A	B	A	B
gr1	1,973	0,456	1,898	0,516
gr2	1,966	0,329	1,941	0,365
gr3	1,966	0,295	1,949	0,378
grA	1,962	0,288	1,897	0,327
grB	1,976	0,395	1,856	0,569
grC	1,977	0,414	1,866	0,575
grD	1,988	0,243	1,95	0,313
grE	2,024	0,313	1,98	0,415
klA	1,995	0,363	1,924	0,374
klB	2,079	0,237	2,079	0,268
klC	1,962	0,246	1,924	0,331
klD	2,041	0,403	1,99	0,515
klE	2,01	0,156	1,985	0,192
klF	2,014	0,275	1,981	0,156
	-----	-----	-----	-----
gem.	1,997	0,315	1,944	0,378
st.dev.	0,037	0,087	0,056	0,125



Bijlage 9

In deze bijlage wordt de onzekerheid van de schatter van de parameter A van de exponentiële verdeling op Bayesiaanse wijze uit de verdelingsfunctie geïntegreerd. Zie ook § 3.3.2.

De exponentiële verdeling met een onzekere parameter A wordt gedefinieerd door de functie:

$$|e = \exp\left(-\frac{x-A-e}{B}\right)$$

De spreiding e is normaal verdeeld met $\mu_e=0$. De gezamenlijke kansdichtheidsfunctie is dan:

$$F(x, e) = \exp\left(-\frac{x-A-e}{B}\right) \cdot \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{e}{\sigma}\right)^2\right)$$

Dit kan herschreven worden als:

$$F(x, e) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left(-\frac{x-A}{B}\right) \cdot \exp\left(-\frac{1}{2}\left(\left(\frac{e}{\sigma}\right)^2 - \frac{2e}{\sigma}\right)\right) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left(-\frac{x-A}{B}\right) \cdot \exp\left(-\frac{1}{2\sigma^2}\left(e - \frac{\sigma^2}{B}\right)^2\right) \cdot \exp\left(\frac{\sigma^2}{2B}\right)$$

Om de onzekerheid e uit deze verdeling te halen, wordt de functie vermenigvuldigd met de kansdichtheidsfunctie van e en over e geïntegreerd.

$$F(x) = \int F(x, e) \cdot f(e) de$$

$$\Rightarrow F(x) = \exp\left(-\frac{x-A}{B}\right) \cdot \exp\left(\frac{\sigma^2}{2B}\right) \cdot \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\left(\frac{e - \frac{\sigma^2}{B}}{\sigma}\right)^2\right) de$$

Het tweede gedeelte van bovenstaande vergelijking is de definitie van een normale verdeling met oppervlakte 1.

Afgeleid in bijlage 4 is de standaardafwijking van de schatter: $\sigma_e = \frac{B}{\sqrt{N}}$

Als dit gesubstitueerd wordt in de verdelingsfunctie dan volgt :

$$F(x) = \exp\left(-\frac{x-A}{B}\right) \cdot \exp\left(\frac{B}{2N}\right)$$

De onzekerheid van A is terug te leiden tot de invloed van het aantal waarnemingen N.

Bijlage 10

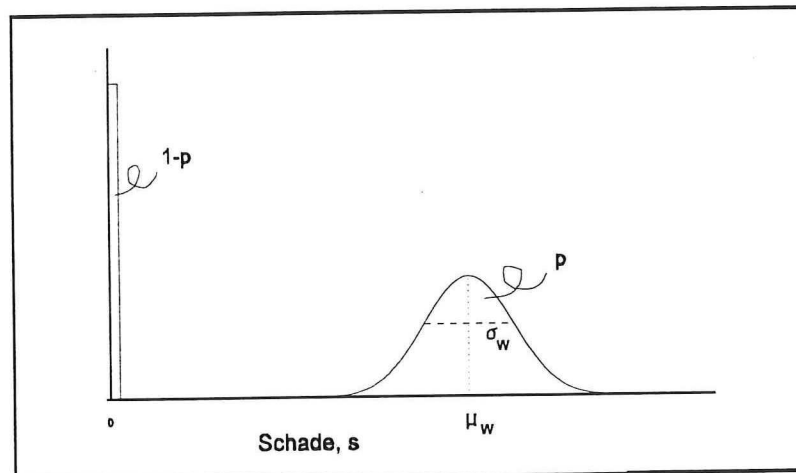
In deze bijlage wordt eerst uitgelegd hoe de verwachtingswaarde van de schade te bepalen is. Vervolgens wordt bekeken hoe de stochast schade contant gemaakt zou kunnen worden.

De schadeverwachting

De kansdichtheidsfunctie van de schade, $f_s(s)$, is geen continue functie. Ze bestaat uit twee delen, het ene deel met een kans p en het andere met een kans $1-p$.

Er is een kans p dat een gebied gaat overstromen. Een uitgangspunt bij de bepaling van de optimale dijkhoogte is dat in het geval van een overstroming alles verloren gaat. Er is dus een kans p , dat een een ramp plaatsvindt, waarbij W verloren gaat. Deze schade is echter geen constante. We gaan er van uit dat de schade in geval van een ramp normaal verdeeld is.

De normale verdeling van de schade bij een ramp, $f_w(s)$ met gemiddelde μ_w en standaardafwijking σ_w . Voor de duidelijkheid: de oppervlakte van de verdeling is p , niet de hoogte. In de berekening van de schadeverwachting wordt de kansdichtheidsfunctie daarom vermenigvuldigd met p .



De verdeling van de schade

Er is een kans $1-p$ dat er niets gebeurt. Een kans van $1-p$ op 0 schade. In de afbeelding lijkt de lijn bij 0 schade, $s=0$, nog enige dikte te hebben. Dat heeft ze dus niet! Deze toestand kan weergegeven worden door de zogenaamde Dirac-delta-functie. Deze 'spijker' met een grootte $1-p$ bij $s=0$ kan als volgt weergegeven worden:

$$\delta(s) = 0, \text{ als } s \neq 0 ; \int_{-\infty}^{\infty} (1-p) \delta(s) ds = 1-p$$

Omdat in het betreffende geval de oppervlakte niet gelijk is aan 1, maar aan $1-p$, is de Dirac-functie met $1-p$ vermenigvuldigd.

Samenvattend kan gesteld worden, dat de totale schadefunctie $f_s(s)$ bestaat uit een Dirac-functie en een normale verdeling, $f_w(s)$.

Voor de verwachtingswaarde van de totale schade, S , geldt het volgende:

$$\begin{aligned} E(S) &= \int_{-\infty}^{\infty} s f_s(s) ds = \int_{-\infty}^{\infty} s (1-p) \delta(s) ds + \int_{-\infty}^{\infty} s p f_w(s) ds \\ &= 0 + p \int_{-\infty}^{\infty} s f_w(s) ds = p \mu_w \end{aligned}$$

Het tweede moment wordt gedefinieerd door:

$$E(S^2) = \int_0^{\infty} f_S(s) s^2 ds = \int_0^{\infty} (1-p) \delta(s) s^2 ds + \int_0^{\infty} p f_W(s) s^2 ds$$

$$= 0 + p \cdot \int_0^{\infty} f_W(s) s^2 ds = 0 + p \cdot E(f_W^2) = p [\text{Var}(f_W) + E^2(f_W)] = p \cdot \sigma_W^2 + p \cdot \mu_W^2$$

Zodat voor de variantie van de totale schade geldt:

$$\text{Var}(S) = E(S^2) - E(S)^2$$

$$= p \cdot \sigma_W^2 + p \cdot \mu_W^2 - p^2 \cdot \mu_W^2 = p \cdot \sigma_W^2 + p(1-p) \cdot \mu_W^2 = p(\sigma_W^2 + (1-p) \cdot \mu_W^2)$$

Uiteindelijk is de toe te passen standaardafwijking van de schade te bepalen door de wortel uit bovenstaande variantie te trekken.

De rentevoet

Om het risico contant te maken, dient ze ieder jaar gedeeld te worden door $(1+r)$. r is de reële rentevoet. De reële rentevoet is de rentevoet verminderd met de groeivoet. Verondersteld wordt dat gedurende de levensduur van de kering de reële rentevoet gemiddeld 1.5% groter is dan de groeivoet van de economie.

De contante waarde van de schade betekent dat bepaald wordt wat nú gereserveerd zou moeten worden, voor een schade over een zeker aantal jaren. Als over 100 jaar bijvoorbeeld 20 miljoen betaald zou moeten worden, dan is dat mogelijk door vandaag 4,5 miljoen op de bank te zetten.

In het geval van de zo juist beschreven schadeverwachting, waren we geïnteresseerd in de schade voor een oneindig aantal jaren.

Voor een schade in het eerste jaar moet de schade gedeeld worden door $(1+r)$, als die in het tweede jaar zal vallen door $(1+r) \cdot (1+r)$. En zo voort. Aangezien de rente niet van invloed is op de verwachtingswaarde van de schade kan dit ook geschreven worden als:

$$\frac{1}{(1+r)^1} * E[s] + \frac{1}{(1+r)^2} * E[s] + \dots = \sum_{i=1}^N \frac{1}{(1+r)^i} * E[s]$$

Als deze sommatie over de levensduur N wordt uitgevoerd voor een onbegrensde toekomst, ($N \rightarrow \infty$), dan is het te vereenvoudigen tot $1/r$. Een oneindige levensduur geldt echter niet voor dergelijke constructies. Aan een oneindig lange levensduur kleven nog twee andere bezwaren. Als een schade erg ver in de tijd plaats zou vinden dan wordt de contante waarde van die schade nu bijzonder klein. Een tweede bezwaar is dat de kans op méér dan één ramp tijdens de levensduur groter wordt. Aan de andere kant kunnen dus hele grote totale kosten ontstaan.

Aangezien de vereenvoudiging van de sommatie tot $1/r$ al niet meer opgaat bij een levensduur langer dan 100 jaar, is opnieuw vereenvoudigd. De invloed van een nog te kiezen levensduur N is daarbij nu meegenomen.

$$\sum_{i=1}^N \frac{1}{(1+r)^i}, \quad \text{met } r > 0; \quad \text{stel } x = \frac{1}{(1+r)}$$

$$\Rightarrow \sum_{i=1}^N x^i = \sum_{i=1}^{\infty} x^i - \sum_{i=N+1}^{\infty} x^i$$

$$= \sum_{i=1}^{\infty} x^i - x^N \sum_{i=1}^{\infty} x^i$$

$$= (1 - x^N) \sum_{i=1}^{\infty} x^i$$

De som van x tot de macht i is als volgt te herschrijven:

$$\sum_{i=0}^{\infty} x^i = \frac{1}{1-x} \quad \text{als } |x| < 1$$

$$\Rightarrow \sum_{i=1}^{\infty} x^i = \frac{1}{1-x} - 1 = \frac{x}{1-x}$$

$$\Rightarrow (1-x^N) \sum_{i=1}^{\infty} x^i = (1-x^N) \frac{x}{1-x}$$

Dit is de oorspronkelijk som met daarin x gesubstitueerd. Dit kan nu dus weer naar r teruggeschreven worden.

$$\begin{aligned} (1-x^N) \frac{x}{1-x} &= \left(1 - \frac{1}{(1+r)^N}\right) \cdot \frac{1}{1+r} \cdot \frac{1}{\left(1 - \frac{1}{1+r}\right)} \\ &= \left(1 - \frac{1}{(1+r)^N}\right) \cdot \frac{1}{1+r} \cdot \frac{1}{\frac{r}{1+r}} \\ &= \left(1 - \frac{1}{(1+r)^N}\right) \cdot \frac{1}{r} \end{aligned}$$

De sommatie is uit de vergelijking verdwenen. Voor grote N vereenvoudigd de vergelijking inderdaad tot $\frac{1}{r}$.

De som van de gekwadraterde rentevoet, die toegepast wordt bij de variantie van de schade, is op een zelfde eenvoudige wijze te herschrijven :

$$\begin{aligned} \sum_{i=1}^N \frac{1}{((1+r)^2)^i} , \quad \text{met } r > 0 \text{ en } x &= \frac{1}{((1+r)^2)} \\ &= \left(1 - \frac{1}{(1+r)^{2N}}\right) \cdot \frac{1}{(1+r)^2 - 1} \end{aligned}$$

Hier geldt opnieuw dat de eerste term wegvalt bij grote N .