Invited chapter in *Frontiers in Queuing: Models, Methods and Problems*, (J. Dshalalow, Ed.), CRC Press, 1995.

THE ASTA PROPERTY

BENJAMIN MELAMED
NEC USA, Inc.
4 Independence Way
Princeton, NJ 08540

DAVID D. YAO*
Columbia University
Department of IE/OR
New York, NY 10027

1 INTRODUCTION

The term "ASTA" (Arrivals See Time Averages) connotes the equality of statistics in a time series to corresponding statistics in an embedded time series, the embedding times being typically random. While ASTA can be mathematically formulated in a fairly general setting, the name, ASTA, reflects its historical origins in queueing theory [7, 26, 31].

The ASTA property in queueing context can be informally described as follows. Consider a stable queueing system and the arrival stream of discrete jobs offered to it (the underlying time parameter may be continuous or discrete, but batch arrivals are precluded for simplicity). Suppose that the state process of the system (typically queue length) has equilibrium statistics equaling the corresponding long-term time averages over sample paths; such time averages will be referred to as *time-observed* statistics. Let now each arriving job observe the state of the queueing system, but with the arriving job excluded from the observed statistic; this is sometimes referred to as the "state of the system just prior to arrival", a notion to be made precise in the sequel. System state observations embedded at job arrival times will be referred to as *job-observed* statistics. In particular, *job averages* are obtained by averaging the sequence of job-observed statistics. In this setting, the ASTA property is said to hold for the state process, if the time-observed and corresponding job-observed statistics coincide asymptotically, as the time interval and number of jobs, respectively, increase to infinity. Two related issues are the existence of these quantities, and whether the existence of one implies the existence of the other. Notice that the need to exclude the arriving job from the job-observed state stems from the simple observation that otherwise, ASTA is a priori precluded in most cases, due to the biasing effect of the arrivals; for example, arriving jobs which are allowed "to see themselves", can never observe an empty system.

Evidently, the setting of the informal discussion above can be generalized at several levels. First, job-observed statistics may be embedded at departure times, i.e., jobs observe the "state of the system just after departure", or in a queueing network context, jobs may observe the state at "traffic times", i.e., at time instants when jobs are routed from one node to another. Second, the embedding times need not be related to the physical movement of jobs; embedding can be effected at any sequence of random time points. However, for ASTA to have a chance of holding, it is generally necessary to define the embedded state, so as to remove any biasing effect introduced by the embedding sequence. And third, the state process and the random embedding times can be

*Supported in part by NSF Grant MSS-92-16490

abstracted to fairly general stochastic processes, entirely divorced from a queueing setting; any pair consisting of a real-valued stochastic process and a point process over a common probability space will do. In this general setting, the queueing terminology is modified: time-observed statistics become *time averages*, and job-observed statistics become *event averages*.

At its most general, the ASTA setting and associated ASTA problem (to be described below) may be profitably viewed simply as a *sampling problem*. To fix the ideas, we couch the discussion in engineering-like terminology. Imagine some random system (e.g., a physical device or network) whose stochastic state evolves in time (continuous or discrete), with the proviso that its state or a function thereof may be observed (sampled) only at a subset of discrete time points. Suppose that a set of state functions has been selected and a class of sampling procedures, obeying this proviso, has been devised. Usually, no other restrictions are imposed on the sampling mechanism. Admissible sampling procedures are allowed to be random or deterministic, and they may or may not interact with the state of the system under observation; the term "interaction" will be used in informal discussion in lieu of stochastic dependence. Next, form the event averages of interest for each admissible sampling procedure, and compare each of them to the corresponding time averages (note that the latter is just another sampling procedure, except that it permits continual observation of the system, without time gaps, and therefore is excluded from the admissible class). Informally, the ASTA problem seeks to answer the following question: For which state functions and admissible sampling procedures, will the limiting event averages coincide with the corresponding limiting time averages? A more refined ASTA problem would seek information on the *estimation bias* (the difference between the event average and time average): Do the event averages overestimate or underestimate the corresponding time averages, and if so, by how much?

To illustrate that the theoretical ASTA problem is of practical interest, consider a computer operating system which collects some useful statistics on its own operation (e.g., CPU utilization, execution times of jobs, etc.) For reasons of efficiency, the system is permitted to collect those statistics only at prescribed time points, and statistics collection is combined with other housekeeping tasks; otherwise, the attendant overhead may well degrade system performance. System designers would then like to know whether the event averages are good approximations of the "true" statistics, and whether the former are hopelessly biased with respect to the latter.

A qualitative first-cut analysis of the ASTA problem can be given in terms of the interaction (dependence) between the sampling procedure and the observed state. Clearly, if the sampling process is stochastically independent of the observed state process, then no bias will ensue. This case is mathematically uninteresting, and may not be feasible in practice (see, e.g., the example above). Conversely, if the sampling procedure "interacts strongly" with the observed state, then biasing effects should be fully expected. This is, in fact, what happens in most queueing systems, where the sampling procedure observes the state at job arrival times. Unlike "outside observers" which are non-intrusive, "job observers" interact with the state of the queue, since each ends up joining the queue. This example makes it clear that the aforementioned requirement that "jobs may not see themselves" is designed to lessen that interaction, which would otherwise preclude ASTA in most cases. Thus, the ASTA problem can be informally viewed as seeking to delineate the "upper limits" of dependence between the sampling procedure and the observed state, within which the ASTA condition holds true.

Remarkably, a host of ASTA conditions have been identified. The most celebrated instance of these is the PASTA Theorem, which is due to R. Wolff [31] (see also [15]). Furthermore, the general ASTA problem (lack of bias), as well as its refined variant (direction of bias), both have a surprisingly simple resolution (see Melamed and Whitt [20]). Ignoring secondary technical details, the bias can be written in terms of the sampling rate and the pointwise covariance between the observed state and the conditional intensity of the counting process which models the sampling procedure (the point process at whose arrival points the state is sampled); the sign and magnitude of the bias are proportional to the the covariance (sign and magnitude) and inversely proportional to the sampling rate. A necessary and sufficient condition for ASTA to hold is zero covariance (correlation).

The rest of this article is organized as follows. Section 2 sets up the ASTA problem formally, and provides related definitions. Section 3 presents a synopsis of ASTA results. Section 4 sketches ASTA-related proofs, using the stochastic Riemann-Stieltjes sum approach, whose merits are its intuitiveness and technical simplicity. Section 5 generalizes the setting and the results, using a martingale approach. Section 6 presents ASTA examples for queueing systems, mainly queueing networks. Finally, Section 7 presents a brief summary of the article.

2 ASTA SETTING AND PROBLEM DEFINTION

The mathematical setting of ASTA will be presented next. Let $(\Omega, \mathcal{F}, \mathsf{P})$ be the underlying probability space over which all stochastic processes will be defined. Let $X \equiv \{X(t) : \ t \geq 0\}$ be a stochastic process taking values in a complete separable metric space, E, with the associated σ-algebra, $\mathcal{E}$, generated by the open sets. Let $N \equiv \{N(t) : \ t \geq 0\}$ be a simple counting process, which is non-null and non-explosive, i.e., $0 < \mathsf{E}[N(t)] < \infty, \ t > 0$. The jump times of N form a stochastic sequence $T \equiv \{T_n : \ n \geq 0\}$, where $T_n = \inf\{t > 0 : \ N(t) \geq n\}$ for $n \geq 1$, and $T_0 = 0$ by convention (0 may not be an arrival point). The process X is referred to as a *state process* and represents the stochastic evolution of the system of interest. The counting process N is referred to as the *arrival process*, and its jump times serve as embedding times of the state process. More specifically, the *embedded (state) process* is the stochastic sequence $X^{(N)} \equiv \{X^{(N)}(n) = X(T_n) : \ n \geq 0\}$.

For technical reasons, it is customary to work with transformed state processes of the form $U \equiv \{U(t) : \ t \geq 0\}$, where $U(t) = f(X(t))$ and f is a bounded real-valued measurable function on $\mathcal{E}$. The function f is designed to extract the relevant state information from the state description of X. For example, in queueing settings, f typically yields the node or network state, excluding the observing job. For arrival streams, this is often attained by taking the left-continuous version, $U(t) = f(X(t-))$, of the original state, while for departure streams, the right-continuous version, $U(t) = f(X(t+)))$, of the original state is taken instead; for traffic streams within a queueing network, f extracts the left-continuous state of the origination node and the right-continuous state of the destination node. In particular, f can be an indicator function, so that the corresponding $U(t)$ can extract detailed stochastic information from X, whose limiting time averages are the state probabilities of X. Both X and U will be referred to as state processes (the ambiguity to be resolved by context), and similarly for the embedded state processes $X^{(N)}$ and $U^{(N)}$. We mention that the distinction between the state processes, X and U, as well as their respective embedded ones, is retained for modeling purposes.

The ASTA problem has several related variants, all of which involve various notions of stochastic convergence. Ultimately, these variants imply the ASTA property in terms of convergence in distribution (weak convergence, denoted $\Longrightarrow$); see Billingsley [3]. Equality in distribution is denoted by $\stackrel{d}{=}$. This fundamental notion of ASTA is defined next.

Definition 2.1 (The ASTA Problem)
Suppose that

$$U(t) \underset{t\to\infty}{\Longrightarrow} U(\infty) \quad \text{and} \quad U^{(N)}(n) \underset{n\to\infty}{\Longrightarrow} U^{(N)}(\infty). \tag{1}$$

The *ASTA problem* is concerned with finding sufficient and/or necessary conditions for

$$U(\infty) \stackrel{d}{=} U^{(N)}_{\infty}. \tag{2}$$

In this case, we say that *ASTA holds* for the pair (U, N). □

Other variants of the ASTA problem are designed to facilitate mathematical manipulation by imposing additional structure which implies (2) in Definition 2.1. The two main variants are described below.

The first variant, referred to as the *sample path framework*, assumes that $U(\infty)$ and $U^{(N)}(\infty)$ correspond to the unique stationary distributions of U and $U^{(N)}$, respectively. Consider the two stochastic processes, $V \equiv \{V(t): \; t \geq 0\}$ and $W \equiv \{W(t): \; t \geq 0\}$, given by

$$V(t) \;\equiv\; \frac{1}{t}\int_0^t U(s)\,ds, \quad t \geq 0, \tag{3}$$

$$W(t) \;\equiv\; \frac{1}{N(t)}\sum_{n=1}^{N(t)} U^{(N)}(n) \;=\; \frac{\int_0^t U(s)\,dN(s)}{N(t)}, \quad t \geq 0, \tag{4}$$

where $V(0) = 0$ in (3), and in (4) it is understood that $W(t) = 0$, whenever $N(t) = 0$. Observe that Eq. (3) is a time average, whereas Eq. (4) is the corresponding N-embedded event average, where the sum admits a simple stochastic Riemann-Stieltjes integral representation. The ASTA condition for the pair (U, N) then assumes the form

$$\lim_{t\to\infty}[W(t) - V(t)] \;=\; 0, \quad \text{w.p.1}, \tag{5}$$

where the existence of the limits is part of the problem.

The second variant, referred to as the *expectations framework*, was first introduced in [20]. It is formulated in terms of the following expected values (instead of random variables),

$$\bar{V}(t) \;\equiv\; \frac{1}{t}\,\mathsf{E}\Big[\int_0^t U(s)\,ds\Big], \quad t \geq 0, \tag{6}$$

$$\bar{W}(t) \;\equiv\; \frac{1}{\mathsf{E}[N(t)]}\,\mathsf{E}\left[\sum_{n=1}^{N(t)} U^{(N)}(n)\right] \;=\; \frac{\mathsf{E}\Big[\int_0^t U(s)\,dN(s)\Big]}{\mathsf{E}[N(t)]}, \quad t \geq 0, \tag{7}$$

where $\bar{V}(0) = 0$ in (6), and in (7) it is understood that $\bar{W}(t) = 0$, whenever $\mathsf{E}[N(t)] = 0$. The ASTA condition for the pair (U, N) is then stated, in terms of ordinary limits, as

$$\lim_{t\to\infty}\bar{V}(t) \;=\; \bar{V}(\infty) \;=\; \bar{W}(\infty) \;=\; \lim_{t\to\infty}\bar{W}(t), \tag{8}$$

where the existence of the limits is part of the problem. This case is of interest when we have the weak convergences,

$$V(t) \;\underset{t\to\infty}{\Longrightarrow}\; V(\infty) \;=\; \bar{V}(\infty) \;=\; \lim_{t\to\infty}\bar{V}(t), \tag{9}$$

$$W(t) \;\underset{t\to\infty}{\Longrightarrow}\; W(\infty) \;=\; \bar{W}(\infty) \;=\; \lim_{t\to\infty}\bar{W}(t), \tag{10}$$

which coincide with the previous formulation. Eqs. (9) – (10) hold in a regenerative framework or in stationary ergodic framework (see Remark 5.9). These are usually sufficiently general for most practical purposes.

3 SYNOPSIS OF ASTA RESULTS

This section presents a synopsis of ASTA results and attendant definitions. Generally, the discussion will proceed from strong sufficient conditions to weaker ones, culminating with necessary and sufficient ones. As usual, U is assumed bounded throughout.

3.1 LAA, PASTA and Anti-PASTA

We start with LAA and PASTA — the two pioneering ASTA-related properties, later refined and generalized beyond their original setting.

Definition 3.1 (LAA)
Let U be left-continuous and let N be right-continuous. The pair (U, N) is said to satisfy the Lack of Anticipation Assumption (LAA), if for all $t \geq 0$, $\{N(t+s) - N(t) : s \geq 0\}$ is independent of $\{U(s) : 0 \leq s \leq t\}$. □

Intuitively, LAA restricts the interaction between U and N, by requiring that at any time point, future arrivals and the past state history be independent; thus, future arrivals cannot be "anticipated" by past states. LAA is the sufficient condition for PASTA in the following celebrated theorem, cast in the sample path framework.

Theorem 3.2 (PASTA Theorem)
Let N be a Poisson Process and suppose that LAA holds for the pair (U, N). Then, $V(t) \underset{t\to\infty}{\rightarrow} V(\infty)$ w.p.1 iff $W(t) \underset{t\to\infty}{\rightarrow} W(\infty)$ w.p.1, in which case $V(\infty) = W(\infty)$ w.p.1. □

An early heuristic variant of LAA and PASTA appears in Strauch [26]. A complete proof, using martingale techniques, is due to R. Wolff [31], which partly draws upon Whitt [30]. A discrete-time analog of Theorem 3.2, in which intervals between points have a geometric distribution, was first derived in Halfin [13]. For more recent work in the discrete-time setting, see Makowski et al. [18].

The PASTA Theorem is often invoked in queueing context, where Poisson arrivals are frequently used to model arrival processes. LAA then holds due to the independence of future Poisson increments of the σ-algebra generated by the state process to date. Consequently, the ASTA Theorem implies that the equilibrium ("outside-observer") queue length distribution coincides with its embedded counterpart (at arrival time points). In particular, for queues with exponential servers, this fact simplifies the computation of equilibrium distributions for job waiting times (time spent in queue only) and job sojourn times (time spent in queue and in service). To see that, note that both the waiting time and sojourn time distributions are completely determined by the queue length *ahead* of the arriving job, that is, by the state "seen" by that job; more specifically, given $U^{(N)}(n) = k$, the waiting and sojourn time distributions are convolutions of k and $k+1$ service times, respectively. The unconditional distributions are then obtained with respect to the known "outside-observer" queue length distribution.

The converse of the PASTA property can be formulated as follows.

Definition 3.3 (The Anti-PASTA Problem)
Given that ASTA holds for a pair, (U, N), find conditions under which N is a Poisson process. □

The term, Anti-PASTA, was coined in Green and Melamed [12]. It is an obvious choice for the converse of PASTA, in view of the inadvertent "culinary" nature of the latter. Some Anti-PASTA results will be given later in a Markovian setting.

3.2 ASTA under the WLAA Condition

As its name suggests, the Weak Lack of Anticipation Assumption (WLAA) is a weaker version of LAA, i.e., it strengthens the interaction between U and N without destroying the ASTA property (see [20]).

Definition 3.4 (WLAA)
Let U be left-continuous and let N be right-continuous. The pair (U, N) is said to satisfy the Weak Lack of Anticipation Assumption (WLAA), if for each $t \geq 0$ there exists $s_0 > 0$, such that $U(t)$ and $N(t+s) - N(t)$ are uncorrelated for all $0 \leq s \leq s_0$. □

The definition of WLAA weakens that of LAA in three ways. First, the entire past state history, $\{U(s) : 0 \leq s \leq t\}$, is replaced by the present state, $U(t)$, alone; second, the entire future history of arrivals, $\{N(t+s) - N(t) : s \geq 0\}$, is replaced by finite-horizon future increments, $\{N(t+s) - N(t) : \ 0 \leq s \leq s_0\}$; and third, stochastic independence is replaced by lack of correlation. Thus, LAA implies WLAA, but not vice versa. The WLAA condition is closely related to an independence condition introduced by König and Schmidt [15]; see Theorem 1.6.6 of Franken et al. [9].

A sufficient condition for ASTA in terms of WLAA is cast in the expectation framework as follows.

Theorem 3.5 (ASTA under WLAA)
Suppose that the pair (U, N) satisfies WLAA. Assume further that at least one of the two following conditions holds:

1. $\mathsf{E}[U(t)]$ is independent of $t \geq 0$
2. for each $t \geq 0$ there exists $s_0 > 0$, such that $\mathsf{E}[N(t+s) - N(t)] = \lambda s$ for $0 \leq s \leq s_0$

Then, $\bar{V}(t) = \bar{W}(t)$ for all $t > s_0$, so that $\bar{V}(t) \underset{t\to\infty}{\to} \bar{V}(\infty)$ iff $\bar{W}(t) \underset{t\to\infty}{\to} \bar{W}(\infty)$, in which case $\bar{V}(\infty) = \bar{W}(\infty)$. □

Clearly, if U is stationary and N has stationary increments, then Conditions 1 and 2 in Theorem 3.5 will be satisfied. However, there is no requirement that N be a Poisson Process, though WLAA imposes a considerable restriction on N. The proof of Theorem 3.5, utilizing the Riemann-Stieltjes Sum approach in the expectations framework, will be sketched in Section 4.

3.3 ASTA under the LBA Condition

The Lack of Bias Assumption (LBA) generalizes both LAA and WLAA essentially to the utmost generality. Ignoring some minor technical details, LBA provides a fairly general necessary and sufficient condition for ASTA, thereby delineating the strongest interaction possible between U and N, which still preserves the ASTA property. The discussion of this section is based on [20].

The LBA condition does not directly involve the counting process, N. Rather, LBA requires the existence of a *conditional stochastic intensity* process, $\eta_U \equiv \{\eta_U(t) : \ t > 0\}$, of N given U. The general definition of the stochastic intensity concept requires highly-technical conditions under the martingale approach, and is deferred until Section 5. In this section, however, we present a more restrictive definition, but one that has the advantage of being intuitive.

Definition 3.6 (Conditional Stochastic Intensity)
Let U be left-continuous and let N be right-continuous. The conditional intensity, $\eta_U(t)$, of N at t given $U(t)$, is the random variable,

$$\eta_U(t) = \lim_{s \to 0} \frac{\mathsf{E}[N(t+s) - N(t) | U(t)]}{s}, \quad t \geq 0, \tag{11}$$

provided the limit exists. □

One motivation for calling η_U an intensity process is suggested by Eq. (11). A more transparent motivation is offered by the fact that if the interchange of the limit and expectation operation is admissible in (11), then

$$\mathsf{E}[\eta_U(t)] \ = \ \mathsf{E}\left[\lim_{s \to 0} \frac{\mathsf{E}[N(t+s) - N(t) | U(t)]}{s}\right] \ = \ \lim_{s \to 0} \mathsf{E}\left[\frac{\mathsf{E}[N(t+s) - N(t) | U(t)]}{s}\right] \ = \ \lambda(t), \tag{12}$$

where $\lambda(t) = (d/dt)\, \mathsf{E}[N(t)]$ is the rate (or intensity) of $N(t)$. It is useful to think of $\eta_U(t)$ as a conditional hazard, that is, as the (random) rate of a new arrival (in the counting process, N) at time t, given that the state process is $U(t)$. Thus, η_U encapsulates the "decision" of whether or not to sample U at time t, having observed its value, $U(t)$. The interaction between U and N should thus be intuitively clear: for ASTA to hold, the sampling decision should not depend "too much" on the state; otherwise, biased statistics would ensue, that is, $\bar{W}(\infty) - \bar{V}(\infty) \neq 0$. It will turn out that lack of bias is ensured by the LBA property, as defined below.

Definition 3.7 (LBA)
Let U be left-continuous and let N be right-continuous. The pair (U, N) is said to satisfy the Lack of Bias Assumption (LBA), if U and η_U are pointwise uncorrelated, i.e., $\mathsf{Cov}[U(t), \eta_U(t)] = 0$, for all $t \geq 0$. □

Formally, LBA weakens WLAA by substituting a pointwise condition for an interval-based condition. This fact together with Definition 3.6 render LBA a local condition in the sense that the interaction between U and N (through η_U) need be known only in an infinitesimal forward interval, and then only through covariances which are a measure of *linear* dependence. The LBA definition and terminology are motivated, in retrospect, by the following Lemma.

Lemma 3.8 (The Bias Formula)
Let U be left-continuous and let N be right-continuous. Suppose that

1. U and N are jointly stationary in the sense that

$$(U(t), N(t+s) - N(t)) \ \stackrel{d}{=} \ (U(t+h), N(t+h+s) - N(t+h)), \quad t, s, h > 0$$

2. η_U is well defined and obeys Eq. (12), so that

$$\mathsf{E}[\eta_U(t)] \equiv \lambda, \quad t \geq 0, \tag{13}$$

where $\lambda = \lim\limits_{s \to 0} \dfrac{N(s)}{s}$.

Then, both $\bar{V}(t)$ and $\bar{W}(t)$ are independent of t and the bias, $\bar{W}(t) - \bar{V}(t)$, obeys the relation

$$\bar{W}(t) - \bar{V}(t) = \frac{\mathsf{Cov}[U(0), \eta_U(0)]}{\lambda}, \quad t \geq 0. \tag{14}$$

□

The Bias Formula (14) states that the ASTA condition can be attained either by decreasing the magnitude of the covariance in the numerator, or by increasing the sampling rate in the denominator. In the first case, the magnitude of the covariance, being proportional to the bias, can be viewed as a direct measure of the interaction between U and N, where the sign of the covariance determines the direction of the bias. Thus, a positive (negative) covariance means that the event average overestimates (underestimates) its time average counterpart. In the second case, sampling at a higher rate decreases the bias inverse-proportionately. ASTA holds when the bias vanishes, i.e., when either the covariance vanishes or the rate is infinite. The latter case is not feasible, but intuitively, sampling U at an infinite rate is informationally equivalent to U itself. Overall, the covariance in the numerator of (14) plays the most fundamental role in the ASTA problem, motivating the LBA condition. This is explicitly summarized in the following theorem.

Theorem 3.9 (ASTA under LBA)
Assume that the conditions of Lemma 3.8 hold for the pair (U, N). Then, $\bar{W}(\infty) = \bar{V}(\infty)$ iff LBA holds for (U, N). □

While the LBA condition is given in terms of covariances, it can be shown that it assumes the form of stochastic independence, provided it holds for a sufficiently rich class of functions f. Specifically, consider the class $\mathcal{U}$ consisting of all stochastic processes of the form $U(t) = f(X(t))$, over all bounded continuous real-valued functions, f. Alternatively, $\mathcal{U}$ may be taken as the class of all indicator functions of events in terms of X.

Theorem 3.10 (ASTA under Lack of Dependence)
Assume that the conditions of Lemma 3.8 hold for each pair (U, N), where $U \in \mathcal{U}$ and N is a fixed counting process. Then the following are equivalent:

1. ASTA holds for (X, N) in the sense that $X(\infty) \stackrel{d}{=} X^{(N)}(\infty)$
2. LBA holds for each pair (U, N), for all $U \in \mathcal{U}$
3. $\eta_X(t) = \mathsf{E}[\eta_X(t)] \equiv \lambda$ w.p.1., for all $t \geq 0$
4. $\eta_X(t)$ is independent of $X(t)$, for all $t \geq 0$ □

Thus, ASTA is equivalent here to the fourth condition above, which is stated in terms of pointwise independence (and therefore is nominally stronger than pointwise lack of correlation); however, being a local (pointwise) condition, it is still weaker than WLAA.

We mention that excluding PASTA, both WLAA and LBA are not easy to verify, except when X or U are Markovian. In that case, the conditional intensity process can be defined rather simply in terms of the Markovian transition intensities (rates) of the corresponding infinitesimal generator. Finally, we mention that WLAA and LBA are not vacuous conditions. One can exhibit non-Poisson counting processes of interest for which ASTA holds. These will be demonstrated in Section 6.

3.4 Conditional ASTA

The term "Conditional ASTA" refers to a generalization of ASTA to conditional statistics; see Van Doorn and Regterschot [27]. The setting requires a modification and augmentation of the old one, but nothing essentially new is required.

The Conditional ASTA setting assumes that X and N are as before, but these are augmented as follows. In addition to X, let $Z \equiv \{Z(t): \ t \geq 0\}$ be another process on the underlying probability space, $(\Omega, \mathcal{F}, P)$, having the same sample path properties as X. Further, in addition to f, let g and h be bounded measurable real-valued functions, where g is defined on the state space of Z and h is defined on the cross product of the state spaces of X and Z. Finally, in addition to U, define two other transformed stochastic processes, $G \equiv \{G(t) = g(Z(t)): \ t \geq 0\}$ and $H \equiv \{H(t) = h(X(t), Z(t)): \ t \geq 0\}$.

Definition 3.11 (The Conditional ASTA Problem)
Suppose that

$$\frac{H(t)}{G(t)} \underset{t\to\infty}{\Longrightarrow} \frac{H(\infty)}{G(\infty)} \quad \text{and} \quad \frac{H^{(N)}(n)}{G^{(N)}(n)} \underset{n\to\infty}{\Longrightarrow} \frac{H^{(N)}(\infty)}{G^{(N)}(\infty)}. \tag{15}$$

The *Conditional ASTA* problem is concerned with finding sufficient and/or necessary conditions for the equality

$$\frac{H(\infty)}{G(\infty)} \stackrel{d}{=} \frac{H^{(N)}(\infty)}{G^{(N)}(\infty)}. \tag{16}$$

In this case, we say that *Conditional ASTA holds* for the triple (G, H, N). □

Calling (16) Conditional ASTA is justified by the observation that the particular choices, $h = 1_{A\times B}$ and $g = 1_B$, where A and B are measurable state sets of X and Z, respectively, correspond to conditional probabilities of the form

$$\mathsf{P}\{X(t) \in A | Z(t) \in B\}, \quad \mathsf{P}\{X^{(N)}(n) \in A | Z^{(N)}(n) \in B\},$$

whose equality in the limit (as $t \to \infty$) is the counterpart of the weak convegence condition (2) in ordinary ASTA setting.

The two variants of the ASTA problem can be extended to their conditional counterparts as well. In analogy to $V(t)$ in (3) and $W(t)$ in (4), the sample path framework for the Conditional ASTA problem considers the conditional time average process, V^*, and conditional event average process, W^*, given by

$$V^*(t) \ = \ \frac{\int_0^t H(s)\,ds}{\int_0^t G(s)\,ds}, \quad t > 0, \tag{17}$$

$$W^*(t) \ = \ \frac{\int_0^t H(s)\,dN(s)}{\int_0^t G(s)\,dN(s)}, \quad t > 0. \tag{18}$$

Similarly, in analogy to $\bar{V}(t)$ in (6) and $\bar{W}(t)$ in (7), the expectations framework for the Conditional ASTA problem considers the corresponding sequences of expectation ratios,

$$\bar{V}^*(t) = \frac{\mathsf{E}[\int_0^t H(s)\,ds]}{\mathsf{E}[\int_0^t G(s)\,ds]}, \quad t > 0, \tag{19}$$

$$\bar{W}^*(t) = \frac{\mathsf{E}[\int_0^t H(s)\,dN(s)]}{\mathsf{E}[\int_0^t G(s)\,dN(s)]}, \quad t > 0. \tag{20}$$

It should be clear that the entire theory presented up to this point will go through, by applying the various conditions, with minor modifications, to the pairs (H, N) and (G, N), separately. In fact, the new process (X, Z) can be viewed as being of the same species as the old process X. Moreover, $H(t) = h(X(t), Z(t))$ and $G(t) = g(Z(t))$ are just two candidate bounded continuous real-valued functions of $(X(t), Z(t))$ of the same species as $U(t)$. Thus, the Conditional ASTA problem can be treated via the tools of the ordinary ASTA problem, and for this reason we proceed here no further. The omitted details may be found in Van Doorn and Regterschot [27] and in Melamed and Whitt [20].

3.5 ASTA Under Time reversal

The PASTA Theorem in Wolff [31] was stated for left-continuous state processes, U, and right-continuous counting processes, N. In queueing context, the semantics of these assumptions correspond to *jobs in an arrival process observing the state of the queue just before joining it*. There is, however, a conceptual duality between these semantics and those corresponding to *jobs in a departure process observing the state of the queue just after departing from it*. The duality is obtained by time reversal, namely, by considering a stochastic process $Y \equiv \{Y(t) : -\infty < t < \infty\}$ and its time-reversed version $\tilde{Y} \equiv \{\tilde{Y}(t) = Y(-t) : -\infty < t < \infty\}$.

Note carefully that "time reversal" is a concept distinct from "time reversibility"; any process can be reversed in time, but only certain processes are time-reversible, namely, their probability law coincides with that of their time-reversed version. In queueing context, time reversal can be visualized as jobs moving through a queue backward in time. Thus, arrivals in Y become departures in $\tilde{Y}$ and vice versa. Moreover, reversing time in a left-continuous U results in a right-continuous $\tilde{U}$, while reversing time in a right-continuous $\tilde{N}$ results in a left-continuous N. However, the arrival time points, T_n, remain unchanged under time reversal, though their enumeration changes per force. More importantly, all limiting time averages, $\bar{V}(\infty)$, and event averages, $\bar{W}(\infty)$, along each sample path remain unchanged under time reversal, provided only that the integrals (sums) considered are invariant under the order of integration (summation). Thus, the foregoing ASTA results can be applied to "new" traffic processes in the time-reversed system and "old" state processes in the original setting, yielding dual ASTA results. Such results typically extend ASTA from arrival-embedded to departure-embedded statistics, and will be referred to as *Reverse ASTA*.

To illustrate this idea, we restate some Reverse ASTA conditions in terms of ASTA conditions for pairs (U, N) in ordinary time; these generally require the replacement of forward time intervals with backward time intervals. (Refer to Serfozo [23] for related conditions and results.)

Definition 3.12 (Reverse ASTA Conditions)
Let U be right-continuous and let N be left-continuous.

The Reverse Lack of Anticipation Assumption (Reverse LAA) holds for (U, N), if for all $t \geq 0$, $\{N(t) - N(t-s) : \; 0 \leq s \leq t\}$ is independent of $\{U(s) : \; s \geq 0\}$.

The Reverse Weak Lack of Anticipation Assumption (Reverse WLAA) holds for (U, N), if for each $t \geq 0$ there exists $s_0 > 0$, such that $U(t)$ and $N(t-s, t)$ are uncorrelated for all $0 \leq s \leq s_0$ with $t - s \geq 0$.

The Reverse Lack of Bias Assumption (Reverse LBA) holds for (U, N), if $U(t)$ and $\tilde{\eta}_U(t)$ are pointwise uncorrelated for all $t \geq 0$, where $\tilde{\eta}_U(t) = \lim_{s \to 0} \dfrac{\mathsf{E}[N(t-s,t)|U(t)]}{s}$ is the reverse conditional intensity of N at t given $U(t)$. □

The results of the previous ASTA theorems, will then hold under the following modifications.

Theorem 3.13 (ASTA under Time Reversal)
Let U be right-continuous and N left-continuous. Then

1. The results of Theorem 3.2 hold, provided LAA in Definition 3.1 is replaced by Reverse LAA in Definition 3.12.
2. The results of Theorem 3.5 hold, provided WLAA in Definition 3.4 is replaced by Reverse WLAA in Definition 3.12, and Condition 2 in Theorem 3.5 is replaced by the condition:
(2') For each $t \geq 0$ there exists positive s_0, such that $\mathsf{E}[N(t) - N(t-s)] = \lambda s$, for $0 \leq s \leq s_0$ with $t - s \geq 0$.
3. The results of Theorem 3.9 hold, provided LBA in Definition 3.7 is replaced by Reverse LBA in Definition 3.12, η_U in Definition 11 is repaced by $\tilde{\eta}_U$ in Definition 3.12, and finally, Condition 1 in Lemma 3.8 is replaced by the condition:
(1') U and N are jointly stationary in the sense that
$$(U(t), N(t) - N(t-s)) \overset{d}{=} (U(t+h), N(t+h) - N(t+h-s)), \quad t, s, h > 0. \qquad \square$$

The time-reversed version of Theorem 3.10 is similar.

3.6 ASTA in Markovian Setting with Transition Counts

When U is a Markovian state process and N is a counting process of transitions in U, a sweeping simplification of the ASTA problem takes place. In this specialized setting, the ASTA, PASTA and Anti-PASTA properties merge and become essentially equivalent. Many useful queueing models fall within the scope of this simplification; see Melamed [19].

The setting of this case is as follows. Let U be a right-continuous Markov process. For simplicity, we assume that the state space, E of U, is countable, though the results go through for the uncountable case as well (see *ibid.*) Assume further that the infinitesimal generator matrix, Q of U, with elements $q(i, j)$, has a bounded diagonal, and that U has a unique long-run distribution vector ξ. Let $\Theta \subset E \times E$ be a so-called *traffic set* of U. Define a sequence of random times, T_n, recursively by

$$T_n = \begin{cases} 0, & n = 0 \\ \inf\{t > T_{n-1} : \; (U(t-), U(t+)) \in \Theta\}, & n > 0 \end{cases}$$

and take N to be the right-continuous counting process of the random times T_n, above. Let ψ^+ and ψ^- be the long run distributions of $\{U(T_n+):\ n = 0, 1, \ldots\}$ and $\{U(T_n-):\ n = 0, 1, \ldots\}$, respectively. It can be shown (see [19]) that

$$\psi^+(k) \;=\; \frac{\displaystyle\sum_{j:\,(j,k)\in\Theta} \xi(j)\, q(j,k)}{\displaystyle\sum_{(i,j)\in\Theta} \xi(i)\, q(i,j)}, \tag{21}$$

$$\psi^-(k) \;=\; \frac{\xi(k) \displaystyle\sum_{j:\,(k,j)\in\Theta} q(k,j)}{\displaystyle\sum_{(i,j)\in\Theta} \xi(i)\, q(i,j)}. \tag{22}$$

When U has an uncountable state space, then additional technical measure-theoretic assumptions are made, such as imposing Harris recurrence on U. However, Eqs. (21) and (22) remain structurally unchanged, with the sums replaced by corresponding integrals. See [19] for details.

Recall that the ASTA problem seeks conditions under which $\xi = \psi^+$ for the right-continuous version of U, or $\xi = \psi^-$ for the left-continuous version of U. The key conditions that bring about this state of affairs are pointwise independence conditions on U and N, which may be viewed as weaker versions of LAA; refer to Section 5.3 for precise details. Intuitively, the Markov property obviates the requirement that future arrivals be independent of the entire past history of a Markovian state process, U. In fact, independence of the present (Markovian) state of U will do. The Markovian analogue of Theorem 3.2 is stated below.

Theorem 3.14 (ASTA Theorem for Markovian State Processes)
Let U be a regular Markov jump process.
Then ASTA holds for the pair (U, N) iff either of the following holds:

1. U is left-continuous and the pair (U, N) is forward-pointwise independent, i.e., for all $t > 0$, $U(t)$ and $\{N(t+s) - N(s):\ s \geq 0\}$ are independent.

2. U is right-continuous and the pair (U, N) is backward-pointwise independent, i.e., for all $t > 0$, $U(t)$ and $\{N(t) - N(t-s):\ 0 \leq s \leq t\}$ are independent. In fact, the latter condition is equivalent to pointwise independence of the pair (U, N), i.e., for all $t > 0$, $U(t)$ and $N(t)$ are independent.

In each case above, PASTA and Anti-PASTA are equivalent to ASTA. □

A proof for the second case in Theorem 3.14 appears in Melamed [19] . This case is a consequence of more general results on traffic-embedded invariant distributions in Markovian jump processes with right-continuous sample paths. The corresponding version for the left-continuous case can be obtained via time-reversal arguments (see e.g. Green and Melamed [12]). Recall that time reversal does not affect ξ, but merely interchanges left-continuity and right-continuity, as well as forward-pointwise and backward-pointwise independence.

Theorem 3.14 is a variant of the Arrival Theorem class of ASTA results. Applications of the Arrival Theorem to queueing networks may be found in Section 6.

4 The Riemann-Stieltjes Sum Approach to ASTA

The Riemann-Stieltjes Sum approach to ASTA uses a stochastic integral representation for the embedded sums used in event averages. Specifically, since the counting process N is nondecreasing, its realizations induce measures of the form $dN(\omega, s) = N(\omega, s+) - N(\omega, s-)$ on the non-negative half line for almost every sample point, ω. Notice that $dN(s) = 1$ whenever $s = T_k$ for some $k > 0$, but $dN(s) = 0$, otherwise. Omitting the ω argument for typographical clarity, we can consequently write w.p.1,

$$\sum_{k=1}^{N(t)} U_k^{(N)} \;=\; \sum_{k=1}^{N(t)} U(T_k) \;=\; \int_0^t U(s)\,dN(s) \;=\; \lim_{n\to\infty} \sum_{k=0}^{n-1} U(kt/n)\,[N((k+1)t/n) - N(kt/n)], \tag{23}$$

where the last sum is the Riemann-Stieltjes sum of the preceding Riemann-Stieltjes integral. From Apostol [1],Chapter 9, the integral in (23) is well defined by virtue of the w.p.1 boundedness and left-continuity of U and the w.p.1 right-continuity of N. We point out that for the Riemann-Stieltjes integral to be well defined, it suffices for U to have w.p.1 no discontinuities from the right (left) where N has its right (left) discontinuities; in fact, limits from the left and right are not actually needed. If, however, U and N have common discontinuities from the left or from the right with positive probability, then there can be difficulties (*ibid.*) This is avoided by the assumption that U is left-continuous (right-continuous), whereas N is assumed to be right-continuous (left-continuous). Note also that the Riemann-Stieltjes sum above need not have equal-size subintervals in the partition of $[0, t]$, so long as the width of the largest interval approaches zero in the limit. Moreover, the k-th term of the n-th sum in (23) can be evaluated for U anywhere in the closed interval $[kt/n, (k+1)t/n]$. However, for our purposes, it is important to use the left endpoint (right endpoint for time-reversed cases).

Recalling that U is bounded (because f is), and $\mathsf{E}[N(t)] < \infty,\ t > 0$, we have w.p.1,

$$\sum_{k=0}^{n-1} U(kt/n)\,[N((k+1)t/n) - N(kt/n)] \;\le\; [\sup_{0\le s\le t} |U(s)|]\; N(t) < \infty. \tag{24}$$

Applying the Lebesgue Dominated Convergence Theorem to interchange the expectation and limit operations in (24), we get

$$\mathsf{E}[\sum_{k=1}^{N(t)} U(T_k)] \;=\; \lim_{n\to\infty} \sum_{k=0}^{n-1} \mathsf{E}[U(kt/n)\,\{N((k+1)t/n) - N(kt/n)\}]. \tag{25}$$

We are now in a position to illustrate the Riemann-Stieltjes Sum approach to the ASTA problem. The proof of two ASTA theorems will be now sketched in the expectations framework.

Proof of Theorem 3.5:
By (25) and WLAA (Definition 3.4),

$$\mathsf{E}[\sum_{k=1}^{N(t)} U(T_k)] \;=\; \lim_{n\to\infty} \sum_{k=0}^{n-1} \mathsf{E}[U(kt/n)]\,\mathsf{E}[N((k+1)t/n) - N(kt/n)].$$

Suppose that Condition 1 in Theorem 3.5 holds, so that $\mathsf{E}[U(t)] = \mathsf{E}[U(0)]$ for all t. Then,

$$\mathsf{E}[\sum_{k=1}^{N(t)} U(T_k)] \;=\; \mathsf{E}[U(0)] \lim_{n\to\infty} \sum_{k=0}^{n-1} \mathsf{E}[N((k+1)t/n) - N(kt/n)] \;=\; \mathsf{E}[U(0)]\,\mathsf{E}[N(t)].$$

Hence,

$$\bar{W}(t) \;=\; \mathsf{E}[U(0)] \;=\; \frac{1}{t}\int_0^t \mathsf{E}[U(s)]\,ds \;=\; \bar{V}(t), \quad t > 0,$$

and ASTA has been established.

Suppose that Condition 2 in Theorem 3.5 holds, so that $\mathsf{E}[N(t+s) - N(t)] = \lambda s$ for $t \geq 0$ and $0 \leq s \leq s_0$. Then

$$\mathsf{E}[\sum_{k=1}^{N(t)} U(T_k)] \;=\; \lim_{n\to\infty} \frac{\mathsf{E}[N(t/n)]}{t/n} \sum_{k=0}^{n-1} \frac{t}{n}\,\mathsf{E}[U(kt/n)] \;=\; \lambda \int_0^t \mathsf{E}[U(s)]\,ds.$$

Since $\lambda = \mathsf{E}[N(t)/t]$, it follows that $\bar{W}(t) = \bar{V}(t)$ for $t > s_0$, whence ASTA follows again. □

Proof of Theorem 3.9:
First, assuming that f is continuous, we have

$$\begin{aligned}
&\mathsf{E}\left[\frac{1}{t}\int_0^t U(s)\,dN(s)\right] \;=\\
&= \mathsf{E}\left[\lim_{n\to\infty} \frac{1}{t}\sum_{k=0}^{n-1} U(kt/n)\,\{N((k+1)t/n) - N(kt/n)\}\right] \quad \text{by (23)}\\
&= \lim_{n\to\infty} \frac{1}{t}\sum_{k=0}^{n-1} \mathsf{E}\left[U(kt/n)\,\{N((k+1)t/n) - N(kt/n)\}\right] \quad \text{by (25)}\\
&= \lim_{n\to\infty} \frac{n}{t}\, E[U(0)\,N(t/n)] \quad \text{by Condition 1 in Lemma 3.8}\\
&= \lim_{n\to\infty} \frac{n}{t}\, \mathsf{E}\Big[U(0)\,\mathsf{E}[N(t/n)|U(0)]\Big] \quad \text{by the calculus of conditioning}\\
&= \mathsf{E}\left[U(0)\,\lim_{n\to\infty}\frac{n}{t}\,\mathsf{E}[N(t/n)|U(0)]\right] \quad \text{by (11) and (12)}\\
&= E[U(0)\,\eta_U(0)] \quad \text{by (11),}
\end{aligned}$$

where in the second to last step, the convergence and uniform integrability in (11) and (12) remain unaltered on multiplication by the bounded random variable $U(0)$. Hence, from (7),

$$\bar{W}(t) \;=\; \frac{t}{\mathsf{E}[N(t)]}\,\mathsf{E}[U(t)\,\eta_U(t)].$$

But by (6) and the assumed stationarity of U,

$$\bar{V}(t) \;=\; \mathsf{E}[U(t)].$$

The result follows from the above for U with continuous f, in view of the relation

$$\mathsf{E}[N(t)] \;=\; \lambda t \;=\; \mathsf{E}[\eta_U(t)]t.$$

The extension to all measurable f, for which the expectations are well defined, is obtained by taking limits; see Billingsley [3], pp. 7–9. □

We conclude this section by pointing out that the Riemann-Stieltjes Sum approach is highly intuitive. In particular, it makes it easy to identify the WLAA and LBA as useful ASTA conditions. To see that, observe that both of these ASTA conditions are based on lack of correlation, or equivalently, on the algebraic equality of expectations of products to the corresponding products of expectations.

5 The Martingale Approach to ASTA

This section employs the martingale approach (refer to [2, 5, 6, 18, 21], among others) to treat the ASTA Problem in the sample path framework. The treatment requires a more elaborate setting, but the results are technically more general than the Riemann-Stieltjes Sum approach. An alternative sample path approach is due to Stidham and El-Taha (see [8, 25]).

5.1 The Martingale Approach Setting

As before, we start with a general state process, $X \equiv \{X(t): \ t \geq 0\}$, and a counting process, $N \equiv \{N(t): \ t \geq 0\}$, over a common probability space, $(\Omega, \mathcal{F}, \mathsf{P})$. Both X and N are assumed to be right-continuous with left limits.

Let $\mathcal{F} \equiv \{\mathcal{F}_t\}_{t \geq 0}$ be the filtration (an increasing family of σ-fields),

$$\mathcal{F}_t \ = \ \sigma\{(X(s), N(s)): \ s \leq t\},$$

generated by the pair, (X, N). Henceforth, all martingales and stochastic intensities, as well as all measurability and predictability properties, are defined with respect to this filtration, unless otherwise specified.

As in Section 2, let $U(t) = f(X(t-))$ be bounded and left-continuous. The left continuity is required as a technical condition to ensure predictability, namely, $U(t) \in \mathcal{F}_{t-}$ for all t, where $\mathcal{F}_{t-} = \cup_{s<t} \mathcal{F}_s$. If $M \equiv \{M(t): \ t \geq 0\}$ is a martingale, then the boundedness and predictability of U imply that the stochastic integral,

$$(U \circ M)(t) \ \equiv \int_0^t U(s)\, dM(s), \tag{26}$$

is also a martingale, a fact that will be used later, in particular, in proving Theorem 5.3 below.

The (unique) Doob-Meyer decomposition of N yields

$$N(t) \ = \ M(t) \ + \ A(t), \tag{27}$$

where $M \equiv \{M(t): \ t \geq 0\}$ is a martingale, and $A \equiv \{A(t): \ t \geq 0\}$ is a predictable, increasing process, with $A(0) = 0$. The process A is known as the *compensator* (or "dual predictable projection") of the counting process N. The martingale, M, is assumed to be square integrable.

Finally, N is assumed to be a *simple* counting process, i.e., of a *unit* jump size. Let

$$\Delta N(t) \ \equiv \ N(t) - N(t-), \quad \text{and} \quad \Delta A(t) \ \equiv \ A(t) - A(t-).$$

Then, based on (27), and by virtue of the martingale property of M and the predictability of A, we have

$$\mathsf{E}[\Delta N(t) | \mathcal{F}_{t-}] \ = \ \mathsf{E}[\Delta A(t) | \mathcal{F}_{t-}] \ = \ \Delta A(t),$$

implying $\Delta A(t) \leq 1$ for all t, since $\Delta N(t) \leq 1$. Consequently, the *quadratic variation* process, $\langle M \rangle$, satisfies for all $t \geq 0$,

$$\langle M \rangle(t) \ = \ \int_0^t [1 - \Delta A(s)]\, dA(s) \ \leq \ A(t). \tag{28}$$

Note that when A is continuous, i.e., $\Delta A(t) = 0$ for all t, we have

$$\langle M\rangle(t) \;=\; A(t), \quad \forall t. \tag{29}$$

We shall need the following SLLN (strong law of large numbers) for the martingale M.

Lemma 5.1 (Special cases in Liptser [17])

(i) If $\langle M\rangle(\infty) < \infty$, then $M(\infty) < \infty$;

(ii) If $\langle M\rangle(\infty) = \infty$, then $\lim_{t\to\infty} M(t)/\langle M\rangle(t) = 0$. $\square$

Another mild regularity condition is $N(\infty) = \infty$.

Lemma 5.2
$N(\infty) = \infty$ if and only if $A(\infty) = \infty$. Furthermore, either condition implies

$$\lim_{t\to\infty} M(t)/A(t) \;=\; 0,$$

or equivalently,

$$\lim_{t\to\infty} N(t)/A(t) \;=\; 1.$$

Proof:
(i) Suppose $N(\infty) = \infty$. If $A(\infty) < \infty$, then $\langle M\rangle(\infty) < \infty$ from (28), implying $M(\infty) < \infty$ (Lemma 5.1). Hence, $N(\infty) < \infty$ follows from (27), a contradiction.

(ii) Suppose $A(\infty) = \infty$. If $\langle M\rangle(\infty) = \infty$, then $M(t)/\langle M\rangle(t) \to 0$, from Part (ii) in Lemma 5.1, and hence (cf. (28)),

$$|M(t)|/A(t) \;\le\; |M(t)|/\langle M\rangle(t) \;\to\; 0.$$

If, however, $\langle M\rangle(\infty) < \infty$, then $M(\infty) < \infty$ (Lemma 5.1) and hence $M(t)/A(t) \to 0$. Thus, in either case $M(t)/A(t) \to 0$, which is equivalent to $N(t)/A(t) \to 1$, following the decomposition in (27). Hence, we conclude that $N(\infty) = \infty$. $\square$

Next, assume that N admits a stochastic intensity, $\nu \equiv \{\nu(t) : \; t \ge 0\}$ of the form

$$\nu(t) \;=\; \lim_{h\to 0} \frac{1}{h}\, \mathsf{E}[N(t+h) - N(t)|\mathcal{F}_t]. \tag{30}$$

Substituting (27) into (30), and again making use of the martingale property of M and the predictability of A, we have

$$\begin{aligned}
\nu(t) \;&=\; \lim_{h\to 0} \frac{1}{h}\, \mathsf{E}[M(t+h) - M(t) + A(t+h) - A(t)|\mathcal{F}_t] \\
&=\; \lim_{h\to 0} \frac{1}{h}\, \mathsf{E}[A(t+h) - A(t)|\mathcal{F}_t] \\
&=\; \lim_{h\to 0} \frac{1}{h}\, [A(t+h) - A(t)].
\end{aligned}$$

That is,

$$\nu(t) \;=\; \frac{d}{dt} A(t), \quad \text{or} \quad A(t) \;=\; \int_0^t \nu(s)\, ds.$$

Hence, $A(t)$ is absolutely continuous, and in particular, (29) holds. Note that ν is non-negative, and predictable, i.e., $\nu(t) \in \mathcal{F}_{t-}$, for all t. Assume that for each t, $\mathsf{E}[\nu(t)] < \infty$, and define

$$r(t) \;\equiv\; \mathsf{E}[\nu(t)], \quad \text{and} \quad R(t) \;\equiv\; \int_0^t r(s)\, ds.$$

Applying Fubini's theorem, we get $R(t) = \mathsf{E}[A(t)]$, for all t.

Another well-known fact (e.g., Brémaud [4]) is that for any bounded, predictable process $\{Y(t) : t \geq 0\}$, one has

$$\mathsf{E}[\int_0^t Y(s)\, dN(s)] \;=\; \mathsf{E}[\int_0^t Y(s)\, \nu(s)\, ds], \quad \forall t. \tag{31}$$

This is so, because from (27), the lefthand side of (31) can be written as

$$\mathsf{E}[\int_0^t Y(s)\, dN(s)] \;=\; \mathsf{E}[\int_0^t Y(s)\, dM(s)] \;+\; \mathsf{E}[\int_0^t Y(s)\, dA(s)].$$

The first term on the righthand side above equals 0, since $\int_0^t Y(s)\, dM(s)$ is a martingale (as discussed earlier), while the second term equals the righthand side of (31).

For the rest of this section, we always assume that N is a simple counting process, with $N(\infty) = \infty$, that admits the decomposition in (27) and the stochastic intensity in (30).

5.2 The General Case

This section establishes a version of ASTA in the sample path framework. Define

$$V(t) \;=\; \frac{\int_0^t r(s)\, U(s)\, ds}{R(t)}, \quad W(t) \;=\; \frac{\int_0^t U(s)\, dN(s)}{N(t)}, \tag{32}$$

and note that while $W(t)$ in (32) remains the same as in (4), $V(t)$ in (32) is slightly more general than the definition in (3): here it is a *weighted* time average, with weight $r(s)$ for each $s \in (0, t]$. The goal is again to prove ASTA in the sense of

$$\lim_{t\to\infty} [W(t) - V(t)] \;=\; 0. \tag{33}$$

Unless otherwise specified, all limits involving stochastic processes are, henceforth, in the almost sure sense.

Letting

$$V^{(A)}(t) = \frac{\int_0^t U(s) dA(s)}{A(t)},$$

we first prove an intermediate, key result.

Theorem 5.3

$$\lim_{t\to\infty} [W(t) - V^{(A)}(t)] \;=\; 0.$$

Proof:

$$\begin{aligned} W(t) \;-\; V^{(A)}(t) \;&=\; \frac{A(t)}{N(t)} \frac{\int_0^t U(s)[dN(s) - dA(s)]}{A(t)} \;+\; \left[\frac{A(t)}{N(t)} - 1\right] \frac{\int_0^t U(s)\, dA(s)}{A(t)} \\ &=\; \frac{A(t)}{N(t)} \frac{\int_0^t U(s)\, dM(s)}{A(t)} \;+\; \left[\frac{A(t)}{N(t)} - 1\right] \frac{\int_0^t U(s)\, dA(s)}{A(t)}, \end{aligned} \tag{34}$$

where the second equality makes use of the Doob-Meyer decomposition in (27). Note that the second term on the righthand side above vanishes as $t \to \infty$, since $A(t)/N(t) \to 1$ by Lemma 5.2 and U is bounded. On the other hand, since U is bounded and predictable, it follows from (26) that

$$\tilde{M}(t) \;\equiv\; (U \circ M)(t) \;=\; \int_0^t U(s)\, dM(s),$$

is itself a martingale, and in view of (29), the associated quadratic variation process is

$$\langle \tilde{M} \rangle(t) \;=\; (U^2 \circ \langle M \rangle)(t) \;=\; \int_0^t U^2(s)\, dA(s).$$

As in the proof of Part (ii) of Lemma 5.2, and with the aid of Lemma 5.1, we may deduce that $\langle \tilde{M} \rangle(\infty) < \infty$ implies $\tilde{M}(\infty) < \infty$, and hence $\tilde{M}(t)/A(t) \to 0$ (since $A(\infty) = \infty$); whereas, $\langle \tilde{M} \rangle(\infty) = \infty$ implies $\tilde{M}(t)/\langle \tilde{M} \rangle(t) \to 0$. Now, the boundedness of U implies $U^2(t) \le B$, for some positive constant B; whence, regardless whether $\tilde{M}(\infty) < \infty$ or $\tilde{M}(\infty) = \infty$, we have

$$\frac{\tilde{M}(t)}{A(t)} \;=\; \frac{\tilde{M}(t)}{\langle \tilde{M} \rangle(t)} \, \frac{\langle \tilde{M} \rangle(t)}{A(t)} \;\le\; B\, \frac{\tilde{M}(t)}{\langle \tilde{M} \rangle(t)} \to 0.$$

It then follows that the first term on the right side of (34) vanishes too as $t \to \infty$. □

A key necessary and sufficient condition for ASTA is presented in the following theorem.

Theorem 5.4
ASTA holds in the sense of (33) iff $\lim_{t\to\infty}[V^{(A)}(t) - V(t)] = 0$, i.e.,

$$\lim_{t\to\infty} \left[\frac{\int_0^t U(s)\, dA(s)}{A(t)} \;-\; \frac{\int_0^t U(s)\, r(s)\, ds}{R(t)} \right] \;=\; 0. \tag{35}$$

□

Suppose now that the intensity ν, and hence the compensator, A, is a deterministic function. That is,

$$r(t) \;\equiv\; \mathsf{E}[\nu(t)] \;=\; \nu(t) \quad \text{and} \quad R(t) \;\equiv\; \mathsf{E}[A(t)] \;=\; A(t), \quad \text{w.p.1, } \forall t.$$

Hence, (35) is trivially satisfied, since $V^{(A)}(t) = V(t)$ w.p.1, for all t. Furthermore, from Watanbe's characterization of Poisson processes (see, e.g., [29]), N is an $\mathcal{F}_t$-Poisson process. In addition, since, $\mathsf{E}[\nu(t)] = \nu(t)$, w.p.1, ν is also an intensity with respect to $\{\mathcal{F}_t^N\}$ — the filtration generated by N alone. Hence, N is also an $\mathcal{F}_t^N$-Poisson process. This reduces to the case of PASTA (cf. Theorem 3.2).

Corollary 5.5
(PASTA) ASTA holds when N has an $\mathcal{F}_t$-intensity ν, satisfying $\nu(t) = \mathsf{E}[\nu(t)]$ w.p. 1. That is, ASTA holds when N is a Poisson process (with a possibly time-dependent intensity function). □

5.3 The Stationary Ergodic Case

An important setting for ASTA is the stationary ergodic case, in which (35) is reduced to

$$\lim_{t\to\infty}[V^{(A)}(t) - V(t)] \;=\; \frac{\mathsf{E}[U(t)\nu(t)]}{\mathsf{E}[\nu(t)]} \;-\; \mathsf{E}[U(t)] \;=\; 0. \tag{36}$$

Since $U(t) = f(X(t-))$, (36) is satisfied iff

$$\mathsf{Cov}[\nu(t), f(X(t-))] \;=\; 0, \tag{37}$$

and because $\mathsf{E}[U(t)\nu(t)] = \mathsf{E}\Big[f(X(t-))\mathsf{E}[\nu(t)|X(t-)]\Big]$, a sufficient condition is

$$\mathsf{E}[\nu(t)|X(t-)] \;=\; \mathsf{E}[\nu(t)]. \tag{38}$$

Corollary 5.6
Under stationary ergodicity, ASTA holds iff (37) holds, or if (38) holds. □

Remark 5.7
Condition (37) coincides with LBA in Definition 3.7. Condition (38) is slightly stronger: it is equivalent to requiring that (37) hold for all bounded and measurable function f; as such it is analogous to Condition 2 in Theorem 3.10. Henceforth, we shall mainly use this slightly stronger version, still referring to it as LBA. □

Remark 5.8
It is easy to verify that LBA in (38) is implied by the following sufficient condition,

$$\mathsf{E}[\Delta N(t)|X(t-)] \;=\; \mathsf{E}[\Delta N(t)], \quad \forall t. \tag{39}$$

Clearly, (39) and hence (38) are implied by LAA and WLAA, and both are satisfied when N is a Poisson process, with respect to $\mathcal{F}_t$. In other words, when N is Poisson (with respect to $\mathcal{F}_t$), stationary or non-stationary, PASTA follows. No additional condition is necessary. In particular, LBA is implied by the Poisson property. (In Wolff's version of PASTA (cf. Theorem 3.2), the condition LAA is required because the Poisson process is defined with respect to the filtration $\mathcal{F}_t^N$, instead of the larger filtration $\mathcal{F}_t$ as is the case here. □

Remark 5.9
In the stationary ergodic case, the sample path and expectation frameworks of ASTA coincide (in the sense of the (41) below), LBA being the key condition for both. Similarly to (6) and (7), denote

$$\bar{V}(t) \;=\; \frac{\mathsf{E}[\int_0^t U(s)\, r(s)\, ds]}{R(t)}, \quad \bar{W}(t) \;=\; \frac{\mathsf{E}[\int_0^t U(s)\, dN(s)]}{\mathsf{E}[N(t)]}. \tag{40}$$

Then in the stationary ergodic case,

$$\bar{V}(\infty) \;=\; V(\infty) = \mathsf{E}[U(t)].$$

Also, applying (30) and (27) to the numerator and the denominator of $\bar{W}(t)$, respectively, one has

$$\bar{W}(t) \;=\; \frac{\mathsf{E}[\int_0^t U(s)\, \nu(s)\, ds]}{\mathsf{E}[A(t)]},$$

whence,

$$\bar{W}(\infty) \;=\; V^{(A)}(\infty) \;=\; \frac{\mathsf{E}[U(t)\, \nu(t)]}{\mathsf{E}[\nu(t)]}.$$

Thus, under LBA, in the sense of (36),

$$\bar{W}(\infty) \;=\; V^{(A)}(\infty) \;=\; V(\infty) \;=\; \bar{V}(\infty).$$

On the other hand, from Theorem 5.3, we have $W(\infty) = V^{(A)}(\infty)$. Thus, in the stationary ergodic case, LBA leads to

$$\bar{W}(\infty) \;=\; \bar{V}(\infty) = V(\infty) \;=\; W(\infty). \tag{41}$$

□

Consider now Corollary 5.6 in Markovian setting, thereby revisiting Section 3.6. Specifically, suppose X is a Markov process (in addition to being stationary and ergodic), and assume that $\mathcal{F}_t = \mathcal{F}_t^X$. The latter implies that $\mathcal{F}_t^N \subset \mathcal{F}_t^X$, which is typically satisfied in queueing applications. For instance, in an M/M/1 queue, with X and N being, respectively, the state process and the arrival process, all the jumps of N correspond to the upward transitions of X. Utilizing the predictability of ν and the Markov property of X, we have

$$\nu(t) \;=\; \mathsf{E}[\nu(t)|\mathcal{F}_{t-}] \;=\; \mathsf{E}[\nu(t)|\mathcal{F}_{t-}^X] \;=\; \mathsf{E}[\nu(t)|X(t-)]. \tag{42}$$

Hence, LBA in (38) becomes $\nu(t) = \mathsf{E}[\nu(t)]$, implying that N is a Poisson process (with respect to both $\mathcal{F}^X$ and $\mathcal{F}^N$). This is the *anti*-PASTA property in Definition 3.3. Since LBA is generally satisfied when N is a Poisson process (see Remark 5.7), LBA and the Poisson property become equivalent in the Markovian setting.

Corollary 5.10
Suppose that the state process X is Markovian (in addition to being stationary and ergodic), and that $\mathcal{F}_t^N \subset \mathcal{F}_t^X$ for all t. Then, LBA in the sense of (38) holds iff N is a Poisson process. □

The relations between ASTA, LBA and the Poisson property are summarized below (here only, the symbol '⇒' denotes logical implication rather than weak convergence).

- In general, N Poisson ⇒ LBA (Remark 5.7), Poisson ⇒ ASTA (Corollary 5.5).
- In the stationary ergodic case, N Poisson ⇒ LBA ⇔ ASTA (Corollaries 5.5 and 5.6).
- In the stationary ergodic Markovian case, N Poisson ⇔ LBA ⇔ ASTA (Corollary 5.10). □

5.4 Connections to Palm Calculus

Some interesting connections become apparent in the stationary setting between Corollary 5.6 and the so-called *Palm calculus* (see Brémaud [5], Baccelli and Brémaud [2], and König and Schmidt [15]).). The stationary setting means that (X, N) are assumed to be jointly stationary on the probability space, $(\Omega, \mathcal{F}, \mathsf{P})$.

let P^0 denote the *Palm probability* associated with N. Let $T_0 < T_1 < \dots$ be the sequence of arrival instants of N, and let N^0 be the *synchronized* version of N, in the sense that $T_0 = 0$ (i.e., the origin is an arrival instant). The *Palm probability (measure)*, P^0, associated with N^0, is defined by

$$\mathsf{P}^0(B) \;=\; \mathsf{P}\{N^0 \in B\}.$$

Henceforth, E and E^0 will denote the expectations with respect to P and P^0, respectively.

A key result of the Palm Calculus is Papangelou's Lemma (see Papangelou [22]).

Lemma 5.11 (Papangelou's Lemma)
$\mathsf{P}^0 \ll \mathsf{P}$ iff N admits an intensity ν with respect to $(\mathsf{P}, \mathcal{F}_t)$, in which case one has the Radon-Nikodym derivative

$$\frac{d\mathsf{P}^0}{d\mathsf{P}} = \frac{\nu}{\mathsf{E}[\nu]},$$

where the time index, t, is omitted from $\nu(t)$, due to the assumed stationarity. □

Making use of Lemma 5.11, one has

$$\begin{aligned}\mathsf{E}^0[f(X)] &= \int_{\omega\in\Omega} f(X(\omega))\,\mathsf{P}^0(d\omega) = \int_{\omega\in\Omega} f(X(\omega))\,\frac{\mathsf{P}^0(d\omega)}{\mathsf{P}(d\omega)}\,\mathsf{P}(d\omega)\\ &= \int_{\omega\in\Omega} f(X(\omega))\,\frac{\nu}{\mathsf{E}[\nu]}\,\mathsf{P}(d\omega) = \frac{\mathsf{E}[f(X)\,\nu]}{\mathsf{E}[\nu]}.\end{aligned}$$

In other words, the event average of the Palm version of the process $U \equiv f(X)$ is equal to $\mathsf{E}[f(X)\,\nu]/\mathsf{E}[\nu]$. Hence, $\mathsf{E}^0[f(X)] = \mathsf{E}[f(X)]$ iff X and ν are uncorrelated, as per (37) in the earlier analysis of stationary ergodic case of Section 5.3; see Theorem 5.3 and Remark 5.9. The connection between ASTA and Papangelou's result in Lemma 5.11 was established in Brémaud [5], which also contains a proof of the latter.

The relation derived above,

$$\mathsf{E}^0[f(X)] = \frac{\mathsf{E}[f(X)\,\nu]}{\mathsf{E}[\nu]}, \tag{43}$$

is commonly known as the *Palm formula*. The converse relation,

$$\mathsf{E}[f(X)] = \mathsf{E}[\nu]\,\mathsf{E}^0\left[\int_0^{T_1} f(X(s-))\,ds\right], \tag{44}$$

is called the *Palm inversion formula*. Taking $f(x) \equiv 1$ in (44) implies that $\mathsf{E}^0[T_1] = 1/\mathsf{E}[\nu]$. Thus, for ASTA to hold (or equivalently, for $\mathsf{E}^0[f(X)] = \mathsf{E}[f(X)]$ to hold), it suffices to have

$$\mathsf{E}^0\left[\int_0^{T_1} f(X(s-))\,ds\right] = \mathsf{E}^0[T_1]\,\mathsf{E}^0[f(X)]. \tag{45}$$

ASTA through the Palm inversion formula is due to König and Schmidt [15].

6 QUEUEING NETWORK APPLICATIONS

It turns out that Corollaries 5.6 and 5.10 have interesting connections to the so-called *quasi-reversible* queues and networks thereof (Kelly [14]). Most of the properties of quasi-reversible queues and networks, to be discussed below, are known results. However, their connection to LBA and ASTA appears to be new.

6.1 Quasi-Reversible Queues

Let $\mathbf{X} = \{\mathbf{X}(t) : -\infty < t < \infty\}$ be a Markov chain with a countable state space S and rate matrix $Q = [q(\mathbf{x}, \mathbf{y})]$. Let $\{\pi(\mathbf{x}) : \mathbf{x} \in \mathrm{S}\}$ be an invariant (ergodic) distribution of $\mathbf{X}$. The time-reversed version of $\mathbf{X}$ is the process $\tilde{\mathbf{X}} = \{\tilde{\mathbf{X}}(t) : -\infty < t < \infty\}$ over $\mathbf{S}$, defined by $\tilde{\mathbf{X}}(t) = \mathbf{X}(-t)$. The notion of a quasi-reversible queue is motivated by the following result, which relates a stationary Markov chain to its time-reversed version (Kelly, [14], p. 28).

Proposition 6.1 (Time Reversal in a Markovian Process)
The time-reversed version, $\tilde{\mathbf{X}}$ of $\mathbf{X}$, is also a stationary Markov chain, governed by a rate matrix, $\tilde{Q} = [\tilde{q}(\mathbf{x}, \mathbf{y})]$, satisfying

$$\pi(\mathbf{x})\, \tilde{q}(\mathbf{x}, \mathbf{y}) = \pi(\mathbf{y})\, q(\mathbf{y}, \mathbf{x}), \quad \forall \mathbf{x}, \mathbf{y} \in \mathbf{S},\ \mathbf{x} \neq \mathbf{y}. \tag{46}$$

Moreover, π is also an invariant distribution of $\tilde{\mathbf{X}}$. □

It is easy to verify that (46) implies the relation,

$$\sum_{\mathbf{y} \neq \mathbf{x}} q(\mathbf{x}, \mathbf{y}) = \sum_{\mathbf{y} \neq \mathbf{x}} \tilde{q}(\mathbf{x}, \mathbf{y}) \quad \forall \mathbf{x} \in \mathbf{S}. \tag{47}$$

Henceforth, we restrict consideration to queueing processes. Specifically, assume that $\mathbf{X}$ represents a queue with job classes in a countable set C, and denote the arrival rate of class-c jobs by $\lambda^{(c)}$. Further, let $\mathbf{A}^{(c)}$ be the traffic set (see Section 3.6), denoting the set of all state pairs $(\mathbf{x}, \mathbf{y})$, such that $\mathbf{y}$ is a state reachable from $\mathbf{x}$ due to a class-c job's arrival. Similarly, $\mathbf{D}^{(c)}$ is the analogous traffic set corresponding to class-c job departures. Clearly, $\mathbf{1}[(\mathbf{x}, \mathbf{y}) \in \mathbf{D}^{(c)}] = \mathbf{1}[(\mathbf{y}, \mathbf{x}) \in \mathbf{A}^{(c)}]$, where $1[\cdot]$ is an indicator function of the condition enclosed in brackets.

Definition 6.2 (Quasi-Reversibility of a Markovian Queue)
$\mathbf{X}$ is called a *Markovian queue* if, as the state process of the queue, it is a stationary and ergodic Markov chain. A Markovian queue, $\mathbf{X}$, is termed a *quasi-reversible queue*, if

$$\lambda^{(c)} = \sum_{\mathbf{y}} q(\mathbf{x}, \mathbf{y})\, \mathbf{1}[(\mathbf{x}, \mathbf{y}) \in \mathbf{A}^{(c)}] = \sum_{\mathbf{y}} \tilde{q}(\mathbf{x}, \mathbf{y})\, \mathbf{1}[(\mathbf{x}, \mathbf{y}) \in \mathbf{A}^{(c)}], \quad \mathbf{x} \in \mathbf{S}, \tag{48}$$

for all job classes, $c \in C$. □

Condition (48) requires that the rate, $\lambda^{(c)}$, be a constant *independent* of the state $\mathbf{x}$. Together with the Markov property of $\mathbf{X}$, this implies that the arrival process of each job class is Poisson. Furthermore, the time-reversed verssion of a quasi-reversible queue is *by definition* just another quasi-reversible queue, with the same Poisson arrivals. Since time-reversed arrivals correspond to departures in the original queue (and vice versa), one can immediately deduce that the departure process is Poisson as well.

Proposition 6.3
A quasi-reversible queue satisfies the following properties:

(i) Each arrival process of class-c jobs forms a Poisson Process with rate $\lambda^{(c)}$ for all $c \in C$, and these are mutually independent.

(ii) Each departure process of class-c jobs forms a Poisson Process with rate $\lambda^{(c)}$ for all $c \in C$, and these are mutually independent. Moreover, past departures up to time t are independent of the state of the system at time t. □

Thus, a quasi-reversible queue has the *Poisson-in-Poisson-out* property, namely, both input and output flows are Poisson.

6.2 Quasi-Reversible Networks

A quasi-reversible network consists of quasi-reversible queues, connected through a very general routing scheme, and in particular, the network itself possesses the Poisson-in-Poisson-out property.

A quasi-reversible network consists of a node set $M = \{1, \ldots, m\}$, and a fictitious node 0 that denotes both the "source" and "sink" of the network. Jobs belong to classes, drawn from a set C. External jobs arrive at the network (from the source node, 0) as a Poisson process of rate λ_0. Each arrival joins node i as a class-c job with probability $r_{0,i}^{(c)}$. On service completion at node i, a class-c is routed to node j and becomes a class-d job, with probability $r_{i,j}^{(c,d)}$, or leaves the network with probability $r_{i,0}^{(c)} = 1 - \sum_{(j,d)\in M\times C} r_{i,j}^{(c,d)}$, where $r_{0,0}^{(c,d)} = 0$, for all $c, d \in C$.

In the context of queueing networks, we consistently denote mathematical objects associated with node i, by appending the subscript i to them. For example, network states have the form $\mathbf{x} = (\mathbf{x}_i)_{i=1}^m$, where $\mathbf{x}_i$ is the state of node i; in a similar vein, $\mathbf{A}_i^{(c)}$ denotes the traffic set of all state pairs, $(\mathbf{x}_i, \mathbf{y}_i)$, such that $\mathbf{y}_i$ is reachable from $\mathbf{x}_i$ due to an arrival of a class-c job at node i from any network node (including network source), and similarly for $\mathbf{D}_i^{(c)}$. Accordingly, let $\lambda_i^{(c)}$ denote the total equilibrium traffic rate of class-c jobs through node i (including both external arrivals and internal transitions). Then the $\lambda_i^{(c)}$ obey the so-called *traffic equations*,

$$\lambda_i^{(c)} \;=\; \lambda_0\, r_{0,i}^{(c)} \;+ \sum_{(j,d)\in M\times C} \lambda_j^{(d)}\, r_{j,i}^{(d,c)}, \quad (i,c) \in M \times C. \tag{49}$$

The service mechanism at each node is immaterial, except that each each queue *in isolation* is required to be quasi-reversible. Specifically, the analog of (48) is assumed to hold for all $(i,c) \in M \times C$, namely,

$$\lambda_i^{(c)} \;=\; \sum_{\mathbf{y}_i\in\mathbf{S}_i} q_i(\mathbf{x}_i,\mathbf{y}_i)\,\mathbf{1}[(\mathbf{x}_i,\mathbf{y}_i)\in\mathbf{A}_i^{(c)}] \;=\; \sum_{\mathbf{y}_i\in\mathbf{S}_i} \tilde{q}_i(\mathbf{x}_i,\mathbf{y}_i)\,\mathbf{1}[(\mathbf{x}_i,\mathbf{y}_i)\in\mathbf{A}_i^{(c)}]. \tag{50}$$

Proposition 6.4
A quasi-reversible network has a product-form invariant distribution

$$\pi(\mathbf{x}) \;=\; \prod_{i=1}^m \pi_i(\mathbf{x}_i),$$

where π_i is the invariant (marginal) distribution of the queue at node i when operated in isolation and fed with a set of idependent Poisson arrivals with rates $\lambda_i^{(c)},\ c \in C$, and the $\lambda_i^{(c)}$ are the solution to the traffic equations in (49). In particular, the departure processes from the network are independent Poisson processes. □

Proposition 6.4 is proved by pairing the network with its time-reversed version, which is itself a network of quasi-reversible queues (each being the time-reversed version of its counterpart in the original network), and then analyzing both networks together based on Propositon 6.1.

Note carefully that when quasi-reversible queues are connected into a quasi-reversible network, total input streams into nodes (consisting of both external arrivals and internal transitions) *may or may not be Poisson processes*, depending on network topology; see, e.g., Walrand [28], Chapter 4. In view of Proposition 6.3, quasi-reversibility may not hold at individual nodes in a quasi-reversible network, though it holds for the network as a whole. Nevertheless, LBA and ASTA still hold, as will be explained below.

6.3 ASTA, LBA and Quasi-Reversibility

Corollaries 5.6 and 5.10 have interesting connections to quasi-reversibility. In particular, as we will demonstrate in Theorem 6.5 below, LBA and quasi-reversibility are equivalent.

In queueing applications, LBA for the arrival stream takes the form

$$\mathsf{E}[\alpha(t)|\mathbf{X}(t-)] \;=\; \mathsf{E}[\alpha(t)], \tag{51}$$

where $\mathbf{X}$ denotes the state process, and $\alpha = \{\alpha(t) : \; t \geq 0\}$ denotes the stochastic intensity of the arrival process, $N_{\mathbf{A}}$, in $\mathbf{X}$, and $\mathbf{A} \subset \mathbf{S} \times \mathbf{S}$ is the corresponding traffic set of state transitions. In quasi-reversible queues, the departure process, $N_{\mathbf{D}}$, in $\mathbf{X}$, can be viewed as the arrival process, $\tilde{N}_{\mathbf{A}}$, in the time-reversed version, $\tilde{\mathbf{X}}$ (refer to Section 3.5). With $\delta = \{\delta(t) : \; t \geq 0\}$ denoting the stochastic intensity of the departure process, LBA for the departure stream takes the analogous form,

$$\mathsf{E}[\delta(t)|\mathbf{X}(t+)] \;=\; \mathsf{E}[\delta(t)], \tag{52}$$

where $\delta(t)$ denotes the stochastic intensity of the departure process, $N_{\mathbf{D}}$, in $\mathbf{X}$, and $\mathbf{D} \subset \mathbf{S} \times \mathbf{S}$ is the corresponding traffic set of state transitions. Note, that we write $\mathbf{X}(t+) = \mathbf{X}(t)$ (even though $\mathbf{X}$ is right-continuous) for emphasis and contrast with arrival processes.

Theorem 6.5
For a stationary Markovian queue, $\mathbf{X}$, possessing an ergodic distribution, π, the following are equivalent:

(i) $\mathbf{X}$ is quasi-reversible, i.e., (50) holds.

(ii) LBA holds for $(\mathbf{X}, N_{\mathbf{A}})$ and $(\mathbf{X}, N_{\mathbf{D}})$, i.e., (51) and (52) are satisfied.

(iii) The arrival and the departure processes are Poisson.

Proof:
We verify that (51) and (52) are equivalent to the two equations in (50) (ignoring the indices c and i in the latter). Making use of the relation between a queue and its time-reversed version (cf. Proposition 6.1), we have,

$$\begin{aligned}
\mathsf{E}[\delta(t)|\mathbf{X}(t+) = \mathbf{y}] \;&=\; \frac{\mathsf{E}[q(\mathbf{X}(t-), \mathbf{y})\mathbf{1}[\mathbf{X}(t+) = \mathbf{y}]]}{\mathsf{E}[\mathbf{1}[\mathbf{X}(t+) = \mathbf{y}]]} \\
&=\; \sum_{\mathbf{x}\in\mathbf{S}} \frac{\pi(\mathbf{x})\, q(\mathbf{x}, \mathbf{y})\, \mathbf{1}[(\mathbf{x}, \mathbf{y}) \in \mathbf{D}]}{\pi(\mathbf{y})} \\
&=\; \pi(\mathbf{y}) \sum_{\mathbf{y}\in\mathbf{S}} \frac{\tilde{q}(\mathbf{y}, \mathbf{x})\, \mathbf{1}[(\mathbf{y}, \mathbf{x}) \in \mathbf{A}]}{\pi(\mathbf{y})} \\
&=\; \sum_{\mathbf{y}\in\mathbf{S}} \tilde{q}(\mathbf{y}, \mathbf{x})\, \mathbf{1}[(\mathbf{y}, \mathbf{x}) \in \mathbf{A}], \quad \mathbf{y} \in \mathbf{S}.
\end{aligned} \tag{53}$$

On the other hand,

$$\alpha(t) \;=\; \sum_{\mathbf{y}\in\mathbf{S}} q(\mathbf{X}(t-), \mathbf{y})\, \mathbf{1}[(\mathbf{X}(t-), \mathbf{y}) \in \mathbf{A}]. \tag{54}$$

Suppose that the queue $\mathbf{X}$ is quasi-reversible. Combining (50) and (53), we get

$$\mathsf{E}[\delta(t)|\mathbf{X}(t+) = \mathbf{y}] \;=\; \mathsf{E}[\delta(t)] \;=\; \lambda_0, \tag{55}$$

whence (52) holds. In addition, (50) and (54) together implies that $\alpha(t) \equiv \lambda_0$, whence (51) holds.

For the converse, (52) implies (55), which on combining with (53), yields one of the two equations in (50). Further, (51) implies that $\alpha(t)$ is a deterministic function, independent of $\mathbf{X}(t-)$. Due to stationarity, $\alpha(t)$ must be a constant, which in turn must be equal to λ_0, due to equilibrium. Hence, the other equation in (50) follows from (54).

Thus, (i) and (ii) are equivalent. The equivalence of (ii) and (iii) follows Corollary 5.10. □

Theorem 6.5 and Corollary 5.5 ensure that PASTA holds in a quasi-reversible queue, so that both arrivals and departures there see time averages. Recall that in quasi-reversible networks, internal Poisson traffic is not guaranteed. However, LBA and ASTA still hold. To see that, pick some node i and let $\alpha_i^{(c)}(t)$ denote the stochastic intensity of the total arrival process of class-c jobs at node i, including both external arrivals and internal transitions (from *other* nodes). Then,

$$\alpha_i^{(c)}(t) \;=\; \lambda_0\, r_{0,i}^{(c)} \;+\; \sum_{d\neq c, j\neq i} \mu_j^{(d)}(\mathbf{X}_j(t-))\, r_{j,i}^{(d,c)},$$

where

$$\mu_j^{(d)}(\mathbf{X}_j(t-)) = \sum_{\mathbf{y}_j \in \mathbf{S}_j} q_j(\mathbf{X}_j(t-), \mathbf{y}_j)\, \mathbf{1}[(\mathbf{X}_j(t-), \mathbf{y}_j) \in \mathbf{D}_j^{(d)}]$$

is the (state-dependent) service rate for class-d jobs at node j, $\mathbf{D}_j^{(d)} \subset \mathbf{S}_j \times \mathbf{S}_j$ is the corresponding traffic set, and $\mathbf{X}_j$ is the state at node j with state space $\mathbf{S}_j$. Since the product-form invariant distribution in Proposition 6.4 implies that $\mathbf{X}_i$ is independent of $\mathbf{X}_j$ for any $j \neq i$, we have

$$\mathsf{E}[\alpha_i^{(c)}(t)|\mathbf{X}_i(t-)] \;=\; \mathsf{E}[\alpha_i^{(c)}(t)],$$

implying that LBA is satisfied. Consequently, arrivals at node i do see time averages there; in particular, arrivals see the invariant distribution, π_i, provided the arriving job is excluded.

Do arrivals at node i see time averages at the other nodes as well? The answer is affirmative, and the argument proceeds through an examination of the departure process at node i, along with its time-reversed version. Let $\delta_i^{(c)} = \{\delta_i^{(c)}(t) : \; t \geq 0\}$ denote the stochastic intensity of the departure process of class-c jobs from node i, excluding those leaving the network. Denoting by $\mathbf{X}_{-i} = (\mathbf{X}_1, \ldots, \mathbf{X}_{i-1}, \mathbf{X}_{i+1}, \ldots, \mathbf{X}_m)$ the partial state process of $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_m)$, with the i-th component omitted, we have

$$\delta_i^{(c)}(t) \;=\; \mu_i^{(c)}(\mathbf{X}_i(t-))\,[1 - r_{i,0}^{(c)}],$$

which is independent of $\mathbf{X}_{-i}(t-)$. Hence, departures from node i that are routed to other parts of the network observe time averages at all the other nodes. Note that here $\mathbf{X}_i(t-)$ denotes the state at node i *just before* the departure of the job in motion, i.e., it includes the departing job. Similarly, $\mathbf{X}_{-i}(t-)$ denotes the state of the rest of the network, which *does not* include the job in motion.

The above argument applies to the time-reversed network as well. Hence, we may conclude that *arrivals* at node i, also see time averages at all the other nodes, *just after* arrival. The following proposition summarizes the foregoing discussion.

Proposition 6.6
In an open network of quasi-reversible queues (such as the one in Proposition 6.4), LBA holds, at each node i,

- for the arrival process $N_{\mathbf{A}_i^{(c)}}$, with stochastic intensity $\alpha_i^{(c)}$, and the left-continuous state process, $\mathbf{X}_i^{(c)}(t-)$,
- for the arrival process $N_{\mathbf{A}_i^{(c)}}$, with stochastic intensity $\alpha_i^{(c)}$, and the right-continuous state process, $\mathbf{X}_{-i}^{(c)}(t+)$,
- for the departure process, $N_{\mathbf{D}_i^{(c)}}$, with stochastic intensity $\delta_i^{(c)}$, and the left-continuous state process, $\mathbf{X}_{-i}^{(c)}(t-)$.

Hence, ASTA holds. In particular, arrivals at node i see time averages at node i, *just before* arrival at node i; they see time averages at all other nodes *just after* arrival at node i. $\square$

Remark 6.7
Note that Proposition 6.6 also implies that any job in *transition* sees the ergodic distribution of the whole network, but with itself excluded from the state description. Intuitively, the job in transition may be considered as visiting an extra node, which is not part of the original network, and from there it observes the ergodic distribution of all other nodes, which constitute the original network. It is crucial that the observation be made *just after* arrival at the extra node to ensure that the job in motion does not "see itself". To make this argument rigorous, it only remains to let the service rate at the extra node go to infinity, so that in the limit, no time is actually spent at the extra node. This argument was pointed out in [20]. $\square$

As noted in [20] through a network example, it is also possible to use Theorem 5.3, or equivalently, the Palm formula in (43), to derive job-averages directly. This class of results falls in the domain of the so-called *Arrival Theorem*, to be illustrated in detail in the next two examples.

6.3.1 *The Arrival Theorem in Quasi-Reversible Networks*

In this example, it will be shown directly (in contrast to the indirect argument in Remark 6.7) that any job in transition sees the ergodic state distribution, *with itself excluded*.

Let $\mathbf{X}$ denote the state of the open network in Proposition 6.6, associated with a job's transition from node i to node j and changing class from c to d. The state does *not* include the job in transition. Let f be a bounded, measurable function on S. Then, conditioning upon the states immediately before and immediately after the transition, and making use of the relations in (50), we have

$$\begin{aligned}\mathsf{E}[f(\mathbf{X})\,\delta_{i,j}^{(c,d)}] \;=\; &\sum_{\mathbf{x}_{-(i,j)}} \prod_{k\neq i,j} \pi_k(\mathbf{x}_k) \sum_{\mathbf{x}_i}\sum_{\mathbf{y}_i} \pi_i(\mathbf{y}_i)\, q_i(\mathbf{y}_i,\mathbf{x}_i)\, \mathbf{1}[(\mathbf{y}_i,\mathbf{x}_i)\in \mathbf{D}_i^{(c)}] \\ &\times \sum_{\mathbf{x}_j} \pi_j(\mathbf{x}_j) \sum_{\mathbf{y}_j} q_j(\mathbf{x}_j,\mathbf{y}_j)\, \mathbf{1}[(\mathbf{x}_j,\mathbf{y}_j)\in \mathbf{A}_j^{(d)}]\, f(\mathbf{x})\, \frac{r_{i,j}^{(c,d)}}{\lambda_j^{(d)}}\end{aligned}$$

$$
\begin{aligned}
&= r_{i,j}^{(c,d)} \sum_{\mathbf{x}_{-(i,j)}} \prod_{k \neq i,j} \pi_k(\mathbf{x}_k) \sum_{\mathbf{x}_j} \pi_j(\mathbf{x}_j)\, f(\mathbf{x}) \\
&\qquad \times \sum_{\mathbf{x}_i} \pi_i(\mathbf{x}_i) \sum_{\mathbf{y}_i} \tilde{q}_i(\mathbf{x}_i, \mathbf{y}_i)\, \mathbf{1}[(\mathbf{x}_i, \mathbf{y}_i) \in \mathbf{A}_i^{(c)}] \\
&= \lambda_i^{(c)}\, r_{i,j}^{(c,d)} \sum_{\mathbf{x}_{-(i,j)}} \prod_{k \neq i,j} \pi_k(\mathbf{x}_k) \sum_{\mathbf{x}_i} \pi_i(\mathbf{x}_i) \sum_{\mathbf{x}_j} \pi_j(\mathbf{x}_j)\, f(\mathbf{x}) \\
&= \mathsf{E}[\delta_{i,j}^{(c,d)}]\, \mathsf{E}[f(\mathbf{X})].
\end{aligned}
$$

Hence, following the Palm formula (43), we have $\mathsf{E}^0[f(\mathbf{X})] = \mathsf{E}[f(\mathbf{X})]$. In particular, letting $f(\mathbf{X}) = \mathbf{1}[\mathbf{X} = \mathbf{x}]$, it follows that a job in transition from node i to node j sees the ergodic network state distribution, but with itself excluded from the state description.

Note that the derivation above does *not* constitute a verification of LBA, since the state, $\mathbf{X}$, seen by the job in transition with itself excluded, is neither the state $\mathbf{X}(t-)$ nor the state $\mathbf{X}(t+)$, as required by LBA. Note carefully that if the above were a verification of LBA, it would imply that the process in question were Poisson, following Corollary 5.10. In particular, this does *not* contradict Theorem 3.14, since the state seen by a job in transition (with itself excluded) does *not* constitute a Markov process.

6.3.2 *The Arrival Theorem in Closed Networks*

In this example, the same approach is applied to a *closed* network. This example is similar to Example 4 in [20], where a closed network with multiple servers at each node is analyzed, but the closed network here is more general.

Consider a single-class network, with N jobs circulating among its m nodes. The (state-dependent) service rate at node i is denoted by $\mu_i(n)$, when there are n jobs there, subject to $\mu_i(0) = 0$ and $\mu_i(n) > 0$ for $n > 0$. The routing matrix is $R = [r_{i,j}]$, which is a stochastic matrix (each rows sums to unity). For any vector $\mathbf{z} = (z_1, \ldots, z_m)$, denote $|\mathbf{z}| = \sum_i z_i$. Let $\mathbf{v} = (v_1, \ldots, v_m)$ be a solution to the system of equations,

$$
v_i \;=\; \sum_{j=1}^{m} v_j\, r_{j,i}, \quad i = 1, ..., m.
$$

Consider the network state process, $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_m)$, of job totals at the nodes in equilibrium. Let $\mathbf{Y} = (\mathbf{Y}_1, ..., \mathbf{Y}_m)$ be a set of *independent* non-negative integer-valued random variables, satisfying

$$
v_i\, \mathsf{P}\{\mathbf{Y}_i = n - 1\} \;=\; \mu_i(n)\, \mathsf{P}\{\mathbf{Y}_i = n\}, \quad n = 1, 2, ...;\; i = 1, ..., m. \tag{56}
$$

It is known that the stationary (invariant) distribution of $\mathbf{X}$ is

$$
\mathsf{P}\{\mathbf{X} = \mathbf{x}\} \;=\; \prod_{i=1}^{m} \frac{\mathsf{P}\{\mathbf{Y}_i = \mathbf{x}_i\}}{\mathsf{P}\{|\mathbf{Y}| = N\}}, \quad \mathbf{x} \in \mathbf{S}_N, \tag{57}
$$

where $\mathbf{S}_N = \{\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_m) : \; \mathbf{x}_i \geq 0,\; i \in M,\; |\mathbf{x}| = N\}$. >From the joint distribution, (57), one may compute the marginal distributions,

$$
\mathsf{P}\{\mathbf{X}_i = \mathbf{x}_i\} \;=\; \mathsf{P}\{\mathbf{Y}_i = \mathbf{x}_i\} \sum_{\mathbf{x}_{-i} \in \mathbf{S}_{N - \mathbf{x}_i}} \frac{\mathsf{P}\{\mathbf{Y}_{-i} = \mathbf{x}_{-i}\}}{\mathsf{P}\{|\mathbf{Y}| = N\}}, \tag{58}
$$

and verify that

$$\mathsf{E}[\mu_i(\mathbf{X}_i)] \;=\; v_i\,\frac{\mathsf{P}\{|\mathbf{Y}| = N-1\}}{\mathsf{P}\{|\mathbf{Y}| = N\}}. \qquad (59)$$

Since the $\mathbf{X}_i$ in the closed network are *not* independent, due to the constraint $|\mathbf{X}(t)| = N$, one cannot expect LBA to hold. However, as mentioned earlier, one may use Theorem 5.3, or equivalently, the Palm formula in (43), to derive event averages, and compare them against their time-average counterparts. To this end, pick a node, say 1, and consider its departure process, with stochastic intensity $\delta_1(t) = \mu_1(\mathbf{X}_1(t-))$. We are interested in the distribution of the state, as seen by a departing job from node 1, but with the departing job excluded from the state description. Now, while $\mathsf{P}\{\mathbf{X}(t-) \in B\} = \mathsf{P}[\mathbf{Y} \in B|\,|\mathbf{Y}| = N]$ (see (57)), the state, $\mathbf{X}(t+)$, left behind by the departing job (with itself excluded), follows the *ergodic* distribution, $\mathsf{P}\{\mathbf{Y} - \mathbf{e}_1|\,|\mathbf{Y}| = N\}$, where $\mathbf{e}_i$ is the unit vector of coordinate i. Making use of (58), one has

$$\begin{aligned}
\mathsf{E}[f(\mathbf{X})\,\mu_1(\mathbf{X}_1)] &= \mathsf{E}[f(\mathbf{Y}-\mathbf{e}_1)\,\mu_1(\mathbf{Y}_1)|\,|\mathbf{Y}| = N] \\
&= \sum_{x_1=1}^{N} \sum_{\mathbf{x}_{-1}\in\mathbf{S}_{N-\mathbf{x}_1}} f(\mathbf{x}_1 - 1, \mathbf{x}_{-1})\,\mu_1(\mathbf{x}_1)\,\frac{\mathsf{P}\{\mathbf{Y}_1 = \mathbf{x}_1\}\,\mathsf{P}\{\mathbf{Y}_{-1} = \mathbf{x}_{-1}\}}{\mathsf{P}\{|\mathbf{Y}| = N\}} \\
&= v_1 \sum_{\mathbf{x}_1=1}^{N} \sum_{\mathbf{x}_{-1}\in\mathbf{S}_{N-\mathbf{x}_1}} f(\mathbf{x}_1 - 1, \mathbf{x}_{-1})\,\frac{\mathsf{P}\{\mathbf{Y}_1 = \mathbf{x}_1 - 1\}\,\mathsf{P}\{\mathbf{Y}_{-1} = \mathbf{x}_{-1}\}}{\mathsf{P}\{|\mathbf{Y}| = N\}} \\
&= v_1 \sum_{\mathbf{x}_1=0}^{N-1} \sum_{\mathbf{x}_{-1}\in\mathbf{S}_{N-1-\mathbf{x}_1}} f(\mathbf{x}_1, \mathbf{x}_{-1})\,\frac{\mathsf{P}\{\mathbf{Y}_1 = \mathbf{x}_1]\,\mathsf{P}\{\mathbf{Y}_{-1} = \mathbf{x}_{-1}\}}{\mathsf{P}\{|\mathbf{Y}| = N\}} \\
&= v_1\,\frac{\mathsf{P}\{|\mathbf{Y}| = N-1\}}{\mathsf{P}\{|\mathbf{Y}| = N\}}\,\mathsf{E}[f(\mathbf{Y}_1, \mathbf{Y}_{-1})|\,|\mathbf{Y}| = N-1] \\
&= \mathsf{E}[\mu_1(\mathbf{X}_1)]\,\mathsf{E}[f(\mathbf{Y}_1, \mathbf{Y}_{-1})|\,|\mathbf{Y}| = N-1].
\end{aligned}$$

Thus, the (event) average observed by jobs departing from node 1 is

$$\mathsf{E}^0[f(\mathbf{X})] \;=\; \mathsf{E}[f(\mathbf{Y}_1, \mathbf{Y}_{-1})|\,|\mathbf{Y}| = N-1].$$

In particular, for $f(\mathbf{X}) = \mathbf{1}[\mathbf{X} = \mathbf{x}]$,

$$\mathsf{P}^0\{\mathbf{X} = \mathbf{x}\} \;=\; \mathsf{P}\{\mathbf{Y} = \mathbf{x}|\,|\mathbf{Y}| = N-1\}.$$

The avove result is the well-known Arrival Theorem in a closed queueing network; see, e.g., Kelly [14], Sevcik and Mitrani [24], Lavenberg and Reiser [16].

Again, we note that the above does not constitute a verification of LBA, since the state $\mathbf{X}$ is associated with a job in transition. It is neither $\mathbf{X}(t-)$ nor $\mathbf{X}(t+)$ as require in the verification of LBA.

To summarize, we have established that for Markovian queues (under stationary ergodicity), LBA is equivalent to quasi-reversibility. While quasi-reversibility is in general lost when the queue is connected into a network, LBA is preserved. For a quasi-reversible queue in isolation, LBA implies Poisson arrivals, and hence PASTA; in a network of quasi-reversible queues, LBA implies (or more precisely, is equivalent to) ASTA. Hence, it appears that LBA, and hence ASTA, are more fundamental properties that characterize quasi-reversible queues.

Finally, we note that in a network of quasi-reversible queues, LBA can also be used to identify Poisson flows. Along this line, conditions that are similar to those in Serfozo [23] and in Walrand [28] (pp. 166-167) can be established.

7 SUMMARY AND OPEN PROBLEMS

Three approaches to the ASTA problem may be discerned in the literature. The first is the elementary approach, which treats event averages as stochastic Rieman-Stieltjes integrals. This is the approach taken in Melamed and Whitt [20], and exemplified here in Section 4. The second is the martingale approach, which exploits properties of the martingale decomposition, stochastic integration, martingale SLLN, and the theory of stochastic intensities and the compensators, leading to ASTA in the most general and direct manner. This approach is taken in Brémaud, Kannurpatti and Mazumdar [6], and in Melamed and Whitt [21], among others. It is also the approach followed in Section 5. The third approach focuses on the stationary case, and views the problem as relating two expectations with respect to the underlying probability measure, P, and its Palm-measure counterpart, P^0, respectively. This approach is briefly summarized in Section 5.4. It is due to König and Schmidt [15] which first presented ASTA in the form of the Palm inversion formula [see (44)]. The connection between this approach and the martingale approach was brought out in Brémaud [5], through Papangelou's Lemma (Lemma 5.11), which relates the stochastic intensity to the Radon-Nykodym derivative $d\mathsf{P}^0/d\mathsf{P}$. This result reveals that the role played by the counting (or, "arrival") process N, through its intensity, is no more than a change of measure, from P to P^0.

The LBA condition, (37) and (38), is central to ASTA, particularly in the stationary ergodic setting and in queueing network applications. The intuition behind LBA is that for ASTA to hold, the time points ("arrival" epochs) selected for making observations, should not bias the observed state process.

To conclude, we comment on some open problems as candidates for further studies.

- To date, the best-known examples of ASTA, *in the absence* of Poisson flows, are all in the area of stationary Markovian queues, and in particular, networks of quasi-reversible queues (also known as "product-form" networks). This is hardly surprising, in view of the equivalence between LBA and quasi-reversibility (Theorem 6.5) and the fact that LBA is preserved in a network of quasi-reversible queues (Section 6.3). On the other hand, it remains an interesting and perhaps challenging open problem to identify ASTA results in other non-stationary and non-Markovian settings.

- The martingale approach to ASTA in Section 5 is based on a martingale strong law of large numbers; refer to Lemma 5.1. It turns out that one can also make use of a martingale central limit theorem to study the ASTA problem (Yao [32], Glynn, Melamed and Whitt [11]). This approach leads to results that are useful in constructing confidence intervals for estimators of job and time averages, and in performing statistical tests. For details refer to [11] and Glasserman [10]. Further studies along this line, we believe, should lead to more fruitful applications of ASTA.

- ASTA was originally motivated, in the queueing context, by the need to relate job (or, customer) based performance (e.g., delay) to system performance (e.g., time-average work load). Hence, the discrete (or, "embedded") observation epochs are naturally chosen as the arrival epochs of jobs or customers. In other applications, however, the choice of such sampling epochs might be far from obvious. A quick example is a fluid network that processes *continuous* flows, instead of dicrete jobs. Formulating and studying an analogous ASTA problem in this setting should be an interesting undertaking.

Acknowledgments

We are grateful to Ward Whitt for his careful reading of the manuscript and his valuable comments. We thank Arif Merchant for reading and commenting on the manuscript.

References

[1] APOSTOL, T.M. 1957. *Mathematical Analysis.* Addison-Wesley, New York.

[2] BACCELLI, F. AND BRÉMAUD, P. 1994. *Elements of Queueing Theory: Palm-Martingale Calculus and Stochastic Recurrences.* Springer-Verlag, New York.

[3] BILLINGSLEY, P. 1968. *Convergence of Probability Measures*. Wiley, New York.

[4] BRÉMAUD, P. 1981. *Point Processes and Queues.* Springer-Verlag, New York.

[5] BRÉMAUD, P. 1989. Characteristics of Queueing Systems Observed at Events and the Connection between Stochastic Intensity and Palm Probability. *Queueing Systems: Theory and Applications* **5**, 99-112.

[6] BRÉMAUD, P., KANNURPATTI, R. AND MAZUMDAR, R. 1992. Event and Time Averages: A Review and Some Generalizations. *Adv. Appl. Prob.* **24**, 377-411.

[7] DESCLOUX, A. 1967. On the Validity of a Particular Subscriber's View. *Proc. Fifth International Traffic Conference*, New York, p. 309.

[8] EL-TAHA, M. AND STIDHAM, S. 1994. Sample-Path Stability Conditions for Multiserver Input-Output Processes. *J. of Applied Mathematics and Stochastic Analysis* **7**, 437–456.

[9] FRANKEN, P., KÖNIG, D., ARNDT U. AND SCHMIDT, V. 1981. *Queues and Point Processes.*, Akademie-Verlag, Berlin. (Also Wiley, Chichester, 1982.)

[10] GLASSERMAN, P. 1993. Filtered Monte Carlo. *Math. Opns. Res.* **18**, 610-634.

[11] GLYNN, P., MELAMED, B. AND WHITT, W. 1993. Estimating Customer and Time Averages. *Operations Res.* **41**, 400-408.

[12] GREEN, L. AND MELAMED, B. 1990. An Anti-Pasta Result for Markovian Systems. *Operations Res.* **38**, 173–175.

[13] HALFIN, S. 1983. Batch Delays Versus Customer Delays. *Bell Syst. Tech. J.* **62**, 2011–2015.

[14] KELLY, F.P. 1979. *Reversibility and Stochastic Networks.* Wiley, New York.

[15] KÖNIG, D. AND SCHMIDT, V. 1980. Imbedded and Non-Imbedded Stationary Characteristics of Queueing Systems with Varying Service Rate and Point Processes. *J. Appl. Prob.* **17**, 753-767.

[16] LAVENBERG, S.S. AND REISER, M. 1980. Stationary State Probabilities at Arrival Instants for Closed Queueing Networks with Multiple Types of Customers. *J. Appl. Prob.* **17**, 1048-1061.

[17] LIPTSER, R.SH. 1980. A Strong Law of Large Numbers for Local Martingales. *Stochastics* **3**, 217-228.

[18] MAKOWSKI, A., MELAMED, B. AND WHITT, W. 1989. On Averages Seen by Arrivals in Discrete Time. *Proceedings of the 28th IEEE CDC*, Tampa, Florida, 1989.

[19] MELAMED, B. 1982. On Markov Jump Processes Imbedded at Jump Epochs and Their Queueing-Theoretic Applications. *Math. Opns. Res.* **7**, 111–128.

[20] MELAMED, B. AND WHITT, W. 1990. On Arrivals That See Time Averages. *Operations Res.* **38**, 156-172.

[21] MELAMED, B. AND WHITT, W. 1990. On Arrivals That See Time Averages: A Martingale Approach. *J. Appl. Prob.* **27**, 376-384.

[22] PAPANGELOU, F. 1972. Integrability of Expected Increments of Point Processes and a Related Change of Time. *Am. Math. Soc.* **165**, 483-506.

[23] SERFOZO, R.F. 1989. Poisson Functionals of Markov Processes and Queueing Networks. *Adv. Appl. Prob.* **21**, 595-611.

[24] SEVCIK, K.C. AND MITRANI, I. 1981. The Distribution of Queueing Network States at Input and Output Instants. *J. Assoc. Comput. Mach.* **28**, 358-371.

[25] STIDHAM, S. AND EL-TAHA, M. 1995. Sample-Path Techniques in Queueing Theory. In this book.

[26] STRAUCH, R.E. 1970. When a Queue Looks the Same to an Arriving Customer as to an Observer. *Mgmt. Sci.* **17**, 140-141.

[27] VAN DOORN, E.A. AND REGTERSCHOT, G.J.K. 1988. Conditional PASTA. *Opns. Res. Lett.* **7**, 229–232.

[28] WALRAND, J. 1988. *An Introduction to Queueing Networks.* Prentice-Hall, Englewood Cliffs, New Jersey.

[29] WATANABE, S. 1964. On Discontinuous Additive Functionals and Lévy Measures of a Markov Process. *Japan J. Math.* **34**, 53-70.

[30] WHITT, W. 1979. A Note on Poisson Arrivals Seeing Time Averages. Unpublished Paper, AT&T Bell Laboratories, Murray Hill, New Jersey.

[31] WOLFF, R.W. 1982. Poisson Arrivals See Time Averages. *Operations Res.* **30**, 223-231.

[32] YAO, D.D. 1992. On Wolff's PASTA Martingale. *Operations Res.* **40**, 352-355.