

Tilburg University

Een toepassing van importance sampling

Kleijnen, J.P.C.

Published in:
Statistica Neerlandica

Publication date:
1968

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):
Kleijnen, J. P. C. (1968). Een toepassing van importance sampling. *Statistica Neerlandica*, 22(3), 179-198.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Een toepassing van "importance sampling"

door J. P. C. KLEIJNEN

Summary This paper describes an experiment with "importance sampling", to show how much reduction of the computation time and sample size can be achieved in comparison with the usual Monte Carlo method. A comparison is made between each of the three methods of "importance sampling" and the usual Monte Carlo method by the determination of the expression

$$\xi(\lambda, v) = \int_v^{\infty} \frac{1}{x} \lambda e^{-\lambda x} dx.$$

Of the three methods A, B and C the first one uses the shifted exponential distribution, the second one uses the gamma distribution, and the third one uses the exponential distribution with modified parameter. These three methods have all smaller variances, ranges and sample sizes than the usual Monte Carlo method. Their order of preference is A, B, C. With respect to computing time only the method A is significantly better. So only the method A is an improvement in respect of both the sample size and the computing time.

1. Inleiding

"Importance sampling" is een methode die vrij lang geleden al ontwikkeld werd¹, maar waarvan zeer weinig toepassingen in de literatuur worden gegeven. Dit onderzoek werd opgezet met het doel een toepassing van deze methode te vinden en aan te tonen tot welke besparingen "importance sampling" kan leiden. Daartoe werd gekozen het probleem van de bepaling van

$$\xi(\lambda, v) = \int_v^{\infty} \frac{1}{x} \lambda e^{-\lambda x} dx,$$

welke integraal optreedt bij het vaststellen van de optimale bestelhoeveelheid in een bepaald voorraadmodel². Uit dat model volgde namelijk, dat de totale voorraadkosten minimaal zijn, indien voldaan is aan

$$\int_0^v f(x) dx + v \int_v^{\infty} \frac{1}{x} f(x) dx = \frac{C_2}{C_1 + C_2} \quad (1)$$

waarin

- v = beginvoorraad, na ontvangst van de bestelde hoeveelheid,
- x = vraag per tijdseenheid,

Notitie nr. 3 van het Rekencentrum van de Katholieke Hogeschool, Tilburg.

Veel dank ben ik verschuldigd aan Prof. Drs. J. KRIENS voor het kritisch doorlezen van dit artikel en aan Ir. D. NEELEMAN en A. J. VAN REEKEN voor hun medewerking bij het oplossen van enige statistische en wiskundige problemen. Uiteraard blijft de schrijver verantwoordelijk voor de inhoud van dit artikel.

¹ H. KAHN en A. W. MARSHALL: Methods of reducing sample size in Monte Carlo computations. (J.O.R.S.A., I, 1953).

² Leergang besliskunde: Deel 5, Inleiding tot de Mathematische Besliskunde, door J. KRIENS en G. DE LEVE, Mathematisch Centrum, Amsterdam, 1966.

$f(x)$ = kansdichtheid van de vraag per tijdseenheid,
 C_1 = voorraadkosten per stuk per tijdseenheid,
 C_2 = naleveringskosten per stuk per tijdseenheid.

Indien $f(x)$ een exponentiële verdelingsdichtheid voorstelt moet in (1) worden bepaald

$$\int_v^{\infty} \frac{1}{x} \lambda e^{-\lambda x} dx = \xi(\lambda, v). \quad 1$$

Deze integraal is moeilijk te berekenen, aangezien partiële integratie en reeksontwikkeling niet zonder meer kunnen worden toegepast². Door middel van de Monte Carlo methode kan $\xi(\lambda, v)$ evenwel op een eenvoudige manier geschat worden. Daartoe loten wij x uit de exponentiële verdelingsdichtheid $\lambda e^{-\lambda x}$ en substitueren de gelote x in

$$g(x) = 0 \text{ als } x < v$$

$$g(x) = \frac{1}{x} \text{ als } x \geq v.$$

Na n lotingen kan $\xi(\lambda, v)$ dan geschat worden door

$$\hat{\xi} = \frac{1}{n} \sum_{i=1}^n g(x_i).$$

Naarmate n groter is, zal de kans op een juiste schatting van $\xi(\lambda, v)$ toenemen.

In dit artikel zal aangegeven worden, hoe door "importance sampling" de gewenste nauwkeurigheid van de schatter van $\xi(\lambda, v)$ met aanzienlijk minder lotingen bereikt kan worden. Reeds nu willen wij erop wijzen, dat in paragraaf 4 zal blijken dat deze besparingen een omvang bereiken (99%) zoals die in de ons bekende literatuur nergens genoemd worden. In paragraaf 2 worden de principes van "importance sampling" gegeven, zoals die door H. KAHN en A. W. MARSHALL zijn afgeleid, waarna in paragraaf 3 uiteengezet wordt, hoe deze principes bij de schatting van $\xi(\lambda, v)$ kunnen worden toegepast.

2. De principes van "importance sampling"

Voor de integraal

$$\xi = \int_{-\infty}^{\infty} g(x)f(x) dx$$

¹ λ en v moeten positief zijn, aangezien anders deze integraal divergeert. $v \leq 0$ geeft immers $\xi(\lambda, 0) = \lambda \Gamma(0)$ en $\Gamma(0)$ divergeert.

² Dr. L. LIPS, lector voor de wiskunde aan de Katholieke Hogeschool, leidde op het einde van dit onderzoek af dat

$$\xi(\lambda, v) = \lambda \left[-c - \ln(\lambda v) + \sum_{i=1}^{\infty} \left\{ (-1)^{i+1} \frac{(\lambda v)^i}{i! i} \right\} \right]$$

waarin c de constante van EULER voorstelt.

waarbij $f(x)$ een verdelingsdichtheid voorstelt, kan ook worden geschreven

$$\xi = \int_{-\infty}^{\infty} \frac{g(x)f(x)}{h(x)} h(x) dx.$$

Indien $h(x)$ een verdelingsdichtheid van x is, kan ξ ook worden geschat door x te loten uit $h(x)$ en te substitueren in

$$g^*(x_i) = \frac{g(x_i)f(x_i)}{h(x_i)}. \quad (2)$$

Na n lotingen uit $h(x)$ kan ξ dan geschat worden door

$$\xi^* = \frac{1}{n} \sum_{i=1}^n g^*(x_i). \quad (3)$$

KAHN en MARSHALL hebben bewezen, dat de variantie van ξ^* minimaal zal zijn, indien $h(x)$ gelijk is aan

$$h_0(x) = \frac{|g(x)f(x)|}{\int_{-\infty}^{\infty} |g(x)f(x)| dx}. \quad (4)$$

Als $g(x) \geq 0$ voor $-\infty < x < \infty$ kan voor (4) worden geschreven

$$h_0(x) = \frac{g(x)f(x)}{\int_{-\infty}^{\infty} g(x)f(x) dx} = \frac{g(x)f(x)}{\xi} \quad (5)$$

en kan worden bewezen, dat de variantie van ξ^* zelfs 0 wordt. Het is evenwel duidelijk, dat wij uit (5) toch niet $h_0(x)$ kunnen bepalen, doordat in (5) ook ξ voorkomt, en die ξ moeten wij juist bepalen.

In de volgende paragraaf zullen wij proberen $h_0(x)$ te benaderen door rekening te houden met (5) én met het feit dat de variantie van ξ^* bij die $h_0(x)$ tot nihil daalt.

3. Enige benaderingen van $h_0(x)$ bij het schatten van $\xi(\lambda, \nu)$

Vergelijking van

$$\xi(\lambda, \nu) = \int_{\nu}^{\infty} \frac{1}{x} \lambda e^{-\lambda x} dx$$

met de in paragraaf 2 gegeven

$$\xi = \int_{-\infty}^{\infty} g(x)f(x) dx$$

toont dat in het beschouwde probleem geldt

$$g(x) = 0 \text{ als } x < v$$

$$g(x) = \frac{1}{x} \text{ als } x \geq v$$

en

$$f(x) = \lambda e^{-\lambda x} \text{ als } x \geq 0$$

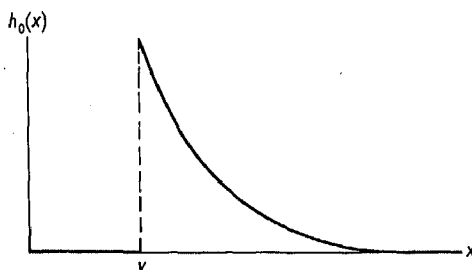
$$f(x) = 0 \text{ als } x < 0.$$

Uit (5) volgt dan de optimale $h(x)$ in dit probleem:

$$h_0(x) = 0 \text{ als } x < 0.$$

$$h_0(x) = \frac{1}{x} \lambda e^{-\lambda x} \frac{1}{\xi} \text{ als } x \geq v.$$

De grafische voorstelling van $h_0(x)$ heeft dus ongeveer de volgende vorm:



Wij zullen nu achtereenvolgens 3 benaderingen van deze $h_0(x)$ behandelen¹.

A. De verschoven exponentiële verdelingsdichtheid

Wij loten eerst x' uit de exponentiële verdeling met de parameter c (die verderop bepaald wordt) en verhogen de geloten x' met de constante v . Dan heeft $x = x' + v$ de volgende kansdichtheid:

$$h_1(x) = ce^{-c(x-v)} = ce^{cv} e^{-cx} \text{ als } x \geq v$$

$$h_1(x) = 0 \text{ als } x < v.$$

Uit (2) volgt dan

$$g_1^*(x) = \frac{1}{x} \lambda e^{-\lambda x} c^{-1} e^{-cv} e^{cx} \text{ als } x \geq v.$$

¹ Een * geeft aan dat de schatter een „importance sampling“-schatter is. De benedenindices 1, 2 en 3 geven aan dat het de eerste, tweede resp. derde benadering is van de optimale verdeling $h_0(x)$ en van de daarmee corresponderende schatter.

Zodat:

$$g_1^*(x) = \frac{\lambda}{c} e^{-cv} \frac{1}{x} e^{-(\lambda-c)x}$$

en uit (3) volgt:

$$\xi_1^* = \frac{1}{n} \sum_{i=1}^n g_1^*(x_i).$$

Wij proberen nu de waarde van c te bepalen. Indien wij c zo wensen te kiezen dat de variantie van ξ_1^* minimaal is, moet de variantie van $g_1^*(x)$ geminimaliseerd worden. In appendix I.3 is echter aangetoond dat deze optimale c niet te berekenen is, doordat bij die berekening een zelfde soort integraal optreedt als $\xi(\lambda, v)$.

Dit probleem is nu als volgt benaderd. In de vorige paragraaf is vermeld dat de variantie van ξ_0^* (en dus van $g_0^*(x)$) nihil is, mits $g(x) \geq 0$ voor $-\infty < x < \infty$. De spreidingsbreedte van $g_0^*(x)$ is dan ook nihil. Omdat wij de parameter c van $h_1(x)$ niet zo kunnen bepalen dat de variantie van $g_1^*(x)$ minimaal wordt, kiezen wij c nu zo, dat de spreidingsbreedte van $g_1^*(x)$ geminimaliseerd wordt. Indien $c > \lambda$ zou de maximale waarde van $g_1^*(x)$ oneindig groot zijn, indien namelijk x oneindig groot is. Derhalve kunnen wij al meteen stellen dat $c \leq \lambda$ opdat de spreidingsbreedte eindig blijft. Voor elke waarde van c in het interval $(0, \lambda]$ geldt dan, dat de minimale waarde van $g_1^*(x)$ 0 bedraagt, indien namelijk $x = \infty$. Teneinde de spreidingsbreedte te minimaliseren moet dus de maximale waarde van $g_1^*(x)$ geminimaliseerd worden, door een juiste keuze van c in het gebied $(0, \lambda]$. Aangezien voor $0 < c \leq \lambda$ $g_1^*(x)$ een afnemende functie van x is, wordt die maximale waarde van $g_1^*(x)$ bereikt voor $x = v$. Dus

$$g_1^*(x)_{\max} = g_1^*(v) = \frac{\lambda}{c} e^{-cv} \frac{1}{v} e^{-(\lambda-c)v} = \frac{\lambda}{c} \frac{1}{v} e^{-\lambda v}.$$

Hieruit blijkt dat $g_1^*(x)_{\max}$ geminimaliseerd wordt door een zo groot mogelijk c te kiezen, in het gebied $(0, \lambda]$, met andere woorden de spreidingsbreedte van $g_1^*(x)$ is minimaal als $c = \lambda$. Hieruit volgt:

$$h_1(x) = \lambda e^{\lambda v} e^{-\lambda x} \text{ als } x \geq v$$

$$h_1(x) = 0 \quad \text{als } x < v.$$

Wij loten derhalve uit de oude kansdichtheid $\lambda e^{-\lambda x}$ maar verhogen de gelote waarde met de constante v . Deze nieuwe x wordt gesubstitueerd in:

$$g_1^*(x) = e^{-\lambda v} \frac{1}{x}.$$

Tot slot van deze benadering vergelijken wij $g_1^*(x)$ met $g(x)$ van de gewone Monte Carlo-methode (zie paragraaf 1):

¹ $x < v$ kan niet uit $h_1(x)$ geloot worden.

- a. $g_1^*(x)$ en $g(x)$ hebben dezelfde minimale waarde, namelijk 0 voor $x = \infty$, zodat de spreidingsbreedte van $g_1^*(x)$ kleiner zal zijn dan die van $g(x)$, indien $g_1^*(x)_{\max} < g(x)_{\max}$. Dat aan deze ongelijkheid voldaan is, volgt onmiddellijk uit

$$g_1^*(x)_{\max} = g_1^*(v) = e^{-\lambda v} \frac{1}{v}$$

en

$$g(x)_{\max} = g(v) = \frac{1}{v}$$

- b. Het bleek niet mogelijk c zo te kiezen dat de variantie van $g_1^*(x)$ minimaal wordt. Het is echter wel mogelijk te bewijzen dat $c = \lambda$ (welke keuze een minimale spreidingsbreedte van $g_1^*(x)$ geeft) tot een variantie van $g_1^*(x)$ leidt die in elk geval kleiner is dan de variantie van $g(x)$. Voor dit bewijs verwijzen wij naar appendix I.4.

B. De Gammaverdeling

De figuur in het begin van paragraaf 3 suggereert dat wij $h_0(x)$ ook kunnen benaderen door middel van een continue verdelingsdichtheid waarvan de modus links van de verwachting ligt. De meeste verdelingsdichtheden blijken echter óf diskreet óf symmetrisch te zijn; de gammaverdeling

$$\frac{b^p}{\Gamma(p)} x^{p-1} e^{-bx} \quad (x, b, p > 0)$$

voldoet evenwel aan de gestelde voorwaarden.

Net als onder A moeten wij na de keuze van de soort verdelingsdichtheid nog de waarde van de parameter(s) bepalen. Dit gebeurt als volgt.

Het loten uit een gammaverdeling is een eenvoudige zaak, mits p een geheel getal is¹. Zoals bekend is neemt de asymmetrie toe, indien de waarde van p afneemt². Verder blijkt uit de formule $[b^p/\Gamma(p)]x^{p-1}e^{-bx}$ dat deze verdelingsdichtheid een continue dalende functie is van x , indien $p \leq 1$. Samenvattend:

- $p > 1$ (opdat de verdeling een „top” heeft)
- p zo klein mogelijk (opdat de asymmetrie groot is)
- p geheel (opdat het loten eenvoudig verloopt).

Daaruit volgt zonder meer: $p = 2$.

Wij kunnen nu dus zeggen: loot x uit $h_2(x) = b^2 x e^{-bx}$ en substitueer de gelote x in

$$g_2^*(x) = 0 \quad \text{indien } 0 < x < v \quad ^3$$

$$g_2^*(x) = \frac{x^{-1} \lambda e^{-\lambda x}}{b^2 x e^{-bx}} = \frac{\lambda}{b^2} x^{-2} e^{-(\lambda-b)x} \quad \text{indien } x \geq v.$$

¹ Door een aantal trekkingen uit een exponentiële verdeling te sommeren, ontstaat nl. een gamma-verdeelde grootheid.

² Voor de gammaverdeling is af te leiden dat de scheefheidscoëfficiënt $\gamma = \mu_3/\sigma_3$ gelijk is aan $2/\sqrt{p}$.

³ Dan geldt immers $g(x) = 0$, zodat $g(x) \cdot f(x)/h_2(x) = 0$.

Omdat niet berekend kan worden, welke waarde van b de minimale variantie van $g_2^*(x)$ geeft (zie appendix II), is b bepaald door de spreidingsbreedte te minimaliseren. Dit gebeurt op analoge wijze als onder A: $b \leq \lambda$ opdat de spreidingsbreedte eindig blijft. Dus:

$$g_2^*(x)_{\max} = g_2^*(v) = \frac{\lambda}{b^2} v^{-2} e^{-(\lambda-b)v}.$$

Zodat:

$$\text{minimale } g_2^*(x)_{\max} \text{ als } \frac{d[g_2^*(x)_{\max}]}{db} = 0.$$

Het is eenvoudig aan te tonen, dat hieraan voldaan is door $b = 2/v$.

x wordt dus geloot uit $h_2(x) = (4/v^2)xe^{-2/v(x)}$, d.w.z. uit een gammaverdeling met $p = 2$ en $b = 2/v$, zodat de modus ligt bij $x = \frac{1}{2}v$ en de verwachting bij $x = v$.

De gelote x wordt gesubstitueerd in:

$$g_2^*(x) = 0 \quad \text{als } 0 < x < v$$

$$g_2^*(x) = \frac{\lambda}{4} v^2 x^{-2} e^{-(\lambda-2/v)x} \quad \text{als } x \geq v$$

en na n lotingen wordt berekend:

$$\xi_2^* = \frac{1}{n} \sum_{i=1}^n g_2^*(x_i).$$

Tot slot van de benadering via de gammaverdeling kunnen wij nog enige opmerkingen maken;

- a. $b = 2/v$ moet ook voldoen aan de voorwaarde $b \leq \lambda$ opdat de spreidingsbreedte van $g_2^*(x)$ eindig blijft. Daaruit volgt $v \geq 2(1/\lambda)$ hetgeen betekent dat de beginvoorraad zo hoog moet zijn dat de *neenverkopen* minder dan 13,5% bedragen. Indien wij zulke grote *neenverkopen* niet a priori uitsluiten, kunnen wij een gammaverdeling met $b < 2/v$ gebruiken.

De variantie van $g_2^*(x)$ bleek echter ook kleiner te zijn dan de variantie van $g(x)$ indien

$$v < 2(1/\lambda) \text{ en toch } b = 2/v \text{ gebruikt wordt. }^1$$

$$v > 2(1/\lambda) \text{ en toch } b < 2/v \text{ gebruikt wordt. }^2$$

Wij moeten hierbij bedenken, dat de afleiding van b geschiedde onder de voorwaarde van een minimale spreidingsbreedte, terwijl het werkelijke criterium de variantie is. De gammaverdeling blijkt dus bij meerdere waarden van b tot een kleinere variantie van de schatter te leiden, vergeleken met de gewone Monte-Carlo-methode. Bepaling

¹ In de testgevallen waarbij $v < 2(1/\lambda)$ bleek $b < 2/v$ inderdaad tot een kleinere variantie van $g_2^*(x)$ te leiden dan $b = 2/v$. Dit resultaat stond niet van tevoren vast, aangezien $b = 2/v$ en $b \leq \lambda$ zijn afgeleid uit de minimalisatie van de spreidingsbreedte.

² Indien $v > 2(1/\lambda)$ geeft $b = 2/v$ een kleinere variantie dan $b < 2/v$.

van b zó, dat een minimale spreidingsbreedte bereikt wordt, blijkt een goede leidraad te zijn (zie ook de voetnoten.)

- b. De spreidingsbreedte van $g_2^*(x)$ blijkt kleiner te zijn dan die van $g(x)$. Dit is eenvoudig te bewijzen uit

$$g_2^*(x)_{\max} = g_2^*(v) = \frac{\lambda}{4} e^{-\lambda v} e^2$$

en

$$g(x)_{\max} = g(v) = \frac{1}{v}.$$

De spreidingsbreedte van $g_2^*(x)$ is echter groter dan die van $g_1^*(x)$ (mits $v > 0,54(1/\lambda)$ aan welke voorwaarde in een realistisch voorraadmodel voldaan zal worden). Dit is eenvoudig te bewijzen uit $g_2^*(x)_{\max}$ en uit $g_1^*(x)_{\max} = (1/v)e^{-\lambda v}$.

Vergelijking van de variantie van $g_2^*(x)$ met die van $g(x)$ en $g_1^*(x)$ blijkt niet mogelijk op de manier zoals dat in appendix I wel voor $g_1^*(x)$ lukte. Wel tonen alle testgevallen duidelijk dat net als bij de spreidingsbreedte geldt:

$$\text{var}[g_2^*(x)] < \text{var}[g(x)]$$

en

$$\text{var}[g_2^*(x)] > \text{var}[g_1^*(x)].$$

C. De exponentiële verdelingsdichtheid met gewijzigde parameter

KAHN en MARSHALL wijzen erop dat eenvoudshalve de optimale $h(x)$ benaderd kan worden, doordat $f(x|\theta)$ vervangen wordt door $f(x|\theta')$ waarbij θ en θ' de vector van parameters voorstellen.¹ Met andere woorden:

$$h_3(x) = \lambda' e^{-\lambda' x} \quad \text{indien } x \geq 0$$

$$h_3(x) = 0 \quad \text{indien } x < 0.$$

$$g_3^*(x) = \frac{x^{-1} \lambda e^{-\lambda x}}{\lambda' e^{-\lambda' x}} = \frac{1}{x} \frac{\lambda}{\lambda'} e^{-(\lambda - \lambda')x} \quad \text{indien } x > v$$

$$= 0 \quad \text{indien } 0 \leq x < v.$$

Net als onder A en B kan de waarde van λ' die $\text{var}[g_3^*(x)]$ minimaliseert niet worden berekend, zodat wij λ' zó bepalen dat de spreidingsbreedte van $g_3^*(x)$ geminimaliseerd wordt: $\lambda' \leq \lambda$ opdat de spreidingsbreedte eindig blijft.

$$g_3^*(x)_{\max} = g_3^*(v) = \frac{1}{v} \frac{\lambda}{\lambda'} e^{-(\lambda - \lambda')v}$$

¹ Dit is eigenlijk al onder B gebeurd: De gammaverdeling met $p = 1$ en $b = \lambda$ (d.w.z. de exponentiële verdelingsdichtheid met parameter λ) is daar vervangen door de gammaverdeling met $p = 2$ en $b = 2/v$.

zodat $\frac{d[g_3^*(x)_{\max}]}{d\lambda'} = 0$ geeft $\lambda' = \frac{1}{v}$.¹

Deze benadering houdt derhalve in: loot x uit

$$h_3(x) = \frac{1}{v} e^{-(1/v)x} \quad \text{indien } x \geq 0$$

$$h_3(x) = 0 \quad \text{indien } x < 0$$

en substitueer de gelote x in

$$g_3^*(x) = \frac{\lambda v}{x} e^{-(\lambda - 1/v)x} \quad \text{indien } x \geq v$$

$$= 0 \quad \text{indien } 0 \leq x < v.$$

Tot slot van deze benadering kan nog opgemerkt worden:

- Ook een waarde van λ' die afwijkt van $1/v$ geeft nog besparingen ten opzichte van de gewone Monte Carlo-methode². Deze besparingen zijn evenwel kleiner dan in het geval $\lambda' = 1/v$.
- De spreidingsbreedte van $g_3^*(x)$ is kleiner dan die van $g(x)$ ¹. Zij is echter groter dan de spreidingsbreedte van $g_2^*(x)$ en van $g_1^*(x)$. Vergelijking van de varianties, zoals dat in appendix I geschiedde, is niet mogelijk. Wel bleek in elk van de testgevallen² dat

$$\text{var}[g_1^*(x)] < \text{var}[g_2^*(x)] < \text{var}[g_3^*(x)] < \text{var}[g(x)].$$

4. Invloed van "importance sampling" op het aantal lotingen

In de vorige paragraaf bleek dat het mogelijk is aan te tonen, dat de variantie van de schatter bij het loten uit de verschoven exponentiële verdeling kleiner is dan de variantie van de gewone Monte Carlo-schatter. Bij de beide andere benaderingsmethoden van $h_0(x)$ bleek een uitspraak over de verhouding van de varianties niet mogelijk. Wel kon bewezen worden, dat de spreidingsbreedte van beide schatters kleiner is dan die van de gewone Monte Carlo-schatter. Natuurlijk kan experimenteel worden nagegaan, hoe de varianties van de diverse schatters zich verhouden. Nog interessanter is de vraag: Hoe beïnvloeden de verschillen in de varianties het aantal lotingen dat nodig is om $\xi(\lambda, v)$ met een bepaalde betrouwbaarheid en nauwkeurigheid te schatten?

Indien wij b.v. met 95% zekerheid wensen te zeggen, dat onze schatting van $\xi(\lambda, v)$ hoogstens 10% fout is, kan voor het gewenste aantal lotingen op de volgende manier een formule worden afgeleid³.

¹ Als $v = 1/\lambda$ dan $\lambda' = 1/v = \lambda$ zodat deze benadering hetzelfde wordt als de gewone Monte Carlo-methode. In een realistisch model zal evenwel gelden: $v > 1/\lambda$, zodat $\lambda' = 1/v < \lambda$.

² Mits $v > 1/\lambda$, hetgeen een weinig stringente voorwaarde is.

³ In deze paragraaf kan $g(x)$ ook een van de in paragraaf 3 besproken vormen hebben: $g_1^*(x)$, $g_2^*(x)$, $g_3^*(x)$. ξ wordt dan vervangen door ξ_1^* resp. ξ_2^* resp. ξ_3^* .

Uit de definitie van de schatter,

$$\hat{\xi} = \frac{1}{n} \sum_{i=1}^n g(x_i),$$

volgt dat

$$\text{var}(\hat{\xi}) = \frac{1}{n} \text{var}[g(x)].$$

Indien n niet te klein is, heeft $\hat{\xi}$ bij benadering een normale verdeling $N(\xi, \sigma)$ waarbij $\xi = \xi(\lambda, \nu)$ en

$$\sigma = \sqrt{\frac{\text{var}[g(x)]}{n}}.$$

Derhalve geldt:

$$P(|\hat{\xi} - \xi| < 0,10\xi) = 0,95$$

indien $\frac{0,10\xi}{\sigma} = 1,96.$

Hieruit volgt onmiddellijk:

$$\frac{0,10\xi}{\sqrt{\text{var} g(x)}} \sqrt{n} = 1,96$$

zodat

$$n = \left[\frac{1,96}{0,10} \right]^2 \frac{\text{var}[g(x)]}{\xi^2}.$$

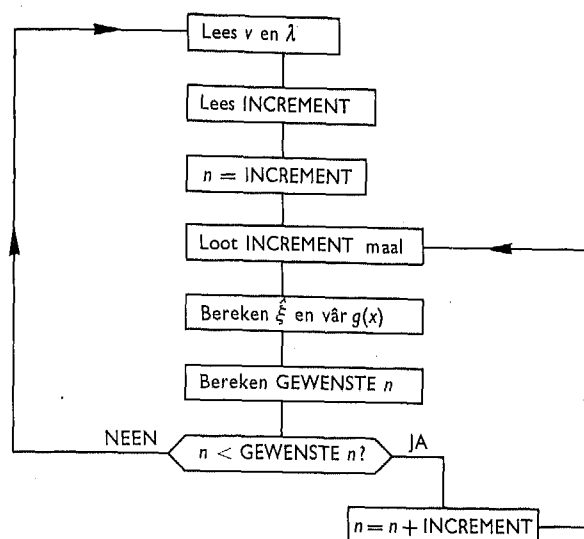
Of algemeen geformuleerd: Als wij met een betrouwbaarheid van $100 \cdot (1 - \alpha)\%$ een schatting van $\xi(\lambda, \nu)$ wensen te maken die hoogstens $\gamma\%$ fout is, dan moet n voldoen aan

$$n = \left[\frac{t_\alpha}{\gamma/100} \right]^2 \frac{\text{var}[g(x)]}{\xi^2}.$$

Hieruit volgt dat een daling van de variantie van de schatter met een bepaald percentage, het gewenste aantal lotingen met hetzelfde percentage doet afnemen.

In dit onderzoek is nu voor diverse waarden van λ en ν nagegaan, wat de grootte van n moet zijn om de gewenste betrouwbaarheid en nauwkeurigheid te bereiken. Omdat evenwel in de zojuist afgeleide formule voor n de grootheden ξ en $\text{var}[g(x)]$ voorkomen die wij niet kennen, is eerst op basis van een steekproef ξ en $\text{var}[g(x)]$ geschat. Dit leidt tot een eerste schatting van het gewenste aantal lotingen. Omdat de steekproef-grootheden $\text{var}[g(x)]$ en ξ evenwel tot een verkeerde waarde van de gewenste n kunnen leiden, is het gewenste aantal lotingen niet meteen gelijk gesteld aan de eerste schatting van n . Het aantal lotingen wordt daarentegen steeds met b.v. 50 waarnemingen (het zgn. increment) vergroot; op basis van de steeds groter wordende steekproef wordt

vervolgens een nieuwe gewenste n berekend, en de benadering wordt beëindigd als de steekproefomvang de waarde van de gewenste n heeft bereikt. Een en ander blijkt ook uit het volgende schema.



Uit de testgevallen bleek dat deze pragmatische aanpak inderdaad de gewenste betrouwbaarheid en nauwkeurigheid oplevert. (Zie ook pagina 193).

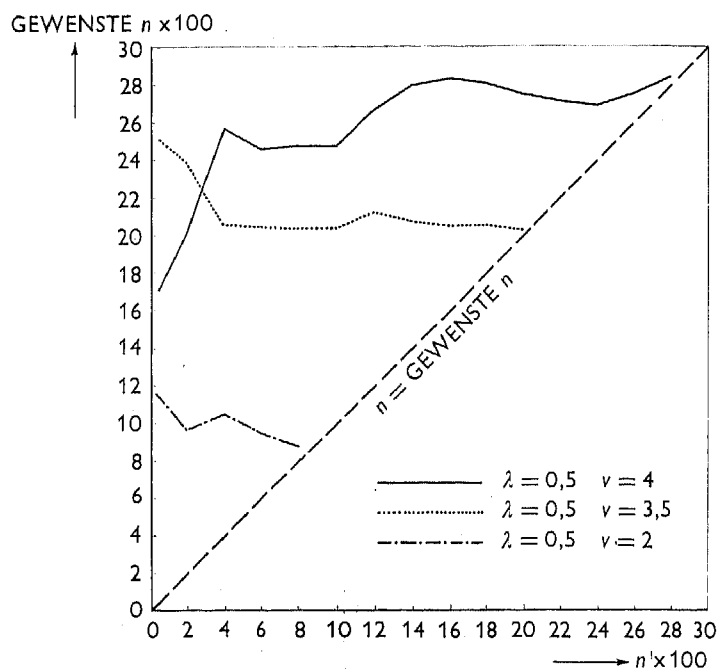
Ter illustratie wordt in grafiek 1 het verloop gegeven van het gewenste aantal lotingen bij toepassing van de gewone Monte Carlo-methode in 3 testgevallen.

Tabel 1. Gewenste aantal lotingen ¹

λ	ν	Gewone Monte Carlo		"Importance sampling" benadering					
		aantal	%	A		B		C	
				aantal	%	aantal	%	aantal	%
0,5	4	2459	100,0	19	0,8	772	31,4	1362	55,4
0,5	3,5	1779	100,0	26	1,5	699	39,3	1217	68,4
0,5	2	812	100,0	40	4,9	523	64,4	812	100,0
0,25	8,5	2844	100,0	17	0,6	828	29,1	1438	50,6

Bij tabel 1. Gegeven is het gewenste aantal lotingen nadat het werkelijke aantal het gewenste aantal lotingen overtreft. In de %-kolommen staan de gewenste aantallen uitgedrukt als een percentage van het gewenste aantal lotingen bij de gewone Monte Carlo-methode. Het gewenste aantal lotingen is een gemiddelde. Bij $\lambda = 0,5$ en $\nu = 4$ resp. 3,5 is dat gemiddelde berekend uit 3 waarden (verkregen door verschillende reeksen aselechte getallen te gebruiken); bij $\lambda = 0,5$ en $\nu = 2$ uit 2 waarden, en bij $\lambda = 0,25$ en $\nu = 8,5$ uit 1 waarde.

¹ Bij elk van deze testgevallen is een andere reeks aselechte getallen gebruikt.



Grafiek 1.

Zoals al eerder werd opgemerkt zijn de gevallen waarin $v = 1/\lambda$ minder relevant voor het trekken van conclusies, aangezien de neenverkoop dan erg groot is. In de tabel is een aparte kolom voor λ én voor v opgenomen. Het is evenwel te bewijzen dat het gewenste aantal lotingen hetzelfde verloop heeft indien $\lambda \times v$ hetzelfde is. Wel hangt ξ af van de waarde van λ . Uit de tabel blijkt duidelijk, dat de verschoven exponentiële verdelingsdichtheid enorme besparingen oplevert. Zoals door het vergelijken in paragraaf 3 van de spreidingsbreedten al gesuggereerd werd, leveren echter ook de beide andere benaderingen aanzienlijke besparingen op.

Naast het aantal lotingen is relevant de benodigde *machinetijd*. Deze besparingen zullen iets kleiner zijn, omdat bij elk van de "importance sampling"-methoden de vorm van $g^*(x_i)$ ingewikkelder is dan die van $g(x_i)$, zodat de computer meer moet rekenen. Bovendien eist de benadering via de gammaverdeling (methode B) dat wij per waarneming x_i twee lotingen doen uit de exponentiële verdeling. Een en ander blijkt uit de volgende tabel waarbij wel moet worden opgemerkt, dat de machinetijden slechts ruw benaderd zijn ¹.

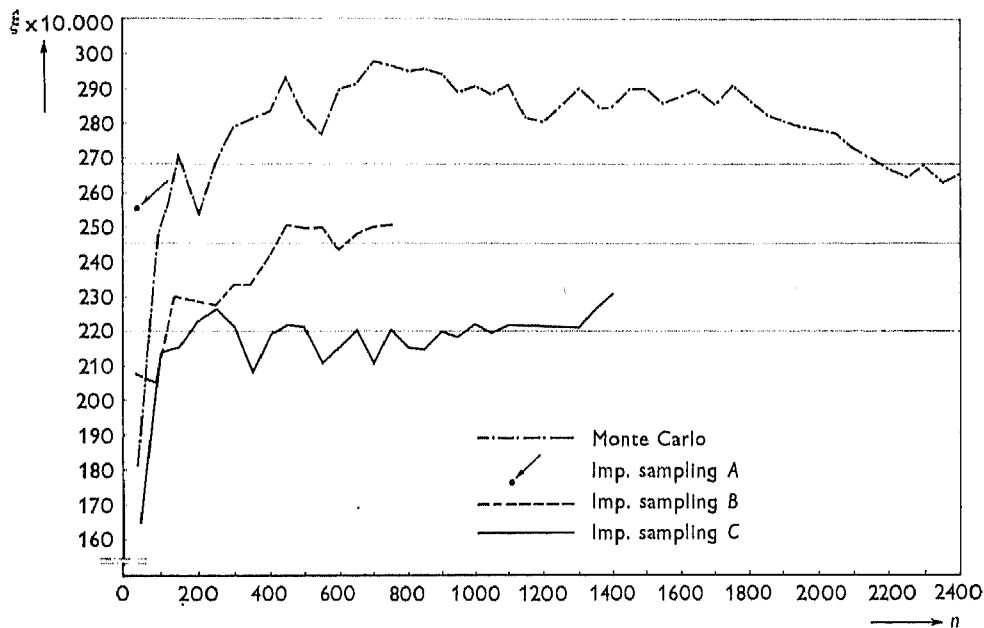
In grafiek 2 wordt aangegeven, hoe ξ verloopt naarmate het aantal lotingen toeneemt (totdat het gewenste aantal is bereikt). In de grafiek is bovendien door horizontale lijnen $\xi(\lambda, v)$ en 110% resp. 90% van $\xi(\lambda, v)$ aangegeven. De waarde van $\xi(\lambda, v)$ is

¹ De snelheid waarmee de operator telkens tussentijden intypt, speelt mee. Bovendien zijn de tijden afgerond en is per λ, v slechts één steekproef genomen.

Tabel 2. Benodigde machinetijd

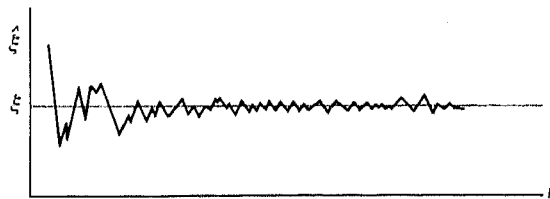
λ	ν	Gewone Monte Carlo-methode		"Importance sampling"					
				A		B		C	
		ged. van 1 uur	%	ged. van 1 uur	%	ged. van 1 uur	%	ged. van 1 uur	%
0,5	4	0,35	100,0	0,01	2,9	0,23	65,7	0,28	80,0
0,5	3,5	0,26	100,0	0,01	3,9	0,24	92,3	0,27	103,9
0,5	2	0,12	100,0	0,02	16,7	0,19	158,3	0,14	116,7
0,25	8,5	0,40	100,0	0,02	5,0	0,28	70,0	0,31	77,5

berekend met behulp van de formule voor de analytische benadering die in de tweede voetnoot op pagina 180 is gegeven. In de overige gevallen toonden de grafische voorstellingen een soortgelijk verloop. Dit verloop geeft aanleiding tot de volgende opmerkingen.



Grafiek 2. Schatting van $\xi = \int_0^{\infty} \frac{x}{\lambda} \cdot e^{-\lambda x} dx$ op basis van waarnemingen ($\lambda = 0.5$; $\nu = 4.0$).

a. Op het eerste gezicht lijkt het misschien vreemd, dat de grafieken er niet als volgt uitzien:



Zo'n verloop zou echter eerder ontstaan, indien wij ξ als volgt zouden berekenen:

Loot x_i ($i = 1, 2, \dots, 50$) en bereken

$$\xi_1 = \frac{1}{50} \sum_{i=1}^{50} g(x_i).$$

Loot vervolgens x_i ($i = 51, 52, \dots, 100$) en bereken

$$\xi_2 = \frac{1}{100} \sum_{i=51}^{100} g(x_i),$$

enz. Omdat ξ telkens op basis van een grotere steekproef wordt berekend, mogen wij verwachten, dat ξ steeds dichterbij ξ komt te liggen. Omdat de x_i onafhankelijk zijn, zullen ook ξ_1, ξ_2, \dots onafhankelijk zijn. De kans dat alle ξ 's groter dan ξ zijn, of dat ze alle kleiner dan ξ zijn, is derhalve klein.¹

Wij hebben ξ echter anders berekend, namelijk als volgt. Loot x_i ($i = 1, 2, \dots, 50$) en bereken

$$\xi_1 = \frac{1}{50} \sum_{i=1}^{50} g(x_i).$$

Daarna echter wordt x_i ($i = 51, 52, \dots, 100$) geloot en wordt

$$\xi_2 = \frac{1}{100} \sum_{i=1}^{100} g(x_i)$$

berekend, enz. Met andere woorden:

$$\begin{aligned} \xi_j &= \frac{1}{50j} \sum_{i=1}^{50j} g(x_i) = \frac{1}{50j} \left[\sum_{i=1}^{50(j-1)} g(x_i) + \sum_{i=50(j-1)+1}^{50j} g(x_i) \right] \\ &= \frac{50(j-1)}{50j} \cdot \frac{\sum_{i=1}^{50(j-1)} g(x_i)}{50(j-1)} + \frac{50}{50j} \cdot \frac{\sum_{i=50(j-1)+1}^{50j} g(x_i)}{50} \\ &= \frac{j-1}{j} \cdot \xi_{j-1} + \frac{1}{j} \frac{\sum_{i=50(j-1)+1}^{50j} g(x_i)}{50}. \end{aligned} \quad (j = 1, 2, \dots, n/50)$$

Door deze berekeningswijze wordt ook van alle voorgaande waarnemingen gebruik gemaakt, hetgeen efficiënter is dan het eerstgenoemde procédé. Nu is ξ_j echter afhan-

$$^1 P[\xi_1 > \xi] = P\left[\frac{1}{50} \sum_{i=1}^{50} g(x_i) > \xi\right] \approx 0,50$$

omdat ξ_1 bij benadering normaal verdeeld is. De normale verdeling zal nog dichter benaderd worden door ξ_2, ξ_3, \dots doordat meer variabelen gesommeerd worden. Daaruit volgt:

$$P(\xi_1 \wedge \xi_2 \wedge \dots \wedge \xi_m > \xi) \approx (0,5)^m.$$

kelijk van de voorgaande schattingen $\hat{\xi}_{j-1}, \hat{\xi}_{j-2}, \dots, \hat{\xi}_2, \hat{\xi}_1$. De kans op schommelingen van de grafiek om ξ is bij deze methode dus kleiner dan bij de eerste methode.

Tot slot van deze opmerking wijzen wij erop, dat weliswaar de $\hat{\xi}_j$ ($j = 1, 2, \dots, n/50$) afhankelijk zijn, maar dat de $g(x_i)$ ($i = 1, 2, \dots, n$) onafhankelijk blijven, doordat de x_i onafhankelijk zijn. Daarom is de formule voor het gewenste aantal lotingen toch geldig. Zoals uit de grafische voorstelling blijkt, zou „loten totdat $\hat{\xi}_j$ zich gestabiliseerd heeft”, in dit geval een slechte leidraad zijn.

b. Zoals uit de grafische voorstelling blijkt, geeft in dit voorbeeld elk van de vier schatters een schatting van ξ die inderdaad niet meer dan 10% onnauwkeurig is. Toch blijft het mogelijk, dat een schatting meer dan 10% fout is, want:

- in de formule voor het gewenste aantal lotingen zijn $\text{var}[g(x)]$ en ξ vervangen door steekproefschattingen.
- de formule is afgeleid onder de voorwaarde dat er een kans van 5% blijft bestaan, dat toch grotere schattingsfouten dan 10% optreden. Dit aspect is nog enigermate getoetst door 10 schattingen uit te voeren. Een van die schattingen bleek buiten de toegestane foutenmarge van 10% te vallen.

Om na te gaan of onze formule voor het gewenste aantal lotingen inderdaad hoogstens een onbetrouwbaarheid van 5% inhoudt, is de nulhypothese $p \leq 5\%$ gesteld. Omdat $P(k \geq 1 | n = 10, p = 0,05) = 0,4013$ hoeft deze hypothese niet te worden verworpen.

c. Omdat elk van de vier schatters zuiver is én omdat zo lang geloot is totdat (met 95% zekerheid) de schattingsfout hoogstens 10% is, kan worden verwacht dat elke schatter even goed is. Dit is nog getoetst met behulp van de „toets van FRIEDMAN”.¹

Tabel 3. Toets van FRIEDMAN²

Experiment	$(\hat{\xi} - \xi) \times 100.000$						Rangorde		
	Monte Carlo	Importance sampling			Monte Carlo	Importance sampling			
		A	B	C		A	B	C	
1 ($\lambda = 0,5 \quad \nu = 4$)	25	55	55	155	1	2,5	2,5	4	
2 ($\lambda = 0,5 \quad \nu = 3,5$)	156	294	44	56	3	4	1	2	
3 ($\lambda = 0,25 \quad \nu = 8,5$)	4	5	2	6	2	3	1	4	

De toetsgrootheid K heeft de waarde 21,5, terwijl bij $\alpha = 0,05$ de waarde van K , welke kritiek is, 37 bedraagt. Wij hoeven dus niet te besluiten tot het verwerpen van de hypothese, dat de diverse schattingsmethoden een aselechte rangorde tonen.

¹ Het computerprogramma impliceert, dat het loten gestopt wordt indien $n > \text{GEWENSTE } n$. Daardoor is iets meer geloot dan de formule voor het gewenste aantal lotingen aangeeft. Dit aantal te veel verrichte lotingen varieert van geval tot geval, hetgeen de vergelijking in tabel 3 iets vertroebelt.

² Per λ, ν -combinatie een andere reeks aselechte getallen gebruikt.

d. In paragraaf 4 bleek dat ξ bij benadering normaal verdeeld is. Derhalve is te verwachten dat de kans op een hogere waarde dan ξ gelijk is aan de kans op een lagere waarde dan ξ . Bij schatting volgens de verschoven exponentiële verdelingsdichtheid bleek in 6 van de 17 gevallen een lagere waarde dan ξ te worden gevonden. Omdat $P(k \leq 6 | n = 17, p = 0,50) = 17\%$ is er geen reden om de hypothese $p = 0,50$ te verwerpen. Bij de andere schattingsmethoden waren te weinig onafhankelijke trekkingen beschikbaar om deze toets uit te voeren.

APPENDIX I

1. De variantie van $g(x)$

Bij toepassing van de gewone Monte Carlo-methode wordt

$$\xi(\lambda, v) = \int_v^{\infty} \frac{1}{x} \lambda e^{-\lambda x} dx$$

geschat door \underline{x} te loten uit

$$\begin{aligned} f(x) &= \lambda e^{-\lambda x} & x \geq 0 \\ f(x) &= 0 & x < 0 \end{aligned}$$

en de gelote \underline{x} te substitueren in:

$$\begin{aligned} g(x) &= 0 & \text{als } x < v \\ g(x) &= \frac{1}{x} & \text{als } x \geq v. \end{aligned}$$

Na n lotingen resulteert dit in de schatting

$$\hat{\xi} = \frac{1}{n} \sum_{i=1}^n g(x_i).$$

Uit deze definitie van $\hat{\xi}$ volgt:

$$\text{var}(\hat{\xi}) = \frac{\text{var } g(x)}{n}$$

waarbij

$$\text{var } g(x) = \int_v^{\infty} \frac{1}{x^2} \lambda e^{-\lambda x} dx - \left[\int_v^{\infty} \frac{1}{x} \lambda e^{-\lambda x} dx \right]^2.$$

Via partiële integratie van de eerste integraal en rekening houdend met de definitie van $\xi(\lambda, v)$ vinden wij ¹

$$\text{var } [g(x)] = \lambda e^{-\lambda v} v^{-1} - \lambda \xi - \xi^2.$$

Aldus is $\text{var } [g(x)]$ uitgedrukt in de onbekende ξ .

2. De variantie van $g_1^*(x)$

Wij loten \underline{x} uit:

$$\begin{aligned} h_1(x) &= c e^{cv} e^{-cx} & (x \geq v) \\ h_1(x) &= 0 & (x < v) \end{aligned}$$

¹ Eenvoudshalve is hier ξ in plaats van $\xi(\lambda, v)$ geschreven.

en berekenen

$$g_1^*(x) = \frac{\lambda}{c} e^{-cv} x^{-1} e^{-(\lambda-c)x} \quad (x \geq v).$$

Na n lotingen berekenen wij

$$\xi_1^* = \frac{1}{n} \sum_{i=1}^n g_1^*(x_i)$$

$$\text{var}(\xi_1^*) = \frac{\text{var}[g_1^*(x)]}{n}.$$

Berekening van $\text{var}[g_1^*(x)]$ levert:

$$\text{var}[g_1^*(x)] = \int_v^{\infty} \frac{1}{x^2} \lambda^2 e^{-2\lambda x} c^{-1} e^{-cv} e^{cx} dx - \left[\int_v^{\infty} \frac{1}{x} \lambda e^{-\lambda x} dx \right]^2.$$

Via partiële integratie van de eerste integraal, en bedenkend dat de tweede integraal $\xi(\lambda, v)$ voorstelt, vinden wij:

$$\text{var}[g_1^*(x)] = \frac{\lambda^2 e^{-2\lambda v}}{cv} + \lambda^2 e^{-cv} \left[1 - \frac{2\lambda}{c} \right] \int_v^{\infty} \frac{1}{x} e^{-(2\lambda-c)x} dx - \xi^2. \quad (c \leq 2\lambda)$$

In deze uitdrukking voor $\text{var}[g_1^*(x)]$ treedt naast $\xi(\lambda, v)$ een integraal op, die ons voor dezelfde problemen stelt als $\xi(\lambda, v)$.

3. De bepaling van de optimale c in $h_1(x)$

De optimale c is de waarde van c die een minimale variantie van $g_1^*(x)$ oplevert. Daarom differentiëren wij $\text{var}[g_1^*(x)]$ naar c en stellen die afgeleide 0. Dit geeft:

$$\frac{d\{\text{var}[g_1^*(x)]\}}{dc} = y_1'(c) + y_2'(c)y_3(c) + y_2(c)y_3'(c) = 0$$

waarbij

$$y_1(c) = \frac{\lambda^2 e^{-2\lambda v}}{cv}$$

$$y_2(c) = \lambda^2 e^{-cv} - \frac{2\lambda^3 e^{-cv}}{c}$$

$$y_3(c) = \int_v^{\infty} \frac{1}{x} e^{-(2\lambda-c)x} dx.$$

In deze afgeleide luidt de tweede term derhalve

$$(-\lambda^2 v e^{-cv} + 2\lambda^3 c^{-2} e^{-cv} + 2v\lambda^3 c^{-1} e^{-cv}) \int_v^{\infty} \frac{1}{x} e^{-(2\lambda-c)x} dx.$$

Aangezien de integraal in de tweede term dezelfde problemen als $\xi(\lambda, v)$ geeft, kan c niet uit de afgeleide worden berekend. Daardoor kan de minimale $\text{var}[g_1^*(x)]$ niet bepaald worden.

4. Vergelijking van $\text{var}[g(x)]$ met $\text{var}[g_1^*(x)]$ indien $c = \lambda$

Indien $c = \lambda$ kunnen wij voor

$$\int_v^{\infty} \frac{1}{x} e^{-(2\lambda-c)x} dx$$

schrijven

$$\int_v^{\infty} \frac{1}{x} e^{-\lambda x} dx = \frac{1}{\lambda} \int_v^{\infty} \frac{1}{x} \lambda e^{-\lambda x} dx = \frac{\xi}{\lambda}$$

Dan geldt dus

$$\text{var}[g_1^*(x)] < \text{var}[g(x)] \quad \text{mits}$$

$$\frac{\lambda e^{-2\lambda v}}{v} + (\lambda^2 e^{-\lambda v} - 2\lambda^2 e^{-\lambda v}) \frac{\xi}{\lambda} - \xi^2 < \lambda e^{-\lambda v} v^{-1} - \lambda \xi - \xi^2.$$

Hieraan is voldaan indien

$$\xi < e^{-\lambda v} v^{-1}.$$

Rekening houdend met de definitie van ξ_1 vinden wij dat inderdaad

$$\xi = \int_v^{\infty} \frac{1}{x} \lambda e^{-\lambda x} dx < \int_v^{\infty} \frac{1}{v} \lambda e^{-\lambda x} dx = -\frac{1}{v} [e^{-\lambda x}]_v^{\infty} = \frac{1}{v} e^{-\lambda v}$$

waarmee bewezen is

$$\text{var}[g_1^*(x)] < \text{var}[g(x)].$$

APPENDIX II

1. De variantie van $g_2^*(x)$

x wordt geloot uit

$$h_2(x) = b^2 x e^{-bx} \quad \text{als } x \geq 0$$

$$h_2(x) = 0 \quad \text{als } x < 0$$

en gesubstitueerd in

$$g_2^*(x) = 0 \quad \text{indien } x < v$$

$$g_2^*(x) = \frac{\lambda}{b^2} x^{-2} e^{-(\lambda-b)x} \quad \text{indien } x \geq v.$$

$$\text{var}[g_2^*(x)] = \int_v^\infty \frac{x^{-2}\lambda^2 e^{-2\lambda x}}{b^2 x e^{-bx}} dx - \left[\int_v^\infty \frac{1}{x} \lambda e^{-\lambda x} dx \right]^2.$$

Zoals in appendix I, paragraaf 2, kan deze uitdrukking worden herleid tot:

$$\begin{aligned} \text{var}[g_2^*(x)] &= \frac{\lambda^2 e^{(b-2\lambda)v}}{b^2 v^2} + \frac{\lambda^2 (b-2\lambda) e^{(b-2\lambda)v}}{b^2 v} + \\ &+ \frac{\lambda^2 (b-2\lambda)^2}{b^2} \int_v^\infty \frac{1}{x} e^{(b-2\lambda)x} dx - \xi^2. \end{aligned}$$

In deze uitdrukking voor $\text{var}[g_2^*(x)]$ treedt, net als bij $\text{var}[g_1^*(x)]$, weer een integraal op die dezelfde moeilijkheden als $\xi(\lambda, v)$ oplevert.

Teneinde de waarde van b te bepalen die de variantie van $g_2^*(x)$ minimaliseert, moet de afgeleide van $\text{var}[g_2^*(x)]$ naar b worden bepaald, en vervolgens 0 worden gesteld. Uit deze gelijkheid moet dan b worden opgelost. Aangezien echter in deze vergelijking de onbekende integraal

$$\int_v^\infty \frac{1}{x} e^{(b-2\lambda)x} dx$$

voorkomt, kan de optimale b niet worden berekend.

Zelfs is het onmogelijk na te gaan, of bij $b = 2/v$ voldaan is aan

$$\text{var}[g_2^*(x)] < \text{var}[g(x)].$$

In appendix I bleek het wel mogelijk te bewijzen dat bij $c = \lambda$ voldaan is aan

$$\text{var}[g_1^*(x)] < \text{var}[g(x)].$$

Daar ging echter bij $c = \lambda$ de onbekende integraal

$$\int_v^\infty \frac{1}{x} e^{-(2\lambda-c)x} dx$$

over in ξ/λ .