

Schatten van kleine kansen door grote afwijkingen

Ad Ridder

Samenvatting

In deze studie behandelen we twee eenvoudige voorbeelden die illustratief zijn voor kwaliteitsgarantie van communicatiesystemen, namelijk een kleine kans op vermindering van verzonden berichten en een kleine kans op verlies van delen van berichten. Deze kansen kunnen geschat worden door bijvoorbeeld het uitvoeren van simulaties. Omdat we hier typisch spreken over kansen van de orde 10^{-6} en kleiner, is het wenselijk variantiereductietechnieken toe te passen. Wij zullen in de voorbeelden laten zien hoe *importance sampling* de simulaties versnelt. Om een optimale nieuwe kansmaat te vinden worden limietresultaten uit de theorie van grote afwijkingen gebruikt.

1 Inleiding

Een typisch ongewenst gedrag van communicatiesystemen is een verminkt ontvangst van het verzonden bericht. Ten gevolge van ruis gedurende transmissie zullen enige 'bits' verkeerd aankomen maar als dat er niet te veel zijn vindt correctie plaats met behulp van de controle informatie. Toekomstige communicatienetwerken leggen geen directe verbinding tussen verzender en ontvanger. Het bericht wordt gesplitst in pakketten die door een netwerk van tussenliggende knooppunten naar de plaats van bestemming gaan. Bij druk verkeer kan opstapeling van pakketten ontstaan in die knooppunten. Bufferruimtes zijn daarom beschikbaar om (niet al te grote) fluctuaties in het verkeer op te vangen. Echter bij 'spitstijden' is het aanbod zo groot dat de buffers vol kunnen raken en pakketten verloren gaan.

Het is onmogelijk om een volledig foutenvrije ontvangst en nulverlies van berichten te garanderen, maar systeembeheerders streven naar een prestatie waarbij de twee bovenervermelde onvolkomenheden minder vaak optreden dan een acceptatieniveau. Vaak

worden simulatiestudies aangewend om inzicht te verkrijgen in het gedrag van een systeem onder verschillende scenario's. Uit deze studies worden de nodige prestatiematen geschat. Maar als deze maten van een zeer kleine gewenste orde zijn—zoals verlieskansen in een communicatienetwerk van 10^{-6} of kleiner—worden de simulatie runs dermate lang dat ze niet alleen veel tijd vergen, maar ook onbetrouwbaar worden. Dit is een klassiek probleem in simulatiestudies, en vele technieken zijn ontwikkeld die oplossingen aanbieden. Ieder tekstboek over simulatie bevat minstens een hoofdstuk dat deze zogenaamde variantiereductietechnieken behandelt (zie bv. [3, 8, 10, 13]).

In onze studie rapporteren we een methode die tegenwoordig veel opgang doet bij simulaties van communicatie- en wachtrijsystemen, namelijk *importance sampling*. Bij deze technieken worden de steekproeven getrokken volgens een andere onderliggende kansmaat doordat systeemparemeters zijn veranderd. Om binnen een klasse van mogelijke kansmaten een 'optimale' te vinden wordt voorgesteld om de theorie van *grote afwijkingen* toe te passen. We zullen aan de hand van twee voorbeelden laten zien hoe dit kan worden aangepakt.

2 Achtergronden

Laat $(\Omega, \mathcal{F}, \mathbb{P})$ een kansruimte zijn en A een een zeldzame gebeurtenis. Dat betekent

$$A \in \mathcal{F} \quad \text{en} \quad \alpha \triangleq \mathbb{P}(A) \text{ is klein.}$$

We willen een schatting van α krijgen door Monte Carlo simulaties. Daartoe doen we m onafhankelijke observaties (of: m onafhankelijke trekkingen) $\omega_1, \omega_2, \dots, \omega_m$ uit Ω volgens kans $\mathbb{P}$ en schatten α met de relatieve frequentie van goede ω 's:

$$\hat{\alpha}_p = \frac{1}{m} \sum_{k=1}^m 1_{\{\omega_k \in A\}}.$$

De gebruikte schatter bij dit experiment is zuiver en heeft een variantie die geschat kan worden door

$$\hat{\sigma}_p^2 = \frac{1}{m-1} \left(\frac{1}{m} \sum_{k=1}^m 1_{\{\omega_k \in A\}} - \hat{\alpha}_p^2 \right).$$

De prestatie van de schatting wordt weergegeven door twee getallen, betrouwbaarheid γ en efficiency ϵ . Laat $t_\gamma > 0$ de $(1 + \gamma)/2$ quantiel zijn van standaard normale verdeling, $\Phi(t_\gamma) = 1 - (\gamma/2)$. Formeler zouden we hier de Student's t verdeling moeten gebruiken maar omdat we met grote steekproeven zullen werken, is er geen groot verschil met de normale verdeling. Een γ -betrouwbaarheidsinterval is dus

$$(\hat{\alpha}_p - t_\gamma \hat{\sigma}_p, \hat{\alpha}_p + t_\gamma \hat{\sigma}_p).$$

Dan wordt de efficiency van de schatting gedefinieerd als de ratio $\epsilon = t_\gamma \hat{\sigma}_p / \hat{\alpha}_p$. Bij een vereist (γ, ϵ) paar kunnen we dus de steekproefgrootte m bepalen uit

$$m \geq \epsilon^{-2} t_\gamma^2 \frac{1 - \hat{\alpha}_p}{\hat{\alpha}_p}. \quad (1)$$

Bijvoorbeeld bij $\gamma = 0.95$, $\epsilon = 0.1$ en $\hat{\alpha}_p \approx \alpha < 1$

$$m \approx 400 \frac{1 - \alpha}{\alpha} \approx 400 \frac{1}{\alpha}.$$

Bij kansen van de orde 10^{-6} (en kleiner) kunnen dus problemen ontstaan om binnen beheersbare tijd de simulaties uit te voeren. Bovendien is het mogelijk dat de cyclenlengte van de gebruikte *random generator* wordt overschreden.

Zoals al gezegd in de introductie is *importance sampling* een techniek om variantiereductie te verkrijgen. Bij deze techniek worden de simulaties uitgevoerd onder een andere kansmaat Q op $(\Omega, \mathcal{F})$. Als we eisen dat de oorspronkelijke maat IP absoluut continu is ten opzichte van de nieuwe kans Q , bestaat er een Radon-Nikodym afgeleide of *likelihood* die we met L aangeven:

$$L \triangleq \frac{dIP}{dQ}. \quad (2)$$

Merk op dat we dan α kunnen herschrijven in $\alpha = \int_A dIP = \int_A L dQ$. Hierin zien we dat de absoluut continuïteitseis afgezwakt kan worden tot de eis dat een afgeleide $L = dIP/dQ$ moet bestaan op de gebeurtenis A .

Laat nu $\omega_1, \omega_2, \dots, \omega_m$ een steekproef uit Ω zijn gegenereerd door uit Q te trekken. We berekenen $\ell_k = L(\omega_k)$, $k = 1, \dots, m$, en schatten α door

$$\hat{\alpha}_q = \frac{1}{m} \sum_{k=1}^m \ell_k \mathbf{1}\{\omega_k \in A\}.$$

De gebruikte schatter is weer zuiver. Echter, nu wordt de variantie geschat door

$$\hat{\sigma}_q^2 = \frac{1}{m-1} \left(\frac{1}{m} \sum_{k=1}^m \ell_k^2 \mathbf{1}\{\omega_k \in A\} - \hat{\alpha}_q^2 \right).$$

Veronderstel dat we op deze wijze een variantiereductie willen bereiken in de simulaties. Uit de expressies van de varianties van van de gebruikte schatters volgt dan noodzakelijk

$$\int_A L^2 dQ \leq \int_A dIP = \alpha.$$

De ongelijkheid van Jensen geeft ons een ondergrens voor het linkerlid:

$$\int_A L^2 dQ \geq \left(\int_A L dQ \right)^2 = \alpha^2 \quad (3)$$

Een optimale verandering van kansmaat zou daarom een Q zijn die een gelijkheid geeft, dus $L = \alpha$ op A , ofwel $Q(A) = 1$. De schatter bij deze kansmaat heeft variantie nul maar kan niet geïmplementeerd worden want dan is kennis van de te schatten onbekende parameter nodig. Bovendien blijven eigenschappen die onder IP gelden—zoals Markoveigenschap—niet behouden onder deze Q . Dus ook uit praktische overwegingen komt Q niet in aanmerking (zie [5]). In de volgende secties zullen we een methodologie bestuderen om een optimale nieuwe kansmaat te vinden die tot een zekere klasse behoort.

Om een indicatie te geven van de simulatieversnellingen, nemen we aan dat de dichtheid L (bijna) constant is, $L(\omega) \approx \ell$, op de zeldzame gebeurtenis A voor zekere $\ell > \alpha$. Dan wordt de steekproefomvang bij een vereist (γ, ϵ) paar,

$$m \geq \epsilon^{-2} t_{\gamma}^2 \frac{\ell - \hat{\alpha}_q}{\hat{\alpha}_q}. \quad (4)$$

Veronderstel verder een vereist (γ, ϵ) paar, en bovendien geschatte waardes $\hat{\alpha}_p \approx \alpha$ en $\hat{\alpha}_q \approx \alpha$, en $\alpha \ll 1$. Kies de steekproefgroottes m_p en m_q volgens de rechterleden van (1) en (4). Dan wordt de reductiefactor

$$\frac{m_q}{m_p} \approx \frac{\ell - \alpha}{1 - \alpha} \approx \ell - \alpha.$$

In de voorbeelden zullen we zien dat als ℓ proportioneel is met α , de reductie enorm kan zijn.

3 Grote afwijkingen

We veronderstellen vanaf nu dat de zeldzame gebeurtenis A nog afhangt van een parameter $n \in \mathbb{N}$, bijvoorbeeld A is de gebeurtenis dat er meer dan n klanten in een wachtrijsysteem aanwezig zijn. We schrijven dan A_n en

$$\alpha_n = IP(A_n)$$

Ook nemen we aan dat α_n het volgende limietgedrag vertoont:

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \alpha_n \leq -I(\cdot), \quad (5)$$

waarbij de functie I gedefinieerd is op een metrische ruimte X zodat

- (i) $I(x) \in [0, \infty]$ (voor alle x),
- (ii) I is convex,
- (iii) I is onder semi-continu,
- (iv) $\{x \in X : I(x) \leq a\}$ is compact (voor alle a).

Dergelijke limieten zijn ontwikkeld in de theorie van grote afwijkingen (*large deviations*). We verwijzen voor een uitgebreide behandeling van dat onderwerp naar Dembo en Zeitouni [7] en Bucklew [4].

De functie I geeft de exponentiële snelheid waarmee α_n naar nul convergeert. Het argument van I in (5) is probleemafhankelijk en zal in de voorbeelden verder gespecificeerd worden. In veel gevallen bestaat naast (5) ook een overeenkomstige ondergrens voor de liminf en onder zekere regulariteitsvoorwaarden zijn onder- en bovengrens gelijk zodat (5) vervangen kan worden door een limiet met gelijkheid.

Zoals we in (3) in de vorige sectie zagen, is bij een verandering (2) van kansmaat het tweede moment

$$v_n(Q) \triangleq \int_{A_n} L^2 dQ$$

van belang om variantiereductie bij simulaties te ontwikkelen. We veronderstellen nu dat we kansmaten Q kunnen vinden zodat de bijbehorende $v_n(Q)$ eveneens een limietgedrag als (5) tonen. Dus

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log v_n(Q) \leq -I_Q(\cdot),$$

voor een zekere functie I_Q . Merk op dat (3) leidt tot

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log v_n(Q) \geq -2I(\cdot).$$

We zeggen daarom dat een kansmaat Q *asymptotisch efficient* is als

$$\lim_{n \rightarrow \infty} \frac{\log v_n(Q)}{\log \alpha_n} = 2, \quad (6)$$

zie [2, 5, 11, 14].

4 Voorbeeld 1

Beschouw een communicatiekanaal om 0 en 1 symbolen (*bits*) te versturen. De verzender stuurt een woord van n bits door het kanaal maar ten gevolge van ruis hoeft het andere eind van het kanaal niet hetzelfde woord te ontvangen. We veronderstellen dat het foutief ontvangen van een bit alleen afhangt of het voorafgaande bit al dan niet correct was verstuurd. Laat de stochastische variabele X_i aangeven of het i -de bit correct ($X_i = 0$) of foutief ($X_i = 1$) ontvangen is. Dan is $X_1, X_2, \dots, X_n$ een Markovketen op $\{0, 1\}$ met matrix $P = (p_{ij})$ van overgangskansen die we noteren als

$$P = \begin{pmatrix} 1-p & p \\ 1-q & q \end{pmatrix}.$$

We nemen aan dat de eerste rij van de matrix de beginverdeling a van de keten is:

$$a_0 = \mathbb{P}(X_1 = 0) = 1 - p, \quad a_1 = \mathbb{P}(X_1 = 1) = p.$$

De stationaire verdeling is eenvoudig gevonden:

$$\pi_0 = \frac{1-q}{1+p-q}, \quad \pi_1 = \frac{p}{1+p-q}.$$

Het aantal fouten in het ontvangen woord na verzending is $S_n = X_1 + X_2 + \dots + X_n$ waarvan we verlangen dat het ‘klein’ is ten opzichte van de lengte n van het woord. Merk op dat als n naar oneindig gaat, $\mathbb{E}_{\mathbb{P}} S_n/n \rightarrow \pi_1$, dus voor grote n verwachten we $n\pi_1$ fouten. Laat $\delta \in (\pi_1, 1)$ en definieer de ‘zeldzame’ gebeurtenis

$$A_n = \left\{ \frac{S_n}{n} \geq \delta \right\}.$$

Het blijkt nu dat het limietgedrag (5) voor Markovketens op een eindige toestandsruimte bewezen kan worden en dat bovendien de functie I bepaald kan worden (zie bijvoorbeeld hoofdstuk V in [4]). In feite is dit een toepassing van de stelling van Gärtner-Ellis (p.15 in [4], p.45 in [7]). Die zegt het volgende. Laat

$$\psi(\theta) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} \exp(\theta S_n), \quad \theta \in \mathbb{R}$$

de asymptotische cumulantgenererende functie zijn van de rij partieelsommen (S_n) en

$$I(x) = \sup_{\theta} [\theta x - \psi(\theta)], \quad x \in \mathbb{R} \quad (7)$$

de Legendre-Fenchel getransformeerde van ψ . Dan geldt onder een aantal voorwaarden (bv. bovenstaande limiet moet bestaan, ψ moet differentieerbaar zijn binnen zijn domein)

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(S_n/n \geq \delta) = -I(\delta), \quad (\delta > \pi_1).$$

Met behulp van eigenschappen van nietnegatieve matrices kan de functie ψ bepaald worden als $\psi(\theta) = \log \rho(\theta)$ waarin $\rho(\theta)$ de grootste eigenwaarde is van de matrix

$$T(\theta) = \begin{pmatrix} 1-p & pe^{\theta} \\ 1-q & qe^{\theta} \end{pmatrix}.$$

We noteren θ^* voor een argument waarvoor het supremum in (7) bereikt wordt wanneer $x = \delta$. Hiervoor zijn geen analytische uitdrukkingen beschikbaar, maar zal in het algemeen numeriek bepaald worden. We merken nog op dat aangetoond kan worden met eigenschappen van de functie I dat θ^* positief en uniek is.

We veronderstellen dat de kansruimte $(\Omega, \mathcal{F}, \mathbb{P})$ de coördinatenruimte is van de keten, dus $\Omega = \{0, 1\}^n$ en de stochastische variabelen X_i zijn de projectiefuncties. Daarom is, als $\omega = (i_1, i_2, \dots, i_n) \in \Omega$,

$$\mathbb{P}(\omega) = a_{i_1} \prod_{k=1}^{n-1} p_{i_k, i_{k+1}}.$$

Hierdoor wordt α_n eenvoudig uitgedrukt en kan exact berekend worden. Tevens zullen we Monte Carlo simulaties uitvoeren en daarmee de gevraagde kans te schatten. Zoals uitgelegd in sectie 2 kunnen we door de kansmaat op een slimme manier te veranderen variantiereductie tot stand brengen. Daartoe beschouwen we een klasse van kansmaten op Ω zodat de rij $X_1, X_2, \dots, X_n$ een Markovketen blijft. Elk lid Q van deze klasse wordt eenduidig gekarakteriseerd door een beginverdeling b (op Ω) en een matrix Q van overgangskansen:

$$b = (b_0, b_1) \quad \text{en} \quad Q = \begin{pmatrix} q_{00} & q_{01} \\ q_{10} & q_{11} \end{pmatrix}.$$

Met andere woorden

$$Q(\omega) = b_{i_1} \prod_{k=1}^{n-1} q_{i_k, i_{k+1}}.$$

We willen een kansmaat Q^* in deze klasse vinden die asymptotisch efficient is (vergelijking (6)). We refereren naar hoofdstuk VIII van [4] voor het uiteindelijke resultaat:

$$Q^* = \frac{1}{\rho(\theta^*)} \begin{pmatrix} 1-p & pe^{\theta^* \frac{x_1}{x_0}} \\ (1-q) \frac{x_0}{x_1} & qe^{\theta^*} \end{pmatrix},$$

waarbij (x_0, x_1) een positieve eigenvector van $T(\theta^*)$ bij eigenwaarde $\rho(\theta^*)$. Immers, dan is de waarschijnlijkheid (na enig rekenwerk)

$$L(\omega) = \frac{IP(\omega)}{Q^*(\omega)} = \frac{a_{i_1}}{b_{i_1}} (\rho(\theta^*))^{n-1} \exp \left(-\theta^* \sum_{k=2}^n i_k \right) \frac{x_{i_1}}{x_{i_n}}.$$

Omdat voor $\omega \in A$ geldt dat $\sum_{k=1}^n i_k \geq n\delta$ en omdat θ^* positief is, kan L^2 naar boven afgeschat worden, en hierdoor ook $v_n(Q^*)$:

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log v_n(Q^*) \leq -2(\theta^* \delta - \log \rho(\theta^*)) = -2I(\delta).$$

Met andere woorden, de nieuwe kansmaat Q^* is asymptotisch efficient. Onder Q^* is $X_1, X_2, \dots, X_n$ weer een Markovketen met Q^* als matrix van overgangskansen, en $E_{Q^*} S_n / n \rightarrow \delta$, als $n \rightarrow \infty$. Hieruit concluderen we dat in de nieuwe simulaties steeds gemiddeld $n\delta$ fouten in de getrokken ω 's zitten.

De methode van snelle simulatie laat nu toe allerlei scenario's door te rekenen om een ontwerp te ontwikkelen met een verlangde maximale fouttolerantie δ . In tabel 1 staan de geschatte overschrijdingskansen bij 20% ($\delta = 0.2$) voor het model met $p = 0.05, q = 0.1$. In dat geval is het gemiddeld aantal fouten in een woord van lengte n ruim 5 procent ($\pi_1 = 0.0526$). De asymptotische vervalparameter $I(0.2) = 0.113$ bij $\theta^* = 1.26$ geeft

$$Q^* = \begin{pmatrix} 0.8267 & 0.1733 \\ 0.6932 & 0.3068 \end{pmatrix}, \quad \pi_1^* = 0.2.$$

	direct			versneld			exact
n	$^{10}\log(\hat{\alpha})$	aantal woorden	CPU sec.	$^{10}\log(\hat{\alpha})$	aantal woorden	CPU sec.	$^{10}\log(\alpha)$
5	-0.6108	1200	0.0	-0.6342	500	0.0	-0.6455
10	-1.0073	3600	0.4	-0.9901	700	0.1	-0.9954
20	-1.5783	14200	3.0	-1.5972	1000	0.3	-1.6052
30	-2.2047	61200	20	-2.1265	1100	0.6	-2.1707
40	-2.6722	181000	77	-2.7203	1400	0.9	-2.7164
50	-3.1244	511400	272	-3.2129	1500	1.3	-3.2507
60	-3.6294	1640000	1046	-3.7159	1500	1.5	-3.7775
70	-4.0699	4522000	3357	-4.2517	1700	2.0	-4.2992

Tabel 1: Overschrijdingskansen, benodigde aantal woorden en CPU seconden van directe en versnelde simulaties van voorbeeld 1.

De simulaties zijn uitgevoerd met een vereiste ($\gamma = 0.95, \epsilon = 0.1$) paar. De bijbehorende aantallen woorden die gesimuleerd zijn met de benodigde CPU tijden (op een NeXT workstation), zijn eveneens vermeld. Ter vergelijking zijn de exacte overschrijdingskansen tevens numeriek berekend.

We onderscheiden duidelijk de exponentiële toenames in aantal gesimuleerde woorden en in executietijd van de directe simulaties en de lineaire toenames in geval van de *importance sampling* simulatie.

5 Voorbeeld 2

Het tweede voorbeeld behandelt een discreet wachtrijmodel. Beschouw een servicefaciliteit met een wachtruimte van omvang n en één bediende. Aankomsten worden gegeneerd door de discrete-tijd Markovketen $(M_t)_{t=0}^{\infty}$ op $\{0, 1\}$ met matrix van overgangskansen

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}.$$

Met de t -de periode bedoelen we de tijdspanne tussen t en $t+1$. Als de keten aan het begin van de t -de periode zich in toestand 1 bevindt, zal er in die periode één klant aankomen bij de servicefaciliteit. In toestand 0 wordt geen klant gegenereerd. Als er aan het begin van de t -de periode een klant of klanten aanwezig zijn bij de servicefaciliteit, wordt de bediening van de voorste (eerst aangekomen) aan het eind van die periode beëindigd met kans μ en verlaat die klant het systeem aan het eind van de periode. Met kans $1 - \mu$ zal de

voorstel klant dus blijven en wordt opnieuw 'geholpen' in de volgende periode. Een klant die in de t -de periode aankomt, kan pas vanaf periode $t + 1$ geholpen worden. Tenslotte, als een klant aankomt bij een vol systeem (n klanten aanwezig), gaat hij verloren. In dit model zijn we geïnteresseerd in de fractie van het aantal klanten dat verloren gaat, aangegeven met α_n .

Als X_t het aantal aanwezige klanten aangeeft aan het begin van de t -de periode, dan is het simultane proces $(M_t, X_t)_{t=0}^{\infty}$ regeneratief met regeneratietijdstippen de momenten waarop $(M_t, X_t) = (1, 0)$. Het is eenvoudig in te zien dat de verlieskans α_n evenredig is met de fractie van regeneratiecycli waarin een verlies optreedt. Geef deze fractie (of overschrijdingskans) aan met ϕ_n . Formeel kunnen we dan de ruimte $(\Omega, \mathcal{F}, \mathbb{P})$ van alle regeneratiecycli invoeren: iedere ω representeert een realisering van een rij (X_t, M_t) 's tussen twee regeneratietijdstippen. De zeldzame gebeurtenis A_n is de deelverzameling van cycli waarin niveau n bereikt wordt.

Schattingen van de verlies- en overschrijdingskansen α_n en ϕ_n kunnen verkregen worden uit regeneratieve simulaties. Onder aanname dat er gemiddeld per periode minder klanten aankomen dan vertrekken, $p/(p+q) < \mu$, nemen deze kansen exponentieel af (in n) met de gebruikelijke lange simulatieruns als gevolg. We zullen een afleiding hiervan geven.

Beschouw de terugkeertijdstippen $R_0 = 0, R_1, R_2, \dots$ naar toestand 1 door de keten (M_t) en definieer aangroeiingen

$$\xi_k = \sum_{t=R_{k-1}}^{R_k-1} (M_t - B_t), \quad k = 1, 2, \dots$$

Hierin is $B_0, B_1, \dots$ een rij van onafhankelijke Bernoulli(μ) stochastische variabelen (en onafhankelijk van de Markovketen). Het is duidelijk dat ξ_k de werkelijke aangroeiing beschrijft van het aantal klanten in het systeem tussen twee terugkeertijdstippen als in die periode het systeem niet leeg raakt of vol loopt. Deze aangroeiingen zijn onafhankelijk en gelijkverdeeld, met andere woorden de rij partiele sommen $(S_k)_{k=0}^{\infty}$ van (ξ_k) , is een *random walk*. Bestuderen we het systeem vanaf de start (op tijdstip 0) van een regeneratiecyclus totdat een van de fysieke grenzen 0 of n bereikt wordt, dan vormt de *random walk* (met $S_0 = 0$) een inbedding van het proces (X_t) op speciale tijdstippen. Tenslotte is eenvoudig in te zien dat als het systeem vol loopt, zeg op tijdstip T , het tijdstip vlak daarvoor een terugkeertijdstip is met $X_{T-1} = n - 1$ en $B_{T-1} = 0$.

We mogen dus concluderen dat als het aantal klanten in een regeneratiecyclus hoge niveaus bereikt, ook de *random walk* deze niveaus bereikt. Zetten we $S_0 = 1$ en definiëren we de stoptijd

$$T = \inf\{k > 0 : S_k \leq 0 \text{ of } S_k \geq n\},$$

dan moet gelden

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \phi_n = \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(S_T \geq n).$$

Om deze laatste te bepalen redeneren we als volgt. Laat $(p(x))_{x \in \mathbb{Z}}$ de kansdichtheid zijn van de aangroeiing ξ . De θ -verdraaide dichtheid wordt gedefinieerd door

$$p_\theta(x) = \frac{e^{\theta x} p(x)}{M(\theta)}, \quad M(\theta) = \mathbb{E} \exp(\theta \xi), \quad \theta \in \mathbb{R},$$

en kan beschouwd worden als de kansdichtheid van de aangroeiingen onder een kansmaat $\mathbb{P}_\theta$ in plaats van de oorspronkelijke $\mathbb{P}$. Dan is

$$\mathbb{P}(S_T \geq n) = \sum_{t=1}^{\infty} \mathbb{E} 1\{T = t, S_t \geq n\} = \sum_{t=1}^{\infty} \mathbb{E}_\theta L_t 1\{T = t, S_t \geq n\},$$

waarin L_t de waarschijnlijkheid (of *likelihood*) van de eerste t waarnemingen is, dus

$$L_t = L_t(\xi_1, \dots, \xi_t) = \frac{p(\xi_1) \cdots p(\xi_t)}{p_\theta(\xi_1) \cdots p_\theta(\xi_t)} = M^t(\theta) \exp(-\theta \sum_{k=1}^t \xi_k) = M^t(\theta) \exp(-\theta(S_t - 1)). \quad (8)$$

Hierdoor wordt voor positieve θ

$$\mathbb{P}(S_T \geq n) = \sum_{t=1}^{\infty} \mathbb{E}_\theta M^t(\theta) \exp(-\theta(S_t - 1)) 1\{T = t, S_t \geq n\} \leq e^{-\theta n} e^\theta,$$

ofwel

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(S_T \geq n) \leq -\theta.$$

Dit geldt voor elke positieve θ . Met behulp van *Slow Markov Walk* technieken (zie bv. [6] of hoofdstuk IV in [4]) kan ook de ondergrens bewezen voor een unieke positieve θ^* , namelijk die waarvoor $M(\theta) = 1$. Nogmaals gebruikmakend van expressie (8) van de *likelihood* kan eenvoudig bewezen worden dat de gevonden kansmaat $\mathbb{P}_{\theta^*}$ asymptotisch efficiënt is.

Onder deze verdraaide kansmaat is het proces (M_t) weer een Markovketen, zeg met overgangsmatrix

$$Q^* = \begin{pmatrix} 1 - p^* & p^* \\ q^* & 1 - q^* \end{pmatrix},$$

en worden de vertrekken gegenereerd door een Bernoulli(μ^*) variabele. Het vinden van deze nieuwe parameters gebeurt aan de hand van de observatie dat voor de kansdichtheden van de aangroeiingen moet gelden

$$p_{\theta^*}(0) = \frac{p(0)}{M(\theta^*)} = p(0). \quad (9)$$

Beschouw daarom ‘minimale’ cycli van toestanden van de Markovketen (M_t, X_t) die $\xi = 0$ leveren:

- (i) $(1, x) \rightarrow (1, x),$
- (ii) $(0, x) \rightarrow (0, x),$
- (iii) $(0, x) \rightarrow (1, x - 1) \rightarrow (0, x),$
- (iv) $(0, x) \rightarrow (1, x - 1) \rightarrow (1, x) \rightarrow (0, x).$

Volgens (9) zijn de nieuwe overgangskansen langs deze cycli gelijk aan de oorspronkelijke, d.w.z.:

- (i) $\mu^*(1 - q^*) = \mu(1 - q)$,
- (ii) $(1 - \mu^*)(1 - p^*) = (1 - \mu)(1 - p)$,
- (iii) $\mu^*p^*(1 - \mu^*)q^* = \mu p(1 - \mu)q$,
- (iv) $\mu^*p^*(1 - \mu^*)(1 - q^*)\mu^*q^* = \mu p(1 - \mu)(1 - q)\mu q$.

Hieruit lossen we de nieuwe parameters op. We merken op dat als $q = 1 - p$ de aankomsten per periode onafhankelijk Bernoulli(p) verdeeld zijn. Dan wordt de optimale verdraaiing verkregen door aankomst- en bedieningsparameters te verwisselen, i.e. $p^* = \mu$, $\mu^* = p$ (en $q^* = 1 - p^*$). Dit komt overeen met het verwisselen van ' λ ' en ' μ ' voor een optimale verdraaide kansmaat in een $M/M/1$ systeem (zie [12]).

In de versnelde simulaties worden regeneratiecycli gesimuleerd door een hybride toepassing van zowel de nieuwe als de oorspronkelijke kansparameters. Zolang een cyclus nog niet tot een verlies heeft geleid, wordt het proces (M_t, X_t) statistisch gerealiseerd door de nieuwe kansmaat IP_{θ^*} . Omdat de drift $p^*/(p^* + q^*) - \mu^*$ dan positief is, zal het proces (X_t) de neiging vertonen snel naar niveau n weg te lopen. Vanaf het moment dat het aantal klanten in de cyclus niveau n heeft bereikt, wordt de simulatie weer uitgevoerd onder de oorspronkelijke parameters, totdat de cyclus eindigt. Dit wordt gedaan omdat anders de 'staart' van de cyclus, dat is het deel vanaf niveau n tot het begin van de volgende cyclus, lang kan zijn. Immers, onder kansmaat Q^* is het bereiken van een leeg systeem vanuit een hoog niveau n een zeldzame gebeurtenis. Voor schattingen van de verliesfractie maken we gebruik van de regeneratieve simulatiemethode en dus hebben we de volledige cycli nodig.

De experimenten zijn uitgevoerd met een model met $p = 0.6$, $q = 0.3$ en $\mu = 0.8$. Dit geeft een systeembelasting van 0.8333. De asymptotische vervalparameter is $\theta^* = 0.6325$ waaruit $p^* = 0.75$, $q^* = 0.1765$, $\mu^* = 0.68$ volgen die een belasting 1.1905 geven onder de nieuwe kansmaat. De simulatieresultaten (met vereist $(\gamma = 0.95, \epsilon = 0.1)$ paar) zijn weergegeven in tabel 2. De exacte verlieskansen zijn berekend door de stationaire verdeling van (M_t, X_t) numeriek te bepalen.

Uit deze tabel blijkt dat de directe simulatieresultaten onbetrouwbaar worden voor grote waarden van de bufferomvang. De versnelde simulaties geven echter steeds een goede schatting van de verlieskansen. We merken ook op dat het aantal benodigde cycli in de versnelde simulaties ongeveer gelijk blijft. Dat kan verklaard worden door uit (8) te beredeneren dat L van de orde α is en dit te substitueren in vergelijking (4).

	direct			versneld			exact
n	$^{10}\log(\hat{\alpha})$	aantal cycli	CPU sec.	$^{10}\log(\hat{\alpha})$	aantal cycli	CPU sec.	$^{10}\log(\alpha)$
5	-1.9641	14100	2.7	-1.9403	1800	1.8	-1.9411
10	-3.4978	527800	110	-3.3371	1800	3.9	-3.3378
15	-5.0579	19805000	4356	-4.7065	2000	6.6	-4.7122
20	-6.6077	680190000	149514	-6.1272	2000	8.7	-6.0858
25				-7.4626	2000	11.6	-7.4584
30				-8.8435	2000	13.4	-8.8239

Tabel 2: Verlieskansen, benodigde aantal cycli en CPU seconden van directe en versnelde simulaties, en exacte verlieskansen van voorbeeld 2.

6 Conclusies

We hebben laten zien hoe *importance sampling* toegepast kan worden bij simulatiestudies voor scenarioanalyse als de mogelijke negatieve effecten uitgedrukt worden in (kleine) kansen op zeldzame gebeurtenissen. De systeemparameters waarmee gesimuleerd wordt, ondergaan een verandering zodat zeldzame gebeurtenissen vaker zullen optreden. Een optimale verandering wordt gevonden door numerieke toepassingen van asymptotische limietstellingen uit de theorie van grote afwijkingen.

De twee voorbeelden illustreren niet alleen deze methode, maar ook de ‘typische’ verschillende toepassingen. In het eerste voorbeeld ligt de lengte van een *run* vast en kan na afloop bepaald worden of er sprake is van een zeldzame gebeurtenis of niet. In de *importance sampling* variant wordt de volledige *run* uitgevoerd met de nieuwe systeemparameters.

Het tweede voorbeeld is een toepassing van regeneratieve simulatie. De lengte van een *run* is dynamisch en hangt af van de toestanden van het systeem tijdens de *run*. Het einde van de *run* wordt gekenmerkt door het optreden van een ‘frequente’ gebeurtenis, namelijk omdat het systeem toestand 0 bereikt. Van de runlengte kan slechts gezegd worden dat die bijna zeker (met kans 1) eindig is. Toestand n is de ‘zeldzame’ gebeurtenis. In de *importance sampling* variant worden zeldzaam en frequent verwisseld. Een optimale simulatie wordt daarom bereikt door de *importance sampling* slechts toe te passen tot het bereiken van de (verwisselde) frequente gebeurtenis.

Omdat we in het eerste voorbeeld slechts geïnteresseerd zijn of het aantal fouten tijdens een *run* al dan niet een drempelwaarde overschrijdt, kunnen we de simulaties nog versnellen door iedere *run* te stoppen zodra een overschrijding plaatsvindt. Dat doen

we zowel voor de directe als de *importance sampling* simulaties. In dat geval zouden we ook kunnen spreken van een dynamische runlengte, met dien verstande dat er een deterministische bovengrens voor geldt.

Dankwoord

De auteur bedankt een anonieme referent voor commentaar en suggesties ter verbetering van een vorige versie.

Referenties

- [1] S. Asmussen Conjugate processes and the simulation of ruin problems. *Stochastic Processes and Their Applications*, 20: 213 – 229, 1985.
- [2] P. Baldi, A. Frigessi and M. Piccioni. Importance sampling for Gibbs random fields. *Annals of Applied probability*, 3: 914 – 933, 1993.
- [3] P. Bratley, B. Fox and L.F. Schrage. *A guide to simulation*. Springer, New York, 1983.
- [4] J.A. Bucklew. *Large Deviation techniques in Decision, Simulation, and Estimation*. Wiley, New York, 1990.
- [5] J.A. Bucklew, P. Ney and J.S. Sadowsky. Monte Carlo simulation and large deviations theory for uniformly recurrent Markov chains. *Journal of Applied Probability*, 27: 44 – 59, 1990.
- [6] M. Cottrell, J-C. Fort and G. Malgouyres. Large deviations and rare events in the study of stochastic algorithms. *IEEE Transactions on Automatic Control*, 9: 907 – 920, 1983.
- [7] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Jones and Bartlett, Boston, 1993.
- [8] G.S. Fishman. *Principles of discrete event simulation*. Wiley, New York, 1978.
- [9] P.W. Glynn and D.L. Iglehart. Importance sampling for stochastic simulations. *Management Science*, 35: 1367 – 1392, 1989.
- [10] A.M. Law and W.D. Kelton. *Simulation modeling and analysis*. McGraw-Hill, New York, 1982.
- [11] T. Lehtonen and H. Nyrhinen. Simulating level-crossing probabilities by importance sampling. *Advances of Applied Probability*, 24: 858 – 874, 1992.
- [12] S. Parekh and J. Walrand. A quick simulation method for excessive backlog in networks of queues. *IEEE Transactions on Automatic Control*, 34: 54 – 66, 1989.
- [13] S.M. Ross. *A course in simulation*. MacMillan, New York, 1990.
- [14] J.S. Sadowsky. On the optimality and stability of exponential twisting in Monte Carlo estimation. *IEEE Transactions on Information Theory*, 39: 119 – 128, 1993.

Ontvangen: 19-10-1994

Geaccepteerd: 15-9-1995

