Skewness Variance Approximation for Dynamic Rate Queues

Jamol Pender
Department of Operations Research and Financial Engineering
Princeton University
jpender@princeton.edu
William A. Massey
ORFE Department, Princeton University

May 16, 2011

Abstract

A fundamental dynamic rate queueing model for large scale service systems is a multi-server queue with non-homogeneous Poisson arrivals and customer abandonment. By scaling the arrival rates and number of servers of such systems, using the fluid and diffusion limit theorems found in Mandelbaum, Massey, and Reiman [2], we can approximate the stochastic behavior of this queueing process by one that is Gaussian. Moreover, the approximations to the mean and variance produced by these limiting processes form a two-dimensional dynamical system. Recent work by Gautam and Ko [1] found a modified version of these differential equations and obtained a better estimates of the mean and variance for the original queueing system. In this paper, we introduce a new three-dimensional dynamical system that surpasses these two approaches.
Keywords: Queues, Delay, Staffing, Skewness, Hermite Polynomials.

Large scale systems such as customer contact centers, like healthcare centers, like hospitals, have customer inflow-outflow dynamics with many common features. First, the customer arrival patterns may have time of day or seasonal effects. Moreover, customer population sizes tend to be large where the individual actions are intrinsic and independent of other customer actions. Second, there are multiple service agents, so many customers have access to services in parallel. In a call center, for example, these agents may be the telephone operators. For a hospital, these agents may be the hospital beds or they may be nurses that attend to the patient body. In both of these examples, arriving customers wanting to engage in service may be delayed if all the available agents are busy. Moreover, these waiting customers may decide to leave the system if they feel that their delay in receiving service is excessively long.

To model these large scale systems, we model the hospital or health care center using a Markovian queueing system with abandonment. One of the main contributions of the paper is to expand our queueing process in terms of a finite number of Hermite polynomials. This would give the true distribution if we used an infinite expansion of Hermite polynomials. The second contribution is to use the forward equations and the distribution of the finite expansion of Hermite polynomials to construct a set of autonomous differential equations that describe the stochastic dynamics of our queueing process. We show that the first order expansion yields a deterministic process which turns out to be equivalent to the the fluid limit derived from the strong law of large numbers limit by Mandelbaum, Massey, and Reiman [2]. A second order expansion, the Gaussian variance approximation (GVA), yields a Gaussian approximation which yields mean and variance equations derived by Gautum and Ko [1]. Finally our third order expansion, the Gaussian skewness approximation (GSA), yields an autonomous set of differential equations involving the mean, variance, and skewness of the queueing process. Figure 1 shows that by adding the skewness to our model or by adding a third Hermite polynomial to the expansion, we are able to accurately approximate the true empirical distribution of the queueing process. Figure 1 also illustrates that GSA has asymmetric tails so it is able to fit both distributional tails of the queue much better than other methods. We will also show by adding the skewness to our model allows us to predict precise delay probabilities for our queueing system. The ability to predict these delay probabilities with excellent precision is extremely important for hospitals and healthcare centers, as understanding the precise behavior of delay probabilities allows one to staff the hospitals with the appropriate number of nurses and beds so that patients are not waiting excessively to receive quality care. Thus, in this paper we show that it is necessary to understand beyond the mean and variance of our queueing process. By exploring other moments such as the skewness, we are able to understand the non-Gaussian behavior of large scale systems not predicted from heavy traffic fluid and diffusion limits.

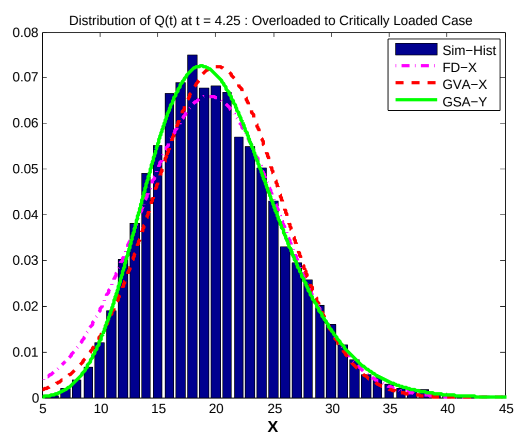

Figure 1: Empirical Distribution of Queueing Process

References

[1] Ko, Y. M. and Gautam, N. (2010). Critically loaded multi-server queues with abandonment, retrials, and time-varying parameters. *Working Paper*

[2] A. Mandelbaum , W.A Massey, M. Reiman (1998). Strong Approximations for Markovian Service Networks. *Queueing Systems*, 30 (1998) pp. 149-201.