Improving adaptive importance sampling simulation of Markovian queueing models using non-parametric smoothing

Pieter-Tjerk de Boer
& Edwin Woudt
& Jan-Kees van Ommeren
University of Twente
Enschede, The Netherlands
ptdeboer@cs.utwente.nl
Back in 1999...

- Reuven was just starting the cross-entropy method.
- I was working towards my PhD at the University of Twente, on rare-event simulation.
- Reuven visited the UT during summer ...
- ... and introduced me to the cross-entropy method.
- We experimented with it on Markovian queueing networks, using state-independent tilting to estimate overflow probabilities.
Later in 1999...

- State-independent tilting seemed unsuitable for some simple networks.
- How to find a good state-dependent tilting?
Later in 1999...

- State-independent tilting seemed unsuitable for some simple networks.
- How to find a good state-dependent tilting?
- Just use CE with separate tilting parameters for each state:

\[
q_{\ell m} = \frac{\sum_{Z=Z_1}^{Z_N} I(Z) L(Z) \sum_i 1_{z_i=\ell} \land z_{i+1}=m}{\sum_{Z=Z_1}^{Z_N} I(Z) L(Z) \sum_i 1_{z_i=\ell}}
\]

new prob. from state \( \ell \) to \( m \)  
sum over samplepaths  
sum over steps within samplepath
Later in 1999...

- State-*independent* tilting seemed unsuitable for some simple networks.

- How to find a good state-*dependent* tilting?

- Just use CE with separate tilting parameters for each state:

\[
q_{\ell m} = \frac{\sum_{Z=Z_1}^{Z_N} I(Z) L(Z) \sum_i 1_{z_i=\ell \land z_{i+1}=m}}{\sum_{Z=Z_1}^{Z_N} I(Z) L(Z) \sum_i 1_{z_i=\ell}}
\]

new prob. from state $\ell$ to $m$  sum over samplepaths  sum over steps within samplepath

- Problem: how to estimate so many parameters with relatively few samples?

Use the fact that adjacent states have similar optimal parameters:

- Local average
- Boundary layers
- Spline smoothing
Later in 1999...

- State-independent tilting seemed unsuitable for some simple networks.
- How to find a good state-dependent tilting?
- Just use CE with separate tilting parameters for each state:

\[
q_{\ell m} = \frac{\sum_{Z=Z_1}^{Z_N} I(Z) L(Z) \sum_i 1_{z_i=\ell \land z_{i+1}=m}}{\sum_{Z=Z_1}^{Z_N} I(Z) L(Z) \sum_i 1_{z_i=\ell}}
\]

- Problem: how to estimate so many parameters with relatively few samples?
  Use the fact that adjacent states have similar optimal parameters:
  - Local average
  - Boundary layers
  - Spline smoothing
  - **2006/2007: non-parametric smoothing**
Iterative procedure

\[
\begin{align*}
(q_{\ell,m})_0 & \quad \text{initial tilting} \\
\text{simulation} & \quad (q_{\ell,m})_1 \\
\text{target prob. estimate} & \quad \text{& relative error} \\
\text{simulation} & \quad (q_{\ell,m})_2 \\
\text{target prob. estimate} & \quad \text{& relative error} \\
\text{simulation} & \quad (q_{\ell,m})_3 \\
\text{target prob. estimate} & \quad \text{& relative error}
\end{align*}
\]
Iterative procedure

\[
\begin{align*}
(q_{\ell,m})_0 & \quad \rightarrow \quad \text{simulation} \\
(q_{\ell,m})_1 & \quad \rightarrow \quad \text{smoothing} \\
(q_{\ell,m})_2 & \quad \rightarrow \quad \text{smoothing} \\
(q_{\ell,m})_3 & \quad \rightarrow \quad \text{simulation}
\end{align*}
\]

Initial tilting

target prob. estimate & relative error

target prob. estimate & relative error

target prob. estimate & relative error
M/M/1/20 example

![Graph showing the relationship between the number of jobs in queue and the tilted probability of an arrival. The graph includes simulation results represented by red circles.](image-url)
M/M/1/20 example

![Graph showing the relationship between the number of jobs in queue and the tilted probability of an arrival. The graph includes simulation results and optimal values.]
M/M/1/20 example

Simulation results
Optimal values
Local linear regression h=1.5

Number of jobs in queue vs. tilted probability of an arrival

UT/EWI/DACS/PTdB 2008
M/M/1/20 example

![Graph showing the tilted probability of an arrival versus the number of jobs in queue. The graph includes simulation results, optimal values, and local linear regression with h=4.](image)

- Simulation results
- Optimal values
- Local linear regression h=4

UT/EWI/DACS/PTdB 2008
M/M/1/20 example

The graph shows the relationship between the number of jobs in queue and the tilted probability of an arrival. The data points represent simulation results, while the dashed line represents the optimal values from local linear regression with h=15.
Choosing the kernel width $h$
Choosing the kernel width $h$

$$h_{\text{good}} = 10^{0.4D+0.15B-1} \times (\sigma^2)^{1/3}$$

(turns out to be not critical)
Example results

level 50
final est. = 2.617e-44
4 bnd.l., 1e5 repl.

level 100
final est. = 3.883e-93
5 bnd.l., 1e5 repl.

level 100
3.959e-93
5 b.l., 1e6 repl.

UT/EWI/DACS/PTdB 2008
Concluding remarks

- Non-parametric smoothing works better than the old spline smoothing.
- Non-parametric smoothing is computationally feasible.
- Perhaps non-parametric smoothing is also useful in other CE problems with many parameters?