International Journal of Production Research
Vol. 48, No. 5, 1 March 2010, 1381–1394

Taylor & Francis
Taylor & Francis Group

Recursive parameter estimation for categorical process control

Kaibo Wang[a] and Fugee Tsung[b]*

[a]*Department of Industrial Engineering, Tsinghua University, Beijing, P. R. China;* [b]*Department of Industrial Engineering and Logistics Management, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong*

(*Received 10 February 2008; final version received 13 October 2008*)

Statistical process adjustment (SPA) is utilised prevalently in novel manufacturing scenarios. When quality characteristics rather than internal process variables are inspected for the purpose of quality control, data with different resolutions may be collected. This paper proposes a Bayesian framework for parameter estimation when only categorical observations are available. The proposed method incorporates categorical information recursively and updates parameter estimates in real time. Simulation results show that the framework is effective in utilising low-resolution information in parameter estimation, model building and process control.

Keywords: categorical observations; statistical process adjustment; statistical quality control

1. Introduction

Timely and accurate measurements are important in most engineering scenarios for process control, quality evaluation and production planning. However, with the fast development of new technologies, manufacturing processes that involve the synthesis of nanocomposites and the fabrication of nanodevices are continuously introducing new challenges to traditional quality control practices. Among others, low-resolution information collected on a categorical scale is frequently seen in these processes; the efficient use of such information for quality control has become an important topic.

There are two possible reasons for categorical observations to be generated in engineering processes. First, timely and accurate numerical measurements are sometimes too costly to be obtained; second, some quality characteristics cannot inherently be measured on a numerical scale. Spanos and Chen (1997) presented an example in which quality characteristics are measured on a discrete scale. Based on the roughness of etched sidewalls, wafers are classified by trained operators into categories such as 'very rough', 'rough', 'smooth' and 'very smooth'. Fasulo *et al*. (2004) studied an extrusion process of the thermoplastic olefin (TPO) nanocomposites. Surfaces generated from each formulation were first checked via *visual ranking*. Based on the surface quality, all panels were ranked from 1 to 5 with 1 being the best and 5 the worst. After that, micrographs of the surfaces were taken with the aid of special equipment for more accurate surface measures. Brondino *et al*. (2006) presented examples that generate multivariate ordinal variables

*Corresponding author. Email: season@ust.hk

ISSN 0020–7543 print/ISSN 1366–588X online

DOI: 10.1080/00207540802570651
http://www.informaworld.com

from both manufacturing processes and social sciences. Wang and Tsung (2007) studied a deep reactive ion etching (DRIE) process, in which categorical observations were collected and used for process adjustment.

In novel manufacturing environments, statistical process adjustment (SPA) is an important way to improve quality and production efficiency and to reduce defects. The purpose of SPA is to model, forecast and control a dynamic process using a set of statistical techniques (Colosimo *et al.* 2005, Del Castillo 2006). Modelling is a vital step, which usually includes model building and parameter estimation. Compared with the automatic process control methods developed in traditional engineering fields, SPA has two distinguishing features. First, SPA is not necessarily implemented fully automatically. Rather, human operators or other components may be included in an SPA loop. Therefore, SPA plays a supervisory role in guiding lower-level production systems. Second, SPA can take quality characteristics as process responses to adjust the processes. As quality characteristics are usually measured in the customer's domain, such as fail or pass, some new data types that are not seen when measuring internal process variables are included. As the foregoing examples show, categorical observations collected on different scales are available for quality control.

However, in order to maintain product quality, most algorithms for modelling and controlling are built on the basis of numerical observations. For example, the most widely used controllers in run-to-run (R2R) processes are the exponentially weighted moving average (EWMA) controllers (Ingolfsson and Sachs 1993) and various extensions of it, such as the double EWMA controller (Butler and Stefani 1994, Del Castillo 1999), triple EWMA controller (Fan *et al.* 2002), variable EWMA controller (Tseng *et al.* 2003) and the self-tuning controller (Del Castillo and Hurwitz 1997, Jen *et al.* 2004). Even though these controllers are designed for compensating different types of noise signals, they all share a common feature. All feedback controllers are model-based and require numerical measurements from process outputs to be known. This simply implies that in order to implement such controllers, two requirements must be fulfilled: one, the measuring procedure must be finished quickly so that a new control action can be calculated by the controller without inferring continuous production; two, quality features of finished products must be able to be measured on a continuous scale. When such assumptions are violated, as in the aforementioned examples, we have to design new algorithms for model building, parameter estimation and process control.

In this study, we take the DRIE process introduced by Wang and Tsung (2007) as an example and propose a new approach for process parameter estimation using categorical observations. Originally developed for micro-electro-mechanical systems (MEMS), DRIE is a process that involves complex chemical-mechanical reactions. As a typical R2R process widely used in semiconductor and nanomanufacturing, DRIE has been successfully utilised in producing photonic crystals, magnetic nanostructures, MEMS resonators (STS 2006), and high aspect ratio ($>50:1$) silicon pillar arrays (Chan *et al.* 2006).

In the DRIE process, the maintenance and control of trench profiles, which is a key quality characteristic, is a challenging problem. Based on its etched profile, each wafer can be classified into different categories labelled as 'negative', 'normal' or 'positive', indicating that the wafers are over-etched, normally etched or under-etched, respectively. In addition, some irregular profiles may be produced in practice, as illustrated in Figure 1. Wafers with irregular profiles can only be judged as over-etched or under-etched and therefore be classified as negative or positive on a categorical scale.

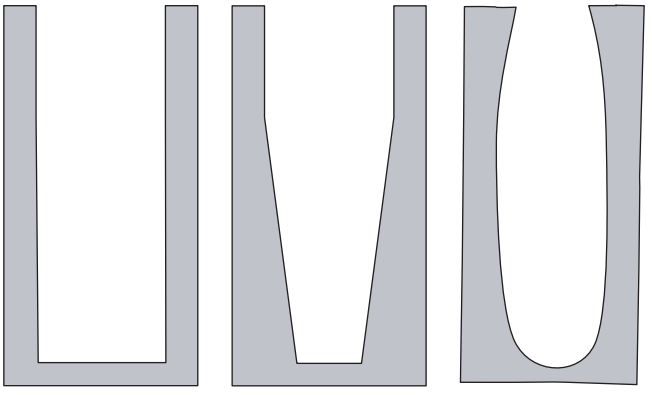

Figure 1. Illustrations of regular and irregular etching profiles from a DRIE process.

The shape of the etching profile is largely dominated by the etch/deposition time ratio parameter of a process, which is a controllable factor during production. In order to achieve an ideal profile, the etch-to-deposition time ratio has to be adjusted at the end of each run to compensate for noise and aging effects. From these numerical input and categorical output, this paper focuses only on the modelling issue of the process. Parameters involved in illustrating this process are estimated using categorical observations.

Categorical variables, which by convention are called linguistic variables, are also found in fuzzy control theory (Yu *et al*. 2003). A fuzzy logic controller converts numerical observations to linguistic variables by referring to specific membership functions. The linguistic variables are then fed into the rule base of the controller to suggest fuzzy control actions. A crisp control action is finally generated by combining the fuzzy actions based on certain decision-making logic. However, a fuzzy logic controller requires the determination of an arbitrary membership function; the controller may assign each observation into more than one linguistic set based on the membership functions, which differ from the mutually exclusive categories considered in this paper. The philosophical and practical differences between fuzzy control theory and statistical methods are discussed by Laviolette *et al*. (1995). Wang and Tsung (2007) proposed a two-phase R2R categorical controller that works on timely categorical measurements. However, they estimated process models in the first phase by assuming that numerical values were available, which is a prerequisite for the categorical controller to be built.

This paper proposes a Bayesian framework for parameter estimation based on categorical observations. The rest of this paper is organised as follows. Section 2 presents the model to be studied; Section 3 proposes the Bayesian framework for parameter estimation using categorical observations; Section 4 studies the performance of the proposed method via simulations. Finally, Section 5 concludes the paper with suggestions for future research.

2. Process modelling and parameter definition

An R2R process is typically illustrated by a linear model (see Del Castillo and Hurwitz 1997, and Apley and Kim 2004 and the references therein), as follows:

$$y_t = \alpha + \beta u_{t-1} + \varepsilon_t, \tag{1}$$

where y_t is the key quality characteristic, u_{t-1} is the process input at step $t-1$ -1 and ε_t is a white noise series. Without loss of generality, we assume that $\varepsilon_t \sim N(0, \sigma^2)$.

Del Castillo (2006) discussed the control issue of processes represented by Equation (1). Essentially, if the process is on target and it suffers from white noise only, there is no need for process adjustment. However, if set up bias exists and initial quality characteristics do not meet the target, which happens frequently since the true process parameters are never known, process adjustment will be necessary. In addition, if the process is not stable and suffers from faults such as deterministic drift, process adjustment should also be considered. A more detailed discussion is found in Del Castillo (2006).

Parameter estimation is the basis for process forecasting and control. Therefore, we focus on parameter estimation issues based on categorical observations. In Equation (1), the variable y_t is used to denote the output at step t measured on a numerical scale. However, y_t is not available in the studied applications. Therefore, we call y_t a latent variable. We further define $Y_t = g(y_t)$ as the categorised value of y_t. Y_t is a measurement collected on a discrete scale at step t. The relationship between Y_t and y_t is adequately modelled by the following mapping function:

$$Y_t = j \text{ if } \gamma_{j-1} < y_t < \gamma_j, \quad j = 1, \ldots, J, \tag{2}$$

where J is the total number of categories to which y_t will be assigned and $\gamma_j, j = 0, 1, \ldots, J$, are the cut-off parameters that separate the output space into J intervals. For certain applications without the lowest and highest boundaries, the cut-off parameters, γ_0 and γ_J, may take values of negative and positive infinity. Following the above mapping function, the probability that an observation falls into either category j or a lower category is given by:

$$\Pr(Y_t \le j) = \Pr(y_t \le \gamma_j) = \Pr(\varepsilon_t \le \gamma_j - \alpha - \beta u_{t-1}).$$

Since $\varepsilon_t \sim N(0, \sigma^2)$, $\Pr(Y_t \le j) = \Phi((\gamma_j - \alpha - \beta u_{t-1})/\sigma)$, where $\Phi(\cdot)$ is the cumulative density function of a standard normal distribution. Applying the probit link function to the above equation and making a simple transformation yields:

$$\sigma \Phi^{-1}(\Pr(Y_t \le j)) = \gamma_j - \alpha - \beta u_{t-1}, \quad j = 1, \ldots, J,$$

which shows a linear relationship between the transformed probability and the process parameters.

Let $\Theta = \{\alpha, \beta\}$ be a set that contains unknown process parameters. In the rest of this paper, we present a Bayesian framework to estimate Θ. The vector $\boldsymbol{\gamma} = [\gamma_0, \gamma_1, \ldots, \gamma_J]^T$, which is in fact the hidden rules used by operators who inspect and classify products, is assumed to be known. The estimation of $\boldsymbol{\gamma}$ using categorical observations deserves separate analysis. (Interested readers are referred to Wang and Tsung (2007) for situations when numerical observations are available during the experimental stage.)

3. A Bayesian framework for parameter estimation using categorical observations

In the following, we will study a recursive Bayesian framework for estimating and updating Θ. Different from the usual estimation procedure that assumes that all observations are obtained at once, the recursive method assumes that observations are collected one by one, which fits the real situation of engineering processes.

Whenever a new sample is collected and measured on a categorical scale, this measurement will be utilised to update previous estimates of the parameters. As the process evolves, the parameters are expected to approach their respective true values.

To implement the above framework, we first derive the joint posterior distributions of unknown parameters. Prior distributions that appear in the joint posterior distributions are then elicited. After that, the full conditional distributions of all parameters are derived, which are used for recursive parameter estimation using Gibbs sampling.

3.1 *Posterior distributions*

As all numerical variables, $y_1, \ldots, y_t$, are not observable, we treat them as latent variables that need to be estimated. According to the Bayes' theorem, given all categorical observations up to step t, the joint posterior distribution of unknown process parameters, Θ, and latent variables, $y_1, \ldots, y_t$, is given by:

$$f(y_t, \ldots, y_1, \Theta | Y_t, \ldots, Y_1) \propto f(\Theta) f(y_t, \ldots, y_1, Y_t, \ldots, Y_1 | \Theta). \tag{3}$$

The second component in Equation (3) represents the likelihood of all categorical and continuous data up to time t, which can be further factorised as:

$$\begin{aligned} f(y_t, \ldots, y_1, Y_t, \ldots, Y_1 | \Theta) &= f(y_1, Y_1 | \Theta) \\ &\quad \times f(y_2, Y_2 | y_1, Y_1, \Theta) \cdots f(y_t, Y_t | y_{t-1}, Y_{t-1}, \ldots, y_1, Y_1, \Theta). \end{aligned}$$

For each $k \in [1, t]$, the joint distribution can be further decomposed as:

$$\begin{aligned} f(y_k, Y_k | y_{k-1}, Y_{k-1}, \ldots, y_1, Y_1, \Theta) &= f(y_k | y_{k-1}, Y_{k-1}, \ldots, y_1, Y_1, \Theta) \\ &\quad \times f(Y_k | y_k, y_{k-1}, Y_{k-1}, \ldots, y_1, Y_1, \Theta). \end{aligned}$$

As the distribution of y_k is solely determined by $y_{k-1}, y_{k-2}, \ldots, y_1$ and Θ, it follows that:

$$f(y_k | y_{k-1}, Y_{k-1}, \ldots, y_1, Y_1, \Theta) = f(y_k | y_{k-1}, \ldots, y_1, \Theta) = N(y_k | \mu_k, \sigma^2), \tag{4}$$

where

$$\mu_k = \alpha + \beta u_{k-1}.$$

Note that u_{k-1} is a constant and μ_k is a function of α and β.

Once the distributions of the latent variables are determined, the density function of Y_k takes one over the interval in which the latent variable falls and takes zero otherwise. That is,

$$f(Y_k | y_k, y_{k-1}, Y_{k-1}, \ldots, y_1, Y_1, \Theta) = f(Y_k | y_k) \propto I_{\gamma_{Y_k - 1} \le y_k < \gamma_{Y_k}}, \tag{5}$$

where

$$I_{\gamma_{Y_k - 1} \le y_k < \gamma_{Y_k}} = \begin{cases} 1 & \text{if} \gamma_{Y_k - 1} \le y_k < \gamma_{Y_k} \\ 0, & \text{otherwise.} \end{cases}$$

Putting the above derivations together, we obtain the posterior distribution of unknown parameters given all categorical observations collected up to step t:

$$f(y_t, \ldots, y_1, \Theta | Y_t, \ldots, Y_1) \propto f(\Theta) \prod_{k=1}^{t} N(y_t | \mu_k, \sigma^2) I_{\gamma_{Y_k - 1} \le y_k < \gamma_{Y_k}}. \tag{6}$$

3.2 *Prior elicitation*

In order to obtain appropriate prior distributions for the unknown parameters Θ, we apply a reasonable assumption that both α and β follow normal distributions and they are independent, which implies that:

$$\begin{cases} f(\Theta) = f(\alpha) f(\beta) \\ f(\alpha) = N(\alpha | \alpha_0, \sigma^2_{\alpha,0}) \\ f(\beta) = N(\beta | \beta_0, \sigma^2_{\beta,0}). \end{cases} \tag{7}$$

By substituting Equation (7) for $f(\Theta)$ in Equation (6), the joint posterior distribution becomes:

$$\begin{aligned} f(y_t, \ldots, y_1, \Theta | Y_t, \ldots, Y_1) \propto\ & N(\alpha | \alpha_0, \sigma^2_{\alpha,0}) N(\beta | \beta_0, \sigma^2_{\beta,0}) \\ & \times \prod_{k=1}^{t} N(y_k | \mu_k, \sigma^2) I_{\gamma_{Y_k - 1} \le y_k < \gamma_{Y_k}}. \end{aligned} \tag{8}$$

The joint distribution is truncated so that the intervals to categorise continuous variables are properly defined.

3.3 *Fully-conditional distributions*

The fully-conditional distribution of each parameter is defined as the distribution when all other parameters are known. In Gibbs sampling, based on samples drawn from the fully-conditional distributions, the marginal distribution of each parameter can be obtained. In this section, we derive the fully-conditional distribution for unknown parameters and use Gibbs sampling to derive the marginal distributions of all parameters.

It should be noted that one fundamental difference between Bayesian process control and Bayesian data analysis is the availability of information. In ordinary data analysis, information is given all at once with a limited number of observations; in process control, a process generates data continuously. As a result, the dimension of the available data is not a constant but a variable that continuously increases. Therefore, in this study, we develop a recursive method to make use of information in the model estimation. This recursive method fits the philosophy of Bayesian analysis. At step t, parameter estimates obtained based on previous $t-1$ observations are treated as the prior for step t. After that, new information collected from the tth step is incorporated into the model by updating the priors and generating a new prior for step $t+1$.

We start by investigating the fully-conditional distribution of y_t after Y_t is observed. As the process is subject to white noise, the distribution of y_t depends on knowledge

collected up to time t only. Therefore,

$$f(y_t|\alpha, \beta, y_{t-1}, \ldots, y_1, Y_t, \ldots, Y_1) \propto N(y_t|\mu_t, \sigma^2) I_{\gamma_{Y_{t-1}} \le y_t < \gamma_{Y_t}}, \tag{9}$$

which tends to be a normal distribution truncated by the boundaries of the category that the observation falls into, as shown in Figure 2. The mean of the truncated distribution, denoted by m_t, is different from the non-truncated mean, μ_t.

It is not difficult to show that m_t satisfies:

$$m_t = \mu_t + \frac{\sigma(\varphi(b_1) - \varphi(b_2))}{\Phi(b_2) - \Phi(b_1)},$$

where $b_1 = (\gamma_{Y_t-1} - \mu_t)/\sigma$, $b_2 = (\gamma_{Y_t} - \mu_t)/\sigma$, and φ is the probability density function of the standard normal distribution and Φ is the cumulative density function of the standard normal distribution.

Based on the joint posterior distribution in Equation (8) and Bayes' theorem, the fully-conditional distribution of α can be written as:

$$f(\alpha|\beta, y_t, \ldots, y_1, Y_t, \ldots, Y_1) \propto f(\alpha, \beta, y_t, \ldots, y_1|Y_t, \ldots, Y_1). \tag{10}$$

We treat Equation (10) as a function of α and move all parameters that are independent of α to a constant term to normalise the density function. The density function of α then reduces to:

$$f(\alpha^{(t)}|\cdot) \propto N\left(\alpha|\alpha_0, \sigma^2_{\alpha,0}\right) \prod_{k=1}^{t-1} N(y_k|\mu_k, \sigma^2) \times N(y_t|\mu_t, \sigma^2), \tag{11}$$

where the superscript (t) indicates that the current estimate is obtained based on historical observations up to step t. For simplicity, we use $f(\alpha^{(t)}|\cdot)$ to represent the fully-conditional density of $\alpha^{(t)}$. Equation (11) establishes that the estimate of α at step t can be expressed recursively based on the estimate obtained at step $(t-1)$, which implies:

$$f(\alpha^{(t)}|\cdot) \propto f(\alpha^{(t-1)}|\beta, \phi, y_{t-1}, \ldots, y_1, Y_{t-1}, \ldots, Y_1) \times N(y_t|\mu_t, \sigma^2), \tag{12}$$

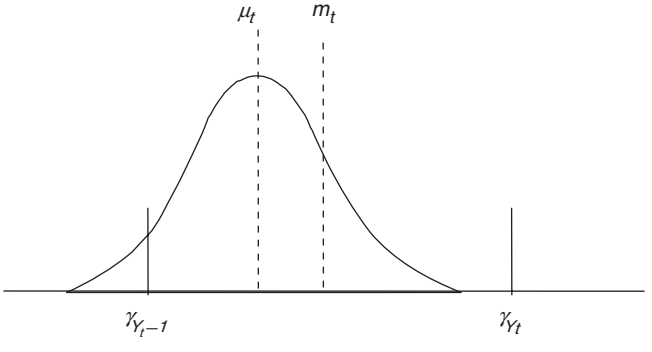

Figure 2. Truncated normal distributions.

The last term in Equation (12), $N(y_t|\mu_t,\sigma^2)$, is the likelihood function of observation y_t, which can be transformed to a function of α:

$$\begin{aligned} N(y_t|\mu_t,\sigma^2) &\propto \exp\left(-\tfrac{(y_t-\mu_t)^2}{2\sigma^2}\right) \\ &\propto \exp\left(-\tfrac{(y_t-(\alpha+\beta u_{t-1}))^2}{2\sigma^2}\right) \\ &\propto \exp\left(-\tfrac{(\alpha-(y_t-\beta u_{t-1}))^2}{2\sigma^2}\right). \end{aligned} \tag{13}$$

Let $\mu_{\alpha,y}=y_t-\beta u_{t-1}$ and $\sigma^2_{\alpha,y}=\sigma^2$. It is not difficult to see that Equation (13) is in fact a normal density function of α with mean $\mu_{\alpha,y}$ and variance $\sigma^2_{\alpha,y}$.

It becomes clear that Equation (11) or, equivalently, Equation (12) is proportional to the product of two normal density functions, $f(\alpha^{(t-1)}|\cdot)$ and $N(y_t|\mu_t,\sigma^2)$. Finally, by denoting the mean and variance of distribution $f(\alpha^{(t)}|\cdot)$ as $\mu_{\alpha,t}$ and $\sigma^2_{\alpha,t}$, the fully-conditional distribution of α obtained at step t can be proved to be a normal distribution with mean:

$$\mu_{\alpha,t}=\left(\frac{\mu_{\alpha,t-1}}{\sigma^2_{\alpha,t-1}}+\frac{\mu_{\alpha,y}}{\sigma^2_{\alpha,y}}\right)\Big/\left(\frac{1}{\sigma^2_{\alpha,t-1}}+\frac{1}{\sigma^2_{\alpha,y}}\right)$$

and variance:

$$\sigma^2_{\alpha,t}=1\Big/\left(\frac{1}{\sigma^2_{\alpha,t-1}}+\frac{1}{\sigma^2_{\alpha,y}}\right).$$

The above analysis shows that when a new observation is sampled, the distribution of α is obtained as the previous estimate calibrated by information conveyed by the latest observation. Therefore, it is feasible for us to implement a recursive algorithm to simplify the calculation while making the estimate of the parameters more accurate.

In a similar fashion, the fully-conditional distribution of β can be obtained as:

$$f(\beta^{(t)}|\cdot)\propto f(\beta^{(t-1)}|\cdot)N(y_t|\mu_t,\sigma^2). \tag{14}$$

Let $\mu_{\beta,y}=(y_t-(1-\phi)\alpha)/u_{t-1}$ and $\sigma^2_{\beta,y}=\sigma^2/|u_{t-1}|$. We can show that the fully-conditional distribution of β is a normal distribution with mean:

$$\mu_{\beta,t}=\left(\frac{\mu_{\beta,t-1}}{\sigma^2_{\beta,t-1}}+\frac{\mu_{\beta,y}}{\sigma^2_{\beta,y}}\right)\Big/\left(\frac{1}{\sigma^2_{\beta,t-1}}+\frac{1}{\sigma^2_{\beta,y}}\right)$$

and variance

$$\sigma^2_{\beta,t}=1\Big/\left(\frac{1}{\sigma^2_{\beta,t-1}}+\frac{1}{\sigma^2_{\beta,y}}\right),$$

where $\mu_{\beta,t-1}$ and $\sigma^2_{\beta,t-1}$ are the estimated mean and variance of β obtained in step $t-1$.

3.4 *Parameter estimation and process control via Gibbs sampling*

Gibbs sampling is a way of estimating marginal distributions of random variables (see Geman and Geman 1984, Colosimo and Del Castillo 2007 for more details). In the

following, a Gibbs sampling procedure for estimating the unknown parameters in Equation (1) is proposed.

When a new categorical observation becomes available, the Gibbs sampling procedure starts sampling y_t, α and β repeatedly until a sufficiently large number is reached. By calculating the sample mean and variance with the initial samples removed, the marginal distributions of α and β can be obtained. The recursive estimation and sampling process is outlined as follows:

Step 1: Sample one y_t from Equation (9).

Step 2: Using the estimate obtained from the previous step as the prior and y_t from Step 1, update the conditional distribution of α in Equation (12).

Step 3: Sample one α from the updated conditional distribution.

Step 4: Using the estimate obtained from the previous step as the prior, y_t from Step 1 and α from Step 3, update the conditional distribution of β in Equation (14).

Step 5: Sample one β from the new conditional distribution.

Step 6: Using the newly sampled α and β, go back to Step 1. Update the conditional distribution of y_t.

Step 7: Repeat Steps 1–6 until a sufficiently large number is reached.

Step 8: Calculate the sample mean and standard deviation of α and β, respectively, which will serve as the prior for the next step; if a feedback controller is in place, use the updated parameters to generate a new recipe; then go ahead to produce a new run, perform an inspection, and repeat Steps 1–8.

The above procedure routinely uses new categorical observations to update previous estimates of the parameters. Therefore, the estimated values are expected to approach their true levels gradually.

In order to control the process against initial bias, we define the objective to minimise the expected one-step-ahead prediction errors given all historical information. Denote the target of process (1) as T_0. At step t, given $Y_t, \ldots, Y_1$, the objective function is given by:

$$E\left[(y_{t+1} - T_0)^2 | Y_t, \ldots, Y_1\right].$$

In this study, as Equation (1) shows, we assume that the process has a white noise disturbance series. Since $y_{t+1} = \alpha + \beta u_t + \varepsilon_t$ and $E(\varepsilon_{t+1}) = 0$, it is not difficult to show that:

$$E\left[(y_{t+1} - T_0)^2 | Y_t, \ldots, Y_1\right] = \beta^2 u_t^2 + (\alpha - T_0)^2 + \sigma^2 + 2(\alpha - T_0)\beta u_t.$$

Taking a partial derivation of the above equation with respect to u_t and equating it to zero yields the optimal control action:

$$u_t = \frac{T_0 - \alpha}{\beta}. \tag{15}$$

As the control algorithm is derived based on Bayesian analysis using categorical information, we name it a Bayesian categorical controller (BCC). It is worth noting that the above control action is analogous to the EWMA controller. However, the EWMA

controller updates α using continuous measurements collected from a process, while the BCC uses categorical information.

It should be noted that if all the process parameters are already known or have been estimated rather accurately, it might not be necessary to continue to update the parameters. A good practice is to use the above procedure to calibrate the initial bias only and turn to more advanced process controllers when the estimated parameters show stability.

4. Simulation study

In this section, we investigate the performance of the proposed estimation procedure using a numerical example. In each simulation, we assume that the process follows Equation (1) with $\alpha = 80.0$, $\beta = 1.0$ and a standard deviation $\sigma = 0.3$. The target value of the process is, $T = 90.0$. Two cut-off parameters, 89 and 91, are set up to classify the process outputs into three categories. The Gibbs sampling set is repeated 110,000 times whenever a new observation is collected. The last 100,000 samples are used to calculate the marginal distributions of the unknown parameters.

In Bayesian analysis, there are many ways to obtain priors for unknown parameters. As we want to investigate the issues caused by initial bias, we assume that the prior mean of α and β are 75.0 and 0.8, respectively, and their variance is 1.0. The difference in these values indicates the existence of initial bias. The BCC controller is set up for initial process adjustment.

In the simulation procedure illustrated above, the purpose of Step 1 is to sample a numerical y when a categorical Y is available. The samples are taken from a truncated normal distribution. This is a critical step that links categorical variables with numerical variables. Figure 3 shows two cases during the Gibbs sampling when the observation falls between a deviant category and the target category. By taking sufficiently large samples, the truncated distribution is reconstructed well.

Figure 4 shows the trajectories of the mean and variance of estimated α and β. It is clearly seen from Figures 4(a) and 4(b) that the initial values have deviated from their respective true levels. When categorical observations are used in the parameter estimation, the estimated parameters approach their true value gradually. In the early stage, oscillation may exist since the information contained in the categorical variables may not be accurate. After around 20 steps, however, the estimates are already rather close to their true values. Figures 4(c) and 4(d) suggest that the variances of the estimated parameters decrease as more observations are accumulated. This trend also shows that the recursive algorithm is effective in making use of categorical observations for parameter estimation.

The sequence of controlled output, y_t, is shown in Figure 5. As is seen, the initial biased estimates of unknown parameters lead to extreme output values. However, as the estimated parameters approach their true values quickly, the process is maintained on target after a short period of time.

As discussed in Section 2, a process that is influenced by a white noise series should not be adjusted by any feedback controller. However, if initial bias exists, the process will stay far from its target forever. In the simulation studies, we equip the process with a simple controller to calibrate its initial bias. After the transient stage, unless sudden parameter changes may happen, the controller could be removed. Alternatively, the controller could

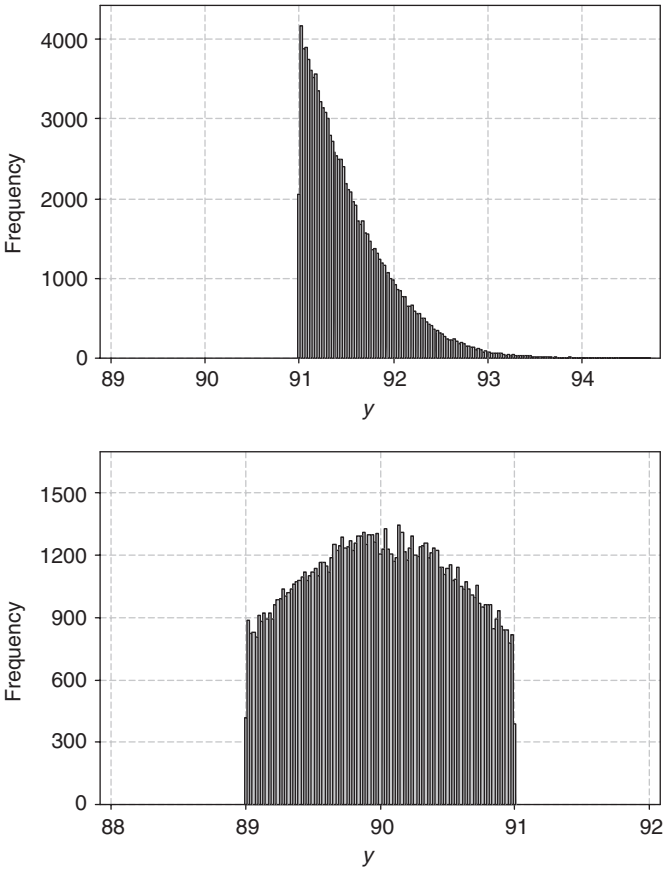

Figure 3. Samples drawn from truncated normal distributions during Gibbs sampling.

be set to work only if the output deviates far from the target value, which serves as an indication of a sudden process shift. This treatment resembles the bounded controller (Del Castillo and Hurwitz 1997). We do not discuss this issue in depth here but leave it as a topic for future research.

5. Conclusions and future research

In novel manufacturing scenarios, it is not uncommon that information with different resolutions is presented. This paper investigated the estimation of process parameters when only categorical observations are available. A Bayesian framework was proposed to estimate the unknown parameters efficiently. This framework takes categorical observations sequentially when they become available and updates the estimates recursively, which is a feasible and efficient way in engineering control settings. Simulation studies showed that even when based on categorical observations, the proposed method can estimate the unknown parameters rather accurately and quickly.

Information with different resolutions is frequently encountered in modern manufacturing scenarios. Conventional quality control theories have to be updated to fit in the

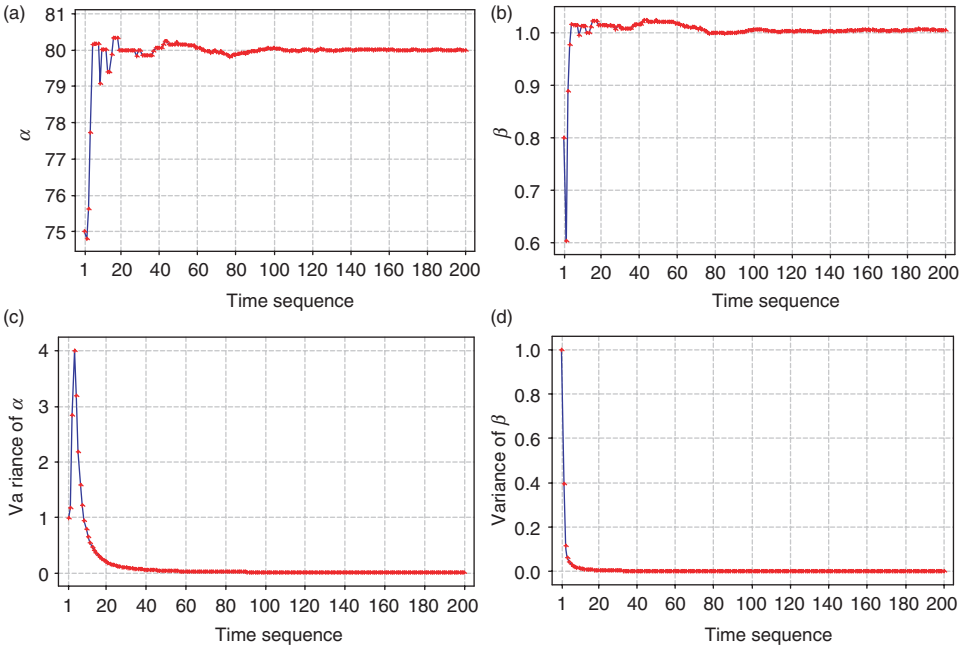

Figure 4. Trajectories of estimated parameters: (a) the mean of α; (b) the mean of β; (c) the variance of α; and (d) the variance of β.

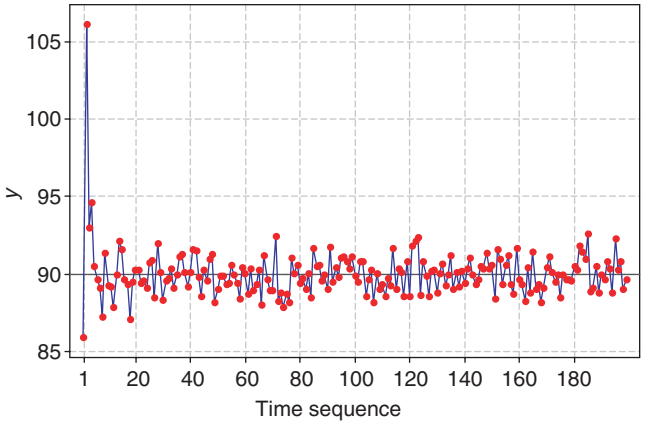

Figure 5. Trajectories of process output.

new settings. This study focused on the estimation issue. When a more complicated disturbance series that follows an IMA(1, 1) or a general ARIMA time series model is presented, more parameters will be involved and the recursive Bayesian framework should be extended. Furthermore, deployment of more advanced adjustment algorithms, including a bounded controller based on categorical or mixed-resolution observations, is also an interesting topic for future research.

In this research, we assumed all cut-off parameters were fixed. While in practice, especially when human factors are involved, misclassification may exist.

Measurement system analysis for attribute data and its integration with process control is another interesting topic for future research.

Acknowledgements

The authors are grateful to the editor and the referees for their valuable comments. Dr Wang's work was supported by the National Natural Science Foundation of China (NSFC) under grant No. 70802034. Professor Tsung's work was supported by the RGC Competitive Earmarked Research Grants 620606 and 620707.

References

Apley, D.W. and Kim, J., 2004. Cautious control of industrial process variability with uncertain input and disturbance model parameters. *Technometrics*, 46 (2), 188–199.

Butler, S.W. and Stefani, J.A., 1994. Supervisory run-to-run control of polysilicon gate etch using *in-situ* ellipsometry. *IEEE Transactions on Semiconductor Manufacturing*, 7 (2), 193–201.

Brondino, G., *et al*., 2006. Synthesis maps for multivariate ordinal variables in manufacturing. *International Journal of Production Research*, 44 (20), 4241–4255.

Chan, Y.C., Lee, Y.K., and Zohar, Y., 2006. High-throughput design and fabrication of an integrated microsystem with high aspect-ratio sub-micron pillar arrays for free-solution micro capillary electrophoresis. *Journal of Micromechanics and Microengineering*, 16 (4), 699–707.

Colosimo, B.M., Pan, R., and Del Castillo, E., 2005. Setup adjustment for discrete-part manufacturing processes with asymmetric cost functions. *International Journal of Production Research*, 43 (18), 3837–3854.

Colosimo, B.M. and Del Castillo, E., 2007. *Bayesian process monitoring, control and optimization*. Boca Raton, FL: Chapman and Hall/CRC.

Del Castillo, E. and Hurwitz, A.M., 1997. Run-to-run process control: literature review and extensions. *Journal of Quality Technology*, 29 (2), 184–196.

Del Castillo, E., 1999. Long run and transient analysis of a double EWMA feedback controller. *IIE Transactions*, 31 (12), 1157–1169.

Del Castillo, E., 2006. Statistical process adjustment: a brief retrospective, current status, and some opportunities for further work. *Statistica Neerlandica*, 60 (3), 309–326.

Fan, S.K.S., *et al*., 2002. SISO run-to-run feedback controller using triple EWMA smoothing for semiconductor manufacturing processes. *International Journal of Production Research*, 40 (13), 3093–3120.

Fasulo, P.D., *et al*., 2004. Extrusion processing of TPO nanocomposites. *Polymer Engineering and Science*, 44 (6), 1036–1045.

Geman, S. and Geman, D., 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6 (6), 721–741.

Ingolfsson, A. and Sachs, E., 1993. Stability and sensitivity of an EWMA controller. *Journal of Quality Technology*, 25 (4), 271–287.

Jen, C.H., Jiang, B.C., and Fan, S.K.S., 2004. General run-to-run (R2R) control framework using self-tuning control for multiple-input multiple-output (MIMO) processes. *International Journal of Production Research*, 42 (20), 4249–4270.

Laviolette, M., *et al*., 1995. A probabilistic and statistical view of fuzzy methods. *Technometrics*, 37 (3), 249–261.

Spanos, C.J. and Chen, R.L., 1997. Using qualitative observations for process tuning and control. *IEEE Transactions on Semiconductor Manufacturing*, 10 (2), 307–316.

STS, 2006. STS in Nanotechnology. STS Innovations.

Tseng, S.T., *et al*., 2003. A study of variable EWMA controller. *IEEE Transactions on Semiconductor Manufacturing*, 16 (4), 633–643.

Wang, K. and Tsung, F., 2007. Run-to-run process adjustment using categorical observations. *Journal of Quality Technology*, 39 (4), 312–325.

Yu, F.J., Low, C.Y., and Cheng, S.S., 2003. Design for an SPRT control scheme based on linguistic data. *International Journal of Production Research*, 41 (6), 1299–1309.