RC23383 (W0410-125) October 20, 2004
Mathematics

IBM Research Report

On Detection of Changes in Count Data

Emmanuel Yashchin
IBM Research Division
Thomas J. Watson Research Center
P.O. Box 218
Yorktown Heights, NY 10598

Research Division
Almaden - Austin - Beijing - Haifa - India - T. J. Watson - Tokyo - Zurich

On Detection of Changes in Count Data

Emmanuel Yashchin

IBM Research Division, Mathematical Sciences Department, Thomas J. Watson Research Ctr.,
P.O. Box 218, Yorktown Heights, NY 10598, USA

Abstract

The problem of monitoring categorical data is encountered in a wide range of practical settings. For example, in the field of manufacturing, problems of this type include monitoring defect rates, levels of contamination or data integrity. This article discusses an approach to monitoring in situations where the underlying parameters of the categorical data are subject to abrupt changes of unpredictable magnitude at some unknown points in time. We derive detection schemes based on the Likelihood Ratio Approach and discuss their performance and issues related to their design and analysis. The paper also discusses problems related to on-line estimation of the parameters in the presence of abrupt changes. It contains several examples that illustrate use of the proposed techniques in semiconductor industry.

Key words and phrases: Cusum, Detection, Likelihood Ratio, Run Length

1. INTRODUCTION

A large number of statistical applications in industrial setting is related to problems involving data streams subject to abrupt changes in time, such as shifts or drifts. Typical problems of this kind are related to detection of unfavorable changes (data monitoring and control charting), estimation of the current process parameters (control and filtering), identifying points of change and regimes (forecasting, process diagnostics) and tests for data homogeneity (modelling, process capability analysis). These problems have been extensively discussed in the literature (eg., see Telksnys (1986), Basseville and Nikiforov (1993), Hawkins and Olwell (1998). Most of the literature on this subject focuses on variable data and attributes data. The early literature (eg. see Shewhart (1931)) emphasized computationally simple ad-hoc techniques; the more recent literature, however, takes into consideration recent advances in computing technology and discusses a number of methods that offer clear advantages in terms of statistical power. For problems involving count data, a number of new methods have been discussed in recent literature, e.g., see Michelson (1996), Woodall (1997), Borror, Champ and Rigdon (1998), Skinner, Montgomery and Runger (2003); several practical guides on working with count data are also available, e.g., see Bzik (1997).

One way to obtain powerful monitoring procedures is based on the concept of the Likelihood Ratio (LR). In this article we discuss several LR based methods in conjunction with the problem of monitoring categorical data. The presentation will center around the problem of detection; however, the reader familiar with the literature related to use of LR techniques will be able to take advantage of the proposed methods in other problems involving serial count data.

Section 2 discusses some basic concepts related to the LR technique. In Section 3 we introduce the basic model and its parameters. We also introduce the parameters that a practitioner would

typically want to monitor in this model. In Section 4 we apply the LR approach do derive control schemes for the parameters of interest. In Section 5 we discuss an example.

2. THE PROBLEM OF DETECTION

Let $\{\boldsymbol{X}_t\}$, $t = 1, 2, \ldots$ be a sequence of (generally multivariate) observations that may represent, for example, proportions of defective items due to various causes or counts of contaminating particles of various types observed in successive periods of time. The stochastic behavior of the sequence is determined by the vector of parameters $\boldsymbol{\theta}$. To simplify the presentation, we assume that all the components of this vector are of primary interest (i.e., no nuisance parameters) and that the observations $\{\boldsymbol{X}_t\}$ are independent. Denote the most recent moment of time by T and the corresponding most recently observed observation by $\boldsymbol{X}_T$. Denote the joint distribution of m most recent observations by $f_{\boldsymbol{\theta}}(\boldsymbol{X}_{T-m+1}, \ldots, \boldsymbol{X}_{T-1}, \boldsymbol{X}_T)$, and denote its natural logarithm (the log-likelihood) by $L_m(\boldsymbol{\theta})$.

To set up the problem of detection, one should first specify the acceptable region Ω_0 in which $\boldsymbol{\theta}$ should reside under normal operating conditions and the unacceptable region, Ω_1. Note that the union of Ω_0 and Ω_0 does not need, in general, to cover the whole parameter space: there will generally exist a "grey" area in between. This three-zone approach is motivated by practical convenience: in many industrial applications an engineer will have no difficulty specifying areas that are distinctly "good" or "bad"; however, dividing the parameter space into two regions to separate "good" values from "bad" ones could prove to be a challenge.

Performance of detection schemes is typically measured in terms of the Run Length (RL), a random variable representing the number of observations taken until a signal is triggered. In general, one would like this variable to be large when $\boldsymbol{\theta} \in \Omega_0$ (i.e., a low false alarm rate) and small when $\boldsymbol{\theta} \in \Omega_1$ (i.e. good sensitivity with respect to out-of-control conditions). The most popular measure is the Average Run Length (ARL); however, design of control schemes based on quantiles and probabilities of RL has also been developed.

Now let us denote, for a given set of last m observations,

$$L^*_{m0} = \max_{\boldsymbol{\theta}\in\Omega_0} L_m(\boldsymbol{\theta}), \qquad L^*_{m1} = \max_{\boldsymbol{\theta}\in\Omega_1} L_m(\boldsymbol{\theta}), \qquad D^*_m = L^*_{m1} - L^*_{m0}. \tag{2.1}$$

Denote by m_0 the minimal depth of data m for which $\boldsymbol{\theta}$ is estimable. Then one can define a general strategy that leads to powerful control schemes as follows.

Likelihood Ratio (LR) Strategy: Trigger an out of control signal at time T if $D^*_m > h$ for some $m \geq m_0$ and pre-specified threshold h.

The LR strategy leads to powerful procedures for a wide class of situations involving control of univariate and multivariate processes with or without serial correlation. Since it is not convenient for practical use (as it requires one to examine the whole data set to reach a decision whether a signal is to be triggered at time T), any practical application of the LR strategy involves choosing a window of size M and triggering a signal only if $D^*_m > h$ for values $m_0 \leq m \leq M$. In effect, this amounts to running a truncated SPRT backwards in time. As shown in Lai (1995), in the univariate case one can achieve asymptotic efficiency of the LR test by examining only a subset of values m. However, this approach could still require a search going deep into history to establish whether a signal is to be triggered.

In Yashchin (1995) an alternative procedure, the Regenerative Likelihood Ratio (RLR) is introduced. In its simple version, this procedure calls for determining the depth M_T dynamically, based on the previous history:

Regenerative Likelihood Ratio (RLR) Strategy: Given that at time T the last regeneration point was registered M_T units of time ago, trigger a signal if $D_m^* > h$ for **some** $m_0 \le m \le M_T$. If $D_m^* \le 0$ for **every** $m_0 \le m \le M_T$, declare T the new regeneration point.

When the observations are univariate and their distribution belongs to an exponential family, the LR and RLR strategies are equivalent, see Yashchin (1995)). The latter paper also discusses the relative merits of these approaches. In what follows, we limit our attention to the RLR schemes.

It is sometimes more convenient to implement the RLR strategy in terms of a scheme $\{s_t, t = 1, 2, \ldots, T\}$ that can be plotted on a chart so as to enable one to assess the state of the process graphically:

RLR Strategy, Alternative Formulation: Define the value of the scheme s_T at time T by

$$s_T = \max_{m_0 \le m \le M_T} D_m^* \tag{2.2}$$

and trigger a signal at time T if $s_T > h$. If $s_T \le 0$ then declare M_T the new regeneration point.

3. PARAMETERS OF INTEREST IN COUNT DATA MODELS

Consider the problem of monitoring level of contamination in a chip manufacturing line. In a typical fab, almost all parts of the line are located in so called "clean rooms" where special precautions are taken to eliminate even very small particles from the air, tools and workers. The air quality is monitored by counting, at regular time intervals, the number of particles of size exceeding a certain threshold that land on the exposed test surface. The particles are further categorized by type to facilitate the diagnosis in case the counts become unacceptably high. The four categories considered in this example are "Metal", "Organic", "Inorganic" and "Other". In what follows, we refer to these categories as ML, GC, IC and OR, respectively. Consider an air filtering system that serves three parts of the line which we will call areas A1, A2 and A3. The purpose of the monitoring system is to trigger a signal as soon as possible after onset of an undesirable event, while maintaining a low rate of false alarms. Furthermore, we will show how one can set up the system in such a way that a signal is helpful in diagnosing the problem.

When deciding on monitoring policy, one must first have a good idea on what can go wrong with the process and, based on that, decide which parameters to monitor and prioritize them. To clarify this statement, note that monitoring multiple parameters inevitably poses a problem of keeping the *overall* rate of false alarms low. By dividing the set of parameters into classes of first and second priority and making sure that the rate of false alarms for the second priority parameters is very low (which will cause an agreeable decrease in detection sensitivity for these parameters), one will be able to greatly reduce the intensity of false alarms and take the maximum advantage of the available data stream.

3.1. The Model

The data used in the monitoring system consists of a sequence of two-way contingency tables $\{\boldsymbol{B}_t\}$, $t = 1, 2, \ldots$. Every table has 3 rows corresponding to areas A1, A2, A3, and 4 columns corresponding to particle types ML, GC, IC and OR. Since it is quite convenient to work in a general setting, we shall denote the number of rows in the table by I and the number of columns - by J. We shall use the following notation:

n is the number of particles in a single observed contingency table.

n_{ij} is the number of particles in the cell (i, j).

$n_{i.}$ number of particles in the i-th row.

$n_{.j}$ number of particles in the j-th column.

λ_{ij} expected number of particles in the cell (i, j)

$\lambda_{i.}$ expected number of particles in the i-th row.

$\lambda_{.j}$ expected number of particles in the j-th column.

λ expected number of particles in the table.

p_{ij} is the probability that the next observed particle will be in the cell (i, j) under the target operating conditions.

$p_{i.}$ is the probability that the next observed particle will be in the i-th row under the target operating conditions.

$p_{.j}$ is the probability that the next observed particle will be in the j-th column under the target operating conditions.

π_{ij} is the probability that the next observed particle will be in the cell (i, j) under the actual operating conditions.

$\pi_{i.}$ is the probability that the next observed particle will be in the i-th row under the actual operating conditions.

$\pi_{.j}$ is the probability that the next observed particle will be in the j-th column under the actual operating conditions.

In this article we assume that the data for an individual two-way table comes from the following model: first, the counts for individual areas are determined based on some discrete (say, Poisson) random variables with means $\lambda_{i.}, i = 1, 2, \ldots, I$. The total number of particles is, therefore, distributed with mean $\lambda = \sum_i \lambda_{i.}$ and the probability that a particle falls in row i is $\pi_{i.} = \lambda_{i.}/\lambda$. Subsequently, these counts are distributed into individual cells in accordance with the probabilities $\pi_{ij}/\pi_{i.}$ which will be considered the basic parameters of the model.

3.2. Area - specific Parameters

It is quite obvious that a sensible monitoring policy should give proper consideration to events pertaining to each individual area: if, for example, a tool in a given area malfunctions, one would like to be able to detect this fact and diagnose the problem by focusing on data exclusively from this area. So, let is focus on the data related to the i-th area. One simple approach to monitoring would be to consider the cell counts as Poisson distributed random variables with means $\{\lambda_{ij}\}$ and monitor data in each cell (i, j) by specifying acceptable and unacceptable regions for the respective λ_{ij}. In the situation of interest, however, such policy would hardly be advisable, the main reason being that it is not clear how to set the acceptable and unacceptable regions. Consider, for example, metal particles in Area 1. An increase in the average count of metal particles may not in itself represent a danger to the process: it is the overall number of particles in Area 1 that determines the process yield. So, in the case where filtering in Area 1 fails we could get a signal that there is an increase in the number of metal particles, which could confuse the diagnostic process: in fact, there is an increase in *every* particle category, but other charts did not yet produce a signal. So, one could prefer in this situation to monitor the expected total particle counts in every area (the acceptable and unacceptable levels for

the underlying parameter can usually be easily established) and, in addition, monitor the proportions of particles of various types in every row.

Now let us focus on the data corresponding to the i-th row. In light of the arguments given above, $\lambda_{i.}$ should be considered a first priority parameter and monitored based on the counts $n_{i.}$. Let us introduce the notation

$$\tilde{p}_{ij} = p_{ij}/p_{i.} \quad \text{and} \quad \tilde{\pi}_{ij} = \pi_{ij}/\pi_{i.}, \tag{3.1}$$

for $j = 1, 2, \ldots, J$. We shall also define the distribution $(\tilde{\pi}_{ij}, \quad j = 1, 2, \ldots, J)$ by $\tilde{\boldsymbol{\pi}}_i$ and the distribution $(\tilde{p}_{ij}, \quad j = 1, 2, \ldots, J)$ - by $\tilde{\boldsymbol{p}}_i$. Since a malfunctioning tool is considered quite likely to produce particles of a single type, it also makes sense to monitor the individual proportions, $\tilde{\pi}_{ij}$ to detect changes upwards from the respective target levels, $\tilde{p}_{ij}$. These target levels and unacceptable deviations are determined from historic data related to past normal and abnormal environments in a specific area.

Another important parameter to monitor is the lack of fit between the distribution $\tilde{\boldsymbol{\pi}}_i$ and the target distribution, $\tilde{\boldsymbol{p}}_i$. Even if the overall particle intensity $\lambda_{i.}$ and the individual proportions $\tilde{\pi}_{ij}$ remain within acceptable limits, a change in the distribution will typically be associated with some new factor affecting the manufacturing line that the engineers need to know about. It occurs quite frequently that changes of this nature have nothing to do with the line itself, but are related to malfunctions of measurement tools, changes in the data acquisition procedure or arrival of a new technician. The situation here is quite similar to that occurring in monitoring multivariate data: it is not only important to detect changes in individual measurements, but also to detect changes that affect all the measurements simultaneously.

To monitor the discrepancy between the actual and target distributions within the i-th row we must first choose the measure of distance between these distributions. One natural choice is to use the Kullback - Leibler information defined by

$$d_i = I(\tilde{\boldsymbol{p}}_i, \tilde{\boldsymbol{\pi}}_i) = \sum_j \tilde{p}_{ij} ln(\tilde{p}_{ij}/\tilde{\pi}_{ij}). \tag{3.2}$$

The measure d_i has a number of attractive theoretical properties and it has been used quite extensively in research related to categorical data (see Kullback (1959), Gokhale and Kullback (1978)). The main reason for statistical power of techniques based on this measure is related to the fact that it is proportional to the expected log-likelihood ratio between the actual and target distributions. In many applications, it is advisable to monitor d_i as a second priority parameter. One of the problems in practical applications is how to determine its acceptable and unacceptable levels. This can usually be achieved by analysis of stretches of historic data corresponding to acceptable process conditions and plotting the resulting estimates $\hat{d}_i$ that for longer stretches could be considered as approximately equal to the underlying values of d_i. The unacceptable level can then be chosen somewhere in the right tail of this distribution; under such conditions, one will be able to state, following a signal, that the discrepancy between the actual and target distributions is much higher than that expected under satisfactory conditions. The acceptable level can be chosen somewhere around the sample mean of $\hat{d}_i$.

3.3. Global Parameters

In the described data model it is natural to monitor several global features related to the sequence of tables $\{\boldsymbol{B}_t\}$. One important (usually first priority) parameter is the rate λ. The reason for monitoring λ in addition to particle rates in individual areas is related to the fact that the latter are driven by different assignable causes. For example, when the chart for monitoring λ triggers a signal, it can typically be associated with the failure of the central filtering system; if, however, a chart for

$\lambda_{i.}$ triggers a signal, the cause is more likely to be within the area itself (of course, such diagnostic statements can only be made if the charts under consideration are properly designed).

Another parameter worth monitoring involves the distance d between the actual data model and the class of two-way classifications for which the conditional distributions within rows are equal to the target vectors ($\tilde{\boldsymbol{p}}_i,\ i = 1, 2, \ldots, I$). In what follows we will refer to this family as the FCR (fixed - conditional - rows) class of two-way distributions and denote it by $FCR\{\tilde{p}_{ij}\}$. The reason for monitoring d in addition to monitoring d_i for every row is again related to our requirement that an out of control signal should also be helpful in diagnosing the problem: the assignable causes leading to increase in an individual d_i can be different from those leading to increase in d.

One simple way to define such a distance is by using the principle of *minimum discrimination information* (MDI) estimation (see Gokhale and Kullback (1978)). This principle states that a distance between a given distribution $\{\pi_{ij}\}$ (typically determined by the data) and some class of models (that are used to explain the data) is defined as the minimal Kullback - Leibler distance measured between all representatives of the class and $\{\pi_{ij}\}$; when $\{\pi_{ij}\}$ is an empirical distribution, the representative of the class for which the minimum is reached is called the MDI estimator. The properties of such an estimator are typically comparable to those of the Maximum Likelihood Estimator (MLE) and in many cases the two coincide.

So, consider the $FCR\{\tilde{p}_{ij}\}$ class of two-way distributions. Each member of this class represents a collection of rows $\{\alpha_i\tilde{\boldsymbol{p}}_i, i = 1, 2, \ldots, I\}$ with some $\alpha_i > 0$ and $\sum_i \alpha_i = 1$. Let us find the representative of this class that is closest to the actual underlying distribution $\{\pi_{ij}\}$. This is achieved by minimizing (by α_i) the distance

$$\sum_{i,j} \alpha_i \tilde{p}_{ij} ln(\alpha_i \tilde{p}_{ij}/\pi_{ij}) \tag{3.3}$$

subject to the constraint $\sum_i \alpha_i = 1$. By using the Lagrange multiplier method we obtain the optimal values of α_i,

$$\alpha_i^* = \frac{\pi_{i.} exp(-d_i)}{\sum_i \pi_{i.} exp(-d_i)}, \tag{3.4}$$

where d_i is given by (3.2). The distance d between the $FCR\{\tilde{p}_{ij}\}$ class of models and the distribution $\{\pi_{ij}\}$ is, therefore,

$$d = -ln\left[\sum_i \pi_{i.} exp(-d_i)\right]. \tag{3.5}$$

Finally, one will also want to monitor some measure of association between the rows and columns in the table. Such measure is an analog of a squared correlation coefficient (it is always between 0 and 1). The value 0 means that the rows and columns are independent, i.e. $\pi_{ij} = \pi_{i.}\pi_{.j}$ for every (i, j). In this article we limit our attention to one such measure - the Theil's η defined by

$$\eta = 1 - \frac{\sum_{i,j} \pi_{ij} ln(\pi_{ij}/\pi_{i.})}{\sum_j \pi_{.j} ln(\pi_{.j})} \tag{3.6}$$

(see Theil (1970)). One can see that this measure does not treat the rows and columns symmetrically: in the setting of our example, the denominator represents the overall entropy of the distribution of particles while the numerator is the expected entropy of the particle type distribution given the area.

The reason why it is useful to monitor η in addition to monitoring other global parameters is related to the fact that its behavior is typically associated with its own set of assignable causes. For example, in the situation where the conditional distributions $\tilde{\boldsymbol{\pi}}_i$ undergo changes of the same kind (say, increase in proportions of ML and OC particles at the expense of IC and OR particles), the chart for d will trigger a signal, but the chart for η will remain comfortably below its threshold.

However, in the case where areas are affected in a non-systematic way, the signal will typically be more likely to be triggered by the chart for η.

4. THE MONITORING SCHEMES

Since our monitoring strategy calls for decomposing the data stream into sequences of univariate statistics (each sequence being targeted to its own set of assignable causes), the LR scheme will be somewhat simpler from the computational standpoint. The easiest to handle is the case where we monitor a basic parameter for which we can identify a sufficient statistic. The main reason for this simplification is related to the fact that in this case the acceptable and unacceptable regions for the monitored parameter are one - dimensional sets and we can typically point out the value $\theta_0 \in \Omega_0$ that is closest (in the sense of Kullback - Leibler distance) to Ω_1. Similarly, we can point out the value $\theta_1 \in \Omega_1$ that is closest to Ω_0. So, if we denote by $\hat{\theta}(m)$ the MLE based on the last m data points (in the case of our example, on data in the tables $\boldsymbol{B}_{T-m+1}, \ldots, \boldsymbol{B}_{T-1}, \boldsymbol{B}_T$), and by L_m^* the corresponding value of the log-likelihood, $L_m^* = L_m(\hat{\theta}(m))$, then the maximum log-likelihoods L_{m0}^* and L_{m1}^* are determined as follows:

(a) If $\hat{\theta}(m) \in \Omega_0$ set $L_{m0}^* = L_m^*$ and $L_{m1}^* = L_m(\theta_1)$.

(b) If $\hat{\theta}(m) \notin \Omega_0$ and $\hat{\theta}(m) \notin \Omega_1$, set $L_{m0}^* = L_m(\theta_0)$ and $L_{m1}^* = L_m(\theta_1)$.

(c) If $\hat{\theta}(m) \in \Omega_1$ set $L_{m1}^* = L_m^*$ and $L_{m0}^* = L_m(\theta_0)$.

The above technique simplifies computations greatly in cases where we monitor the only parameter of the distribution (it can also be extended to the case where nuisance parameters are present, see Yashchin (1995)). Note that in the case (a) there is no need to compute the score D_m^* because it is negative.

4.1. Monitoring the Basic Model Parameters

In the context of categorical data, this technique can be used to monitor the basic model parameters $\lambda_{i.}$, $\tilde{\pi}_{ij}$ or a simple function like λ. We will only give here schemes related to λ and to parameter p of a binomial population with variable sample sizes. The reader will find it quite easy to use the presented formulas to derive schemes for $\tilde{\pi}_{ij}$ and $\lambda_{i.}$.

Suppose that the acceptable and unacceptable levels of λ are $\lambda \leq \underline{\lambda}$ and $\lambda \geq \overline{\lambda}$, respectively ($\underline{\lambda} < \overline{\lambda}$). Let us define the *reference value* $k_\lambda = (\overline{\lambda} - \underline{\lambda})/(ln\overline{\lambda} - ln\underline{\lambda})$ and let $\hat{\lambda}_{(m)} = (1/m)\sum_{T-m+1}^{T} n_t$, i.e., the average sample size for the last m tables. Then, one can show that the score is

$$D_m^* = \begin{cases} m[\hat{\lambda}_{(m)}(ln\overline{\lambda} - ln\hat{\lambda}_{(m)}) - (\overline{\lambda} - \hat{\lambda}_{(m)})] & \text{if} \quad \hat{\lambda}_{(m)} \leq \underline{\lambda} \\ m(\ln\overline{\lambda} - \ln\underline{\lambda})(\hat{\lambda}_{(m)} - k_\lambda) & \text{if} \quad \underline{\lambda} < \hat{\lambda}_{(m)} \leq \overline{\lambda} \\ m[\hat{\lambda}_{(m)}(ln\hat{\lambda}_{(m)} - ln\underline{\lambda}) - (\hat{\lambda}_{(m)} - \underline{\lambda})] & \text{if} \quad \hat{\lambda}_{(m)} \geq \overline{\lambda} \end{cases}$$

Since $D_m^* \leq 0$ when $\hat{\lambda}_{(m)} \leq k_\lambda$, we can reduce the upper LR control scheme to the following form:

Upper LR scheme for λ : Trigger an out of control signal at time T if for some $m \geq 1$

$$\begin{array}{lll} k_\lambda \leq \hat{\lambda}_{(m)} \leq \overline{\lambda} & \text{and} & m(\ln\overline{\lambda} - \ln\underline{\lambda})(\hat{\lambda}_{(m)} - k_\lambda) > h_\lambda \quad \text{or} \\ \hat{\lambda}_{(m)} \geq \overline{\lambda} & \text{and} & m[\hat{\lambda}_{(m)}(ln\hat{\lambda}_{(m)} - ln\underline{\lambda}) - (\hat{\lambda}_{(m)} - \underline{\lambda})] > h_\lambda, \end{array} \tag{4.1}$$

where h_λ is a suitably chosen signal level.

One can easily present the above scheme in an RLR form. Instead of doing this, we will give a simplified version of this scheme, which uses the given acceptable and unacceptable regions only

when computing scores based on $m = 1$. For $m > 1$ the score is computed based on the regions collapsing into a single point, namely, $\Omega_0 = \{\underline{\lambda}\}$ and $\Omega_1 = \{\overline{\lambda}\}$. This leads to what we call a *mixed* scheme:

Mixed Upper RLR scheme for λ : Trigger an out of control signal when the process s_t defined by

$$s_0 = 0, \quad s_t = max\{0, s_{t-1} + (n_t - k_\lambda)\}, \quad t = 1, 2, \ldots \tag{4.2}$$

exceeds the value $\tilde{h}_\lambda$ defined by $\tilde{h}_\lambda = h_\lambda/(\ln \overline{\lambda} - \ln \underline{\lambda})$. In addition, when $\overline{\lambda} < \tilde{h}_\lambda + k_\lambda$, also trigger a signal when $n_t > x(\underline{\lambda}, h_\lambda)$, where $x(\underline{\lambda}, h_\lambda)$ is the solution of the equation

$$x(\ln x - \ln \underline{\lambda}) - (x - \underline{\lambda}) = h_\lambda. \tag{4.3}$$

This scheme is somewhat similar to a Cusum scheme supplemented with a Shewhart limit (see Lucas (1982)) with one important difference: the need for supplemental Shewhart limit and its magnitude are determined automatically from other parameters of the problem. Therefore, to design a mixed scheme we need to adjust only one parameter, h_λ.

The mixed procedure defined above is of Markovian type in the sense that under the assumption of independence the values of the scheme form a Markov chain. Therefore, the ARL performance of the scheme can be analyzed by using the theory of Markov chains, see Brook and Evans (1972). Analysis of performance of the RLR scheme is more complex and will typically require simulation runs.

Now consider the problem of monitoring a proportion of particles of a given type in a given area. In essence, this is a problem of monitoring binomial proportions based on variable sample sizes. To simplify the notation, we will denote the proportion of interest π and the sample sizes corresponding to the *combined* data from last m tables by $n_{(m)}$. Let us denote the estimate of π based on the sample size $n_{(m)}$ by $\hat{\pi}_{(m)}$. Let us also define $I(p, \pi)$ as the Kullback - Leibler distance between two binomial distributions with probabilities of success p and π:

$$I(p, \pi) = pln(\pi/p) + (1 - p)ln[(1 - p)/(1 - \pi)]. \tag{4.4}$$

Let the acceptable and unacceptable levels of π be $\pi \leq \underline{\pi}$ and $\pi \geq \overline{\pi}$ ($\underline{\pi} < \overline{\pi}$). Then, the score D_m^* is given by

$$D_m^* = \begin{cases} -n_{(m)} I(\hat{\pi}_{(m)}, \overline{\pi}) & \text{if} \quad \hat{\pi}_{(m)} \leq \underline{\pi} \\ n_{(m)}[I(\hat{\pi}_{(m)}, \underline{\pi}) - I(\hat{\pi}_{(m)}, \overline{\pi})] & \text{if} \quad \underline{\pi} < \hat{\pi}_{(m)} \leq \overline{\pi} \\ n_{(m)} I(\hat{\pi}_{(m)}, \underline{\pi}) & \text{if} \quad \hat{\pi}_{(m)} \geq \overline{\pi} \end{cases}$$

Define the reference value for π by

$$k_\pi = \frac{ln[(1 - \overline{\pi})/(1 - \underline{\pi})]}{ln[(1 - \overline{\pi})/(1 - \underline{\pi})] - ln(\overline{\pi}/\underline{\pi})}$$

Then, as before, the LR scheme can be reduced to a quite simple form:

Upper LR scheme for π : Trigger an out of control signal at time T if for some $m \geq 1$

$$\begin{array}{lll} k_\pi \leq \hat{\pi}_{(m)} \leq \overline{\pi} & \text{and} & n_{(m)} \left[ln\frac{\overline{\pi}}{1-\overline{\pi}} - ln\frac{\underline{\pi}}{1-\underline{\pi}} \right] (\hat{\pi}_{(m)} - k_\pi) > h_\pi \quad \text{or} \\ \hat{\pi}_{(m)} \geq \overline{\pi} & \text{and} & n_{(m)} I(\hat{\pi}_{(m)}, \underline{\pi}) > h_\pi, \end{array} \tag{4.5}$$

where h_π is a signal level.

Now denote by $\{\hat{\pi}_t\}$ the proportions corresponding to individual tables and by $\{n_t\}$ - the corresponding sample sizes ($n_{(m)} = \sum_{T-m+1}^{T} n_t$). Then the *mixed* version of the scheme can be formulated as follows:

Mixed Upper RLR scheme for π : Trigger an out of control signal when the process s_t defined by

$$s_0 = 0, \quad s_t = max\{0, s_{t-1} + n_t(\hat{\pi}_t - k_\pi)\}, \quad t = 1, 2, \ldots \tag{4.6}$$

exceeds the value $\tilde{h}_\pi$ defined by $\tilde{h}_\pi = h_\lambda / \left[ln\frac{\overline{\pi}}{1-\overline{\pi}} - ln\frac{\underline{\pi}}{1-\underline{\pi}} \right]$. In addition, when $\overline{\pi} < \tilde{h}_\pi/n_t + k_\pi$, also trigger a signal when $p_t > x(\underline{\pi}, h_\pi)$, where $x(\underline{\pi}, h_\pi)$ is the solution of the equation

$$I(x, \underline{\pi}) = h_\pi / n_t. \tag{4.7}$$

Design and analysis of the above mixed scheme is much easier than that of the RLR scheme and it can be used not only as a scheme in its own right, but also as a tool to provide initial assessment of the RLR scheme performance. It is closely related to the concept of Weighted Cusum Schemes (see Yashchin (1989), Bissell (1973)). To design and analyze this scheme one will typically base computations on a fixed (most prevalent) sample size.

4.2. Monitoring Functions of Parameters

A more complex situation arises when monitoring d_i, d or η because in these parameters are non-trivial functions of the basic model parameters. Consider, for example, the parameter d and suppose that the acceptable and unacceptable levels of this parameter are $d \leq \underline{d}$ and $d \geq \bar{d}$, respectively ($\underline{d} < \bar{d}$). Denoting the set of basic parameters by $\boldsymbol{\theta}$ we will abbreviate the relationship (3.5) by writing $d = d(\boldsymbol{\theta})$. Then Ω_0 contains all the values $\boldsymbol{\theta}$ for which $d(\boldsymbol{\theta}) \leq \underline{d}$ and Ω_1 - the values $\boldsymbol{\theta}$ for which $d(\boldsymbol{\theta}) \geq \overline{d}$.

A substantial simplification occurs here because one prove the following

Statement 4.1: If for the MLE $\hat{\boldsymbol{\theta}}$ based on some data set we have $d(\hat{\boldsymbol{\theta}}) \geq \underline{d}$ then the maximum value of likelihood in Ω_0 is achieved for some value $\boldsymbol{\theta}$ for which $d(\boldsymbol{\theta}) = \underline{d}$. Similarly, if $d(\hat{\boldsymbol{\theta}}) \leq \overline{d}$ then the maximum value of likelihood in Ω_1 is achieved for some value $\boldsymbol{\theta}$ for which $d(\boldsymbol{\theta}) = \overline{d}$.

Now define the MLE of $\boldsymbol{\theta}$ based on data for the last m periods by $\hat{\boldsymbol{\theta}}(m)$, the corresponding MLE estimate $d(\hat{\boldsymbol{\theta}}(m))$ of d – by $\hat{d}_{(m)}$ and the corresponding maximal value $L_m(\hat{\boldsymbol{\theta}}(m))$ of the log-likelihood – by L_m^*. In light of Statement 4.1, the procedure for computing L_{m0}^* and L_{m1}^* for the problem of monitoring d is then as follows:

(a) If $\hat{d}_{(m)} \leq \underline{d}$ set $L_{m0}^* = L_m^*$ and $L_{m1}^* = \max_{d(\boldsymbol{\theta})=\overline{d}} L_m(\boldsymbol{\theta})$.

(b) If $\underline{d} \leq \hat{d}_{(m)} \leq \overline{d}$ set $L_{m0}^* = \max_{d(\boldsymbol{\theta})=\underline{d}} L_m(\boldsymbol{\theta})$ and $L_{m1}^* = \max_{d(\boldsymbol{\theta})=\overline{d}} L_m(\boldsymbol{\theta})$.

(c) If $\hat{d}_{(m)} \geq \overline{d}$ set $L_{m1}^* = L_m^*$ and $L_{m0}^* = \max_{d(\boldsymbol{\theta})=\underline{d}} L_m(\boldsymbol{\theta})$.

Once again, the optimization in the case (a) is unnecessary because in this case $D_m^* < 0$.

We shall illustrate application of the above procedures to monitor d and η. Since the d_i can be considered as a special case corresponding to a single-row table, our ability to design a monitoring scheme for d implies directly ability to design schemes for d_i as well.

Let us denote by $n_{ij(m)}$ the counts corresponding to the *combined* data from last m tables. The log-likelihood L_m of the data corresponding to the distribution $\{\pi_{ij}\}$ is given (up to an additive constant) by

$$L_m = \sum_{i,j} n_{ij(m)} ln(\pi_{ij}). \tag{4.8}$$

Its maximum, L_m^*, is attained at the MLE estimates,

$$\hat{\pi}_{ij(m)} = n_{ij(m)} / \sum_{i,j} n_{ij(m)}. \tag{4.9}$$

and the MLE estimate of d (i.e., $\hat{d}_{(m)} = d(\hat{\boldsymbol{\theta}}(m))$ used in the above procedure for finding L_{m0}^* and L_{m1}^*) is obtained by substituting $\hat{\pi}_{ij(m)}$ for π_{ij} in the formula (3.5).

To find the values L_{m0}^* and L_{m1}^* we must be able to solve the following problem:

$$\begin{array}{l} \textit{Maximize } L_m \\ \\ \textit{Subject to: } \quad \sum_{i,j} \pi_{ij} = 1 \quad \textit{and} \\ \sum_i \pi_{i.} exp(-d_i) = exp(-d), \end{array} \tag{4.10}$$

where d_i are given by (3.2) and d is some fixed value. It follows from the Statement 4.1 that the values L_{m0}^* and L_{m1}^* are either the maximum values of the objective function corresponding to the solution of the above problem for $d = \underline{d}$ and $d = \overline{d}$, respectively, or are equal to L_m^*. The outline of the optimization process for the problem (4.10) is given in Appendix A.

An approach to monitoring the coefficient of association, η is quite similar. The acceptable and unacceptable regions for this parameter are frequently of type $\eta \leq \underline{\eta}$ and $\eta \geq \overline{\eta}$, respectively ($\underline{\eta} < \overline{\eta}$). The MLE estimate $\hat{\eta}_{(m)}$ is obtained by substituting (4.9) into (3.6) and the corresponding L_m^* is the same as in the problem of monitoring d.

We obtain $L_{m0}^* = L_m^*$ when $\hat{\eta}_{(m)} < \underline{\eta}$ and $L_{m1}^* = L_m^*$ when $\hat{\eta}_{(m)} > \overline{\eta}$. In other cases, we find the values L_{m0}^* and L_{m1}^* by solving the problem:

$$\begin{array}{l} \textit{Maximize } L_m \\ \\ \textit{Subject to: } \quad \sum_{i,j} \pi_{ij} = 1 \quad \textit{and} \\ \sum_{i,j} \pi_{ij} ln(\pi_{ij}/\pi_{i.}) = (1-\eta)(\sum_j \pi_{.j} ln(\pi_{.j})), \end{array} \tag{4.11}$$

for $\eta = \underline{\eta}$ and $\eta = \overline{\eta}$. The outline of the optimization process for the problem (4.11) is given in Appendix B.

In practical applications one may also want to consider some simpler procedures for monitoring d and η. For example, one could consider a weighted Geometric Cusum (GC) scheme based on the MLE estimates $\hat{d}_t$ corresponding to individual tables $\boldsymbol{B}_t$:

Upper GC scheme for d : Trigger an out of control signal when the process s_t defined by

$$s_0 = 0, \quad s_t = max\{0, \gamma s_{t-1} + n_t(\sqrt{\hat{d}_t} - k_d)\}, \quad t = 1, 2, \ldots \tag{4.12}$$

exceeds the threshold h_d.

In this scheme the reference value k_d is chosen about midway between the acceptable and unacceptable levels for $\sqrt{d}$ and the parameter γ is typically between 0.7 and 0.9. The scheme (4.12) is statistically less powerful than the RLR, but it is somewhat easier to design.

A procedure similar to (4.12) can also be used for monitoring η based on the estimates $\hat{\eta}_t$ obtained by substitution of the MLE estimates of $\{\pi_{ij}\}$ into (3.6). As in (4.12), we recommend to use a square-root transform to reduce skewness.

5. AN EXAMPLE

Consider once again the situation described in Section 3. We will now describe a monitoring system for the parameters discussed in this article. To simplify the presentation, we will limit our attention to GC procedures only. As shown in Yashchin (1989), procedures of this type can be deployed by either using schemes like (4.12) directly or via the V-mask approach that calls for plotting weighted deviations from some centering constant (usually target) and triggering a signal when a trajectory falls outside a V-mask. Our display will be based on the latter approach.

Let us suppose that the parameter values prevalent in the recent history are $\lambda_{i.} = 70, 50$ and 40 (for $i = 1, 2, 3$) and the prevalent proportions of particles of various types are $\tilde{\boldsymbol{\pi}}_i = (0.4, 0.3, 0.2, 0.1)$, for every area i. The prevalent values for $\sqrt{d_i}$ are 0.1 for every i. We will use the above values as centering constants in our charts. Note that these values are used exclusively for charting purposes - they do not play a role in a detection scheme which is based on the acceptable and unacceptable regions for each parameter. With $\gamma = 0.8$ the GC trajectories are shown in Fig. 5.1. Every chart occupies a sector on this plot and an Exponentially Weighted Moving Average (EWMA) estimate of the underlying parameter corresponds to the point on the scale to which the GC trajectory is pointing. The abrupt changes, however, correspond to a change in slope of the trajectory (like as in a Cusum display). Such form of display is called a Tunnel Chart.

For example, for a binomial proportion π_t the corresponding plot contains at time T the points (0, 0) and

$$\{\sum_{r=1}^{t} \gamma^{T-t} n_r, \sum_{r=1}^{t} \gamma^{T-r} n_r(\hat{\pi}_r - t_0)\}, \quad t = 1, 2, \ldots, T, \tag{5.1}$$

where t_0 is the respective centering constant. For the counts n_t (used to monitor λ) the GC plot is analogous to (5.1), but there is no weighting by sample sizes. The plot labels give the total observed and expected particle counts for each area.

We will not discuss deriving control schemes for the parameters of the individual areas. The only scheme we will show is a GC (4.12) with $h_d = 0.15$ and $k_d = 0.05$ to detect a change in d (here the acceptable and unacceptable levels of $\sqrt{d}$ are taken to be $d < 0.1$ and $d > 0.2$, respectively. The one-sided V-mask shown on the plot indicates an out of control signal. The on-target ARL of this scheme is about 1000.

The plot indicates that there was a major change in d five days ago. This change is mostly related to an increase in the proportion of OR particles at the expense of ML and OC particles (the proportion of IC particles remains basically unchanged). Figure 5.1 shows that the impact of the change is seen in every area, but individual area charts are not likely to detect it because these changes do not appear high for the area sample sizes. The global chart, however, detects this change and pinpoints its location. Since the chart for η remains below its target level, 0.1, the underlying assignable cause tends to affect all three areas in about the same way.

Another event detected by the global chart is increase in the overall particle intensity, λ. For illustrative purposes we show how one can also deploy an EWMA chart in which a signal is triggered when the EWMA value exceeds a threshold. Since under target conditions the variance of the EWMA is $\sqrt{160(1-0.8)/(1-0.8)} = 4.22$, the control limits corresponding to on-target ARL = 1000 are $160 \pm (3.19)(4.22) = 160 \pm 13.5$. These bounds are shown on the plot for λ. The chart shows an out of control condition associated with a change in λ that happened five days ago. This change also affects all areas but is too small to be detected in every area individually.

The Area 1 data used in the Example is shown in Table 5.1. The data for other areas is omitted.

Table 5.1. Particle Counts for Area 1

Date	Total	Metal	Organic	Inorgc	Other
06/01/04	30	13	8	5	4
06/02/04	38	18	6	8	6
06/03/04	60	24	12	14	10
06/04/04	49	20	18	8	3
06/05/04	69	27	20	13	9
06/06/04	50	22	11	9	8
06/07/04	30	15	6	7	2
06/08/04	57	23	16	10	8
06/09/04	76	30	21	13	12
06/10/04	43	18	9	12	4
06/11/04	57	25	15	10	7
06/12/04	73	29	20	13	11
06/13/04	73	33	18	15	7
06/14/04	66	30	17	8	11
06/15/04	41	15	14	9	3
06/16/04	67	26	12	16	13
06/17/04	46	22	11	8	5
06/18/04	85	36	22	15	12
06/19/04	80	30	19	16	15
06/20/04	58	24	17	11	6
06/21/04	83	29	23	15	16
06/22/04	86	33	23	16	14
06/23/04	79	32	21	13	13
06/24/04	75	26	15	20	14
06/25/04	77	24	20	19	14

6. CONCLUSIONS

The problems of monitoring count data occur frequently in industrial applications. In many practical situations the models that drive the data involve abrupt changes and can be handled by methods especially developed for such models in the last decade.

In the area of detection, the Regenerative Likelihood Ratio technique looks very promising. This technique combines the statistical power of inference based on Likelihood Ratios with practical convenience associated with an efficient mechanism for discarding older data that is no longer relevant for the purpose of control. It can be applied to various types of data, including multivariate data and serially correlated data.

In this article we showed how the RLR technique can be used to derive monitoring schemes for parameters of categorical data that are of interest in many industrial applications. We also illustrated how the technique can be used to derive simpler schemes (such as mixed schemes) that offer a somewhat easier design process at the expense of some statistical power. In practical applications, one may also want to implement a graphical display of a simpler scheme in combination with decision - making based on the RLR.

Though our discussion has centered around two-way classifications, the methods can be generalized to handle more complex contingency arrays - for example, the concept of the FRC class of

models can be generalized for more complex arrays and the distance d from such a class makes sense in many problems involving multi-way classifications. In many of these problems one will also find it useful to monitor a generalized coefficient of association, η.

APPENDIX A: MAXIMIZING LIKELIHOOD UNDER CONSTRAINT ON d.

To maximize L_m given by (4.8) under the constraints in (4.10) we introduce the Lagrangian

$$G = \ln L_m + \alpha \bullet \left(1 - \sum_{i,j} \pi_{ij}\right) + \beta \bullet \left[\sum_i exp\left\{-\sum_j \tilde{p}_{ij} ln(\tilde{p}_{ij}/\pi_{ij})\right\} - c\right] \tag{A.1}$$

where, for reasons of convenience, we use the notation $c = exp(-d)$. By setting the derivatives of G by π_{ij} to zero we obtain a system of equations

$$\frac{\partial G}{\partial \pi_{ij}} = \frac{n_{ij(m)}}{\pi_{ij}} - \alpha + \beta \left[exp\left\{-\sum_j \tilde{p}_{ij} ln(\tilde{p}_{ij}/\pi_{ij})\right\}\right] \bullet \frac{\tilde{p}_{ij}}{\pi_{ij}} = 0, \quad \text{for every } (i,j). \tag{A.2}$$

This implies, for every (i,j), the relation

$$n_{ij(m)} = \alpha \pi_{ij} - \beta \tilde{p}_{ij} \pi_{i.} exp(-d_i), \tag{A.3}$$

where d_i are defined by (3.2). Let us introduce the notation $n = \sum_{i,j} n_{ij(m)}$ and $n_{i.} = \sum_j n_{ij(m)}$. By adding all the equations in (A.3) we obtain $\alpha = n + \beta c$, which enables us to represent the system (A.2) in the form

$$n_{ij(m)} = (n + \beta c)\pi_{ij} - \beta \tilde{p}_{ij} \pi_{i.} exp(-d_i). \tag{A.4}$$

After summing the above system by j we obtain, for every i,

$$\beta \pi_{i.} exp(-d_i) = (n + \beta c)\pi_{i.} - n_{i.}, \tag{A.5}$$

and our system of equations can be written in the form

$$\pi_{ij} = \pi_{i.} \tilde{p}_{ij} + (n_{ij(m)} - \tilde{p}_{ij} n_{i.})/(n + \beta c), \quad \text{for every } (i,j). \tag{A.6}$$

The above equations show that the system (A.4) can be reduced to a much smaller system for finding the marginal probabilities $\pi_{i.}$. In particular, after substitution of (A.6) into (A.5) we obtain a system of equations

$$\Psi(\pi_{i.}) = \sum_j \tilde{p}_{ij} \ln\left[\pi_{i.} - \frac{n_{i.} - n_{ij(m)}/\tilde{p}_{ij}}{n + \beta c}\right] - \ln \frac{(n + \beta c)\pi_{i.} - n_{i.}}{\beta} = 0, \text{ for every } i. \tag{A.7}$$

The method for solving the above system is as follows: we perform a one-dimensional search for β. For a given value of β we solve the equations of the system (A.7) *one by one* by using the Newton-Raphson method with an initial point

$$\pi_{i.}{}^{(0)} = n_{i.}/[n + \beta(c - 1)] \tag{A.8}$$

obtained from the Taylor series expansion of the sum in (A.7). If it happens that the distance of the distribution $\{\hat{\pi}_{ij(m)}\}$ defined by the MLE's (see (4.9)) from the $FCR\{\tilde{p}_{ij}\}$ family is exactly d then

the only solution of the system is $\{\hat{\pi}_{ij(m)}\}$ and it corresponds to $\beta = 0$. Otherwise, one can show that there is only one value of β for which the solution of (A.7) satisfies the relation $\sum_i \pi_{i.} = 1$, and finding it solves our optimization problem; to get the complete set of $\{\pi_{ij}\}$ we substitute the found values of β and $\{\pi_{i.}\}$ into (A.6).

If the distance (3.5) of the two-way distribution $\{\hat{\pi}_{ij(m)}\}$ from the $FCR\{\tilde{p}_{ij}\}$ family is less than d then the solution is in the region $\beta > 0$; otherwise, it is in the region $-(n/c) < \beta < 0$. In both cases, we recommend to start the search from $\beta^{(0)} = -n/(2c)$. The Newton-Raphson procedure for β is greatly facilitated by the fact that we have a closed form expression for the derivative of the solution $\pi_{i.}$ of (A.7) by β:

$$\frac{\partial \pi_{i.}}{\partial \beta} = \frac{-1}{n + \beta c} \left\{ \frac{n}{\beta \Psi'(\pi_{i.})} + c\pi_{i.} \right\}. \tag{A.9}$$

APPENDIX B: MAXIMIZING LIKELIHOOD UNDER CONSTRAINT ON η.

To maximize L_m given by (4.8) under the constraints in (4.11) we introduce the Lagrangian

$$G = \ln L_m + \alpha \bullet \left(1 - \sum_{i,j} \pi_{ij} \right) + \beta \bullet \left[\sum_{i,j} \pi_{ij} ln(\pi_{ij}/\pi_{i.}) - (1-\eta)(\sum_j \pi_{.j} ln(\pi_{.j})) \right]. \tag{B.1}$$

By setting the derivatives of G by π_{ij} to zero we obtain a system of equations

$$n_{ij(m)} = \alpha \pi_{ij} - \beta \pi_{ij} \ln(\pi_{ij}/\pi_{i.}) + \beta(1-\eta)\pi_{ij}(1 + \ln \pi_{.j}), \text{ for every } (i,j). \tag{B.2}$$

After introducing the notation $n = \sum_{i,j} n_{ij(m)}$, $n_{i.} = \sum_j n_{ij(m)}$ and adding all the equations in (B.2) we obtain $\alpha = n - \beta(1-\eta)$. Therefore, the system (B.2) can be written in the form

$$n_{ij(m)} = n\pi_{ij} - \beta \pi_{ij} \ln(\pi_{ij}/\pi_{i.}) + \beta(1-\eta)\pi_{ij} \ln(\pi_{.j}), \text{ for every } (i,j). \tag{B.3}$$

It is not difficult to see that one solution of the system (B.3) is

$$\beta = 0, \qquad \pi_{ij} = \hat{\pi}_{ij(m)}, \tag{B.4}$$

where $\hat{\pi}_{ij(m)}$ are the MLE's given by (4.9); the corresponding MLE of η is $\hat{\eta}_{(m)}$. If $\hat{\eta}_{(m)} = \eta$ then (B.4) is the only solution of (B.3) and it gives the solution to our optimization problem. Otherwise, (B.4) corresponds to a saddle point and we must take special precautions to prevent our iterative search procedure from converging to this point. One simple and robust procedure is to perform a one - dimensional search by β until a value is found for which $\sum_{i,j} \pi_{ij} = 1$. This gives the solution to our optimization problem (note that (B.3) implies that it also satisfies the second constraint of (4.11)).

One can show that if $\hat{\eta}_{(m)} < \eta$ then the optimal Lagrange multiplier is in the region $\beta > 0$; otherwise, it is in the region $\beta < 0$.

Consider first the case $\hat{\eta}_{(m)} > \eta$. Once can see that the optimal solution corresponding to the constraint $\eta = 0$ is

$$\bar{\pi}_{ij(m)} = \pi_{i.(m)} \pi_{.j(m)}, \tag{B.5}$$

where $\pi_{i.(m)} = \sum_j \hat{\pi}_{ij(m)}$ and $\pi_{.j(m)} = \sum_i \hat{\pi}_{ij(m)}$. For other values of η we propose to use the initial point

$$\pi_{ij}{}^{(0)} = a\bar{\pi}_{ij(m)} + (1-a)\hat{\pi}_{ij(m)}, \tag{B.6}$$

where a is chosen so as to satisfy both constraints of the problem. The initial value $\beta^{(0)}$ is then chosen as the least squares approximation to the solution of (B.3) corresponding to ${\pi_{ij}}^{(0)}$. The proposed iterative process for a fixed β is then

$$\pi_{ij}^{(k+1)} = n_{ij(m)}/n + (\beta/n)\pi_{ij}^{(k)} \left\{ \ln(\pi_{ij}^{(k)}/\pi_{i.}^{(k)}) - (1-\eta)\ln \pi_{.j}^{(k)} \right\}, \text{ for every } (i,j). \tag{B.7}$$

The subsequent value of β is chosen to the left or right of the current value depending on whether the sum of the two-way distribution obtained as the limit of (B.7) is less or greater than 1.

In the case $\hat{\eta}_{(m)} < \eta$ one can use an initial approximation

$${\pi_{ij}}^{(0)} = \hat{\pi}_{i.}(\hat{\pi}_{ij(m)})^a / \sum_{i,j} (\hat{\pi}_{ij(m)})^a, \tag{B.8}$$

where $a > 1$ is chosen so as to satisfy both constraints of the problem (this method will not work, however, in cases where largest values of $\hat{\pi}_{ij(m)}$ in every row all fall in the same column; in this case one can simply start from the MLE's). The initial value $\beta^{(0)}$ is then chosen as the least squares approximation to the solution of (B.3) corresponding to ${\pi_{ij}}^{(0)}$. The proposed iterative process for a fixed β is

$$\pi_{ij}^{(k+1)} = \frac{n_{ij(m)}}{n - \beta \ln(\pi_{ij}^{(k)}/\pi_{i.}^{(k)}) + \beta(1-\eta)\ln \pi_{.j}^{(k)}}, \text{ for every } (i,j). \tag{B.9}$$

Similarly to the previous case, the next value of β is chosen to the left or right of the current value depending on whether the sum of the two-way distribution obtained as the limit of (B.9) is less or greater than 1.

When β is in the vicinity of its optimal value, the processes (B.7) and (B.9) can be adjusted so as to lead directly to the solution of the optimization problem. This is achieved by changing the value of β as part of the iteration process. Specifically, after each iteration one can select new β equal to the least squares approximation to the solution of (B.3) corresponding to $\pi_{ij}^{(k)}$.

REFERENCES

1. Agresti, A. (2002), *Categorical Data Analysis*, Wiley, New York.

2. Basseville, M., and Nikiforov, I. (1994), *Detection of Abrupt Changes - Theory and Application*, Prentice Hall, Englewood Cliffs, New Jersey.

3. Bissell, A.F. (1973), "Process Monitoring With Variable Element Sizes", *Applied Statistics* **22**, 226–238.

4. Bissell, A.F. (1973), "Monitoring Event Rates using Variable Sample Element Sizes", *The Statistician* **22**, 43–58.

5 Borror, C. M., Champ, C. W. and Rigdon, S. E. (1998), "Poisson EWMA control charts", *Journal of Quality Technology* **30**, 352–361.

6 Brook, D. and Evans, D.A. (1972), "An Approach to the Probability Distribution of Cusum Run Length", *Biometrika*, **59**, 539-549.

7 Bzik, T.J. (1997), "Considerations Regarding the Statistics of Particle Counts", *J. of the Inst. of Environmental Sciences* **22**, 43–58.

8. Gokhale, D.V., and Kullback, S. (1978), *The Information in Contingency Tables*, Marcel Dekker, New York.

9. Hawkins, D. M., and Olwell, D. H. (1998), *Cumulative Sum Control Charts and Charting for Quality Improvement*, Springer, New York.

10. Kullback, S. (1959), *Information Theory and Statistics*, Wiley, New York.

11. Lai, T. L. (1995), "Sequential Changepoint Detection in Quality Control and Dynamic Systems", *J. Royal Stat. Soc. B* July/August, 29–33.

12. Lucas, J. (1982), "Combined Shewhart-Cusum Quality Control Schemes", *Journal of Quality Technology* **14**, 51-59.

13. Michelson, D. (1996), "A Data Driven Method for Calculating Limits for Particle Control Charts", *Proc. SPIE - Int. Soc. Optic. Eng* **2876**, 38-44.

14. Reynolds, M.R. and Stoumbos, Z.G. (2000), "A General Approach to Modeling Cusum Charts for as Proportion", *IIE Transactions* **32**, 515-535.

15. Reynolds, M.R. and Stoumbos, Z.G. (2001), "Monitoring a Proportion Using CUSUM and SPRT Control Charts", *Frontiers in Statistical Quality Control*, **6**, 155-175.

16. Shewhart, W.A. (1931), *Economic Control of Quality of Manufactured Product*, Van Nostrand Co., New York.

17. Skinner, K.R., Montgomery, D.R., Runger, G.C. (2003), "Process Monitoring for Multiple Count Data using Generalized Linear Model-based Control Charts", *Intl. J. of Production Res.* **41**, 1167-80.

18. Stimson WA , Mastrangelo CM (1996) Monitoring Serially-Dependent Processes with Attribute Data, *Journal of Quality Technology* **28**, 279-288

19. Telksnys, L. (ed.) (1986), *Detection of Changes in Random Processes*, Optimization Software, Inc., New York.

20. Theil, H. (1970), "On the Estimation of Relationships Involving Qualitative Variables", *Amer. J. Sociol.* **76**, 103–154.

21 Woodall, W. H. (1997), "Control Charting Based on Attribute Data: Bibliography and Review," *Journal of Quality Technology* **57**, 172-183.

22 Yashchin, E. (1989), "On Weighted Cusum Technique", *Technometrics* **31**, 321-338.

23. Yashchin, E. (1995), "Likelihood Ratio Methods for Monitoring Parameters of a Nested Random Effect Model", *J. Amer. Statist. Assoc.* **90**, 729–738.

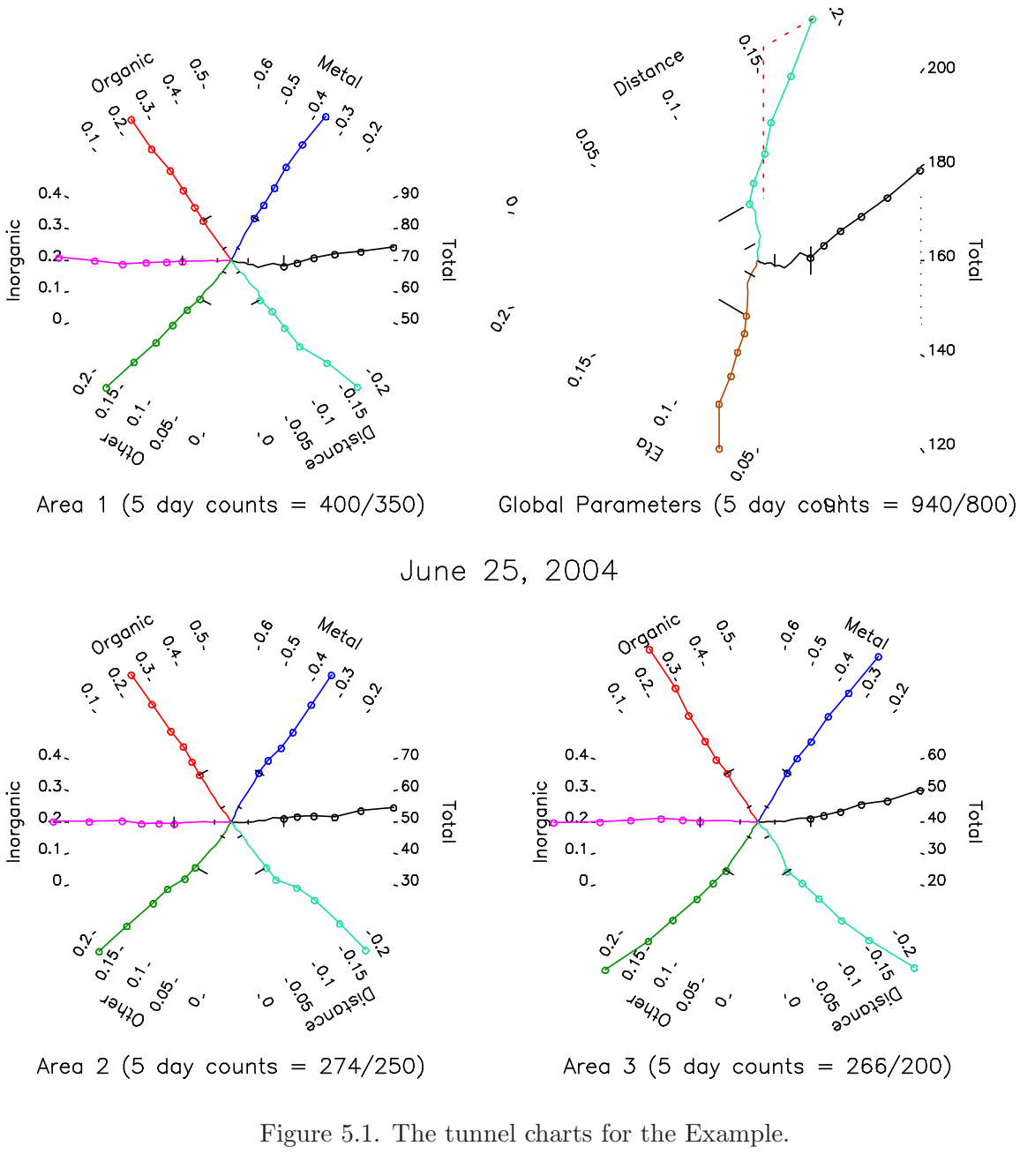

Figure 5.1. The tunnel charts for the Example.