Basic Theory and Some Applications of Martingales

Richard A. Hayden
Department of Computing
Imperial College London
London, SW7 2AZ, United Kingdom
rh@doc.ic.ac.uk

ABSTRACT

This tutorial surveys the fundamental results of the theory of martingales from the perspective of the performance engineer. We will present the fundamental results and illustrate their power through simple and elegant proofs of important and well-known results in performance analysis. The remainder of the tutorial will introduce the martingale functional central limit theorem and semi-martingale decomposition methodology for the characterisation and proof of heavy-traffic limit results for Markovian queueing systems.

Categories and Subject Descriptors

G.3 [**Probability and Statistics**]: Markov processes, Queueing theory, Stochastic processes

Keywords

stochastic processes, queueing theory, martingales, heavy-traffic

1. OVERVIEW

A stochastic process $X = \{X(t)\}_{t \geq 0}$ is a *martingale*[1] if:

$$\mathbb{E}[X(t+s)|(X(u))_{u \leq t}] = X(t) \text{ with probability } 1$$

for all $s, t \geq 0$.

Martingales constitute a very important class of stochastic processes about which very strong, and often intuitively surprising, general statements can be made. In addition to their interest from a purely mathematical point of view, they have key applications in applied probability, perhaps most crucially to the proof of so-called *heavy-traffic limits* of queueing systems [9, 5], and, relatedly, *mean-field* or *fluid* limits in other settings [3, 1, 6].

The aim of this tutorial is to provide an accessible introduction to the theory of martingales for the non-expert, with a specific emphasis on applications of interest to practitioners in the performance analysis community.

We will first cover the basic definitions and fundamental results, including the *optional stopping theorem*, the *maximal inequality*, the *upcrossings inequality* and the *martingale convergence theorems*. We will illustrate the practical power of even these basic results by showing how they yield simple and elegant proofs of some of the fundamental results of performance modelling, for example, the famous *PASTA (Poisson Arrivals See Time Averages)* property [12] and the fact that the best phase-type approximation to a deterministic delay is Erlang [4].

In the second part of the tutorial, our focus will be on the powerful martingale approach for establishing heavy-traffic limits of queueing networks [9, 10, 8, 11, 7]. This will rest on the decomposition of the components of the queueing processes into sums of martingales and *predictable*[2] processes according to a version of the *Doob–Meyer decomposition theorem*:

Theorem 1 (D–M decomp. of submarts. in $\mathbb{R}_{\geq 0}$ [9]). *If X is a submartingale with paths in $\mathbb{R}_{\geq 0}$ and $\mathbb{E}[X(t)] < \infty$ for each $t \geq 0$, then there exists a predictable process A, called the* compensator *of X, such that A has non-negative, non-decreasing sample paths, $\mathbb{E}[A(t)] < \infty$ for each $t \geq 0$ and $M := X - A$ is a martingale. The compensator is unique in the sense that the sample paths of any two versions must be equal with probability 1.*

This will then allow application of the *martingale functional central limit theorem (FCLT)*, which concerns the convergence of appropriate martingales to (multi-dimensional) Brownian motion:

Theorem 2 (Martingale FCLT [9]). *For $n \geq 1$, let $\mathbf{M}^n$ be a sequence of (local) martingales in D^k with $\mathbf{M}_0^n = \mathbf{0}$.*[3] *Let $\mathbf{C} = (c_{ij})$ be a $k \times k$ non-negative-definite symmetric real matrix. Assume further that each $\mathbf{M}^n$ is (locally) square integrable so that the* predictable quadratic co-variation *processes:*

$$\langle M_i^n, M_j^n \rangle := \frac{1}{4}(\langle M_i^n + M_j^n \rangle - \langle M_i^n - M_j^n \rangle)$$

can be defined, where the predictable quadratic variation $\langle M \rangle$ of a martingale M is defined to be the unique compensator of the non-negative submartingale M^2 given in Theorem 1.

Then if the expected values of the maximum jump of $\langle M_i^n, M_j^n \rangle$ and maximum squared jump of $\mathbf{M}^n$ are asymptotically negligible as $n \to \infty$, and $\langle M_i^n, M_j^n \rangle(t) \Rightarrow c_{ij}t$ as $n \to \infty$ for each $t \geq 0$, we have:

$$\mathbf{M}^n \Rightarrow \mathbf{B} \text{ weakly in } D^k \text{ as } n \to \infty$$

[1] Or a *submartingale* if equality is replaced with $\geq$ in the definition.

[2] The value of a predictable process at time t is known given the information about the process up to (but not including) time t.

[3] D^k is the space of right-continuous functions with left limits taking values in $\mathbb{R}^k$.

SIGMETRICS'12, June 11–15, 2012, London, England, UK.
ACM 978-1-4503-1097-0/12/06.

where $\mathbf{B}$ *is a* k*-dimensional Brownian motion having mean vector* $\mathbb{E}[\mathbf{B}(t)] = \mathbf{0}$ *and covariance matrix* $\mathbb{E}[\mathbf{B}(t)\mathbf{B}(t)^T] = \mathbf{C}t$ *for* $t \geq 0$.

More concretely, in this tutorial we consider sequences of Markovian many-server queues which can be described in terms of *thinnings* of (possibly infinite) sets of Poisson processes [2]. The characterisation and proof of heavy-traffic limits then follows approximately the following recipe:

1. Decomposition of the queueing processes into sums of martingales and predictable processes (*semi-martingale decomposition*). Proof of the representation by exploiting properties of stochastic integrals such as those in [2, Page 10].

2. Verification of conditions of the martingale FCLT: proof of convergence of the *predictable quadratic variation* — usually relies on an auxiliary fluid limit result that follows from an appropriate stochastic boundedness lemma. Again, martingale inequalities are the powerful tools of choice for establishing these boundedness results.

3. Convergence to the limit process then follows by combining the FCLT with the *continuous mapping theorem (CMT)*.

We will illustrate this approach with examples taken from the literature.

2. BRIEF SPEAKER BIOGRAPHY

Richard Hayden obtained his MSci degree in mathematics and computer science from the Department of Computing at Imperial College London in 2007. He received his Ph.D. in 2011, for which he was a finalist for the 2012 INFORMS Doctoral Dissertation Award for Operations Research in Telecommunications. He currently holds a postdoctoral position at Imperial College and his research interests lie in the efficient analysis and optimisation of large-scale performance models of massively-parallel systems using mean-field analysis and related asymptotic techniques.

3. REFERENCES

[1] M. Benaïm and J.-Y. Le Boudec. A class of mean field interaction models for computer and communication systems. *Performance Evaluation*, 65(11-12):823–838, Nov. 2008.

[2] P. Brémaud. *Point processes and queues: martingale dynamics*. Springer, 1981.

[3] R. W. R. Darling and J. R. Norris. Differential equation approximations for Markov chains. *Probability Surveys*, 5:37–79, 2008.

[4] A. David and S. Larry. The least variable phase type distribution is Erlang. *Stochastic Models*, 3(3):467–473, 1987.

[5] S. Halfin and W. Whitt. Heavy-Traffic Limits for Queues with Many Exponential Servers. *Operations Research*, 29(3):567–588, May 1981.

[6] R. A. Hayden, A. Stefanek, and J. T. Bradley. Fluid computation of passage time distributions in large Markov models. *Theoretical Computer Science*, 413(1):106–141, 2012.

[7] H. Kaspi and K. Ramanan. Law of large numbers limits for many-server queues. *The Annals of Applied Probability*, 21(1):33–114, Feb. 2011.

[8] R. S. Lipster and A. N. Shiryayev. *Theory of martingales*. Kluwer, 1989.

[9] G. Pang, R. Talreja, and W. Whitt. Martingale proofs of many-server heavy-traffic limits for Markovian queues. *Probability Surveys*, 4:193–267, 2007.

[10] A. A. Puhalskii and M. I. Reiman. The multiclass GI/PH/N queue in the Halfin-Whitt regime. *Advances in Applied Probability*, 32(2):564–595, June 2000.

[11] W. Whitt. Heavy-Traffic Limits for the G/$\mathrm{H}_2^*/n/m$ Queue. *Mathematics of Operations Research*, 30(1):1–27, Feb. 2005.

[12] R. W. Wolff. Poisson Arrivals See Time Averages. *Operations Research*, 30(2):223–231, Oct. 1982.