C.P. van Splunter

Grote afwijkingen

Bachelorscriptie, 21 april 2010

Scriptiebegeleiders:
prof.dr. F. Redig
prof.dr. E.A. Verbitskiy

Mathematisch Instituut, Universiteit Leiden

Inhoudsopgave

1 Inleiding **3**

2 Bovengrens **6**

3 Ondergrens **9**

4 Methode van Cramer **12**

5 Markov afhankelijke variabelen **14**

6 Voorbeeld: 2x2 matrix **21**

7 Voorbeeld: 3x3 matrix **24**

8 Conclusie **30**

9 Referenties **31**

10 Bijlage **32**

1 Inleiding

In mijn onderzoek zal ik mij bezig houden met ongewone gebeurtenissen, gebeurtenissen die slechts zeer zelden optreden. Voor een idee van dit soort gebeurtenissen kun je denken aan bijvoorbeeld aan het 100 keer opgooien van een geldstuk. Je verwacht dat dit geldstuk ongeveer 50 keer op kop valt en 50 keer op munt. De gebeurtenis dat het geldstuk 10 keer op munt valt en 90 keer op kop, is zeer zeldzaam. Deze gebeurtenis is dus een grote afwijking ten opzichte van de verwachting dat het geldstuk ongeveer 50 keer op kop valt en 50 keer op munt.
In deze scriptie zal ik mij gaan richten op grote afwijkingen. Nadat ik deze term gedefinieerd heb, zullen we zien dat de kans op een dergelijke grote afwijking naar 0 zal gaan. De gebeurtenis dat zo'n afwijking toch optreedt, is dus zeer zeldzaam.
Ergens begin 20^e eeuw begon men zich te interesseren voor dit onderwerp. In de afgelopen vier decennia heeft het een grote vlucht genomen. Op 22 maart 2007 kreeg 67-jarige Srinivasa S.R. Varadhan in Oslo de Abelprijs uitgereikt voor zijn fundamentele bijdragen in de kansrekening, met in het bijzonder zijn werk aan de theorie van de grote afwijkingen. Het onderwerp is dus zeer actueel en er wordt nog steeds veel onderzoek gedaan op dit gebied. Andere grote namen die veel onderzoek hebben verricht op dit gebied zijn Donsker, Freidlin en Wentzell. De theorie van grote afwijkingen heeft onder andere toepassingen in de economie, natuurkunde en biologie.

Ik zal beginnen met het maken van een aantal aannames (waaronder identiek en onafhankelijk verdeelde variabelen) en het definiëren van een aantal functies. Vervolgens zal ik een stelling bewijzen die zegt dat de kans op een grote afwijking zelfs exponentieel snel naar 0 gaat. Daarna zal ik iets zeggen over de zogenaamde Cramér theorie en tot slot ga ik het geval bekijken waar de variabelen een specifieke Markov-verdeling hebben. Hierbij zal ik ook een tweetal voorbeelden geven.

Maar eerst is het belangrijk om te weten wat nu precies een grote afwijking is. De definitie hiervan komt voort uit de wet van de grote aantallen. Om te zorgen dat we deze wet kunnen toepassen, hebben we een aantal aannames nodig:
We nemen $X_1, X_2, ..., X_n$ onderling onafhankelijk en identiek verdeeld (in het vervolg zal ik dit i.i.d. (independent and identically distributed) noemen). We nemen nu aan dat $\mathbb{E}(X_i) = \mu$ (eindige verwachting) en $\mathbb{E}(X_i^2) < \infty$ (eindige variantie). Onder deze aannames geldt:

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i = \mu$$

Deze wet van de grote aantallen wordt afgeleid uit de ongelijkheid van Chebyshev. Die zegt dat voor alle $\epsilon > 0$:

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geq \epsilon\right) \leq \frac{\sigma^2}{n\epsilon^2}$$

Als we n naar oneindig laten lopen, geldt dus voor elke $\epsilon > 0$:

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geq \epsilon\right) \leq \lim_{n\to\infty} \frac{\sigma^2}{n\epsilon^2} \to 0$$

Omdat $\mathbb{P}(|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu| \geq \epsilon)$ voor willekeurig kleine ϵ naar nul gaat als $n \to \infty$, weten we dat $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i = \mu$.
De gebeurtenis $|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu| \geq \epsilon$ noemen we een grote afwijking. De theorie van grote afwijkingen houdt zich bezig met het berekenen van kansen op dit soort afwijkingen.

Het eerste wat ik in mijn scriptie zal doen, is bewijzen dat de kans op een grote afwijking **exponentieel snel** naar 0 gaat als $n \to \infty$. We gaan hierbij weer uit van iid variabelen. Dus voor alle $\epsilon > 0$:

$$\lim_{n\to\infty} \mathbb{P}(|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu| \geq \epsilon) = \lim_{n\to\infty} e^{-nI(\mu+\epsilon)} \to 0 \tag{1}$$

Hierbij is I een of andere functie van $\mu + \epsilon$, onafhankelijk van n, die groter dan 0 is. Deze functie zullen we later specificeren.
Om het schrijfwerk wat te reduceren, definiëren we:

$$S_n := \sum_{i=1}^{n} X_i,$$
$$a := \mu + \epsilon = \mathbb{E}X_i + \epsilon > \mathbb{E}X_i.$$

Omdat we deze later nodig zullen hebben definiëren we ook onderstaande functies:

Definitie 1

$$\varphi(t) = \mathbb{E}(e^{tX_i}) = \mathbb{E}(e^{tX_1}) < \infty, t \in \mathbb{R}.$$

Definitie 2

$$F(t) = \log \varphi(t).$$

Als we nu in (1) links en rechts de natuurlijke logaritme nemen, en vervolgens delen door n, hebben we de volgende stelling te bewijzen:

Stelling 1 *Voor alle $a > \mathbb{E}X_i$ geldt:*

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq na) = -I(a)$$

waarbij $I(x) = \sup_{t\in\mathbb{R}}(tx - F(t))$.

Hoe we aan deze $I(x)$ komen zal blijken uit het bewijs.

2 Bovengrens

Zoals eerder aangegeven begin ik deze scriptie met het bewijzen van stelling 1. Dit ga ik doen door te laten zien dat $-I(a)$ zowel een bovengrens is, als een ondergrens voor $\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq na)$. In deze paragraaf zullen we laten zien dat het een bovengrens is:

Stelling 2 *Voor alle* $a > \mathbb{E}X_i$ *geldt:*

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq na) \leq -I(a)$$

waarbij $I(x) = \sup_{t\in\mathbb{R}}(tx - F(t))$.

Voor het bewijs van deze stelling hebben we nog iets meer informatie nodig. We hebben nodig dat $F(t)$ convex is, en we zullen $\varphi(t)$, zoals gedefinieerd in definitie 1, omschrijven in een Taylorreeks in $t = 0$. We beginnen met het bewijs dat $F(t)$ convex is:

Stelling 3 *F(t) is convex:*

$$F(\lambda t + (1-\lambda)s) \leq \lambda F(t) + (1-\lambda)F(s)$$

voor alle $0 \leq \lambda \leq 1$.

Bewijs:
Allereerst merken we op dat voor $\lambda = 0$ en $\lambda = 1$ het bewijs triviaal is.
Vervolgens gebruiken we de ongelijkheid van Hölder, die alsvolgt luidt:

$$\mathbb{E}(XY) \leq \mathbb{E}(|X|^p)^{\frac{1}{p}} \mathbb{E}(|Y|^q)^{\frac{1}{q}}$$

waarbij $p, q > 1$ en

$$\frac{1}{p} + \frac{1}{q} = 1$$

We kiezen nu p en q alsvolgt:

$$p = \frac{1}{\lambda} \qquad \text{en} \qquad q = \frac{1}{1-\lambda}$$

(dan geldt $\frac{1}{p} + \frac{1}{q} = \lambda + 1 - \lambda = 1$ en omdat $0 < \lambda < 1$, geldt ook $p, q > 1$)

Nu zien we:

$$F(\lambda t + (1-\lambda)s) = \log \mathbb{E}(e^{\lambda tX + (1-\lambda)sX}) = \log \mathbb{E}(e^{\lambda tX} \cdot e^{(1-\lambda)sX}) \leq$$
$$\log((\mathbb{E}(e^{tX}))^{\lambda}(\mathbb{E}(e^{sX}))^{(1-\lambda)}) = \lambda \log \mathbb{E}(e^{tX}) + (1-\lambda) \log \mathbb{E}(e^{sX}) =$$
$$\lambda F(t) + (1-\lambda)F(s)$$

Dus $F(t)$ is convex. $\square$

Als we vervolgens $\varphi(t)$ omschrijven in een Taylor reeks in $t = 0$, zien we, omdat

$$\frac{\partial^n \varphi(t)}{\partial t^n}|_{t=0} = \frac{\partial^n \mathbb{E}(e^{tX_1})}{\partial t^n}|_{t=0} = \mathbb{E}(\frac{\partial^n}{\partial t^n} e^{tX_1}|_{t=0}) = \mathbb{E}(X_1^n)$$

, het volgende:

$$\varphi(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} \mathbb{E}(X_1^n)$$

Dit is de reden dat $\varphi(t)$ ook wel de momentgenererende functie wordt genoemd ($\mathbb{E}(X_1^n)$ zijn de momenten).

We gaan nu over tot het bewijzen van stelling 2.

Bewijs (Stelling 2):
Eerst schrijven we $\mathbb{P}(S_n \geq na)$ anders:

$$\begin{aligned} &\mathbb{P}(S_n \geq na) = \mathbb{P}(S_n - na \geq 0) \\ &= \mathbb{P}(t(S_n - na) \geq 0) = \mathbb{P}(e^{t(S_n - na)} \geq 1) \end{aligned}$$

$\forall t \geq 0$.
Nu gaan we gebruik maken van de Markovongelijkheid: $\mathbb{P}(Y \geq 1) \leq \mathbb{E}(Y)$ als $Y \geq 0$ (in ons geval hebben we te maken met een e-macht, en is deze regel dus toe te passen). Het korte bewijs van de Markovongelijkheid volgt verderop.
Als we de Markovongelijkheid toepassen, krijgen we:

$$\begin{aligned} &\mathbb{P}(S_n \geq na) = \mathbb{P}(e^{t(S_n - na)} \geq 1) \\ &\leq \mathbb{E}(e^{t(S_n - na)}) = \varphi(t)^n e^{-nta} = e^{-n(ta - F(t))} \end{aligned}$$

Wat we nu zien is dat als we van beide kanten de natuurlijke logaritme nemen, vervolgens delen door n en dan van de rechterkant ook nog het supremum nemen over t (dit kunnen we doen omdat de ongelijkheid geldt voor alle $t \geq 0$), we onze functie $I(a)$ gevonden hebben:

$$\frac{1}{n} \log \mathbb{P}(S_n \geq na) \leq -\sup_{t \geq 0}(ta - F(t)) = -\sup_{t \in \mathbb{R}}(ta - F(t)) = -I(a)$$

De tweede stap kunnen we nemen omdat F convex is en (zoals we later zullen zien) $F'(0) = \mu$. Dit betekent dat als $F'(x) = \mu + \epsilon$, moet gelden dat $x > 0$ (vanwege de convexiteit van F). Het supremum van $t(\mu + \epsilon) - F(t)$ (als $(\mu + \epsilon) - F'(t) = 0$) wordt dus aangenomen voor een of andere t groter dan 0.
Het blijkt dus dat onze functie I precies de Legendre getransformeerde is. Deze is gedefinieerd als:

$$I(x) = \sup_t (tx - F(t))$$

Wat we nu nog moeten doen, is laten zien dat deze functie $I(a)$ strikt positief is. We zullen kijken hoe de afgeleide van $F(t)$ in het punt 0 eruit ziet:

$$
\begin{aligned}
F'(t) &= (\log \mathbb{E}(e^{tX}))' = \frac{1}{\mathbb{E}(e^{tX})}\mathbb{E}(Xe^{tX}) \\
F'(0) &= \frac{1}{\mathbb{E}(1)}\mathbb{E}(X) = \mathbb{E}X = \mu
\end{aligned}
$$

Als we nu $F(t)$ gaan ontwikkelen met Taylorontwikkeling rond $t = 0$, zien we:

$$
F(t) = F(0) + tF'(0) + \frac{1}{2}t^2F''(0) + \mathcal{O}(t^3) = t\mu + \frac{1}{2}t^2F''(0) + \mathcal{O}(t^3)
$$

Invullen in I geeft:

$$
\begin{aligned}
&I(a) = \sup_t(ta - F(t)) = \sup_t(t(\mu + \epsilon) - (t\mu + \frac{1}{2}t^2F''(0) + \mathcal{O}(t^3))) \\
&= \sup_t(t\epsilon - (\frac{1}{2}t^2F''(0) + \mathcal{O}(t^3)))
\end{aligned}
$$

We kunnen zeggen dat deze functie groter dan nul is, omdat voor t klein genoeg (zo klein dat $(\frac{1}{2}t^2F''(0) + \mathcal{O}(t^3)) < t\epsilon$) de functie positief is. Het supremum is dus zeker positief. Dus:

$$
I(a) > 0
$$

Nu we hebben bewezen dat $I(a) > 0$ hebben we dus een bovengrens gevonden voor $\frac{1}{n}\log \mathbb{P}(S_n \geq na)$:

$$
\frac{1}{n}\log \mathbb{P}(S_n \geq na) \leq -I(a)
$$

en dus ook:

$$
\limsup_{n\to\infty} \frac{1}{n}\log \mathbb{P}(S_n \geq na) \leq -I(a)
$$

$\square$

Rest ons alleen nog de Markovongelijkheid te bewijzen. Dit bewijs luidt alsvolgt:

$$
\begin{aligned}
&\mathbb{E}(Y) = \mathbb{E}(Y \cdot \mathbb{I}(Y \geq 1)) + \mathbb{E}(Y \cdot \mathbb{I}(Y < 1)) \\
&\geq \mathbb{E}(Y \cdot \mathbb{I}(Y \geq 1)) \geq 1 \cdot \mathbb{E}(\mathbb{I}(Y \geq 1)) = \mathbb{P}(Y \geq 1)
\end{aligned}
$$

($\mathbb{I}$ is hier de indicatorfunctie, en dus niet de eerdergenoemde I)
$\square$

3 Ondergrens

We hebben nu gezien dat $e^{-nI(a)}$ een bovengrens is voor $\mathbb{P}(\frac{1}{n}S_n \geq a)$. Wat we nu zullen laten zien, is dat $e^{-nI(a)}$ ook een ondergrens is voor $\mathbb{P}(S_n \geq na)$, als $n \to \infty$:

Stelling 4

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\frac{1}{n}S_n \geq a) \geq -I(a)$$

met $I(a) = \sup_t(ta - F(t))$.

Als we deze stelling hebben bewezen kunnen we met stelling 2 zeggen:

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\frac{1}{n}S_n \geq a) = -I(a)$$

Bewijs (Stelling 4):
Als eerste definiëren we Y_i als $Y_i := X_i - a$. Dan zien we:

$$\begin{aligned}
&\mathbb{E}_X(e^{tX}) = \mathbb{E}_Y(e^{t(Y+a)}) \Rightarrow \\
&\mathbb{E}_X(e^{tX}) = e^{ta}\mathbb{E}_Y(e^{tY}) \Rightarrow \\
&\log \varphi_X(t) = ta + \log \varphi_Y(t) \Rightarrow \\
&ta - F_X(t) = -F_Y(t) \Rightarrow \\
&\sup_t(ta - F_X(t)) = \sup_t(t \cdot 0 - F_Y(t)) \Rightarrow \\
&I_X(a) = I_Y(0)
\end{aligned}$$

Dan komt de te bewijzen ongelijkheid er alsvolgt uit te zien:

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\sum_{i=1}^{n} Y_i \geq 0) \geq I_Y(0)$$

$(\sum_{i=1}^n X_i \geq na \Leftrightarrow \sum_{i=1}^n X_i - na \geq 0 \Leftrightarrow \sum_{i=1}^n (X_i - a) \geq 0 \Leftrightarrow \sum_{i=1}^n Y_i \geq 0)$

Hierbij geldt dat $\mathbb{E}(Y_i) = \mathbb{E}(X_i - a) = \mathbb{E}(X_i - \mu - \epsilon) = -\epsilon < 0$.

Het is dus voldoende om te bewijzen:

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq 0) \geq I(0)$$

met $\mathbb{E}(X_i) < 0$.
Om dit te bewijzen zullen we eerst $\mathbb{P}(S_n \geq 0)$ gaan benaderen met de zogenaamde Cramér-truc. Deze truc is gebaseerd op het exponentieel veranderen van de maat, zodanig dat de atypische gebeurtenis $S_n \geq 0$ (deze is atypisch

omdat $\mathbb{E}(X_i) < 0$), typisch wordt onder de nieuwe maat.
We schrijven $\mathbb{P}(S_n \geq 0)$ eerst om:

$$\mathbb{P}(S_n \geq 0) = \mathbb{E}(I(S_n \geq 0)) = \frac{\mathbb{E}(I(S_n \geq 0)e^{-tS_n}e^{tS_n})}{\mathbb{E}(e^{tS_n})}\mathbb{E}(e^{tS_n})$$
$$= \tilde{\mathbb{E}}_t(I(S_n \geq 0)e^{-tS_n})\mathbb{E}(e^{tS_n})$$

We hebben nu een nieuwe kansmaat $\tilde{\mathbb{P}}_t$ met een verwachting $\tilde{\mathbb{E}}_t$ gemaakt, die we alsvolgt definiëren:

$$\tilde{\mathbb{E}}_t(Y) = \frac{\mathbb{E}(Ye^{tS_n})}{\mathbb{E}(e^{tS_n})}$$

Voor de nieuwe kansmaat geldt:

$$\frac{d\tilde{\mathbb{P}}_t}{d\mathbb{P}} = \frac{e^{tS_n}}{\mathbb{E}(e^{tS_n})}$$

Merk op dat onder deze nieuwe kansmaat de variablen $\tilde{X_1}, \ldots, \tilde{X_n}$ weer iid verdeeld zijn:

$$\tilde{\mathbb{E}}_t(f_1(\tilde{X_1}), \ldots, f_n(\tilde{X_n})) = \frac{\mathbb{E}(f_1(X_1), \ldots, f_n(X_n)e^{tX_1} \ldots e^{tX_n})}{\mathbb{E}(e^{tX_1} \ldots e^{tX_n})}$$
$$= \frac{\mathbb{E}(f_1(X_1)e^{tX_1})}{\mathbb{E}(e^{tX_1})} \ldots \frac{\mathbb{E}(f_n(X_n)e^{tX_n})}{\mathbb{E}(e^{tX_1})} = \tilde{\mathbb{E}}_t(f_1(\tilde{X_1})) \ldots \tilde{\mathbb{E}}_t(f_n(\tilde{X_n}))$$

Vervolgens zullen we kijken wat er gebeurt als we X_1 invullen in de nieuwe verwachting:

$$\tilde{\mathbb{E}}_t(X_1) = \frac{\mathbb{E}(X_1e^{tS_n})}{\mathbb{E}(e^{tS_n})} = \frac{\mathbb{E}(X_1e^{tX_1})\mathbb{E}(e^{tX_2})\mathbb{E}(e^{tX_3}) \ldots}{\mathbb{E}(e^{tX_1})\mathbb{E}(e^{tX_2})\mathbb{E}(e^{tX_3}) \ldots} = \frac{\mathbb{E}(X_1e^{tX_1})}{\mathbb{E}(e^{tX_1})} = \frac{\frac{d}{dt}\varphi(t)}{\varphi(t)} = \frac{d}{dt}F(t)$$

Omdat $X_1, \ldots, X_n$ iid zijn, weten dus nu dat voor alle i:

$$\tilde{\mathbb{E}}_t(X_i) = \frac{d}{dt}F(t)$$

We nemen nu $t = t^*$, zodanig dat $\frac{d}{dt}\varphi(t^*) = 0$ en dus $\frac{\frac{d}{dt}\varphi(t)}{\varphi(t)} = \frac{d}{dt}F(t^*) = \tilde{\mathbb{E}}_{t^*}(X_i) = 0$. We kiezen t^* op deze manier zodat $\frac{1}{n}S_n \simeq 0$ onder de nieuwe kansmaat en zodat dus onder de nieuwe kansmaat, de kans $\tilde{\mathbb{P}}_{t^*}(\frac{1}{n}S_n \geq 0)$ typisch wordt. Verder is t^* ook het unieke minimum van de functie $\varphi(t)$. Deze bestaat omdat $\varphi(t)$ convex is en omdat

$$\lim_{t \to \infty} \varphi(t) = \lim_{t \to -\infty} \varphi(t) = \infty$$

We kunnen nu $\mathbb{P}(S_n \geq 0)$ als volgt berekenen:

$$\mathbb{P}(S_n \geq 0) = \tilde{\mathbb{E}}_{-t^*}(I(S_n \geq 0)e^{t^*S_n})\mathbb{E}(e^{t^*S_n}) = \tilde{\mathbb{E}}_{-t^*}(I(S_n \geq 0)e^{t^*S_n})e^{-nF(t^*)}$$
$$= \tilde{\mathbb{E}}_{-t^*}(I(S_n \geq 0)e^{t^*S_n})e^{-n \cdot \sup_x(x \cdot 0 - F(x))} = \tilde{\mathbb{E}}_{-t^*}(I(S_n \geq 0)e^{t^*S_n})e^{-nI(0)}$$

Wat we nu nog moeten doen, is laten zien dat $\tilde{\mathbb{E}}_{t^*}(I(S_n \geq 0)e^{-t^* S_n})$ verwaarloosbaar is op deze schaal als we de $\liminf_{n\to\infty}$ nemen, dus:

$$\liminf_{n\to\infty} \frac{1}{n} \log \tilde{\mathbb{E}}_{t^*}(I(S_n \geq 0)e^{-t^* S_n}) = 0$$

Dat doen we als volgt:

$$\begin{aligned}
&\tilde{\mathbb{E}}_{-t^*}(I(S_n \geq 0)e^{t^* S_n}) \geq \tilde{\mathbb{E}}_{t^*}(I(S_n \geq 0)I(S_n \leq a\sqrt{n})e^{-t^* S_n}) \\
&\geq e^{-t^* a\sqrt{n}} \tilde{\mathbb{P}}_{t^*}(0 \leq \frac{1}{\sqrt{n}} S_n \leq a)
\end{aligned}$$

Omdat de verdeling van $\frac{1}{\sqrt{n}} S_n$ voor $n \to \infty$ naar de normale verdeling gaat volgens de centrale limietstelling ($\mu = \tilde{\mathbb{E}}_{t^*}(X_i) = 0$ en $\sigma^2 = \tilde{\mathbb{E}}_{t^*}(X_i)^2 < \infty$), is de kans groter dan nul en onafhankelijk van n. Nu zien we:

$$\mathbb{P}(S_n \geq 0) \geq e^{-t^* a\sqrt{n}} \tilde{\mathbb{P}}_{t^*}(0 \leq S_n \leq a\sqrt{n}) \cdot e^{-nI(0)}$$

Als we van beide kanten de logaritme nemen, vervolgens delen door n en tot slot de lim inf van $n \to \infty$ nemen, kunnen we schrijven:

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq 0) \geq -I(0)$$

Omdat we hebben laten zien dat dit voldoende was om te bewijzen geldt nu:

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(S_n \geq na) \geq -I(a)$$

□

Door de ondergrens en de bovengrens te combineren, kunnen we nu stelling 1 bewijzen.

Bewijs Stelling 1:
Combineer Stelling 2 en Stelling 4. □

4 Methode van Cramer

In het bewijs van stelling 4 hebben we gebruik gemaakt van de zogenaamde Cramér truc. Bij deze truc hebben we gebruikt dat de variabelen iid waren. De Cramér truc is echter ook algemeen toepasbaar. In het algemene geval zijn de variabelen niet iid verdeeld. De truc is nu gebaseerd op het exponentieel veranderen van maat zodanig dat de atypische kans $\mathbb{P}(\frac{1}{n}S_n \in [x-\delta, x+\delta])$ met δ klein, typisch wordt. Deze kans zullen we voortaan noteren als $\mathbb{P}(\frac{1}{n}S_n \approx x)$. We schrijven weer eerst deze kans om totdat we onze nieuwe verwachting (dezelfde als hiervoor) hebben gevonden:

$$\mathbb{P}(\frac{1}{n}S_n \approx x) = \mathbb{E}(I(\frac{1}{n}S_n \approx x)) = \frac{\mathbb{E}[(I(\frac{1}{n}S_n \approx x))e^{-tS_n}e^{tS_n}]}{\mathbb{E}(e^{tS_n})}\mathbb{E}(e^{tS_n})$$
$$= \tilde{\mathbb{E}}_t(I(\frac{1}{n}S_n \approx x)e^{-tS_n}))\mathbb{E}(e^{tS_n})$$

De nieuwe verwachting $\tilde{\mathbb{E}}_t$ en de nieuwe kans $\tilde{\mathbb{P}}_t$ voldoen dus weer aan:

$$\tilde{\mathbb{E}}_t(Y) = \frac{\mathbb{E}(Ye^{tS_n})}{\mathbb{E}(e^{tS_n})}$$
$$\frac{d\tilde{\mathbb{P}}_t}{d\mathbb{P}} = \frac{e^{tS_n}}{\mathbb{E}(e^{tS_n})}$$

Net zoals eerder kunnen we weer $t = t^*$ kiezen zoals we willen, omdat de vergelijking geldt voor alle t. In dit geval kiezen we t^* zo, dat $\liminf_{n\to\infty} \tilde{\mathbb{E}}_{t^*}(\frac{1}{n}S_n) = x$ (op deze manier wordt $\mathbb{P}(\frac{1}{n}S_n \approx x)$ typisch als n groot is, en dat is wat we willen). Dan kunnen we zeggen:

$$\liminf_{n\to\infty} \mathbb{P}(\frac{1}{n}S_n \approx x) = \liminf_{n\to\infty} \tilde{\mathbb{E}}_{t^*}(I(\frac{1}{n}S_n \approx x)e^{-t^*S_n}))\mathbb{E}(e^{t^*S_n})$$
$$= \liminf_{n\to\infty} e^{-t^*nx}\mathbb{E}(e^{t^*S_n}) = \liminf_{n\to\infty} e^{(-n(t^*x - \frac{1}{n}F_n(t^*)))}$$

Nu stuiten we op een probleem: Eerder konden we zeggen dat, omdat de variabelen iid waren

$$\frac{1}{n}F_n = \frac{1}{n}\log \mathbb{E}(e^{tS_n}) = \log \mathbb{E}(e^{tX_1}) = F$$

en was $\frac{1}{n}F_n$ dus onafhankelijk van n. Dat dit nu ook het geval is, moeten we eerst nog maar bewijzen. Wat we dus eigenlijk willen is dat voor $n \to \infty$ $\frac{1}{n}F_n(t^*) \to F(t^*)$ met $F(t)$ een of andere functie onafhankelijk van n. Dit is in het algemeen niet het geval. Wel kunnen we dit laten zien voor variabelen met een bepaalde afhankelijkheid. In het volgende hoofdstuk zullen we dit laten zien voor Markov afhanklijke variabelen. Als we dit bewezen hebben voor variabelen met een bepaalde afhankelijkheid, kunnen we ook nog verder redeneren.

Als we gaan kijken naar de afgeleide van $\frac{1}{n}F_n(t^*)$ blijkt:

$$\liminf_{n\to\infty} \frac{1}{n} F_n'(t^*) = \liminf_{n\to\infty} \frac{1}{n} \frac{\mathbb{E}(S_n e^{t^* S_n})}{\mathbb{E}(e^{t^* S_n})}$$
$$= \liminf_{n\to\infty} \frac{1}{n} \tilde{\mathbb{E}}_{t^*}(S_n) = x$$

Dus als we hebben bewezen dat $\frac{1}{n}F_n(t^*) \to F(t^*)$ voor $n \to \infty$, dan weten we ook dat $F'(t^*) = x$. Wat we hiermee kunnen zien, is dat

$$(xt^* - F(t^*))' = x - F'(t^*) = 0$$

Dit betekent dus dat de functie $xt - F(t)$ zijn supremum aanneemt voor $t = t^*$. Nu zien we dus, als we de logaritme nemen en delen door n:

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\frac{1}{n} S_n \approx x) = \liminf_{n\to\infty} -(t^* x - \frac{1}{n} F_n(t^*))$$
$$= -(t^* x - F(t^*)) = -\sup_t (tx - F(t)) = -I(x)$$

5 Markov afhankelijke variabelen

We weten dat we met de wet van de grote aantallen kunnen zien, dat als $\mu = \mathbb{E}(X_i)$ en $\frac{1}{n}S_n \to \mu$, dan

$$\mathbb{P}(\frac{1}{n}S_n \geq \mu + \epsilon) \to 0$$

Verder hebben we tot nu toe gezien dat als de variabelen X_i onafhankelijk verdeeld zijn, de kans op een grote afwijking zelfs exponentieel snel naar 0 gaat. In deze sectie gaan we uit van Markov afhankelijke variabelen $X_1, X_2, .., X_n$. We willen iets kunnen zeggen over een bepaalde functie f van deze Markov afhankelijk variabelen. We defieren

$$Z_k = f(X_k)$$

We gaan uit van een Markov keten met een eindige toestandsruimte Ω (dus $X_i \in \Omega$), die aperiodiek en irreducibel is. Er is dan dus een unieke stationaire maat op de Markov keten. Het eerste wat we zien met de wet van de grote aantallen, is dat voor $n \to \infty$:

$$\frac{1}{n}S_n = \frac{1}{n}\sum_{i=1}^{n} Z_i \to \sum_{x\in\Omega} \mu(x)f(x) = \mathbb{E}_\mu(f(X_i))$$

Zonder verlies van algemeenheid kunnen we deze verwachting 0 nemen:

$$\mathbb{E}_\mu(f(X_i)) = 0$$

Om te kijken of $\mathbb{P}(\frac{1}{n}\sum_{i=1}^{n} Z_i > \epsilon) \to 0$ als $n \to \infty$, gebruiken we weer de ongelijkheid van Chebyshev. Echter in dit geval moeten we eerst laten zien dat $\mathrm{Var}(\sum f(X_i))$ niet van orde n^2 of hoger is. De wet van de grote aantallen geldt immers niet voor oneindig grote variantie. Dat is het eerste wat we zullen doen:

$$\begin{aligned}
\mathrm{Var}(\sum_{i=1}^{n} f(X_i)) &= \mathbb{E}_\mu((\sum_{i=1}^{n} f(X_i))^2) - \mathbb{E}_\mu(\sum_{i=1}^{n} f(X_i))^2 = \mathbb{E}_\mu((\sum_{i=1}^{n} f(X_i))^2) \\
&= \mathbb{E}_\mu(\sum_{i=1}^{n}\sum_{j=1}^{n} f(X_i)f(X_j)) = 2\mathbb{E}_\mu(\sum_{i=1}^{n}\sum_{j\geq i} f(X_i)f(X_j)) \\
&= 2\mathbb{E}_\mu(\sum_{i=1}^{n}(f(X_i))^2) + 2\mathbb{E}_\mu(\sum_{i=1}^{n}\sum_{j>i} f(X_i)f(X_j)) = 2(I + II)
\end{aligned}$$

Wat we nu zullen laten zien, is dat deze beide termen van orde n zijn:

$$I = \mathbb{E}_\mu(\sum_{i=1}^{n}(f(X_i))^2) = n\mathbb{E}_\mu(f(X_i)^2) = n\mathbb{E}_\mu(f(X_1)^2) = O(n)$$

$$II = \sum_{i=1}^{n}\sum_{j>i} \mathbb{E}_\mu(f(X_i)f(X_j)) = \sum_{i=1}^{n}\sum_{j>i} \mathbb{E}_\mu(f(X_0)f(X_{j-i}))$$

We hebben hierbij gebruikt dat het voor de verwachting niet uitmaakt of we een stuk naar links of naar rechts in de Markovketen opschuiven. Dit volgt uit de stationariteit van de Markov keten:

$$\forall k : \mathbb{P}(X_i = a, X_j = b) = \mathbb{P}(X_{i+k} = a, X_{j+k} = b)$$

Wat we vervolgens zullen laten zien is dat geldt:

$$\mathbb{E}_\mu(f(X_0)f(X_{j-i})) \leq e^{-c(j-i)}$$

met c een of andere positieve constante. Als we dit namelijk hebben bewezen kunnen we zeggen:

$$II = \sum_{i=1}^{n}\sum_{j>i}\mathbb{E}_\mu(f(X_0)f(X_{j-i})) \leq \sum_{i=1}^{n}\sum_{j>i} e^{-c(j-i)}$$
$$\leq \sum_{i=1}^{n}\sum_{r=0}^{\infty} e^{-cr} = \frac{n}{1-e^{-c}} = O(n)$$

We hebben nu alleen nog nodig dat

$$\mathbb{E}_\mu(f(X_0)f(X_{j-i})) \leq e^{-c(j-i)}$$

We maken hierbij de aanname dat we te maken hebben met een primitieve matrix $\mathbf{B}$. In dat geval is bovenstaande ongelijkheid een direct gevolg van de zogenaamde **spectral gap** die ontstaat tussen de grootste eigenwaarde en alle kleinere eigenwaarden van de matrix. Dat deze **spectral gap** er is, zal ik bewijzen in Stelling 5 deel (b).

Wat we nu gaan doen, is naar een stelling toe werken die iets zegt over $\mathbb{P}(\frac{1}{n}S_n \approx x)$. Hierbij bedoel ik met $\approx x$ weer $\in [x-\delta, x+\delta]$. Om deze stelling te kunnen bewijzen, heb ik eerst een andere stelling nodig. Deze stelling van Perron-Frobenius beschrijft wat eigenschappen van irreducibele matrices:

Stelling 5 (Perron-Frobenius) *Laat $\mathbf{B} = B(i,j)_{i,j=1}^{|\Omega|}$ een irreducibele, primitieve matrix (iedere staat in de Markovketen is te bereiken vanuit iedere willekeurige staat en $\mathbf{B}^k > 0$ voor zekere k). Dan bevat $\mathbf{B}$ een eigenwaarde ρ (de Perron-Frobenius eigenwaarde) zodat:*
(a) $\rho > 0$ is reëel.
(b) Voor iedere eigenwaarde $\lambda \neq \rho$ van $\mathbf{B}$ geldt: $|\lambda| < \rho$.
(c) Er bestaan linker en rechter eigenvectoren bij de eigenwaarde ρ met alleen maar strikt postitieve coördinaten.
(d) De linker en de rechter eigenvectoren μ en θ bij de eigenwaarde ρ zijn uniek op vermenigvuldiging met een constante na.
(e) Voor alle $i \in \Omega$ en iedere $\phi = (\phi_1, \ldots, \phi_{|\Omega|})$ zodat $\phi_j > 0$ voor alle j geldt:

$$\lim_{n\to\infty}\frac{1}{n}\log[\sum_{j=1}^{|\Omega|} B^n(i,j)\phi_j] = \lim_{n\to\infty}\frac{1}{n}\log[\sum_{j=1}^{|\Omega|}\phi_j B^n(j,i)] = \log\rho$$

Zoals eerder aangegeven zullen we deel (b) van deze stelling bewijzen. Voor het bewijs van deel (b) heb ik ook het bewijs van deel (a) nodig. Omdat ik voor het bewijs van stelling 6 deel (e) van deze stelling nodig heb geef ik dus in totaal drie bewijzen: deel (a), (b) en (e).
Bewijs (Stelling 5 deel (a)):
Neem een rijvector $\mathbf{x}$, $\mathbf{x}' \neq \mathbf{0}'$ en $\mathbf{x}' \geq \mathbf{0}$.
Laat

$$\rho(\mathbf{x}) = \min_j \frac{\sum_i x_i b_{ij}}{x_j} \qquad 0 \leq \rho(\mathbf{x}) < \infty$$

De rechterkant van deze vergelijking interpreteren we als ∞ als $x_j = 0$. Dan geldt:

$$\begin{aligned} x_j \rho(\mathbf{x}) &\leq \sum_i x_i b_{ij} \qquad \text{voor alle j} \\ \mathbf{x}'\rho(\mathbf{x}) &\leq \mathbf{x}'\mathbf{B} \\ \mathbf{x}'\mathbf{1}\rho(\mathbf{x}) &\leq \mathbf{x}'\mathbf{B}\mathbf{1} \end{aligned}$$

We definieren

$$K = \max_i \sum_j b_{ij}$$

Dan moet gelden $\mathbf{B}\mathbf{1} \leq K\mathbf{1}$ en dus

$$\rho(\mathbf{x}) \leq \frac{\mathbf{x}'\mathbf{1}K}{\mathbf{x}'\mathbf{1}} = K = \max_i \sum_j b_{ij}$$

Dus $\rho(\mathbf{x})$ is begrensd van boven.
Omdat $\mathbf{B}$ primitief is, bevat $\mathbf{B}$ geen kolommen met alleen nullen. Dus ook $\rho(\mathbf{1}) > 0$. We kunnen nu ρ alsvolgt definieren:

$$\rho = \sup_{\mathbf{x} \geq 0; \mathbf{x} \neq 0} \min_j \frac{\sum_i x_i b_{ij}}{x_j}$$

Hieruit zien we dat

$$0 < \rho(\mathbf{1}) \leq \rho \leq K < \infty.$$

Omdat zowel de teller, als de noemer niet veranderen met de norm van $\mathbf{x}$ (met de definitie van $\rho(\mathbf{x})$ zien we $\rho(c\mathbf{x}) = \rho(\mathbf{x})$), kunnen we ook een $\mathbf{x}$ nemen met norm 1:

$$\rho = \sup_{\mathbf{x} \geq 0; \mathbf{x}'\mathbf{x} = 1} \min_j \frac{\sum_i x_i b_{ij}}{x_j}$$

Nu is het gebied $\{\mathbf{x}; \mathbf{x} \geq 0, \mathbf{x}'\mathbf{x} = 1\}$ compact in de Euclidische n-ruimte $\mathbb{R}^n$ en $\rho(\mathbf{x})$ is de half-continue projectie van dit gebied op $\mathbb{R}^1$. Het supremum wordt dus aangenomen voor een zekere $\mathbf{x}$, zeg $\hat{\mathbf{x}}$. Er bestaat dus een $\hat{\mathbf{x}}$ zodat

$$\min_j \frac{\sum_i \hat{x}_i b_{ij}}{\hat{x}_j} = \rho \qquad \forall j \qquad \text{dus}$$

$$\sum_i \hat{x}_i b_{ij} \geq \rho \hat{x}_j. \qquad \text{Kortom} \qquad \hat{\mathbf{x}}'\mathbf{B} \geq \rho \hat{\mathbf{x}}'.$$

voor iedere $j = 1, 2, \ldots, n$ en met gelijkheid voor een zeker element van $\hat{\mathbf{x}}$.
Bekijk nu:

$$\mathbf{z}' = \hat{\mathbf{x}}\mathbf{B} - \rho \hat{\mathbf{x}}' \geq \mathbf{0}'$$

Nu geldt $\mathbf{z}' = \mathbf{0}'$ of niet. Stel dat $\mathbf{z}' \neq \mathbf{0}'$. We weten dat voor $k \geq k_0$, $\mathbf{B}^k > 0$ (omdat $\mathbf{B}$ primitief is). Omdat $\mathbf{z}' \geq \mathbf{0}'$, $\mathbf{z}' \neq \mathbf{0}'$ en $\mathbf{B}^k > 0$ geldt:

$$\mathbf{z}'\mathbf{B}^k > 0$$

Dus ook

$$\mathbf{z}'\mathbf{B}^k = (\hat{\mathbf{x}}'\mathbf{B}^k)\mathbf{B} - \rho(\hat{\mathbf{x}}'\mathbf{B}^k) > \mathbf{0}' \qquad \text{dus}$$

$$\Big(\frac{(\hat{\mathbf{x}}'\mathbf{B}^k)\mathbf{B}}{\hat{\mathbf{x}}'\mathbf{B}^k}\Big)_j > \rho \qquad \text{voor elke } j$$

Hierbij geeft het subscript j, het j-de element aan.
Dit is in tegenstelling met de definitie van ρ. Er geldt dus altijd $\mathbf{z} = \mathbf{0}$, dus

$$\mathbf{x}'\mathbf{B} = \rho \mathbf{x}'$$

□

Bewijs (Stelling 5 deel (b)):
Laat λ een eigenwaarde van $\mathbf{B}$. We moeten laten zien dat als $\lambda \neq \rho$, dan geldt: $|\lambda| < \rho$. Dit zullen we doen door eerst te laten zien dat $|\lambda| \leq \rho$ en vervolgens dat als $|\lambda| = \rho$, dan $\lambda = \rho$ (wat in tegenstelling is met de aanname $\lambda \neq \rho$.
Voor een zekere $\mathbf{x} \neq \mathbf{0}$ (mogelijk complex) geldt voor alle j's:

$$\sum_i x_i b_{ij} = \lambda x_j \tag{2}$$

$$|\lambda x_j| = |\sum_i x_i b_{ij}| \leq \sum_i |x_i b_{ij}| \leq \sum_i |x_i| b_{ij}$$

$$|\lambda| \leq \frac{\sum_i |x_i| b_{ij}}{|x_j|} \qquad \text{(de rechterkant is hier } \infty \text{ als } x_j \text{ gelijk is aan 0)}$$

$$|\lambda| \leq \min_j \frac{\sum_i |x_i| b_{ij}}{|x_j|} \leq \rho \qquad \text{met de definitie van } \rho \text{ in het bewijs van (a)}$$

We moeten nu alleen nog laten zien dat $|\lambda| \neq \rho$. Stel $|\lambda| = \rho$. Dan

$$\sum_i |x_i| b_{ij} \geq |\lambda||x_j| = \rho |x_j|$$

Dit is een soortgelijke situatie die we zagen in het bewijs van (a):

$$\sum_i |x_i| b_{ij} = \rho |x_j|, \qquad > 0 \qquad j = 1, 2, \ldots, n$$
$$\sum_i |x_i| b_{ij}^{(k)} = \rho^k |x_j|, \qquad > 0 \qquad j = 1, 2, \ldots, n$$

Nu zien we:

$$|\sum_i x_i b_{ij}^{(k)}| = |\lambda^k x_j| = \sum_i |x_i b_{ij}^{(k)}| = \sum_i |x_i| b_{ij}^{(k)}$$

Als voor twee getallen γ en δ geldt $|\gamma + \delta| = |\gamma| + |\delta|$, dan hebben deze twee getallen dezelfde richting in het complexe vlak. Als we dus x_j schrijven als $x_j = |x_j| e^{\theta_j i}$, dan geldt $\theta_j = \theta$. Als we x_j in deze vorm invullen in vergelijking (2), krijgen we

$$\sum_i |x_i| b_{ij} = \lambda |x_j|$$

Hierbij geldt: $|x_i| > 0$ voor alle i en λ is reeel en positief. Omdat $|\lambda| = \rho$, moet nu gelden $\lambda = \rho$. Dit is in tegenstelling met onze aanname $\lambda \neq \rho$. Dus $|\lambda| < \rho$.
□

Bewijs (Stelling 5 deel (e)):
We definiëren α, β, γ en δ als volgt:

$$\alpha = \sup_i \theta_i \qquad \beta = \inf_i \theta_i > 0 \qquad \gamma = \sup_i \phi_i \qquad \delta = \inf_i \phi_i > 0$$

Hierbij is θ de rechter eigenvector bij eigenwaarde ρ en ϕ en ρ zoals in de stelling. Dan geldt voor alle $i, j \in \Omega$:

$$\frac{\gamma}{\beta} B^n(i,j)\theta_j \geq \frac{\theta_j}{\beta} B^n(i,j)\phi_j \geq B^n(i,j)\phi_j$$
$$\geq \frac{\theta_j}{\alpha} B^n(i,j)\phi_j \geq \frac{\delta}{\alpha} B^n(i,j)\theta_j$$

We kunnen dus nu zeggen:

$$\lim_{n\to\infty} \frac{1}{n} \log[\sum_{j=1}^{|\Omega|} B^n(i,j)\phi_j] = \lim_{n\to\infty} \frac{1}{n} \log[\sum_{j=1}^{|\Omega|} B^n(i,j)\theta_j]$$
$$= \lim_{n\to\infty} \frac{1}{n} \log(\rho^n \theta_i) = \lim_{n\to\infty} (\frac{1}{n} \cdot n \log(\rho) + \frac{1}{n}\theta_i) = \log \rho$$

Zo kunnen we op vergelijkbare wijze laten zien dat ook:

$$\lim_{n\to\infty} \frac{1}{n} \log[\sum_{j=1}^{|\Omega|} \phi_j B^n(j,i)] = \log \rho$$

$\square$

Nu we het bewijs hebben gezien van stelling 5 deel (e) kunnen we bijna overgaan tot het bewijzen van de hoofdstelling van dit hoofdstuk. Maar voordat we dat doen moeten we eerst nog een aantal definities geven.

Noem $\Pi = \pi(i,j)_{i,j=1}^{|\Omega|}$ de transitie matrix van onze Markov keten met $\pi(i,j)$ de kans dat er vanuit punt i naar punt j gesprongen wordt. Laat vervolgens $\mathbb{P}_\sigma^\pi$ de kansmaat horende bij deze transitie matrix Π met als startpunt σ. Dus

$$\mathbb{P}_\sigma^\pi(Y_1 = y_1, \ldots, Y_n = y_n) = \pi(\sigma, y_1)\Pi_{i=1}^{n-1}\pi(y_i, y_{i+1}). \tag{3}$$

Verwachtingen horende bij $\mathbb{P}_\sigma^\pi$ noteren we als $\mathbb{E}_\sigma^\pi$.

We definiëren Π_t als:

Definitie 3

$$\Pi_t = \left(\pi_t(i,j)\right)_{i,j=1}^{|\Omega|} = \left(\pi(i,j)e^{tf(j)}\right)_{i,j=1}^{|\Omega|}$$

Omdat we te maken hebben met een irreducibele Markovketen, is de transitiemartix Π irreducibel. Omdat $e^{tf(j)} > 0$ is Π_t dus ook irreducibel. Verder noemen we $\rho(\Pi_t)$ de Perron-Frobenius eigenwaarde van Π_t. We hebben nu genoeg informatie om te kunnen overgaan tot het bewijzen van de hoofdstelling, die als volgt luidt.

Stelling 6 *Laat Y_k een eindige Markov keten met een irredubibele transitiematrix Π. Voor alle $x \in \mathbb{R}$ definiëren we:*

$$I(x) = \sup_{t\in\mathbb{R}}(tx - \log\rho(\Pi_t))$$

(Hierin speelt $\log\rho(\Pi_t)$ dus de rol van $F(t)$.)

Dan geldt:

$$\liminf_{n\to\infty} \frac{1}{n}\log\mathbb{P}_\sigma^\pi(\frac{1}{n}S_n \approx x) = -I(x)$$

Bewijs:

We definiëren net zoals eerder:

$$F_n(t) = \log\mathbb{E}_\sigma^\pi(e^{tS_n}).$$

Zoals we eerder zagen is het voldoende om te laten zien dat

$$F(t) = \lim_{n\to\infty}\frac{1}{n}F_n(t) = \lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}_\sigma^\pi(e^{tS_n})$$

bestaat voor alle $t \in \mathbb{R}$, dat F eindig is, dat F differentieerbaar is in heel $\mathbb{R}$ en dat $F(t) = \log\rho(\Pi_t)$.

Om te beginnen kijken we naar $F_n(t)$ (onder de maat $\mathbb{P}_\sigma^n$):

$$\begin{aligned}
F_n(t) \quad &= \log \mathbb{E}_\sigma^\pi(e^{tS_n}) \\
&= \log \sum_{y_1,\ldots,y_n} \mathbb{P}_\sigma^\pi(Y_1 = y_1, \ldots, Y_n = y_n)\Pi_{k=1}^n e^{tf(y_k)} \\
&= \log \sum_{y_1,\ldots,y_n} \pi(\sigma, y_1)e^{tf(y_1)} \ldots \pi(y_{n-1}, y_n)e^{tf(y_n)} \\
&= \log \sum_{y_n=1}^{|\Omega|} (\Pi_t)^n(\sigma, y_n).
\end{aligned}$$

Omdat Π_t irreducibel is kunnen we deel (e) van stelling 5 gebruiken. Daarbij nemen we $\phi = (1, \ldots, 1)$:

$$F(t) = \lim_{n\to\infty} \frac{1}{n} F_n(t) = \lim_{n\to\infty} \frac{1}{n} \log \sum_{y_n=1}^{|\Omega|} (\Pi_t)^n(\sigma, y_n) = \log \rho(\Pi_t)$$

En omdat $|\Omega|$ eindig is, is $\rho(\Pi_t)$ (een oplossing van de karakteristieke functie van Π_t) positief, eindig en differentieerbaar met betrekking tot t. Met de Cramér truc zien we nu dat de stelling klopt. $\square$

6 Voorbeeld: 2x2 matrix

We bekijken in deze sectie een willekeurig voorbeeld met Markov-afhankelijke variabelen. De Markovketen bij dit voorbeeld is uiteraard aperiodiek en irreducibel. We bekijken de Markovketen met transitiematrix

$$\Pi = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

Volgens definitie 3 geldt dan:

$$\Pi_t = \left(\pi(i,j)e^{tf(j)}\right)_{i,j=1}^{2}$$

Dus

$$\Pi_t = \begin{bmatrix} \frac{1}{2}e^{tf(1)} & \frac{1}{2}e^{tf(2)} \\ \frac{1}{2}e^{tf(1)} & \frac{1}{2}e^{tf(2)} \end{bmatrix}$$

Nu we de matrix Π_t hebben gevonden kunnen de Perron-Frobenius eigenwaarde ervan uitrekenen. We berekenen de eigenwaarden van Π_t als volgt:

$$\begin{aligned} |\Pi_t - \lambda I| \quad &= (\frac{1}{2}e^{tf(1)} - \lambda)(\frac{1}{2}e^{tf(2)} - \lambda) - \frac{1}{2}e^{tf(1)} * \frac{1}{2}e^{tf(2)} \\ &= \frac{1}{4}e^{t(f(1)+f(2))} - \frac{1}{2}\lambda(e^{tf(1)} + e^{tf(2)}) + \lambda^2 - \frac{1}{4}e^{t(f(1)+f(2))} \\ &= \lambda(\lambda - \frac{1}{2}(e^{tf(1)} + e^{tf(2)}) = 0 \end{aligned}$$

Hieruit volgt:

$$\lambda_1 = 0 \quad \text{of} \quad \lambda_2 = \frac{1}{2}(e^{tf(1)} + e^{tf(2)})$$

Omdat $\lambda_2 > \lambda_1 = 0$, is λ_2 de Perron-Frobenius eigenwaarde:

$$\rho_t = \rho(\Pi_t) = \lambda_2 = \frac{1}{2}(e^{tf(1)} + e^{tf(2)})$$

$I(x)$ uit stelling 6 kunnen we nu schrijven als:

$$I(x) = \sup_{t\in\mathbb{R}}\left(tx - \log(\frac{1}{2}(e^{tf(1)} + e^{tf(2)}))\right) = \sup_{t\in\mathbb{R}}\left(tx + \log(2) - \log(e^{tf(1)} + e^{tf(2)})\right)$$

Deze uitdrukking van $I(x)$ kunnen we vervolgens verder uitwerken, door t te berekenen. Voor t^* waarvoor $tx - \log(\frac{1}{2}(e^{tf(1)} + e^{tf(2)}))$ zijn supremum aanneemt, moet gelden:

$$(t^*x - \log(\frac{1}{2}(e^{t^*f(1)} + e^{t^*f(2)})))' = 0$$

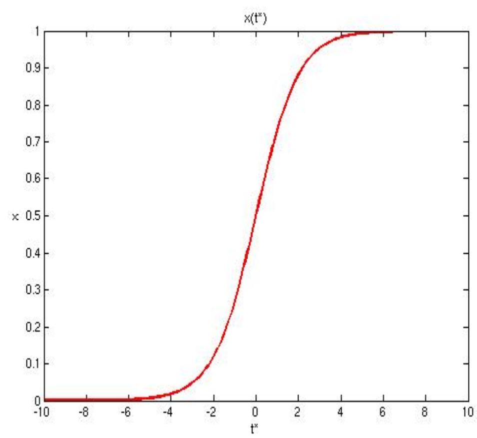
Dus:

$$\begin{aligned} x &= \frac{1}{\frac{1}{2}(e^{t^* f(1)} + e^{t^* f(2)})} \cdot \frac{1}{2}(f(1)e^{t^* f(1)} + f(2)e^{t^* f(2)}) \\ &= \frac{f(1)e^{t^* f(1)} + f(2)e^{t^* f(2)}}{e^{t^* f(1)} + e^{t^* f(2)}} \\ &= \frac{f(1) + f(2)e^{t^*(f(2)-f(1))}}{1 + e^{t^*(f(2)-f(1))}} \end{aligned}$$

Als $f(1) = 0$ en $f(2) = 1$, geldt

$$x = \frac{e^{t^*}}{1 + e^{t^*}}$$

Deze functie ziet er alsvolgt uit:



Figuur 1: x(t*) behorende bij de 2x2 matrix met f(1)=0 en f(2)=1

We substitueren nu $y = e^{t^*(f(2)-f(1))}$ en zien dan:

$$x = \frac{f(1) + f(2)y}{1 + y} \quad \Rightarrow \quad x(1 + y) = f(1) + f(2)y \quad \Rightarrow$$

$$y(x - f(2)) = f(1) - x \quad \Rightarrow \quad y = \frac{f(1) - x}{x - f(2)} \quad \Rightarrow$$

$$e^{t^*(f(2)-f(1))} = \frac{f(1) - x}{x - f(2)} \quad \Rightarrow \quad t^*(f(2)-f(1)) = \log \frac{f(1) - x}{x - f(2)} \quad \Rightarrow$$

$$t^* = \frac{1}{f(2) - f(1)} \log(\frac{f(1) - x}{x - f(2)})$$

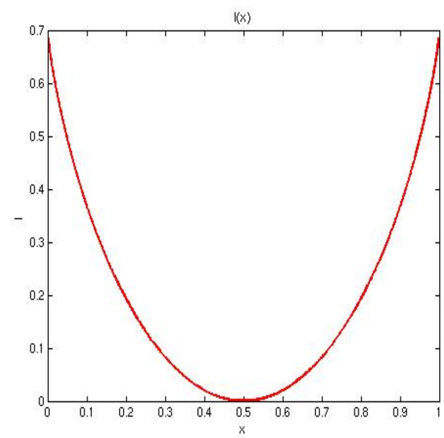
Invullen geeft:

$$I(x) = \frac{x}{f(2)-f(1)} \log(\frac{f(1)-x}{x-f(2)}) + \log(2) - \log(\frac{f(1)-x}{x-f(2)} e^{\frac{f(1)}{f(2)-f(1)}} + \frac{f(1)-x}{x-f(2)} e^{\frac{f(2)}{f(2)-f(1)}})$$

Als we weer $f(1) = 0$ en $f(2) = 1$ nemen, geldt

$$I(x) = x\log(\frac{x}{1-x}) + \log(2) - \log(1 + \frac{x}{1-x}) = x\log(\frac{x}{1-x}) + \log(2) + \log(1-x)$$

$I(x)$ is alleen gedefinieerd op $(0,1)$ en ziet er alsvolgt uit:



Figuur 2: I(x) behorende bij de 2x2 matrix met f(1)=0 en f(2)=1

7 Voorbeeld: 3x3 matrix

Dan gaan we nu kijken naar een iets uitgebreidere keten met 3 staten. Bij deze keten hoort dus een 3x3 transitiematrix. Deze is dus van de vorm:

$$\Pi = \begin{bmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{bmatrix}$$

Voor de rijsommen van Π moet natuurlijk gelden, dat deze gelijk zijn aan 1.
Eerst zullen we even kijken wat we nu precies gaan berekenen. Bij deze Markov keten met 3 staten nemen we:

$$f(y) = I(y = x_1)$$

met $I(.)$ de indicator-functie.
Dan geld:

$$S_n = \sum_{i=1}^{n} f(y_i) = \sum_{i=1}^{n} I(y_i = x_1) = \text{aantal bezoeken aan } x_1$$

Omdat

$$\mathbb{E}(S_n) \simeq \mu(x_1) \cdot n \qquad \text{en} \qquad \text{Var}(S_n) = \mathcal{O}(n)$$

kunnen we nu weer met de wet van de grote aantallen zeggen dat voor $n \to \infty$:

$$\frac{1}{n} S_n \to \mu(x_1)$$

Dus nu zien we

$$\mathbb{P}(|\frac{1}{n} S_n - \mu(x_1)| \geq \epsilon) \approx e^{-n(\inf_{|x-\mu(x_1)| \geq \epsilon} I(x))}$$

met

$$I(x) = \sup_t (tx - \log \rho(\Pi_t))$$

Als we nu kijken naar $f(y)$ zien we dat deze gelijk is aan 1 voor de eerste staat en 0 voor de tweede en derde staat. Π_t ziet er dus alsvolgt uit:

$$\Pi_t = \begin{bmatrix} P_{11}e^t & P_{12} & P_{13} \\ P_{21}e^t & P_{22} & P_{23} \\ P_{31}e^t & P_{32} & P_{33} \end{bmatrix}$$

Zoals we hebben gezien in voorbeeld 1 moeten we een aantal stappen nemen om de uiteindelijke $I(x)$ te kunnen berekenen. Dezelfde stappen zullen we bij een keten met 3 staten weer moeten nemen. Echter zouden deze stappen wel een stuk moeilijker kunnen zijn bij een 3x3 matrix.

Om dit rekenwerk wat makkelijker te maken, kunnen we bepaalde eisen stellen aan de transitiematrix. We kunnen deze bijvoorbeeld symmetrisch nemen. Ik ben met Mathematica aan de slag gegaan om te kijken bij welke transitiematrices, ik een Perron-Frobenius eigenwaarde kreeg, waar ik mee verder kon rekenen. Uiteindelijk kon ik pas echt goed verder werken bij een hele makkelijke matrix:

$$\Pi = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$$

Op deze manier is Π irreducibel en zijn de rijsommen 1.
Π_t ziet er dan alsvolgt uit:

$$\Pi_t = \begin{bmatrix} \frac{1}{3}e^t & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3}e^t & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3}e^t & \frac{1}{3} & \frac{1}{3} \end{bmatrix}$$

Van deze matrix heb ik met behulp van Mathemica de eigenwaarden berekend. Deze zijn:

$$\lambda_1 = 0, \qquad \lambda_2 = 0, \qquad \lambda_3 = \frac{1}{3}(2 + e^t)$$

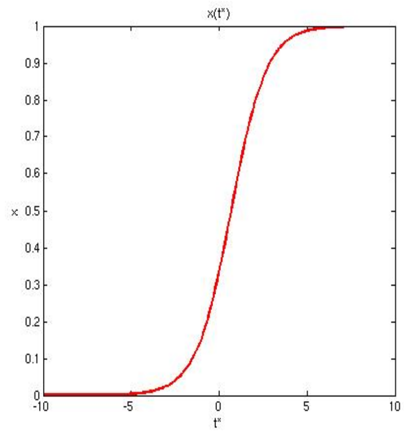
De Perron-Frobenius (grootste) eigenwaarde is dus duidelijke λ_3. Invullen in $I(x)$ geeft:

$$I(x) = \sup_t (tx - \log(\frac{1}{3}(2 + e^t))) = \sup_t (tx + \log(3) - \log(2 + e^t))$$

Om het supremum over t te berekenen leiden we weer af naar t en stellen dan gelijk aan 0. Dan komen we op

$$x = \frac{e^{t^*}}{2 + e^{t^*}}$$

Deze functie ziet er alsvolgt uit:



Figuur 3: x(t) behorende bij de eerste 3x3 matrix

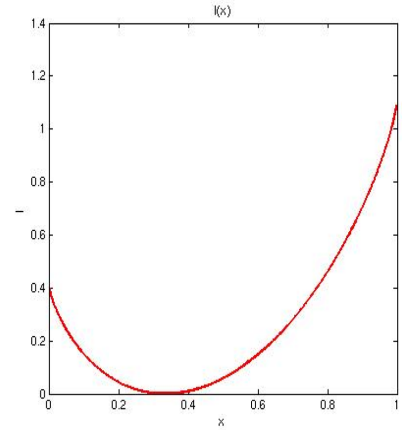
Vervolgens drukken we t^* uit in x:

$$(2 + e^{t^*})x = e^{t^*} \qquad \Rightarrow \qquad e^{t^*}(1-x) = 2x \qquad \Rightarrow \qquad t^* = \log(\frac{2x}{1-x})$$

Tot slot vullen we t^* in:

$$\begin{aligned} I(x) \qquad &= x\log(\frac{2x}{1-x}) - \log(\frac{1}{3}(2 + e^{\log(\frac{2x}{1-x})})) \\ &= x\log(\frac{2x}{1-x}) - \log(\frac{1}{3}(2 + \frac{2x}{1-x})) \\ &= x\log(x) + x\log(2) - x\log(1-x) + \log(3) - \log(\frac{2}{1-x}) \\ &= (1-x)\log(1-x) + x\log(x) + \log(3) - (1-x)\log(2) \end{aligned}$$

Deze functie ziet er alsvolgt uit:



Figuur 4: I(x) behorende bij de eerste 3x3 matrix

Als laatste bekijken we nog een symmetrische transititiematrix. Hiervoor kunnen we geen exacte oplossing vinden, maar kunnen we de oplossing wel numeriek benaderen.
We kiezen $P_{ii} = \frac{1}{2}$ voor $i = 1, 2, 3$ en $P_{ij} = \frac{1}{4}$ met $i \neq j$ en $i, j = 1, 2, 3$. Alle rijsommen zijn dan 1 en de overgangs matrix Π is irreducibel en symmetrisch. Π_t ziet er nu alsvolgt uit:

$$\Pi_t = \begin{bmatrix} \frac{1}{2}e^t & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4}e^t & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4}e^t & \frac{1}{4} & \frac{1}{2} \end{bmatrix}$$

We berekenen wederom de eigenwaarden met Mathematica. We krijgen dan de volgende eigenwaarden:

$$\lambda_1 = \frac{1}{4} \qquad \lambda_2 = \frac{1}{8}(3 + 2e^t - \sqrt{9 - 4e^t + 4e^{2t}}) \qquad \lambda_3 = \frac{1}{8}(3 + 2e^t + \sqrt{9 - 4e^t + 4e^{2t}})$$

Het is duidelijk dat λ_3 de grootste eigenwaarde is, dus

$$\rho(\Pi_t) = \frac{1}{8}(3 + 2e^t + \sqrt{9 - 4e^t + 4e^{2t}})$$

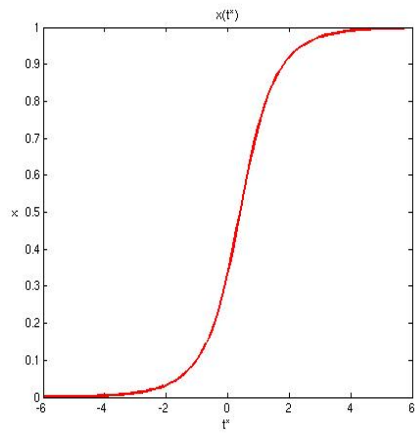
Voor $I(x)$ geldt dus

$$I(x) = \sup_t (tx + \log(8) - \log(3 + 2e^t + \sqrt{9 - 4e^t + 4e^{2t}}))$$

Door af te leiden naar t en gelijk te stellen aan 0, vinden we het supremum voor een of andere t^*. Dus

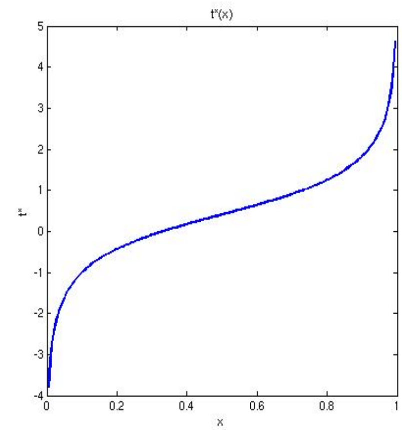
$$(t^*x + \log(8) - \log(3 + 2e^{t^*} + \sqrt{9 - 4e^{t^*} + 4e^{2t^*}}))' = 0 \qquad \Rightarrow$$
$$x = \frac{\lambda_3'}{\lambda_3} = \frac{2e^{t^*} + \frac{-4e^{t^*}+8e^{2t^*}}{2\sqrt{9-4e^{t^*}+4e^{2t^*}}}}{3 + 2e^{t^*} + \sqrt{9 - 4e^{t^*} + 4e^{2t^*}}}$$

Deze functie ziet er alsvolgt uit:



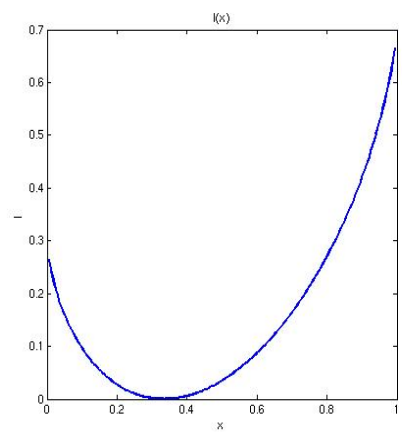
Figuur 5: x(t) behorende bij de tweede 3x3 matrix

Omdat we in dit voorbeeld niet via de algebraische weg t^* kunnen uitdrukken in x, zullen we dit numeriek gaan doen. Met behulp van het computerprogramma Matlab krijgen we dan de grafiek van t^* die te zien is in figuur 6 (de bijbehorende code staat in de bijlage).



Figuur 6: t*(x) behorende bij de tweede 3x3 matrix

Met deze numerieke berekende waarden van t^* hebben we vervolgens ook de grafiek van $I(x)$ kunnen maken:



Figuur 7: I(x) behorende bij de tweede 3x3 matrix

We zien dus dat, als we te maken hebben met variabelen uit een irreducibele Markovketen, we ook dan de Legendre getransformeerde functie $I(x)$ kunnen berekenen (expliciet of numeriek). Met deze functie $I(x)$ kunnen we het bewijs leveren dat de kans op een grote afwijking exponentieel snel naar 0 gaat:

$$\liminf_{n\to\infty} \mathbb{P}_\sigma^\pi(\frac{1}{n}S_n \approx x) = \liminf_{n\to\infty} e^{-n\cdot I(x)} \to 0$$

8 Conclusie

Voor mijn Bachelor-onderzoek ben ik gedoken in de theorie van de grote afwijkingen. Voordat ik met dit onderzoek begon, had ik nog nooit gehoord van een dergelijke theorie en wist ik nog niet eens wat er bedoeld werd met een "grote afwijking". Gedurende mijn onderzoek kreeg ik een steeds beter beeld van wat deze theorie precies inhield en kon ik ook steeds meer bedenken in welke richtingen nog verder gewerkt kan worden. Ik denk dan ook dat ik op dit moment slechts een zeer klein gedeelte van de wereld van de grote afwijkingen heb bestudeerd. Zo heb ik in de eerste hoofdstukken de aanname gedaan dat de variabelen onafhankelijk en identiek verdeeld waren. Deze aanname maakte het werk een stuk makkelijker. En ik denk dat ik op dit moment ook nog te weinig kennis van allerlei gebieden van de wiskunde heb, om te werken zonder deze aanname (of met een verzwakking van deze aanname). Op advies van mijn begeleider heb ik nog wel gekeken naar variabelen die een bepaalde Markov-afhankelijkheid hadden. Met deze verzwakte aanname heb ik met mijn huidige kennis van wiskunde nog wel wat onderzoek kunnen doen. Tot slot heb ik zelfs wat voorbeelden kunnen geven van dit onderzoek met betrekking tot Markov-afhankelijke variabelen.
Kortom, ik heb na deze scriptie enig beeld gekregen van wat de theorie van de grote afwijkingen inhoudt. Daarnaast kan ik een aantal richtingen aangeven waarin er op dit gebied onderzoek wordt gedaan of kan worden gedaan.

9 Referenties

- F.W. Redig, ”Handouts”, Leiden 2007-2009
- Amir Dembo en Ofer Zeitouni, ”Large Deviations Techniques and Applications”, 2^e editie
- S.R.S. Varadhan, ”Large Deviations and Applications
- Piet van Oostrum, ”Handleiding LaTeX”
- E. Seneta, ”Non-negative Matrices and Markov Chains”, 2^e editie

10 Bijlage

```
i=1;
for i=1:200
z(i)=i./201;
y=@(x)
(2.*exp(x)+(-4.*exp(x)+8.*exp(2.*x)) ./ (2.*sqrt(9-4.*exp(x)+4.*exp(2.*x)))))./(3+2.*exp(x)+sqrt(9-
4.*exp(x)+4.*exp(2.*x)))-z(i);
[w(i),fval]=fsolve(y,0);
x(i)=i./201;
p(i)=w(i).*x(i)+log(8)-log(3+2.*exp(w(i))+sqrt(9-4.*exp(w(i))+4.*exp(2.*w(i))));
i=i+1;
end
plot(x,w,'linewidth',2)
plot(x,p,'linewidth',2)
```