A New Look at Detection of Abrupt Changes

Yajun Mei
Fred Hutchinson Cancer Research Center
Seattle, WA, USA
ymei@fhcrc.org

Abstract

Detecting abrupt changes in a stochastic system on the basis of sequential observations is a major problem of the statistical reliability theory. The traditional methods specify a required frequency of false alarms for a given pre-change distribution. However, these methods are inappropriate in the new applications such as detecting terrorism attack and intrusion attack in communication networks. The reasons are: (1) a required frequency of false alarms cannot control the risk of an attack; and (2) when there are no attacks, it is possible that the true distribution of the observations could be any of a range of "acceptable" pre-change distributions. In this paper we present a different formulation by specifying a required quickness of detection and seeking to minimize the frequency of false alarms for all possible pre-change distributions. We also offer asymptotically optimal procedures that are easy to implement.

1 Introduction

Detecting abrupt changes in a stochastic system on the basis of sequential observations is a major problem of the statistical reliability theory. Extensive research has been done during last few decades. For recent reviews, we refer readers to Basseville and Nikiforov (1993), Lai (1995), and the references therein.

The standard methods in parametric approach assume that the parameter θ in the probability density f_θ of the data is equal to some known value θ_0 before a change occurs. Some popular procedures are Shewhart's control charts, moving average control charts, Page's CUSUM procedure, and the Shiryayev-Roberts procedure. When the true θ is unknown before a change occurs, it is typical to assume that a training sample is available so that one can use the method of "point estimation" to obtain a value θ_0. However, it is well-known that the performances of such procedures are very sensitive to the error in estimating θ, see, for example, Stoumbos, Reynolds, Ryan, and Woodall (2000). Thus we need to study change-point problems for composite pre-change hypotheses, which allow a range of "acceptable" values of θ.

There are many practical situations where the need to take action in response to a change in a parameter θ is definable by a fixed threshold value. For instance, consider the surveillance of the incidence of rare health events. If the underlying disease rate is greater than some specified level, we want to detect it quickly so as to enable early intervention from a public health point of view and to avoid a much greater tragedy. Another example occurs in the applications such as detecting terrorism attack and intrusion attack in communication networks. Parameters θ associated with the activities in networks could be any of a range of "acceptable" values before the attack occurs. Moreover, the standard formulation which specifies a required frequency of false alarms cannot control the risk of the attack.

In this paper we present a different formulation by specifying a required quickness of detection and seeking to minimize the frequency of false alarms for all possible pre-change distributions. As an illustration, we will focus on the problem of detecting a change of the parameter value θ in the exponential distribution $f_\theta(x) = \theta \exp(-\theta x) I(x \geq 0)$.

2 Problem Formulation

Suppose one observes a sequence of independent exponential random variables $X_1, X_2, \ldots$. At some unknown time ν, which is usually called a change-point, the parameter value θ in the exponential distribution f_θ of X_i

changes to another value λ. In other words, $X_1, \ldots, X_{\nu-1}$ are distributed according to a "pre-change" density function f_θ and $X_\nu, X_{\nu+1}, \ldots$ are distributed according to a "post-change" density function f_λ. Denote by $P^{\nu}_{\theta,\lambda}$ and $E^{\nu}_{\theta,\lambda}$ the corresponding probabilities and expectations. We shall also use P_θ and E_θ to denote the probability measure and expectation, respectively, under which $X_1, X_2, \ldots$ are i.i.d. with density f_θ.

Mathematically, a detection procedure is defined as a stopping time τ with respect to $\{X_i\}_{i\geq 1}$. The interpretation of τ is that, when $\tau = n$, we stop taking observations at time n and declare that a change has occurred somewhere in the first n observations. The performance of a stopping time τ is usually evaluated by two criteria: the long and short Average Run Lengths (ARL). The long ARL is defined by $E_\theta \tau$. Imagining repeated applications of such procedures, practitioners refer to the frequency of false alarms as $1/E_\theta\tau$ and the mean time until a false alarm as $E_\theta \tau$. The short ARL can be defined by the following worse case detection delay, proposed by Lorden (1971),

$$\overline{E}_\lambda \tau = \sup_{\nu \geq 1} \Big(\operatorname{ess\,sup} E^{\nu}_{\theta,\lambda} [(\tau - \nu + 1)^+ | X_1, \ldots, X_{\nu-1}] \Big).$$

Note that the definition of $\overline{E}_\lambda \tau$ does not depend upon the pre-change distribution f_θ by virtue of the essential superum, which takes the "worst-possible X's before the change."

Our problem can be stated as follows: Find a stopping time τ such that the mean time until a false alarm, $E_\theta \tau$, is as large as possible for all $0 < \theta \leq \lambda$, subject to the constraint

$$\overline{E}_\lambda \tau \leq \gamma, \tag{1}$$

where γ is a given constant.

3 Asymptotically Optimal Procedures

For each θ, Page's CUSUM procedure stops when for some $k \geq 1$ the last k observations satisfy

$$\sum_{i=k}^{n} \Big((\log \lambda - \log \theta) - (\lambda - \theta) X_i \Big) \geq C_\theta.$$

In order to satisfy (1), it is well-known that $C_\theta \approx I(\lambda, \theta)\gamma$, where $I(\lambda, \theta) = \theta/\lambda - 1 - \log(\theta/\lambda)$ is the Kullback-Leibler information number. See, for example, Page 26 of Siegmund (1985).

For our problem, it is natural to consider simultaneous Page's CUSUM procedures. Since the pre-change hypothesis is a union of the individual pre-change hypothesis, the intersection-union method, see for example Berger and Hsu (1996), suggest a procedure which stops when for some $k \geq 1$ the last k observations satisfy

$$\sum_{i=k}^{n} \Big((\log \lambda - \log \theta) - (\lambda - \theta) X_i \Big) \geq I(\lambda, \theta) a \quad \text{for all } \ 0 < \theta \leq \lambda.$$

A routine calculation shows that this procedure can be written as

$$M(a, \lambda) = \inf \Big\{ n \geq a : \max_{1 \leq k \leq n-a+1} \sum_{i=k}^{n} \Big(1 - \lambda X_i \Big) \geq 0 \Big\}. \tag{2}$$

In order to implement $M(a, \lambda)$ numerically, we can express $M(a, \lambda)$ as

$$M(a, \lambda) \quad = \quad \inf \Big\{ n \geq b : \ W_{n-b} + \sum_{i=n-b+1}^{n} \Big(1 - \lambda X_i \Big) \geq 0 \Big\}, \tag{3}$$

where $b = [a], W_0 = 0$, and $W_k = \max\{W_{k-1}, 0\} + \big(1 - \lambda X_k\big)$. Since W_k can be calculated recursively, this form reduces the memory requirements at every stage n from the full data set $\{X_1, \ldots, X_n\}$ to the data set of size $b+2$, i.e., $\{X_{n-b-1}, X_{n-b}, \ldots, X_n\}$. It is easy to see that this form involves only $O(a)$ computations at every stage n.

By the definition (2), we can show that $\overline{E}_{\theta_1} M(a,\lambda) = E_{\theta_1} M(a,\lambda)$. Thus in order the study the properties (i.e., long and short ARLs) of $M(a,\lambda)$, it suffices to study $E_\theta M(a,\lambda)$ for different θ.

It is interesting to note that if $a=1$, then $M(a,\lambda)$ becomes Page's CUSUM procedure with boundary 0. In that case, $M(a,\lambda)$ can be rewritten as

$$M(1,\lambda) = \inf\Big\{n \geq 1 : X_i \leq \frac{1}{\lambda}\Big\},$$

which is one of Shewhart's control charts. Thus, the ARL of $M(1,\lambda)$ is given by

$$E_\theta M(1,\lambda) = \frac{1}{P_\theta\big(X_i \leq 1/\lambda\big)} = \frac{1}{1-\exp(\theta/\lambda)}.$$

For $a \geq 2$, it is difficult to derive the exact formula for $E_\theta M(a,\lambda)$. In Mei (2003), it was demonstrated that

$$E_\theta M(a,\lambda) \geq \exp\big(I(\lambda,\theta)a\big) \quad \text{for all } 0 \leq \theta < \lambda,$$

where $I(\lambda,\theta) = \theta/\lambda - 1 - \log(\theta/\lambda)$. The strong law of large numbers shows that if a (or θ) is large, then

$$\frac{E_\theta M(a,\lambda)}{a} \approx 1 \quad \text{for} \quad \theta > \lambda.$$

Therefore, as θ increases from 0 to ∞, the value of $E_\theta M(a,\lambda)$ decreases from ∞ to a. Moreover, since the distribution of $(\lambda/p)X_i$ under $P_{\theta/p}$ is same for all $p>0$, we have

$$E_\theta M(a,\lambda) = E_{\theta/p}\Big(M(a,\frac{\lambda}{p})\Big) = E_{\theta/\lambda} M(a,1).$$

These properties can help us to select a and λ to achieve a prescribed $E_{\theta_1} M(a,\lambda) \approx \gamma$. We can first choose a to be the maximal number of observations allowed from the post-change distribution f_λ with large λ. Next we can use Monte Carlo methods to find θ_0 such that $E_{\theta_0} M(a,1) \approx \gamma$. Finally, the choice of $\lambda = \theta_1/\theta_0$ will lead to $E_{\theta_1} M(a,\lambda) \approx \gamma$ and $E_\theta M(a,\lambda) \approx a$ for large θ.

4 Numerical Results

We now describe the results of a Monte Carlo experiment designed to check the performance of our procedure $M(a,\lambda)$. Table 1 reports the sampled average run lengthes of $M(a,\lambda)$ for $a=5$ and $\lambda=1$. We ran 10,000-repetitions to simulate $E_\theta M(a,\lambda)$ for different θ. Each result in Table 1 is recorded as the Monte Carlo estimate $\pm$ standard error.

If we want to design a procedure $M(5,\lambda)$ satisfying $E_{\theta_1=1} M \approx 20$, then Table 1 tells us that $E_{\theta_0 \approx 0.5} M(5,1) \approx 20$. Thus $\lambda = \theta_1/\theta_0 = 2$ will be desired values. Furthermore, $E_{\theta=0.4} M(5,2) = E_{0.2} M(5,1) \approx 377$ while $E_{\theta=0.3} M(5,2) = E_{0.15} M(5,1) \approx 1191$.

Table 1: Average run length (sampled)

	Values of θ										
	0.15	0.2	0.25	0.3	0.4	0.5	0.7	0.8	1	1.5	2
$E_\theta M(5,1)$	1191	377	165	89	37.02	20.94	10.76	8.69	6.71	5.28	5.05
	±12	±4	±2	±1	±0.33	±0.17	±0.07	±0.05	±0.03	±0.01	±0.00

5 Conclusion

We have studied a change-point problem where the pre-change distribution involves unknown parameters. In the problem of detecting a change of the parameter value θ in exponential distribution f_θ, we have presented a new formulation by specifying a required average time to detection after the value of θ shifts to a specified λ while minimizing the frequency false alarms over a range of possible values of θ before a change occurs. We have also proposed asymptotically optimal procedures which are easy to implement.

Acknowledgements

This work was supported in part by the National Institutes of Health under Grant R01 AI055343.

References

Basseville, M. and Nikiforov, I. (1993). *Detection of Abrupt Changes: Theory and Applications.* Englewood Cliffs, Prentice-Hall.

Berger, R. L. and Hsu, J. C. (1996). Bioequivalence trials, intersection-union tests and equivalence confidence sets (with discussion). *Statist. Sci.* 11, 283-319.

Lai, T. L. (1995). Sequential change-point detection in quality control and dynamical systems. *J. Roy. Statist. Soc. Ser. B* 57, 613-658.

Lorden, G. (1971). Procedures for reacting to a change in distribution. *Ann. Math. Statist.* 42, 1897-1908.

Mei, Y. (2003). Asymptotically optimal methods for sequential change-point detection. *Ph.D. thesis*, California Institute of Technology. Available at <http://resolver.caltech.edu/CaltechETD:etd-05292003-133431> .

Siegmund, D. (1985). *Sequential Analysis: Tests and Confidence Intervals.* New York: Springer-Verlag.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika* 41, 100-115.

Stoumbos, Z., Reynolds, M. R. Jr., Ryan, T. P. and Woodall, W. H. (2000). The state of statistical process control as we proceed into the 21st century. *J. Amer. Statist. Assoc* 95, 992-998.