



BSc verslag TECHNISCHE WISKUNDE

“Limietstellingen voor het aantal lussen in een random graaf”

(Engelse titel: “Limit theorems for the number of loops in a random graph”)

Nathan Meibergen

Technische Universiteit Delft

Begeleider

Dr. G. Hooghiemstra

Overige commissieleden

Dr.ir. L.E. Meester

Dr. J.L.A. Dubbeldam

Dr. J.G. Spandaw

Juni, 2011

Delft

Inhoudsopgave

Inleiding	2
1 Configuratie model	3
2 Kansen op lussen in een knoop	3
2.1 Geen lussen	4
2.2 Één lus	4
2.3 Meerdere lussen	5
3 Limietstellingen voor graad van $O(1)$ en $O(\sqrt{l})$	5
3.1 Bewijs: $P(S_1 = 0) \rightarrow 1$	6
3.2 Simulaties naar de Poisson- en normale verdeling	6
3.3 Convergentie naar de Poisson verdeling	7
4 Limietstelling voor graad van $O(l)$	9
4.1 Martingaal	10
4.2 Convergentie van martingaal-coëfficiënten	11
4.3 Verwachte aantal lussen	13
4.4 Convergentie van de stoptijd	14
4.5 Bewijs voor convergentie naar de normale verdeling	15
Samenvatting	19
Appendices	20
A Simulatie van de Poisson- en normale verdeling	20
B Convergentie van ν_l/l	20
Symbolenlijst	22

Inleiding

Een random graaf is een graaf die willekeurig gevormd wordt. Dit random proces kan op verscheidene manieren. Een voorbeeld is het Erdős-Rényi $G(n, p)$ model. Dit is een graaf met n knopen waarin tussen ieder paar *verschillende* knopen met kans p een kant wordt gevormd. Wanneer $p = 1$ zullen dus alle knopen direct met elkaar verbonden zijn en wanneer $p = 0$ zal de graaf geen kanten bevatten. Merk op dat de resulterende graaf sowieso geen kant bevat die zijn begin en eindpunt in dezelfde knoop heeft, dit noemen we een lus.

Wij zullen gebruik maken van het configuratiemodel. Dit model staat lussen wel toe. Het is zelfs mogelijk dat er meerdere lussen ontstaan in een knoop. Het aantal lussen dat gevormd wordt blijft uiteraard random.

In het configuratiemodel kennen we aan iedere knoop een gegeven aantal verbindingen toe. Vervolgens kiezen we geheel willekeurig twee verbindingen en deze vormen samen een kant. Het aantal verbindingen van een knoop noemen we de graad.

In dit verslag zullen we uitvoerig kijken naar de kansverdeling van *het aantal lussen in een knoop met graad d* , in het bijzonder wanneer de som van graden naar oneindig gaat.

In hoofdstuk 1 zullen we het configuratiemodel uitgebreider bespreken en uitleggen hoe we deze gebruiken.

We laten dan de som van graden, l , eindig. Om een beetje grip te krijgen op het configuratiemodel en de probleemstelling beginnen we in hoofdstuk 2 met het berekenen van enkele kansen voor het aantal lussen in een knoop. We berekenen de kans op geen lussen, daarna op precies 1 lus en vervolgens op $k \in \mathbb{N}$ lussen.

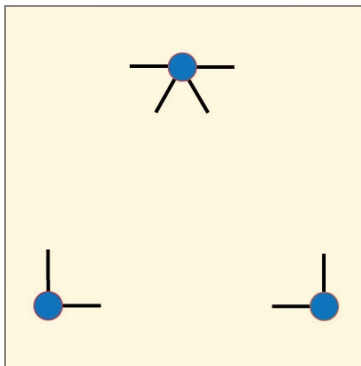
Vervolgens zullen we l naar oneindig laten gaan en bespreken we in hoofdstuk 3 de drie situaties die we zullen onderscheiden, namelijk: $d = O(1)$, $d = O(\sqrt{l})$ en $d = O(l)$. Dan zullen we de verdeling voor het aantal lussen voor $d = O(1)$ direct bewijzen in sectie 3.1. Omdat de 2e en 3e situaties wat ingewikkelder zijn zullen we in sectie 3.2 enkele simulaties doen. Daarna kunnen we in sectie 3.3 direct de verdeling van het aantal lussen behorende bij $d = O(\sqrt{l})$ bepalen en bewijzen.

De derde situatie is echter nog gecompliceerder, deze zullen we daarom apart beschouwen in hoofdstuk 4. Het bewijs voor de bijbehorende verdeling voor het aantal lussen zal ondersteund moeten worden door enkele lemma's. Pas in sectie 4.5 zullen we toekomen aan het uiteindelijke bewijs.

1 Configuratie model

Een random graaf wordt aan de hand van een model gevormd. In dit verslag zullen we gebruik maken van het configuratiemodel. Dit model werkt als volgt:

We beginnen met een stelsel met n knopen en van iedere knoop is de graad d_i , $i = 1, 2, \dots, n$ bekend, dit zijn het aantal verbindingen van de knoop. Een voorbeeld van zo een stelsel is te zien in figuur 1. Dit stelsel heeft drie knopen, de onderste twee hebben graad 2 en de bovenste heeft graad 4.



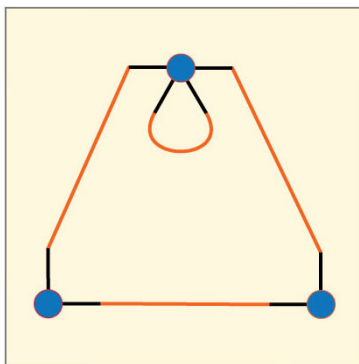
Figuur 1: Een nog onvolledige graaf in het configuratiemodel.

Nu beginnen we met, wat we het proces van trekkingen zullen noemen. Daartoe kiezen we een knoop, zeg knoop 1, met graad d_1 en gaan nu een verbinding van deze knoop willekeurig koppelen aan een *andere* verbinding in het stelsel. Deze vormen samen een kant. Nu kiezen we de volgende vrije verbinding in deze knoop en gaan deze ook koppelen, enz. Tot knoop 1 geen vrije verbinding meer heeft - met vrije verbinding bedoelen we een verbinding die nog niet gekoppeld is aan een andere. Daarna gaan we naar de volgende knoop, waar we hetzelfde proces herhalen. Dit levert uiteindelijk een graaf. Zoals bijvoorbeeld de graaf in figuur 2.

Merk op dat we voor dit proces maar één eis stellen, namelijk dat een verbinding nooit aan zichzelf mag worden gekoppeld.

Voor het totaal aantal verbindingen - de som van graden - noteren we l (in het voorbeeld hebben we $l = 8$). Merk op dat l een even getal is omdat iedere kant precies 2 verbindingen nodig heeft.

Het kan tijdens een trekking voorkomen dat een verbinding in een bepaalde knoop gekoppeld wordt aan een andere verbinding in dezelfde knoop. De kant die dan gevormd wordt noemen we een lus. Het aantal lussen dat uiteindelijk gevormd worden in knoop i , noteren we door S_i . En precies in S_i zijn we geïnteresseerd.



Figuur 2: Een graaf gevormd met behulp van het configuratiemodel. Als gevolg van figuur 1

2 Kansen op lussen in een knoop

We zullen vanaf nu voor het gemak werken met knoop 1. Dit had uiteraard iedere andere knoop kunnen zijn. De kansen die we nu bepalen voor knoop 1 kunnen dus algemeen gemaakt worden voor iedere knoop.

Verder is het belangrijk om op te merken dat we tijdens het proces van trekkingen in een knoop iteratief langs iedere verbinding zullen lopen. We zullen dus als het ware alle verbindingen nummeren en dan beginnen met een trekking in verbinding 1 en daarna in 2 (behalve als verbinding 1 en 2 gekoppeld worden, dan gaan we verder naar 3), enz.

2.1 Geen lussen

De graad van knoop 1 is genoteerd als d_1 . We starten het proces van trekkingen. De kans dat knoop 1 geen lussen bevat is gelijk aan: de kans dat de eerste verbinding gekoppeld wordt met geen verbinding van dezelfde knoop, keer de kans dat de tweede verbinding niet gekoppeld wordt aan een verbinding in dezelfde knoop, enz.

De kans dat de eerste verbinding niet gekoppeld wordt met dezelfde knoop, is gelijk aan de kans dat deze verbinding koppelt met een verbinding van knoop i , $i = 2, 3, \dots, n$, dus: $\frac{l-d_1}{l-1}$. Nu zijn er buiten knoop 1 nog $l - d_1 - 1$ beschikbare verbindingen en in totaal nog $l - 3$. De kans dat de volgende verbinding van knoop 1 nu ook gekoppeld wordt met een verbinding buiten knoop 1 is gelijk aan $(l - d_1 - 1)/(l - 3)$. We blijven koppelen tot er geen vrije verbindingen meer zijn in knoop 1. Dit herhalen we dus d_1 keer. Dan volgt:

$$P(S_1 = 0) = \frac{l - d_1}{l - 1} \cdot \frac{l - d_1 - 1}{l - 3} \cdots \frac{l - 2d_1 + 1}{l - 2d_1 + 1} = \prod_{i=0}^{d_1-1} \frac{l - d_1 - i}{l - 2i - 1}. \quad (1)$$

Uiteindelijk eindigt het product bij $\frac{1}{1}$ voor de kans dat de laatste verbinding van knoop 1 gekoppeld wordt met een verbinding buiten deze knoop.

2.2 Één lus

We bepalen nu de kans dat een knoop precies één lus bevat. We zullen daartoe enkele gevallen moeten onderscheiden, namelijk: de kans dat de eerste verbinding koppelt met een van de andere vrije verbindingen in dezelfde knoop, de kans dat de tweede verbinding koppelt met een van de resterende verbindingen, enz.

Omdat we iteratief door de verbindingen heen gaan betekent dit dat de eerste verbinding een lus kan vormen met precies $d_1 - 1$ vrije verbindingen. De tweede verbinding met nog maar $d_1 - 2$ vrije verbindingen, enz. Dit levert dus:

$$\begin{aligned} P(S_1 = 1) = & \left[\frac{d_1 - 1}{l - 1} \right] \left[\frac{l - d_1}{l - 3} \cdot \frac{l - d_1 - 1}{l - 5} \cdots \frac{l - 2d_1 + 3}{l - 2d_1 + 3} \right] + \\ & \left[\frac{l - d_1}{l - 1} \right] \left[\frac{d_1 - 2}{l - 3} \right] \left[\frac{l - d_1 - 1}{l - 5} \cdots \frac{l - 2d_1 + 3}{l - 2d_1 + 3} \right] + \\ & \vdots \\ & \left[\frac{l - d_1}{l - 1} \cdot \frac{l - d_1 - 1}{l - 3} \cdots \frac{l - 2d_1 + 3}{l - 2d_1 + 5} \right] \left[\frac{d_1 - (d_1 - 1)}{l - 2d_1 + 3} \right] \end{aligned}$$

De enkele breuk tussen haakjes representeert telkens de kans op een lus.

Bovenstaande kans kunnen we omschrijven naar:

$$\sum_{n=1}^{d_1-1} L_n, \quad L_n = \left[\frac{(l - d_1)(l - d_1 - 1) \cdots (l - 2d_1 + 3)}{(l - 1)(l - 3) \cdots (l - 2d_1 + 3)} \right] (d_1 - n).$$

We vinden nu dus:

$$\begin{aligned} P(S_1 = 1) &= \frac{(l - d_1)(l - d_1 - 1) \cdots (l - 2d_1 + 3)}{(l - 1)(l - 3) \cdots (l - 2d_1 + 3)} \sum_{n=1}^{d_1-1} d_1 - n \\ &= \frac{P(S_1 = 0)}{l - 2d_1 + 2} \sum_{n=1}^{d_1-1} d_1 - n \\ &= \frac{P(S_1 = 0)}{l - 2d_1 + 2} \cdot \binom{d_1}{2} \end{aligned}$$

2.3 Meerdere lussen

Voor k algemeen kunnen we net als voor $k = 1$ de kans uitschrijven, namelijk:

$$\begin{aligned}
 P(S_1 = k) = & \left[\frac{d_1 - 1}{l - 1} \cdot \frac{d_1 - 3}{l - 3} \cdots \frac{d_1 - 2k + 1}{l - 2k + 1} \right] \left[\frac{l - d_1}{l - 2k - 1} \cdot \frac{l - d_1 - 1}{l - 2k - 3} \cdots \frac{l - 2d_1 + 2k + 1}{l - 2d_1 + 2k + 1} \right] + \\
 & \left[\frac{d_1 - 1}{l - 1} \cdot \frac{d_1 - 3}{l - 3} \cdots \frac{d_1 - 2k + 3}{l - 2k + 3} \right] \left[\frac{l - d_1}{l - 2k + 1} \right] \left[\frac{l - 2k + 1}{l - 2k - 1} \right] \left[\frac{l - d_1 - 1}{l - 2k - 3} \cdots \frac{l - 2d_1 + 2k + 1}{l - 2d_1 + 2k + 1} \right] + \\
 & \left[\frac{d_1 - 1}{l - 1} \cdot \frac{d_1 - 3}{l - 3} \cdots \frac{d_1 - 2k + 3}{l - 2k + 3} \right] \left[\frac{l - d_1}{l - 2k + 1} \cdots \frac{l - 2d_1 + 2k + 1}{l - 2d_1 + 2k + 3} \right] \left[\frac{d_1 - (d_1 - 1)}{l - 2d_1 + 2k + 1} \right] + \\
 & \vdots \\
 & \left[\frac{l - d_1}{l - 1} \cdots \frac{l - 2d_1 + 2k + 1}{l - 2d_1 + 4k + 3} \right] \left[\frac{d_1 - 1}{l - 2d_1 + 4k + 1} \cdot \frac{d_1 - 3}{l - 2d_1 + 4k - 1} \cdots \frac{d_1 - (d_1 - 1)}{l - 2d_1 + 2k + 1} \right]
 \end{aligned}$$

In bovenstaande formule zie je dat we uiteindelijk alle mogelijke combinaties van k lussen af zijn gegaan. Eerst verschuiven we lus k een plaats verder op, tot aan het einde. Vervolgens doen we hetzelfde voor lus $k - 1$, enz. Merk op dat lus $l \leq k$ altijd na lus $l - 1$ komt en dus automatisch een plek op moet schuiven indien we lus $l - 1$ op willen schuiven. Indien dit goed begrepen is zal de volgende kortere notatie niet heel verbazend zijn:

$$P(S_1 = k) = L_k \cdot \left(\sum_{\alpha_1=1}^{d_1-2k+1} \sum_{\alpha_2=\alpha_2+2}^{d_1-2k+3} \cdots \sum_{\alpha_k=\alpha_{k-1}+2}^{d_1-1} (d_1 - \alpha_1)(d_1 - \alpha_2) \cdots (d_1 - \alpha_k) \right), \quad (2)$$

met

$$L_k = \frac{(l - d_1)(l - d_1 - 1) \cdots (l - 2d_1 + 2k + 1)}{(l - 1)(l - 3) \cdots (l - 2d_1 + 2k + 1)}. \quad (3)$$

Deze sommaties zijn ingewikkeld en onhandig. Gelukkig kunnen we deze anders interpreteren. Zoals we net al aangaven zijn we alle mogelijke combinaties afgegaan voor k lussen. De sommaties zijn gelijk aan dit totaal aantal mogelijkheden. Dus rest een eenvoudigere notatie hiervoor. Deze vinden we en is gelijk aan:

$$\binom{d_1}{2k} (2k - 1)!!. \quad (4)$$

Met $(2n - 1)!! = (2n - 1) \cdot (2n - 3) \cdots 1$ voor iedere $n \in \mathbb{N}$, waarin $(-1)!! = 1$. De combinatie in (4) geeft het aantal verschillende manieren waarop we $2k$ verbindingen kunnen isoleren binnen knoop 1. Dit zijn precies het aantal verbindingen die nodig zijn om k lussen te vormen. Als deze $2k$ verbindingen gekozen zijn kunnen we de lussen hierin plaatsen op $(2k - 1)!!$ manieren. Hiermee vinden we de formule:

$$P(S_1 = k) = L_k \cdot \binom{d_1}{2k} (2k - 1)!! \quad (5)$$

Nu kunnen we L_k nog omschrijven naar een functie met faculteiten en semifaculteiten, zodat de formule wat duidelijker wordt:

$$L_k = \frac{(l - d_1)!}{(l - 2d_1 + 2k)!} \cdot \frac{(l - 2d_1 + 2k - 1)!!}{(l - 1)!!}.$$

Hiermee vinden we ten slotte:

$$P(S_1 = k) = \frac{(l - d_1)!}{(l - 2d_1 + 2k)!} \cdot \frac{(l - 2d_1 + 2k - 1)!!}{(l - 1)!!} \cdot \binom{d_1}{2k} (2k - 1)!!, \quad k = 0, 1, \dots, \left\lfloor \frac{d_1}{2} \right\rfloor.$$

Merk op dat $\left\lfloor \frac{d_1}{2} \right\rfloor$ inderdaad het maximaal aantal lussen is in knoop 1.

3 Limietstellingen voor graad van $O(1)$ en $O(\sqrt{l})$

Tot nu toe hebben we kansen op lussen bepaald in het geval dat l en d_1 eindig natuurlijke getallen zijn. Maar hoe gedraagt de verdeling van het aantal lussen in knoop 1 zich wanneer $l \rightarrow \infty$. Dit is uiteraard afhankelijk van hoe d_1 zich gedraagt als functie van l . Als d_1 namelijk een constante blijft is deze verdeling volkomen anders dan wanneer d_1 mee loopt naar oneindig.

We onderscheiden hiertoe de volgende drie situaties.

- (i) $d_1 = O(1)$,
- (ii) $d_1 = O(\sqrt{l})$,
- (iii) $d_1 = O(l)$.

Op situatie (i) kunnen we eigenlijk direct een intuïtief antwoord geven. Er zijn in het gehele stelsel oneindig veel verbindingen en in knoop 1 maar een eindig aantal. De kans op een lus, dus dat een verbinding uit knoop 1 gekoppeld wordt aan een andere verbinding in dezelfde knoop, zal naar 0 gaan omdat er simpelweg veel meer koppelmogelijkheden zijn buiten de knoop zelf. Het bewijs hiervoor vind u in de volgende sectie.

Wanneer d_1 mee loopt naar oneindig zal het antwoord niet zo eenvoudig zijn.

Merk op dat het proces van trekkingen betekent dat we telkens een Bernoulli variabele beschouwen waarin 1 staat voor een lus en 0 voor geen lus. De som van al deze Bernoulli variabelen levert dan het totaal aantal lussen. De verdeling van deze som lijkt dus erg veel de binomiale verdeling. Merk op dat een binomiale verdeling bestaat uit identieke en onafhankelijke Bernoulli variabelen. Beide eigenschappen gelden niet in het trekkings proces.

Echter ontstaat hierdoor wel het idee dat situatie (ii) convergeert naar een Poisson verdeling. Er geldt namelijk dat de $\text{Bin}(n, p)$ -verdeling met $np = \mu \in \mathbb{R}_{>0}$ voor $n \rightarrow \infty$ convergeert naar de Poisson verdeling met parameter μ . Kiezen we dan $n = \sqrt{l}$ en $p = \frac{\sqrt{l}}{l}$, dan volgt $np = 1$.

Voor situatie (iii) vermoeden we een normale verdeling. Nu geldt in de limiet dat de kans op een lus redelijk aanwezig is - in tegenstelling tot de kans $\frac{\sqrt{l}}{l}$ die naar 0 -, zeg met kans p . Toepassen van de centrale limietstelling op de binomiale verdeling $\text{Bin}(n, p)$ voor $n \rightarrow \infty$ levert direct een normale verdeling.

We zullen de drie gevallen achtereenvolgens bekijken.

3.1 Bewijs: $P(S_1 = 0) \rightarrow 1$

Theorema 1. *Beschouw een random graaf uit het configuratie model met een totaal aan l verbindingen en een knoop met $d_1 = O(1)$ verbindingen. Als $l \rightarrow \infty$, dan*

$$P(S_1 = 0) \rightarrow 1.$$

Bewijs.

$$P(S_1 = 0) = \prod_{i=0}^{d_1-1} \frac{l - d_1 - i}{l - 1 - 2i}.$$

De termen in dit product zijn strikt stijgend, daarom gelden de volgende twee ongelijkheden:

$$\begin{aligned} P(S_1 = 0) &\leq \prod_{i=0}^{d_1-1} \frac{l - 2d_1 + 1}{l - 2d_1 + 1} = \left(\frac{l - 2d_1 + 1}{l - 2d_1 + 1} \right)^{d_1}, \\ P(S_1 = 0) &\geq \prod_{i=0}^{d_1-1} \frac{l - d_1}{l - 1} = \left(\frac{l - d_1}{l - 1} \right)^{d_1}. \end{aligned}$$

Als $l \rightarrow \infty$ volgt direct met behulp van de insluitstelling:

$$P(S_1 = 0) \rightarrow 1.$$

□

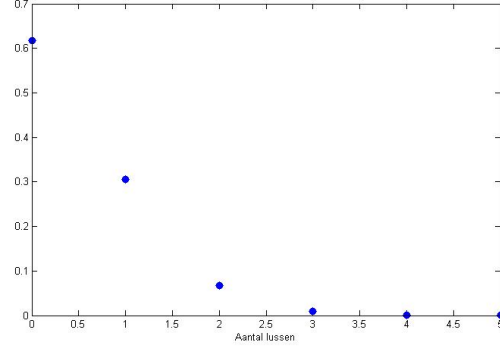
3.2 Simulaties naar de Poisson- en normale verdeling

We hebben eerder aangegeven dat we voor de gevallen $d_1 = O(\sqrt{l})$ en $d_1 = O(l)$ respectievelijk een Poisson- en normale verdeling als limiet verwachten. Voor we dit daadwerkelijk bewijzen gaan we na of onze verwachting overeenkomt met een simulatie.

Voor de simulaties hebben we $l = 2000$ gekozen. We hebben vervolgens 1000 keer het trekkingsproces uitgevoerd en iedere keer het totaal aantal lussen in een vector opgeslagen. Voor de code zie Appendix A.

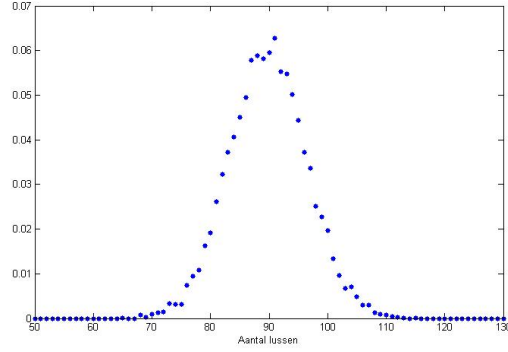
Voor het simuleren van de Poissonverdeling kiezen we $d_1 = 44$ ($\sqrt{l} \approx 44.7$). Het bijbehorende plot vind

u in figuur 3. Inderdaad volgt de grafiek het dalende karakter van een Poissonverdeling met een kleine parameter.



Figuur 3: Simulatie voor $l = 2000$ en $d_1 = 40$

Voor de simulatie van de normale verdeling hebben we gekozen voor $d_1 = 0.3 \cdot 2000 = 600$. Figuur 4 toont het resulterende plotje. Ook nu lijkt onze verwachting te kloppen. Het resultaat levert ongeveer een normale verdeling met verwachting ongeveer 90. Later zullen we zien dat de bijbehorende verwachting inderdaad precies 90 is!



Figuur 4: Simulatie voor $l = 2000$ en $d_1 = 600$

3.3 Convergentie naar de Poisson verdeling

Theorema 2. *Beschouw een random graaf uit het configuratie model met in totaal l verbindingen en een knoop met $d_1 = O(\sqrt{l})$ verbindingen. Voor het aantal lussen, S , in deze knoop geldt als $l \rightarrow \infty$*

$$P(S_1 = k) = \frac{1}{k!} \cdot \left(\frac{1}{2}\right)^k \cdot e^{-1/2} \quad \text{als } n \rightarrow \infty,$$

m.a.w. Poisson verdeeld met parameter $\frac{1}{2}$.

Bewijs. Als eerst gaan we de kans $P(S_1 = k)$ noteren met behulp van $P(S_1 = 0)$, dit doen we omdat $P(S_1 = 0)$ netjes convergeert naar $e^{-1/2}$, en dus al een deel van de Poisson verdeling vast legt. Uit (5) en (3) weten we nog dat:

$$P(S_1 = k) = L_k \cdot \binom{d_1}{2k} (2k-1)!!, \quad \text{met} \quad L_k = \frac{(l-d_1)(l-d_1-1) \cdots (l-2d_1+2k+1)}{(l-1)(l-3) \cdots (l-2d_1+2k+1)}.$$

Met behulp van $P(S_1 = 0)$ uit (1) zien we dat er geldt:

$$L_k = \frac{P(S_1 = 0)}{\frac{(l-2d_1+2k)(l-2d_1+2k-1) \cdots (l-2d_1+1)}{(l-2d_1+2k-1)(l-2d_1+2k-3) \cdots (l-2d_1+1)}}.$$

We kunnen in de onderste breuk alle oneven getallen wegstrepen:

$$\frac{(l-2d_1+2k)(l-2d_1+2k-1)\cdots(l-2d_1+1)}{(l-2d_1+2k-1)(l-2d_1+2k-3)\cdots(l-2d_1+1)} = (l-2d_1+2k)(l-2d_1+2k-2)\cdots(l-2d_1+2),$$

deze term is op zijn beurt te schrijven als:

$$\frac{2^{l/2-d_1+k} \cdot (l/2-d_1+k)!}{2^{l/2-d_1} \cdot (l/2-d_1)!} = 2^k \frac{(l/2-d_1+k)!}{(l/2-d_1)!}.$$

Dit levert

$$P(S_1 = k) = P(S_1 = 0) \cdot 2^{-k} \frac{(l/2-d_1)!}{(l/2-d_1+k)!} \cdot \binom{d_1}{2k} (2k-1)!! \quad (6)$$

We kunnen $P(S_1 = k)$ nog een stapje verder vereenvoudigen door de combinatie in bovenstaande vergelijking als volgt om te schrijven

$$\binom{d_1}{2k} (2k-1)!! = \frac{d_1!}{2^k \cdot k! \cdot (d_1-2k)!} \quad (7)$$

Met behulp van (7) gaat (6) over in

$$P(S_1 = k) = P(S_1 = 0) \cdot \frac{(l/2-d_1)!}{(l/2-d_1+k)!} \cdot \frac{d_1!}{k! \cdot (d_1-2k)!} \cdot 2^{-2k}. \quad (8)$$

Nu zijn we in ieder geval van de dubbele faculteiten en de combinatie af. Dat maakt het rekenwerk eenvoudiger.

De tweede stap is om aan te tonen dat geldt:

$$P(S_1 = 0) \sim e^{-\frac{1}{2}}.$$

Merk op dat als er inderdaad sprake is van convergentie naar een Poisson verdeling dat hiermee meteen volgt dat de parameter $\lambda = \frac{1}{2}$.

Er geldt:

$$P(S_1 = 0) = \prod_{j=0}^{d_1-1} \left[\frac{l-d_1-j}{l-1-2j} \right] = \prod_{j=0}^{d_1-1} \left[1 + \frac{1+j-d_1}{l-1-2j} \right] = \prod_{j=0}^{d_1-1} \left[1 - \frac{j}{l-2(d_1-j)+1} \right].$$

Nemen van het logaritme levert en toepassen van de Taylorreeks van het natuurlijke logaritme levert:

$$\ln[P(S_1 = 0)] = \sum_{j=0}^{d_1-1} \ln \left[1 - \frac{j}{l-2(d_1-j)+1} \right] \sim - \sum_{j=0}^{d_1-1} \left[\frac{j}{l-2(d_1-j)+1} \right].$$

Het nemen van de limiet van l naar oneindig in deze laatste term geeft ons

$$\lim_{l \rightarrow \infty} \frac{j}{l-2(d_1-j)+1} = \frac{j}{l}.$$

Omdat $d_1 \sim \sqrt{l}$, volgt

$$\ln[P(S_1 = 0)] \sim - \sum_{j=0}^{d_1-1} \left[\frac{j}{l} \right] = \frac{1}{2l} d_1(d_1-1) \sim -\frac{1}{2}.$$

Dit resulteert in

$$P(S_1 = 0) \sim e^{-\frac{1}{2}}.$$

Vullen we dit in in (8) dan levert dit de volgende uitdrukken:

$$P(S_1 = k) \sim \frac{e^{-\frac{1}{2}}}{k!} \cdot \left[\frac{(l/2-d_1)!}{(l/2-d_1+k)!} \cdot \frac{d_1!}{(d_1-2k)!} \cdot 2^{-2k} \right].$$

Omdat we convergentie naar $Pois(0.5)$ verwachten, verwachten we dus ook dat er zal gelden:

$$\frac{(l/2-d_1)!}{(l/2-d_1+k)!} \cdot \frac{d_1!}{(d_1-2k)!} \sim 2^k. \quad (9)$$

Inderdaad kunnen we (9) eenvoudig aantonen met behulp van de volgende eigenschap van de gamma functie:

$$\frac{\Gamma(z+a)}{\Gamma(z+b)} \sim z^{b-a}, \quad \text{als } z \rightarrow \infty. \quad (10)$$

We splitsen (9) als volgt op:

$$\underbrace{\frac{(l/2 - d_1)!}{(l/2 - d_1 + k)!}}_{\alpha} \cdot \underbrace{\frac{d_1!}{(d_1 - 2k)!}}_{\beta}. \quad (11)$$

- (i) Beschouw $z = 1/2 \cdot l - \sqrt{l}$, $a = 0$ en $b = k$. z gaat naar oneindig dus kunnen we (10) toepassen. Hieruit volgt:

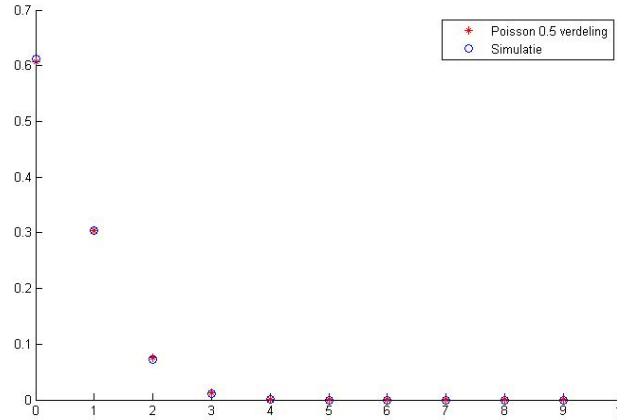
$$\alpha \sim \left(\frac{1}{2}l - \sqrt{l}\right)^{-k} \sim 2^k l^{-k}$$

- (ii) Beschouw $z = \sqrt{l}$, $a = 0$ en $b = -2k$. z gaat naar oneindig dus kunnen we (10) toepassen. Hieruit volgt:

$$\beta \sim (\sqrt{l})^{2k} \sim l^k.$$

Dus vinden we inderdaad de geldigheid van (9). □

Simulatie We zijn theorema 2 nagegaan met behulp van een simulatie. Zie Appendix A voor de bijbehorende code. We lieten $d = 44$ en $l = 2000$. Vervolgens hebben we $10 \cdot 10^4$ keer het trekkingsproces uitgevoerd. Het resulterende plotje in figuur 5 toont zowel de exacte Poissonverdeling met parameter $\frac{1}{2}$ en onze simulatie. Inderdaad overlappen beide elkaar heel netjes.



Figuur 5: Simulatie Poisson verdeling voor $l = 2000$ en $d_1 = 44$

4 Limietstelling voor graad van $O(l)$

In onze laatste situatie laten we $d_1 = O(l)$, de hoeveelheid verbindingen in knoop 1 loopt dus net zo hard mee naar oneindig als het totaal aantal verbindingen. Omdat $d_1 = O(l)$, betekent dit $d_1 \leq \alpha \cdot l$, met $\alpha \in \mathbb{R}_{>0}$. Echter zullen we ons beperken tot $\alpha < \frac{1}{2}$, omdat $\alpha > \frac{1}{2}$ direct impliceert dat knoop 1 lussen *moet* hebben.

In Sectie 3 hebben we gesproken over de reeks Bernoulli variabelen die we verkrijgen tijdens het proces van trekkingen. Deze reeks is te manipuleren tot een martingaal.

We noemen R_k het aantal lussen in knoop 1 na k trekkingen. Na ν_l , een stochastisch, aantal trekkingen zijn er geen lussen meer mogelijk (in knoop 1), omdat daarvoor onvoldoende verbindingen beschikbaar zijn. Omdat we nu alleen geïnteresseerd zijn in het aantal lussen stoppen we het proces na ν_l trekkingen. We willen het proces van trekkingen manipuleren tot een martingaal. Daarom zullen we eerst een martingaal, $\{X_k\}_{k=1}^{\nu_l}$, moeten construeren. Hierin geldt $X_k = a_k R_k + b_k$. We zullen dus de martingaalcoëfficiënten a_k en b_k moeten bepalen.

X_{ν_l} zal met behulp van de centrale limietstelling voor martingalen convergeren naar een normale verdeling. Maar uiteindelijk zijn we geïnteresseerd in R_{ν_l} . Daartoe is het dus van belang de limieten van a_{ν_l} en b_{ν_l} te bepalen. Deze limieten zullen uiteraard afhankelijk zijn van het stochastisch aantal trekkingen ν_l . We zullen daarom ook ν_l moeten bepalen.

4.1 Martingaal

Een discrete tijd martingaal is een stochastisch proces $(X_1, X_2, \dots)$ met de volgende eigenschap:

$$\mathbb{E}[X_k | X_{k-1}] = X_{k-1}.$$

We manipuleren $\{R_k\}_{k=1}^{\nu_l}$ tot een martingaal. Kies daartoe $X_k = a_k R_k + b_k$. We bepalen de coëfficiënten a_k en b_k .

Recursieve formules Zij R_{k-1} gegeven. De kans op een lus in de k -de trekking is dan gelijk aan:

$$P(R_k = R_{k-1} + 1 | R_{k-1}) = \frac{d_1 - k - R_{k-1}}{l - 2k + 1} = p_k.$$

De kans op geen lus in de k -de trekking is $P(R_k = R_{k-1} | R_{k-1}) = 1 - p_k$. Als X_k een martingaal dan $\mathbb{E}[X_k | R_{k-1}] = X_{k-1}$. Dus we dienen op te lossen:

$$a_k \cdot E[R_k | R_{k-1}] + b_k = a_{k-1} \cdot R_{k-1} + b_{k-1}. \quad (12)$$

Het linkerdeel van (12) kunnen we als volgt schrijven:

$$\begin{aligned} \mathbb{E}[R_k | R_{k-1}] + b_k &= a_k [p \cdot (R_{k-1} + 1) + (1 - p) \cdot R_{k-1}] + b_k \\ &= a_k \cdot R_{k-1} + p \cdot a_k + b_k \\ &= a_k \cdot R_{k-1} + \frac{d_1 - k - R_{k-1}}{l - 2k + 1} \cdot a_k + b_k \\ &= R_{k-1} \left[a_k - \frac{a_k}{l - 2k + 1} \right] + a_k \frac{d_1 - k}{l - 2k + 1} + b_k \end{aligned}$$

Hieruit kunnen we a_k recursief oplossen:

$$a_k - \frac{a_k}{l - 2k + 1} = a_{k-1} \quad \rightarrow \quad a_k = a_{k-1} \cdot \frac{l - 2k + 1}{l - 2k}.$$

Voor b_k vinden we de volgende recursieve oplossing, afhankelijk van a_{k-1} :

$$a_k \frac{d_1 - k}{l - 2k + 1} + b_k = b_{k-1} \quad \rightarrow \quad b_k = b_{k-1} - a_{k-1} \frac{d_1 - k}{l - 2k}.$$

We vinden dus de martingaal $\{X_k\}_{k=1}^{\nu_l}$ waarin

$$X_k = a_k \cdot R_k + b_k, \quad \text{met} \begin{cases} a_k = a_{k-1} \cdot \frac{l-2k+1}{l-2k}, & a_0 = 1 \\ b_k = b_{k-1} - a_{k-1} \frac{d_1-k}{l-2k}, & b_0 = 0 \end{cases}$$

Directe formules We kunnen a_k en b_k ook schrijven als functies van k .

Omdat a_k iedere keer vermenigvuldigd wordt met $(l - 2k + 1)/(l - 2k)$, volgt direct:

$$a_k = \prod_{i=1}^k \frac{l - 2i + 1}{l - 2i}.$$

Voor b_k geldt:

$$\begin{aligned}
b_k &= b_{k-1} - a_{k-1} \frac{d_1 - k}{l - 2k} \\
&= b_{k-2} - a_{k-2} \frac{d_1 - (k-1)}{l - 2(k-1)} - a_{k-1} \frac{d_1 - k}{l - 2k} \\
&\vdots \\
&= - \sum_{n=1}^k a_{n-1} \frac{d - n}{l - 2n} \\
&= - \sum_{n=1}^k \frac{d - n}{l - 2n} \cdot \left(\prod_{i=1}^{n-1} \frac{l - 2i + 1}{l - 2i} \right)
\end{aligned}$$

4.2 Convergentie van martingaal-coëfficiënten

Zoals eerder opgemerkt zijn we geïnteresseerd in het limiet van de zojuist bepaalde martingaal-coëfficiënten a_k en b_k , wanneer $k \rightarrow \infty$. Hiertoe hebben we de volgende twee lemma's.

Lemma 1. *Laat $0 < x < \frac{1}{2}$, l het totaal aantal verbindingen en $k = xl$. Als $l \rightarrow \infty$ dan*

$$a_k \sim \frac{1}{\sqrt{1 - 2x}}.$$

Bewijs. We weten $a_k = \prod_{n=1}^k \left(\frac{l-2n+1}{l-2n} \right)$. We nemen het logaritme, dit levert:

$$\ln(a_k) = \sum_{n=1}^k \ln \left(1 + \frac{1}{l - 2n} \right).$$

Als $l \rightarrow \infty$, dan

$$\frac{1}{l - 2n} \leq \frac{1}{(1 - 2x)l} \rightarrow 0, \quad \text{als } l \rightarrow \infty.$$

We mogen daarom met behulp van de Maclaurin reeks van het natuurlijk logaritme de volgende benadering doen:

$$\ln(a_k) = \sum_{n=1}^{xl} \ln \left(1 + \frac{1}{l - 2n} \right) = \sum_{n=1}^{xl} \left(\frac{1}{l - 2n} \right).$$

We kunnen van de laatste reeks een Riemann som maken, dat doen we als volgt:

$$\begin{aligned}
\sum_{n=1}^{xl} \left(\frac{1}{l - 2n} \right) &= \sum_{n=1}^{xl} \left(\frac{x}{xl - 2xn} \right) \\
&= x \sum_{n=1}^{xl} \frac{1}{xl} \left(\frac{1}{1 - 2x \cdot n/xl} \right).
\end{aligned}$$

Bovenstaande expressie is gelijk aan een riemann som, daarom is de volgende afleiding tot een integraal geldig:

$$\lim_{l \rightarrow \infty} x \sum_{n=1}^{xl} \frac{1}{xl} \left(\frac{1}{1 - 2x \cdot n/xl} \right) = \lim_{l \rightarrow \infty} x \sum_{n=1}^l \frac{1}{l} \left(\frac{1}{1 - 2x \cdot n/l} \right) = x \int_0^1 \frac{1}{1 - 2xy} dy.$$

Bovenstaande integraal is eenvoudig op te lossen voor $0 < x < \frac{1}{2}$:

$$\begin{aligned}
x \int_0^1 \frac{1}{1 - 2xy} dy &= -\frac{1}{2} \int_1^{1-2x} \frac{1}{y} dy \\
&= \frac{1}{2} \int_{1-2x}^1 \frac{1}{x} dx \\
&= \frac{1}{2} \ln \left(\frac{1}{1 - 2x} \right)
\end{aligned}$$

We vinden dus: $\lim_{l \rightarrow \infty} \ln(a_{x \cdot l}) = \frac{1}{2} \ln \left(\frac{1}{1-2x} \right)$. Wegwerken van het logaritme levert:

$$\lim_{l \rightarrow \infty} a_k = e^{\frac{1}{2} \ln \left(\frac{1}{1-2x} \right)} = \frac{1}{\sqrt{1-2x}}.$$

□

Lemma 2. Laat l het totaal aantal verbindingen, $d_1 = \alpha \cdot l$, met $0 < x \leq \alpha \leq \frac{1}{2}$ en $k = xl$. Als $l \rightarrow \infty$ dan

$$b_k \sim l \cdot \left(\frac{1 - \alpha - x}{\sqrt{1-2x}} + \alpha - 1 \right).$$

Bewijs. Laat $y = \frac{k}{l}$ en definiëer $\tilde{b}(y) = b_k$ dan geldt

$$\frac{\tilde{b}(y) - \tilde{b}\left(y - \frac{1}{l}\right)}{1/l} = \frac{b_k - b_{k-1}}{1/l} = -a_{k-1} \frac{d_1 - k}{l - 2k} \cdot l.$$

Omdat $k = yl$ en $d_1 = \alpha l$ volgt:

$$\frac{\tilde{b}(y) - \tilde{b}\left(y - \frac{1}{l}\right)}{1/l} = -l \cdot a_{yl-1} \frac{\alpha - y}{1 - 2y}.$$

Nemen we de limiet van $l \rightarrow \infty$ vinden we een benadering voor de afgeleide van $\tilde{b}(y)$:

$$\frac{d}{dy} \tilde{b}(y) \sim -l \cdot a_{yl-1} \frac{\alpha - y}{1 - 2y}. \quad (13)$$

Met behulp van de hoofdstelling van de integraalrekening volgt direct:

$$\tilde{b}(y) \sim \int_0^y -l \cdot a_{\bar{y}l-1} \frac{\alpha - \bar{y}}{1 - 2\bar{y}} d\bar{y}. \quad (14)$$

Voor iedere $\epsilon \in \mathbb{R}_{>0}$ geldt met behulp van Lemma 1 voor l voldoende groot:

$$a_{\bar{y}l-1} \sim \frac{1}{\sqrt{1-2\bar{y}}}.$$

Veranderen we de ondergrens in (14) van 0 naar ϵ dan vinden we:

$$\int_\epsilon^y -l \cdot a_{\bar{y}l-1} \frac{\alpha - \bar{y}}{1 - 2\bar{y}} d\bar{y} \sim \int_\epsilon^y -l \cdot \frac{1}{\sqrt{1-2\bar{y}}} \frac{\alpha - \bar{y}}{1 - 2\bar{y}} d\bar{y}.$$

Omdat dit laatste geldt voor alle $\epsilon > 0$, mogen we ook schrijven:

$$\tilde{b}(y) \sim \int_0^y -l \cdot \frac{1}{\sqrt{1-2\bar{y}}} \frac{\alpha - \bar{y}}{1 - 2\bar{y}} d\bar{y}.$$

Deze integraal kunnen we eenvoudig oplossen:

$$\begin{aligned} \int_0^y -l \cdot \frac{1}{\sqrt{1-2\bar{y}}} \frac{\alpha - \bar{y}}{1 - 2\bar{y}} d\bar{y} &= \int_0^y -l \cdot \frac{\alpha - \bar{y}}{(1 - 2\bar{y})^{3/2}} d\bar{y} \\ &= -l \cdot \left(\frac{-1}{2} \int_1^{1-2y} \left(\alpha - \frac{1}{2} + \frac{1}{2}u \right) \cdot u^{-3/2} du \right) \\ &= -l \cdot \left(\frac{1}{2} \left(\alpha - \frac{1}{2} \right) \int_{1-2y}^1 u^{-3/2} du + \frac{1}{4} \int_{1-2y}^1 u^{-1/2} du \right) \\ &= -l \cdot \left(\left(\frac{1}{2} - \alpha \right) u^{1/2} \Big|_{1-2y}^1 + \frac{1}{2} u^{1/2} \Big|_{1-2y}^1 \right) \\ &= -l \cdot \left(\frac{\alpha - \frac{1}{2}}{\sqrt{1-2y}} - \frac{1}{2} \sqrt{1-2y} + 1 - \alpha \right) \\ &= l \cdot \left(\frac{1 - \alpha - y}{\sqrt{1-2y}} + \alpha - 1 \right) \end{aligned}$$

We vinden nu dus voor $l \rightarrow \infty$:

$$\tilde{b}(y) \sim l \cdot \left(\frac{1 - \alpha - y}{\sqrt{1 - 2y}} + \alpha - 1 \right) \quad (15)$$

Verder weten we dat $k = xl \leftrightarrow y = x$, dus volgt:

$$b_k \sim l \cdot \left(\frac{1 - \alpha - x}{\sqrt{1 - 2x}} + \alpha - 1 \right).$$

□

4.3 Verwachte aantal lussen

We geven nu twee lemma's met betrekking tot de verwachting $\mathbb{E}[R_{xl}]$. Het eerste lemma bepaalt de verwachting in de limiet van $l \rightarrow \infty$. De tweede lemma laat zien dat $\mathbb{E}\left[\frac{R_k}{l}\right]$ een goede benadering is voor $\frac{R_k}{l}$ als $l \rightarrow \infty$.

Lemma 3. Laat $d_1 = \alpha \cdot l$, met $0 < x \leq \alpha \leq \frac{1}{2}$ en $k = xl$. Als $l \rightarrow \infty$, dan:

$$\mathbb{E}[R_k] \sim l \cdot (\alpha + x - 1 - (\alpha - 1)\sqrt{1 - 2x}).$$

Bewijs. X_k is een martingaal dus:

$$0 = \mathbb{E}[X_0] = \mathbb{E}[X_k] = \mathbb{E}[\mathbb{E}[X_k | R_{k-1}]] = \mathbb{E}[X_k] = a_k \mathbb{E}[R_k] + b_k.$$

Hiermee volgt:

$$\mathbb{E}[R_k] = -\frac{b_k}{a_k}.$$

Toepassen van Lemma 1 en Lemma 2 levert voor $l \rightarrow \infty$:

$$\begin{aligned} \mathbb{E}[R_k] &\sim l\sqrt{1 - 2x} \left(\frac{\alpha - \frac{1}{2}}{\sqrt{1 - 2x}} - \frac{1}{2}\sqrt{1 - 2x} + 1 - \alpha \right) \\ &= l \left(\alpha - \frac{1}{2} - \frac{1}{2}(1 - 2x) - \sqrt{1 - 2x}(\alpha - 1) \right) \\ &= l (\alpha + x - 1 - (\alpha - 1)\sqrt{1 - 2x}) \end{aligned}$$

□

Lemma 4. Laat $d_1 = \alpha \cdot l$, met $0 < \alpha \leq \frac{1}{2}$ en $k \in \mathbb{N}$. Als $l \rightarrow \infty$ dan

$$P \left(\left| \frac{R_k}{l} - \mathbb{E} \left[\frac{R_k}{l} \right] \right| \geq a \right) = O \left(\frac{1}{l} \right)$$

Bewijs. We maken gebruik van Chebyshev's ongelijkheid [1]:

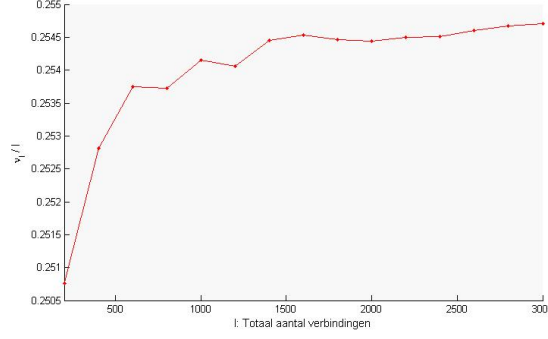
$$P \left(\left| \frac{R_k}{l} - \mathbb{E} \left[\frac{R_k}{l} \right] \right| \geq a \right) \leq \frac{1}{a^2 l^2} \cdot \text{Var}(R_k).$$

We hoeven dus alleen aan te tonen dat $\text{Var}(R_k) = O(l)$. We hebben:

$$\text{Var}(R_k) = \mathbb{E}[\text{Var}(R_k | R_{k-1})] + \text{Var}(\mathbb{E}[R_k | R_{k-1}]). \quad (16)$$

Omdat X_k een martingaal $\mathbb{E}[a_k R_k + b_k | R_{k-1}] = a_{k-1} R_{k-1} + b_{k-1}$. Verder geldt er $\text{Var}(R_k | R_{k-1}) = \text{Var}(R_{k-1} + \text{Ber}(p_k) | R_{k-1})$. Dit vullen we in in (16).

$$\begin{aligned} \text{Var}(R_k) &= \mathbb{E}[\text{Var}(R_k | R_{k-1})] + \text{Var} \left(\frac{a_{k-1}}{a_k} R_{k-1} \right) \\ &= \mathbb{E}[p_k(1 - p_k)] + \text{Var} \left(\frac{a_{k-1}}{a_k} R_{k-1} \right) \\ &\leq \mathbb{E}[p_k] + \text{Var}(R_{k-1}) \end{aligned}$$



Figuur 6: Convergentie van ν tov l

We maken de volgende afchatting voor $\mathbb{E}[p_k]$:

$$\mathbb{E}[p_k] = \mathbb{E}\left[\frac{d_1 - k - R_{k-1}}{l - 2k + 1}\right] \leq \mathbb{E}\left[\frac{d_1}{l}\right] \leq \alpha.$$

We vinden dus de volgende ongelijkheid voor de variantie van R_k :

$$\text{Var}(R_k) \leq \alpha + \text{Var}(R_{k-1}).$$

Hiermee volgt direct $\text{Var}(R_k) \leq \alpha k = O(l)$. □

4.4 Convergentie van de stoptijd

Het proces van trekkingen stopt wanneer er onvoldoende verbindingen over zijn om nog een lus te kunnen vormen in knoop 1. Het aantal trekkingen dat hiervoor nodig zijn is stochastische en noemen we de stoptijd, ν_l . Dan is R_{ν_l} dus het totaal aantal lussen in deze knoop.

We definiëren de stoptijd als volgt:

$$\nu_l = \min\{k \mid k + R_k \geq d_1 - 1\}.$$

We vermoeden dat als het aantal connecties, l , naar oneindig gaat dat de stoptijd gedeeld door l in waarschijnlijkheid naar een positieve constante convergeert. Dus:

$$\frac{\nu_l}{l} \xrightarrow{p} \gamma, \quad \gamma \in \mathbb{R}_{>0}.$$

Voordat we dit bewijzen, gaan we ons vermoeden na met behulp van een simulatie. De hiervoor gebruikte code in matlab is terug te vinden in appendix B.

We kiezen voor de simulatie $d_1 = 0.3 \cdot l$ en laten l lopen van 200 tot en met 3000 met een stapgrootte 200. In Figuur 6 kun je zien dat naarmate l groter wordt ν_l/l lijkt te convergeren - naar ongeveer 0.255. Dit kan dus goed duiden op convergentie naar een constante. Merk ten slotte nog op dat $0 < \gamma < \alpha$.

Nu zullen we dit bewijzen.

Lemma 5. *Laat $d_1 = \alpha \cdot l$. Als $l \rightarrow \infty$, dan*

$$\frac{\nu_l}{l} \rightarrow \gamma = \alpha - \frac{1}{2} \alpha^2.$$

Bewijs. De stoptijd ν_l is gedefiniëerd als $\nu_l = \min\{k \mid k + R_k \geq d_1 - 1\}$. Hieruit volgt $\nu_l + R_{\nu_l} \geq d_1 - 1$ en $\nu_l + R_{\nu_l-1} < d_1$ en dus

$$\frac{\nu_l}{l} + \frac{R_{\nu_l}}{l} \sim \alpha.$$

Met behulp van Lemma 4 en 3 weten we dat

$$\frac{R_{\nu_l}}{l} \rightarrow \alpha + \gamma - 1 - (\alpha - 1)\sqrt{1 - 2\gamma}, \quad (17)$$

indien $\frac{\nu_l}{l} \rightarrow \gamma$.

De rij ν_l/l is beperkt tot het interval $[0, 1]$ en dus uniform strak, m.a.w. voor iedere $\epsilon > 0$ bestaat er een K zodanig dat

$$P\left(\left|\frac{\nu_l}{l}\right| > K\right) \leq \epsilon.$$

Volgens Prohorovs theorema [2] geldt dan dat iedere deelrij l_k nog een deelrij l'_k heeft zodanig dat $\nu_{l'_k}/l'_k$ in waarschijnlijkheid convergeert, zeg naar γ . Langs deze deelrij geldt:

$$\gamma + \frac{R_{\nu_{l'_k}}}{l'_k} \sim \alpha.$$

Met behulp van (17) volgt

$$\gamma = \lim_{k \rightarrow \infty} -\frac{R_{\nu_{l'_k}}}{l'_k} + \alpha = 1 - \gamma + (\alpha - 1)\sqrt{1 - 2\gamma}$$

Nu bepalen we eenvoudig γ :

$$\begin{aligned} 2\gamma - 1 + (1 - \alpha)\sqrt{1 - 2\gamma} &\sim 0 \\ \rightarrow (1 - \alpha)^2(1 - 2\gamma) &\sim (1 - 2\gamma)^2 \\ \xrightarrow{\gamma < 1/2} (1 - \alpha)^2 &\sim (1 - 2\gamma) \\ \rightarrow \gamma &\sim \alpha - \frac{1}{2}\alpha^2 \end{aligned}$$

Dus iedere deelrij heeft dezelfde limiet waaruit volgt dat

$$\frac{\nu_l}{l} \rightarrow_p \alpha - \frac{1}{2}\alpha^2$$

□

Eerder zagen we dat bij $\alpha = 0.3$, γ leek te convergeren naar 0.255. Lemma 5 bevestigt dit vermoeden: $0.3 - \frac{1}{2} \cdot 0.3^2 = 0.255$.

4.5 Bewijs voor convergentie naar de normale verdeling

De centrale limietstelling voor martingalen [2] luidt als volgt:

Stel $X_1, X_2, \dots$ een martingaal relatief aan $R_1, R_2, \dots$ en laat $Y_{l,k} = \frac{X_k - X_{k-1}}{\sqrt{l}}$ een geschaald verschilmartingaal. Definieer

$$\sigma_{l,k}^2 = \mathbb{E}[Y_{l,k}^2 | R_k].$$

Stel nu dat

$$\sum_{k=1}^{\infty} \sigma_{l,k}^2 \rightarrow_p \sigma^2 \tag{18}$$

waarin σ een positieve constante, en dat de Lindeberg's conditie geldt:

$$\sum_{k=1}^{\infty} \mathbb{E}[Y_{l,k}^2 I_{|Y_{l,k}| > \epsilon}] \rightarrow 0 \tag{19}$$

voor iedere ϵ . Dan geldt dat $\sum_{k=1}^{\infty} Y_{l,k} \rightarrow N(0, \sigma^2)$.

We willen nu gebruik maken van deze stelling. Daartoe moeten we dus aantonen dat de geschaalde verschilmartingaal $Y_{l,k}$ voldoet aan (18) en (19). Dit doen we in Lemma 6 en 7 respectievelijk.

Lemma 6. Laat $d_1 = \alpha \cdot l$ en $k = \gamma \cdot l$, waarin $\gamma = \alpha - \frac{1}{2}\alpha^2$. Als $l \rightarrow \infty$, dan

$$\sum_{k=1}^{\gamma l} \mathbb{E}_{k-1} \left[\left(\frac{X_k}{\sqrt{l}} - \mathbb{E} \left[\frac{X_k}{\sqrt{l}} \right] \right)^2 \right] \sim \frac{\alpha^2}{2},$$

met $\mathbb{E}_{k-1}[f(X_k)] = \mathbb{E}[f(X_k) | R_{k-1}]$.

Bewijs. Laat

$$\bar{\sigma}_{l,k}^2 = \mathbb{E}_{k-1} \left[\left(\frac{X_k}{\sqrt{l}} - \mathbb{E} \left[\frac{X_k}{\sqrt{l}} \right] \right)^2 \right].$$

Als eerst berekenen we wat er gebeurd met iedere term in de sommatie als $l \rightarrow \infty$.

$$\begin{aligned} \bar{\sigma}_{l,k}^2 &= \frac{1}{l} \text{Var}(X_k \mid R_{k-1}) \\ &= \frac{a_k^2}{l} \text{Var}(R_k \mid R_{k-1}) \\ &= \frac{a_k^2}{l} \text{Var}(R_{k-1} + \text{Ber}(p_k) \mid R_{k-1}) \\ &= \frac{a_k^2}{l} p_k(1 - p_k). \end{aligned}$$

Laat $x = \frac{k}{l} \rightarrow k = xl$. Als $l \rightarrow \infty$, dan $k - 1 \sim xl$ en dus ook

$$p_k = \frac{d_1 - k - R_{k-1}}{l - 2k + 1} \sim \frac{\alpha - x - \frac{R_{xl}}{l}}{1 - 2x}.$$

Met behulp van Lemma 4 volgt:

$$p_k \sim \frac{1 - 2x - (\alpha - 1)\sqrt{1 - 2x}}{1 - 2x} = 1 + \frac{\alpha - 1}{\sqrt{1 - 2x}}.$$

Dus als $l \rightarrow \infty$, dan

$$p_k(1 - p_k) \sim \left(1 + \frac{\alpha - 1}{\sqrt{1 - 2x}} \right) \left(-\frac{\alpha - 1}{\sqrt{1 - 2x}} \right) = \left(\frac{1 - \alpha}{\sqrt{1 - 2x}} - \frac{(\alpha - 1)^2}{1 - 2x} \right)$$

en voor a_{xl} met behulp van Lemma 1:

$$a_{xl}^2 \sim \frac{1}{1 - 2x}.$$

Hiermee volgt dus voor $\bar{\sigma}_{l,k}^2$:

$$\bar{\sigma}_{l,k}^2 \sim \frac{1}{l} \cdot \frac{1}{1 - 2x} \cdot \left(\frac{1 - \alpha}{\sqrt{1 - 2x}} - \frac{(\alpha - 1)^2}{1 - 2x} \right).$$

Invullen van de gevonden $\bar{\sigma}_{l,k}^2$ in de sommatie en $l \rightarrow \infty$ levert:

$$\sum_{k=1}^{\gamma l} \left(\frac{1}{l} \cdot \frac{1}{1 - 2x} \cdot \left(\frac{1 - \alpha}{\sqrt{1 - 2x}} - \frac{(\alpha - 1)^2}{1 - 2x} \right) \right) = \int_0^\gamma \frac{1}{1 - 2x} \cdot \left(\frac{1 - \alpha}{\sqrt{1 - 2x}} - \frac{(\alpha - 1)^2}{1 - 2x} \right) dx. \quad (20)$$

Merk op: de integraal bevat a_{xl} en b_{xl} en de gebruikte benadering voor deze termen geldt voor iedere $x \in \mathbb{R}_{>0}$. De integraal mag dus vanaf een willekeurige ondergrens $\epsilon \in \mathbb{R}_{>0}$ beginnen. Omdat deze ϵ willekeurig is mogen we de integraal laten lopen vanaf 0. Ditzelfde argument is ook gebruikt voor het bepalen van b_k wanneer $l \rightarrow \infty$.

We lossen nu de integraal uit (20) op:

$$\begin{aligned} &\int_0^\gamma \left(\frac{1 - \alpha}{(1 - 2x)^{3/2}} - \frac{(\alpha - 1)^2}{(1 - 2x)^2} \right) dx \\ &= (1 - \alpha) \int_0^\gamma \frac{1}{(1 - 2x)^{3/2}} dx - (\alpha - 1)^2 \int_0^\gamma \frac{1}{(1 - 2x)^2} dx \\ &= \frac{1 - \alpha}{2} \int_{1-2\gamma}^1 u^{-3/2} - \frac{(\alpha - 1)^2}{2} \int_{1-2\gamma}^1 u^{-2} \\ &= \frac{1 - \alpha}{2} \left[-2u^{-1/2} \right]_{1-2\gamma}^1 - \frac{(\alpha - 1)^2}{2} \left[-u^{-1} \right]_{1-2\gamma}^1 \\ &= (1 - \alpha) \left(\frac{1}{\sqrt{1 - 2\gamma}} - 1 \right) - \frac{(\alpha - 1)^2}{2} \left(\frac{1}{1 - 2\gamma} - 1 \right) \end{aligned}$$

We vervangen γ voor $\alpha - \frac{1}{2}\alpha^2$ dan $1 - 2\gamma = \alpha^2 - 2\alpha + 1 = (\alpha - 1)^2$. Merk verder op dat $1 - \alpha > 0$. Dan volgt:

$$\begin{aligned}\sum_{k=1}^{\gamma l} \bar{\sigma}_{l,k}^2 &\sim (1 - \alpha) \left(\frac{1}{1 - \alpha} - 1 \right) - \frac{(\alpha - 1)^2}{2} \left(\frac{1}{(\alpha - 1)^2} - 1 \right) \\ &= 1 - (1 - \alpha) - \frac{1}{2} + \frac{1}{2}(\alpha - 1)^2 \\ &= \frac{1}{2} \alpha^2\end{aligned}$$

□

Lemma 7. *Laat $d_1 = \alpha \cdot l$ en $\gamma \cdot l$ de stoptijd. Dan geldt:*

$$\lim_{l \rightarrow \infty} P(|X_k - X_{k-1}| = O(1)) = 1, \quad 0 \leq k \leq \gamma l.$$

Bewijs. We weten dat $R_k \leq R_{k-1} + 1$, dus

$$\begin{aligned}|X_k - X_{k-1}| &= |a_k R_k - a_{k-1} R_{k-1} + b_k - b_{k-1}| \\ &\leq |a_k(R_{k-1} + 1) - a_{k-1} R_{k-1} + b_k - b_{k-1}| \\ &\leq R_{k-1} |a_k - a_{k-1}| + |a_k| + |b_k - b_{k-1}|\end{aligned}$$

Voor a_k geldt $a_k \leq \frac{1}{\sqrt{1-2\gamma}} = O(1)$ en voor $|b_k - b_{k-1}|$ hebben we

$$|b_k - b_{k-1}| = \left| a_{k-1} \frac{d_1 - k}{l - 2k} \right| \leq \left| a_{k-1} \frac{\alpha l - \gamma l}{l - 2\gamma l} \right| = O(1).$$

Nu rest aan te tonen dat $R_{k-1} |a_k - a_{k-1}| = O(1)$. We weten dat $R_{k-1} \leq \alpha l = O(l)$. Dit samen met

$$\begin{aligned}|a_k - a_{k-1}| &= \left| a_{k-1} \frac{l - 2k + 1}{l - 2k} - a_{k-1} \right| \\ &= \left| \frac{a_{k-1}}{l - 2k} \right| \\ &= O(l^{-1})\end{aligned}$$

levert het gewenste resultaat. □

Theorema 3. *Beschouw een random graaf uit het configuratie model met een totaal aan l verbindingen en een knoop met $d_1 = \alpha l$ verbindingen. Voor het aantal lussen, R , in deze knoop geldt als $l \rightarrow \infty$*

$$P\left(\frac{R - \frac{\alpha^2 l}{2}}{\sqrt{\frac{1}{2}(\alpha - 1)^2 \alpha^2 l}} \leq x\right) \rightarrow \Phi(x).$$

Bewijs. Definiëer $Y_{l,k} = \frac{X_k - X_{k-1}}{\sqrt{l}}$ als $0 < k \leq \nu_l = \gamma l$ en anders $Y_{l,k} = 0$. Volgens Lemma 6 hebben we voor $l \rightarrow \infty$:

$$\sum_{k=1}^{\infty} \sigma_{l,k}^2 \rightarrow_p \frac{1}{2} \alpha^2.$$

Merk hiertoe op dat

$$\mathbb{E}_{k-1} \left[\left(\frac{X_k}{\sqrt{l}} - \mathbb{E} \left[\frac{X_k}{\sqrt{l}} \right] \right)^2 \right] = \mathbb{E}_{k-1} \left[\left(\frac{X_k}{\sqrt{l}} - \frac{X_{k-1}}{\sqrt{l}} \right)^2 \right] = \mathbb{E}_{k-1} [Y_{l,k}^2]$$

Voor alle $\epsilon > 0$ bestaat er volgens Lemma 7 een $L \in \mathbb{N}$ zodanig dat $\forall k \in \mathbb{N}_{\geq 0}$ en $\forall l > L$: $|Y_{l,k}| < \epsilon$. En dus als $l \rightarrow \infty$ wordt er voldaan aan de Lindeberg conditie:

$$\sum_{k=1}^{\infty} \mathbb{E} [Y_{l,k}^2 I_{|Y_{l,k}| \geq \epsilon}] \rightarrow 0.$$

Volgens de centrale limietstelling voor martingalen geldt nu dat

$$\sum_{k=1}^{\infty} Y_{l,k} \rightarrow N\left(0, \frac{1}{2} \alpha^2\right).$$

Dit laatste kunnen we nu verder uitwerken:

$$\sum_{k=1}^{\infty} Y_{l,k} = \sum_{k=1}^{\gamma l} \left(\frac{X_k - X_{k-1}}{\sqrt{l}} \right) = \frac{X_{\gamma l}}{\sqrt{l}}.$$

Hiermee volgt dat als $l \rightarrow \infty$, $X_{\gamma l}$ asymptotisch normaal verdeeld is:

$$X_{\gamma l} \sim N\left(0, \frac{1}{2} \alpha^2 l\right).$$

We weten dat $X_{\gamma l} = a_{\gamma l} R_{\gamma l} + b_{\gamma l} \rightarrow R_{\gamma l} = \frac{1}{a_{\gamma l}} X_{\gamma l} - \frac{b_{\gamma l}}{a_{\gamma l}}$, dus geldt

$$R_{\gamma l} \sim N\left(-\frac{b_{\gamma l}}{a_{\gamma l}}, \frac{1}{a_{\gamma l}^2} \frac{1}{2} \alpha^2 l\right). \quad (21)$$

Voor de verwachting van $R_{\gamma l}$ vinden we met behulp van Lemma 3 en 5

$$\mathbb{E}[R_k] \sim \frac{1}{2} \alpha^2 l. \quad (22)$$

Volgens Lemma 1 geldt $a_{\gamma l}^2 = \frac{1}{1-2\gamma}$ als $l \rightarrow \infty$ en toepassen van Lemma 5 geeft

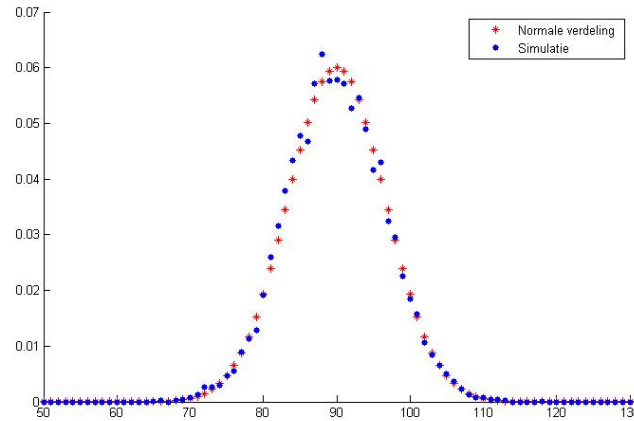
$$a_{\gamma l}^2 = \frac{1}{(\alpha - 1)^2}. \quad (23)$$

Invullen van (22) en (23) in (21) levert het eindresultaat

$$R_{\gamma l} \sim N\left(\frac{1}{2} \alpha^2 l, \frac{(\alpha - 1)^2 \alpha^2 l}{2}\right).$$

□

Simulatie Met behulp van een simulatie zijn we theorema 3 nagegaan. Zie Appendix A voor de bijbehorende code. We lieten $d = 600$ en $l = 2000$. Vervolgens hebben we $10 \cdot 10^4$ keer het trekkingsproces uitgevoerd. In Figuur 7 ziet u het resulterende plotje. Inderdaad volgt de verdeling lussen netjes de bijbehorende normale verdeling.



Figuur 7: Simulatie Poisson verdeling voor $l = 2000$ en $d_1 = 600$

Samenvatting

In het 1e hoofdstuk hebben we het configuratiemodel uitgebreid besproken. Vervolgens hebben we in hoofdstuk 2 een algemene formule bepaald voor de kans op $k \in \mathbb{N}$ lussen in een knoop. Deze expressie was zeer ingewikkeld vanwege een reeks sommaties die erin voor kwam. Deze konden we echter vereenvoudigen door in te zien dat deze gelijk stond aan het totaal aantal manieren waarop we k lussen konden combineren in de knoop.

Hoofdstuk 2 was voornamelijk bedoeld om het configuratiemodel en de probleemstelling beter te begrijpen. In hoofdstuk 3 begonnen we aan de limietstellingen, waarvoor we l naar oneindig lieten gaan. Hiertoe beschouwen we de volgende drie situaties: $d = O(1)$, $d = O(\sqrt{l})$, $d = O(l)$.

Het eerste geval was niet erg ingewikkeld. Wanneer $l \rightarrow \infty$ zal de kans op een lus in een eindige knoop verwaarloosbaar zijn. Dus de kans op 0 lussen gaat naar 1. Het bewijs hiervoor was slechts een toepassing van twee ongelijkheden.

Voor de tweede situatie verwachtten we dat het aantal lussen Poisson verdeeld zou zijn, en in de derde situatie normaal verdeeld. Deze situaties waren echter een stuk ingewikkelder dan de eerste, daarom hebben we onze verwachtingen eerst gecontroleerd met behulp van een simulatie. Inderdaad werden deze bevestigd.

We bewezen meteen dat als $d = O(\sqrt{l})$ dat het aantal lussen Poisson verdeeld was met parameter $\frac{1}{2}$. Om dit te bewijzen schreven we eerst de kans op k lussen om naar een formule afhankelijk van de kans op 0 lussen. We zagen dat de kans op 0 lussen in de limiet convergeert naar $e^{-1/2}$. Vervolgens maakten we gebruik van een eigenschap van de Gamma-functie om de rest van de Poisson verdeling aan te tonen.

Wanneer $d = O(l)$ hebben we gezegd dat dit betekent $d = \alpha \cdot l$, waarin $\alpha \leq \frac{1}{2}$. Daarna hebben we heel hoofdstuk 4 gebruikt om uiteindelijk te bewijzen dat er in dit geval sprake is van convergentie naar de normale verdeling met verwachting $\frac{1}{2}\alpha^2 l$ en variantie $\frac{(\alpha-1)^2 \alpha^2 l}{2}$.

Het bewijs maakt gebruik van de centrale limietstelling voor martingalen. Hiertoe hebben we eerst het random proces in een knoop van het configuratiemodel om moeten zetten naar een martingaal. Dit leverde een martingaal met een stochastische stoptijd. In de limiet van $l \rightarrow \infty$ bewezen we dat de stoptijd in kans convergeert naar een constante keer l . Verder was het essentieel om enkele kansen te bepalen met betrekking tot de verwachting van het aantal lussen tijdens het martingale proces.

Uiteindelijk konden we de centrale limietstelling voor martingalen toepassen op onze eerder gevonden martingaal. Vervolgens hebben we deze omgezet naar de stochast voor het aantal lussen in de knoop. Dit leverde het gewenste resultaat.

Appendices

A Simulatie van de Poisson- en normale verdeling

Hieronder ziet u de gebruikte code voor het simuleren van de normale verdeling. Dezelfde code is gebruikt voor de Poisson verdeling. Hiertoe hoeft u alleen d aan te passen.

Onderaan ziet men een stukje code voor het weergeven van de exacte normale verdeling. Dit wordt gebruikt om de simulatie te vergelijken.

```
1  %%% Simulatie Verdeling van lussen %%%
2  clear all;
3
4  N=2000;
5  M=1e4;
6  A=0.3;
7  B=1-A;
8  l=(A+B)*N;
9  d=A*l;
10
11 % We herhalen het trekkingsproces M keer
12 TotR=0;
13 for i=1:M
14
15     R=0;
16     k=1; % aantal trekkingen
17     dLeft=d; % aantal beschikbare verbindingen
18     while(dLeft>2)
19
20         p=(d-k-R)/(1-2*k+1); % kans op een nieuwe lus
21
22         if(binornd(1,p)==1)
23             R=R+1; % een nieuwe lus
24             dLeft=dLeft-1;
25         end
26
27         k=k+1; % aantal trekkingen + 1
28         dLeft=dLeft-1;
29
30     end
31
32     TotR(i)=R;
33 end
34
35 for i=50:1:130
36     PN(i-49)=sum(TotR==i)/length(TotR);
37 end
38
39 x=[50:1:130];
40
41 % Normale verdeling (exact)
42 mu=0.5*A^2*l;
43 sigma=(0.5*(A-1)^2*A^2*l);
44 y=normpdf(x,mu,sqrt(sigma));
45 hold on;
46 plot(x,y,'r*');
47 hold on;
48
49 plot(x,PN,'.','MarkerSize',15);
```

B Convergentie van ν/l

```

1  %%% Convergentie Nu_l / l %%%
2  clear all;
3
4  k=0;
5  l=200:200:3000;
6  M=1e3;
7
8  i=1;
9  for n=1:length(l)
10     ln=l(n);
11     k=k+1;
12     A=0.3;
13     d=round(A*ln);
14     nu=0;
15
16     for j=1:M
17         nuCount=0;
18         dleft=d;
19         toleft=ln;
20
21         % We gaan de trekkingen uitvoeren tot de knoop 'vol' zit.
22         while(dleft>1)
23             % Nieuwe verbinding
24             toleft=toleft-1;
25             dleft=dleft-1;
26
27             p=(dleft)/(toleft);
28
29             if(binornd(1,p)==1) % Lus
30                 dleft=dleft-1;
31             end
32
33             toleft=toleft-1;
34             nuCount=nuCount+1;
35         end
36
37         nu(j)=nuCount/ln;
38     end
39
40     m(i)=mean(nu);
41     i=i+1;
42 end
43
44 plot(l,m);

```

Symbolenlijst

d_i	Graad van knoop i	3
l	Som van graden	3
S_i	Aantal lussen in knoop i	3
α	Een constante $< 1/2$ die de grootte van d bepaald.	9
R_k	Aantal lussen na k trekkingen	9
ν_l	Stoptijd van de martingaal	9
X_k	Martingaal voor het trekkingsproces	10
a_k	Martingaalcoëfficiënt	10
b_k	Martingaalcoëfficiënt	10
p_k	Kans op een lus in de k -de trekking	10
γ	Convergentieconstante van de stoptijd	14
$Y_{l,k}$	Geschaalde verschilmartingaal voor X_k	15

Referenties

- [1] F.M. Dekking, C. Kraaikamp, H.P. Lopuhaä, L.E. Meester (2005), *A Modern Introduction to Probability and Statistics*. Springer.
- [2] P. Billingsley (1995), *Probability and Measure*. Third edition, J.Wiley & Sons.
- [3] S.Janson (2009). *The Probability That a Random Multigraph is Simple*. Combinatorics, Probability and Computing, 18, pp 205-225 doi:10.1017/S0963548308009644
- [4] R. van der Hofstad (2011), *Random Graphs and Complex Networks (lecture notes)*.