# A Bayesian Scheme to Detect Changes in the Mean of a Short Run Process

Panagiotis Tsiamyrtzis and Douglas M. Hawkins

School of Statistics, University of Minnesota

Technical Report #642

September 25, 2002

# A Bayesian Scheme to Detect Changes in the Mean of a Short Run Process

**Panagiotis Tsiamyrtzis**
Department of Statistics
Athens University of Economics
and Business
76 Patission Str, 10434
Athens, Greece
pt@aueb.gr

**Douglas M. Hawkins**
School of Statistics
University of Minnesota
313 Ford Hall
224 Church Street S.E.
Minneapolis, MN 55455
doug@stat.umn.edu

**Abstract**

In this paper we propose a statistical model for short production runs. We wish to detect on line whether the mean of the process has exceeded a prespecified upper threshold value. The theoretic development of the model is based on a Bayesian update of a mixture of normal distributions. Issues of decisions about whether the process is within specification and forecasting will be treated. The Kalman filter model is shown to be related to a special case of our model. The calculations are illustrated with a clinical chemistry example. Other particularly relevant settings include tool wear problems.

**Key Words**: Normal Mixtures, Bayesian Statistical Process Control, Kalman Filter, Tool Wear.

# 1 Introduction

A sequence $X_i$ of random variables is observed over time, and as each observation becomes available to us we are faced with the question of whether the mean has been crossed an upper threshold value. If the values are independent and identically distributed, and if we have a long historical series to calibrate the process, then conventional SPC methodologies such as Shewhart or cusum charts solve the problem. We however will be concerned with short runs, and will also move away from the independence setting.

There are many proposals for short-run SPC; some are sketched in Wheeler (1992). Bothe (1989) and Burr (1989) proposed scaling the quality characteristic by target values. Quesenberry (1990) and (1991) also in a purely frequentist way developed the Q-charts, where running process estimates of the mean and the variance are used. The Q-charts can be severely affected if assignable cause variation is present at the start up. Del Castillo and Montgomery (1992) proposed a control chart based upon a first order adaptive Kalman filter model. Wasserman and Sudjianto (1993) extended these ideas to a control chart based on the second order dynamic linear model. Wright et al. (2001) proposed joint estimation to detect outliers in short run autocorrelated data. Finally Woodward and Naylor (1993) addressed the short run problem with a simple Bayesian paradigm. In the case where we do not have necessarily short runs and we are not interested in detecting the mean drifts on line, Chang and Fricker (1999) proposed use of a CUSUM, while from the Bayesian change point methodology Chernoff and Zacks (1964) proposed the use of mixture normal distributions.

In some circumstances, we need to have a well defined inferential procedure even when we have available a single observation, and this leads inevitably to a Bayesian framework. Here, at any given time we have available prior information about the mean of the process. As each data point

becomes available, through Bayes theorem we get the posterior distribution for the mean and use this to draw some inference (of whether the mean has drifted above the threshold or not). If we decide that no 'significant' change has been occurred we use this posterior as prior for the next stage.

## 2 Statistical Modeling

At time 0, prior to any process readings, we have a distribution for the parameter of interest:

$$\theta_0 \sim N(\zeta, \sigma_0^2)$$

where $\zeta, \sigma_0^2$ are the prior mean and variance, assumed specified. We will use the model that when moving from time $n-1$ to time $n$

$$\theta_n|\theta_{n-1} \sim \begin{cases} N(\theta_{n-1}, \quad \sigma 2) & \text{with probability} \quad p \\ N(\theta_{n-1}+\delta, \ \sigma 2) & \text{with probability } 1-p \end{cases}$$

In other words, the parameter drifts according to a normal random walk, and is also subject to occasional shocks. Initially, we assume that $\sigma^2, \delta$ and $p$ are known. The $\sigma^2$ represents the model's variability and $\delta$ is the size of a positive jump that occurs with probability $1-p$. If $\delta = 0$ (or $p = 1$) then it is a pure random walk with no shocks.

At each time $n$ of our process we are not able to observe $\theta_n$ directly, but have a measurement $X_n$ with distribution

$$X_n|\theta_n \sim N(\theta_n, \ \tau 2)$$

Here, $\tau 2$ (assumed known) represents the variability due to measurement error.

The accumulated data observed up to time $n$: $x_1, x_2, \ldots x_n$ will be denoted

$$\mathbf{X}_n = \{x_1, x_2, \ldots, x_n\}$$

At each stage $i$ of our process, once $X_i = x_i$ is available, we can obtain the posterior distribution of $\theta_i|\mathbf{X}_i$ and draw some inference for the parameter $\theta_i$. If the process continues to operate then we use the posterior distribution of $\theta_i|\mathbf{X}_i$ as a prior distribution for $\theta_{i+1}$ and so on.

At each stage of the process we have two possible scenarios for $\theta$ (jump or no jump), so the posterior distribution of $\theta_n|\mathbf{X}_n$ is a mixture of $2^n$ distributions. More specifically we have:

**Theorem 1** *At time $n$ the posterior distribution of $\theta_n|\mathbf{X}_n$ is a mixture of $2^n$ Normal distributions. All the components have identical variance:*

$$\hat{\sigma}_n 2 = (1 - K_n)\tau 2 = K_n(\sigma 2 + \hat{\sigma}^2_{n-1})$$

*where $K_n = \tau 2/(\tau 2 + \sigma 2 + \hat{\sigma}^2_{n-1})$ and $\hat{\sigma}^2_{n-1}$ is the common posterior variance of the components at time $n-1$. The $2^n$ Normal components have means given by the formula:*

$$\hat{\theta}_n + f(K_i, \delta, \jmath, \ell)$$

*where $\hat{\theta}_n = K_n\hat{\theta}_{n-1} + (1 - K_n)x_n$ refers to the no jump case at the end of the $n$-th time point and $f(K_i, \delta, \jmath, \ell)$ is a function depending on $K_i$, the size of the jump $\delta$, the number of previous jumps ($\jmath = 0, 1, \ldots, n$) and the time points at which they occurred ($\ell$ can take as values all the possible subsets of the set $\{1, 2, \ldots, n\}$). The $2^n$ different values of $f(K_i, \delta, \jmath, \ell)$ are given in Table 1:*

**Table 1 about here**

The proof of theorem 1 is given in Appendix 1.

At time $n$ we can easily show that $\hat{\theta}_n$ will be given by:

$$\hat{\theta}_n = K_n \ldots K_1\zeta + K_n \ldots K_2(1 - K_1)x_1 + \cdots + K_n(1 - K_{n-1})x_{n-1} + (1 - K_n)x_n$$

Therefore $\hat{\theta}_n$ is a convex combination of the prior mean $\zeta$ and the data points $x_1, x_2, \ldots, x_n$. The coefficients in the convex combination are determined by the constants $K_i$, $(1 \leq i \leq n)$, which turn

out (see the Appendix 1) to satisfy

$$K_i = \frac{\tau 2}{\tau 2 + \sigma 2 + \hat{\sigma}^2_{i-1}}, \quad \text{where} \quad \hat{\sigma}^2_{i-1} = (1 - K_{i-1})\tau 2$$

These constants depend only on the values of the variances $\tau 2, \sigma 2$ and the initial prior variance $\sigma 2_0$. First of all we will show that this sequence converges.

Define $c = \tau 2/\sigma 2$. Then

$$K_n = \frac{\tau 2}{\tau 2 + \sigma 2 + (1 - K_{n-1})\tau 2} = \frac{c\sigma 2}{c\sigma 2 + \sigma 2 + (1 - K_{n-1})c\sigma 2} = \frac{c}{1 + (2 - K_{n-1})c} = \varphi(K_{n-1})$$

where $K_i \in [0,1], \forall i$.

It is easy to see that the function $\varphi : [0,1] \to [0,1]$ satisfies a Lipschitz condition. Thus the sequence of $K_n$ has a unique fixed point in $[0,1]$ (Kress (1998)) $K^* \in [0,1]$ satisfying $\varphi(K^*) = K^*$. A closed-form expression for $K^*$ (see Appendix 2) is:

$$K^* = 1 + \frac{1}{2c} - \sqrt{\frac{1}{4c2} + \frac{1}{c}}$$

The convergence of the sequence of $K_n$'s to a value $K^*$ is of particular interest, because it connects our model to the familiar Exponentially Weighted Moving Average (EWMA) technique (Roberts (1959)). Following Montgomery (1997) or Montgomery and Mastrangelo (1991), the EWMA is defined as:

$$z_0 = \mu_0$$

$$z_i = \lambda z_{i-1} + (1 - \lambda)x_i \qquad i = 1, 2, \ldots$$

where $\mu_0$ is the starting value, $0 < \lambda \leq 1$ is a constant.

In our model convergence of $K_n$'s to $K^*$ gives the asymptotic recursion

$$\hat{\theta}_n = K^*\hat{\theta}_{n-1} + (1 - K^*)X_n$$

which is an EWMA. Therefore as $n$ increases $\hat{\theta}_n$ approaches an EWMA and so our model gives an EWMA with special start-up and with smoothing parameter an explicit function of the ratio of the drift to the measurement variances.

Turning to the unconditional process mean, we have

$$\theta_n = \theta_{n-1} + u_n \quad \text{where} \quad u_n \sim \left\{ \begin{array}{ll} N(0,\ \sigma 2) & \text{with prob.} \quad p \\ N(\delta,\ \sigma 2) & \text{with prob. } 1-p \end{array} \right\}$$

thus we have:

$$\theta_n = \theta_0 + \sum_{i=1}^{n} u_i$$

from which it follows that

$$E(\theta_n) = E(\theta_0) + \sum_{i=1}^{n} E(u_i) = \zeta + \sum_{i=1}^{n} (1-p)\delta = \zeta + n(1-p)\delta$$

and

$$Var(\theta_n) = Var(\theta_0) + Var\left(\sum_{i=1}^{n} u_i\right) = \sigma_0 2 + nVar(u_i)$$

but

$$\begin{aligned} Var(u_i) &= E(u_i 2) - [E(u_i)]^2 \\ &= p\sigma 2 + (1-p)(\sigma 2 + \delta 2) - (1-p)2\delta 2 \\ &= p\sigma 2 + \sigma 2 - p\sigma 2 + \delta 2[(1-p) - (1-p)2] \\ &= \sigma 2 + \delta^2 p(1-p) \end{aligned}$$

Thus

$$Var(\theta_n) = \sigma_0 2 + n[\sigma 2 + \delta^2 p(1-p)]$$

So we have that as $n \to \infty$ we get $E(\theta_n) \to \infty$ and $Var(\theta_n) \to \infty$. Clearly our model can not describe a system that tends to a steady state, and it not suitable for all correlated processes

(Tsiamyrtzis (2000)). It is however attractive as a model for tool wear problems in which the wear incorporates a random step change as well as drift.

Of course, in short run processes, the lack of an asymptotic distribution may be academic and then the model will be successful to the extent that it reasonably describes actual process readings.

# 3 Inference

## 3.1 Action rules

In problems with random deterioration, the concern at each time point is whether the mean has degraded to the point that some corrective action is needed. We will make this specific as the decision of whether the parameter of interest $\theta_n$ has crossed some upper threshold value $M$. This leads to the sequence of decisions between

$$\left\{ \begin{array}{l} H_0: \ \theta_n \leq M \\ H_1: \ \theta_n > M \end{array} \right\}$$

If $H_0$ is not rejected at time $n$, the process continues to operate, while if it is rejected some corrective action is taken.

In a Bayesian framework we can make this decision using the posterior distribution of $\theta_n|\mathbf{X}_n$. Once the data $x_n$ is available, calculate the posterior probability $P_n = P(\theta_n \leq M|\mathbf{X}_n)$ and accept $H_0$ iff $P_n \geq c$ where $c$ is a specified cutoff value.

A helpful chart can be made by plotting a suitable function of the posterior probabilities $P_n$ – such as a normal deviate – in time order. This chart functions much like a conventional Shewhart chart.

If we will use the 'generalized $0-1$ loss' with $C_I$ and $C_{II}$ being the costs of type I and type II

error respectively then the test of the form:

$$\left\{ \begin{array}{l} \text{accept } H_0 : \theta_n \leq M, \quad \text{if} \quad P(\theta_n \leq M | \mathbf{X}_n) > \frac{C_{II}}{C_I + C_{II}} \\ \qquad \text{and} \\ \text{reject } H_0 : \theta_n \leq M, \quad \text{if} \quad P(\theta_n \leq M | \mathbf{X}_n) < \frac{C_{II}}{C_I + C_{II}} \end{array} \right\}$$

is a Bayes rule, or Bayes test (Casella and Berger (1990)).

We could also implement the decision rule using Bayes factors (Jeffrey (1948)). If we denote $P(H_i|x)$, $i = 0, 1$ the posterior probabilities of the hypothesis $H_i$, $i = 0, 1$ when the data $x$ were observed and $P(H_i)$ denote the prior probabilities of $H_i$, $i = 0, 1$ then the Bayes factor, $B$, is defined as the ratio of the posterior odds of $H_0$ to the prior odds of $H_0$:

$$B = \frac{P(H_0|x)/P(H_1|x)}{P(H_0)/P(H_1)}$$

Jeffreys (1948) provides a table of cutoff values for B, to be used when we decide about rejecting $H_0$ or not. In our study at every time $n$ we have available both the posterior $\pi(\theta_n|\mathbf{X}_n)$ and the prior $\pi(\theta_n)$ which we can use to calculate the Bayes factor $B$ and decide whether the mean has been shifted above the upper threshold value $M$ or not.

## 3.2 Forecasting

Another interesting issue that we will develop is forecasting. More precisely based on the available observations $\mathbf{X}_{n-1}$ we can obtain the predictive distribution of a future observation $X_n$ and use it to do forecasting (Geisser (1993)). The predictive distribution of $X_n|\mathbf{X}_{n-1}$ is given by:

$$P(X_n|\mathbf{X}_{n-1}) = \int f(X_n|\theta_n)\pi(\theta_n|\mathbf{X}_{n-1})\, d\theta_n$$

where $f(X_n|\theta_n)$ is the likelihood $N(\theta_n, \tau 2)$ at stage $n$ and $\pi(\theta_n|\mathbf{X}_{n-1})$ is the prior distribution of $\theta_n$ at stage $n$. Under the original model the form of $\pi(\theta_n|\mathbf{X}_{n-1})$ was derived in the proof of Theorem

1 (Appendix 1). Using that distribution we can easily show that the predictive distribution of $X_n$ given the available data $\mathbf{X}_{n-1}$ is a mixture of $2^n$ Normal components. The distribution function is given in Appendix 3.

# 4 Relationship of the Model to Kalman Filter

The Kalman filter method (Kalman 1960) is a visibly somewhat related method with potential use in short-run problems. Meinhold and Singpurwalla (1983) described the Kalman filter methodology as follows:

We wish to infer a parameter $\theta_n$ at time $n$ of a running process. We are not able to observe $\theta_n$ directly, but observe $Y_n$ which is related to $\theta_n$ by the *observation equation*:

$$Y_n = F_n\theta_n + v_n$$

where $F_n$ is known and the observation error $v_n$ is assumed $N(0,\ V_n)$, with known variance $V_n$.

The parameter of interest $\theta_n$ is not constant over time, but is characterized by the *system equation*

$$\theta_n = G_n\theta_{n-1} + w_n$$

where $G_n$ is known and the error $w_n$ is assumed $N(0,\ W_n)$, with $W_n$ being known.

The two error terms $v_n$ and $w_n$ are assumed internally and mutually independent. The major assumptions in the Kalman Filter is that the *observation error* $v_t$ and the *system equation error* $w_t$ are both independently Normally distributed with known variances $V_t$ and $W_t$ respectively.

Comparing Kalman filter model with our model we observe that they share the observation equation but the system equation is different. The Kalman filter does not allow any jumps of the

parameter of interest $\theta_n$, but does allow the coefficient $G_n$ to differ from 1. In the special case where $p = 1$ or $\delta = 0$ and $G_n = 1$, the two models are the same.

# 5 Numerical Example

We illustrate the methodology with a data set of size 10 (Table 2), kindly provided by Dr Daniel Schultz of the Rogasin Institute, which illustrates a common laboratory quality control situation.

**Table 2 about here**

The laboratory performs routine blood chemistry assays on patients. As part of the quality control procedures, a weekly measurement is made of the cholesterol level of a control sample. Because of sample deterioration with age, the true cholesterol level of the control sample changes drifts over time. It is also possible for shocks (such as environmental stresses) to add a jump to the true mean. In addition to the variation in the true mean, there is an independent measurement random variability. Once the true cholesterol level of the control sample has drifted 'too much', then that sample needs to be discarded and replaced with a fresh one. The quality control scheme therefore has to check each week's reading to come to a decision whether the control sample has exceeded the level indicating the end of its useful life.

To put this problem within the framework of the current methodology, we need values for the model parameters. By a combination of statistical analysis of historical measurements and judgment of how large a drift would have to be to necessitate replacing the control sample, we arrived at the following values:

$$\zeta = 144, \qquad \sigma_0 2 = \sigma 2 = 12, \qquad \tau 2 = 4, \qquad p = 0.9, \qquad \delta = 4\sigma$$

Finally, we set the upper threshold defining unacceptable deterioration to $M = 150mg/dL$. Then applying the original model to the data we obtained the following posterior probabilities (Table 3):

**Table 3 about here**

For the first eight data points the posterior probabilities suggest confidently that the mean $\theta_i$ is below the upper threshold. At the ninth observation we get an indication that the mean may be approaching the threshold. The tenth data point confirms this suggestion. It implies that the threshold has been crossed and it is time for a new standard cholesterol sample.

We investigated the sensitivity of our conclusions to the values of these parameters by repeating the calculations while varying the hyperparameters. We can summarize the results of this by noting that the conclusions hardly moved if we altered $\zeta$, $\sigma_0 2$, $p$ and $\delta$ but were sensitive to the value selected for $\sigma 2$ and $\tau 2$. The parameters of the prior at time 0 ($\zeta$ and $\sigma_0 2$) have little effect because at every stage of the process we are updating our prior using the posterior of the previous stage, thus even for very poor choices, the posterior distribution will be affected at the first stage only. The probability of not having a jump, $p$, has also a small effect, as long as it is not getting small enough to be almost equal to 0.5. In the context of this problem, we know jumps are rare, so $p$ must be well above 0.7. The size of the jump, $\delta$, is not well quantified by historical data, so the value chosen was a matter of judgment rather than empirical estimation. We used a multiple of the model's standard deviation ($\sigma$) to be able to discriminate between random variability and jump. Using a larger multiple of $\sigma$ would have a very small impact on our analysis. On the other hand the posterior probabilities are more sensitive on the choices of the $\sigma 2$ and $\tau 2$. If $\sigma 2$ is large and/or $\tau 2$ is small, the results are sensitive to the data while in the opposite situation the prior predominates.

## 6 Concluding Remarks

We have developed a dynamic model to handle short production runs where our main concern is ensuring that the mean stays below an upper threshold. The case of a lower threshold follows immediately from symmetry. The use of the Bayesian sequentially updated mixture of normal distributions led us to a generalized Kalman filter model (where the model error terms are allowed to have jumps). This model has a wide range of applications in quality (new machines, different raw materials and tool wear, for example). It is also attractive for such problems as monitoring vital signs of patients in intensive care, where we also have short series of measurement, serial correlation plausibly modelable by random walks, and upper thresholds for intervention. Apart from these short-run scenarios though, this model can be employed at the start up phase of any process, where absence of historical data prevents use of the classical charting tools.

## 7 Acknowledgments

We are grateful to Dr Daniel Schultz from Rogasin Institute for the data provided in section 5. This work was partially supported by grant DMS 9803622 from the National Science Foundation.

## Appendix 1

The proof will be given via induction.

For $n = 1$ we have:

$$\left\{ \begin{array}{rcl} \theta_1|\theta_0 & \sim & pN(\theta_0,\ \sigma 2) + (1-p)N(\theta_0+\delta,\ \sigma 2) \\ \theta_0 & \sim & N(\zeta,\ \sigma_0 2) \end{array} \right\} \Rightarrow$$

$$\Rightarrow \theta_1 \sim pN(\zeta,\ \sigma 2 + \sigma_0 2) + (1-p)N(\zeta+\delta,\ \sigma 2 + \sigma_0 2)$$

So after observing $X_1 = x_1$ we have:

$$\left\{ \begin{array}{rcl} X_1|\theta_1 & \sim & N(\theta_1,\ \tau 2) \\ \theta_1 & \sim & p\underbrace{N(\zeta,\ \sigma 2+\sigma_0 2)}_{\pi_1(\theta_1)} + (1-p)\underbrace{N(\zeta+\delta,\ \sigma 2+\sigma_0 2)}_{\pi_2(\theta_1)} \end{array} \right\} \Rightarrow$$

$$\Rightarrow \theta_1|X_1 = x_1 \sim \alpha_1^{(1)} N\left(\frac{\tau 2\zeta + (\sigma 2+\sigma_0 2)x_1}{\tau 2+\sigma 2+\sigma_0 2},\ \frac{\tau 2(\sigma 2+\sigma_0 2)}{\tau 2+\sigma 2+\sigma_0 2}\right)$$

$$+\alpha_2^{(1)} N\left(\frac{\tau 2(\zeta+\delta) + (\sigma 2+\sigma_0 2)x_1}{\tau 2+\sigma 2+\sigma_0 2},\ \frac{\tau 2(\sigma 2+\sigma_0 2)}{\tau 2+\sigma 2+\sigma_0 2}\right)$$

where $\alpha_i^{(j)}$ is the weight for the $i$-th component of the mixture of the Normals, at time $j$. For this first time point we have:

$$\alpha_1^{(1)} = \frac{pm_1(x_1)}{pm_1(x_1)+(1-p)m_2(x_1)}, \qquad \alpha_2^{(1)} = 1-\alpha_1^{(1)}$$

where $m_1(x_1)$, $m_2(x_2)$ are the marginal densities of $X_1 = x_1$ with respect to the prior densities $\pi_1(\theta_1)$ and $\pi_2(\theta_1)$ given by:

$$m_i(x_1) = \int_{-\infty}^{\infty} f(x_1|\theta_1)\pi_i(\theta_1)\,d\theta_1, \qquad i = 1,2$$

If we will call:

$$K_1 = \frac{\tau 2}{\tau 2+\sigma 2+\sigma_0 2},$$

$$\hat{\theta}_1 = K_1\zeta + (1-K_1)x_1 \quad \text{and} \quad \hat{\sigma}_1 2 = K_1(\sigma 2 + \sigma_0 2) = (1-K_1)\tau 2$$

then the posterior distribution of $\theta_1|\mathbf{X}_1$ is given by:

$$\theta_1|\mathbf{X}_1 \sim \alpha_1^{(1)} N(\hat{\theta}_1,\ \hat{\sigma}_1 2) + \alpha_2^{(1)} N(\hat{\theta}_1 + K_1\delta,\ \hat{\sigma}_1 2)$$

Thus for $n = 1$ the theorem holds. We will assume that it is true for $n-1$, and we will prove it for $n$. We know that:

$$\begin{aligned}
\theta_n|\theta_{n-1} &\sim pN(\theta_{n-1},\ \sigma 2) + (1-p)N(\theta_{n-1}+\delta,\ \sigma 2) \\
\theta_{n-1}|\mathbf{X}_{n-1} &\sim \alpha_1^{(n-1)} N(\hat{\theta}_{n-1},\ \hat{\sigma}_{n-1}^2) \\
&\quad + \alpha_2^{(n-1)} N(\hat{\theta}_{n-1} + K_{n-1}\delta,\ \hat{\sigma}_{n-1}^2) \\
&\quad \vdots \\
&\quad + \alpha_n^{(n-1)} N(\hat{\theta}_{n-1} + K_{n-1}\ldots K_1\delta,\ \hat{\sigma}_{n-1}^2) \\
&\quad \vdots \\
&\quad + \alpha_{2^{n-1}}^{(n-1)} N(\hat{\theta}_{n-1} + K_{n-1}\delta + \cdots + K_{n-1}\ldots K_1\delta,\ \hat{\sigma}_{n-1}^2)
\end{aligned}$$

Thus the prior distribution of $\theta_n$ at stage $n$ is given by:

$$\begin{aligned}
\theta_n|\mathbf{X}_{n-1} \sim\ & p\ \alpha_1^{(n-1)} N(\hat{\theta}_{n-1},\ \sigma 2 + \hat{\sigma}_{n-1}^2) \\
& + (1-p)\alpha_1^{(n-1)} N(\hat{\theta}_{n-1}+\delta,\ \sigma 2 + \hat{\sigma}_{n-1}^2) \\
& + p\ \alpha_2^{(n-1)} N(\hat{\theta}_{n-1} + K_{n-1}\delta,\ \sigma 2 + \hat{\sigma}_{n-1}^2) \\
& \vdots \\
& + p\ \alpha_n^{(n-1)} N(\hat{\theta}_{n-1} + K_{n-1}\ldots K_1\delta,\ \sigma 2 + \hat{\sigma}_{n-1}^2) \\
& + (1-p)\alpha_2^{(n-1)} N(\hat{\theta}_{n-1} + \delta + K_{n-1}\delta,\ \sigma 2 + \hat{\sigma}_{n-1}^2) \\
& \vdots
\end{aligned}$$

$$+(1-p)\alpha_{n}^{(n-1)}N(\hat{\theta}_{n-1}+\delta+K_{n-1}\dots K_1\delta,\ \ \sigma 2+\hat{\sigma}^2_{n-1})$$

$$\vdots$$

$$+\quad p\ \ \alpha_{2^{n-1}}^{(n-1)}N(\hat{\theta}_{n-1}+K_{n-1}\delta+\dots+K_{n-1}\dots K_1\delta,\ \ \sigma 2+\hat{\sigma}^2_{n-1})$$

$$+(1-p)\alpha_{2^{n-1}}^{(n-1)}N(\hat{\theta}_{n-1}+\delta+K_{n-1}\delta+\dots+K_{n-1}\dots K_1\delta,\ \ \sigma 2+\hat{\sigma}^2_{n-1})$$

The prior distribution of $\theta_n$ has $2^n$ terms. The order (from first to last) that these Normal components were written is increasing as a function of $\jmath$ (number of jumps) and in the class of constant values of $\jmath$ they are written in decreasing order as a function of $\ell$ (the location that the jump(s) occurred) from the most recent to the earliest one. So for example when we have $\jmath = 2$ we will write the $\binom{n}{2}$ Normal components in the following order (from first to last) as functions of $\ell$:

| $\ell$ | respective mean |
|---|---|
| $\{n, n-1\}$ | $\hat{\theta}_{n-1}+\delta+K_{n-1}\delta$ |
| $\vdots$ | $\vdots$ |
| $\{n, 1\}$ | $\hat{\theta}_{n-1}+\delta+K_{n-1}\dots K_1\delta$ |
| $\{n-1, n-2\}$ | $\hat{\theta}_{n-1}+K_{n-1}\delta+K_{n-1}K_{n-2}\delta$ |
| $\vdots$ | $\vdots$ |
| $\{2, 1\}$ | $\hat{\theta}_{n-1}+K_{n-1}\dots K_2\delta+K_{n-1}\dots K_1\delta$ |

At time $n$ the likelihood is: $X_n|\theta_n \sim N(\theta_n,\ \tau 2)$. Then:

$$\theta_n|\mathbf{X}_n \sim \sum_{i=1}^{2^n}\alpha_i^{(n)}p_i(\theta_n|\mathbf{X}_n)$$

where $p_i(\theta_n|\mathbf{X}_n)$ are the $2^n$ posterior Normal components of the mixture. So if we define $K_n = \tau 2/(\tau 2+\sigma 2+\hat{\sigma}^2_{n-1})$, then $\forall\ i = 1,\dots,2^n$

$$Var[p_i(\theta_n|\mathbf{X}_n)] = \frac{\tau 2(\sigma 2+\hat{\sigma}^2_{n-1})}{\tau 2+\sigma 2+\hat{\sigma}^2_{n-1}} = K_n(\sigma 2+\hat{\sigma}^2_{n-1}) = (1-K_n)\tau 2$$

If we call $\pi_i(\theta_n)$, $i = 1, \ldots, 2^n$ the $2^n$ Normal components of the prior distribution of $\theta_n$ then

$$E[p_i(\theta_n|\mathbf{X}_n)] = \frac{\tau^2 E[\pi_i(\theta_n)] + (\sigma 2 + \hat{\sigma}^2_{n-1})x_n}{\tau 2 + \sigma 2 + \hat{\sigma}^2_{n-1}} = K_n E[\pi_i(\theta_n)] + (1 - K_n)x_n$$

| $\jmath$ | $E[\pi_i(\theta_n)]$ | $K_n E[\pi_i(\theta_n)] + (1 - K_n)x_n$ |
|---|---|---|
| 0 | $\hat{\theta}_{n-1}$ | $\hat{\theta}_n = K_n\hat{\theta}_{n-1} + (1 - K_n)x_n$ |
| | $\hat{\theta}_{n-1} + \delta$ | $\hat{\theta}_n + K_n\delta$ |
| | $\hat{\theta}_{n-1} + K_{n-1}\delta$ | $\hat{\theta}_n + K_nK_{n-1}\delta$ |
| 1 | $\vdots$ | $\vdots$ |
| | $\hat{\theta}_{n-1} + K_{n-1}\ldots K_1\delta$ | $\hat{\theta}_n + K_nK_{n-1}\ldots K_1\delta$ |
| $\vdots$ | $\vdots$ | $\vdots$ |
| | $\hat{\theta}_{n-1} + \delta + \cdots + K_{n-1}\ldots K_{n-m+1}\delta$ | $\hat{\theta}_n + K_n\delta + \cdots + K_nK_{n-1}\ldots K_{n-m+1}\delta$ |
| $m$ | $\vdots$ | $\vdots$ |
| | $\hat{\theta}_{n-1} + K_{n-1}\ldots K_m\delta + \cdots + K_{n-1}\ldots K_1\delta$ | $\hat{\theta}_n + K_nK_{n-1}\ldots K_m\delta + \cdots + K_n\ldots K_1\delta$ |
| $\vdots$ | $\vdots$ | $\vdots$ |
| $n$ | $\hat{\theta}_{n-1} + \delta + K_{n-1}\delta + \cdots + K_{n-1}\ldots K_1\delta$ | $\hat{\theta}_n + K_n\delta + K_nK_{n-1}\delta + \cdots + K_n\ldots K_1\delta$ |

where $2 \leq m \leq n - 1$. Therefore we have shown that:

$$E[p_i(\theta_n|\mathbf{X}_n)] = K_n E[\pi_i(\theta_n)] + (1 - K_n)x_n = \hat{\theta}_n + f(K_i, \delta, \jmath, \ell)$$

## Appendix 2

Since $K^*$ is the fixed point of $K_n$ it will satisfy:

$$K^* = \varphi(K^*) \Rightarrow K^* = \frac{c}{1 + (2 - K^*)c} \Rightarrow K^* + K^*(2 - K^*)c = c \Rightarrow$$

$$\Rightarrow K^* + 2cK^* - c(K^*)2 - c = 0 \Rightarrow c(K^*)2 - (1 + 2c)K^* + c = 0 \qquad \textbf{(I)}$$

which is a quadratic. The discriminant is:

$$\Delta = (1+2c)2 - 4c2 = 1 + 4c2 + 4c - 4c2 = 1 + 4c$$

Since $\Delta > 0$ $(c > 0)$ the equation **(I)** has two real roots:

$$K^*_{1,2} = \frac{(1+2c) \pm \sqrt{1+4c}}{2c} = 1 + \frac{1}{2c} \pm \sqrt{\frac{1}{4c2} + \frac{1}{c}}$$

but given that $0 \leq K_n \leq 1 \; \forall \, n \in N$ we accept only the root

$$K^* = 1 + \frac{1}{2c} - \sqrt{\frac{1}{4c2} + \frac{1}{c}}$$

as the unique fixed point of $\varphi(K_n)$ in [0,1], $\forall \, c \in (0, +\infty)$.

## Appendix 3

An immediate consequence of the mixture distribution of $\theta_n | \mathbf{X}_{n-1}$ (Appendix 1) is the predictive distribution of the next observation $X_n$ from the accumulated history up to instant $n-1$. This is:

$$\begin{aligned}
X_n|\mathbf{X}_{n-1} \sim \quad & p \;\; \alpha_1^{(n-1)} N(\hat{\theta}_{n-1}, \;\; \tau 2 + \sigma 2 + \hat{\sigma}^2_{n-1}) \\
& + (1-p)\alpha_1^{(n-1)} N(\hat{\theta}_{n-1} + \delta, \;\; \tau 2 + \sigma 2 + \hat{\sigma}^2_{n-1}) \\
& + \quad p \;\; \alpha_2^{(n-1)} N(\hat{\theta}_{n-1} + K_{n-1}\delta, \;\; \tau 2 + \sigma 2 + \hat{\sigma}^2_{n-1}) \\
& \quad \vdots \\
& + \quad p \;\; \alpha_n^{(n-1)} N(\hat{\theta}_{n-1} + K_{n-1} \ldots K_1 \delta, \;\; \tau 2 + \sigma 2 + \hat{\sigma}^2_{n-1}) \\
& + (1-p)\alpha_2^{(n-1)} N(\hat{\theta}_{n-1} + \delta + K_{n-1}\delta, \;\; \tau 2 + \sigma 2 + \hat{\sigma}^2_{n-1}) \\
& \quad \vdots \\
& + (1-p)\alpha_n^{(n-1)} N(\hat{\theta}_{n-1} + \delta + K_{n-1} \ldots K_1 \delta, \;\; \tau 2 + \sigma 2 + \hat{\sigma}^2_{n-1}) \\
& \quad \vdots \\
& + \quad p \;\; \alpha_{2^{n-1}}^{(n-1)} N(\hat{\theta}_{n-1} + K_{n-1}\delta + \cdots + K_{n-1} \ldots K_1 \delta, \;\; \tau 2 + \sigma 2 + \hat{\sigma}^2_{n-1})
\end{aligned}$$

$$+(1-p)\alpha_{2^{n-1}}^{(n-1)} N(\hat{\theta}_{n-1} + \delta + K_{n-1}\delta + \cdots + K_{n-1}\ldots K_1\delta, \;\; \tau 2 + \sigma 2 + \hat{\sigma}_{n-1}^2)$$

## References


[1] Bothe, D., R., (1989), "A Powerful New Control Chart for Job Shops", *43rd Annual Quality Congress Transactions*, 265-270.

[2] Burr, J., T., (1989), "SPC in the Short Run" *43rd Annual Quality Congress Transactions*, 776-780.

[3] Casella, G., and Berger, R., L., (1990), *Statistical Inference*, Wadsworth & Brooks-Cole, Pacific Grove, California.

[4] Chang, J., T., and Fricker R., D., (1999) "Detecting When a Monotonically Increasing Mean Has Crossed a Threshold" *Journal of Quality Technology*, 31, 217-234.

[5] Chernoff, H., and Zacks, S., (1964), "Estimating the current mean of a Normal distribution which is subjected to changes in time", *The Annals of Mathematical Statistics*, 40, 999-1018.

[6] Del Castillo, E., and Montgomery, D., C., (1992) "Average Run Length Properties of Q-Charts for Short Run Statistical Control: Enhancements and Alternative Method", *Quality and Reliability Group, Arizona State University*, Publication Number 92-30.

[7] Geisser, S., (1993), *Predictive Inference: An Introduction*, Chapman & Hall, London.

[8] Jeffreys, H., (1948), *Theory of Probability*, Second Edition, University Press, Oxford.

[9] Kalman, R., E., (1960), "A new approach to linear filtering and prediction problems", *Journal of Basic Engineering*, 82, 35-45.

[10] Kress, R., (1998), *Numerical Analysis*, Springer, New York.

[11] Montgomery, C., D., (1997), *Introduction to Statistical Quality Control*, Wiley, New York.

[12] Montgomery, C., D., and Mastrangelo, C., M., (1991), "Some Statistical Process Control Methods for Autocorrelated Data (with discussion)", *Journal of Quality Technology*, 23, 179-204.

[13] Meinhold, R., J., and Singpurwalla, N., D., (1983), "Understanding the Kalman Filter", *The American Statistician*, 37, 123-127.

[14] Quesenberry, C., P., (1990), "SPC Variables Q-Charts for Short Runs", *Quality Concepts '90 Conference proceedings*, 223-236.

[15] Quesenberry, C., P., (1991), "SPC Q Charts for Start-Up Processes and Short or Long Runs", *Journal of Quality Technology*, 23, 213-224.

[16] Roberts, S., W., (1959), "Control Chart Tests Based on Geometric Moving Average", *Technometrics*, 1, 239-250.

[17] Tsiamyrtzis, P., (2000), "A Bayesian Approach to Quality Control Problems", *Ph.D. Dissertation*, School of Statistics, University of Minnesota.

[18] Wasserman, G., S., and Sudjianto, A., (1993), "Short Run SPC Based Upon the Second Order Dynamic Linear Model for Trend Detection", *Communication in Statistics A*, 22, 1011-1036.

[19] West, M., and Harrison, J., (1996), *Bayesian Forecasting and Dynamic Models*, Second Edition, Springer, New York.

[20] Wheeler, D., J., (1992), *Short Run SPC*, SPC Press, Knoxville.

[21] Woodward, P., W., and Naylor, J., C., (1993), "An application of Bayesian methods in SPC", *The Statistician*, 42, 461-469.

[22] Wright, C., M., Booth, D., E. and Hu, M., Y., (2001), "Joint Estimation: SPC Method for Short-Run Autocorellated Data", *Journal of Quality Technology*, 33, 365-378.

# TABLES

Table 1: The $2^n$ different values of the function $f(K_i, \delta, \jmath, \ell)$ at time $n$ of the original model

| $\jmath$ | $\ell$ | $f(K_i, \delta, \jmath, \ell)$ |
|---|---|---|
| 0 | $\emptyset$ | 0 |
| | $\{n\}$ | $K_n\delta$ |
| | $\{n-1\}$ | $K_nK_{n-1}\delta$ |
| 1 | $\vdots$ | $\vdots$ |
| | $\{2\}$ | $K_n \ldots K_2\delta$ |
| | $\{1\}$ | $K_n \ldots K_1\delta$ |
| | $\{n, n-1\}$ | $K_n\delta + K_nK_{n-1}\delta$ |
| | $\vdots$ | $\vdots$ |
| 2 | $\{n, 1\}$ | $K_n\delta + K_n \ldots K_1\delta$ |
| | $\vdots$ | $\vdots$ |
| | $\{2,1\}$ | $K_n \ldots K_2\delta + K_n \ldots K_1\delta$ |
| $\vdots$ | $\vdots$ | $\vdots$ |
| $n$ | $\{1,2,\ldots,n\}$ | $K_n\delta + K_nK_{n-1}\delta + \cdots + K_n \ldots K_1\delta$ |

Table 2: The 10 consecutive measurements of the cholesterol level (in mg/dL)

| Time | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Chol. Level | 144 | 146 | 148 | 147 | 146 | 147 | 147 | 146 | 149 | 151 |

Table 3: The posterior probabilities that $P(\theta_i \leq M)$ at the end of the $i$th stage

| Stage $i$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $P(\theta_i \leq M)$ | .999 | .993 | .919 | .948 | .983 | .962 | .956 | .984 | .812 | .397 |