Conditional Monte Carlo Gradient Estimation in Economic Design of Control Limits[1]

Michael C. Fu, **Shreevardhan Lele**

Robert H. Smith School of Business
University of Maryland
College Park, MD 20742-1815

Thomas W.M. Vossen

Leeds School of Business
University of Colorado
Boulder, CO 80309-0419

July 2001; revised July 2002, December 2003

Abstract

The economic approach to determining the optimal control limits of control charts requires estimating the derivative of the expected cost function. Simulation is a very general methodology used for estimating the expected costs in the economic design of control charts, but for the gradient estimation problem, finite difference estimators can be inefficient. We demonstrate an alternative approach based on smoothed perturbation analysis (SPA), also known as conditional Monte Carlo. Numerical results and consequent design insights are obtained in determining the optimal control limits for EWMA and Bayes charts. The results indicate that the SPA gradient estimators can be significantly more efficient than finite difference estimators, and that a simulation approach using these estimators provides a viable alternative to other numerical solution techniques for the economic design problem.

Keywords: gradient estimation, control charts, economic design, Monte Carlo simulation, smoothed perturbation analysis

[1]M.C. Fu was supported in part by the National Science Foundation under Grants DMI-9713720, DMI-9988867, and DMI-0323220. S. Lele was supported in part by a grant from the Graduate Research Board of the University of Maryland, College Park. We thank the referee, Associate Editor, and Department Editor for their many suggestions and comments which have led to a much improved paper.

1 Introduction

The economic approach to the design of quality control charts distinguishes itself from the purely statistical approach in that it explicitly addresses the process failure mechanism and accounts for the various costs involved in the continuous operation of the chart (e.g., see survey papers by Montgomery 1980, Svoboda 1991, and Ho and Case 1994). The optimal chart parameters (e.g., the control limit, the sampling interval, the sample size) minimize the expected cost function. In the earliest models, the cost function could be computed analytically, and thus the optimization problem solved either analytically or numerically through some simple iterative (search) procedure. However, for charts of practical interest, even under standard process assumptions, optimization over the space of chart parameters is rarely analytically tractable. In particular, for important classes of charts that are more sensitive than the $\bar{X}$-chart, such as the exponentially weighted moving average (EWMA) chart and the Bayes chart, analytical expressions for the expected cost function are not available. In such settings, Monte Carlo simulation is frequently used (e.g., Albin et al. 1997 and Grimshaw and Alt 1997), with Barish and Hauser (1963) being a very early example of where simulation is used in economic design of control charts.

Our focus on the Bayes and EWMA charts arises from the characterization of the former as yielding optimal control policies and the latter as being a simple extension of the widely used $\bar{X}$-chart that significantly improves its sensitivity to small shifts in the process mean. Girshick and Rubin (1952), Bather (1963) and Taylor (1965, 1967) provided the early results on the optimality of the Bayesian technique. More recent developments are presented in Tagaras (1994, 1996, 1998), Tagaras and Nikolaidis (2002), Calabrese (1995) and Porteus and Angelus (1997). The EWMA chart was proposed by Roberts (1959) and is explained at length in Hunter (1986) and Montgomery (1996). Al-Sultan and Rahim (1997), and the papers therein, present various issues in applying optimization techniques to problems in quality control.

Under fairly standard assumptions, the dynamics of a very broad class of control charts can be modeled as an ergodic Markov process (e.g., Pollock and Alden 1992, Lele 1996). Once the steady-state distribution is computed, the expected cost per unit time can be found for a given

control limit, and then the optimal control limit determined. For the case of the $\bar{X}$-chart under a normal sampling distribution, the steady-state distribution can be found analytically, because the Markov process reduces to a simple Markov chain. However, for EWMA and Bayes charts, the steady-state distributions require numerical solution techniques, e.g., discretization of the resulting continuous state space or simulation. For design purposes, it is desirable to obtain an entire curve of optimal control limits as a function of cost ratios. If the gradient of the steady-state distribution with respect to the control limits are available, then this curve can be generated in an efficient manner that avoids enumeration and search. Lele (1996) used simulation-based finite differences to estimate these gradients for EWMA and Bayes charts, but found them to be quite noisy.

In this paper, we apply the technique of smoothed perturbation analysis (SPA) or conditional Monte Carlo, introduced by Gong and Ho (1987), to estimate the desired gradients. This technique was first applied to statistical quality control by Fu and Hu (1997, 1999). They only considered simple average run length (ARL) performance measures, and did not address any optimal design problem, nor did they use an economic cost model. Furthermore, they did not consider Bayes control charts. This paper significantly extends the work in Fu and Hu (1999) by showing that the solution methodology of SPA can be fruitfully employed to obtain economically optimal control limits for a broad class of control charts. The approach offers an alternative to discretizing the state space and numerically solving the resulting Markov chain. We compare the two approaches on some numerical examples, which indicates that the SPA/simulation design approach is quite competitive, its relative performance depending on the level of discretization chosen in solving the Markov chain versus the statistical precision controlled by the number of simulation replications. We note that one advantage of the simulation approach is the availability of a measure of precision through standard error estimates. Another advantage of a simulation-based approach is that essentially a single code can be used for a number of various models, e.g., $\bar{X}$, EWMA, and Bayes, with just a few changes in the program. It is for this reason that Albin et al. (1997) chose simulation in their comparison of a large number of control charts.

The rest of the paper is organized as follows. In Section 2, we provide the basic control chart

model and state the economic design problem. In Section 3, we apply SPA to the economic design problem and derive two distinct gradient estimators. Section 4 contains the computational results, which includes comparisons both with finite difference estimators in terms of simulation efficiency, and with the numerical solution of a discretized Markov chain for the control chart design problem. Summary and conclusions are given in Section 5.

2 Problem Setting

2.1 Process Model

The standard process assumptions in the economic design literature (e.g., see the textbook by Montgomery 1996), which we adopt in this paper as well, are (a) a single known out-of-control state, and (b) a memoryless failure time distribution (i.e., either exponential or geometric, depending on whether a continuous-time or discrete-time setting is employed). It has been shown by Banerjee and Rahim (1988), McWilliams (1989) and Collani (1997) that the optimal design is relatively insensitive to the exact distribution of the failure time as long as the expected failure time is efficiently estimated. Ho and Case (1994) discuss the robustness of the above assumptions. Krishnamoorthi (1985) provides a thorough application of the economic design approach, including a discussion of the estimation of the necessary inputs.

Consider a production process whose state at time i ($i = 1, 2, \ldots$) is described by $\mu_i \in \{0, \delta\}$. When $\mu_i = 0$ the production process is in control, and when $\mu_i = \delta$ it is out of control. The process shifts from the in-control to the out-of-control state at a random failure time $T \in \{1, 2, \ldots\}$, assumed to follow a geometric distribution with parameter λ, so that $E[T] = 1/\lambda$. The quantities λ and δ are the process parameters of the model, and can be estimated from past data.

2.2 Control Chart Model

The process state μ_i is only partially observable. At time i, a sample measurement X_i is drawn from the output of the process, with sampling density/distribution either $f^{(0)}/F^{(0)}$ or $f^{(\delta)}/F^{(\delta)}$, depending on whether the process is in control or out of control, respectively.

At each time i, the process can either be declared to be in control, in which case it is allowed to

continue production, or it can be declared to be out of control, in which case an alarm is sounded, the process is stopped, repaired and reset. In general, the time for repairing and resetting can be any multiple of the basic unit of time, which is the time between measurements. However, it can be shown that without loss of generality, the repair and resetting time can be set to 1. The repair/resetting mechanism is such that the process is probabilistically renewed each time an alarm is sounded.

The decision to sound an alarm at time i is based on the control statistic at time i, denoted Y_i, obtained from the available product quality measurements, $X_1, X_2, \ldots, X_i$. The process is declared to be in control if Y_i is in some interval $[l, u]$; otherwise it is declared to be out of control. l and u are the lower and upper control limits, respectively, whose optimal values are to be determined.

We restrict our attention to control charts having the following three properties:

1. Y_i $(i = 1, 2, \ldots)$ depends on $X_1, X_2, \ldots, X_{i-1}$, only through Y_{i-1}, i.e., Y_i can be completely specified as $\psi(X_i, Y_{i-1})$, where Y_0 is given.

2. ψ is invertible in its first argument, i.e., given Y_{i-1}, there is a 1-1 mapping between the sample observation, X_i, and the control statistic, Y_i.

3. The control limits, l and u, are independent of the sample observations.

These properties are far from restrictive and allow for the study of a very broad family of control charts including EWMA and Bayes charts. For the EWMA chart, the control statistic evolves according to $Y_i = \alpha X_i + (1 - \alpha)Y_{i-1}$, where $\alpha \in (0, 1]$ is a smoothing parameter. The $\bar{X}$-chart is a special case of the EWMA chart when $\alpha = 1$. The control limits of the EWMA chart are customarily symmetric around the centerline $Y_i = 0$, and therefore, only one control limit has to be determined. For the Bayes chart, the control statistic is the posterior probability of being out of control, which for the given process model evolves according to

$$Y_i = \frac{f^{(\delta)}(X_i)[\lambda + (1 - \lambda)Y_{i-1}]}{f^{(\delta)}(X_i)[\lambda + (1 - \lambda)Y_{i-1}] + f^{(0)}(X_i)(1 - \lambda)(1 - Y_{i-1})}.$$

For this chart, the lower control limit is fixed at zero, and u is the sole control limit to be determined.

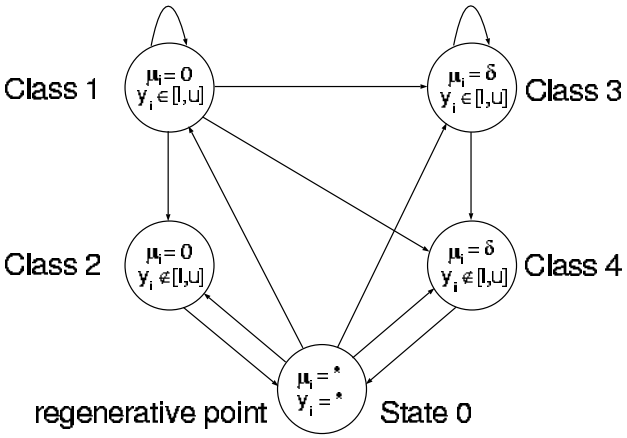

Figure 1: (Class) Transition Diagram for W_t.

When the process is stopped for repair/resetting, let μ_i and Y_i be each assigned the symbol "$*$". Consider the pair (μ_i, Y_i) and an aggregate *class* descriptor W_i taking the following possible values:

$$W_i = \begin{cases} 0 & \text{if } \mu_i = *, Y_i = *; \\ 1 & \text{if } \mu_i = 0, Y_i \in [l,u]; \\ 2 & \text{if } \mu_i = 0, Y_i \notin [l,u]; \\ 3 & \text{if } \mu_i = \delta, Y_i \in [l,u]; \\ 4 & \text{if } \mu_i = \delta, Y_i \notin [l,u]. \end{cases}$$

Class 2 comprises false alarm states, while class 4 comprises true alarm states. Classes 1 and 3 comprise non-alarm states with the process actually in control or out of control, respectively. Since "class" 0 consists of the single repair/reset state, we will usually refer to it as a state. The transitions among the classes of W_i are shown in Figure 1. Note that $\{W_i\}$ is not in general Markov, because transitions generally depend on the actual value of Y_i (the $\bar{X}$-chart is an exception). Aside from the state $W_i = 0$, only the full specification (μ_i, Y_i) constitutes a true Markov state. Under very mild conditions (cf. Pollock and Alden 1992), the pair (μ_i, Y_i) evolves as an aperiodic, irreducible Markov process, so there exists a steady-state distribution for (μ_i, Y_i), and consequently for W_i, as well. Let $\boldsymbol{\pi} = (\pi_0, \pi_1, \pi_2, \pi_3, \pi_4)$ be the steady-state probability mass function corresponding to the five classes of W_i. Define three aggregate probabilities: $p_0 \equiv \pi_1 + \pi_2$, $p_\delta \equiv \pi_3 + \pi_4$ and $p_r \equiv \pi_0$, so that $p_0 = \lim_{i\to\infty} \mathrm{P}(\mu_i = 0)$, $p_\delta = \lim_{i\to\infty} \mathrm{P}(\mu_i = \delta)$ and $p_r = \lim_{i\to\infty} \mathrm{P}(\mu_i = *)$. In other words, p_0, p_δ and p_r are the long-term probabilities of finding the process in control, out of control or in repair, respectively.

2.3 Cost Model

Let k_0 be the "cost" incurred per unit time when the process is in control, k_δ the cost per unit time when the process is out of control, and k_r the cost per unit time when the process is being repaired and reset. (Typically, k_0 is negative, whereas k_δ and k_r are positive.) Then, the steady-state expected cost per unit time, C, is given by

$$C = p_0 k_0 + p_\delta k_\delta + p_r k_r = k_0 + (k_\delta - k_0)p_\delta + (k_r - k_0)p_r. \tag{1}$$

Using this cost function, the economic performance of a control chart is completely characterized by the steady-state probabilities p_δ and p_r.

Let θ denote either of the two control limits, l or u. The design problem is to find the value of θ that minimizes the steady-state expected cost per unit time, C. From Equation (1), the first-order condition at the optimal value of θ is

$$\frac{dC}{d\theta} = (k_\delta - k_0)\frac{dp_\delta}{d\theta} + (k_r - k_0)\frac{dp_r}{d\theta} = 0.$$

Defining the *cost ratio,* $R \equiv (k_r - k_0)/(k_\delta - k_0)$, which is the ratio of the "opportunity cost" of being in repair to the "opportunity cost" of being out of control, we have

$$-\frac{dp_\delta}{dp_r} = -\frac{dp_\delta/d\theta}{dp_r/d\theta} = -\frac{\frac{d\pi_3}{d\theta} + \frac{d\pi_4}{d\theta}}{\frac{d\pi_0}{d\theta}} = R \tag{2}$$

at the optimal value of θ. Thus, determining the optimal value of θ requires calculation of $d\pi_i/d\theta$.

Although not Markov, $\{W_i\}$ is regenerative with repair/reset constituting regenerative points. Following the (class) transition diagram shown in Figure 1, a regenerative cycle with respect to the process $\{W_i\}$ starts with a transition out of state 0 into any of classes 1, 2, 3 or 4 and ends with a transition into state 0, either via class 2 (a false alarm) or via class 4 (a true alarm). Let $\tau \in \{1, 2, \ldots\}$ denote the time at which an alarm is sounded: τ is a stopping time with respect to $\{Y_i\}$, with $Y_i \in [l, u]$ for $i < \tau$, $Y_\tau \notin [l, u]$, and either $W_\tau = 2$ (true alarm) or $W_\tau = 4$ (false alarm). Thus, the length of a regenerative cycle is $\tau + 1$, and $W_{\tau+1} = 0$.

Define S_k, $k = 0, \ldots, 4$, as the number of visits to class k in a regenerative cycle. Then

$S_0 = S_2 + S_4 = 1$ and $\tau = S_1 + S_2 + S_3 + S_4$. The steady-state probabilities are given by

$$\pi_k = \frac{E[S_k]}{E[\tau]+1}, \tag{3}$$

which on differentiation with respect to the parameter θ yields

$$\frac{d\pi_k}{d\theta} = \frac{dE[S_k]}{d\theta}\frac{1}{E[\tau]+1} - E[S_k]\frac{1}{(E[\tau]+1)^2}\frac{dE[\tau]}{d\theta} = \frac{dE[S_k]/d\theta}{E[\tau]+1} - \frac{dE[\tau]/d\theta}{E[\tau]+1}\pi_k. \tag{4}$$

Note that $dE[S_0]/d\theta = 0$, $dE[\tau]/d\theta = \sum_{k=1}^{4} dE[S_k]/d\theta$, and that $\sum_{i=0}^{4} \frac{d\pi_k}{d\theta} = 0$, since $\sum_{i=0}^{4} \pi_k = 1$. Thus, the problem reduces to calculating $dE[S_k]/d\theta, k = 1, 2, 3, 4$.

3 SPA Gradient Estimators

In this section, we derive sample path-based estimators for $dE[S_k]/d\theta$. By definition,

$$S_k = \sum_{i=1}^{\tau} \mathbf{1}\{W_i = k\},\ k = 1, \ldots, 4,$$

where $\mathbf{1}\{\cdot\}$ denotes the indicator function. A sample path will be denoted by $\omega = \{Y_0, T, X_1, \ldots\}$. Note that ω is independent of θ, whereas $\{W_i\}$ clearly depends on θ. Also, Y_i is determined by (measurable with respect to) $\{Y_0, X_1, \ldots, X_i\}$, and W_i depends only on $\{Y_0, T, \ldots Y_i\}$, so with a slight abuse of notation, we may refer to $\{Y_0, T, Y_1, \ldots\}$ as the sample path.

The sample path derivative $dS_k/d\theta$, known as the infinitesimal perturbation analysis (IPA) estimator (see Glasserman 1991), indicates a derivative taken on a fixed sample path. Since S_k is discrete valued, it is piecewise constant for any fixed sample path (e.g., S_1 is a decreasing step function with respect to the upper control limit, and an increasing step function with respect to the lower control limit), and so the sample path derivative $dS_k/d\theta$ is identically zero, hence

$$E\left[\frac{dS_k}{d\theta}\right] \neq \frac{dE[S_k]}{d\theta},\ k \neq 0.$$

As a result, the IPA estimator is incorrect (biased), so we apply an extension of IPA called smoothed perturbation analysis (SPA) (see Gong and Ho 1987), which employs conditional Monte Carlo to "smooth out" the discontinuities in the performance measure. The basic idea is to select a set of quantities on the sample path, say $\mathcal{Z}$, such that

$$E\left[\frac{dE[S_k|\mathcal{Z}]}{d\theta}\right] = \frac{dE[S_k]}{d\theta}$$

and $dE[S_k|\mathcal{Z}]/d\theta$ can be efficiently estimated. Within the framework of Fu and Hu (1997), $dE[S_k|\mathcal{Z}]/d\theta$ is represented as the product of a probability jump rate and an associated jump in the performance measure. In our setting, the jump rate is the (probability) rate at which the number of visits to a certain class in a regenerative cycle jumps with respect to a control limit perturbation, and the jump in the performance measure is the corresponding expected change in the number of visits to that class. For simplicity, we will first assume that the upper control limit (UCL), u, and the lower control limit (LCL), l, can be set independently. The final estimators are then obtained by taking $u = \theta$ and $l = -\theta$ for the EWMA charts, and $u = \theta$ and $l = 0$ for the Bayes chart, as application of the chain rule gives

$$\frac{dE[S_k]}{d\theta} = \frac{dE[S_k]}{du}\frac{du}{d\theta} + \frac{dE[S_k]}{dl}\frac{dl}{d\theta},$$

where in the two cases of interest, $du/d\theta = 1$ and $dl/d\theta = -1$ or 0. For each of $dE[S_k]/du$ and $dE[S_k]/dl$, we derive two distinct estimators (conditioned on a different set of sample paths) called the right-hand (RH) derivative estimator and the left-hand (LH) derivative estimator. Intuitively, the RH derivative with respect to u considers the possibility that an increase in the upper control limit increases the length of a regenerative cycle, thereby increasing the number of visits to a certain class, whereas the LH derivative with respect to u considers the possibility that a decrease in the upper control limit decreases the length of a regenerative cycle, thereby decreasing the number of visits to a given class. We provide detailed derivations of the RH and LH estimators with respect to u. The RH and LH estimators with respect to l are analogous to the LH and RH estimators with respect to u, respectively.

3.1 RH Estimator w.r.t. UCL

We consider the RH UCL estimator, where $\Delta u > 0$. By definition of the alarm time τ, we know $Y_i \leq u < u + \Delta u$ for $i < \tau$ and $Y_\tau > u$. If it is also true that $Y_\tau > u + \Delta u$, then the perturbation has no effect on the sequence $\{W_i\}$ in the regenerative cycle, and therefore $S_k(u + \Delta u) = S_k(u)$, so

$$E[S_k(u + \Delta u) - S_k(u)] = E[(S_k(u + \Delta u) - S_k(u))\mathbf{1}\{u < Y_\tau \leq u + \Delta u\}].$$

In other words, a positive perturbation Δu can affect a regenerative cycle only at the point at which an alarm is sounded because of a breach of the upper control limit, i.e., when $Y_\tau > u$.

Now, conditioning on all the samples taken in a regenerative cycle up to, but not including, the sample that sounds the alarm, i.e., $\mathcal{Z} = \{T, \tau, Y_0, Y_1, Y_2, \dots, Y_{\tau-1}\}$, we write

$$\begin{aligned}
&E[(S_k(u+\Delta u) - S_k(u))\mathbf{1}\{u < Y_\tau \le u+\Delta u\}] \\
&\quad = \quad E[E[(S_k(u+\Delta u) - S_k(u))\mathbf{1}\{u < Y_\tau \le u+\Delta u\} \mid \mathcal{Z}]] \\
&\quad = \quad E[E[(S_k(u+\Delta u) - S_k(u)) \mid \mathcal{Z}, u < Y_\tau \le u+\Delta u] \\
&\qquad\quad \times \mathbf{1}\{Y_\tau > u\} P(u < Y_\tau \le u+\Delta u \mid \mathcal{Z}, Y_\tau > u)].
\end{aligned}$$

Dividing by Δu and taking the limit as $\Delta u \downarrow 0$, we have

$$\begin{aligned}
\frac{dE[S_k]}{du} &= \lim_{\Delta u \downarrow 0} \frac{E[S_k(u+\Delta u) - S_k(u)]}{\Delta u} \\
&= \lim_{\Delta u \downarrow 0} E\Big[E\Big[(S_k(u+\Delta u) - S_k(u)) \mid \mathcal{Z}, u < Y_\tau \le u+\Delta u]\mathbf{1}\{Y_\tau > u\} \\
&\qquad \times P(u < Y_\tau \le u+\Delta u \mid \mathcal{Z}, Y_\tau > u)/\Delta u\Big]\Big] \\
&= E\Big[\lim_{\Delta u \downarrow 0} \frac{P(u < Y_\tau \le u+\Delta u \mid \mathcal{Z}, Y_\tau > u)}{\Delta u}\mathbf{1}\{Y_\tau > u\} \\
&\qquad \times \lim_{\Delta u \downarrow 0} E[S_k(u+\Delta u) - S_k(u) \mid \mathcal{Z}, u < Y_\tau \le u+\Delta u]\Big], \qquad (5)
\end{aligned}$$

assuming the validity of the interchange of limit and expectation, to be formally addressed shortly.

The first limit term is the probability (rate) that an out-of-control alarm is silenced due to a perturbation in the upper control limit, for which explicit calculation gives

$$\begin{aligned}
\lim_{\Delta u \downarrow 0} \frac{P(u < Y_\tau \le u+\Delta u \mid \mathcal{Z}, Y_\tau > u)}{\Delta u} &= \lim_{\Delta u \downarrow 0} \frac{P(\psi^{-1}(u, Y_{\tau-1}) < X_\tau \le \psi^{-1}(u+\Delta u, Y_{\tau-1}) \mid Y_{\tau-1})}{\Delta u \cdot P(X_\tau > \psi^{-1}(u, Y_{\tau-1}) \mid Y_{\tau-1})} \\
&= \frac{f_\tau(\psi^{-1}(u, Y_{\tau-1}))}{1 - F_\tau(\psi^{-1}(u, Y_{\tau-1}))} \frac{d\psi^{-1}(u, Y_{\tau-1})}{du}, \qquad (6)
\end{aligned}$$

where $f_\tau(\cdot)$ and $F_\tau(\cdot)$ are the density function and the cumulative distribution function, respectively, of the alarm-causing sample, X_τ. Recalling that $f^{(0)}/F^{(0)}$ and $f^{(\delta)}/F^{(\delta)}$ are the in-control and out-of-control densities/distributions, respectively, and that T is the failure time, it follows that

$$f_\tau = f^{(0)}\mathbf{1}\{\tau < T\} + f^{(\delta)}\mathbf{1}\{\tau \ge T\}, \quad F_\tau = F^{(0)}\mathbf{1}\{\tau < T\} + F^{(\delta)}\mathbf{1}\{\tau \ge T\}.$$

To summarize, we have the following estimator:

$$\left(\frac{dE[S_k]}{du}\right)_{RH} = \frac{f_\tau(\psi^{-1}(u, Y_{\tau-1}))}{1 - F_\tau(\psi^{-1}(u, Y_{\tau-1}))} \frac{d\psi^{-1}(u, Y_{\tau-1})}{du} \times \mathbf{1}\{Y_\tau > u\} \lim_{\Delta u \downarrow 0} E[S_k(u + \Delta u) - S_k(u) \mid \mathcal{Z}, u < Y_\tau \leq u + \Delta u], \quad (7)$$

The following conditions suffice to establish unbiasedness:

(A1) $\psi^{-1}(\cdot, \cdot)$ is differentiable w.r.t. its first argument.

(A2) $\psi^{-1}(\cdot, \cdot)$ is a decreasing function w.r.t. its second argument, $\left|\frac{d\psi^{-1}(x, \cdot)}{dx}\right| < K_1 \ \forall x$, where $K_1 > 0$ is a constant, and $F^{(0)}(\psi^{-1}(u, l)), F^{(\delta)}(\psi^{-1}(u, l)) < 1$.

(A3) $|f^{(0)}(x)|, |f^{(\delta)}(x)| < K_2 \ \forall x$, where $K_2 > 0$ is a constant.

(A4) $E[\tau(u + \Delta u)] < K_3 \ \ \forall Y_0, \ 0 \leq \Delta u \leq \epsilon$, where $K_3 > 0$ is a constant, for some $\epsilon > 0$.

EWMA control charts and the Bayes control charts that we consider in the next section satisfy the conditions on ψ^{-1} in (A1) and (A2). (A3) holds for the most commonly used distributions such as the normal distribution. (A4) requires that the regenerative cycle be bounded in some small neighborhood of the upper control limit for any initial condition on the control chart statistic.

Proposition 1. Under (A1)–(A4), (7) is an unbiased estimator for $dE[S_k]/du$.

Proof. see Appendix.

Although (7) provides a theoretically unbiased estimator, practical implementation requires a way to estimate the remaining limiting term

$$\lim_{\Delta u \downarrow 0} E[S_k(u + \Delta u) - S_k(u) \mid \mathcal{Z}, u < Y_\tau \leq u + \Delta u]. \quad (8)$$

The natural approach is to estimate $S_k(u+\Delta u)$ and $S_k(u)$ using the same $\mathcal{Z}$ coupled to the original sample path, which we call the *nominal* path, and for which $Y_i \in [l, u]$ (for $i = 1, \ldots, \tau - 1$) and $Y_\tau > u$. Estimation of $S_k(u+\Delta u)$ under condition $Y_\tau \leq u+\Delta u$ means that an out-of-control alarm is not declared at τ in this path. Under the limit $\Delta u \downarrow 0$, this is equivalent to setting $Y_\tau = u_-$ (the "$-$" subscript here denoting infinitesimally below u) and then extending the path appropriately, with a new alarm time of $\tau' > \tau$. We call this path the *perturbed* path (denoted by "PP" in Figure

2, where "NP" denotes the nominal path). Estimation of $S_k(u)$ under conditions $u < Y_\tau$ and the limit $\Delta u \downarrow 0$ is equivalent to setting $Y_\tau = u_+$ (the "+" subscript here denoting infinitesimally above u), which terminates the sample path at the same point as in the nominal path; hence $\{W_i\}$ can be estimated from the nominal path. Thus, (8) represents the expected change in the number of visits to class k, given that the regenerative cycle experiences a state sequence change in the sample path at τ under the limiting condition. Since the paths are coupled before time τ, the difference in S_k for the two paths is also 0 up to then, so (8) can be estimated using a separate sample path starting with $Y_0 = u$, the residual time of T (if any), and W_0 depending on W_τ (through T) in the nominal path. Referring to Figures 1 and 2, there are two cases to consider, depending on how the regenerative state ($W_{\tau+1} = 0$) is entered in the nominal path:

(a) if in the nominal path $W_\tau = 2$, then in the perturbed path $W_\tau = 1$.

Recall that class 2 represents the case where the process quality state is actually in control but the control statistic sounds an alarm; thus, in the perturbed path where no alarm is sounded, $W_\tau = 1$.

(b) if in the nominal path $W_\tau = 4$, then in the perturbed path $W_\tau = 3$.

Since class 4 represents the case where the process quality state is in fact out of control and the control statistic sounds an alarm, in the perturbed path where no alarm is sounded, $W_\tau = 3$.

Since the distribution of the failure time is assumed to be geometric (and hence memoryless), we can ignore dependence on the residual failure time, so the difference term (8) is equal to

$$E[S_k \mid Y_0 = u, W_0 = 1]\mathbf{1}\{W_\tau = 2\} + E[S_k \mid Y_0 = u, W_0 = 3]\mathbf{1}\{W_\tau = 4\} - \mathbf{1}\{W_\tau = k\}, \qquad (9)$$

where the term $\mathbf{1}\{W_\tau = k\}$ accounts for the NP alarm state no longer visited in PP (see Figure 2). Combining (6), (8), and (9) into Equation (7) yields the final estimator:

$$\begin{aligned}\left(\frac{dE[S_k]}{du}\right)_{RH} &= \frac{f_\tau(\psi^{-1}(u, Y_{\tau-1}))}{1 - F_\tau(\psi^{-1}(u, Y_{\tau-1}))} \frac{d\psi^{-1}(u, Y_{\tau-1})}{du} \Big(E[S_k \mid Y_0 = u, W_0 = 1]\mathbf{1}\{W_\tau = 2\} \\ &\quad + E[S_k \mid Y_0 = u, W_0 = 3]\mathbf{1}\{W_\tau = 4\} - \mathbf{1}\{W_\tau = k\} \Big) \mathbf{1}\{Y_\tau > u\}. \qquad (10)\end{aligned}$$

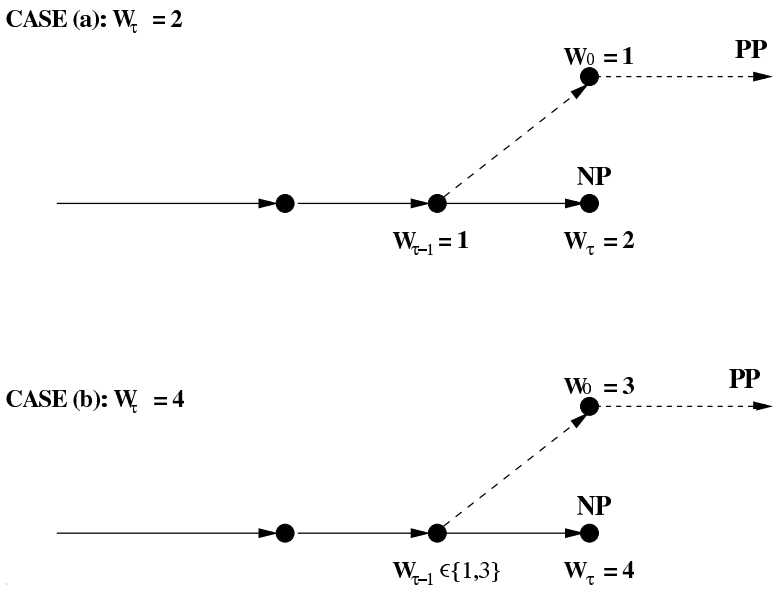

Figure 2: Possible Perturbations for Right-Hand Estimator Term (8).
Note that it is also possible that $W_{\tau-1} = 0$, if and only if $\tau = 1$.

3.2 LH Estimator w.r.t. UCL

The LH UCL estimator is derived considering negative perturbations, i.e., $\Delta u < 0$. In this case, a decrease in the control limit may cause the regenerative cycle to be shortened to any of the earlier samples. Unlike the right hand case, we partition the sample paths before conditioning. Formally, for $\Delta u < 0$, we define the sets $\mathcal{A}_i(\Delta u)$ (which are events in the probability sense) as follows:

$$\mathcal{A}_i(\Delta u) = \{\omega : Y_j \leq u + \Delta u, j = 1, ..., i-1, Y_i > u + \Delta u\},$$

for $i = 1, \ldots, \tau - 1$. In other words, $\mathcal{A}_i(\Delta u)$ consists of all those sample paths in which a perturbation $\Delta u < 0$ causes an alarm to be sounded for the first time at time i, $(1 \leq i < \tau)$ instead of at time τ. This allows us to write

$$\frac{dE[S_k]}{du} = \sum_{1 \leq i < \tau} \lim_{\Delta u \uparrow 0} \frac{E[(S_k(u + \Delta u) - S_k(u))\mathbf{1}(\mathcal{A}_i)]}{\Delta u}. \tag{11}$$

(The $i = \tau$ term provides no contribution since the class visit sequence before and after the perturbation are identical). Each of the terms in the summation is handled separately by selecting a corresponding conditioning set $\mathcal{Z}_i$. For the ith term in Equation (11), we condition on all samples up through the alarm-triggering sample *except* the ith sample, i.e.,

$$\mathcal{Z}_i = \{T, \tau, Y_0, X_1, \ldots, X_\tau\} \backslash \{X_i\}.$$

This allows us to express the left-hand estimator as

$$\left(\frac{dE[S_k]}{du}\right)_{LH} = \sum_{1\le i<\tau} \lim_{\Delta u\uparrow 0} \frac{P(u+\Delta u < Y_i \le u \mid \mathcal{Z}_i, l \le Y_i \le u)}{\Delta u} \times \lim_{\Delta u\uparrow 0} E[S_k(u+\Delta u) - S_k(u) \mid \mathcal{Z}_i, u+\Delta u < Y_i \le u]. \quad (12)$$

As in the right hand case, the first term in the summation equals a probability jump rate, whereas the second term represents the expected change in the number of visits to class k, *given* that the perturbed cycle terminates because of an alarm being sounded by Y_i. A negative perturbation Δu can affect a regenerative cycle at any point prior to the out-of-control declaration at τ. Thus, the probability jump rate term is the probability (rate) that an in-control signal changes to an out-of-control signal (which terminates the regenerative cycle) due to a negative perturbation in the upper control limit. This hazard rate-like term is given by

$$\lim_{\Delta u\uparrow 0} \frac{P(u+\Delta u < Y_i \le u \mid \mathcal{Z}_i, l \le Y_i \le u)}{\Delta u} = \frac{f_i(\psi^{-1}(u, Y_{i-1}))}{F_i(\psi^{-1}(u, Y_{i-1})) - F_i(\psi^{-1}(l, Y_{i-1}))} \frac{d\psi^{-1}(u, Y_{i-1})}{du},$$

so the estimator (12) becomes

$$\left(\frac{dE[S_k]}{du}\right)_{LH} = \sum_{1\le i<\tau} \frac{f_i(\psi^{-1}(u, Y_{i-1}))}{F_i(\psi^{-1}(u, Y_{i-1})) - F_i(\psi^{-1}(l, Y_{i-1}))} \frac{d\psi^{-1}(u, Y_{i-1})}{du} \times \lim_{\Delta u\uparrow 0} E[S_k(u+\Delta u) - S_k(u) \mid \mathcal{Z}_i, u+\Delta u < Y_i \le u]. \quad (13)$$

In order to establish unbiasedness for the LH estimator w.r.t. UCL, conditions (A2) and (A4) need to be altered slightly as follows:

(A2′) $\psi^{-1}(\cdot,\cdot)$ is a decreasing function w.r.t. its second argument, $\left|\frac{d\psi^{-1}(x,\cdot)}{dx}\right| < K_1 \ \forall x$, where $K_1 > 0$ is a constant, and $F_\tau(\psi^{-1}(l,l)) < F_\tau(\psi^{-1}(u,u))$.

(A4′) $E[\tau^2(u+\Delta u)|Y_0 = y] < K_3 \ \ \forall y$, for $0 \ge \Delta u \ge \epsilon$, where $K_3 > 0$ is a constant, for some $\epsilon < 0$.

As an example for (A2′), for the EWMA chart, $\psi^{-1}(y, y) = y$, so the condition on F becomes the easily verifiable $F_\tau(l) < F_\tau(u)$. The need in (A4′) for the bound to be on the second moment is due to the additional summation in the RH estimator, which is not present in the LH estimator. With these modifications, we have the following result, the proof of which is omitted, since it follows exactly the same lines as in the proof of Proposition 1. Analogous results can also be established for the LCL derivative estimators (16) and (17).

Proposition 2. Under (A1),(A2′),(A3), and (A4′), (13) is an unbiased estimator for $dE[S_k]/du$.

As for the RH estimator, implementation for the LH estimator requires an estimation scheme for

$$\lim_{\Delta u \uparrow 0} E[S_k(u + \Delta u) - S_k(u) \mid \mathcal{Z}_i, u + \Delta u < Y_i \leq u]. \tag{14}$$

Parallel to the RH estimator, we estimate this difference using $\mathcal{Z}_i$ coupled to the nominal path. To estimate $S_k(u + \Delta u)$ under condition $Y_\tau > u + \Delta u$ in the limit $\Delta u \uparrow 0$, is equivalent to setting $Y_\tau = u_+$, ending the regenerative cycle ($\tau = i$); this is the perturbed path. To estimate $S_k(u)$ under conditions $Y_i \leq u$ and the limit $\Delta u \uparrow 0$ is equivalent to setting $Y_i = u_-$, which implies that the regenerative cycle continues ($\tau > i$). Note, however, that this continuation differs from the continuation on the nominal path, where $Y_i < u$. Thus, (14) represents the expected change (under the limiting condition) in the number of visits to class k, given that the regenerative cycle terminates early in the sample path at $i < \tau$ due to a negative perturbation of the upper control limit. Since the paths are coupled before time i, the difference in S_k for the two paths is also 0 up to then, so (14) can be estimated using a separate sample path starting with $Y_0 = u$, the residual time of T (if any), and W_0 depending on W_i (through T) in the nominal path.

Referring to Figures 1 and 3, the two cases depend on how the regenerative cycle ends in the perturbed path, and can be viewed as "dual" to the cases for the RH estimator:

(a) $i < T$: the perturbed path terminates in class 2 (false alarm state), so the continuation path starts in class 1 (in control, no alarm state), and the difference is given by

$$\mathbf{1}\{k = 2\} - E[S_k \mid Y_0 = u, W_0 = 1].$$

(b) $i \geq T$: the perturbed path terminates in class 4 (true alarm state), so the continuation path starts in class 3 (out of control, no alarm state), and the difference is given by

$$\mathbf{1}\{k = 4\} - E[S_k \mid Y_0 = u, W_0 = 3].$$

Since the distribution of the failure time is assumed to be geometric (and hence memoryless), we

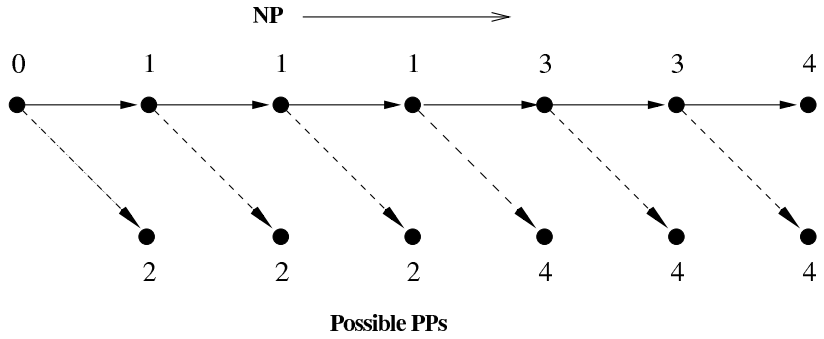

Figure 3: Possible Perturbations for Left-Hand Estimator.

can ignore dependence on the residual failure time, so the difference term (14) is equal to

$$\mathbf{1}\{i < T\}\left(\mathbf{1}\{k=2\} - E[S_k \mid Y_0 = u, W_0 = 1]\right)$$
$$+\mathbf{1}\{i \geq T\}\left(\mathbf{1}\{k=4\} - E[S_k \mid Y_0 = u, W_0 = 3]\right).$$

Substituting this in Equation (12) leads to the final estimator:

$$\begin{aligned}\left(\frac{dE[S_k]}{du}\right)_{LH} &= \sum_{1\leq i<\tau} \frac{f_i(\psi^{-1}(u,Y_{i-1}))}{F_i(\psi^{-1}(u,Y_{i-1})) - F_i(\psi^{-1}(l,Y_{i-1}))}\frac{d\psi^{-1}(u,Y_{i-1})}{du} \\ &\quad \times \Big[\mathbf{1}\{i<T\}\left(\mathbf{1}\{k=2\} - E[S_k \mid Y_0 = u, W_0 = 1]\right) \\ &\quad + \mathbf{1}\{i \geq T\}\left(\mathbf{1}\{k=4\} - E[S_k \mid Y_0 = u, W_0 = 3]\right)\Big]. \qquad (15)\end{aligned}$$

3.3 Estimators w.r.t. LCL

The RH and LH estimators with respect to l can be obtained in a manner analogous to the LH and RH estimators with respect to u, respectively. They are given by:

$$\begin{aligned}\left(\frac{dE[S_k]}{dl}\right)_{LH} &= \frac{f_\tau(\psi^{-1}(l,Y_{\tau-1}))}{F_\tau(\psi^{-1}(l,Y_{\tau-1}))}\frac{d\psi^{-1}(l,Y_{\tau-1})}{dl}\Big(E[S_k \mid Y_0 = l, W_0 = 1]\mathbf{1}\{W_\tau = 2\} \\ &\quad + E[S_k \mid Y_0 = l, W_0 = 3]\mathbf{1}\{W_\tau = 4\} - \mathbf{1}\{W_\tau = k\}\Big)\mathbf{1}\{Y_\tau < l\}, \qquad (16) \\ \left(\frac{dE[S_k]}{dl}\right)_{RH} &= \sum_{1\leq i<\tau} \frac{f_i(\psi^{-1}(l,Y_{i-1}))}{F_i(\psi^{-1}(u,Y_{i-1})) - F_i(\psi^{-1}(l,Y_{i-1}))}\frac{d\psi^{-1}(l,Y_{i-1})}{dl} \\ &\quad \Big[\mathbf{1}\{i<T\}\left(\mathbf{1}\{k=2\} - E[S_k \mid Y_0 = u, W_0 = 1]\right) \\ &\quad + \mathbf{1}\{i \geq T\}\left(\mathbf{1}\{k=4\} - E[S_k \mid Y_0 = u, W_0 = 3]\right)\Big]. \qquad (17)\end{aligned}$$

4 Computational Results

The derivative estimators described in the previous section were implemented for the Bayes chart, as well as EWMA charts of various smoothing parameters. The process parameters were set in turn at $\delta = 1, 2, 3$, and $\lambda = 0.01, 0.05, 0.10$. The sampling distribution of the X_i was assumed to be normal with mean μ_i and unit variance, and we took $Y_0 = 0$ in all cases.

Estimating the derivative dp_δ/dp_r involves both products and ratios of the various $dE[S_k]/d\theta$ and $E[S_k]$. From Equations (2) and (4), we can rewrite the derivative as

$$\frac{dp_\delta}{dp_r} = \frac{(E[\tau]+1)(dE[S_3+S_4]/d\theta) - (dE[\tau]/d\theta)(E[S_3+S_4])}{(E[\tau]+1)(dE[S_0]/d\theta) - (dE[\tau]/d\theta)E[S_0]}. \tag{18}$$

To estimate this quantity, we first obtained two estimates for the numerator of the right-hand side of the above equation from two independent regenerative cycles. Armed with two estimates for each of the numerator and the denominator, we used the jackknife method (see Law and Kelton, 1991) to obtain a less biased estimator for the ratio.

4.1 Optimal Control Limit Curves

Recall that the control statistic for the Bayes chart is the posterior probability of being out of control. The lower control limit of the Bayes chart is fixed at zero, and our goal is to determine the upper control limit, u. This control limit (or the threshold out-of-control probability) was set at 37 different values from 0.05 to 0.9995 as follows (quantities in parentheses represent step sizes): 0.05 (0.05) 0.95 (0.005) 0.995 (0.0005) 0.9995. This range of values was found to be sufficient to observe the limiting behavior of the chart. For each value of u, the derivative $-dp_\delta/dp_r$ was estimated using the RH and LH versions of the SPA estimators of the previous section.

Figure 4 shows some of the SPA(LH) estimator results for $\lambda = 0.05$ and three different values of δ. Instead of plotting the control limits (at which the simulations were run) along the x-axis and the corresponding estimated $-dp_\delta/dp_r$ along the y-axis, we swap the axes so that the control limit is along the y-axis and the corresponding estimated $-dp_\delta/dp_r$ is along the x-axis. Also, we re-label the x-axis as the cost ratio, $R \equiv (k_r - k_0)/(k_\delta - k_0)$. The rationale for this is that (as may be recalled from Section 2.4) $-dp_\delta/dp_r = R$ when the control limit is at its optimal value. With this

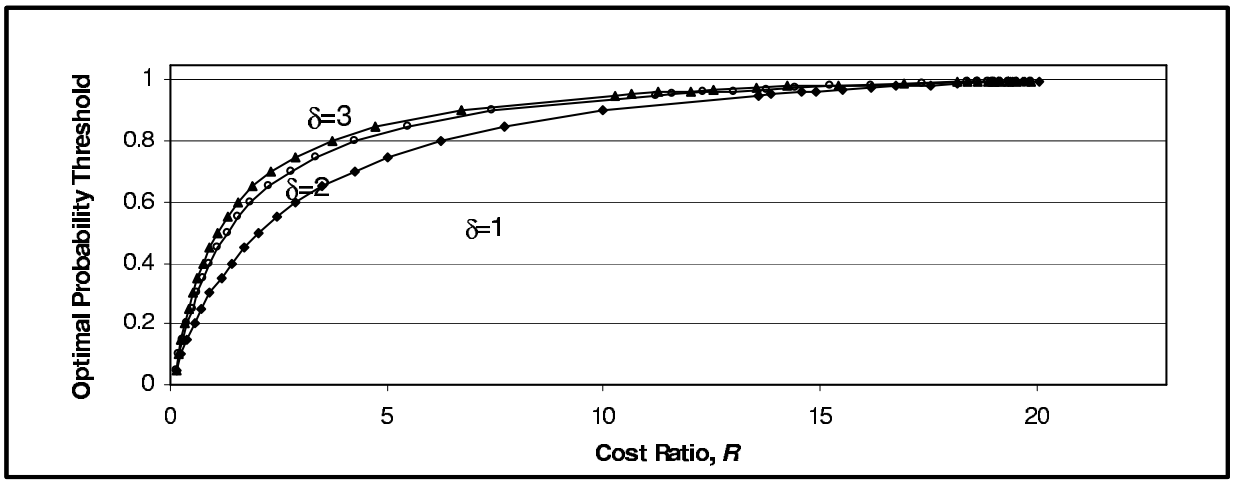

Figure 4: Economically Optimal Control Limit Curves for Bayes Chart using SPA, $\lambda = 0.05$.

orientation of the axes, the chart can be used to read off the optimal control limit corresponding to any given value of the cost ratio, R. For example, suppose we wish to design a Bayes chart that will optimally detect a one standard deviation shift for a process characterized by $\lambda = 0.05$. From Figure 4, we see that if the cost ratio for the process is around 10, the optimal control limit (or the threshold out-of-control probability) should be set at 0.9. On the other hand, if the cost ratio is only half of that (around 5), indicating a lower repair cost relative to the cost of producing product in the out-of-control state, the control limit should be decreased to about 0.75.

As might be expected, the optimal threshold probability increases with the cost ratio for a fixed level of mean shift, δ. Furthermore, we see that from Figure 4 that the optimal threshold probability increases with δ for a given value of the cost ratio. However, the effect on the optimal threshold probability of changing δ is fairly small, especially at higher values of δ (e.g., going from $\delta = 2$ to $\delta = 3$.)

We also see from Figure 4 that as the cost ratio approaches $E[T]$ $(= 1/\lambda = 20)$, the optimal threshold probability approaches 1. This phenomenon, wherein the range of feasible cost ratios for which the there exist non-trivial control limits is limited by the process reliability, is anticipated and explained in Lele (1996).

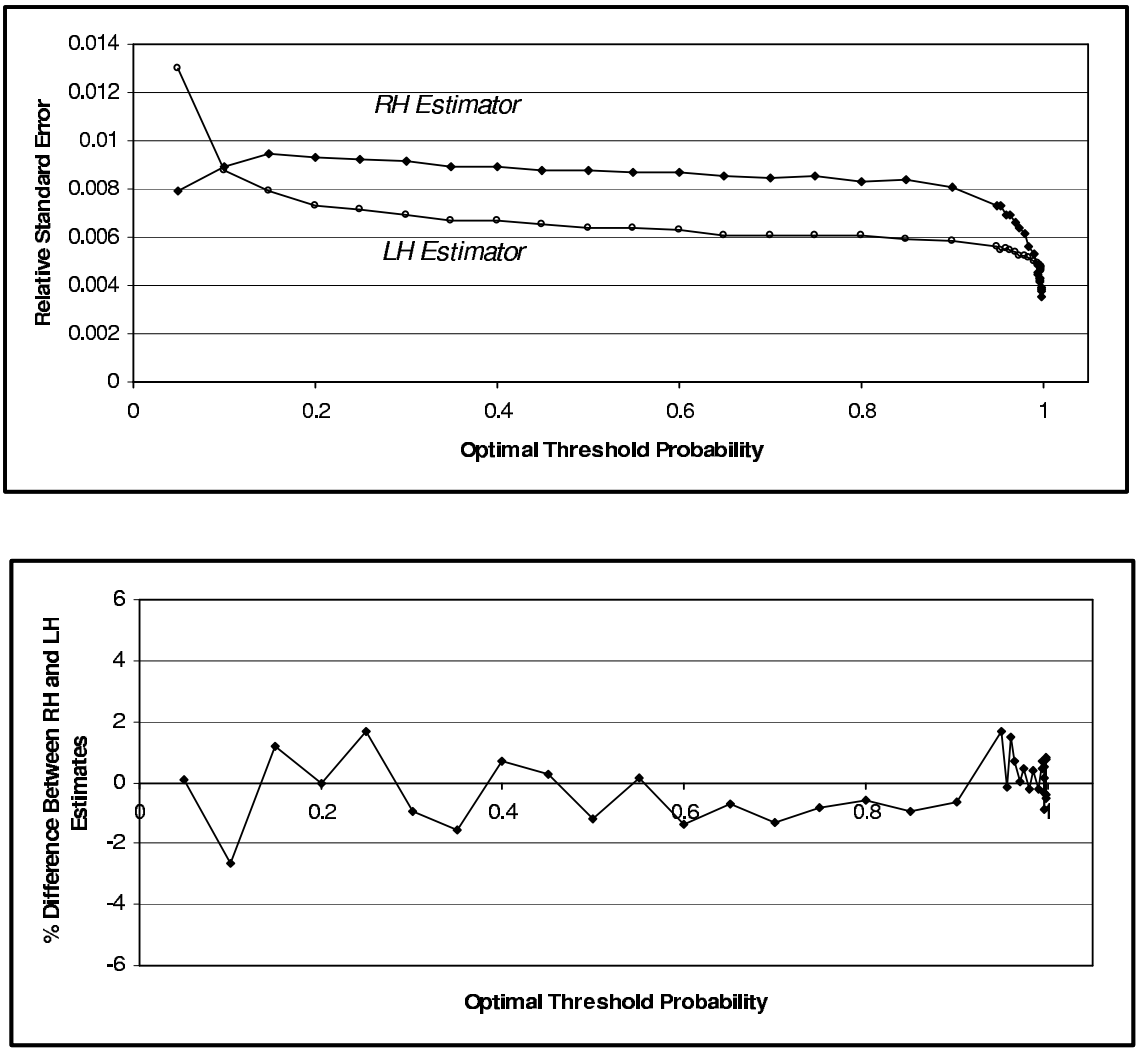

Figure 5: Comparing LH and RH SPA estimators for Bayes Charts, $\lambda = 0.05, \delta = 1$.

In Figure 5, we compare the two versions of the SPA estimator by plotting the percentage difference between the RH and LH estimators — which are generally within $\pm 3\%$ of each other, and comparing the efficiencies of the two versions, finding that the LH estimator generally outperforms the RH estimator in determining the optimal control limit for a Bayes' chart.

The EWMA charts were implemented for various values of the smoothing parameter, α, ranging from 0.2 to 1.0 in steps of 0.2. Although the EWMA chart is a two-sided chart, in practice, the chart is constructed with symmetric limits ($l = -u$) and hence the design problem reduces to determining the optimal upper control limit. This control limit can be expressed in multiples of

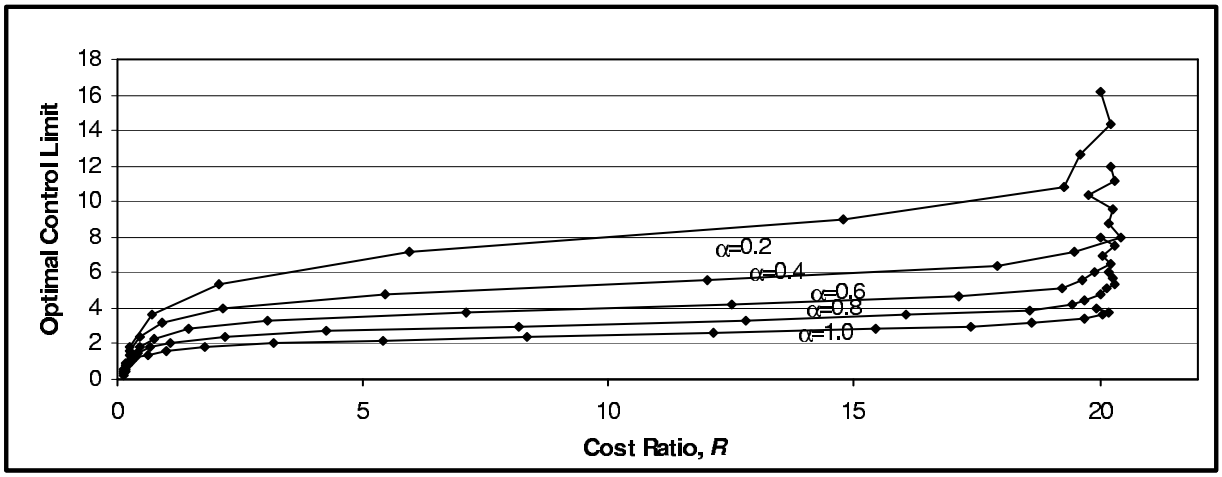

Figure 6: Economically Optimal Control Limit Curves for EWMA Charts using SPA, λ=0.05, δ=1.

the standard deviation of the product quality, X_i, or in multiples of the standard deviation of the control statistic, Y_i. If the variance of X_i is denoted by σ_X^2 (which is assumed, without loss of generality, to be unity in the model presented here), then the variance of the control statistic can be easily seen to approach its asymptotic (in i) value, σ_Y^2, given by $\sigma_Y^2 = \sigma_X^2 \cdot \alpha/(2-\alpha)$. In the EWMA literature, it is customary to state the control limit as a multiple of σ_Y rather than as a multiple of σ_X, and we have followed this convention here.

In Figure 6, we show the optimal control limit as a function of the cost ratio, R, for each of the selected five values of the smoothing parameter. These curves were obtained using the LH version of the SPA estimator. We note that for a given value of α, the control limit is fairly robust with respect to the cost ratio. Furthermore, it is evident that the robustness of the optimal control limit with respect to the cost ratio is stronger at larger values of α (i.e., when it is more like an $\bar{X}$-chart), than at smaller values of α. As in the case of the Bayes chart, the range of cost ratios supported by the EWMA chart is limited by the expected failure time of the underlying production process.

4.2 Comparison with Finite Difference Method

In this section, we compare the efficiency of the SPA estimators with those obtained using a finite difference (FD) method. To increase the efficiency of the estimator, we used the technique of

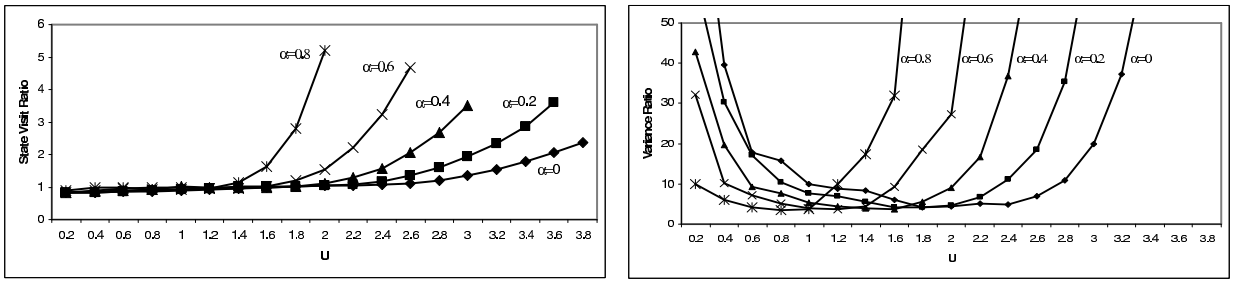

Figure 7: Comparing SPA and FDCRN: EWMA Charts, $\lambda = 0.05, \delta = 2$.

common random numbers (CRN) by synchronizing the simulation between the two runs in each regenerative cycle (i.e., with control limits $u+\Delta u$ and $u-\Delta u$). Specifically, within each regenerative cycle the sample observations used in both runs were identical up through the point where the run with control limit $u - \Delta u$ reaches an alarm state. The resulting estimator is denoted by FDCRN.

As a first step, we compared the relative efficiency of both estimators for a wide range of process parameters. To compare their performance, the number of replications for both estimators was set at $10,000$. In addition, the number of replications for the perturbed path simulations was set at $1,000$. Initially, the increment used in the FDCRN estimators was set at $\Delta u = 0.1u$. Typical results for both EWMA and Bayes charts are shown in Figures 7 and 8. For the EWMA chart, the process parameters used were $\lambda = 0.05$ and $\delta = 2$ and results are shown for different values of the smoothing parameter α. For the Bayes chart, the process parameter used was $\lambda = 0.01$ and results are shown for different values of δ. The first graph in each figure depicts, for each control limit (shown on the x-axis), the state visit ratio, which is the ratio of number of state visits used by the FDCRN estimator to the number used by the SPA estimator. The second graph in each figure depicts the variance ratios, which is the ratio of the sample variance obtained with the FDCRN estimator to the sample variance obtained with the SPA estimator. In both cases, *a ratio greater than 1 indicates less computational effort required for the SPA estimator*, and the product of the two ratios would provide an overall ratio estimating the relative additional computational effort required by the FDCRN estimator to achieve the same level of precision as the SPA estimator.

Since the effectiveness of the FDCRN estimator depends on the increment, we carried out some sensitivity analysis with respect to the increment Δu. Figure 9 shows a typical result, graphs

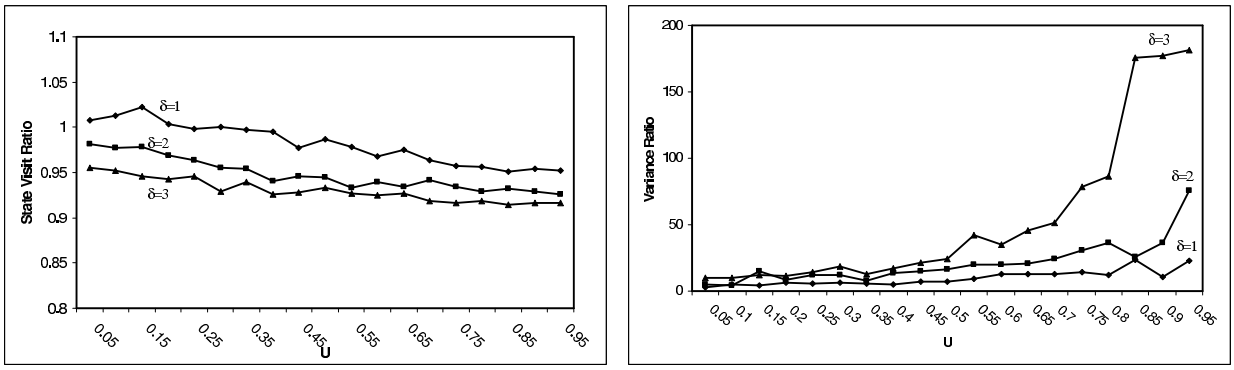

Figure 8: Comparing SPA and FDCRN: Bayes Charts, $\lambda = 0.01$.

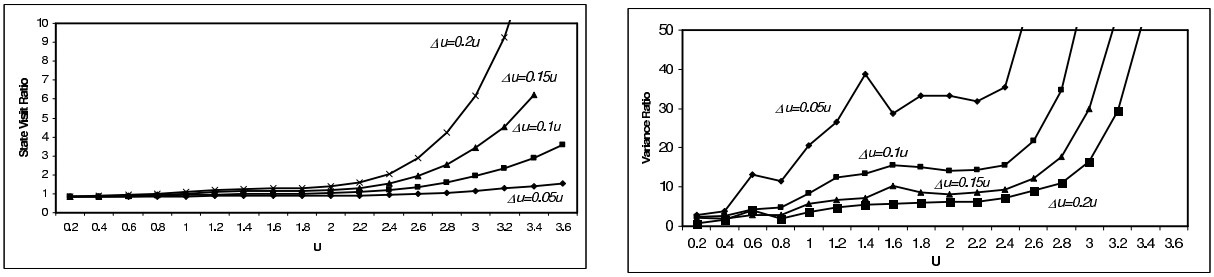

Figure 9: Comparing SPA and FDCRN. Sensitivity Analysis of Difference Increment Δu for EWMA charts, $\lambda = 0.05, \delta = 1, \alpha = 0.2$.

depicting state visit ratios and variance ratios for an EWMA chart with process parameters $\lambda = 0.05$, $\delta = 1$, and $\alpha = 0.2$, for various values of the increment Δu (e.g., $\Delta u = 0.05u, \Delta u = 0.1u, \Delta u = 0.15u$, and $\Delta u = 0.2u$). The graphs indicate a trade-off in increasing the increment: it decreases the variance ratio but increases the state visit ratio. This is reflective of the well-known difficulty in choosing an appropriate increment for simulation-based finite difference estimates, trading off between variance (noise) and bias.

Since the state visit ratios are relatively close to 1 (above 0.9), whereas the variance ratios can be orders of magnitude greater than 1, these results indicate that SPA estimators present a substantial improvement over the use of FD estimators with common random numbers. The number of state visits that would be required to achieve similar levels of precision with FD estimators would be substantially larger than the amount required by SPA estimators. While the precision of the estimators could be improved by changing the increment in the FD estimators, this may come at a substantial cost in terms of computational effort for larger values of the control limit.

4.3 Comparison with Discretized Markov-Chain Models

Recent approaches to the economic design of control charts, such as Tagaras (1994, 1996, 1998), Tagaras and Nikolaidis (2002), and Calabrese (1995), use dynamic programming to determine optimal chart parameters. In our setting, however, dynamic programming is not necessary, as it straightforward to solve a discretized Markov chain to determine the stationary probabilities for a given control limit, so that for a given cost ratio, a simple search could be used to determine the optimal control limit. This approach was taken by Tagaras (1996) for the "static chart" case, where he states (p.44): "The additional complexity and computational requirements of the dynamic programming approach over the static chart can be justified only if the dynamic chart results in substantial quality related cost reductions." We even avoid the search step here, because we are interested in obtaining entire optimal threshold-ratio curves, which depict the optimal control limits over a wide range of cost parameters. Details of the Markov chain model for Bayes charts, as well as the solution procedure used to approximate optimal threshold-ratio curves, are described in more detail in the Appendix. The corresponding model for EWMA charts can be defined in a similar manner, and is not considered here.

To compare this solution approach with the simulation design approach, we first considered the optimal threshold-ratio curves such as those shown in Figure 4. As detailed in the online companion to this paper, over a wide range of parameter values, a Markov chain discretization into $K = 100$ states provides a very good match to the curves obtained using 10,000 simulation replications. However, if we consider the precision of the cost estimates for any given control limit, a discretization bias becomes evident. This is illustrated in Figure 10, which shows the cost estimates under different discretizations of the Markov chain ($K = 10, 20, 50, 100, 1000$) as a function of control limit values compared to simulation (using $1,000$ replications; the results for 10,000 replications are nearly indistinguishable). The graph is representative across a range of the other parameter values (failure probabilities and cost ratios), which indicates that the cost estimates may be substantially different for different degrees of discretization, in this case resulting in a cost difference of approximately 25% between a discretization of 100 states and one of 1000

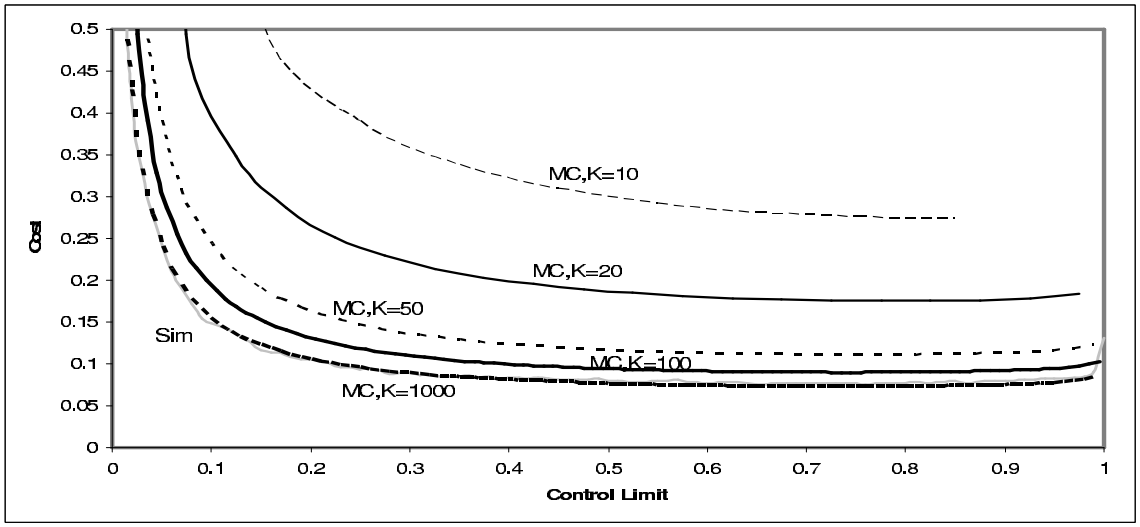

Figure 10: Cost Estimates as a Function of the Control Limit, $\lambda = 0.01, \delta = 2, R = 4$.

states.

We now provide a rough comparison of the computational effort required by the two design methods — SPA-based simulation and numerical solution of the discretized Markov chain model — using two levels of precision for each. The Markov chain calculations were carried out using both CPLEX and MATLAB. The results are shown in Table 1. For the Markov chain solution, the time needed to generate the transition probabilities (*not included in the times reported in the table*) may become non-negligible, averaging approximately 0.3 seconds for a discretization of $K = 100$ and 27 seconds for a discretization of $K = 1000$. Overall, the experiments indicate that the computation times are comparable. For the $K = 100$ discretization, the Markov chain approach takes slightly less overall computation time than simulation (1,000 replications), but for the $K = 1000$ discretization, the required effort increases dramatically (exceeding even 10,000 simulation replications; see Table 1). For both discretizations, the optimal control limits for both cases are quite comparable to the simulation results, but the cost estimates between the two discretizations differ greatly for some parameter values (see Figure 10). Thus, there is a trade-off between discretization bias, which is difficult to quantify, and statistical precision due to simulation sampling, where, however, standard error estimates can serve as a surrogate for estimating precision. Regarding the discretization choice, Tagaras (1996) writes (p.47): "the specific needs of an actual problem with different process

Table 1: Comparison of Computational Effort (CPU seconds) as a function of precision (# simulation replications or degree of Markov chain state discretization).

λ	δ	CL	# simulation reps		# MC states (CPLEX/MATLAB)	
			1,000	10,000	100	1000
0.1	2.0	0.25	0.09	0.73	0.10/0.41	32/52
0.1	2.0	0.50	0.12	1.1	0.12/0.34	88/52
0.1	2.0	0.75	0.15	1.3	0.17/0.34	112/52
0.1	2.0	0.95	0.18	1.6	0.20/0.33	129/53
0.05	2.0	0.25	0.16	1.4	0.07/0.33	32/61
0.05	2.0	0.50	0.24	2.1	0.11/0.33	87/55
0.05	2.0	0.75	0.30	2.7	0.15/0.35	110/53
0.05	2.0	0.95	0.33	3.0	0.18/0.33	131/53
0.01	2.0	0.25	0.8	7.3	0.10/0.34	37/54
0.01	2.0	0.50	1.2	11	0.12/0.38	73/62
0.01	2.0	0.75	1.4	13	0.15/0.35	99/62
0.01	2.0	0.95	1.6	14	0.18/0.34	131/54

and cost parameters will generally necessitate careful experimentation for the appropriate choice of discretization." For example, in Table 1 of Tagaras (1996), the % cost differences between the optimal static control chart and optimal dynamic control chart (one of the main contributions of the paper was to show the advantage of the latter over the former) are generally significantly smaller than the % cost differences between a $K = 100$ and $K = 1000$ discretization, a clear indication of the potential impact of the degree of discretization.

5 Summary and Conclusions

In the economic design of control charts, the gradient of the expected cost function with respect to the control limits plays a significant role. In adopting a simulation-based approach to tackling the problem, we have applied a gradient estimation method that makes use of the smoothing property of conditioning to derive both right-hand and left-hand derivative estimators, which condition on different sample path quantities and lead to final forms that appear to be quite different. For the problem of the economic design of Bayes and EWMA charts, we are able to obtain gradient estimates that can be considerably more efficient than estimators based on finite difference approximations, and have shown that this leads to a simulation-based approach that provides a viable alternative to other numerical approaches.

Our method enables us to efficiently determine the economically optimal control limits for a broad class of control charts, which includes the Bayes and EWMA (and $\bar{X}$) charts as special cases. By providing an efficient technique for solving the optimization problem that is at the heart of the economic design approach, we hope to facilitate a wider acceptance of the economic design approach to control charts, and in particular, of sensitive control charts such as the EWMA and Bayes charts.

References

[1] Al-Sultan, K.S. and Rahim, M.A. (eds.) *Optimization in Quality Control*, Kluwer, 1997.

[2] Albin, S.A., Kang, L., and Shea, G., "An X and EWMA Chart for Individual Observations," *Journal of Quality Technology*, 29, 41-48, 1997.

[3] Banerjee, P.K. and Rahim, M.A., "Economic Design of $\bar{X}$ Control Charts Under Weibull Shock Models," *Technometrics*, 30, 407-414, 1988.

[4] Barish, N.N. and Hauser, N., "Economic Design for Control Decisions," *The Journal of Industrial Engineering*, Vol.14, 125-134, 1963.

[5] Bather, J.A., "Control Charts and Minimization of Costs," *Journal of the Royal Statistical Society*, Series B, 25, 49-80, 1963.

[6] Calabrese, J.M., "Bayesian Process Control for Attributes," *Mgmt Sci*, 41, 637-645, 1995.

[7] v. Collani, E., "Economic Design of Control Charts Simplified," in *Optimization in Quality Control*, edited by Al-Sultan, K.S. and Rahim, M.A., Kluwer, 1997.

[8] Fu, M.C. and Hu, J.Q., *Conditional Monte Carlo: Gradient Estimation and Optimization Applications*, Kluwer, 1997.

[9] Fu, M.C. and Hu, J.Q., "Efficient Design and Sensitivity Analysis of Control Charts Using Monte Carlo Simulation," *Management Science*, 45, 395-413, 1999.

[10] Girshick, M.A. and Rubin, H., "A Bayes' Approach to a Quality Control Model," *Annals of Mathematical Statistics*, 23, 114-125, 1952.

[11] Glasserman, P., *Gradient Estimation Via Perturbation Analysis*, Kluwer, 1991.

[12] Gong, W.B. and Ho, Y.C., "Smoothed Perturbation Analysis of Discrete-Event Dynamic Systems," *IEEE Transactions on Automatic Control*, 32, 858-867, 1987.

[13] Grimshaw, S.D. and Alt, F.B., "Control Charts for Quantile Function Values," *Journal of Quality Technology*, 29, 1-7, 1997.

[14] Ho, C. and Case, K.E., "Economic Design of Control Charts: A Literature Review for 1981-1991," *Journal of Quality Technology*, 26, 39-53, 1994.

[15] Hunter, J.S. "The Exponentially Weighted Moving Average," *Journal of Quality Technology,* 18, 203-210, 1986.

[16] Krishnamoorthi, K.S., "Economic Control Charts-An Application," *ASQC Quality Congress Transactions*, 385-391, 1985.

[17] Law, A.M. and Kelton, W.D., *Simulation Modeling and Analysis*, McGraw-Hill, 1991.

[18] Lele, S., "Steady State Analysis of Three Process Monitoring Procedures in Quality Control," (Ph.D. dissertation, University of Michigan, Ann Arbor) University Microfilms Inc., 1996.

[19] Lorenzen T.J. and Vance, L.C., "The Economic Design of Control Charts: A Unified Approach," *Technometrics*, Vol. 28, 3-10, 1986.

[20] McWilliams, T.P., "Economic Control Chart Designs and the In-Control Time Distribution: A Sensitivity Analysis," *Journal of Quality Technology*, 21, 103-110, 1989.

[21] Montgomery, D.C., "The Economic Design of Control Charts: A Review and Literature Survey," *Journal of Quality Technology*, 12, 75-87, 1980.

[22] Montgomery, D.C., *Introduction to Statistical Quality Control*, 3rd ed., Wiley, 1996.

[23] Pollock, S.M. and Alden, J., "Operating Characteristics for Detection of Process Change, Part 1: Theoretical Development," Technical Report No. 92-34, Dept. of Industrial and Operations Engineering, University of Michigan, 1992.

[24] Porteus, E.L. and Angelus, A., "Opportunities for Improved Statistical Process Control," *Management Science*, 43, 1214-1228, 1997.

[25] Roberts, S.W. "Control Chart Tests Based on Geometric Moving Averages," *Technometrics,* 1, 239-250, 1959.

[26] Svoboda, L., "Economic Design of Control Charts: A Review and Literature Survey (1979-1989)," in *Statistical Process Control in Manufacturing*, ed. J.B. Keats and D.C. Montgomery, Marcel Dekker, 1991.

[27] Tagaras, G., "A Dynamic Programming Approach to the Economic Design of $\bar{X}$-Charts," *IIE Transactions*, 26, 48-56, 1994.

[28] Tagaras, G., "Dynamic Control Charts for Finite Production Runs," *European Journal of Operational Research*, 91, 38-55, 1996.

[29] Tagaras, G., "A Survey of Recent Developments in the Design of Adaptive Control Charts," *Journal of Quality Technology*, 30, 212-231, 1998.

[30] Tagaras, G. and Nikolaidis, Y., "Comparing the Effectiveness of Various Bayesian $\bar{X}$Control Charts," *Operations Research*, 50, 878-888, 2002.

[31] Taylor, H.M., "Markovian Sequential Replacement Processes," *Annals of Mathematical Statistics*, 36, 1677-1694, 1965.

[32] Taylor, H.M., "Statistical Control of a Gaussian Process," *Technometrics*, 9, 29-41, 1967

Appendix

Proof of Proposition 1

(A1) ensures that the estimator (7) is well defined, i.e., that (6) holds, so we have

$$\begin{aligned} E\left(\frac{dE[S_k]}{du}\right)_{RH} &= E\Big[\lim_{\Delta u\downarrow 0} E[S_k(u+\Delta u)-S_k(u)\mid \mathcal{Z}, u<Y_\tau\le u+\Delta u] \\ &\quad \times \mathbf{1}\{Y_\tau>u\}\frac{f_\tau(\psi^{-1}(u,Y_{\tau-1}))}{1-F_\tau(\psi^{-1}(u,Y_{\tau-1}))}\frac{d\psi^{-1}(u,Y_{\tau-1})}{du}\Big], \\ &= E\Big[\lim_{\Delta u\downarrow 0} E[S_k(u+\Delta u)-S_k(u)\mid \mathcal{Z}, u<Y_\tau\le u+\Delta u] \\ &\quad \times \mathbf{1}\{Y_\tau>u\}\frac{P(u<Y_\tau\le u+\Delta u\mid \mathcal{Z}, Y_\tau>u)}{\Delta u}\Big]. \end{aligned} \tag{19}$$

As usual, the key technical step is the exchange of limit and expectation — the justification of the step giving (5) — which requires application of the dominated convergence theorem. Under the remaining conditions, (A2)-(A4), this is easily established. We divide it into two separate parts, i.e.,

$$\begin{aligned} & E\Big[\lim_{\Delta u\downarrow 0} E[S_k(u+\Delta u) \mid \mathcal{Z}, u<Y_\tau\le u+\Delta u]\mathbf{1}\{Y_\tau>u\}\frac{P(u<Y_\tau\le u+\Delta u \mid \mathcal{Z}, Y_\tau>u)}{\Delta u}\Big] \\ = & \lim_{\Delta u\downarrow 0} E\Big[E\Big[S_k(u+\Delta u) \mid \mathcal{Z}, u<Y_\tau\le u+\Delta u]\mathbf{1}\{Y_\tau>u\}\frac{P(u<Y_\tau\le u+\Delta u \mid \mathcal{Z}, Y_\tau>u)}{\Delta u}\Big]\Big], \end{aligned}$$

$$\begin{aligned} & E\Big[\lim_{\Delta u\downarrow 0} E[S_k(u) \mid \mathcal{Z}, u<Y_\tau\le u+\Delta u]\mathbf{1}\{Y_\tau>u\}\frac{P(u<Y_\tau\le u+\Delta u \mid \mathcal{Z}, Y_\tau>u)}{\Delta u}\Big] \\ = & \lim_{\Delta u\downarrow 0} E\Big[E\Big[S_k(u) \mid \mathcal{Z}, u<Y_\tau\le u+\Delta u]\mathbf{1}\{Y_\tau>u\}\frac{P(u<Y_\tau\le u+\Delta u \mid \mathcal{Z}, Y_\tau>u)}{\Delta u}\Big]\Big]. \end{aligned}$$

To establish each of these interchanges, the following bounds suffice: for some $\epsilon>0$,

$$E\left[\sup_{0\le\Delta u\le\epsilon} E[S_k(u+\Delta u) \mid \mathcal{Z}, u<Y_\tau\le u+\Delta u]\mathbf{1}\{Y_\tau>u\}\frac{P(u<Y_\tau\le u+\Delta u \mid \mathcal{Z}, Y_\tau>u)}{\Delta u}\right]<\infty.$$

$$E\left[\sup_{0\le\Delta u\le\epsilon} E[S_k(u) \mid \mathcal{Z}, u<Y_\tau\le u+\Delta u]\mathbf{1}\{Y_\tau>u\}\frac{P(u<Y_\tau\le u+\Delta u \mid \mathcal{Z}, Y_\tau>u)}{\Delta u}\right]<\infty.$$

Since $S_k\le\tau$, we have by (A4), for some $\epsilon>0$:

$$\begin{aligned} E\left[\sup_{0\le\Delta u\le\epsilon} E[S_k(u+\Delta u) \mid \mathcal{Z}, u<Y_\tau\le u+\Delta u]\right] &\le E\left[\sup_{0\le\Delta u\le\epsilon} E[\tau(u+\Delta u) \mid \mathcal{Z}, u<Y_\tau\le u+\Delta u]\right] \\ &\le E\left[E[\tau(u+\epsilon) \mid \mathcal{Z}, u<Y_\tau\le u+\Delta u]\right] \\ &\quad \text{(since } \tau(\cdot) \text{ is increasing for a fixed sample path)} \\ &\le E[\tau(u+\epsilon)]<K_3, \end{aligned} \tag{20}$$

$$\begin{aligned} E\left[\sup_{0\le\Delta u\le\epsilon} E[S_k(u) \mid \mathcal{Z}, u<Y_\tau\le u+\Delta u]\right] &\le E\left[\sup_{0\le\Delta u\le\epsilon} E[\tau(u) \mid \mathcal{Z}, u<Y_\tau\le u+\Delta u]\right] \\ &\le E\left[E[\tau(u) \mid \mathcal{Z}, u<Y_\tau\le u+\Delta u]\right] \\ &= E[\tau(u)]<K_3. \end{aligned} \tag{21}$$

Next, since $\psi^{-1}(\cdot,\cdot)$ is decreasing w.r.t. its second argument due to (A2) and $l\le Y_{\tau-1}\le u$,

$$\begin{aligned} \left|\frac{P(u<Y_\tau\le u+\Delta u \mid \mathcal{Z}, Y_\tau>u)}{\Delta u}\right| &= \left|\frac{1}{\Delta u}\frac{F_\tau(\psi^{-1}(u+\Delta u, Y_{\tau-1}))-F_\tau(\psi^{-1}(u, Y_{\tau-1}))}{1-F_\tau(\psi^{-1}(u, Y_{\tau-1}))}\right| \\ \text{(using (A2) \& (A3))} &\le \frac{1}{\Delta u}\frac{K_2|\psi^{-1}(u+\Delta u, Y_{\tau-1})-\psi^{-1}(u, Y_{\tau-1})|}{1-F_\tau(\psi^{-1}(u,l))}, \\ &\le \frac{K_1K_2}{1-F_\tau(\psi^{-1}(u,l))} \quad \text{via (A2)} \\ &\le \frac{K_1K_2}{1-\max(F^{(0)}(\psi^{-1}(u,l)), F^{(\delta)}(\psi^{-1}(u,l)))} \quad \text{via (A2).} \end{aligned}$$

Combining the two bounds, we have

$$
\begin{aligned}
& E\left[\sup_{0\le\Delta u\le\epsilon} E[S_k(u+\Delta u) \mid \mathcal{Z}, u<Y_\tau\le u+\Delta u]\mathbf{1}\{Y_\tau>u\}\frac{P(u<Y_\tau\le u+\Delta u \mid \mathcal{Z}, Y_\tau>u)}{\Delta u}\right] \\
\le\ & E\left[\sup_{0\le\Delta u\le\epsilon} E[S_k(u+\Delta u) \mid \mathcal{Z}, u<Y_\tau\le u+\Delta u]\frac{P(u<Y_\tau\le u+\Delta u \mid \mathcal{Z}, Y_\tau>u)}{\Delta u}\right] \\
\le\ & E\left[\sup_{0\le\Delta u\le\epsilon} E[S_k(u+\Delta u) \mid \mathcal{Z}, u<Y_\tau\le u+\Delta u]\sup_{0\le\Delta u\le\epsilon}\frac{P(u<Y_\tau\le u+\Delta u \mid \mathcal{Z}, Y_\tau>u)}{\Delta u}\right] \\
\le\ & E\left[\sup_{0\le\Delta u\le\epsilon} E[S_k(u+\Delta u) \mid \mathcal{Z}, u<Y_\tau\le u+\Delta u]\frac{K_1K_2}{1-F_\tau(\psi^{-1}(u,l))}\right] \\
\le\ & \frac{K_1K_2}{1-\max(F^{(0)}(\psi^{-1}(u,l)),F^{(\delta)}(\psi^{-1}(u,l)))}E\left[\sup_{0\le\Delta u\le\epsilon} E[S_k(u+\Delta u) \mid \mathcal{Z}, u<Y_\tau\le u+\Delta u]\right] \\
\le\ & \frac{K_1K_2K_3}{1-\max(F^{(0)}(\psi^{-1}(u,l)),F^{(\delta)}(\psi^{-1}(u,l)))}<\infty.
\end{aligned}
$$

The same bound can be established for the other term (involving $S_k(u+\Delta u)$), so now we can exchange limit and expectation in (19), and simply follow the equations leading to (5) backwards, i.e.,

$$
\begin{aligned}
E\left(\frac{dE[S_k]}{du}\right)_{RH} &= \lim_{\Delta u\downarrow 0} E\Big[E\Big[(S_k(u+\Delta u)-S_k(u)) \mid \mathcal{Z}, u<Y_\tau\le u+\Delta u]\mathbf{1}(Y_\tau>u) \\
&\quad \times P(u<Y_\tau\le u+\Delta u \mid \mathcal{Z}, Y_\tau>u)/\Delta u\Big]\Big] \\
&= \lim_{\Delta u\downarrow 0}\frac{E[S_k(u+\Delta u)-S_k(u)]}{\Delta u}=\frac{dE[S_k]}{du}.
\end{aligned}
$$

Markov Chain Model for Bayes Chart

To define a Markov chain model for Bayes charts, we assume a given control limit θ, as well as the process parameters λ and δ. To determine the long-run average costs given this control limit using a Markov chain, we first need to specify the state space $\mathcal{S}$ and transition probabilities $\{p_{ij}\}$. For Bayes charts, the states represent the posterior probabilities of being out of control. When discretized, this yields

$$\mathcal{S}=\left\{\frac{1}{2K}+\frac{k}{K}: k=0,\ldots,K-1\right\}\cup\{R\},$$

with K representing the degree of discretization of states and R representing the repair state. The transition probabilities p_{ij} can be defined using the approach proposed in Tagaras (1994) and Calabrese (1995); see also Tagaras and Nikolaidis (2002). We note that, in accordance with our definition of the process model, $p_{RR}=0$, $p_{iR}=0$ for $i\le\theta$, and $p_{iR}=1$ for $i>\theta$ (so $p_{ij}=0$ for $i>\theta, j\ne R$). Now we define the remaining transition probabilities.

First, we consider starting from the repair state R, and define the density function of the next sample X, given the current state R:

$$f(X|R) = (1-\lambda)f_0(X) + \lambda f_1(X).$$

Next, we define the transition $T(R|X)$ which represents the transformation of R given a sample X:

$$T(R|X) = \frac{\lambda f_1(X)}{f(X|R)},$$

i.e., $T(R|X)$ equals the posterior probability given that we start in the repair state and that the sample equals X. Given these terms, we can calculate the likelihood that the posterior probability equals j (disregarding, for now, any discretization) as follows:

$$p(j|R) = \int \mathbf{1}\{j = T(R|X)\} f(X|R) dX.$$

Since we assume sample distributions that are normally distributed,

$$\begin{aligned}
& j = T(R|X) = \frac{\lambda f_1(X)}{(1-\lambda)f_0(X) + \lambda f_1(X)} \\
\Leftrightarrow \quad & \frac{1}{j} = 1 + \frac{(1-\lambda)f_0(X)}{\lambda f_1(X)} \\
\Leftrightarrow \quad & \frac{1-j}{j}\frac{\lambda}{1-\lambda} = \frac{f_0(X)}{f_1(X)} = \frac{e^{-\frac{1}{2}X^2}}{e^{-\frac{1}{2}(X-\delta)^2}} = e^{\frac{1}{2}\delta^2 - \delta X} \\
\Leftrightarrow \quad & \ln(\frac{1-j}{j}) + \ln(\frac{\lambda}{1-\lambda}) = \frac{1}{2}\delta^2 - \delta X \\
\Leftrightarrow \quad & X = \frac{1}{2}\delta - \frac{1}{\delta}\ln(\frac{1-j}{j}) - \frac{1}{\delta}\ln(\frac{\lambda}{1-\lambda}).
\end{aligned}$$

Thus,

$$p(j|R) = f\left(\frac{1}{2}\delta - \frac{1}{\delta}\ln\left[\frac{1-j}{j}\right] - \frac{1}{\delta}\ln\left[\frac{\lambda}{1-\lambda}\right]\middle| R\right),$$

and the discretized transition probability is defined by

$$p_{Rj} = \int_{h=j-\frac{1}{2K}}^{h=j+\frac{1}{2K}} f\left(\frac{1}{2}\delta - \frac{1}{\delta}\ln\left[\frac{1-h}{h}\right] - \frac{1}{\delta}\ln\left[\frac{\lambda}{1-\lambda}\right]\middle| R\right) dh.$$

For the transition probabilities for $i \leq \theta$ and $j \neq R$, we follow the same steps as before.

$$\begin{aligned}
f(X|i) &= (1-\lambda)(1-i)f_0(X) + (\lambda + (1-\lambda)i)f_1(X), \\
T(i|X) &= \frac{(\lambda + (1-\lambda i))f_1(X)}{f(X|R)}, \\
p(j|i) &= \int \mathbf{1}\{j = T(i|X)\} f(X|i) dX \\
&= f\left(\frac{1}{2}\delta - \frac{1}{\delta}\ln\left[\frac{1-j}{j}\right] - \frac{1}{\delta}\ln\left[\frac{(\lambda + (1-\lambda)i)}{(1-\lambda)(1-i)}\right]\middle| i\right).
\end{aligned}$$

Putting this all together, we have ($j \neq R$)

$$p_{ij} = \begin{cases} \int_{h=j-\frac{1}{2K}}^{h=j+\frac{1}{2K}} f\left(\frac{1}{2}\delta - \frac{1}{\delta}\ln\left[\frac{1-h}{h}\right] - \frac{1}{\delta}\ln\left[\frac{\lambda}{1-\lambda}\right] \middle| R\right) dh, & i = R, \\ \int_{h=j-\frac{1}{2K}}^{h=j+\frac{1}{2K}} f\left(\frac{1}{2}\delta - \frac{1}{\delta}\ln\left[\frac{1-h}{h}\right] - \frac{1}{\delta}\ln\left[\frac{(\lambda+(1-\lambda)i)}{(1-\lambda)(1-i)}\right] \middle| i\right) dh, & i \leq \theta. \end{cases}$$

Given these components, we can solve the resulting Markov chain to obtain the steady-state probabilities $\pi_i(\theta)$ for $i \in \mathcal{S}$. This yields a long-run steady-state expected cost per unit time, for the given control limit θ:

$$C(\theta) = k_r \pi_R(\theta) + k_\delta \sum_{i \in \mathcal{S}, i \neq R} i\pi_i(\theta) + k_0(1 - \sum_{i \in \mathcal{S}, i \neq R} i\pi_i(\theta)).$$

As in the simulation-based approach, we discretize the control limits and solve the Markov chain for each discretized control limit value over the range $\theta \in (0, 1)$, which yields steady-state probabilities $\pi_j(\theta)$ that can be used to determine the probability of being in the repair state, i.e., $p_R(\theta) = \pi_R(\theta)$, as well as the long-run probabilities of being in the out-of-control, $p_\delta(\theta) = \sum_{j \in \mathcal{S}: j \neq R} j\pi_j(\theta)$, and hence approximate the threshold cost ratios

$$C(\theta) = -\frac{dp_\delta(\theta)}{dp_R(\theta)} \approx -\frac{p_\delta(\theta) - p_\delta(\theta - 1/K)}{p_R(\theta) - p_R(\theta - 1/K)}.$$

In using CPLEX to solve the discretized Markov chain model, the fact that the Markov chains for the various control limits are solved sequentially in this manner can be exploited to decrease the overall computational time to generate the entire curve, as the previous solution can be partially used, e.g., in Table 1, the incremental time needed to solve an individual case will decrease substantially, to approximately 5 seconds per instance for a discretization of $K = 1000$. A similar incremental procedure could be used in the simulation-based approach: given that control limits are sequentially increased, information from previous runs could be reused.