Change-Point Problems
IMS Lecture Notes – Monograph Series (Volume 23, 1994)

AN EFFICIENT NONPARAMETRIC DETECTION SCHEME AND ITS APPLICATION TO SURVEILLANCE OF A BERNOULLI PROCESS WITH UNKNOWN BASELINE

BY CHARLES BELL, LOUIS GORDON, AND MOSHE POLLAK

San Diego State University, University of Southern California, and Hebrew University of Jerusalem

Suppose $X_1, X_2, \cdots$ are independent observations, having an unknown continuous initial distribution, possibly becoming stochastically larger after an unknown point of time ν. We develop a nonparametric detection scheme which has 100% asymptotic relative efficiency for detecting a change of scale of exponential observations. We apply the scheme to detection of change of the residual variance in a regression model. We use the scheme to construct a very effective method for detecting a change of the success probability p of a Bernoulli process with unknown baseline.

1. Introduction. Suppose observations accumulate sequentially, and one is on the lookout for their becoming stochastically larger (or smaller). If the initial distribution F_0 is known, then one can apply classical surveillance methodology, such as Shewhart (1931), Cusum (Page, 1954) or Shiryayev (1963) and Roberts (1966) control charts. However, in many practical situations, the baseline F_0 is not known at the onset of surveillance. For instance, if one were interested in monitoring the thickness of the ozone layer in the atmosphere over a certain region, previous knowledge may not be available to provide a baseline. Recently, the Shiryayev-Roberts approach has been successfully employed to deal with such situations. Pollak and Siegmund (1991) studied detection of a change of a normal mean when the initial mean is unknown. Gordon and Pollak (1990) provide a theorem for evaluating the ARL to false alarm for a wide class of surveillance problems which admit an invariance structure. In addition, they study (1989, 1991) nonparametric procedures for detecting a change of stochastic order. These procedures are highly effective for detecting a change of the mean in a sequence of normally or approximately normally distributed observations. Their importance is three-fold: they enable

AMS 1991 Subject Classification: Primary 62L10, Secondary 61N10

Key words and phrases: Change-point, cusum, Shiryayav-Roberts, average run length.

a nonparametric treatment in case little is known about the distribution of the observations, they do not require knowledge of baseline center and scale, and they constitute robust methods in a wide variety of cases where a parametric family is thought to be approximately appropriate. It should be noted that the ARL to false alarm of parametric surveillance methods is notoriously sensitive to misspecification of distribution characteristics (cf. van Dobben de Bruyn, 1968, Section 2.3).

Though highly efficient (well above 90% relative efficiency) for detecting a change in the mean of a normal distribution, the aforementioned nonparametric schemes are not generally as efficient for detecting a change in the scale of exponential-like distributions, as shown by Table 1. (We will describe in Section 3 how the figures in Table 1 are obtained.)

Table 1: Asymptotic relative efficiencies of Gordon and Pollak's (1991) NPSRI scheme geared to detect a change from an $\exp(\theta)$ to an $\exp(\gamma\theta)$ distribution (γ fixed, θ unknown) with respect to the parametric Cusum scheme (γ fixed, θ unknown), when the post-change distribution actually is $\exp(\gamma\theta)$.

γ	0	$\frac{1}{3}$	$\frac{1}{2}$	≈ 1	2	3
A.R.E.	1	.9994	.9986	.9940	.9759	.9499
γ	5	10	100	1000	10000	∞
A.R.E.	.8926	.7977	.7426	.7738	.8014	1

It is apparent from Table 1 that Gordon and Pollak's (1991) NPSRI scheme is asymptotically very efficient for detecting a change from $\exp(\theta)$ to $\exp(\gamma\theta)$ when $\gamma < 1$. It is also fairly efficient for $\gamma > 1$ if γ is not large. When γ is moderately large (or larger) – and we will argue in Section 4 that this case is of applied interest – efficiency decreases.

Here we develop a nonparametric surveillance scheme which has a 100% asymptotic relative efficiency for detecting a γ-fold change in the parameter of an exponential distribution. This procedure is simpler than the NPSRI, both conceptually and technically. Because of this, and because of its higher asymptotic relative efficiency in the exponenetial case, one would prefer this procedure whenever observations are close to being exponenetial, such as geometric observations or certain gamma-like observations. We develop operating characteristics of this scheme. We apply the scheme to detection of a change of the residual variance in a regression model. We also adapt this procedure to the problem of detecting a change in the success probability p of a Bernoulli sequence by regarding the geometrically distributed times between successes as the basic observations. This application is especially useful because it enables a treatment of the problem without knowledge of the baseline p. We find that the procedure is very efficient, even when the initial value of p is not small.

2. The Procedure and its ARL to False Alarm. We follow Gordon and Pollak's (1989, 1991) approach in constructing a Shiryayev-Roberts type of stopping rule.

Our basic setup is one where a sequence of independent random variables $\{X_i\}_{1\le i<\infty}$ is being observed sequentially, such that $X_1,\ldots,X_{\nu-1} \sim F_0$ and $X_\nu, X_{\nu+1},\cdots \sim F_1$. None of F_0, F_1, ν is known, but it is assumed that F_0 and F_1 are continous and the $X_i, \nu \le i < \infty$, are stochastically larger than the $X_j, 1 \le j < \nu$. Let r_n be the nth sequential rank; i.e. $r_n = \sum_{i=1}^n 1(X_i \le X_n)$. We will base the surveillance on the sequence $r_1, r_2, \cdots$.

Let P_k and E_k denote probability and expectation under this setup when $\nu = k$. When throughout the sequence there is no change, probability and expectation will be denoted by P_∞ and E_∞. The Shiryayev-Roberts approach calls for computing the sequence of likelihood ratios

$$\Lambda_k^n = \frac{P_k(r_1, r_2, \cdots, r_n)}{P_\infty(r_1, r_2, \cdots, r_n)}$$

and the sequence of statistics

$$R_n = \sum_{k=1}^n \Lambda_k^n$$

and declaring at

$$N_A = \min\{n | R_n \ge A\}$$

that a change is in effect. The technical difficulty lies in computing Λ_k^n: although the denominator of Λ_k^n is obviously $1/n!$ whatever F_0 be, computation of the numerator is obviously much more complex.

Gordon and Pollak (1989, 1991) proposed that one should try to find any two distributions F_0^*, F_1^* such that if $F = F_0^*$ and $F_1 = F_1^*$, then it is possible to compute the Λ_k^n's. The Λ_k^n's (hence the R_n's and N_A) will thus become well-defined statistics, their values depending on the values of the X_i's only through their sequential ranks. One may study the distribution of the sequence $R_1, R_2, \cdots$ and of N_A even if in reality the true pre-change and post-change distributions differ from F_0^*, F_1^*. This distribution is essential for computation of $E(N_A|\nu = \infty)$, the average run length (ARL) to false alarm, which is the standard index for quantifying the propensity for false alarms of a detection scheme. The crucial point in the argument is that when there is no change ($\nu = \infty$), the distribution of $R_1, R_2, \cdots$ is the same for all F_0's: even if F_0 differs from F_0^*, the increasing transformation $(F_0^*)^{-1}F_0$ applied to the observations would transform them into F_0^*-distributed observations without changing their ranks. Hence, whatever the true F_0 be, the ARL to false alarm will be the same as when the true pre-change distribution is F_0^*. Therefore, without loss of generality, the ARL to false alarm can be computed under the assumption that $F_0 = F_0^*$.

As mentioned in the introduction, Gordon and Pollak (1989, 1991) chose F_0^*, F_1^* with densities $f_0^*(x) = \exp\{-|x|/2\}/2$, $f_1^*(x) = p\alpha\exp\{-\alpha x\}1(x > 0) + (1-p)\beta\exp\{\beta x\}1(x < 0)$, because they had in mind the problem of detecting a shift of mean of normal-like observations. Here we are contemplating detection of a change of scale of exponential-like observations. Our choice is therefore

$$f_0^*(x) = \exp\{-x\}1(x > 0)$$
$$f_1^*(x) = \alpha\exp\{-\alpha x\}1(x > 0)$$

where $\alpha \neq 1$ is a fixed (known) constant. (f_1^* is a representative of the post-change distribution. We will deal later with specification of α.) To see that this works, we need the following lemma (which can be proven directly or by induction):

LEMMA 2.1 (Savage, 1956). *Let $Y_1, Y_2, \cdots$ be* $\exp(1)$ *iid random variables. Let $x_1, x_2, \cdots$ be a sequence of positive constants. Then*

$$P\left(\frac{Y_1}{x_1} < \frac{Y_2}{x_2} \cdots < \frac{Y_n}{x_n}\right) = \prod_{i=1}^{n} \frac{x_i}{\sum_{j=i}^{n} x_j}.$$

Now regard the numerator of Λ_k^n:

$$\begin{aligned} P_k(r_1,\cdots,r_n) &= P_k\left(X_{\substack{\text{index of}\\ \text{smallest}\\ \text{obs}}} < X_{\substack{\text{index of}\\ \text{2nd smallest}\\ \text{obs}}} < \cdots < X_{\substack{\text{index of}\\ \text{largest}\\ \text{obs}}}\right) \\ &= P\left(\frac{Y_1}{1 \text{ or } \alpha} < \frac{Y_2}{1 \text{ or } \alpha} < \cdots < \frac{Y_n}{1 \text{ or } \alpha}\right) \\ &= \prod_{i=1}^{n} \frac{(1 \text{ or } \alpha)_i}{\sum_{j=i}^{n}(1 \text{ or } \alpha)_j} \end{aligned}$$

where $(1 \text{ or } \alpha)_i$ equals 1 or α depending on whether the serial index number of the ith smallest observation is less than k or not. Formally, denote $\rho(i,n) = \sum_{j=1}^{n} 1(x_j \le x_i)$ and define the inverse permutation $\tau(\cdot, n)$ via

$\rho(\tau(i,n),n) = i$ for $i = 1,\cdots,n$. Define

$$\gamma(j,k) = \begin{cases} 1 & \text{if } j < k \\ \alpha & \text{if } j \geq k. \end{cases}$$

Thus

$$P_k(r_1,\cdots,r_n) = \prod_{i=1}^{n} \frac{\gamma(\tau(i,n),k)}{\sum_{j=i}^{n}\gamma(\tau(j,n),k)} = \frac{\alpha^{n-k+1}}{\prod_{i=1}^{n}\sum_{j=i}^{n}\gamma(\tau(j,n),k)}$$

and so

$$\Lambda_k^n = \frac{a^{n-k+1}}{\prod_{i=1}^{n}[\frac{1}{n-i+1}\sum_{j=i}^{n}\gamma(\tau(j,n),k)]}.$$

As mentioned above, our stopping time is

$$N_A = \min\{n \mid R_n \geq A\} \tag{1}$$

where

$$R_n - \sum_{k=1}^{n}\Lambda_k^n.$$

The P_∞-characteristics of N_A are summarized in Theorem 2.2, whose proof is given in Section 6.

THEOREM 2.2. *Let $\alpha \neq 1$, $0 < \alpha < \infty$ be a specified parameter and let N_A be as in (1). If, when $\nu = \infty$, $X_1, X_2, \cdots$ are iid and their distribution is continuous, then*

$$E_\infty N_A \geq A$$

and

$$\triangle \stackrel{\text{def}}{=} \lim_{A\to\infty} E_\infty N_A/A = \begin{cases} \frac{1}{\alpha} & \text{if } \alpha < 1 \\ \frac{\alpha\log\alpha-\alpha+1}{\alpha-1-\log\alpha} & \text{if } \alpha > 1. \end{cases}$$

We are now ready to describe our procedure, which we will denote by NPSRE (for Non Parametric Shiryayev-Roberts geared to the Exponenetial distribution).

The NPSRE Procedure. Suppose observations $X_1, X_2, \cdots$ are independent, initially having an identical continuous distribution, and that at an unknown time ν they become stochastically larger (or smaller). Suppose further that the requirement of a monitoring procedure for change-point detection is that its ARL to false alarm be no less than a prespecified bound B. The NPSRE scheme requires that the statistician

1. Compute the statistic R_n after each observation.

2. Stop and declare after the first time that R_n exceeds $A = B/\triangle$ (where $\triangle$ is defined in Theorem 2.2) that a change is in effect. (A conservative choice for A is $A = B$.)

We discuss the choice of the tuning parameter α in the next section.

Note that the computations involved in implementing the NPSRE procedure are conveniently programmable. The following (Figure 1) was programmed in the MATLAB programming language, and is the basic code used in the application in Section 4. See Math Works (1989) for a description of the MATLAB language.

3. Speed of Detection, Choice of Parameter and Relative Efficiency. Suppose the change is at time ν and no false alarm was raised. If N is the stopping time used to monitor the process, the lag in detecting the change is $N-(\nu-1)$. We adopt $E_\nu(N-\nu+1|N \geq \nu)$ as a basic index of the speed of detection. (See Lorden, 1971, for a different index.)

Because ν is unknown, one's summary index of the speed of detection has to be a functional of the sequence $\{E_k(N-k+1|N \geq k)\}; k = 1, 2, \cdots$. Note that the case $\nu = 1$ cannot be differentiated from the case $\nu = \infty$. Therefore, $E_1 N = E_\infty N \approx B$. One would expect $E_k(N-k+1|N \geq k)$ to be comparable to B when k is close to 1. In other words, one cannot expect the procedure to detect an early change quickly. We follow Roberts (1966) in choosing $\lim_{k\to\infty} E_k(N-k+1|N \geq k)$ as our primary index of expected lag. Frequently, $E_k(N-k+1|N \geq k)$ is well-approximated by its limit as $k \to \infty$, even for values of k which are small relative to B (cf. Gordon and Pollak, 1989 and 1991; Pollak and Siegmund, 1991).

Suppose that $G_0(x)$ is the real (continuous) initial c.d.f. of the observations, and that $G_1(x)$ (continuous) is the c.d.f. of the observations after a change. Without affecting the ranks, one can transform the observations to make their distribution prior to a change exp(1). The transformation is

$$Q(x) = -\log(1 - G_0(x)). \tag{2}$$

With this notation, we will state a result concerning the speed of detection of the NPSRE.

THEOREM 3.1. *Let G_0 and G_1 be distributions as specified above where it is known that $G_0 \geq G_1$ (or alternatively $G_0 \leq G_1$), and let f_0^* and f_1^* be as in Section 2. Define*

$$\xi = \log\alpha + (1-\alpha)\int_{-\infty}^{\infty} Q(x)dG_1(x) \tag{3}$$

where the transformation $Q(x)$ is specified in (2) and where $\alpha < 1$ (or $\alpha > 1$ if $G_0 \leq G_1$). Denote by B_ε the set $(4\varepsilon, -\log(6\varepsilon))$. Let $\nu(A)$ and $\varepsilon(A)$ be functions such that as $A \to \infty$

(i) $(\log A)P_1\{Q(X_1) \notin B_{\varepsilon(A)}\} \to 0$

(ii) $(\log A)\varepsilon(A)\log(\varepsilon(A)) \to 0$

(iii) $\nu(A)\varepsilon^4(A)/\log A \to \infty$.

If $0 < \xi < \infty$, *then*

$$\lim_{A\to\infty} \sup \sup_{\nu \geq \nu(A)} \frac{E_\nu(N_A - \nu + 1 \mid N_A \geq \nu)}{\log A} \leq \frac{1}{\xi}.$$

The proof of Theorem 3.1 is analogous to that of Theorem 3 of Gordon and Pollak (1991) and will be omitted. Note that Theorem 3.1 gives only an upper bound to the expected lag. We conjecture that this upper bound is a limit, but have not been able to prove it for change-points whose magnitude is very large.

By virtue of Theorem 3.1, if G_0 and G_1 are the suspected before and after change distributions, then one should choose α so as to minimize ξ. Differentiating (2) yields

$$\alpha = 1/\int_{-\infty}^{\infty} Q(x)dG_1(x).$$

For example, if $G_0 = \exp(\theta)$ and $G_1 = \exp(\gamma\theta)$ then $\alpha = \gamma$. As another example, if $G_0 = N(0, \sigma^2)$ and $G_1 = N(\delta\sigma, \sigma^2)$, then $\alpha = 1/\int_{-\infty}^{\infty} -\log(1 - \Phi(x))\varphi(x - \delta)dx$. For $\delta = 1$ numerical evaluation yields $\alpha = .45$.

We obtain a measure of the asymptotic relative efficiency (ARE) of the NPSRE procedure as $A \to \infty$ by comparing $\xi^{-1}\log A$ to $\lim_{k\to\infty} E_k(N^P - k + 1|N^P \geq k)$ where N^P is the stopping time which would have been used in the case that G_0 and G_1 were definitely known to be the pre-change and post-change distributions, and N^P is designed to have the same false alarm rate as N_A (i.e.: $E_\infty N^P = E_\infty N_A$). By the results of Lorden (1971) (or Pollak, 1985), $\lim_{k\to\infty} E_k(N^P - k + 1|N^P \geq k) = (\log A)/\int_{-\infty}^{\infty} \log((dG_1/dG_0)(x))\, dG_1(x) + O(1)$, where $O(1)$ is bounded as $A \to \infty$. Therefore, the ARE of the NPSRE relative to both Cusum and Shiryayev-Roberts parametric procedures is

$$ARE = \lim_{A\to\infty} \frac{(\log A)/\int_{-\infty}^{\infty} \log((dG_1/dG_0)(x))\, dG_1(x) + O(1)}{\xi^{-1}\log A + O(1)}$$
$$= \frac{\xi}{\int_{-\infty}^{\infty} \log((dG_1/dG_0)(x))\, dG_1(x)}.$$

When $G_0 = \exp(\theta)$, $G_1 = \exp(\gamma\theta)$ and $\alpha = \gamma$, the ARE of the NPSRE procedure is 100%. Therefore, if the observations are approximately exponential both before and after a change, the NPSRE procedure is a robust surveillance scheme, and would be better than Gordon and Pollak's (1991) NPSRI scheme (whose ARE is given in Table 1).

In the case $G_0 = N(0, \sigma^2)$, $G_1 = N(\sigma, \sigma^2)$, the ARE of the NPSRE (with $\alpha = .45$) is 85%. So, if the problem is detecting a change in a normal mean, the NPSRI scheme (which has an ARE of 97%) is preferable to the NPSRE.

4. An Application to Mass Calibration. One of the activities of the National Institute of Standards and Technology (NIST) is precision measurement of mass standards. Mass standards are calibrated at the NIST by comparison measurements which relate the mass of a client's standard to the NIST standard kilograms. The NIST has a large stake in monitoring its calibration process to ensure its validity and the validity of NIST statements regarding the process. Surveillance is maintained by a series of check standards which are calibrated with the client's weights. The kilogram level is the critical level in the calibration process because weights of higher and lower denominations are calibrated relative to the NIST kilograms through a series of intercomparison designs. The calibration design involves 6 intercomparison measurements: y_1 = the difference between the two NIST's 1 kg standards, y_2 = the difference between one of the NIST's two 1 kg standards and the client's, y_4 = the difference between the other of the NIST's two 1 kg standards and the client's, y_3 = the difference between the one of the NIST's two 1 kg standards and the sum of the client's 500, 300, 200 gr standards, y_5 = the difference between the other of the NIST's two 1 kg standards and the sum of the client's 500, 300, 200 gr standards, and y_6 = the difference between the client's 1 kg standard and the sum of the 500, 300, 200 gr standards. One can write $y_1 = \mu_1 + \epsilon_1$, $y_2 = \mu_2 + \epsilon_2$, $y_3 = \mu_3 + \epsilon_3$, $y_4 = \mu_2 - \mu_1 + \epsilon_4$, $y_5 = \mu_3 - \mu_1 + \epsilon_5$, $y_6 = \mu_3 - \mu_2 + \epsilon_6$ where the ϵ_j are independent and identically distributed. The standard assumption is that the ϵ_j have a $N(0, \sigma^2)$ distribution, σ unknown. Least squares estimates $\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3, \hat{\sigma}$ of $\mu_1, \mu_2, \mu_3, \sigma$ are easily obtained. Note that under the normality assumption, $3\hat{\sigma}^2/\sigma^2 \sim \chi^2_{(3)}$. Croarkin, Hagwood and Pollak (1993) applied Shiryayev-Roberts control charts to a sequence of $\hat{\mu}_1$'s and $\hat{\sigma}$'s. Their analysis of the $\hat{\sigma}$ sequence is based on the parametric assumption of normality. In this section we will present a nonparametric analysis. The data are based on check standard determinations made at a sequence of 217 (nonequally spaced) time points between 1975 and 1988, and are given in Table 2. They are plotted in Figure 2.

Table 2: Estimates in Milligrams of the Standard Deviations of Weight Difference Measurements, made at the NIST between 1975 and 1988. The figures are to be read from top to bottom.

.0217	.0339	.0275	.0179	.0182	.0321	.0260	.0307	.0102
.0118	.0250	.0376	.0423	.0173	.0162	.0291	.0237	.0445
.0232	.0391	.0482	.0228	.0305	.0296	.0415	.0381	.0454
.0210	.0365	.0290	.0211	.0340	.0663	.0267	.0279	.0240
.0265	.0164	.0326	.0256	.0392	.0266	.0326	.0459	.0409
.0317	.0203	.0338	.0170	.0407	.0241	.0147	.0142	.0570
.0194	.0274	.0381	.0221	.0367	.0204	.0578	.0504	.1492
.0316	.0317	.0373	.0196	.0124	.0140	.0355	.0310	.0469
.0274	.0338	.0245	.0192	.0387	.0233	.0323	.0241	.0275
.0361	.0399	.0212	.0237	.0207	.0222	.0323	.0349	.0180
.0362	.0399	.0194	.0294	.0094	.0267	.0582	.0233	.0106
.0320	.0275	.0216	.0422	.0430	.0267	.0309	.0175	.0252
.0096	.0362	.0322	.0315	.0167	.0117	.0102	.0539	.0523
.0238	.0177	.0482	.0424	.0167	.0500	.0521	.0349	.0275
.0224	.0356	.0165	.0303	.0351	.0102	.0459	.0246	.0376
.0117	.0424	.0272	.0159	.0412	.0267	.0514	.0503	.0215
.0175	.0487	.0336	.0289	.0134	.0306	.0498	.0206	.0403
.0314	.0413	.0452	.0363	.0143	.0147	.0191	.0388	
.0445	.0463	.0459	.0423	.0226	.0324	.0191	.0332	
.0122	.0431	.0359	.0140	.0320	.0079	.0483	.0226	
.0132	.0275	.0251	.0132	.0373	.0143	.0309	.0456	
.0409	.0724	.0254	.0290	.0252	.0398	.0590	.0156	
.0206	.0297	.0243	.0132	.0454	.0205	.0427	.0377	
.0295	.0054	.0284	.0251	.0226	.0205	.0590	.0471	
.0391	.0340	.0094	.0234	.0241	.0247	.0486	.0592	

It is of interest to detect a change in the standard deviation - both an increase and a decrease should be detectable. We chose to represent an increase by a change from σ to 2σ, and a decrease by a change from σ to $\sigma/2$. If we were to choose an NPSRI statistic (Gordon and Pollak, 1991), then the optimal choice of parameters is .2056, 1.2439, .8984 for α, β, p when monitoring for an increase and 4.7435, .5316, .0237 for α, β, p when monitoring for a decrease; the ARE's (relative to an underlying normal distribution of the observations) are .9985 and .8828 for increase and decrease. If we choose the NPSRE method of this paper, calculation of the optimal choice of α and the ARE by the formulae of Section 3 yields α = .1992 and ARE = .9939 for detecting an increase and $\alpha = 5.9207$ and ARE $= .9926$ for a decrease. The difference in ARE is negligible when monitoring for an increase, but the NPSRE clearly has an advantage when monitoring for a decrease. Since we are interested in detecting any change, our statistic is $R_n = \frac{1}{2}(R_n(\alpha = .1992) + R_n(\alpha = 5.9207))$ (where $R_n(\alpha = x)$ is the R_n of Section 2 with x as the parameter value). By Theorem 2.2, $\Delta_{\alpha=.1992} = 4.6703$ and $\Delta_{\alpha=5.9207} = 1.8415$. By appealing to Theorem 2 of Pollak (1987), we would expect that for our (two-sided) procedure $E_\infty N_A/A \approx 1/(.5/\Delta_{\alpha=.1992} + .5/\Delta_{\alpha=5.9207}) = 2.6415$. For instance, if we'd require $E_\infty N_A \geq 370$ (the ARL to false alarm of the 2-sided Shewhart scheme), we would choose $A = 370/2.6415 = 140$. A plot of R_n for the data of Table 2 is given in Figure 3. (For diagnostics regarding underlying assumptions, see Croarkin, Hagwood and Pollak, 1993. Among other things, there is no evidence of serial correlation.)

For the sake of comparison, we bring the parametric analog in Figures 4 and 5 (see Croarkin, Hagwood and Pollak, 1993). Here $E_\infty N_A$ is about the same as in the nonparametric scheme for $A = 140$. (The parametric analog is a 2-sided scheme for a change assuming normality and representing an increase by a change from σ to 2σ and a decrease by a change from σ to $\sigma/2$.) For an ARL to false alarm of 370, the NPSRE scheme would have stopped just a little before the parametric scheme (at the 47th observation). The difference between the schemes would be greater for larger ARL to false alarm; the NPSRE discerns the change just before the 50th observation as a much clearer one. Also, the NPSRE is much less sensitive to the outlier 207th observation.

5. Surveillance of a Bernoulli Process with Unknown Baseline. A problem often encountered in practice is surveillance for an increase in the success probability p in a sequence of Bernoulli trials. For example, when conducting surveillance of congenital malformations in newborn infants, one may monitor for a rise in the incidence rate of a particular malformation. As another example, consider the percentage of washing machines requiring service under the terms of a 3-year guarantee. The company manufacturing the machine may be on the lookout for a deterioration (or an improvement) of the service percentage. If the congenital malformation being monitored is of a newly discerned type (for which no data have been collected in the past), or if surveillance is started on a regional basis (where existing data do not specify regionality), then the baseline incidence rate is unknown. Nonetheless, surveillance for an increase in the incidence rate may be of interest. (The same applies to the washing machine example.) The technical difficulty is that classical procedures do not allow an unknown baseline, and Pollak and Siegmund's (1991) approach will not work (due to Bernoulli variables lacking an invariance structure).

Here we will primarily regard the problem of surveillance for an increase in the success probably p. In this case, it is quite clear that one will raise alarms only right after a success occurs. (One would not conceive raising an alarm right after the birth of a healthy baby.) Therefore, one may regard the geometrically distributed number of Bernoulli trials between successes as one's basic set of observations.

When p is small – as is often the case; e.g. surveillance of congenital malformations – the geometric (p) distribution can be approximated by an exp(p) distribution. When p is not very small – as may be the case in the washing machine example – the quality of the approximation deteriorates. Since, as mentioned in Section 1, the ARL to false alarm is very sensitive to misspecification of the baseline distribution, employing parametric surveillance schemes based on the exponential distribution may be undesirable.We propose to employ a NPSRE scheme instead. We show that the ARE of the NPSRE relative to the schemes based on the geometric distribution with known baseline is very high, even for relatively large values of p.

In order to apply the NPSRE, the observations must be continuous. To overcome the discreteness of the geometric distribution, we propose that a $U(0,1)$-distributed value be subtracted from each geometric observation, independently for each observation. (This may be interpreted as a way of dealing with ties between observations.) We obtain Theorem 5.1, whose proof is deferred to Section 6.

THEOREM 5.1 *When monitoring a sequence of Bernoulli trials with an unknown success probability p for an increase of the success probability, the asymptotic relative efficiency (ARE) of the NPSRE scheme with parameter $\alpha > 1$ (applied to the geometric–$U(0,1)$ observations described above) relative to the parametric Cusum scheme monitoring a sequence of geometric(p) observations for a change to geometric(αp) with known $p(< 1/\alpha)$ is*

$$ARE = \frac{\log\alpha + (1-\alpha)(1 + (1-\frac{1}{\alpha})\frac{\log(1-p)}{p})}{\log\alpha + (\log\frac{1-\alpha p}{1-p})(\frac{1}{\alpha p} - 1)}.$$

(The theorem is also valid for $\alpha < 1$ - when the change is a decrease – but in this case there is no justification to base the procedure on the geometric observations.)

Figure 6 indicates that the ARE is very high for most practical purposes. For example, when using a NPSRE with $\alpha = 2$ for monitoring for a doubling of the success rate, even at $p = .25$ the ARE is above 90%.

6. Proofs.

PROOF OF THEOREM 2.2. To prove that $E_\infty N_A \geq A$, note that for fixed k, Λ_k^n is a P_∞ - martingale with unit expectation, so that $R_n - n$ is a P_∞ - martingale with zero expectation. Apply the optional sampling theorem to obtain $E_\infty(R_{N_A} - N_A) = 0$. Hence $E_\infty N_A = E_\infty R_{N_A}$. But by definition $R_{N_A} \geq A$. Hence $E_\infty N_A \geq A$.

To prove the second part of Theorem 2.2, we need to validate Conditions 1A, 1B and 1C of Theorem 1 of Gordon and Pollak (1990). By virtue of this theorem, the values of $\triangle$ appearing in the theorem are a result of the memorylessness of the exponential distribution when $\alpha < 1$ and standard renewal theory when $\alpha > 1$.

VERIFICATION OF CONDITION 1A. Rewrite Λ_k^n as

$$\Lambda_k^n = \exp\left\{ng\left(\frac{n-k+1}{n}\right) - \sum_{i=1}^{n}\log\left(\frac{\frac{1}{n-i+1}\sum_{j=i}^{n}\gamma(\tau(j,n),k)}{\frac{k-1}{n}+\alpha\frac{n-k+1}{n}}\right)\right\}.$$

Here $g(x) = x\log\alpha - \log(1+(\alpha-1)x)$. Note that $g(0) = g(1) = 0$ and that g is convex, so that $g(x) < 0$ for $0 < x < 1$. Hence, given $\epsilon_1, \epsilon_2 > 0$, there exists $\epsilon_3 > 0$ such that $g((n-k+1)/n) < -\epsilon_3$ for all $\epsilon_1 n \leq k \leq (1-\epsilon_2)n$. Note that

$$\log\left(\frac{\frac{1}{n-i+1}\sum_{j=i}^{n}\gamma(\tau(j,n),k)}{\frac{k-1}{n}+\alpha\frac{n-k+1}{n}}\right) \geq \log\left(\frac{1\wedge\alpha}{1\vee\alpha}\right) = -|\log\alpha|.$$

Let $m = n\epsilon_3/(4|\log\alpha|)$ and note that

$$\sum_{i=n-m+1}^{n}\log\left(\frac{\frac{1}{n-i+1}\sum_{j=i}^{n}\gamma(\tau(j,n),k)}{\frac{k-1}{n}+\alpha\frac{n-k+1}{n}}\right) \geq -\frac{\epsilon_3}{4}n.$$

Denote:

$$S_i = \sum_{j=n+1-i}^{n}\gamma(\tau(j,n),k); \quad i = 1,\cdots,n$$

$$H = \left\{\frac{\frac{1}{i}S_i - E_\infty S_1}{E_\infty S_1} > -\epsilon \text{ for all } i \geq m\right\}.$$

Note that there exist positive constants C_1 and C_2 such that for all $\epsilon > 0$

$$P_\infty\left(\frac{\frac{1}{i}S_i - E_\infty S_1}{E_\infty S_1} < -\epsilon\right) \leq C_1 e^{-C_2 i\epsilon^2}.$$

Therefore, there exist postive constants C_3 and C_4 such that

$$P_\infty(H^c) \leq C_3 e^{-C_4 m\epsilon^2}$$

where H^c is the complement of H. On H, for small enough ϵ,

$$\sum_{i=1}^{n-m} \log\left(\frac{\frac{1}{n-i+1}\sum_{j=i}^{n}\gamma(\tau(j,n),k)}{\frac{k-1}{n}+\alpha\frac{n-k+1}{n}} - 1 + 1\right) > \sum_{i=1}^{n-m}\log(1-\epsilon) > -2n\epsilon.$$

So, in taking $\epsilon < \epsilon_3/4$, we get on H that

$$\Lambda_k^n < e^{-\epsilon_3 n + \frac{1}{4}\epsilon_3 n + \frac{1}{2}\epsilon_3 n} = e^{-\frac{1}{4}\epsilon_3 n}.$$

VERIFICATION OF CONDITION 1B. This is completely analogous to the verification of Condition 1B in Gordon and Pollak (1989), and is therefore omitted.

VERIFICATION OF CONDITION 1C. If $\alpha > 1$, then $\Lambda_k^{n+1}/\Lambda_k^n \leq \alpha$ for all $1 \leq k \leq n+1$, from which Condition 1C is clearly seen to hold. If $\alpha < 1$, the proof is completely analogous to the verification of Condition 1C in Gordon and Pollak (1989), and is therefore omitted.

PROOF OF THEOREM 5.1. Here

$$G_0(x) = P(\text{geometric}(p) - U(0,1) \leq x)$$
$$G_1(x) = P(\text{geometric}(\alpha p) - U(0,1) \leq x).$$

Clearly, α is the correct tuning parameter value for the NPSRE. In order to compute ξ of (3), we must evaluate $\int Q(x)dG_1(x)$. Here

$$\frac{dG_0(x)}{dx} = pq^{[x]} \text{ for } x > 0$$
$$G_0(x) = 1 - q^{[x]} + pq^{[x]}(x - [x]) \text{ for } x > 0$$
$$\frac{dG_1(x)}{dx} = \alpha p(1-\alpha p)^{[x]} \text{ for } x > 0$$

where $q = 1 - p$. Hence

$$\begin{aligned}\int Q(x)dG_1(x) &= \int -\log(1 - G_1(x))dG_1(x)\\ &= \sum_{k=0}^{\infty}\int_k^{k+1} -\log(q^k) - pq^k(x-k))\alpha p(1-\alpha p)^k dx\\ &= -\sum_{k=0}^{\infty}\alpha p(1-\alpha p)^k\Big\{x\log(q^k - pq^k(x-k)) - x+\\ &\quad \frac{(q^k + kpq^k)\log(q^k - pq^k(x-k)}{-pq^k}\Big\}\Big|_k^{k+1}\\ &= -\sum_{k=0}^{\infty}\alpha p(1-\alpha p)^k\Big\{((k+1)^2 - k^2)\log q - 1 - \Big(\frac{1}{p}+k\Big)\log q\Big\}\\ &= -\sum_{j=1}^{\infty}\alpha p(1-\alpha p)^{j-1}\Big(j\log q - \frac{1}{p}\log q - 1\Big)\\ &= 1 + (\log q)\Big(\frac{1}{p} - \frac{1}{\alpha p}\Big)\\ &= 1 + \Big(1 - \frac{1}{\alpha}\Big)\frac{\log(1-p)}{p}.\end{aligned}$$

Now

$$\begin{aligned}E_{G_1}\log\Big(\frac{dG_1}{dG_0}(X)\Big) &= E_{\text{geometric}(\alpha p)}\log\Big(\frac{\alpha p(1-\alpha p)^{X-1}}{p(1-p)^{X-1}}\Big)\\ &= \log\alpha + \Big(\log\frac{1-\alpha p}{p}\Big)\Big(\frac{1}{\alpha p} - 1\Big).\end{aligned}$$

This accounts for Theorem 5.1.

REFERENCES

Croarkin, M. C., Hagwood, C. and Pollak, M. (1993). "Surveillance schemes with applications to mass calibration." NISTIR 5158, Technical Report of the Statistical Engineering Division, the National Institute of Standards and Technology, Gaithersburg, MD 20899.

van Dobben de Bruyn, C. S. (1968). "Cumulative Sum Tests." Hafner, New York.

Gordon, L. and Pollak, M. (1989). "An efficient sequential nonparametric scheme for detecting a change of distribution."Technical Report. To appear in *Ann. Statist.*

GORDON, L. and POLLAK, M. (1990). "Average run length to false alarm for surveillance schemes designed with partially specified pre-change distribution." Technical Report. To appear in *Ann. Statist.*

GORDON, L. and POLLAK, M. (1991). "A robust surveillance scheme for stochastically ordered alternatives." Technical report. To appear in *Ann. Statist.*

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

MATH WORKS (1989). *Pro-Matlab User's Guide.* The Math Works, South Natick, Mass.

PAGE, E. S. (1954). "Continuous inspection schemes." *Biometrika* **41**, 100–115.

POLLAK, M. (1985). "Optimal detection of a change in distribution." *Ann. Statist.* **13**, 206–227.

POLLAK, M. (1987). "Average run lengths of an optimal method of detecting a change in distribution." *Ann. Statist.* **15**, 749–779.

POLLAK, M. and SIEGMUND, D. (1991). "Sequential detection of a change in a normal mean when the initial value is unknown." *Ann. Statist.* **19**, 394–416.

ROBERTS, S. W. (1966). "A comparison of some control chart procedures." *Technometrics* **8**, 411–430.

SAVAGE, I. R. (1956). "Contributions to the theory of rank order statistics – the two-sample case." *Ann. Math. Statist.* **27**, 590-615.

SHEWHART, W. A. (1931). "The Economic Control of the Quality of Manufactured Product." Macmillan, New York.

SHIRYAYEV, A. N. (1963). "On optimum methods in quickest detection problems." *Theory Probab. Appl.* **8**, 22–46.

MATHEMATICAL SCIENCES DEPARTMENT
SAN DIEGO STATE UNIVERSITY
SAN DIEGO, CA 92175

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SOUTHERN CALIFORNIA
DRB-155, MC 1113
LOS ANGELES, CA 90089-1113

DEPARTMENT OF STATISTICS
THE HEBREW UNIVERSITY
JERUSALEM, ISRAEL

```
%   This is a program for computing the sequence of
%         statistics of a one-sided Shiryayev-Roberts
%         nonparametric detection scheme based on Lehmann
%         alternatives (Bell-Gordon-Pollak)
%   Input:  data     row of en data points
%           alpha    representative parameter of post-change
%                    distribution.  alpha < 1 if post-change
%                    distribution is stochastically larger
%                    than pre-change
%   Output:  RBGP    row of en Shiryayev-Roberts statistics
en = length(data');
RBGP = zeros(1:en);
lambdank = zeros(en);
for n=1:en,
    datan = data(1:n);
    [dummy,invrankt] = sort(datan');
    invrank = invrankt';
    for k=1:n,
        timegek = (invrank > =k);
        timelk = ones(1:n) - timegek ;
        g = alpha*timegek + timelk ;
        reverse = cumsum(ones(1:n)) ;
        verse = reverse(n:-1:1) ;
        sumg = cumsum(g(n:-1:1)) ;
        isumg = sumg(n:-1:1) ;
        iavg = log(isumg./verse) ;
        lndenom = sum(iavg') ;
        lambdank(n,k) = exp((n-k+1)*log(alpha)-lndenom) ;
    end
end
RBGP = (sum((lambdank)'))
```

Figure 1: A MATLAB program for computing the statistics R_n

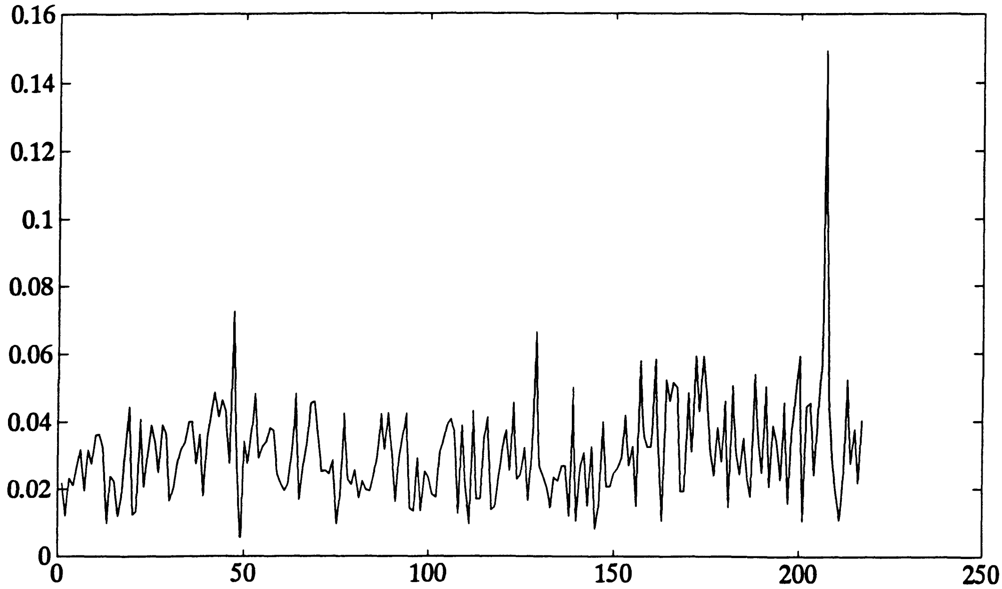

Figure 2: Plot of the data appearing in Table 2

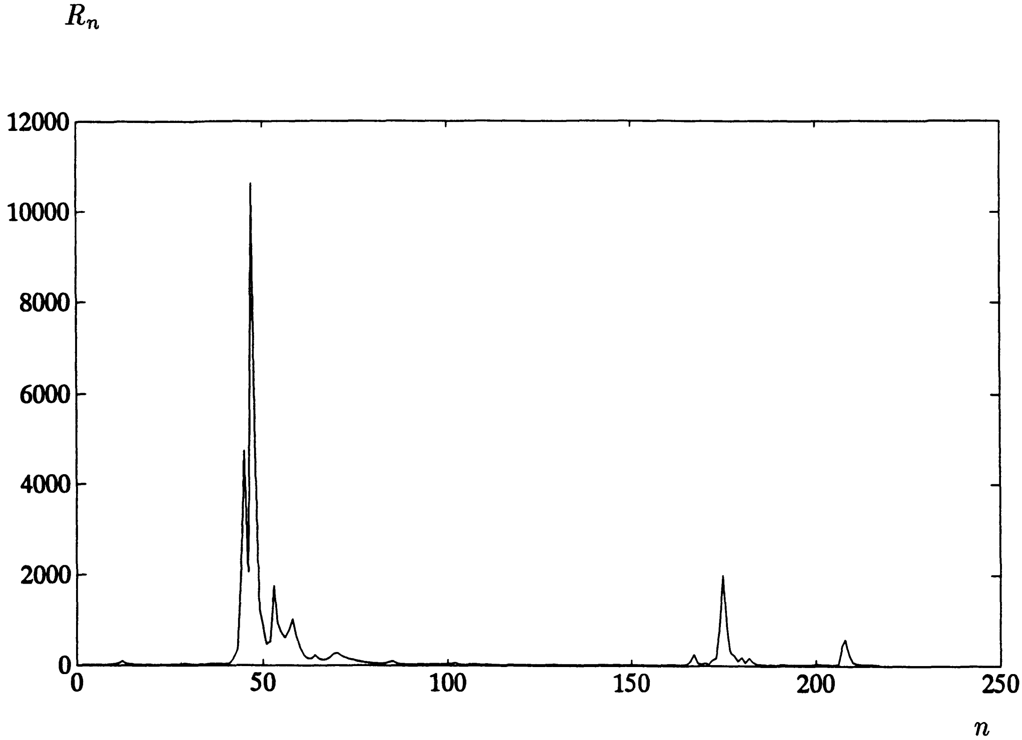

Figure 3: A Plot of R_n for the 2-sided NPSRE (mass of .50 each for $\alpha = .1992$ and $\alpha = 5.9207$) applied to the data of Table 2

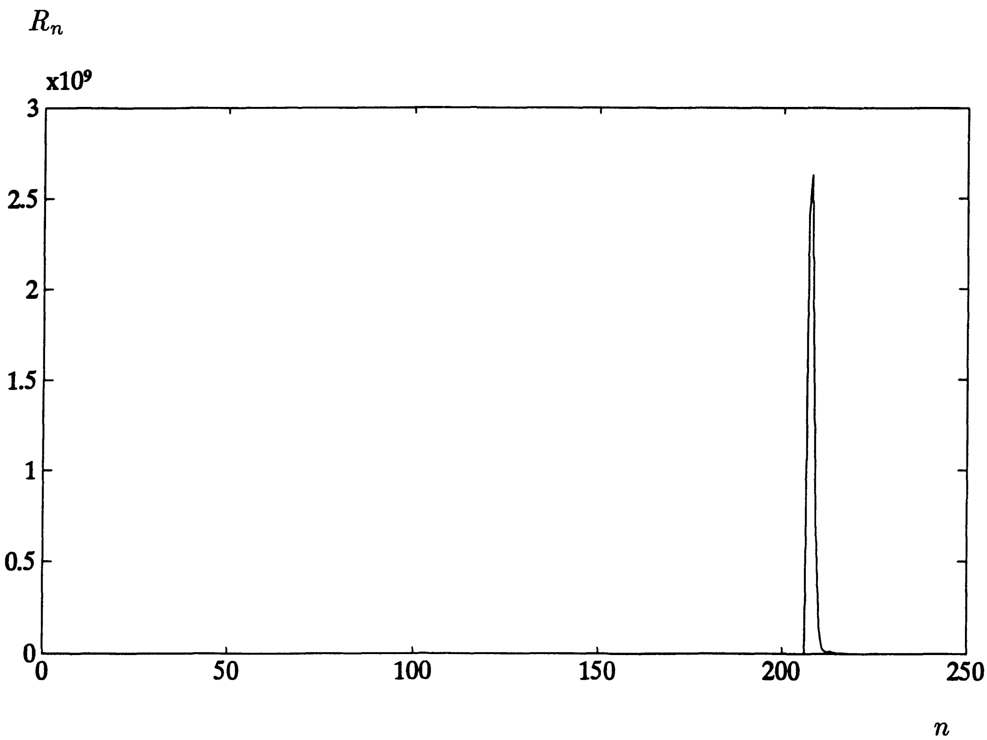

Figure 4: A plot of R_n for the 2-sided parametric Shiryayev-Roberts scheme (mass of .50 each for a change to 2σ and to $\sigma/2$) applied to the data of Table 2

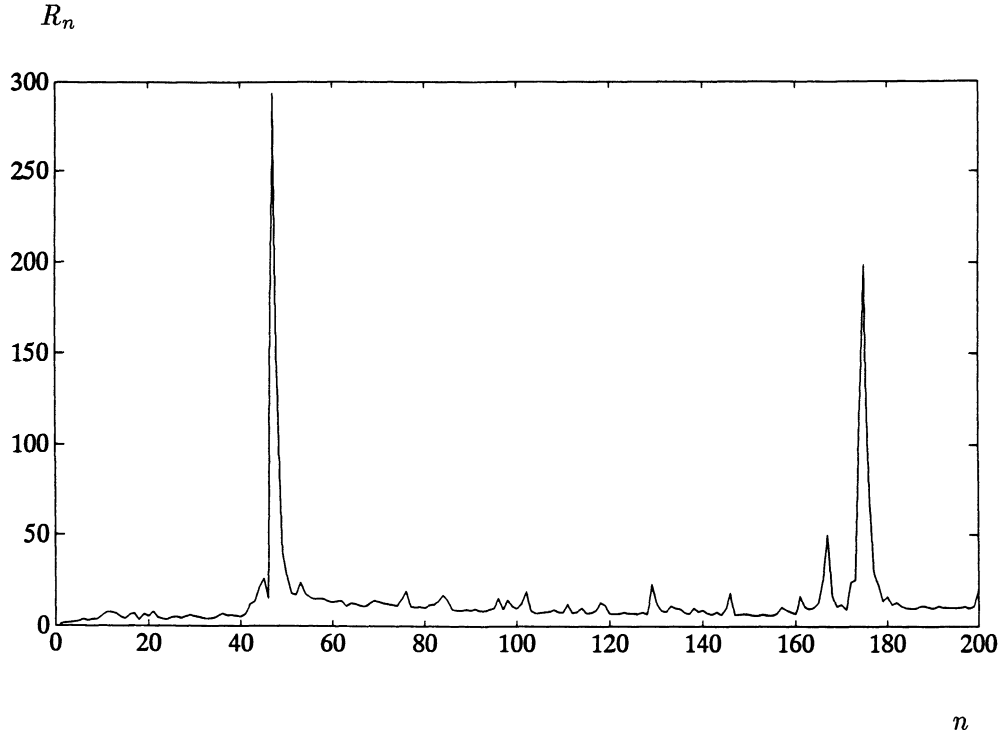

Figure 5: Detail of Figure 4

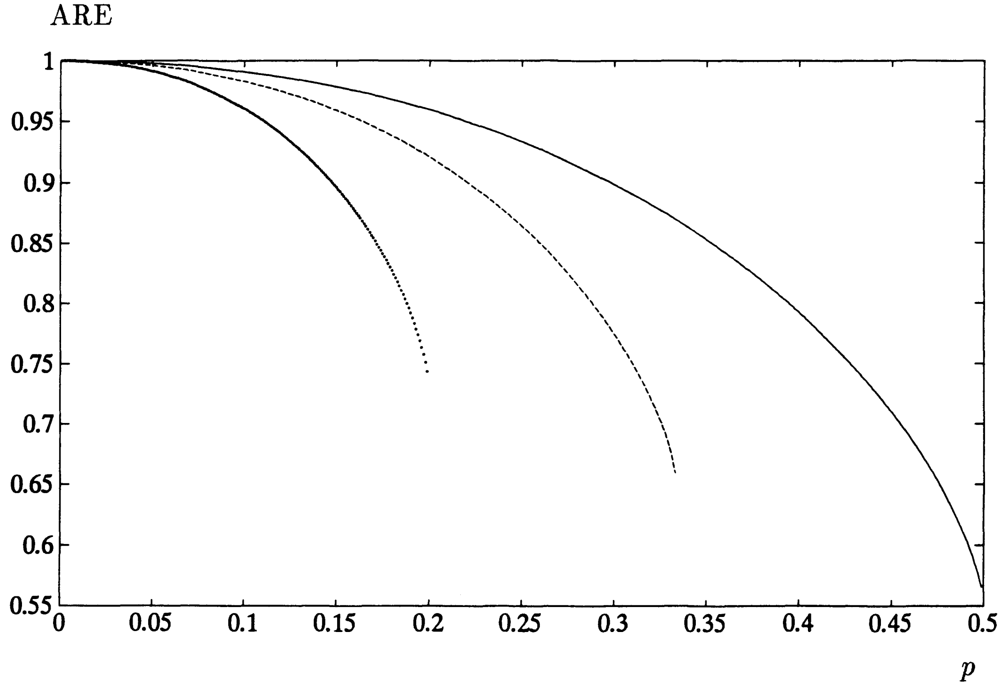

Figure 6: AREE of NPSRE for detecting an increase of the success probability p of a Bernoulli process, when $\alpha = 1/5$ ($\cdots\cdots$), $\alpha = 1/3$(– – – – – ––) and $\alpha = 1/2$ (——).