

Efficient Approximation of Response Time Densities and Quantiles in Stochastic Models

Susanna W.M. Au-Yeung, Nicholas J. Dingle and William J. Knottenbelt
Department of Computing, Imperial College London
South Kensington Campus, London SW7 2AZ, United Kingdom
{swa02,njd200,wjk}@doc.ic.ac.uk

ABSTRACT

Response time densities and quantiles are important performance and quality of service metrics, but their analytical derivation is, in general, very expensive. This paper presents a technique for determining approximate response time densities in Markov and semi-Markov stochastic models that requires two orders of magnitude less computation than exact Laplace transform-based techniques. The method computes the first four moments of the desired response time and then makes use of Generalised Lambda Distributions to obtain an approximation of the corresponding density. Numerical results show good agreement over a range of response time curves, particularly for those that are unimodal.

1. INTRODUCTION

Software and hardware system architects are increasingly required to consider response time guarantees as key quality of service metrics. Indeed, response time quantiles are routinely specified in Service Level Agreements (SLAs) and it is therefore important to quantify the risk of violating response time targets. Analytical methods based on the numerical inversion of Laplace transforms have recently been developed to extract response time densities and quantiles from high-level stochastic modelling formalisms based on Markov and semi-Markov chains [7, 5, 3]. However, these methods are computationally expensive, and large models require the availability of a cluster of workstations to calculate results in reasonable time.

This paper presents a low-cost technique which seeks to approximate response time densities and quantiles from the corresponding first four moments. In contrast to the exact technique which requires the solution of many hundreds of sets of linear equations, calculation of the moments requires the solution of just four sets. In both cases the dimension of the linear equations is given by the number of states in the model. A Generalised Lambda Distribution (GLD) is then fitted to these moments as an approximation to the exact density. Fitting the moments is a rapid operation and its complexity does not depend on the number of states. The corresponding cumulative distribution function (used to determine response time quantiles) is then obtained by numerical integration.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

WOSP '04, January 14-16, 2004, Redwood City, CA.
Copyright 2004 ACM 1-58113-673-0/04/0001 ...\$5.00.

The rest of this paper is organised as follows. Section 2 outlines the computation of response time densities and their corresponding moments in Markov and semi-Markov models. Section 3 introduces Generalised Lambda Distributions and describes how a GLD is fitted to a given set of moments. Section 4 compares exact and approximate response time densities for a range of models. Section 5 concludes and discusses future work.

2. RESPONSE TIME ANALYSIS

2.1 Semi-Markov Processes

Consider a Markov renewal process [9] $\{(X_n, T_n) : n \geq 0\}$ where T_n is the time of the n th transition ($T_0 = 0$) and $X_n \in \mathcal{S}$ is the state at the n th transition. Let the kernel of this process be:

$$R(n, i, j, t) = \mathbb{P}(X_{n+1} = j, T_{n+1} - T_n \leq t \mid X_n = i)$$

for $i, j \in \mathcal{S}$. The continuous time semi-Markov process (SMP), $\{Z(t), t \geq 0\}$, defined by the kernel R , is related to the Markov renewal process by:

$$Z(t) = X_{N(t)}$$

where $N(t) = \max\{n : T_n \leq t\}$, i.e. the number of state transitions that have taken place by time t . Thus $Z(t)$ represents the state of the system at time t . We consider time-homogeneous SMPs, in which $R(n, i, j, t)$ is independent of any previous state except the last. Thus R becomes independent of n :

$$\begin{aligned} R(i, j, t) &= \mathbb{P}(X_{n+1} = j, T_{n+1} - T_n \leq t \mid X_n = i) \\ &= p_{ij} H_{ij}(t) \end{aligned}$$

where $p_{ij} = \mathbb{P}(X_{n+1} = j \mid X_n = i)$ is the state transition probability between states i and j and $H_{ij}(t) = \mathbb{P}(T_{n+1} - T_n \leq t \mid X_{n+1} = j, X_n = i)$, is the sojourn time distribution in state i when the next state is j .

2.2 First passage times

Consider a finite, irreducible, continuous-time semi-Markov process with N states $\{1, 2, \dots, N\}$. Recalling that $Z(t)$ denotes the state of the SMP at time t ($t \geq 0$), the first passage time from a source state i at time t into a non-empty set of target states \vec{j} is:

$$P_{i\vec{j}}(t) = \inf\{u > 0 : Z(t+u) \in \vec{j}, N(t+u) > N(t), Z(t) = i\}$$

For a stationary time-homogeneous SMP, $P_{i\vec{j}}(t)$ is independent of t and we have:

$$P_{i\vec{j}} = \inf\{u > 0 : Z(u) \in \vec{j}, N(u) > 0, Z(0) = i\}$$

$P_{i\vec{j}}$ has an associated probability density function $f_{i\vec{j}}(t)$. In general, the Laplace transform of $f_{i\vec{j}}$, $L_{i\vec{j}}(s)$, can be computed by

solving a set of N linear equations:

$$L_{i\vec{j}}(s) = \sum_{k \notin \vec{j}} r_{ik}^*(s) L_{k\vec{j}}(s) + \sum_{k \in \vec{j}} r_{ik}^*(s) \quad : \text{ for } 1 \leq i \leq N \quad (1)$$

where $r_{ik}^*(s)$ is the Laplace-Stieltjes transform (LST) of $R(i, k, t)$ from Section 2.1 and is defined by:

$$r_{ik}^*(s) = \int_0^\infty e^{-st} dR(i, k, t)$$

When there are multiple source states, denoted by the vector \vec{i} , the Laplace transform of the passage time density at steady-state is:

$$L_{\vec{i}\vec{j}}(s) = \sum_{k \in \vec{i}} \alpha_k L_{k\vec{j}}(s)$$

where the weight α_k is the probability at equilibrium that the system is in state $k \in \vec{i}$ at the starting instant of the passage. If $\vec{\pi}$ denotes the steady-state vector of the embedded discrete-time Markov chain (DTMC) with one-step transition probability matrix $P = [p_{ij}, 1 \leq i, j \leq N]$, then α_k is given by:

$$\alpha_k = \begin{cases} \vec{\pi}_k / \sum_{j \in \vec{i}} \vec{\pi}_j & \text{if } k \in \vec{i} \\ 0 & \text{otherwise} \end{cases}$$

2.3 Moments

Let $M_{i\vec{j}}(n)$ denote the n th moment of the first passage time between a given source state i and set of target states \vec{j} , and let $m_{ik}(n)$ denote the n th moment of the holding time in state i with next state k . Assuming the derivatives of $r_{ik}^*(s)$ exist at the origin, we have

$$m_{ik}(n) = (-1)^n \left. \frac{d^n r_{ik}^*(s)}{ds^n} \right|_{s=0}$$

Hence, using Leibnitz' rule,

$$\begin{aligned} M_{i\vec{j}}(n) &= \sum_{k \notin \vec{j}} \sum_{r=0}^n \binom{n}{r} m_{ik}(r) M_{k\vec{j}}(n-r) + \sum_{k \in \vec{j}} m_{ik}(n) \\ &= \sum_{k \notin \vec{j}} \sum_{r=1}^n \binom{n}{r} m_{ik}(r) M_{k\vec{j}}(n-r) + \\ &\quad \sum_{k \notin \vec{j}} p_{ik} M_{k\vec{j}}(n) + \sum_{k \in \vec{j}} m_{ik}(n) \end{aligned} \quad (2)$$

for $i \notin \vec{j}$ and $M_{i\vec{j}}(n) = 0$ for $i \in \vec{j}$, where $p_{ik} = r_{ik}^*(0) \equiv m_{ik}(0)$. The first and third terms on the right hand side will be known prior to an iteration, facilitating a straightforward iteration that solves a set of linear equations at each step.

For a Markov chain with generator matrix Q , Eq. 2 reduces to:

$$-q_{ii} M_{i\vec{j}}(n) = \sum_{k \notin \vec{j}} q_{ik} M_{k\vec{j}}(n) + n M_{i\vec{j}}(n-1) \quad (3)$$

for $i \notin \vec{j}$ and $M_{i\vec{j}}(n) = 0$ for $i \in \vec{j}$. For $n = 0$, we have $M_{i\vec{j}}(0) = 1$ and so each set of moments can be computed iteratively.

3. GENERALISED LAMBDA DISTRIBUTION

3.1 Description

The Generalised Lambda Distribution (GLD) is a family of curves which has ability to assume a wide variety of shapes including the standard distribution types exponential, normal, χ^2 , uniform, log-normal etc. Because of this flexibility, GLDs have been extensively

used to fit and model continuous probability distributions in diverse application areas such as meteorology, medical trials, financial data modelling and Monte Carlo simulation studies [8].

A GLD is defined as an inverse cumulative distribution (quantile) function $F^{-1}(u)$ (where u takes values between 0 and 1) that yields the value of x such that $F(x) = u$. It has form:

$$F^{-1}(u) \equiv Q_{\lambda_1, \lambda_2, \lambda_3, \lambda_4}(u) \quad (4)$$

where λ_1 the location parameter, λ_2 is the scale parameter and λ_3 and λ_4 are the shape parameters. If $\lambda_3 = \lambda_4$ then the distribution is symmetric. The function Q can take one of two forms, both of which are multi-parameter generalisations of the one-parameter Tukey-Lambda distribution. For notational simplicity in what follows we will omit the λ subscripts and simply write $Q(u)$.

Using the relationships $Q(u) = x$ and $F(x) = u$ and Eq. (4), the probability density function $f(x)$ may be derived as:

$$f(x) = \frac{du}{dx} = \frac{du}{dQ(u)} = \left(\frac{dQ(u)}{du} \right)^{-1} \quad (5)$$

A plot of the density function $f(x)$ can thus be obtained parametrically by plotting $Q(u)$ against $f(Q(u))$ for $0 \leq u \leq 1$.

As we will be fitting the GLD by moments, we note that the k th raw moment of a quantile function $Q(u)$ is:

$$\begin{aligned} E[X^k] &= \int_0^\infty x^k f(x) dx \\ &= \int_0^1 (Q(u))^k \frac{du}{dQ(u)} dQ(u) \\ &= \int_0^1 (Q(u))^k du \end{aligned} \quad (6)$$

3.2 Parameterization

As mentioned, the function Q in Eq. (4) can take on one of two forms. In the original Ramberg-Schmeiser (RS) [15] parameterisation,

$$Q(u) = \lambda_1 + \frac{u^{\lambda_3} - (1-u)^{\lambda_4}}{\lambda_2}$$

However, this parameterisation does not result in a well defined pdf for certain values of λ_3 and λ_4 [8]. This limitation can be partially overcome by introducing Generalised Beta Distributions (GBDs) to extend the defined area [8]. The later FMKL [6] parameterisation due to Freimer *et al* defines

$$Q(u) = \lambda_1 + \frac{1}{\lambda_2} \left(\frac{u^{\lambda_3} - 1}{\lambda_3} - \frac{(1-u)^{\lambda_4} - 1}{\lambda_4} \right) \quad (7)$$

which is well defined over the entire λ_3, λ_4 plane. For this reason, we adopt this FMKL parameterisation. Using Eq. (5) and Eq. (7) we have:

$$f(Q(u)) = \frac{\lambda_2}{u^{\lambda_3-1} + (1-u)^{\lambda_4-1}}$$

3.3 Fitting via moment matching

We wish to find GLD parameters $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ such that the mean μ , variance σ^2 , skewness α_3 and kurtosis α_4 of the GLD correspond to a given mean $\hat{\mu}$, variance $\hat{\sigma}^2$, skewness $\hat{\alpha}_3$ and kurtosis $\hat{\alpha}_4$. Matching these four measures of distribution is adequate to determine $\lambda_1, \lambda_2, \lambda_3$ and λ_4 .

First, we need to determine the central moments of the quantile

function $Q(u)$. Eq. (7) can be expanded as [10]:

$$\begin{aligned} Q(u) &= \left(\lambda_1 - \frac{1}{\lambda_2 \lambda_3} + \frac{1}{\lambda_2 \lambda_4} + \frac{1}{\lambda_2} \left(\frac{u^{\lambda_3}}{\lambda_3} - \frac{(1-u)^{\lambda_4}}{\lambda_4} \right) \right) \\ &= a + bR(u) \end{aligned} \quad (8)$$

where

$$R(u) = \left(\frac{u^{\lambda_3}}{\lambda_3} - \frac{(1-u)^{\lambda_4}}{\lambda_4} \right)$$

Let \hat{q}_k denote the k th central moment of $Q(u)$ and r_k the k th raw moment of $R(u)$. Then, from Eq. (8), the first four central moments of $Q(u)$, can be expressed in terms of the raw moments of $R(u)$ as:

$$\begin{aligned} \hat{q}_1 &= \lambda_1 - 1/(\lambda_2 \lambda_3) + 1/(\lambda_2 \lambda_4) + r_1/\lambda_2 \\ \hat{q}_2 &= \frac{1}{\lambda_2^2} (r_2 - r_1^2) \\ \hat{q}_3 &= \frac{1}{\lambda_2^3} (r_3 - 3r_1 r_2 + 2r_1^3) \\ \hat{q}_4 &= \frac{1}{\lambda_2^4} (r_4 - 4r_1 r_3 + 6r_1^2 r_2 - 3r_1^4) \end{aligned} \quad (9)$$

From Eq. (6), r_k is given by:

$$r_k = \int_0^1 \left(\frac{u^{\lambda_3}}{\lambda_3} - \frac{(1-u)^{\lambda_4}}{\lambda_4} \right)^k du$$

By binomial expansion on r_k , we have [10]:

$$\begin{aligned} r_k &= \int_0^1 \sum_{j=0}^k \binom{k}{j} (-1)^j \frac{u^{\lambda_3(k-j)}}{\lambda_3^{k-j}} \frac{(1-u)^{\lambda_4 j}}{\lambda_4^j} du \\ &= \sum_{j=0}^k \frac{(-1)^j}{\lambda_3^{k-j} \lambda_4^j} \binom{k}{j} \beta(\lambda_3(k-j) + 1, \lambda_4 j + 1) \end{aligned} \quad (10)$$

where

$$\beta(a, b) = \int_0^1 u^{a-1} (1-u)^{b-1} du$$

The beta function is defined only for positive arguments, so (in common with all GLD moment matching methods) we require

$$\min(\lambda_3, \lambda_4) > -\frac{1}{4}.$$

Now, from Eq. (9) and Eq. (10), it follows that the skewness $\alpha_3 \equiv \hat{q}_3/\hat{q}_2^{3/2}$ and and kurtosis $\alpha_4 \equiv \hat{q}_4/\hat{q}_2^2$ are functions of λ_3 and λ_4 only. By solving the set of two simultaneous non-linear equations

$$\begin{aligned} \alpha_3 &= \hat{\alpha}_3 \\ \alpha_4 &= \hat{\alpha}_4 \end{aligned} \quad (11)$$

we obtain values for λ_3, λ_4 . Then λ_2 and λ_1 are computed as:

$$\begin{aligned} \lambda_2 &= \frac{\sqrt{r_2 - r_1^2}}{\hat{\sigma}} \\ \lambda_1 &= \hat{\mu} + \frac{1}{\lambda_2} \left(\frac{1}{\lambda_3 + 1} - \frac{1}{\lambda_4 + 1} \right) \end{aligned}$$

The non-linear equations of Eq. (11) do not have a closed-form solution. However, it is possible to apply numerical methods such as the Nelder-Mead simplex method [11, 13] and Powell's method [12]. Computer software libraries which implement these methods and which perform multi-variable optimization are available; for our results we used Besset's Java library [1]. Corresponding cdfs can be obtained by numerical integration of the pdf (e.g. using the

Trapezoidal rule or Simpson's rule). Note that, since the system of non-linear equations is small and of fixed size irrespective of the number of states in the model, fitting the moments is typically very rapid (cf. Section 4.5).

4. NUMERICAL RESULTS

This section presents results obtained from a variety of Markov and semi-Markov models. In each case, we extract a response time probability density function (pdf) and corresponding cumulative distribution function (cdf) using the exact Laplace transform approach of Section 2.2 [7, 5, 3], and compare it to the approximations calculated using the GLD approach outlined above. We also present timing results and compare the quality of the results with those provided by the WinMoments tool [14].

4.1 GSPN models

We first apply our technique to two GSPN models. The Courier model is a 45-place GSPN representing the ISO Application, Session and Transport layers of a sliding-window communication protocol (see [16] for full details). The response time of interest is the time taken from the start of a transport-layer send to the arrival of the corresponding acknowledgement. The FMS model is a 22-place GSPN representing an assembly line composed of three types of machines and four types of parts (see [4] for full details). Starting with 4 unprocessed parts of types 1, 2 and 3, we are interested in measuring the time to complete the first processed part of type 4. Fig. 1 shows the exact response time pdf and cdf, calculated using the Laplace-transform based technique of [7], as well as the approximate GLD pdf and cdf for the Courier (left) and FMS (right) models. In both cases, we observe good agreement between the approximate and exact pdfs, and excellent agreement between the approximate and exact cdfs. The latter is particularly useful for accurately estimating response time quantiles.

4.2 Queueing network model

We now apply our method to approximate a cycle time density for a path in a closed tree-like network with 8 customers shown in Fig. 2 (left) (see [7] for full details). The cycle time of interest is measured from when a tagged customer arrives at the back of the first queue, and ends when the customer returns to the queue. Fig. 2 (right) shows moderate agreement between the exact and approximate pdfs but excellent agreement between the cdfs.

4.3 Bimodal models

To test the ability of the GLD method to approximate response time densities that are not unimodal, we show results for the cycle time in a branching Erlang model (see [7]). This model is composed of two equiprobable branches, one of which results in an Erlang(3, λ_1) delay, and the other of which results in a Erlang(12, λ_2) delay. Setting $\lambda_1 = \lambda_2 = 1$, we obtain a bimodal density curve, as shown on the left in Fig. 3. The GLD approximation for the pdf does not capture its bimodal nature; however the cdf still shows good agreement. Setting $\lambda_1 = 1$ and $\lambda_2 = 2$, we obtain the almost unimodal curve shown on the right in Fig. 3. The GLD approximation now shows a much better fit (for both pdf and cdf).

4.4 SMSPN models

Moving on to semi-Markov Stochastic Petri net models (see [2] for details of this formalism), we consider a model of a web content authoring service (see [3] for more details). Authors publish content on a number of web servers; there is also a pool of readers who submit requests to the servers to be provided with content. The servers are unreliable and can fail and then recover. A system of

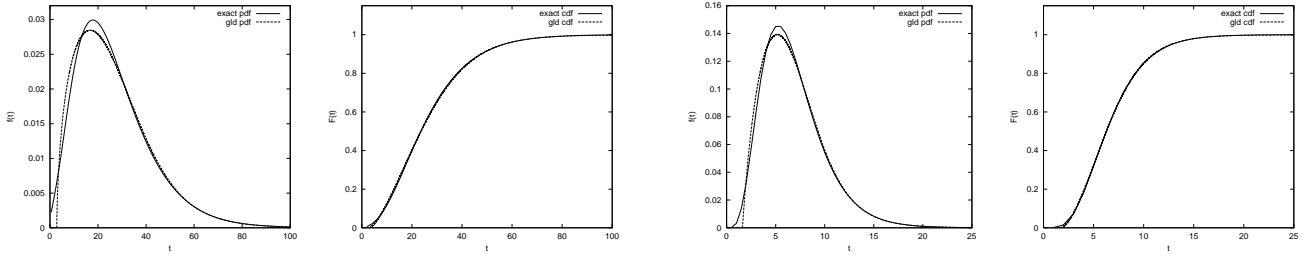


Figure 1: Courier GSPN-model pdf and corresponding cdf (left) and FMS GSPN-model pdf and corresponding cdf (right)

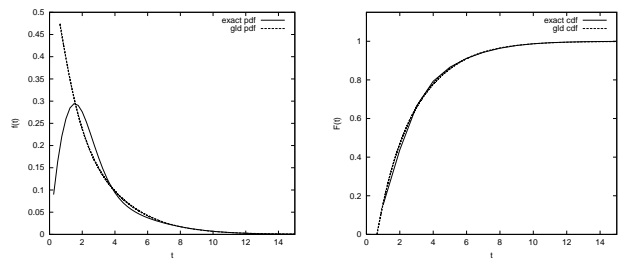
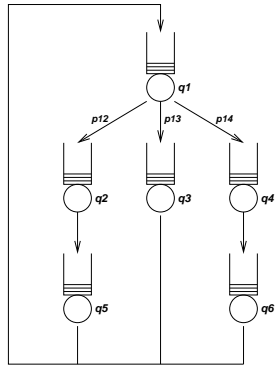


Figure 2: Tree-like queuing network (left) with cycle time pdf and corresponding cdf (right)

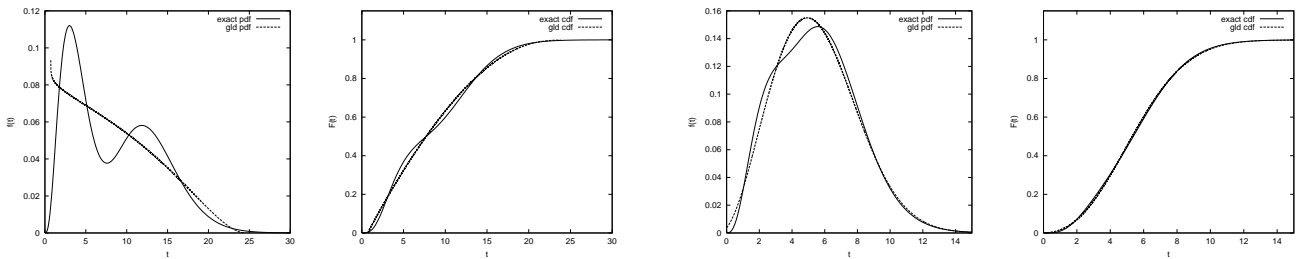


Figure 3: Bimodal branching Erlang pdf and corresponding cdf (left) and almost unimodal branching Erlang pdf and corresponding cdf (right)

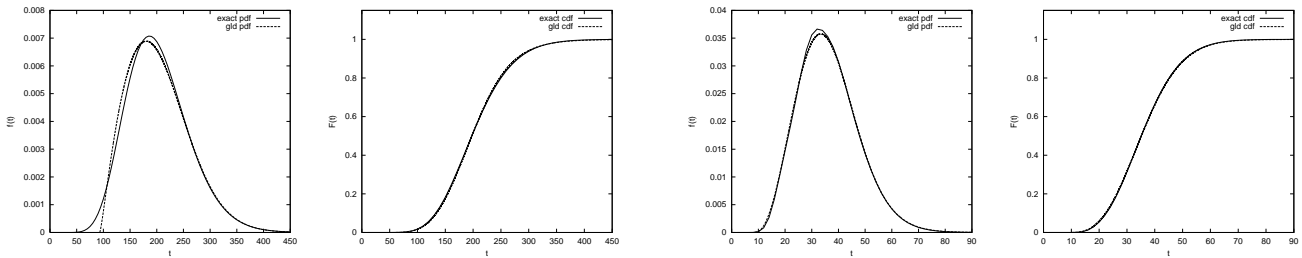


Figure 4: Webserver SMSPN-model Response Time 1 pdf and corresponding cdf (left) and Webserver SMSPN-model Response Time 2 pdf and corresponding cdf (right)

12 writers and 24 readers yields a 15 257 state semi-Markov chain. Fig. 4 (left) shows the distribution of time taken for all writers to commit their updates and all readers to receive their requested content, while Fig. 4 (right) represents the time taken only for the readers to receive their requested content. The exact distributions were calculated using the iterative algorithm presented in [3].

4.5 Timing results

The table below compares the time (in seconds) taken to calculate the above pdfs using the GLD method and the exact Laplace transform-based method.

Model	States	Moment Calc.	Moment Matching	GLD Total	Laplace Total
courier	11 700	0.66	0.28	0.94	134
fms	35 910	7.64	0.42	8.06	834
tree	20 592	2.71	0.61	3.32	366
bi-erlang	32	0.05	0.27	0.32	14.6
uni-erlang	32	0.02	0.12	0.14	13.2
webserver-1	15 257	3.62	0.37	3.99	2374
webserver-2	15 257	1.91	1.08	2.99	1078

4.6 Comparison with WinMoments tool

Finally, Fig. 5 compares the cdf approximations produced by the GLD method with the upper and lower bounds computed by Rácz's WinMoments tool [14] for the Webserver Response Time 2 cdf and the first branching Erlang cdf. Given a finite number of moments and a t point, WinMoments calculates upper and lower bounds on the value of $F(t)$. In both cases, the GLD approximation lies well within the WinMoment-calculated bounds, and provides a better approximation to the actual cdf than the mid-point of the bounds.

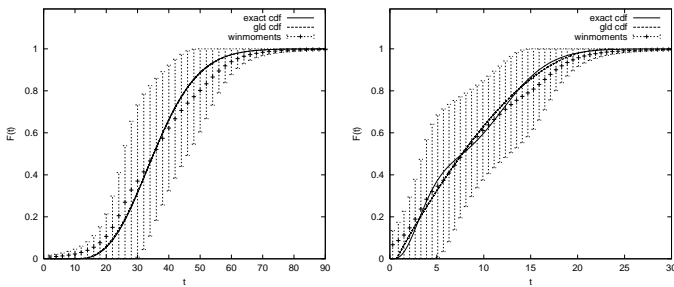


Figure 5: Webserver SMSPN-model Response Time 2 cdf (left) and first branching Erlang cdf (right) compared with the bounds generated by the WinMoments tool.

5. CONCLUSIONS

We have conducted a study into the rapid approximation of response time densities and quantiles in Markov and semi-Markov models using Generalised Lambda Distributions. The results presented demonstrate that this method provides a good estimation of pdfs and excellent estimation of cdfs from the first four moments of response time, most notably where the response time densities are unimodal. The approximations produced compare favourably with the bounds generated by the WinMoments tool.

The GLD-based estimation technique offers significant predictive insights at low cost when compared to an exact Laplace transform-based approach. The exact technique requires the solution of a large number of systems of linear equations (typically greater than 400), the complexity of which is a function of the number of states in the stochastic model. The approximation technique presented here,

however, requires the solution of only four sets of these linear equations in order to calculate the first four moments of the response time distribution, plus the time taken to use these moments to perform the estimation (which is independent of the number of states). As future work, we intend to investigate the sensitivity of the GLD method to perturbations in the moments. For bimodal densities, it may be possible to improve the accuracy of the GLD method by considering them as the superposition of two unimodal densities.

6. ACKNOWLEDGEMENTS

We would like to thank Pete Harrison for his helpful comments.

7. REFERENCES

- [1] D.H. Besset. *Object-Oriented Implementation of Numerical Methods: An Introduction with Java and Smalltalk*. Morgan Kaufmann, 2001.
- [2] J.T. Bradley, N.J. Dingle, W.J. Knottenbelt, and P.G. Harrison. Performance queries on semi-Markov stochastic Petri nets with an extended Continuous Stochastic Logic. In *Proc. Petri Nets and Performance Models (PNPM'03)*, pages 62–71, 2003.
- [3] J.T. Bradley, N.J. Dingle, W.J. Knottenbelt, and H.J. Wilson. Hypergraph-based parallel computation of passage time densities in large semi-Markov models. In *Proc. 4th Int. Meet. on Numerical Solution of Markov Chains (NSMC '03)*, pages 99–120, 2003.
- [4] G. Ciardo and K.S. Trivedi. A decomposition approach for stochastic reward net models. *Performance Evaluation*, 18(1):37–59, 1993.
- [5] N.J. Dingle, P.G. Harrison, and W.J. Knottenbelt. Response time densities in Generalised Stochastic Petri Net models. In *Proc. 3rd Int. Workshop on Software and Performance (WOSP 2002)*, pages 46–54, 2002.
- [6] M. Freimer, G. Mudholkar, G. Kollia, and C. Lin. A study of the generalized Tukey Lambda family. *Communications in Statistics: Theory and Methods*, 17(10):3547–3567, 1988.
- [7] P.G. Harrison and W.J. Knottenbelt. Passage time distributions in large Markov chains. In *Proc. ACM SIGMETRICS 2002*, Marina Del Rey, California, June 2002.
- [8] Z. Karian and E. Dudewicz. *Fitting Statistical Distributions: The Generalized Lambda Distribution and Generalized Bootstrap Methods*. CRC Press, Boca Raton, 2000.
- [9] V.G. Kulkarni. *Modeling and Analysis of Stochastic Systems*. Chapman & Hall, 1995.
- [10] A. Lakhany and H. Mausser. Estimating the parameters of the Generalized Lambda Distribution. *Algo Research Quarterly*, 3(3):47–58, 2000.
- [11] J.A. Nelder and R. Mead. A simplex method for function minimization. *Computer Journal*, 7:308–313, 1965.
- [12] M.J.D. Powell. An iterative method for finding stationary values of a function of several variables. *Computer Journal*, 5(2):147–151, 1962.
- [13] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, 2nd edition, 1993.
- [14] S. Rácz. *Numerical Analysis of Communication Systems through Markov Reward Models*. PhD thesis, Budapest University of Technology and Economics, 2002.
- [15] J.S. Ramberg and B.W. Schmeiser. An approximate method for generating asymmetric random variables. *Communications of the ACM*, 17(2):78–82, 1974.
- [16] C.M. Woodside and Y. Li. Performance Petri net analysis of communication protocol software by delay-equivalent aggregation. In *Proc. 4th Int. Workshop on Petri nets and Performance Models (PNPM '91)*, pages 64–73, 1991.