Detection of Changes in a Multinomial Process

Marlo Brown[1]

[1] Department of Mathematics, Niagara University, New York, USA

Correspondence: Marlo Brown, Department of Mathematics, Niagara University, New York, USA. Tel: 1-716-286-8192. E-mail: mbrown@niagara.edu

Abstract

We look at a multinomial distribution where the probabilities of landing in each category change at some unknown integer. We assume that the probability structure both before and after the change is known, and the problem is to find the probability that the probability structure has changed. For a loss function consisting of the cost of late detection and a penalty for early stopping, we develop, using dynamic programming, the one and two steps look ahead Bayesian stopping rules. We provide some numerical results to illustrate the effectiveness of the detection procedures. We show that the two step ahead procedure is a slight improvement over the one step ahead procedure. However the two procedures are very consistant in their stopping times.

Keywords: Bayesian stopping rules, change-point detection, dynamic programming, multinomial, risk

1. Introduction

Detection of changes in the distribution of random variables is an active area of research with many important applications. There are several published papers dealing with the topic of detection of changes in the distribution of data. In 1926, Shewhart created control charts for means and standard deviations. In 1954, Page created continuous inspection schemes, which are process control procedures designed to detect changes in the distributions of sequences of observed data. Chernoff and Zacks (1964) investigated the Bayesian detection procedures for fixed sample sizes. See Zacks (1983) for a survey of papers on change-point problems. In 1978, Shiryaev studied optimal Bayesian procedures for detecting changes when the process is stopped as soon as a change is detected. Brown and Zacks (2006) investigated the detection of changes in a Poisson process. Zacks and Barzily (1981) extended the work of Shiryaev for the case of Bernoulli sequences. Optimal stopping rules based on Dynamic Programming procedures were discussed.

There were several papers recently published dealing with multivariate data. Brown (2008) investigated how to monitor a series of Poisson processes in several categories at once in such a way that at some unknown time point, there is a change in each of several categories. In that paper, the time point changes from category to category. Thus, we obtain k independent Poisson processes. In this paper we change the assumptions to look at the case where the probability structure changes at the same time for each category. Thus we have a multinomial process. While Zacks and Barzily studied Bernoulli sequences, we extend the number of categories to consider a multinomial sequence in k categories. The probability structure of these sequences change in several of k categories. Robbins, Lund and Gallagher (2011) discussed categorical change-points and their application to climate change. Other applications include modeling a change in the likelihood that a customer will buy one of several brands of a product, and the likelihood that a drug will have a certain effect. When there are two categories (success or failure), we have the results of Zacks and Barzily. In this paper we generalize their results for an arbitrary number of categories. Cho (2009) investigated how to determine which of several categories has the largest probability.

In the current paper we use a Bayesian approach, putting Shiryaev (1978) geometric prior on the change-point τ. In Section 2, we discuss the posterior process of calculating the probability that a change has already occurred by time n. In Section 3, we discuss the optimal stopping rule based on Dynamic Programming procedures. We assume there is a cost associated with stopping early and a cost associated with late detection. We develop optimal stopping rules that minimize the expected cost. In Section 4, we give an explicit formulation of the two step ahead stopping rule. In Section 5, we develop more user friendly calculations for the case where only one category decreases. We conclude with a numerical example showing how our method works.

2. The Bayesian Framework

We have a multinomial distribution with k categories. For each observation, the probability that the observation lands in category i is $\theta_{1,i}$ for $i = 1, ..., k$. At some unknown time point, τ, these probabilities change from $\theta_{1,i}$ to $\theta_{2,i}$ for each category. We arrange the categories in such a way that the probabilities increase for the first k_1 categories and decrease for the last $k_2 = k - k_1$ categories. Thus, $\theta_{1,i} < \theta_{2,i}$ for $i \leq k_1$ and $\theta_{1,i} > \theta_{2,i}$ for $i > k_1$. For each time point, we note the category where the observation occurs. The observations up to time n are given by $\mathbf{X_n} = \{X_1, \ldots, X_n\}$. The joint pdf of $(X_1, X_2, \ldots, X_n)$ when the change-point$= \tau$ is $f(\mathbf{X_n})$.

$$f(\mathbf{X_n}) = I(\tau \leq 1) f_2(X_1, \ldots, X_n) + \sum_{r=2}^{n} I(\tau = r) f_1(X_1, \ldots, X_{r-1}) f_2(X_r, \ldots, X_n)$$
$$+ \quad I(\tau > n) f_1(X_1, \ldots, X_n).$$

Here f_1 and f_2 are defined based on the multinomial probability before and after the change point. For $i = 1, 2$, we have $f_i(X_1, \ldots, X_r) = \prod_{j=1}^{k} \theta_{i,j}^{T_{r,j}}$. $T_{r,j} = \sum_{i=1}^{r} I(X_i = j)$ is the sum of the number of times the first r observations land in category j. To simplify notation, let $f_i^{(r,s)} = f_i(X_r, \ldots, X_s)$ and $f_i^{(r)} = f_i(X_r) = \prod_{j=1}^{k} \theta_{i,j}^{I(X_r=j)}$. Since τ is unknown, we put Shiryaev's geometric prior distribution h on the change-point τ.

$$h(\tau = i) = \begin{cases} \pi & \text{for} \quad i = 0, \\ (1-\pi) p (1-p)^{i-1} & \text{for} \quad i \geq 1. \end{cases}$$

Thus we define the posterior probability function that a change has occurred by time n as

$$\psi_n(\mathbf{X_n}, \pi) = 1 - \frac{(1-\pi)(1-p)^n f_1^{(1,n)}}{D_n(\mathbf{X_n})}. \tag{1}$$

Here $D_n(\mathbf{X_n})$ is calculated as

$$D_n(\mathbf{X_n}) = (\pi + (1-\pi) p) f_2^{(1,n)} + \sum_{r=1}^{n-1} (1-\pi) p (1-p)^r f_1^{(1,r)} f_2^{(r+1,n)} + (1-\pi)(1-p)^n f_1^{(1,n)}. \tag{2}$$

Let $\pi_n = \psi_n(\mathbf{X_n}, \pi)$ be the realization of the posterior probability that a change has occurred by time n.

Theorem 1 *The posterior probability function that a change happens by time $n + 1$, ψ_{n+1}, depends only on the posterior probability that a change happened at time n, π_n, and the category where the next observation lands. Thus, $\psi_{n+1}(\mathbf{X_{n+1}}; \pi) = \psi_1(X_{n+1}; \pi_n)$.*

Proof. The posterior probability function at time $n + 1$ is calculated according to (1) as

$$\psi_{n+1}(\mathbf{X_{n+1}}; \pi) = 1 - \frac{(1-\pi)(1-p)^{n+1} f_1^{(1,n+1)}}{D_{n+1}(\mathbf{X_{n+1}})}, \tag{3}$$

and $D_{n+1}(\mathbf{X_{n+1}})$ is calculated according to (2) as

$$D_{n+1}(\mathbf{X_{n+1}}) = (\pi + (1-\pi) p) f_2^{(1,n+1)} + \sum_{r=1}^{n-1} (1-\pi) p (1-p)^r f_1^{(1,r)} f_2^{(r+1,n+1)}$$
$$+ \quad (1-\pi) p (1-p)^n f_1^{(1,n)} f_2^{(n+1)} + (1-\pi)(1-p)^{n+1} f_1^{(1,n+1)}.$$

By (1) we have that $D_n(\mathbf{X_n})(1 - \pi_n) = (1-\pi)(1-p)^n f_1^{(1,n+1)}$. Using this substitution, we redefine (3) as $\psi_{n+1}(\mathbf{X_{n+1}}, \pi) = 1 - \frac{(1-\pi_n)(1-p) f_1^{(n+1)} D_n(\mathbf{X_n})}{D_{n+1}(\mathbf{X_{n+1}})}$, and

$$D_{n+1}(\mathbf{X_{n+1}}) = f_2^{(n+1)}(D_n(\mathbf{X_n}) - (1-\pi_n) D_n(\mathbf{X_n})) + (1-\pi_n) D_n(\mathbf{X_n})\left(p f_2^{(n+1)} + (1-p) f_1^{(n+1)}\right)$$
$$= D_n(\mathbf{X_n})\left(\pi_n f_2^{(n+1)} + (1-\pi_n) p f_2^{(n+1)} + (1-\pi_n)(1-p) f_1^{(n+1)}\right).$$

Therefore the probability that a change has taken place at time $n + 1$ is given by

$$\pi_{n+1} = \psi_{n+1}(X_{n+1}, \pi_n) = 1 - \frac{(1-\pi_n)(1-p) f_1^{(n+1)}}{\pi_n f_2^{(n+1)} + (1-\pi_n) p f_2^{(n+1)} + (1-\pi_n)(1-p) f_1^{(n+1)}}.$$

The posterior probability of change by time $n+1$ depends only on π_n and the what happens at the next observation. Suppose at the next arrival time, we observe category j. The posterior probability of a change by time $n+1$ is given by $\pi_{n+1} = 1 - \frac{(1-\pi_n)(1-p)\theta_{1,j}}{\pi_n\theta_{2,j}+(1-\pi_n)p\theta_{2,j}+(1-\pi_n)(1-p)\theta_{1,j}}$. The predicted distribution of X_{n+1} is

$$P(X_{n+1} = j) = \pi_n\theta_{2,j} + (1-\pi_n)p\theta_{2,j} + (1-\pi_n)(1-p)\theta_{1,j} \tag{4}$$

for all categories $j = 1,...,k$. After observing the observations up to time n, we compute a posterior probability of the change-point occurring by time n or at a future time $n+1, n+2, \ldots$ to be $h_n(\tau = n + i|\pi_n)$.

$$h_n(\tau = n + i|\pi_n) = \begin{cases} \pi_n & \text{for } i = 0 \\ (1-\pi_n)p(1-p)^{i-1} & \text{for } i \geq 1 \end{cases}. \tag{5}$$

3. Optimal Stopping Rule

At each time point, we calculate π_n, the probability that a change has occurred at time n. Suppose we have a cost of c_1 for stopping early and a cost of c_2 per time unit we stop late. Our goal is to develop a procedure that minimizes the expected loss. Without loss of generality, we assume $c_1 = 1$ and $c = \frac{c_2}{c_1}$. Thus c is the cost per time unit to stop late compared to the cost of stopping early. At each time we consider the cost of stopping and the cost of continuing. The cost of stopping at time n is $1 - \pi_n$. The cost of continuing at time n is given by $c\pi_n + E(R_{n+1}|\pi_n)$, where R_{n+1} is the additional risk at time $n+1$. Therefore, the risk at time n is

$$R_n(\pi_n) = \min(1-\pi_n, c\pi_n + E(R_{n+1}(\pi_{n+1})|\pi_n)). \tag{6}$$

According to (6), we opt to stop and declare there is a change whenever the risk $R_n(\pi_n) = 1 - \pi_n$. Equation (6) is simplified to

$$R_n(\pi_n) = 1 - \pi_n + [c - (1-\pi_n)(c+1) + E(R_{n+1}(\pi_{n+1})|\pi_n)]^-. \tag{7}$$

Here $[a]^- = \min(a, 0)$. According to equation (7) we stop sampling and declare a change in the probability structure when

$$\pi_n \geq \frac{1 - E(R_{n+1}(\pi_{n+1})|\pi_n)}{c+1}. \tag{8}$$

To find an expression for (8), we consider a truncated rule. We utilize the Dynamic Programming procedure developed by Zacks and Barzily (1981). We assume we must stop after observing n^* observations if we have not stopped earlier. Let j be the number of future observations allowed until we reach the truncation point. Thus $j = n^* - n$. Let $R_n^{(j)}(\pi_n)$ be the risk at time n when only j more observations are allowed. Note that $R_n^{(0)}(\pi_n) = 1 - \pi_n$, and $R_n^{(1)}(\pi_n) = 1 - \pi_n + [\pi_n(c+1) - E(\pi_{n+1}|\pi_n)]^-$.

Lemma 1 *The sequence $\{\pi_n, n \geq 1\}$ is a submartingale.*

Proof. We calculate $E(\pi_{n+1}|\pi_n)$ using the posterior distribution of a change by time $n+1$.

$$E(\pi_{n+1}|\pi_n) = \pi_n + (1-\pi_n)p > \pi_n. \tag{9}$$

The conditional expectation in (9) is taken with respect to the posterior distribution of π_{n+1} defined in (5) for all j. Thus the sequence $\{\pi_n; n \geq 1\}$ is a submartingale.

Corollary 1

$$\lim_{n\to\infty} \pi_n = 1 \quad a.s. \tag{10}$$

Proof. According to equation (9), $E\{\pi_n\} = 1 - (1-p)(1-\pi_{n-1}), n \geq 1$. Thus $E\{\pi_n\} = 1 - (1-p)^n(1-\pi)$. Hence (10).

The risk looking one step ahead is $R_n^{(1)}(\pi_n) = 1 - \pi_n + [\pi_n(c+p) - p]^-$. Therefore, the one step ahead procedure is to stop sampling and declare a change when $\pi_n > \frac{p}{c+p}$. We define $\pi^* = \frac{p}{c+p}$ to be the one step ahead boundary. Therefore the one step ahead stopping random variable is $N^{(1)} = \min\{n : \pi_n > \pi^*\}$. Now we look at j steps ahead stopping rules. Define for $j \geq 0$, $M_n^{(j)}(\pi_n)$ recursively as the expected risk of continuing at the next observation.

$$M_n^{(j)}(\pi_n) = E\left[\left[\pi_{n+1}(c+p) - p + M_{n+1}^{(j-1)}(\pi_{n+1})\right]^- | \pi_n\right], \tag{11}$$

and $M_n^{(0)} \equiv 0$. Therefore the risk looking j steps ahead is $R_n^{(j)}(\pi_n) = 1 - \pi_n + \left[\pi_n(c+p) - p + M_n^{(j)}(\pi_n)\right]^-$. The j steps ahead procedure is to stop sampling and declare that a change has occurred the first time that $\pi_n > \pi^* - \frac{M_n^{(j-1)}(\pi_n)}{c+p}$. Define the j steps ahead boundary function $b_n^{(j)}$ as $b_n^{(j)}(\pi_n) = \min\left(\pi^* - \frac{M_n^{(j-1)}(\pi_n)}{c+p}, 1\right)$. Therefore, the j steps ahead stopping random variable is $N^{(j)} = \min\left\{n \geq 1 : \pi_n > b_n^{(j)}\right\}$.

Lemma 2 $b_n^{(j)}$ are increasing functions of j.

Proof. It suffices to show by induction that $M_n^{(j)}$ are decreasing functions of j. By definition of $M_n^{(j)}$, we have that $M_n^{(1)} \leq 0$. For $j \geq 2$,

$$
\begin{aligned}
M_n^{(j)}(\pi_n) - M_n^{(j-1)}(\pi_n) &= E\left[\left[\pi_{n+1}(c+p) - p + M_{n+1}^{(j-1)}(\pi_{n+1})\right]^- |\pi_n\right] \\
&- E\left[\left[\pi_{n+1}(c+p) - p + M_{n+1}^{(j-2)}(\pi_{n+1})\right]^- |\pi_n\right].
\end{aligned}
$$
(12)

Calculating the expected value of (12) we obtain the difference $M_n^{(j)}(\pi_n) - M_n^{(j-1)}(\pi_n)$ as

$$
\sum_{i=1}^{k} P(X_{n+1} = i)\left[\left[\psi_{n+1}(i;\pi_n)(c+p) - p + M_{n+1}^{(j-1)}(\psi_{n+1}(i,\pi_n))\right]^-\right.
$$

$$
\left. - \left[\psi_{n+1}(i,\pi_n)(c+p) - p + M_{n+1}^{(j-2)}(\psi_{n+1}(i,\pi_n))\right]^-\right].
$$

Let $I_a = \left\{i : \psi_{n+1}(i,\pi_n)(c+p) - p + M_{n+1}^{(j-a)}(\psi_{n+1}(i,\pi_n)) < 0\right\}$ for $a = 1, 2$. By the induction hypothesis, $M_{n+1}^{(j-1)}(.) \leq M_{n+1}^{(j-2)}(.)$. Thus, $I_2 \subset I_1$. We express (12) as follows

$$
\sum_{i \in I_2} P(X_{n+1} = i)\left[M_{n+1}^{(j-1)}(\psi_{n+1}(i,\pi_n)) - M_{n+1}^{(j-2)}(\psi_{n+1}(i,\pi_n))\right]
$$

$$
+ \sum_{i \in (I_1 \setminus I_2)} P(X_{n+1} = i)\left[\psi_{n+1}(i,\pi_n)(c+p) - p + M_{n+1}^{(j-1)}(\pi_{n+1}(i,\pi_n))\right] < 0.
$$

The first term is negative by the induction hypothesis, and the second term is also negative by the definition of I_1.

Since the boundary functions are increasing functions of j and bounded above by 1, by the monotone convergence theorem, there exists an optimal stopping rule defined as $b_n^{\infty}(\pi_n) = \lim_{j \to \infty} b_n^{(j)}(\pi_n)$. Thus the optimal stopping rule is to stop the first time that $\pi_n > b_n^{\infty}(\pi_n)$.

4. Explicit Formulation of the Two Step Boundary

In this section we calculate the two step ahead boundary function. First we calculate $M_n^{(1)}(\pi_n)$. By equation (11), we have

$$
M_n^{(1)}(\pi_n) = E\left[\left[\pi_{n+1}(c+p) - p\right]^- |\pi_n\right] = \sum_{i=1}^{k} P(X_{n+1} = i)\left[\psi_{n+1}(i,\pi_n)(c+p) - p\right]^-.
$$
(13)

Since $M_n^{(1)} \leq 0$, the two step ahead boundary $b_n^{(2)}(\pi_n) > \pi^*$ for all π_n. Thus $N^{(2)} \geq N^{(1)}$ with probability 1. Thus we only consider the case where $\pi_n > \pi^*$.

Lemma 3 $\psi_{n+1}(i,\pi_n) > \pi_n$ when $i \leq k_1$.

Proof. Let $X_{n+1} = i \leq k_1$. By equation (6), $\pi_{n+1} = \psi_n(i,\pi_n) = 1 - \frac{(1-\pi_n)(1-p)}{\pi_n\left(\frac{\theta_{2,i}}{\theta_{1,i}}\right) + (1-\pi_n)p\left(\frac{\theta_{2,i}}{\theta_{1,i}}\right) + (1-\pi_n)(1-p)}$. Since in category i, $\theta_{1,i} < \theta_{2,i}$. Thus $\pi_{n+1} > \pi_n$.

Therefore, the sum in equation (13) can be taken over those categories for which the probabilities decrease.

$$
M_n^{(1)}(\pi_n) = \sum_{i=k_1+1}^{k} P(X_{n+1} = i)\left[\psi_{n+1}(i,\pi_n)(c+p) - p\right]^-.
$$
(14)

$P(X_{n+1} = i)$ is calculated using the predicted distribution of X_{n+1} defined in (4). From equations (4), (5) and (14), we calculate $M_n^{(1)}(\pi_n)$ as follows

$$
M_n^{(1)}(\pi_n) = \sum_{i=k_1+1}^{k} c\left[(\pi_n + (1-\pi_n)p)\theta_{2,i} - p(1-\pi_n)(1-p)\theta_{1,i}\right]^-.
$$
(15)

The two step boundary is $b_n^{(2)}(\pi_n) = \pi^* - \sum_{i=k_1+1}^{k} \left[[\pi_n + (1-\pi_n)p](1-\pi^*)\theta_{2,i} - \pi^*(1-\pi_n)(1-p)\theta_{1,i} \right]^-$. Therefore the two step ahead boundary states that we stop sampling for the first time when $\pi_n > b_n^{(2)}(\pi_n)$.

5. The Two Step Boundary When Only One Category That Decreases

In the case where there is only one category that decreases in probability, we simplify the two step ahead boundary to make the procedure more user-friendly. This has applications to situations where the last category is "Other" or "None of the above". For example, suppose a car dealership is anticipating an increase in the sales of all their vehicles on their lot. When a customer enters the dealership, the categories could be "Buys a Sedan", "Buys a Coupe", "Buys a truck" or "Does not buy a vehicle". If an increase occurs in all vehicles, the only category that would have a decrease is the "Does not buy a vehicle" category. In this case, the two step ahead boundary becomes more simple. In calculating the two step ahead boundary, equation (15) becomes

$$\frac{M_n^{(1)}(\pi_n)}{c+p} = \left[(\pi_n + (1-\pi_n)p)\theta_{2,k}(1-\pi^*) - \pi^*(1-\pi_n)(1-p)\theta_{1,k} \right]^-.$$

Here $b_n^{(2)}$ is itself a function of the posterior probabilities π_n. We want to determine which posterior probabilities are such that $\pi_n > \pi^* - \frac{M_n^{(1)}(\pi_n)}{c+p}$. Solving this equation we obtain

$$\pi_n > \frac{\pi^* - (1-\pi^*)p\theta_{2,k} + \pi^*(1-p)\theta_{1,k}}{1+(1-\pi^*)(1-p)\theta_{2,k} + \pi^*(1-p)\theta_{1,k}}. \tag{16}$$

Thus the two step ahead boundary is $B^{(2)} = \frac{\pi^* - (1-\pi^*)p\theta_{2,k} + \pi^*(1-p)\theta_{1,k}}{1+(1-\pi^*)(1-p)\theta_{2,k} + \pi^*(1-p)\theta_{1,k}}$. The two step ahead stopping variable $N^{(2)}$ is defined as $N^{(2)} = \min\left\{ n : \pi_n > B^{(2)} \right\}$.

6. Numerical Results

We now show some numerical results to see how the procedure works. For both examples, we look at a trinomial. In example one, the probability of landing in each category before the change is even, $\frac{1}{3}$ for each category. After the change, we have a high probability of 0.8 landing in category 1, while categories 2 and 3 have a much lower probability of 0.1. The prior probability is $p = \pi = 0.01$ and the cost is 0.06. We assume the change takes place at $\tau = 10$. We calculate the one step ahead procedure to be $\pi^* = 0.1429$. The results are given in Table 1. From this table, we can see that we would stop using both the one step ahead procedure and the two step ahead procedure at time $n = 12$.

Table 1. (Example 1) The probability structure changes from $\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$ to $(0.8, 0.1, 0.1)$

Trial	T_1	T_2	T_3	π_n	$b_n^{(2)}$
1	0	0	1	0.0202	0.2557
2	0	0	2	0.0161	0.2598
3	0	0	3	0.0148	0.2610
4	0	1	3	0.0145	0.2614
5	0	2	3	0.0143	0.2615
6	1	2	3	0.0434	0.2328
7	1	3	3	0.0233	0.2527
8	1	4	3	0.0170	0.2589
9	1	4	4	0.0151	0.2607
10	2	4	4	0.0452	0.2310
11	3	4	4	0.1110	0.2061
12	4	4	4	0.2382	0.1725
13	5	4	4	0.4344	0.1429
14	5	5	4	0.1954	0.1838
15	6	5	4	0.3745	0.1429

The second simulation contains a decrease in just the last category. The probabilities of landing in each category before the change are $(0.3, 0.1, 0.6)$ and after the change are $(0.4, 0.2, 0.4)$. Using equation (16) it is we see that $B^{(2)} = 0.1575$. The results from this simulation are given in Table 2. In this simulation, the one step ahead

procedure stops at time $n = 9$, which is a false alarm. The two step ahead procedure corrects for this and stops at $n = 12$. We also see that whenever $\pi_n > b_n^{(2)}$, we also have the posterior probability greater than 0.1575. That illustrates the fact that in the case where all but one category increases, we can use equation (16) which will simplify the calculations.

Table 2. (Example 2) The probability structure changes from $(0.3, 0.1, 0.6)$ to $(0.4, 0.2, 0.4)$

Trial	T_1	T_2	T_3	π_n	$b_n^{(2)}$
1	0	0	1	0.0254	0.2505
2	1	0	1	0.0432	0.2329
3	1	0	2	0.0390	0.2371
4	1	0	3	0.0361	0.2400
5	1	1	3	0.0789	0.1997
6	1	1	4	0.0635	0.2129
7	1	1	5	0.0528	0.2234
8	1	2	5	0.1093	0.1779
9	2	2	5	0.1492	0.1610
10	2	2	6	0.1136	0.1761
11	3	2	6	0.1545	0.1587
12	4	2	6	0.2040	0.1429
13	4	2	7	0.1544	0.1588
14	5	2	7	0.2039	0.1429
15	6	2	7	0.2620	0.1429

We proceeded by running 1000 Monte Carlo simulations, using various choices for $\theta_{i,j}$. In each of the simulations we assume three categories, a prior probability of $\pi = p = .01$, and a change-point happening at $\tau = 10$. The results are given in Table 3. In this table, we compare the cost for the one step and the two step ahead procedure, we see that the two step ahead procedure has a slight improvement over the one step ahead procedure. The last column marked consistent shows how often the two procedures stop at the same time. For over half of the simulations run, the two step ahead procedure stops at the exact same spot as the one step ahead procedure.

Table 3. Compares the one and two step ahead procedures for various probability structures

$\theta_{1,1}$	$\theta_{1,2}$	$\theta_{2,1}$	$\theta_{2,2}$	1 Step Ahead			2 Step Ahead			Consistent
				Alarms	Delay 1	Cost	Alarm	Delay 1	Cost	
$\frac{1}{3}$	$\frac{1}{3}$	0.8	0.1	0.289	3.139	0.4773	0.258	3.629	0.4757	0.680
0.1	0.1	0.4	0.4	0.273	1.521	0.3643	0.267	1.524	0.3584	0.970
0.15	0.15	0.45	0.45	0.296	1.866	0.4080	0.156	2.097	0.2818	0.652
0.1	0.2	0.3	0.4	0.273	2.694	0.4356	0.230	3.315	0.4289	0.691
0.2	0.3	0.5	0.2	0.317	3.773	0.5434	0.259	4.016	0.5000	0.806
0.3	0.1	0.4	0.2	0.233	4.598	0.5089	0.176	5.266	0.4920	0.723

In Table 3, we compare the cost for the one step and the two step ahead procedure, we see that the two step ahead procedure has a slight improvement over the one step ahead procedure. The last column marked consistent shows how often the two procedures stop at the same time. For over half of the simulations run, the two step ahead procedure stops at the exact same spot as the one step ahead procedure.

7. Conclusions

We have generalized the detection procedure introduced by Zacks and Barzily to include multinomial sequences. We have shown optimal stopping rules when the probability structure changes at some unknown time point. We gave explicit calculations of both the one and two step ahead procedures. There is a slight improvement in cost of the two step ahead procedure over the one step ahead procedure. However our simulations show that over half the time, the one and two step procedures are very consistent stopping at the same time. We also provide a user-friendly variation of the two step ahead stopping bound that works when only one category has decreasing probabilities.

Acknowledgments

References

Brown, M. (2008). Monitoring a Poisson Process in Several Categories Subject to Changes in the Arrival Rates. *Statistics and Probability Letters, 78*, 2637-2643. http://dx.doi.org/10.1016/j.spl.2008.03.005

Brown, M., & Zacks, S. (2006). Note on optimal Stopping for Possible Change in The Intensity of an Ordinary Poisson Process. *Statistics and Probability Letters, 76*, 1417-1425. http://dx.doi.org/10.1016/j.spl.2006.02.011

Chernoff, H., & Zacks, S. (1964). Estimating the current mean of a normal distribution which is subjected to changes in time. *Annals of Math and Statistics, 35*, 999-1018. http://dx.doi.org/10.1214/aoms/1177700517

Cho, H. (2009). Statistical identification in multinomial models with sequential sampling. *American Journal of Mathematical and Management Sciences, 29*, 139-156.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika, 41*, 100-114. http://dx.doi.org/10.2307/2333009

Robbins, M., Lund, R. B, Gallagher, C, & Lu, Q. (2011). Changepoints in the North Atlantic Tropical Cyclone Record. *Journal of the American Statist. Association, 106*, 89-99. http://dx.doi.org/10.1198/jasa.2011.ap10023

Shewhart, W. A. (1926). Quality control charts. *Bell System Technical Journal, 5*, 593-603.

Shiryaev, A. N. (1978). *Statistical Sequential Analysis: Optimal Stopping Rules*. Translations Math. Monographs, 38, Amer. Math. Soc., Providence, R. I.

Zacks, S., & Barzily, Z. (1981). Bayes procedures for detecting a shift in the probability of success in a series of Bernoulli trials. *J. Statist. Planning and Inference, 5*, 107-119. http://dx.doi.org/10.1016/0378-3758(81)90021-5

Zacks, S. (1983). A survey of classical and Bayesian approaches to the change-point problem: fixed sample and sequential procedures for testing and estimation. *Recent Advances in Statistics* (pp. 245-269). New York: Academic Press.