MATHEMATICS OF OPERATIONS RESEARCH
Vol. 18, No. 3, August 1993
Printed in U.S.A.

FILTERED MONTE CARLO

PAUL GLASSERMAN

By a *filtered Monte Carlo* estimator we mean one whose constituent parts—summands or integral increments—are conditioned on an increasing family of σ-fields. Unbiased estimators of this type are suggested by compensator identities. Replacing a point-process integrator with its intensity gives rise to one class of examples; exploiting Lévy's formula gives rise to another.

We establish variance inequalities complementing compensator identities. Among estimators that are (Stieltjes) stochastic integrals, we show that filtering reduces variance if the integrand and the increments of the integrator have conditional positive correlation. We also provide more primitive hypotheses that ensure this condition, making use of stochastic monotonicity properties. Our most detailed conditions apply in a Markov setting where monotone, up-down, and convex generators play a central role. We give examples. As an application of our results, we compare certain estimators that do and do not exploit the property that Poisson arrivals see time averages.

1. Introduction. Compensator identities equate expectations. Viewed with a Monte Carlo bent, they suggest competing estimators of a common mean and lead one to ask which of two quantities, equal in expectation, has smaller variance. No simple answer applies in all settings; but we show here how variance comparisons can be made in the presence of special structure. Stochastic monotonicity conditions are central to our results.

To fix ideas, let $(\Omega, \mathscr{F}, P)$ be a probability space and $\{\mathscr{F}_t,\ t \geqslant 0\}$ an increasing family of sub-σ-fields of $\mathscr{F}$—a filtration. Take $\{\mathscr{F}_t,\ t \geqslant 0\}$ to be right-continuous, in the sense that, for all $t \geqslant 0$, $\mathscr{F}_t$ is the intersection of $\{\mathscr{F}_s,\ s > t\}$. Let $\{A_t,\ t \geqslant 0\}$ be an adapted, process with right-continuous paths of bounded variation, almost surely. Under a variety of additional technical conditions, there exists a *predictable* process $\{\tilde{A}_t,\ t \geqslant 0\}$ such that

$$\mathbf{E}\left[\int_0^T Z_s\,dA_s\right] = \mathbf{E}\left[\int_0^T Z_s\,d\tilde{A}_s\right], \tag{1}$$

for all predictable processes $\{Z_s,\ s \geqslant 0\}$ and stopping times T, the integrals inside the expectations existing pathwise in the Stieltjes sense. See, for example, Brémaud (1981, p. 245), Jacod and Shiryaev (1987, p. 33), and Sharpe (1988, p. 392). Typically, $\{A_t - \tilde{A}_t,\ t \geqslant 0\}$ is a local martingale, and $\tilde{A}$ is called the *compensator* of A. Thus, (1) is an example of a compensator identity.

Now consider the problem of computing the expectations in (1) through simulation; problems of this type arise naturally in queueing theory, for example. When samples of each of the two stochastic integrals appearing in (1) can be generated with reasonable effort, they provide alternative estimators of the same quantity. The effort required to obtain a given precision using either estimator depends critically on its

*Received May 1, 1991; revised May 27, 1992.
AMS 1980 subject classification. Primary: 65C05, 60H99; Secondary: 68J10.
IAOR 1973 subject classification. Main: Simulation. Cross references: Stochastic Processes.
OR/MS Index 1978 subject classification. Primary: 764 Simulation/Efficiency; Secondary: 561 Probabaility.
Key words. Variance reduction, association, associated increments, monotone and convex generators, stochastic integrals.

0364-765X/93/1803/0610/$01.25

variance. Thus, we would like to have a variance inequality to complement the expectation identity (1).

A comparison of the stochastic integrals appearing in (1) begins by noting that conditional expectations play a role: informally, $d\tilde{A}_s = \mathbf{E}[dA_s | \mathcal{F}_{s-}]$, suggesting that

$$\text{(2)} \qquad \operatorname{Var}\left[\int_0^T Z_s \, d\tilde{A}_s\right] \leqslant \operatorname{Var}\left[\int_0^T Z_s \, dA_s\right].$$

However, $\int Z_s \, d\tilde{A}_s$ is not an ordinary *conditional Monte Carlo* estimator of $\mathbf{E}[\int Z_s \, dA_s]$ because different increments are conditioned on different σ-fields in passing from $\int Z_s \, dA_s$ to $\int Z_s \, d\tilde{A}_s$. Hence, correlations among the increments factor in the comparison, and variance reduction is not automatically guaranteed. When different terms are conditioned on different σ-fields, Bratley, Fox, and Schrage (1987) call the method *extended* conditional Monte Carlo. When the σ-fields form an increasing family, it seems appropriate to call it *filtered Monte Carlo*. Thus, in our terminology, $\int Z_s \, d\tilde{A}_s$ is a *filtered Monte Carlo* estimator of $\mathbf{E}[\int Z_s \, dA_s]$.

Equation (1) may look too abstract to be useful in practice, so we mention some important special cases. (Indeed, most of our results are formulated in settings less general than (1)—(2).) Let $\{X_t, \ t \geqslant 0\}$ be a right-continuous, pure-jump Markov process on a countable subset of $\mathbf{R}^d$. A discontinuous additive functional of X takes the form

$$D_t = \sum_{0 < s \leqslant t} h(X_{s-}, X_s), \qquad t \geqslant 0,$$

for some $h\colon \mathbf{R}^d \times \mathbf{R}^d \to \mathbf{R}$ satisfying $h(x, x) = 0$ for all x. This functional assigns a cost or reward $h(x, y)$ to a transition from x to y. If X has generator Q, and if we define $h_Q(x) = \sum_y Q(x, y) h(x, y)$, a version of (1) yields

$$\text{(3)} \qquad \mathbf{E}\left[\sum_{0 < s \leqslant T} h(X_{s-}, X_s)\right] = \mathbf{E}\left[\int_0^T h_Q(X_s) \, ds\right],$$

known as *Lévy's formula*. When h satisfies $h(x, y) = f(x)[g(y) - g(x)]$ for some functions f and g, this reduces to

$$\text{(4)} \qquad \mathbf{E}\left[\int_0^T f(X_{s-}) \, dg(X_s)\right] = \mathbf{E}\left[\int_0^T f(X_s) Qg(X_s) \, ds\right],$$

with $Qg(x) = \sum_y Q(x, y) g(y)$. When $f \equiv 1$, (4) yields *Dynkin's formula*,

$$\text{(5)} \qquad \mathbf{E}[g(X_T)] = \mathbf{E}\left[g(X_0) + \int_0^T Qg(X_s) \, ds\right].$$

Identities (3) and (4) are relevant when a cost or reward is associated with transitions. Identity (5) is useful in evaluating a *terminal* cost or reward $\mathbf{E}[g(X_T)]$.

Related results apply to a right-continuous point process $\{N_t, \ t \geqslant 0\}$ with predictable stochastic intensity $\{\lambda_t, \ t \geqslant 0\}$. By definition,

$$\text{(6)} \qquad \mathbf{E}\left[\int_0^\infty Z_s \, dN_s\right] = \mathbf{E}\left[\int_0^\infty Z_s \lambda_s \, ds\right]$$

for all nonnegative, predictable processes Z, and $N_t - \int_0^t \lambda_s \, ds$ is a local martingale;

see Brémaud (1981, p. 27). To see how (6) can be used, let N be the (right-continuous) arrival process to a queue and let $\{Z_t,\ t \geqslant 0\}$ be the corresponding (left-continuous) queue-length process. Then

$$\mathbf{E}\left[\int_0^T Z_s\, dN_s\right] \Big/ \mathbf{E}[N_T]$$

is the mean queue length found by arrivals in the interval $[0, T]$. Alternative estimators replace N with $\int \lambda_s\, ds$ in the numerator, denominator or both.

When N is a rate-λ Poisson process and PASTA holds (i.e., Poisson arrivals see time averages), a related problem compares arrival averages and time averages as estimators of a common limit. The time average is obtained from the arrival average by replacing dN_t/λ with dt. Our results identify some cases in which the time average has smaller asymptotic variance.

We obtain variance comparisons complementing (1) and (3)—(6) by imposing additional structure. Our key conditions ensure that the correlation among terms of the form $Z_{t_1}[\tilde{A}_{t_2} - \tilde{A}_{t_1}]$, $0 \leqslant t_1 < t_2$ is less than that among corresponding terms $Z_{t_1}[A_{t_2} - A_{t_1}]$. From this we establish variance reduction, as in (2). To ensure that "filtering" the estimator reduces correlation among its constituent parts, we impose stochastic order relations and put *association* properties to use, building on a variance comparison of Glynn and Iglehart (1988) for discrete-time Markov chains. We also draw on conditions from Lindqvist (1988) under which a process is associated, and conditions from Glasserman (1992) under which a process has associated increments. Indeed, the results here are generalizations of the asymptotic variance comparison in §6 of Glasserman (1992).

Section 2 spells out hypotheses implying (2) in a general setting. These provide a template for subsequent, more specific results. Section 3 develops the Markov case, providing variance inequalities for (3)—(5). Point-process integrators are taken up in §4, with special attention given to Markov (Z, N). Included there are implications for estimators that use PASTA. Section 5 illustrates our results with specific examples.

Convention: Throughout this paper, $\int_0^T$ means $\int_{[0,T]}$.

2. A general variance comparison. Theorem 2.1, below, provides high-level conditions for the variance comparison (2). This theorem points to the additional structure needed to get variance comparisons in specific cases.

To streamline our initial discussion, we consider a deterministic time horizon $t > 0$. Thus, we seek to compare the variances of

$$\int_0^t Z_s\, dA_s \quad \text{and} \quad \int_0^t Z_s\, d\tilde{A}_s. \tag{7}$$

Our first assumption is that Riemann-Stieltjes sums for these integrals converge in an L^2 sense. Consider a sequence of partitions $\{s_0^n = 0,\ s_1^n, \ldots, s_n^n = t\}$, $n = 1, 2, \ldots,$ of $[0, t]$, each refining the previous one, with mesh decreasing to zero as $n \to \infty$. We assume that for some such sequence of partitions we have

$$\lim_{n \to \infty} \mathbf{E}\left[\left(\sum_{i=0}^{n-1} Z_{s_i^n}\left[A_{s_{i+1}^n} - A_{s_i^n}\right] - \int_0^t Z_s\, dA_s\right)^2\right] = 0 \tag{8}$$

and

$$(9)\qquad \lim_{n\to\infty} \mathbf{E}\left[\left(\sum_{i=0}^{n-1} Z_{s_i^n}\left[\tilde{A}_{s_{i+1}^n} - \tilde{A}_{s_i^n}\right] - \int_0^t Z_s\, d\tilde{A}_s\right)^2\right] = 0.$$

Next, we impose a compensator-like condition in an L^2 sense. Specifically, we require that for the sequence of partitions used in (8) and (9) we have

$$(10)\qquad \lim_{n\to\infty} \mathbf{E}\left[\left(\sum_{i=0}^{n-1} \mathbf{E}\left[Z_{s_i^n}\left(A_{s_{i+1}^n} - A_{s_i^n}\right)\middle|\mathscr{F}_{s_i^n}\right] - Z_{s_i^n}\left(\tilde{A}_{s_{i+1}^n} - \tilde{A}_{s_i^n}\right)\right)^2\right] = 0;$$

compare Theorem 2.17 of Karr (1986) for an ordinary (point-process) compensator. By assuming (8)—(10) we dispense with the requirement that Z be predictable.

Finally—and most importantly—we need a conditional positive correlation property. We assume that for all $0 \leqslant u_1 < u_2 \leqslant t_1 < t_2 \leqslant t$,

$$(11)\qquad \mathbf{E}\left[Z_{u_1}\left(A_{u_2} - A_{u_1}\right)\middle|\mathscr{F}_{u_1}\right]\mathbf{E}\left[Z_{t_1}\left(A_{t_2} - A_{t_1}\right)\middle|\mathscr{F}_{u_1}\right]$$

$$\leqslant \mathbf{E}\left[Z_{u_1}\left(A_{u_2} - A_{u_1}\right)\cdot Z_{t_1}\left(A_{t_2} - A_{t_1}\right)\middle|\mathscr{F}_{u_1}\right], \quad \text{a.s.}$$

We can now prove

THEOREM 2.1. *Suppose that*
(i) *Z is adapted to* $\{\mathscr{F}_s,\ s \geqslant 0\}$;
(ii) *Riemann-Stieltjes sums for* (7) *converge in L^2, as in* (8)—(9);
(iii) *the L^2 compensator-like condition* (10) *holds; and*
(iv) *the conditional positive correlation property* (11) *holds.*
Then the identity (1) *is valid with $T = t$ and*

$$\operatorname{Var}\left[\int_0^t Z_s\, d\tilde{A}_s\right] \leqslant \operatorname{Var}\left[\int_0^t Z_s\, dA_s\right].$$

In other words, the filtered Monte Carlo estimator has lower variance.

PROOF. Let S_n and $\tilde{S}_n$ be the nth sum appearing in (8) and (9), respectively. Let I_t and $\tilde{I}_t$ be the integrals to which they converge (in mean-square). We will show that $\mathbf{E}[\tilde{I}_t^2] \leqslant \mathbf{E}[I_t^2]$ and $\mathbf{E}[\tilde{I}_t] = \mathbf{E}[I_t]$. Define

$$\Delta_n = \sum_{i=0}^{n-1} Z_{s_i^n}\left\{\left[\tilde{A}_{s_{i+1}^n} - \tilde{A}_{s_i^n}\right] - \mathbf{E}\left[A_{s_{i+1}^n} - A_{s_i^n}|\mathscr{F}_{s_i^n}\right]\right\};$$

then adaptedness of Z and condition (10) together imply that $\mathbf{E}[\Delta_n^2] \to 0$ as $n \to \infty$. By hypothesis, $\mathbf{E}[(\tilde{S}_n - \tilde{I}_t)^2] \to 0$, so also $\mathbf{E}[(\tilde{S}_n - \Delta_n - \tilde{I}_t)^2] \to 0$. Moreover, $\mathbf{E}[I_t^r] = \lim_{n\to\infty} \mathbf{E}[S_n^r]$ and $\mathbf{E}[\tilde{I}_t^r] = \lim_{n\to\infty} \mathbf{E}[(\tilde{S}_n - \Delta_n)^r]$, $r = 1, 2$. Thus, to prove $\mathbf{E}[\tilde{I}_t] = \mathbf{E}[I_t]$ and $\mathbf{E}[\tilde{I}_t^2] \leqslant \mathbf{E}[I_t^2]$ it is enough to show that $\mathbf{E}[\tilde{S}_n - \Delta_n] = \mathbf{E}[S_n]$ and $\mathbf{E}[(S_n - \Delta_n)^2] \leqslant \mathbf{E}[S_n^2]$ for all n.

For the first moments, note that

$$\mathbf{E}\left[\tilde{S}_n - \Delta_n\right] = \mathbf{E}\left[\sum_{i=0}^{n-1} Z_{s_i^n}\mathbf{E}\left[A_{s_{i+1}^n} - A_{s_i^n}|\mathscr{F}_{s_i^n}\right]\right]$$

$$= \mathbf{E}\left[\sum_{i=0}^{n-1} \mathbf{E}\left[Z_{s_i^n}\left(A_{s_{i+1}^n} - A_{s_i^n}\right)|\mathscr{F}_{s_i^n}\right]\right]$$

$$= \sum_{i=0}^{n-1} \mathbf{E}\left[Z_{s_i^n}\left(A_{s_{i+1}^n} - A_{s_i^n}\right)\right]$$

$$= \mathbf{E}[S_n],$$

where the second equation uses the adaptedness of Z.

Since the expectations are equal, we can compare variances rather than second moments. The variance of S_n is

$$(12) \qquad 2\sum_{i=0}^{n-1}\sum_{j=i+1}^{n-1} \operatorname{Cov}\left\{Z_{s_i^n}\left(A_{s_{i+1}^n} - A_{s_i^n}\right), Z_{s_j^n}\left(A_{s_{j+1}^n} - A_{s_j^n}\right)\right\}$$

$$+ \sum_{i=0}^{n-1} \operatorname{Var}\left\{Z_{s_i^n}\left(A_{s_{i+1}^n} - A_{s_i^n}\right)\right\},$$

and that of $\tilde{S}_n - \Delta_n$ is

$$(13) \qquad 2\sum_{i=0}^{n-1}\sum_{j=i+1}^{n-1} \operatorname{Cov}\left\{\mathbf{E}\left[Z_{s_i^n}\left(A_{s_{i+1}^n} - A_{s_i^n}\right)|\mathscr{F}_{s_i^n}\right], \mathbf{E}\left[Z_{s_j^n}\left(A_{s_{j+1}^n} - A_{s_j^n}\right)|\mathscr{F}_{s_j^n}\right]\right\}$$

$$+ \sum_{i=0}^{n-1} \operatorname{Var}\left\{\mathbf{E}\left[Z_{s_i^n}\left(A_{s_{i+1}^n} - A_{s_i^n}\right)|\mathscr{F}_{s_i^n}\right]\right\}.$$

Each of the variance terms in (13) is smaller than the corresponding term in (12) because conditional expectations reduce variance. The comparison of the covariance terms is analogous to that of Glynn and Iglehart (1988, Theorem 12) for discrete-time Markov chains. We need to show that, for all $i < j$,

$$\mathbf{E}\left[\mathbf{E}\left[Z_{s_i^n}\left(A_{s_{i+1}^n} - A_{s_i^n}\right)|\mathscr{F}_{s_i^n}\right] \cdot \mathbf{E}\left[Z_{s_j^n}\left(A_{s_{j+1}^n} - A_{s_j^n}\right)|\mathscr{F}_{s_j^n}\right]\right]$$

$$\leqslant \mathbf{E}\left[Z_{s_i^n}\left(A_{s_{i+1}^n} - A_{s_i^n}\right) \cdot Z_{s_j^n}\left(A_{s_{j+1}^n} - A_{s_j^n}\right)\right].$$

By first conditioning both sides on $\mathscr{F}_{s_i^n}$ and using $\mathscr{F}_{s_i^n} \subseteq \mathscr{F}_{s_j^n}$, this inequality follows from (11). □

Theorem 6.1 of Glasserman (1992) is a similar comparison for deterministic integrands and an infinite horizon; i.e., it compares the variance constants in central limit theorems for $t^{-1}A_t$ and $t^{-1}\tilde{A}_t$.

We view hypotheses (i)—(iii) of Theorem 2.1 as regularity conditions. Slightly different assumptions could be substituted without materially changing the proof. In subsequent sections, we elaborate (i)—(iii) but most of our attention centers on conditional positive correlation, as in (iv) and (11). We pursue more primitive conditions that ensure (11). Often, the processes Z and A are built up from underlying processes, and it is on these that we put conditions. In particular, we impose positive dependence. Various notions of positive dependence present themselves, but an important closure property makes *association* the most appropriate tool: increasing functions of associated random vectors are associated. Building Z and A monotonically from associated underlying processes we guarantee (11).

3. The Markov case. The Markov setting outlined around (3)—(5) is of special importance because of the theoretical simplifications that it permits and because of its applicability. Indeed, the alternative estimators provided by (1) are computationally relevant only in the presence of *some* Markov structure: processes that depend on their entire past generally cannot be simulated.

The following conventions are in force throughout this section: $X = \{X_t, t \geqslant 0\}$ denotes a right-continuous, pure-jump Markov process on a countable subset $\mathbf{S}$ of $\mathbf{R}^d$, $1 \leqslant d < \infty$. The process X has *bounded* generator Q: there exists a constant $\lambda < \infty$ for which $-\sup_x Q(x, x) < \lambda$. The filtration $\{\mathscr{F}_t, t \geqslant 0\}$ is the one generated by X; i.e., $\mathscr{F}_t = \sigma(\{X_s, 0 \leqslant s \leqslant t\})$ for all $t \geqslant 0$.

3.1. *Association properties for Markov processes.* Our goal is to obtain variance inequalities complementing (3)—(5). We do so, as explained at the end of §2, by imposing association properties on X. We first review, briefly, conditions for a Markov process to be associated (following Lindqvist (1988)) and conditions for it to have associated increments (reviewing some results from Glasserman (1992)).

Let $\mathbf{R}^d$ be endowed with a partial order, $\leqslant$, and assume this order is closed: for all x_n, y_n, x and y in $\mathbf{R}^d$, if $x_n \to x$ and $y_n \to y$ and $x_n \leqslant y_n$ for all n, then $x \leqslant y$. Also assume that $\leqslant$ is compatible with addition in $\mathbf{R}^d$: if $x \leqslant y$ then $x + z \leqslant y + z$ for all z. For $n = 1, 2, \ldots,$ let $\mathbf{R}^{nd} = (\mathbf{R}^d)^n$ have the product order (also denoted $\leqslant$) generated by $\leqslant$; i.e., if $x_i, y_i \in \mathbf{R}^d$, $i = 1, \ldots, n$, write $(x_1, \ldots, x_n) \leqslant (y_1, \ldots, y_n)$ if and only if $x_i \leqslant y_i$ for all $i = 1, \ldots, n$. A subset C of $\mathbf{R}^{nd}$ is an *upper set* if $x \in C$ and $x \leqslant y$ together imply $y \in C$. A measure μ on $\mathbf{R}^{nd}$ is *associated* if $\mu(C_1 \cap C_2) \geqslant \mu(C_1)\mu(C_2)$ for all (measurable) upper sets C_1, C_2. Equivalently, μ is associated if $\int fg\, d\mu \geqslant \int f\, d\mu \cdot \int g\, d\mu$ for all increasing functions $f, g\colon \mathbf{R}^{nd} \mapsto \mathbf{R}$ for which the integrals exist. (Henceforth, all sets and functions mentioned are assumed measurable without comment.) A unit mass is always associated, and on a totally ordered state space all measures are associated. If μ is associated and if the random vector Y has distribution μ, then we also call Y associated. See Esary, Proschan, and Walkup (1967) and Lindqvist (1988) for background on association, Fortuin, Kasteleyn, and Ginibre (1971) for related ideas.

A Markov transition kernel P is called *upward* if $P(x, \cdot\,)$ gives probability one to $\{y\colon y \geqslant x\}$. The kernel P is called *monotone* if $P(x, C)$ is an increasing function of x for all upper sets C. Call a Markov chain $\{Y_0, Y_1, \ldots\}$ monotone if its kernel is monotone. (For background, see Daley (1968), Keilson and Kester (1977), Kamae, Krengel, and O'Brien (1977), and Massey (1987).) Following Lindqvist (1988), call P *associated* if $P(x, \cdot\,)$ is an associated measure for all x. Lindqvist shows that monotone, associated kernels preserve association of measures: if P is monotone and associated and if μ is associated then so is $\mu * P(\cdot) = \int P(x, \cdot\,)\mu(dx)$. From this he proves that a Markov chain $\{Y_n, n \geqslant 0\}$ forms an associated sequence if its transition kernel is monotone and associated and if Y_0 is associated. The first condition is necessary if the chain is to be associated for *all* associated initial distributions.

Call P *convex* if $P(x, x + C)$ is increasing in x for all upper sets C; see Shaked and Shanthikumar (1988), Meester and Shanthikumar (1990) and Glasserman (1992). Convex kernels are monotone. Glasserman (1992, Corollary 3.4) shows that a Markov chain with associated initial distribution and a convex, associated kernel has *associated increments*: for all $0 < m < \infty$ and all $0 \leqslant n_0 < n_1 < \cdots < n_m$, the vectors $\{Y_{n_0}, Y_{n_1} - Y_{n_0}, \dots, Y_{n_m} - Y_{n_{m-1}}\}$ are associated.

Analogous results hold in continuous time, where a process is called associated if all its finite dimensional distributions are associated. A generator Q is said to be monotone if $Q(x, C) \leqslant Q(y, C)$ whenever C is an upper set, $x \leqslant y$ and either $x \in C$ or $y \notin C$; see Keilson and Kester (1977) and Massey (1987). Also, Q is upward if $Q(x, y) > 0 \Rightarrow x \leqslant y$. Call Q *up-down* if $Q(x, y) > 0$ implies $x \leqslant y$ or $y \leqslant x$. Harris (1977) shows that a monotone Markov process on a finite state space is associated (for all associated initial distributions) if and only if its generator is up-down. The same is true on a countable state space if the generator is bounded; see Anantharam (1991) and the reference there to Cox (1984).

Call Q convex if $Q(x, x + C)$ is increasing in x for all upper sets C; this property is introduced in Shaked and Shanthikumar (1988) and also used in Meester and Shanthikumar (1990) and Glasserman (1992). Convex generators are monotone. Glasserman (1992) shows that if X_0 is associated and if Q is convex and up-down, then for all $0 < m < \infty$ and all $0 \leqslant t_0 < t_1 < \cdots < t_m$,

$$\{X_{t_0}, X_{t_1} - X_{t_0}, \dots, X_{t_m} - X_{t_{m-1}}\}$$

are associated; i.e., X has associated increments.

3.2. *Variance comparisons: Fixed horizon.* With the tools of §3.1 in place, we can specialize Theorem 2.1 to the Markov setting and replace its hypotheses with simpler ones. We begin with a variance inequality to complement the Lévy formula (3). Initially, we take the time horizon to be a fixed t, $0 < t < \infty$.

For (3) to make sense, we need $h(x, x) = 0$ for all x in $\mathbf{S}$, the state space of X. We are, in fact, only concerned with the restriction of h to

$$\mathbf{S}_2 = \bigcup_{x \in \mathbf{S}} \{x\} \times \{y : Q(x, y) > 0\}.$$

Take h to be increasing in the product order on $\mathbf{S}_2$: for $(x_1, y_1), (x_2, y_2) \in \mathbf{S}_2$,

$$x_1 \leqslant x_2 \quad \text{and} \quad y_1 \leqslant y_2 \Rightarrow h(x_1, y_1) \leqslant h(x_2, y_2). \tag{14}$$

We also make use of the following integrability conditions:

$$\mathbf{E}\left[\left(\sum_{0 < s \leqslant t} |h(X_{s-}, X_s)|\right)^2\right] < \infty, \tag{15}$$

and

$$\mathbf{E}\left[\left(\int_0^t \sum_y Q(X_s, y)|h(X_s, y)|\, ds\right)^2\right] < \infty. \tag{16}$$

With this set-up, we prove

THEOREM 3.1. *Suppose the integrability conditions* (15) *and* (16) *hold. Suppose also that*

(a) h: $\mathbf{S}_2 \mapsto \mathbf{R}$ *is increasing; and*

(b) Q *is bounded, monotone, and up-down.*

Then (3) *holds* (*with* $T = t$) *and*

$$\mathrm{Var}\left[\int_0^t h_Q(X_s)\,ds\right] \leqslant \mathrm{Var}\left[\sum_{0<s\leqslant t} h(X_{s-}, X_s)\right]. \tag{17}$$

In other words, the filtered Monte Carlo estimator has lower variance.

PROOF. The outline of the proof follows that of Theorem 2.1. Take a sequence of evenly spaced partitions of $[0, t]$; thus, $s_i^n = it/n$, $i = 0, \dots, n$. To lighten notation, we sometimes drop the superscript, writing simply s_i and letting context determine the appropriate n.

Assume, for now, that h is bounded. The boundedness of Q implies that X has left limits in addition to being right-continuous, and that X makes only finitely many jumps in $[0, t]$, almost surely. It follows that

$$\lim_{n\to\infty} \sum_{i=0}^{n-1} h(X_{s_i}, X_{s_{i+1}}) = \sum_{0<s\leqslant t} h(X_{s-}, X_s) \tag{18}$$

and

$$\lim_{n\to\infty} \sum_{i=0}^{n-1} h_Q(X_{s_i})[s_{i+1} - s_i] = \int_0^t h_Q(X_s)\,ds, \tag{19}$$

almost surely. Simple bounds take us from almost-sure convergence to convergence in mean square. For bounded h, each sum in (18) is stochastically bounded by $\|h\| = \sup_{x,y}|h(x, y)|$ times a Poisson random variable with mean λt. Therefore, each sum has finite (absolute) moments of all orders bounded uniformly in n. Via Chung (1974, Theorems 4.5.2 and 4.5.4), this gives us L^2 convergence in (18). For (19), note that $\|h_Q\| \leqslant \lambda\|h\|$ so the sums are uniformly bounded by $\lambda\|h\|t$, giving convergence in L^2. These L^2 limits play the role of (8) and (9).

Next, we verify an analog of (10), namely

$$\lim_{n\to\infty} \mathbf{E}\left[\left(\sum_{i=0}^{n-1} \mathbf{E}\left[h(X_{s_i}, X_{s_{i+1}})\middle|X_{s_i}\right] - \int_0^t h_Q(X_s)\,ds\right)^2\right] = 0. \tag{20}$$

If P_t denotes the time-t transition kernel for X (i.e., $P_t(x, C) = P(X_t \in C|X_0 = x)$), then, almost surely,

$$\begin{aligned}
\mathbf{E}\left[h(X_{s_i}, X_{s_{i+1}})\middle|X_{s_i}\right] &= \sum_y P_{s_{i+1}-s_i}(X_{s_i}, y)h(X_{s_i}, y) \\
&= \sum_y Q(X_{s_i}, y)h(X_{s_i}, y)[s_{i+1} - s_i] + o(s_{i+1} - s_i) \\
&= h_Q(X_{s_i})[s_{i+1} - s_i] + o(s_{i+1} - s_i).
\end{aligned}$$

Substituting this into (20) and using the fact that (19) holds in L^2 as well as almost surely, we find that (20) is proved, but for the $o(s_{i+1} - s_i)$ terms. Thus, to establish (20) it remains to show that the contribution of these terms vanishes in L^2. By a

Taylor series argument, each $|o(s_{i+1} - s_i)|$ term is bounded by

$$\left|\sum P_{s_{i+1}-s_i}(X_{s_i}, y)h(X_{s_i}, y) - \sum Q(X_{s_i}, y)h(X_{s_i}, y)[s_{i+1} - s_i]\right|$$
$$\leqslant \lambda^2\|h\|(s_{i+1} - s_i)^2.$$

Moreover, for an evenly spaced partition, $s_{i+1} - s_i = t/n$. Thus,

$$\mathbf{E}\left[\left(\sum_{i=0}^{n-1} |o(s_{i+1} - s_i)|\right)^2\right] \leqslant \sum_{i=0}^{n-1} \lambda^2\|h\|(t/n)^2 = \lambda^2\|h\|t^2/n,$$

which indeed vanishes as $n \to \infty$.

Putting together (20) and mean-square convergence in (18)—(19) just as in Theorem 2.1, we find that

$$\operatorname{Var}\left[\sum_{0<s\leqslant t} h(X_{s-}, X_s)\right] = \lim_{n\to\infty} \operatorname{Var}\left[\sum_{i=0}^{n-1} h(X_{s_i}, X_{s_{i+1}})\right]$$

and

$$\operatorname{Var}\left[\int_0^t h_Q(X_s)\, ds\right] = \lim_{n\to\infty} \operatorname{Var}\left[\sum_{i=0}^{n-1} \mathbf{E}\left[h(X_{s_i}, X_{s_{i+1}})|X_{s_i}\right]\right],$$

and analogous expressions for the means. Equality of the means is immediate (mimic Theorem 2.1) so we proceed to the variances. Expanding the variances as in (12) and (13) and arguing as we did there, it becomes enough to show that

$$\mathbf{E}\left[h(X_{s_i}, X_{s_{i+1}})|X_{s_i}\right]\mathbf{E}\left[h(X_{s_j}, X_{s_{j+1}})|X_{s_i}\right] \tag{21}$$
$$\leqslant \mathbf{E}\left[h(X_{s_i}, X_{s_{i+1}})h(X_{s_j}, X_{s_{j+1}})|X_{s_i}\right], \quad \text{a.s.},$$

for all $i < j$. Since Q is monotone and up-down, X is associated for all associated initial conditions. Thus, given X_{s_i}, $\{X_{s_i}, X_{s_{i+1}}, X_{s_j}, X_{s_{j+1}}\}$ are conditionally associated. By hypothesis, h is an increasing function of pairs. Hence, given X_{s_i}, $h(X_{s_i}, X_{s_{i+1}})$ and $h(X_{s_j}, X_{s_{j+1}})$ have positive covariance, which is (21).

To conclude the proof, we must drop the condition that h is bounded. For any h and any $k \geqslant 0$ let h^k be h truncated at $+k$ and $-k$; i.e., $h^k(x, y) = h(x, y)$ if $|h(x, y)| < k$, but if $h(x, y) \geqslant k$ or $h(x, y) \leqslant -k$ then $h^k(x, y)$ is k or $-k$ accordingly. Clearly,

$$\lim_{k\to\infty} \sum_{0<s\leqslant t} h^k(X_{s-}, X_s) = \sum_{0<s\leqslant t} h(X_{s-}, X_s), \quad \text{a.s.},$$

and

$$\lim_{k\to\infty} \int_0^t h_Q^k(X_s)\, ds = \int_0^t h_Q(X_s)\, ds, \quad \text{a.s.},$$

where $h_Q^k(x) = \Sigma_y Q(x, y)h^k(x, y)$. Now take second moments. Dominated convergence, based on (15) and (16), justifies the interchange of limit and expectation. Since the variance inequality (17) holds for every h^k it holds for h as well. □

We can relax the monotonicity condition on h when X has associated increments in addition to being associated. This translates to a stronger condition on Q. Suppose that h is increasing in the following sense: for $(x_1, y_1), (x_2, y_2) \in \mathbf{S}_2$,

$$x_1 \leqslant x_2 \text{ and } y_1 - x_1 \leqslant y_2 - x_2 \Rightarrow h(x_1, y_1) \leqslant h(x_2, y_2). \tag{22}$$

The left side of (22) is a stronger condition than assuming (x_2, y_2) dominates (x_1, y_1) in the product order on $\mathbf{R}^d \times \mathbf{R}^d$; hence, (22) is weaker than (14). Using (22) and making Q convex we get

THEOREM 3.2. *Under the integrability assumptions* (15)—(16), *the conclusion of Theorem* 3.1 *holds if*
(a) $h\colon \mathbf{S}_2 \mapsto \mathbf{R}$ *satisfies* (22);
(b) Q *is bounded, convex, and up-down.*

PROOF. The proof of Theorem 3.1 applies until we get to (21). Condition (b) gives X associated increments for all associated initial conditions. Moreover, $\{X_{s_i}, X_{s_{i+1}} - X_{s_i}, X_{s_j}, X_{s_{j+1}} - X_{s_j}\}$ are increasing functions of X_{s_i} and increments starting from X_{s_i}; so, they are conditionally associated, given X_{s_i}. Condition (22) makes h an increasing function of the first two of these and the second two of these. Increasing functions of associated vectors have positive covariance, so (21) holds. □

This form of the result adapts to identity (4), i.e., to the case

$$h(x, y) = f(x)[g(y) - g(x)].$$

As in Shaked and Shanthikumar (1990) call $g\colon \mathbf{R}^d \mapsto \mathbf{R}$ *directionally convex* if

$$\left.\begin{aligned} x_1 \leqslant x_2 \wedge x_3;\ x_2 \vee x_3 \leqslant x_4 \\ x_1 + x_4 = x_2 + x_3 \end{aligned}\right\} \Rightarrow g(x_1) + g(x_4) \geqslant g(x_2) + g(x_3).$$

When $d = 1$ this coincides with ordinary convexity; more generally, directional convexity is equivalent to convexity in each coordinate plus supermodularity. With this we get the following generalization of a result in §6 of Glasserman (1992):

COROLLARY 3.3. *Suppose that* (15)—(16) *hold with* $h(x, y) = f(x)[g(y) - g(x)]$ *and* $h_Q(x) = f(x)Qg(x)$. *Suppose also that*
(a) f *is nonnegative and increasing*;
(b) g *is increasing and directionally convex*;
(c) Q *is bounded, convex and upward.*
Then (4) *holds* (*with* $T = t$) *and*

$$\operatorname{Var}\left[\int_0^t f(X_s) Qg(X_s)\, ds\right] \leqslant \operatorname{Var}\left[\int_0^t f(X_{s-})\, dg(X_s)\right]. \tag{23}$$

PROOF. The function $h(x, y) = f(x)[g(y) - g(x)]$ is zero on $\{x = y\}$. The result follows from Theorem 3.2 if we can show that h satisfies (22). Since Q is upward, $\mathbf{S}_2$ contains only points (x, y) with $x \leqslant y$. Thus, it is enough to check (22) under the additional condition that $x_i \leqslant y_i$, $i = 1, 2$. In this case, condition (b) gives $g(x_1) + g(y_2) \geqslant g(x_2) + g(y_1)$; that is, $g(y_2) - g(x_2) \geqslant g(y_1) - g(x_1)$. Condition (a) now gives

$$f(x_2)[g(y_2) - g(x_2)] \geqslant f(x_1)[g(y_1) - g(x_1)],$$

which is (22). □

3.3. *Variance comparison: Random horizon.* The results of §3.2 continue to hold if the time horizon t is replaced with the passage time to a *lower set*—a set C for which $x \in C$ and $y \leqslant x$ together imply $y \in C$. We make this precise in

THEOREM 3.4. *Suppose conditions* (a)—(b) *of either Theorem* 3.1 *or Theorem* 3.2 *hold. Suppose that* T *is the passage time to a lower set and that* (15)—(16) *hold with* t *replaced by* T. *Then* (3) *holds and*

$$\operatorname{Var}\left[\int_0^T h_Q(X_s)\,ds\right] \leqslant \operatorname{Var}\left[\sum_{0<s\leqslant T} h(X_{s-}, X_s)\right].$$

PROOF. First consider a truncated horizon $T \wedge t$ for any fixed $0 < t < \infty$. Writing

$$\sum_{0<s\leqslant T\wedge t} h(X_{s-}, X_s) = \sum_{0<s\leqslant t} \mathbf{1}\{s \leqslant T\} h(X_{s-}, X_s)$$

and

$$\int_0^{T\wedge t} h_Q(X_s)\,ds = \int_0^t \mathbf{1}\{s \leqslant T\} h_Q(X_s)\,ds$$

makes the time horizon deterministic, allowing us to use the arguments of §3.2. Let T be the passage time to C, a lower set. Without loss of generality, take C to be *absorbing*; this only effects the evolution of X after T. We now have $\mathbf{1}\{s \leqslant T\} = \mathbf{1}\{X_s \notin C\}$. Moreover, $\mathbf{1}\{x \notin C\}$ is an increasing function of x. Thus, $\mathbf{1}\{x \notin C\}h(x, y)$ satisfies condition (a) of Theorems 3.1 or 3.2 whenever h does. The result follows for horizon $T \wedge t$.

The result extends to unbounded stopping times through an additional limit. Almost surely,

$$\lim_{t\to\infty} \sum_{0<s\leqslant T\wedge t} h(X_{s-}, X_s) = \sum_{0<s\leqslant T} h(X_{s-}, X_s)$$

and

$$\lim_{t\to\infty} \int_0^{T\wedge t} h_Q(X_s)\,ds = \int_0^T h_Q(X_s)\,ds.$$

These sums and integrals are bounded by

$$\sum_{0<s\leqslant T} |h(X_{s-}, X_s)| \quad \text{and} \quad \int_0^T \sum_y Q(X_s, y)|h(X_s, y)|\,ds;$$

by assumption, these bounds are square integrable. Thus, Var and the limit as $t \to \infty$ can be interchanged to conclude the proof. □

3.4. *Variance comparison: Infinite horizon.* Consider next the normalized functionals

$$H_t = t^{-1} \sum_{0<s\leqslant t} h(X_{s-}, X_s), \quad \text{and} \quad H_t^Q = t^{-1} \int_0^t h_Q(X_s)\,ds.$$

We compare their asymptotic variances, assuming $\{H_t,\ t \geqslant 0\}$ and $\{H_t^Q,\ t \geqslant 0\}$ are consistent estimators of the same quantity. Suppose, then, that there are constants σ,

$\sigma_Q > 0$ and α for which

$$t^{1/2}[H_t - \alpha] \Rightarrow \sigma\mathcal{N}(0,1) \tag{24}$$

and

$$t^{1/2}[H_t^Q - \alpha] \Rightarrow \sigma_Q\mathcal{N}(0,1), \tag{25}$$

where $\mathcal{N}(0, 1)$ denotes a standard normal random variable and $\Rightarrow$ is convergence in distribution. In other words, suppose H and H^Q satisfy central limit theorems with the same means and respective variance constants σ^2 and σ_Q^2. As will become clear, this particular normalization and form of the limiting distribution are not essential to the results; however, (24)—(25) represent the canonical case.

Our variance inequalities are preserved as $t \to \infty$ if we assume

$$\left\{t\left[(H_t - \alpha)^2 + (H_t^Q - \alpha)^2\right], t > 0\right\} \text{ is uniformly integrable.} \tag{26}$$

Combining these assumptions we obtain

PROPOSITION 3.5. *Suppose the conditions of either Theorem* 3.1 *or Theorem* 3.2 *hold and that* (24)—(26) *are in effect. Then* $\sigma_Q \leqslant \sigma$.

PROOF. Assumptions (24)—(26) imply

$$\sigma^2 = \lim_{t\to\infty} t \operatorname{Var}[H_t]$$

and an analogous connection between σ_Q^2 and $\operatorname{Var}[H_t^Q]$. Under the conditions of either Theorem 3.1 or Theorem 3.2, $\operatorname{Var}[H_t^Q] \leqslant \operatorname{Var}[H_t]$. □

The ergodic case merits special attention. Let $\{X_t, t \geqslant 0\}$ be ergodic with invariant distribution π. Standard regenerative theory for Markov processes gives

$$\lim_{t\to\infty} t^{-1}\int_0^t h_Q(X_s)\,ds = \sum_x \pi(x)h_Q(x) \equiv \alpha \tag{27}$$

for π-integrable h_Q. Also (see, e.g., Çinlar (1975, pp. 269–271)),

$$\lim_{t\to\infty} t^{-1} \sum_{0<s\leqslant t} h(X_{s-}, X_s) = \sum_{x,y} \pi(x)Q(x,y)h(x,y), \tag{28}$$

which is α by the definition of h_Q. Our general result for the ergodic case is this:

COROLLARY 3.6. *Suppose that X is ergodic and let T_x be the passage time to a fixed state x. Assume hypotheses* (a)—(b) *of either Theorem* 3.1 *or Theorem* 3.2 *hold and*

$$\mathbf{E}_x\left[\left(\sum_{0<s\leqslant T_x} |h(X_{s-}, X_s)|\right)^2\right] < \infty$$

and

$$\mathbf{E}_x\left[\left(\int_0^{T_x}\sum_y Q(X_s, y)|h(X_s, y)|\,ds\right)^2\right] < \infty$$

where $\mathbf{E}_x$ is expectation with $X_0 = x$. Then (24)—(26) *hold and* $\sigma_Q \leqslant \sigma$.

PROOF. Under the square-integrability conditions, regenerative theory (as in Asmussen (1987, §V.3 and VI.4)) gives (24)—(26). Proposition 3.5 now gives $\sigma_Q \leqslant \sigma$. □

3.5. *Discrete-time conversion.* Let $Y = \{Y_n,\ n \geqslant 0\}$ be the discrete-time Markov chain embedded at the jumps of X. Conditioning either a standard or filtered estimator on Y guarantees variance reduction, essentially without restrictions on X. This is *discrete-time conversion* (see Hordijk, Iglehart, and Schassberger (1976), and Fox and Glynn (1986, 1990)), an implementation of conditional Monte Carlo for Markov processes.

Let $\{\tau_n,\ n > 0\}$ be the jump epochs of X, $\tau_0 \equiv 0$ and $Y_n = X_{\tau_n}$, $n \geqslant 0$. Chain Y has transition matrix

$$P_Y(x, y) = Q(x, y)q^{-1}(x)\mathbf{1}\{y \neq x\}, \qquad x, y \in \mathbf{S},$$

where $q(x) = -Q(x, x)$. If T is a passage time for X, then $T = \tau_K$, a.s., for some Y-stopping time K. In this setting, much as in Fox and Glynn (1990, p. 1461), we have

$$\text{(29)} \qquad \mathbf{E}\left[\sum_{0<s\leqslant T} h(X_{s-}, X_s)\middle| Y\right] = \sum_{n=1}^{K} h(Y_{n-1}, Y_n) = \sum_{0<s\leqslant T} h(X_{s-}, X_s),$$

and

$$\text{(30)} \qquad \mathbf{E}\left[\int_0^T h_Q(X_s)\,ds\middle| Y\right] = \sum_{n=1}^{K} h_Q(Y_{n-1})q^{-1}(Y_{n-1}).$$

Thus, in (29) there is no change, but in (30) conditioning *strictly* reduces variance by integrating out holding times.

With a deterministic time horizon t, conditioning on Y is difficult: the number of jumps in $[0, t]$ is unknown, given Y. In this case Fox and Glynn (1990) uniformize, increasing variance but often reducing computational effort. Let $\{N_t,\ t \geqslant 0\}$ be a Poisson process with rate λ and let $\{Y_n^\lambda,\ n \geqslant 0\}$ be a Markov chain with transition matrix $I + Q/\lambda$ (I is the identity matrix). Then $\{X_t,\ t \geqslant 0\}$ and $\{Y_{N_t}^\lambda,\ t \geqslant 0\}$ are equal in law. Directly from Fox and Glynn (1990, equation (14)) we get

$$\mathbf{E}\left[\int_0^t h_Q(X_s)\,ds\middle| Y^\lambda\right] = t\sum_{k=0}^{\infty}\sum_{n=0}^{k} h_Q(Y_n^\lambda)e^{-\lambda t}\frac{(\lambda t)^k}{(k+1)!}.$$

The same argument shows

$$\text{(31)} \qquad \mathbf{E}\left[\sum_{0<s\leqslant t} h(X_{s-}, X_s)\middle| Y^\lambda\right] = \sum_{k=1}^{\infty}\sum_{n=0}^{k-1} h(Y_n^\lambda, Y_{n+1}^\lambda)e^{-\lambda t}\frac{(\lambda t)^k}{k!}.$$

Rather than truncate infinite sums, Fox and Glynn (1990) propose further conditioning—reducing work at the expense of increased variance. Their results give

$$\mathbf{E}\left[\int_0^t h_Q(X_s)\,ds\middle| Y^\lambda, N_t\right] = \lambda^{-1}\sum_{n=0}^{N_t} h_Q(Y_n^\lambda)\frac{t}{N_t + 1}.$$

Further conditioning (31) on N_t reproduces the original estimator $\sum h(X_{s-}, X_s)$.

Returning to (30), we see that under the conditions of Theorem 3.4, we have

$$\mathrm{Var}\left[\sum_{n=1}^{K} h_Q(Y_{n-1})q^{-1}(Y_{n-1})\right] \leqslant \mathrm{Var}\left[\sum_{0<s\leqslant T} h(X_{s-}, X_s)\right].$$

This can be derived directly from a discrete-time analog of Theorem 3.4. Note that

$$\begin{aligned} h_Q(x)q^{-1}(x) &= \sum_y Q(x,y)h(x,y)q^{-1}(x) \\ &= \sum_y P_Y(x,y)h(x,y) \\ &\equiv h_P(x); \end{aligned}$$

and

$$\sum_{n=1}^{K} \mathbf{E}\left[h(Y_{n-1}, Y_n)|\mathscr{F}_{\tau_{n-1}}\right] = \sum_{n=1}^{K} h_P(Y_{n-1}), \quad \text{a.s.},$$

is a (discrete-time) filtered Monte Carlo estimator. Using this notation we get

THEOREM 3.7. *Suppose the discrete-time analogs of* (15) *and* (16) *hold and either* (i) *h satisfies condition* (a) *of Theorem* 3.1 *and* P_Y *is monotone and associated*; *or* (ii) *h satisfies condition* (a) *of Theorem* 3.2 *and* P_Y *is convex and associated. Then*

$$\mathrm{Var}\left[\sum_{n=1}^{K} h_P(Y_{n-1})\right] \leqslant \mathrm{Var}\left[\sum_{n=1}^{K} h(Y_{n-1}, Y_n)\right],$$

and the corresponding expectations are equal.

The proof is essentially the same as that of Theorem 3.4. In the special case $h(x,y) = f(y)$, with $P_Y f(x) = \sum_y P_Y(x,y)f(y)$ we get

$$\mathrm{Var}\left[\sum_{n=1}^{K} P_Y f(Y_{n-1})\right] \leqslant \mathrm{Var}\left[\sum_{n=1}^{K} f(Y_n)\right],$$

when f is increasing and P is monotone and associated. This is proved by Glynn and Iglehart (1988, Theorem 12) for a totally ordered state space (on which all measures and transition kernels are associated).

4. Point-process integrators. We now turn to variance comparisons complementing (6). Three subsections consider, respectively, a general setting, the Markov case, and Poisson arrivals.

4.1. *A general setting.* Throughout this section, $\{N_t,\ t \geqslant 0\}$ denotes a point process with stochastic intensity $\{\lambda_t,\ t \geqslant 0\}$ based on a filtration $\{\mathscr{F}_t,\ t \geqslant 0\}$ rich enough to include the history of N; see Brémaud (1981) for background. Take N to be right-continuous. Let $\{Z_t,\ t \geqslant 0\}$ be an adapted, right-continuous process with state space $\mathbf{S}$, a countable subset of $\mathbf{R}^d$. Assume the sample paths of Z are step functions, almost surely. By the definition of stochastic intensity (Brémaud (1981, p. 27)),

$$\mathbf{E}\left[\int_0^t f(Z_{s-})\,dN_s\right] = \mathbf{E}\left[\int_0^t f(Z_s)\lambda_s\,ds\right]$$

for all bounded f: $\mathbf{S} \to \mathbf{R}$ and $0 < t < \infty$. Tailoring Theorem 2.1 gives us a corresponding variance comparison:

THEOREM 4.1. *With the notation above, suppose that*
(a) *f is bounded, nonnegative, and increasing*;
(b) $\{\lambda_t,\ t \geqslant 0\}$ *is bounded and left-continuous with right limits*; *and*
(c) *for all* $0 \leqslant u_1 < u_2 \leqslant s_1 < s_2$, $\{Z_{u_1}, N_{u_2} - N_{u_1}, Z_{s_1}, N_{s_2} - N_{s_1}\}$ *are associated given* $\mathscr{F}_{u_1}$, *a.s.*
Then

$$\operatorname{Var}\left[\int_0^t f(Z_s)\lambda_s\,ds\right] \leqslant \operatorname{Var}\left[\int_0^t f(Z_{s-})\,dN_s\right]. \tag{32}$$

PROOF. Riemann-Stieltjes sums for $\int f(Z_{s-})\,dN_s$ converge almost surely because $\{f(Z_{s-}),\ 0 \leqslant s \leqslant t\}$ and $\{N_s,\ 0 \leqslant s \leqslant t\}$ are step functions and one has discontinuities only from the right, the other only from the left. Since $\{f(Z_s),\ 0 \leqslant s \leqslant t\}$ is right-continuous with left limits and $\{\lambda_s,\ 0 \leqslant s \leqslant t\}$ is left-continuous with right limits, each has countably many discontinuities. The same is therefore true of $\{f(Z_s)\lambda_s,\ 0 \leqslant s \leqslant t\}$, which is then Riemann integrable, almost surely. Following Theorem 2.1, we need to verify L^2 convergence of the corresponding Riemann-Stieltjes and Riemann sums for the two integrals.

Let λ^* be a (deterministic) bound on the intensity: $\lambda_s \leqslant \lambda^*$ for all s, a.s. Then for any partition $\{0 = s_0, s_1, \dots, s_n = t\}$ of $[0, t]$, we have

$$\sum_{i=0}^{n-1} f(Z_{s_i})\lambda_{s_i}[s_{i+1} - s_i] \leqslant \|f\|\lambda^* \cdot t.$$

Also,

$$\sum_{i=0}^{n-1} f(Z_{s_i})\left[N_{s_{i+1}} - N_{s_i}\right] \leqslant \|f\| \cdot N_t,$$

and N_t is stochastically bounded by a Poisson random variable with mean $\lambda^* t$. Thus, in both cases the approximating sums have moments of all orders, bounded uniformly in the partition of $[0, t]$. Convergence in L^2 follows.

We next verify an analog of (20); namely,

$$\lim_{n\to\infty} \mathbf{E}\left[\left(\sum_{i=0}^{n-1} \mathbf{E}\left[f(Z_{s_i})(N_{s_{i+1}} - N_{s_i}) \middle| \mathscr{F}_{s_i}\right] - \int_0^t f(Z_s)\lambda_s\,ds\right)^2\right] = 0, \tag{33}$$

for the even partitions of $[0, t]$ defined by $s_i = s_i^n = it/n$, $i = 0, 1, \dots, n$. Adaptedness of Z and N, Brémaud (1981, II.3.4), and boundedness of f and $\{\lambda_s,\ s \geqslant 0\}$ give

$$\begin{aligned}\mathbf{E}\left[f(Z_{s_i})(N_{s_{i+1}} - N_{s_i}) \middle| \mathscr{F}_{s_i}\right] &= f(Z_{s_i})\mathbf{E}\left[\int_{s_i}^{s_{i+1}} \lambda_s\,ds \middle| \mathscr{F}_{s_i}\right] \\ &= f(Z_{s_i})\lambda_{s_i}[s_{i+1} - s_i] + o(s_{i+1} - s_i),\end{aligned}$$

where

$$|o(s_{i+1} - s_i)| \leqslant (\lambda^*)^2 \|f\| \cdot t^2/n^2.$$

This proves (33) by the argument used for (20).

Hypothesis (c) and monotonicity and nonnegativity of f give

$$\mathbf{E}\big[f(Z_{s_i})(N_{s_{i+1}} - N_{s_i})\big|\mathscr{F}_{s_i}\big]\mathbf{E}\big[f(Z_{s_j})(N_{s_{j+1}} - N_{s_j})\big|\mathscr{F}_{s_i}\big]$$

$$\leqslant \mathbf{E}\big[f(Z_{s_i})(N_{s_{i+1}} - N_{s_i})f(Z_{s_j})(N_{s_{j+1}} - N_{s_j})\big|\mathscr{F}_{s_i}\big], \quad \text{a.s.,}$$

for all $0 \leqslant i < j < n$. Putting the pieces together just as in Theorem 2.1 now proves (32). □

Clearly, the most substantive hypothesis of Theorem 4.1 is (c); hence, our attention now turns to conditions that make Z and the increments of N associated. We begin with a fairly general result, adapting a framework used in Kamae, Krengel, and O'Brien (1977) for monotonicity, in Lindqvist (1988) for association, and in Glasserman (1992) for association of increments.

Suppose the process $\{(Z_s, N_s),\ 0 \leqslant s \leqslant t\}$ is governed by a family of transition kernels in the following sense: for any $n = 1, 2, \ldots$ and any $0 \leqslant t_0 < t_1 < \cdots < t_n \leqslant t$ there is a transition kernel $K_{t_0 \cdots t_n}$ such that

$$K_{t_0 \cdots t_n}(z_0, k_0, \ldots, z_{n-1}, k_{n-1}, B)$$

$$= P\big((Z_{t_n}, N_{t_n}) \in B | Z_{t_i} = z_i, N_{t_i} = k_i, i = 0, \ldots, n-1\big),$$

for all $B \subseteq \mathbf{S} \times \{0, 1, \ldots\}$, all $z_0, \ldots, z_{n-1} \in \mathbf{S}$ and all $k_0, \ldots, k_{n-1} \in \{0, 1, \ldots\}$. For each $n = 1, 2, \ldots,$ give $\mathbf{S}^n \times \{0, 1, \ldots\}^n$ the product order; in other words, $(z_1, \ldots, z_n, k_1, \ldots, k_n) \leqslant (z_1', \ldots, z_n', k_1', \ldots, k_n')$ iff $z_i \leqslant z_i'$ and $k_i \leqslant k_i'$ for all $i = 1, \ldots, n$. Then $K_{t_0 \cdots t_n}$ is monotone if $K_{t_0 \cdots t_n}(\cdot, C)$ is an increasing function for all upper sets C, and convex if $K_{t_0 \cdots t_n}(\cdot, \cdot + C)$ is increasing. The kernel is termed associated if $K_{t_0 \cdots t_n}(z_0, \ldots, z_{n-1}, k_0, \ldots, k_{n-1}, \cdot\,)$ is an associated measure for all z_i, k_i, $i = 0, \ldots, n-1$. Theorem 4 of Kamae, Krengel, and O'Brien (1977) shows that monotone kernels generate stochastically monotone processes; Corollary 6.4 of Lindqvist (1988) shows that monotone associated kernels generate associated processes; and Theorem 4.1 of Glasserman (1992) shows that associated convex kernels generate processes with associated increments. We need an intermediate result: roughly speaking, we want (Z, N) to be associated in the first d coordinates and to have associated increments in the last coordinate.

To state the result, we use the following notation: For any $k = 0, 1, 2 \ldots$ and any $C \subseteq \mathbf{S} \times \{0, 1, 2 \ldots\}$,

$$k \oplus C = \{(z, n) \in \mathbf{S} \times \{0, 1, 2 \ldots\} : (z, n - k) \in C\}.$$

We now have

Proposition 4.2. *Suppose $\{(Z_s, N_s),\ 0 \leqslant s \leqslant t\}$ is generated by a family of associated kernels. Suppose that for all n, all $0 \leqslant t_0 < \cdots < t_n \leqslant t$, and all upper sets $C \subseteq \mathbf{S} \times \{0, 1, \ldots\}$,*

$$(34) \qquad K_{t_0 \cdots t_n}(z_0, \ldots, z_{n-1}, k_0, \ldots, k_{n-1}, k_{n-1} \oplus C)$$

is an increasing function of $z_0, \ldots, z_{n-1}, k_0, \ldots, k_{n-1}$. *Then for all n and all* $0 \leqslant u_1 < \cdots < u_n \leqslant t$,

$$\{Z_{u_1}, N_{u_2} - N_{u_1}, \ldots, Z_{u_n}, N_t - N_{u_n}\}$$

are associated whenever (Z_0, N_0) *is associated.*

The proof is a straightforward adaptation of arguments in Lindqvist (1988) (along the lines of Glasserman (1992)), so we omit it. However, one consequence is worth noting: A particular type of upper set is $\{(z, n): z \geqslant z', n \geqslant k'\}$ for some (z', k'). In this case, the main hypothesis (34) of Proposition 4.2 states that

$$P\left(Z_{t_n} \geqslant z', N_{t_n} - N_{t_{n-1}} \geqslant k' | Z_{t_i} = z_i, N_{t_i} = k_i, i = 0, \ldots, n-1\right)$$

is increasing in z_i, k_i, $i = 0, \ldots, n-1$, for all z', k'. In other words, Z and the increments of N are stochastically increasing in past values of Z and N.

It is difficult to use Proposition 4.2 without some simplification in the joint distribution of $\{(Z_s, N_s), 0 \leqslant s \leqslant t\}$. Markov dependence leads to manageable results, as we show next.

4.2. *The Markov case.* Taking (Z, N) to be Markov puts us back in the setting of §3: let $f(Z, N) = f(Z)$ and $g(Z, N) = N$ and use Corollary 3.3. However, integrating with respect to a counting process is a sufficiently interesting special case to merit separate treatment, and the approach here differs from that of §3. Also, the results here sometimes impose less stringent requirements on examples, as we will see in §5.

Let (Z, N) have generator Q. Without technical complications, we may allow N to be a *compound* point process—i.e., N may have jumps of size k for any $k = 1, 2, \ldots$. In state (z, n), the intensity of an N-jump of size k is $Q((z, n), \{(z', n + k), z' \in \mathbf{S}\})$. For any subset B of $\mathbf{S} \times \{0, 1, \ldots\}$ and any $k = 0, 1, 2 \ldots$, define

$$B_k = \{z \in \mathbf{S}: (z, k) \in B\}.$$

With this notation, if C is an upper set then (i) so is every C_k, $k = 0, 1, 2 \ldots$, and (ii) $C_k \subseteq C_{k'}$ whenever $k \leqslant k'$. We now have

LEMMA 4.3. *Suppose that*
(a) *Q is bounded and up-down*;
(b) *for any upper set* $C \subseteq \mathbf{S} \times \{0, 1, \ldots\}$ *and any* $(x, m) \leqslant (y, n)$,

$$Q((x, m), m \oplus C) \leqslant Q((y, n), n \oplus C)$$

whenever $x \in C_0$ *or* $y \notin C_0$; *and*
(c) (Z_0, N_0) *is associated.*
Then, for any n and any $0 \leqslant t_0 < t_1 < \cdots < t_n$,

$$\{Z_{t_0}, N_{t_1} - N_{t_0}, \ldots, Z_{t_{n-1}}, N_{t_n} - N_{t_{n-1}}\}$$

are associated.

SKETCH OF PROOF. Let λ^* be a bound on Q and define the transition kernel $\Lambda = Q/\lambda^* + I$. With our conditions on Q, Λ is monotone and associated and satisfies

(34). Moreover, for any $t > 0$,

$$K_t((z,k),B) \equiv P((Z_{t+s}, N_{t+s}) \in B | Z_s = z, N_s = k)$$
$$= \sum_{n=0}^{\infty} e^{-\lambda^* t} \frac{(\lambda^* t)^n}{n!} \Lambda^n((z,k),B).$$

The argument of Glasserman (1992, Theorem 4.7) now applies: K_t satisfies (34) because Λ does. The result thus follows from Proposition 4.2. □

For any $B \subseteq \mathbf{S} \times \{0, 1, 2, \dots\}$, we have

$$B = \bigcup_{k=0}^{\infty} B_k \times \{k\}.$$

Suppose, now, that Q admits the following decomposition:

$$Q((z,m),B) = Q_0(z, B_m) + \sum_{k=1}^{\infty} \mu_k(z,m) P_k(z, B_{m+k}), \tag{35}$$

where Q_0 is a generator, the μ_k's are nonnegative functions, and the P_k's are transition kernels. Think of $\mu_k(z,m)$ as the intensity of an N-jump of size k; of $P_k(z, \cdot)$ as the conditional distribution of Z following an N-jump of size k starting from state z; and of Q_0 as governing the motion of Z in the absence of N-jumps. We now have

COROLLARY 4.4. *Suppose Q is bounded and admits the decomposition* (35). *Suppose also that*

(a) *Q_0 is monotone and up-down*;
(b) *every P_k is monotone and upward*;
(c) *every μ_k is increasing*; *and*
(d) *$\lambda(\cdot,\cdot) \equiv \sum_k k\mu_k(\cdot,\cdot)$ is bounded.*

Then Q satisfies the (a)—(b) *of Lemma* 4.3, *and*

$$\mathrm{Var}\left[\int_0^t f(Z_s)\lambda(Z_s, N_s)\,ds\right] \leqslant \mathrm{Var}\left[\int_0^t f(Z_{s-})\,dN_s\right]$$

whenever f is bounded, nonnegative, and increasing. The expectations of these integrals are equal.

PROOF. Q is up-down because Q_0 is up-down and because at every jump of N, Z jumps up (since the P_k's are upward). To see that Q satisfies (b) of Lemma 4.3, let C be an upper set and take $(x,m) \leqslant (y,n)$ with either $x \in C_0$ or $y \notin C_0$. Then

$$Q((x,m), m \oplus C) = Q_0(x, C_0) + \sum_{k=1}^{\infty} \mu_k(x,m) P_k(x, C_k)$$
$$\leqslant Q_0(y, C_0) + \sum_{k=1}^{\infty} \mu_k(y,n) P_k(y, C_k)$$
$$= Q((y,n), n \oplus C).$$

The inequality follows from the monotonicity of Q_0, P_k and μ_k, and the fact that C_k, $k = 0, 1, 2 \dots$ are upper sets. Given Lemma 4.3 and condition (d), the result follows by the argument used in Theorem 4.1, with $\mathcal{F}_s = \sigma(\{Z_u, N_u, 0 \leqslant u \leqslant s\})$. □

The extension to unbounded f is made by adding square-integrability conditions analogous to (15)—(16) and using the truncation argument of Theorem 3.1.

4.3. *Poisson arrivals.* Straightforward modification of the arguments in §§3.3 and 3.4 yields random-horizon and infinite-horizon analogs of the results above. When the horizon is infinite, one case has particular interest: Poisson integrators. If conditions for PASTA (Poisson arrivals see time averages) are met (Wolff (1982)), averages with respect to dN_t and dt converge to the same limit but may have different asymptotic variances. We compare the variance constants when the arrival rate is known and used in the arrival average. This is in the spirit of indirect estimation via Little's law, as investigated in Glynn and Whitt (1989).

By a time average we mean, e.g.,

$$A_t \equiv t^{-1} \int_0^t f(Z_s)\, ds.$$

The corresponding arrival average is

$$A_t^N \equiv N_t^{-1} \int_0^t f(Z_{s-})\, dN_s.$$

When the arrival rate λ is known, one also considers

$$A_t^\lambda \equiv (\lambda t)^{-1} \int_0^t f(Z_{s-})\, dN_s.$$

We compare the first and last of these.

Assume that for some ν and $\sigma_A, \sigma_\lambda > 0$ we have, for all (Z_0, N_0),

(36) $$t^{1/2}[A_t - \nu] \Rightarrow \sigma_A \mathcal{N}(0,1)$$

and

(37) $$t^{1/2}\left[A_t^\lambda - \nu\right] \Rightarrow \sigma_\lambda \mathcal{N}(0,1).$$

Conditions for limits like these are developed in Glynn and Whitt (1989) and Glynn, Melamed, and Whitt (1993). Assume also that

(38) $$\left\{t\left[(A_t - \nu)^2 + \left(A_t^\lambda - \nu\right)^2\right], t > 0\right\} \text{ is uniformly integrable.}$$

Next, consider the case of Markov (Z, N) with (bounded) generator Q admitting the following decomposition:

(39) $$Q((z,m), B) = Q_0(z, B_m) + \lambda P_1(z, B_{m+1}).$$

This gives us

THEOREM 4.5. *With the notation above, suppose that*
(a) Q_0 *is monotone and up-down*; P_1 *is monotone and upward*;
(b) (36)—(38) *hold*; *and*
(c) f *is nonnegative and increasing*.

Then $\sigma_A \leqslant \sigma_\lambda$; *i.e., the time average has smaller asymptotic variance than the arrival average with known arrival rate.*

PROOF. From Corollary 4.4 we get

$$\mathrm{Var}\left[\int_0^t f(Z_s)\lambda\, ds\right] \leqslant \mathrm{Var}\left[\int_0^t f(Z_{s-})\, dN_s\right];$$

hence,

$$\mathrm{Var}\left[t^{-1}\int_0^t f(Z_s)\, ds\right] \leqslant \mathrm{Var}\left[(\lambda t)^{-1}\int_0^t f(Z_{s-})\, dN_s\right].$$

Given uniform integrability, the inequality extends to σ_A and σ_λ by letting $t \to \infty$. □

REMARK. Condition (a) is reasonable. The first part states that in the absence of N-jumps the process evolves as an associated, monotone Markov process. The second part states that the N-jumps are indeed "arrivals": they force the process upward, monotonically.

A comparison with A_t^N is not as readily available. Since this estimator has a stochastic denominator this comparison requires a bivariate central limit theorem; specifically, suppose that

$$t^{1/2}\left(A_t^\lambda - \nu, t^{-1}N_t - \lambda\right) \Rightarrow \mathscr{N}(0,\Sigma), \tag{40}$$

with

$$\Sigma = \begin{pmatrix} \sigma_\lambda^2 & \sigma_{\lambda N} \\ \sigma_{\lambda N} & \sigma_{NN}^2 \end{pmatrix}.$$

Since N is Poisson, $\sigma_{NN}^2 = \lambda$. A standard ratio-estimator argument now shows that

$$t^{1/2}\left[A_t^N - \nu\right] \Rightarrow \sigma_N \mathscr{N}(0,1),$$

with

$$\sigma_N^2 = \sigma_\lambda^2 + \lambda^{-1}\nu^2 - 2\lambda^{-1}\nu\sigma_{\lambda N}. \tag{41}$$

In passing from a pure arrival average to a time average (i.e., from A^N to A) we replace dN_s with $\lambda\, ds$ in both numerator and denominator. This gives us automatic variance reduction in the denominator: $\mathrm{Var}[\lambda t] = 0 < \mathrm{Var}[N_t]$. Using Corollary 4.4, we can impose conditions for variance reduction in the numerator as well—conditions for $\sigma_A^2 \leqslant \sigma_\lambda^2$. However, these same conditions make positive the covariance between the numerator and denominator in A^N and force $\sigma_{\lambda N} \geqslant 0$. Since this covariance term is *subtracted* to obtain the asymptotic variance σ_N^2, the overall change in variance is unknown *a priori*.

Glynn and Whitt (1989) discuss related issues in comparing direct and indirect estimators based on Little's law. Glynn, Melamed, and Whitt (1993) analyze estimators of customer and time averages but do not order the corresponding variances.

5. Examples. We now develop in more specific settings some of the ideas from previous sections.

5(a). *Extent of variance reduction*. Filtering an estimator can reduce or magnify variance by arbitrarily large amounts. Let $\{N_t, t \geqslant 0\}$ be a Poisson process with rate λ; then $\mathrm{Var}[N_t] > 0$. Filtering, we pass from

$$N_t = \int_0^t dN_s \quad \text{to} \quad \int_0^t \lambda\, ds = \lambda t,$$

and $\mathrm{Var}[\lambda t] = 0$—infinite variance reduction. On the other hand, if τ_1 is the epoch of the first jump of N, then $\int_0^{\tau_1} dN_s = 1$ has zero variance while the variance of $\lambda\tau_1$ is strictly positive.

Comparing time averages and Poisson-arrival averages, Glynn, Melamed, and Whitt (1993) discuss a related example. Let N be Poisson with unit rate; let $\{Y_n, n \geqslant 0\}$ be a Markov chain; and define $X_t = Y_{N_t}$, $t \geqslant 0$. If τ_n is the epoch of the nth jump of N, then

$$\mathrm{Var}\left[\int_0^{\tau_n} f(X_{s-})\, dN_s\right] \leqslant \mathrm{Var}\left[\int_0^{\tau_n} f(X_s)\, ds\right]$$

because

$$\mathbf{E}\left[\int_0^{\tau_n} f(X_s)\, ds \middle| Y\right] = \sum_{i=0}^{n-1} f(Y_i)$$

$$= \int_0^{\tau_n} f(X_{s-})\, dN_s$$

and conditional expectations reduce variance. On the other hand, if Y is upward and monotone then (X, N) has a convex, upward generator; so, if f is increasing and nonnegative,

$$\mathrm{Var}\left[\int_0^t f(X_s)\, ds\right] \leqslant \mathrm{Var}\left[\int_0^t f(X_{s-})\, dN_s\right]$$

for all fixed $t > 0$, by Corollary 4.4. As discussed in §3.5, when the horizon is fixed these estimators are not related through a conditional expectation.

The cases above are not necessarily representative. In practice, we believe filtering typically reduces variance and that the degree of variance reduction is typically moderate.

5(b). *Birth-death processes*. With suitable parameters, birth-death processes exhibit many of the conditions considered in this paper. We illustrate with a few different cases.

A birth-death process is automatically monotone. To make it convex, we need to extend the state space to $\{\ldots, -1, 0, 1, \ldots\}$ from the more usual $\{0, 1, \ldots\}$. Writing

$$Q(x, x + B) = \lambda_x \mathbf{1}\{1 \in B\} + \mu_x \mathbf{1}\{-1 \in B\} - (\lambda_x + \mu_x)\mathbf{1}\{0 \in B\},$$

indicates that Q is convex if λ_x increases and μ_x decreases in x. It is automatically up-down because the state space is totally ordered. Boundedness of the birth and death rates makes Q bounded.

If we let $h(x, y) = \mathbf{1}\{y > x\}$, then

$$\mathbf{E}\left[\sum_{0<s\leqslant t} h(X_{s-}, X_s)\right] = \mathbf{E}\left[\int_0^t h_Q(X_s)\, ds\right]$$

is the expected number of births in $[0, t]$. By Theorem 3.2, the smaller variance belongs to

$$\int_0^t h_Q(X_s)\, ds = \int_0^t \lambda_{X_s}\, ds.$$

By Theorem 3.4, the same is true if t is replaced with T_0, the passage time to $\{\dots, -1, 0\}$ (a lower set), provided $\mathbf{E}[T_0^2] < \infty$. This holds, for example, if the birth rates are bounded from above and μ_x is bounded away from zero for $x \geqslant 1$. Notice that if $X_0 > 0$, then the law of $\{X_t, 0 \leqslant t \leqslant T_0\}$ is unchanged by our enlargement of the state space.

In modeling a queue with (possibly) state-dependent arrival and service rates, we can use Theorem 3.4 with horizon T_0, as above; for a deterministic horizon, we need results from §4. Consider the bivariate Markov process (X, N) in which X is the queue length, evolving on $\{0, 1, \dots\}$, and N counts arrivals—i.e., up-jumps of X. Let $\tilde{Q}$ be the corresponding generator. Decompose $\tilde{Q}$ as in (39) to get

$$\tilde{Q}((x, n), C) = Q_0(x, C_n) + \lambda_x P_1(x, C_{n+1}).$$

The motion of X in the absence of arrivals is a pure-death process; hence, Q_0 is monotone. By assumption λ_x increases in x. Also, P_1 is upward and monotone because $P_1(x, \cdot\,)$ is the unit mass at $x + 1$. The conditions on $\tilde{Q}$ needed in Corollary 4.4 are thus satisfied.

Let $f(x) = x$ so $f(X_t)$ is the queue length at time t. Then $\int_0^t f(X_{s-})\, dN_s$ is the queue length found by arrivals in $[0, t]$. With the same expectation, $\int_0^t f(X_s)\lambda_{X_s}\, ds$ has smaller variance. As a special case, consider a stable $M/M/1$ queue: $\lambda_x \equiv \lambda$ and $\mu_x \equiv \mu > \lambda$. Theorem 4.5 applies, so $t^{-1}\int_0^t f(X_s)\, ds$ has smaller asymptotic variance than $(\lambda t)^{-1}\int_0^t f(X_{s-})\, dN_s$. By explicit calculation of variance constants, Glynn, Melamed, and Whitt (1993) make a related comparison: the time-averaged workload has smaller asymptotic variance than the average workload found by arrivals.

5(c). *Tandem queues*. Consider d queues in tandem fed by a Poisson arrival process of rate λ. Jobs bring independent, unit-mean, exponential service requirements to each node. Server i works at rate $\nu_i(n)$ when there are n jobs in its queue, $\nu_i(0) = 0$. Let X record the vector of queue lengths and let N count arrivals.

We apply Corollary 4.4 to (X, N). Between jumps of N only service completions can occur; thus, the Q_0 for (39) is given by

$$Q_0(x, B) = \sum_{i=1}^{d-1} \mathbf{1}\{x - e_i + e_{i+1} \in B\}\nu_i(x_i)$$
$$+ \mathbf{1}\{x - e_d \in B\}\nu_d(x_d) - \mathbf{1}\{x \in B\}(\nu_1(x_1) + \cdots + \nu_d(x_d)),$$

where the e_i's are the standard unit vectors and $x = (x_1, \dots, x_d)$ is a vector of nonnegative integers. Apply the following partial order to $\mathbf{Z}^d$: $(x_1, \dots, x_d) \precsim (y_1, \dots, y_d)$ if and only if $x_1 + \cdots + x_i \precsim y_1 + \cdots + y_i$ for all $i = 1, \dots, d$. Under $\precsim$, X is an up-down process; in particular, every arrival triggers a jump up and every service completion triggers a jump down. From our representation of $Q_0(x, B)$ we see that Q_0 is monotone if each $\nu_i(\cdot)$ is decreasing. At each jump of N, X increases by e_1, so the conditions of Corollary 4.4 are in effect.

5(d). *Branching processes.* Let $\{\xi_{ni},\ n = 1,2,\dots, i = 1,2,\dots\}$ be an array of iid, nonnegative, integer-valued random variables. Let $Y_0 = 1$, a.s., and let

$$Y_{n+1} = Y_n + \sum_{i=1}^{Y_n} \xi_{ni}, \qquad n = 0,1,2\dots.$$

Also, let N be a rate-λ Poisson process and define $X_t = Y_{N_t}$, $t \geqslant 0$; then X is a continuous-time branching process. If every ξ_{ni} has mass function ϕ, then the generator of X satisfies $Q(j,\{n \geqslant j+k\}) = \lambda\sum_{n=0}^{\infty}\phi^{*j}(k+n)$, for any j and k, where $*j$ denotes j-fold convolution. Since this is increasing in j for all k, Q is convex. In addition, Q is upward and bounded so it meets the conditions of Corollary 3.3.

Applying Q to an increasing convex function g gives

$$Qg(j) = \lambda \sum_{k=0}^{\infty} \phi^{*j}(k) g(j+k).$$

Suppose $\mathbf{E}[(g(X_t))^2] < \infty$ and

$$\mathbf{E}\left[\left(\int_0^t \sum_y Q(X_s, y)|g(y)|\,ds\right)^2\right] < \infty.$$

Then

$$\mathrm{Var}\left[\int_0^t Qg(X_s)\,ds\right] \leqslant \mathrm{Var}[g(X_t)]$$

and both sides have the same expectation.

5(e). *Martingale integration-by-parts.* This topic is not an example of results proved in previous sections but an illustration of how related ideas can be used in other settings. Let $\{M_t,\ t \geqslant 0\}$ be a martingale with respect to $\{\mathscr{F}_t,\ t \geqslant 0\}$, and $\{A_t,\ t \geqslant 0\}$ an adapted, increasing process. Under various additional hypotheses, one has

$$\mathbf{E}[M_T A_T] = \mathbf{E}\left[\int_0^T M_s\,dA_s\right] \tag{42}$$

for stopping times T. Specific instances of this integration-by-parts formula appear, for example, in Brémaud (1981, p. 302), Jacod and Shiryaev (1987, p. 29), and Protter (1990, p. 89). Since $M_s = \mathbf{E}[M_T|\mathscr{F}_s]$, the integral on the right side of (42) is a filtered Monte Carlo estimator; now it is the integrand, rather than the integrator, that is "filtered".

Paralleling the proof of Theorem 2.1, we could establish a variance inequality for (42) with a dependence condition: M_T and the increments of A should be (conditionally) positively correlated. Suppose, for example, that

$$A_t = \int_0^t f(X_s)\,ds,$$

where X is Markov and f is increasing and nonnegative. Then a variance inequality would follow if we required that M_T and X be associated.

In many examples we obtain a Markov process X associated with a rate-λ Poisson process N. It is well known that for any $\mu > 0$, the likelihood ratio

$$L_t = \left(\frac{\mu}{\lambda}\right)^{N_t} e^{-(\mu-\lambda)t}, \qquad t \geqslant 0,$$

is a martingale. If $\mu > \lambda$ then L_t is an increasing function of N_t. Thus, conditions and arguments like those of §3 give

$$\text{Var}\left[\int_0^T L_s f(X_s)\,ds\right] \leqslant \text{Var}\left[L_T \int_0^T f(X_s)\,ds\right] \tag{43}$$

and equality of the expectations. Integrals of this type arise in simulation through *importance sampling*. Comparisons like (43) are studied in Glasserman (1993).

6. Concluding remarks. Our results order the variances of objects with a common expectation. In general, given two unbiased estimators ξ_1 and ξ_2, say, an optimal convex combination attains lower variance than either estimator separately. If the ξ_i's have variances σ_i^2, $i = 1, 2$, and covariance σ_{12}, the variance of $\alpha\xi_1 + (1-\alpha)\xi_2$ is minimized at $\alpha = (\sigma_2^2 - \sigma_{12})/(\sigma_1^2 + \sigma_2^2 - 2\sigma_{12})$, as a simple calculation shows.

A different perspective on combining estimators is available when the original (unfiltered) estimator is a *semimartingale*, i.e., the sum of its initial condition, a process of bounded variation, and a local martingale: generically, $Z_t = Z_0 + B_t + M_t$. Consider, for example, a Markov process X with generator Q. A process $\{g(X_t), t \geqslant 0\}$ admits the semimartingale representation

$$g(X_t) = g(X_0) + \int_0^t Qg(X_s)\,ds + M_t \tag{44}$$

with

$$M_t = g(X_t) - g(X_0) - \int_0^t Qg(X_s)\,ds$$

a mean-zero martingale, subject to integrability conditions. The alternative estimators suggested by (5) thus correspond to including or not including the martingale part of a semimartingale. The same is true for (3)—(4) and (6). But no bias is introduced if M_t is replaced with αM_t for any constant α; M is a readily available *control*, in the simulation sense. Generically, the variance of $Z_0 + B_t + \alpha M_t$ is minimized at $\alpha = -\text{Cov}[Z_0 + B_t, M_t]/\text{Var}[M_t]$. The cases we have compared correspond to $\alpha = 0, 1$. A further examination of compensator identities for variance reduction should therefore investigate the optimal use of martingale control variates.

Acknowledgement. Thanks go to a referee for a detailed reading of the manuscript and valuable comments and corrections.

References

Anantharam, V. (1991). The Stability Region of the Finite-User Slotted ALOHA Protocol, *IEEE Trans. Inform. Theory*. **37** 535–540.

Asmussen, S. (1987). *Applied Probability and Queues*, Wiley, New York.

Bratley, P., Fox, B. L. and Schrage, L. (1987). *A Guide to Simulation*. 2nd. ed., Springer, New York.

Brémaud, P. (1981). *Point Processes and Queues*, Springer, New York.

Chung, K. L. (1974). *A Course in Probability Theory*. 2nd. ed., Academic Press, New York.

Çinlar, E. (1975). *Introduction to Stochastic Processes*. Prentice-Hall, Englewood Cliffs, NJ.

Cox, T. J. (1984). An Alternate Proof of a Correlation Inequality of Harris. *Ann. Probab.* **12** 272–273.

Daley, D. (1968). Stochastically Monotone Markov Chains, *Z. Wahrsch. verw. Geb.* **10** 305–317.

Esary, J. D., Proschan, F. and Walkup, D. W. (1967). Association of Random Variables. *Ann. of Math. Stat.* **38** 1466–1474.

Fortuin, C. M., Kasteleyn, P. W. and Ginibre, J. (1971). Correlation Inequalities on Some Partially Ordered Spaces. *Comm. Math. Phys.* **22** 89–103.

Fox, B. L. and Glynn, P. W. (1986). Discrete-Time Conversion for Simulating Semi-Markov Processes. *Oper. Res. Lett.* **5** 191–196.

______ and ______. (1990). Discrete-Time Conversion for Simulating Finite-Horizon Markov Processes. *SIAM J. Appl. Math.* **50** 1457–1473.

Glasserman, P. (1992). Processes with Associated Increments. *J. Appl. Probab.* **29** 313–333.

______. (1993). Stochastic Monotonicity and Conditional Monte Carlo for Likelihood Ratios. *Adv. in Appl. Probab.* (to appear).

Glynn, P. W. and Iglehart, D. L. (1988). Simulation Methods for Queues: An Overview. *Queueing Systems Theory Appl.* **3** 221–256.

______, Melamed, B. and Whitt, W. (1993). Estimating Customer Averages and Time Averages. *Oper. Res.* (to appear).

______ and Whitt, W. (1987). Sufficient Conditions for Functional-Limit-Theorem Versions of $L = \lambda W$. *Queueing Systems Theory Appl.* **1** 279–287.

______ and ______. (1989). Indirect Estimation via $L = \lambda W$. *Oper. Res.* **37** 82–103.

Harris, T. E. (1977). A Correlation Inequality for Markov Processes on Partially Ordered Spaces. *Ann. Probab.* **5** 451–454.

Hordijk, A., Iglehart, D. L. and Schassberger, R. (1976). Discrete Time Methods for Simulating Continuous Time Markov Chains. *Adv. in Appl. Probab.* **8** 772–788.

Jacod, J. and Shiryaev, A. N. (1987). *Limit Theorems for Stochastic Processes*. Springer-Verlag, Berlin, Heidelberg, New York.

Kamae, T., Krengel, U. and O'Brien, G. L. (1977). Stochastic Inequalities on Partially Ordered Spaces. *Ann. Probab.* **5** 899–912.

Karr, A. F. (1986). *Point Processes and Their Statistical Inference*. Marcel Dekker, Inc., New York and Basel.

Keilson, J. and Kester, A. (1977). Monotone Matrices and Monotone Markov Processes. *Stochastic Process Appl.* **5** 231–241.

Lindqvist, B. H. (1988). Association of Probability Measures on Partially Ordered Sets. *J. Multivariate Anal.* **26** 111–132.

Massey, W. A. (1987). Stochastic Orderings for Markov Processes on Partially Ordered Spaces. *Math. Oper. Res.* **12** 350–367.

Meester, L. E. and Shanthikumar, J. G. (1990). Stochastic Convexity on General Space. University of California, Berkeley.

Protter, P. (1990). *Stochastic Integration and Differential Equations*. Springer, New York.

Shaked, M. and Shanthikumar, J. G. (1988). Temporal Stochastic Convexity and Concavity. *Stochastic Process. Appl.* **27** 1–20.

______ and ______. (1990). Parametric Stochastic Convexity and Concavity of Stochastic Processes. *Ann. Inst. Statist. Math.* **42** 509–531.

Sharpe, M. (1988). *General Theory of Markov Processes*. Academic Press, New York.

Wolff, R. (1982). Poisson Arrivals See Time Averages. *Oper. Res.* **30** 223–231.

P. Glasserman: 403 Uris Hall, Graduate School of Business, Columbia University, New York, New York 10027; e-mail: pglasser@research.gsb.columbia.edu.