The Annals of Statistics
1994, Vol. 22, No. 2, 763–804

AN EFFICIENT SEQUENTIAL NONPARAMETRIC SCHEME FOR DETECTING A CHANGE OF DISTRIBUTION[1]

BY LOUIS GORDON AND MOSHE POLLAK

University of Southern California and Hebrew University of Jerusalem

Suppose that a system in its standard state produces i.i.d. observations whose distribution is symmetric about zero. At an unknown time the system may leave its standard state, and the observations would subsequently be stochastically larger. Subject to a bound on the rate of false alarms, one wants to detect quickly such a departure from the standard state.

We present a robust method of detection which is computationally feasible and remarkably efficient. The method is based on the sequential vectors of signs and ranks of the observations. The methodology is one of likelihood ratio; a sequence of likelihood ratios for these vectors is computed, and the Shiryayev–Roberts approach to changepoint detection is then applied to yield a class of statistics and associated stopping rules. Inequalities and asymptotic approximations for the operating characteristics of these rules are developed. These are found to be valid also for small average run lengths and early changepoints as well. The relative efficiency of these schemes (with respect to a normal parametric shift detection policy) is very high, making them a robust alternative to parametric methods.

1. Introduction and summary. Consider a situation where a process yields observations $X_1, X_2, \ldots$. Initially, the observations follow a distribution F_0. At ν, an unknown point in time, something happens to the process, causing the distribution of the observations to change to F_1. Subject to a bound on the rate of false alarms, one wants to detect a true change quickly.

This problem is common in process quality control and arises whenever a process is being monitored for potential change. Detection schemes which are in use typically are constructed under assumptions that the observations are independent, that F_0 and F_1 belong to a parametric family and that at least F_0 is completely known. A commonly used simple scheme is due to Shewhart (1931). Efficient schemes in common use today are Cusum procedures, introduced by Page (1954). Let f_i denote the density of F_i, for $i = 0, 1$. The Cusum procedure is to compute

$$\max_{1 \le k \le n} \frac{f_1(x_k) \cdot f_1(x_{k+1}) \cdot \cdots \cdot f_1(x_n)}{f_0(x_k) \cdot f_0(x_{k+1}) \cdot \cdots \cdot f_0(x_n)} \tag{1}$$

sequentially at each point n in time, and to assert that a change has occurred as soon as the statistic (1) exceeds a specified critical level.

Received August 1989; revised May 1993.

[1]Research supported in part by NIH Grant GM-32230, NSF Grant DMS-90-05833 and U.S.–Israel Binational Science Foundation Grant 89-00115.

AMS 1991 *subject classifications.* Primary 62L10; secondary 62N10.

Key words and phrases. Cusum, changepoint, disruption.

Moustakides (1986) shows that if the density f_1 used in computing (1) is indeed the true density after the change has occurred, then the Cusum procedure most rapidly detects a change in distribution among all procedures with a common bound specifying an acceptable rate of false alarms. Moustakides' result requires a specific technical formulation of the procedure's operating characteristics. Ritov (1990) shows that in Moustakides' formulation the Cusum procedure is a Bayes rule.

Another statistic has been proposed by Shiryayev (1963) and Roberts (1966). This procedure requires the statistician to specify an f_1 as above, and then to compute

$$\sum_{k=1}^{n} \frac{f_1(x_k) \cdot f_1(x_{k+1}) \cdot \cdots \cdot f_1(x_n)}{f_0(x_k) \cdot f_0(x_{k+1}) \cdot \cdots \cdot f_0(x_n)} \tag{2}$$

sequentially at each point n in time and again to assert that a change has occurred as soon as this statistic exceeds a specified level A.

Under a formulation of operating characteristic different from that of Moustakides, Pollak (1985) shows that the Shiryayev–Roberts procedure is asymptotically optimal as $A \to \infty$, when the density f_1 of (2) is indeed the true density after the change has occurred. Pollak and Siegmund (1985) show that the difference in the speed of detection between the Cusum and Shiryayev–Roberts procedures is small. The Shiryayev–Roberts approach has technical advantages which enable one to deal with more complex situations, such as when F_1 is unknown and the statistician does not wish to specify f_1. [See Pollak (1985), (1987).] It is because of these technical advantages that we here adopt the Shiryayev–Roberts approach.

Nonparametric schemes which have been studied and reported in the literature are somewhat different in nature. They generally call for constructing a sequence of statistics based on the signs or on the signed ranks of the observations. One again asserts that a change has been detected when the sequence first exceeds a critical level. Exact computation of operating characteristics is typically not feasible, so one is forced to use approximation by Brownian motion. The procedures are generally constructed with contiguous alternatives in mind; one expects to see many observations before the change, and expects to see a small change, if it occurs at all. [For an overview see Bhattacharya and Frierson (1981).]

Recently, McDonald (1990) has proposed a Cusum procedure based on sequential ranks. McDonald's idea is to combine the approximate uniformity of sequential ranks with a Cusum procedure for detecting a change from uniformity to a noncontiguous alternative.

Neither the Cusum nor the Shiryayev–Roberts approaches have been directly applied to the nonparametric likelihood-ratio-based versions of the problem, in spite of both their claims to optimality in the parametric case, as well as their suitability in handling larger changes in distribution. We attribute this lack to the technical difficulty involved in computing likelihood ratios such as needed in computing (1) or (2) for signed ranks. The only sequential nonparametric likelihood ratio approach of which we are aware is due to Savage and Sethuraman (1966). [See also Miller (1970) and Woodroofe (1983).]

Here we explore a setting in which one can compute likelihood ratios for the signed ranks, and so one can try to adapt the Cusum or Shiryayev–Roberts procedures in a nonparametric context. The setting is one where the observations are independent, being initially symmetrically distributed about the origin, and are stochastically larger after the change than before. We will require our procedure to be a function only of the sequence of signs and of the sequence of the ranks of the observations' absolute values. Our procedure will be analyzed for noncontiguous alternatives.

Suppose, for the moment, that the initial distribution of the observations is known to be double exponential: $f_0(x) = \frac{1}{2}\exp(-|x|)$. Suppose further that after the change the density is known to be $f_1(x) = p\alpha\exp(-\alpha x)I_{\{x>0\}} + q\beta\exp(\beta x)$ $I_{\{x<0\}}$, where α, β, p and q are all positive parameters with $p + q = 1$. As will be shown in Section 2, when the changepoint is k, a result of Savage (1956) enables us to compute the likelihood of any finite sequence of signs and ranks of absolute values. This lets us compute the likelihood ratio

$$\Lambda_k^n = \frac{\text{likelihood}_k(\text{first } n \text{ observations' signs and ranks of absolute values})}{\text{likelihood}_\infty(\text{first } n \text{ observations' signs and ranks of absolute values})},$$

where we denote by ∞ the absence of a change. The natural analog of the Cusum statistic is $\max_{1\le k\le n}\Lambda_k^n$; the analog of the Shiryayev–Roberts statistic

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

We here study the latter statistic, which we call the nonparametric Shiryayev –Roberts statistic (NPSR). Consider the stopping rule

$$N_A = \min\{n \mid R_n \ge A\}.$$

Given a stopping rule N, the standard index for the rate of false alarms is $\mathbb{E}_\infty\{N\}$. One typically controls the level of false alarms by considering only stopping rules N which satisfy

$$\mathbb{E}_\infty N \ge B, \tag{3}$$

for some specified level B.

In our case, this means that A has to be set so that $\mathbb{E}_\infty N_A = B$. This requires an analysis of $\mathbb{E}_\infty N_A$, which we provide in Sections 2, 4 and 6. We emphasize that $\mathbb{E}_\infty N_A$ is independent of the actual distribution of the observations, as long as that distribution is continuous and symmetric about the origin. The double exponential distribution discussed in the previous paragraph is merely an artifice for us to define the statistic R_n. The sequence of statistics $\{R_n\}$ is a function only of signs and ranked absolute values; its distribution when $\nu = \infty$ is the same for all continuous distributions symmetric about 0.

There are three parameters (α, β, p) involved in our proposed scheme. They can be chosen arbitrarily. For the problem considered, it would seem appropriate to choose $0 < \alpha \le 1 \le \beta < \infty$ and $\frac{1}{2} \le p < 1$. They can be selected with specific alternative distributions F_1 in mind. For example, in the normal parametric analog of the problem, suppose that one believes the observations to be distributed $\mathcal{N}(0, 1)$ before the change and distributed $\mathcal{N}(\mu, 1)$ after the change, with $\mu > 0$. One could choose $\delta > 0$, set $f_0(x) = \phi(x)$ (the standard normal density), set $f_1(x) = \phi(x-\delta)$ and employ parametric Cusum or Shiryayev–Roberts procedures using statistic (1) or (2). If, however, the statistician were reluctant to specify normality, he could consider a nonparametric Shiryayev–Roberts scheme. In that case the test parameters α, β and p should be selected to make the nonparametric likelihood ratios closely resemble the parametric likelihood ratios. We show how to do this in Section 3. In addition, we there show that the relative efficiency of the parametric and nonparametric Shiryayev–Roberts procedures for normal shifts as $B \to \infty$ is very high, suggesting that our nonparametric scheme is a robust alternative to parametric methods.

A Monte Carlo study of our procedures for normal distributions with shift alternatives is the subject of Section 5. We there present evidence indicating that our asymptotics provide reasonable guidance in situations of practical interest.

2. The procedure and its average run length (ARL) to false alarm. We first present some background and notation. Lemma 2.1 is found in Savage (1956).

LEMMA 2.1. *Let $Y_1, \dots, Y_n$ be i.i.d. exponential random variables, and let $x_1, \dots, x_n$ be arbitrary positive constants. Then*

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

PROOF. Directly, as in Savage (1956), or by induction. □

Now let $X_1, X_2, \dots$ be independent and, until otherwise specified, assume that $X_1, X_2, \dots, X_{\nu-1}$ have common density f_0, while $X_\nu, X_{\nu+1}, \dots$ have common density f_1, where $1 \le \nu \le \infty$ and

$$f_0(x) = \tfrac{1}{2}\exp\bigl(-|x|\bigr), \tag{4}$$

$$f_1(x) = p\alpha \exp(-\alpha x) I_{\{x>0\}} + q\beta \exp(\beta x) I_{\{x<0\}}. \tag{5}$$

The parameters α, β, p and q are assumed positive with $p + q = 1$. We denote the statistics required by the parametric Shiryayev–Roberts statistic with a superscript "p":

$$R_n^p = \sum_{k=1}^{n} \frac{f_1(x_k) \cdot f_1(x_{k+1}) \cdot \cdots \cdot f_1(x_n)}{f_0(x_k) \cdot f_0(x_{k+1}) \cdot \cdots \cdot f_0(x_n)}$$

and $N_A^p = \min\{n \mid R_n^p \geq A\}$. From Pollak [(1987), Theorem 1] the following limit Δ exists:

$$\Delta = \lim_{A \to \infty} \mathbb{E}_\infty \left\{ \frac{N_A^p}{A} \right\}.$$

We shall compute its value in Section 4.

Define $\sigma_i = I_{\{X_i > 0\}}$, giving the sign of the ith observation. We denote the number of positive observations among $X_k, \ldots, X_n$ by $U(k,n) = \sum_{j=k}^n \sigma_j$. The corresponding count of negative observations is $V(k,n) = \sum_{j=k}^n (1 - \sigma_j)$. When $k > n$ we set $U(k,n) = V(k,n) = 0$, following the convention that summation over an empty set is 0.

Denote the rank of the absolute value of the ith observation among the first n absolute values observed by

$$\rho(i,n) = \sum_{j=1}^n I_{\{|X_j| \leq |X_i|\}}.$$

Note that $\rho(\cdot,\cdot)$ is well defined for $i > n$. Write Z_i^n for the 2-vector $(\rho(i,n), \sigma_i)$. The n 2-vectors $Z_1^n, \ldots, Z_n^n$ contain all the information we use from the first n observations.

Because we assume the distributions of the X_i are continuous, the ranked absolute values of the first n observations determine the random permutation $\rho(\cdot, n)$. Denote its inverse permutation $\tau(\cdot, n)$, so that $\rho(\tau(i,n),n) = i$ for $i = 1, \ldots, n$. Finally, define

$$\gamma(j,k) = \begin{cases} 1, & \text{if } j < k, \\ \alpha, & \text{if } j \geq k \text{ and } \sigma_j = 1, \\ \beta, & \text{if } j \geq k \text{ and } \sigma_j = 0. \end{cases}$$

By conditioning on the signs of $X_1, \ldots, X_n$, we obtain from Lemma 2.1 an explicit form for the likelihood $h_k(Z_1^n, \ldots, Z_n^n)$ for the signs and ranks of absolute values when the changepoint is k, the prechange density is f_0 and the postchange density is f_1:

$$h_k(Z_1^n, \ldots, Z_n^n) = \left(\frac{1}{2}\right)^{k-1} p^{U(k,n)} q^{V(k,n)} \prod_{i=1}^n \frac{\gamma(\tau(i,n),k)}{\sum_{j=i}^n \gamma(\tau(j,n),k)},$$

valid for $1 \leq k \leq n+1$. Clearly, $h_\infty(Z_1^n, \ldots, Z_n^n) = 2^{-n}/n!$. Hence, the nonparametric likelihood ratio based on signs and ranked absolute values is

$$\begin{aligned} \Lambda_k^n &= \frac{h_k(Z_1^n, \ldots, Z_n^n)}{h_\infty(Z_1^n, \ldots, Z_n^n)} \\ &= (2p)^{U(k,n)} (2q)^{V(k,n)} n! \prod_{i=1}^n \frac{\gamma(\tau(i,n),k)}{\sum_{j=i}^n \gamma(\tau(j,n),k)} \\ &= (2p)^{U(k,n)} (2q)^{V(k,n)} \prod_{i=1}^n \frac{\gamma(\tau(i,n),k)}{[1/(n-i+1)] \sum_{j=i}^n \gamma(\tau(j,n),k)}, \end{aligned} \tag{6}$$

for $1 \le k \le n+1$. Note that (6) yields $\Lambda_{n+1}^n = 1$.

Finally, we define our nonparametric analogs to the Shiryayev–Roberts statistics by

$$R_n = \sum_{k=1}^{n} \Lambda_k^n \tag{7}$$

and

$$N_A = \min\{n \mid R_n \ge A\}.$$

Note that R_n and N_A are both well-defined statistics, even if the data are not generated under the hypothesized densities. Assertion (9) of the following theorem tells us that the NPSR procedure shares very similar false alarm rates with the parametric version, when the false alarm rate is required to be low.

THEOREM 2.2. *Suppose that when $\nu = \infty$ the observations $X_1, X_2, \dots$ are i.i.d., with double exponential distribution F_0. Let $0 < \alpha < 1 < \beta < \infty$ be specified parameters. Then the parametric and nonparametric Shiryayev–Roberts procedures have comparable ARL to false alarm. In particular,*

$$\mathbb{E}_\infty\{N_A\} \ge A \tag{8}$$

and

$$\lim_{A\to\infty} \frac{\mathbb{E}_\infty\{N_A\}}{A} = \lim_{A\to\infty} \frac{\mathbb{E}_\infty\{N_A^p\}}{A} = \Delta > 1. \tag{9}$$

If $2p\alpha \vee 2q\beta \le 1$, then $\Delta = 1/\alpha$.

We defer the proof to Sections 4.1 and 6.

We are now ready to describe our procedure:

Suppose observations $X_1, X_2, \dots$ are independent, initially having a continuous distribution symmetric about the origin, and that at an unknown time ν they become stochastically larger. Suppose further that one wishes a monitoring procedure for change point detection whose ARL to false alarm is to be no less than the bound B, as in (3). The NPSR procedure requires that the statistician do the following:

1. specify tuning parameters α, β, p and B, the desired ARL to false alarm;
2. compute at each observation the statistic R_n;
3. stop and declare a change has been detected at the first time N_A that R_n exceeds $A = B/\Delta$, and Δ is given by (9) of Theorem 2.2. (A conservative choice for A is to take $A = B$.)

We discuss the choice of tuning parameters α, β and p in Section 3. We show how to compute the limit Δ in Section 4. As an example of the calculations, we give in Table 1 the optimal values of α, β, p and corresponding Δ for various normal shift alternatives. In Section 5, we present the results of simulation studies

```
                % compute NPSR statistic for n observations
                % data assumed in the column n-vector nobsrvd
                % program assumes there are no ties among observations
sigma           = (nobsrvd > = 0);      % 1 if observation j positive
gammas          = beta.*(1-sigma)
                 + alpha.*sigma;
fctrial         = (1:1:n);              % column vector of indices
decr            = n:-1:1;               % indices in reverse order
lambdank        = zeros(1:n, 1);
[dummy, rhon]   = sort(abs(nobsrvd));   % index of smallest in rhon(1)
                                        % index of largest in rhon(n)
rhon            = rhon(decr);           % time of largest in rhon(1)
                                        % time of smallest in rhon(n)
[dummy, taun]   = sort(rhon);           % get inverse ranks
gamman          = gammas(rhon);         % gammas in decreasing rank order
                                        % gamman(taun(k)) is gamma of kth
                                        % observed
sigman          = sigma(rhon);          % signs in decreasing rank order
for k = 1:n,
    lambdank (k, 1) = ···
      ((2*p*alpha)    .^ (               sum(sigman)) ) ···
      .* ((2*q*beta)  .^ ((n + 1 - k) - sum(sigman)) ) ···
      ./ prod( cumsum(gamman) ./ fctrial);
    gamman(taun(k))=1;
    sigman(taun(k))=0;
end;
srn = sum(lambdank);
```

FIG. 1. *MATLAB program for computing NPSR.*

which complement the theoretical development for large samples. Simulation results suggest that the high efficiency seen in theory may be achieved in practical situations. The computations are conveniently programmed. Figure 1 is a program written in the MATLAB programming language and is the basic code used in the simulation results reported in Section 5. See MathWorks (1989) for a description of the MATLAB language. Finally, we provide in Section 6 those proofs deferred from previous sections.

3. Speed of detection, choice of parameters and relative efficiency. Suppose the change is at time ν and one did not raise a false alarm. If N is the stopping time used to monitor the process for change, the lag in detecting the change is $N+1-\nu$. We adopt $\mathbb{E}_\nu\{N+1-\nu \mid N \geq \nu\}$ as an index of the speed of detection. See Lorden (1971) for a different formulation. The asymptotic results and hence the recipe for choice of parameters α, β and p to be presented below do not depend on which of these formulations is adopted.

Because ν is unknown, one's index has to be a function of the sequence $\{\mathbb{E}_k(N+1-k\,|\,N \geq k)\}$, for $k = 1,2,\ldots$. The standard approach is to use $\sup_{1 \leq k < \infty} \mathbb{E}_k(N+1-k\,|\,N \geq k)$. The value $\lim_{k \to \infty} \mathbb{E}_k(N+1-k\,|\,N \geq k)$ is of special interest because it expresses the expected lag in detection when the change takes place a long time after surveillance was started. As a matter of convenience, one replaces the vaguely defined "long time after starting surveillance" with a limit as the time of change gets arbitrarily large.

This replacement creates the substantial technical difficulties we encounter in Sections 6.4 and 6.5 because it requires us to compute conditional expectations over events whose probability tends to 0. [See Roberts (1966).] The value $\mathbb{E}_1(N \mid N \geq 1)$ is also particularly interesting because one might worry that a change has been in effect from the very outset of surveillance. [See Lucas and Crosier (1982).] For many stopping rules, the latter expected lag coincides with $\sup_{1 \leq k < \infty} \mathbb{E}_k(N+1-k \mid N > k)$.

In the parametric problem, the indices

$$\sup_{1 \leq k < \infty} \mathbb{E}_k\big(N+1-k \mid N \geq k\big), \qquad \lim_{k \to \infty} \mathbb{E}_k\big(N+1-k \mid N \geq k\big)$$

or

$$\mathbb{E}_1\big(N+1-k \mid N \geq k\big)$$

all typically grow like (constant) $\cdot \log(A) + o(\log A)$ as $A \to \infty$, where the constant depends on the stopping rule but is the same for all three choices of sensitivity index. The $o(\log A)$ term of course depends on the selected sensitivity index, as well as on the stopping time.

When using N_A in our nonparametric setting, the situation is different. Although $\log(A)$ is still the dominant order of growth, the constant of proportionality depends as well on the time k at which the change occurs. Therefore consideration of the choice of α, β and p which define the NPSR procedure involves a preliminary choice of k.

Following Roberts (1966), we choose

$$\limsup_{k \to \infty} \mathbb{E}_k\big(N+1-k \mid N \geq k\big)$$

as our primary sensitivity index of expected lag. Often $\mathbb{E}_k(N+1-k \mid N \geq k)$ is well approximated by its limit as $k \to \infty$, even for values of k which are small relative to A. Such behavior is also seen in the simulation results presented below. Therefore, if one is not unduly concerned that the process might be out of control from the beginning of surveillance, then $\limsup_{k \to \infty} \mathbb{E}_k(N+1-k \mid N \geq k)$ is a reasonable choice of index.

Suppose that $G_0(x)$ is the real initial c.d.f. of the observations, known to be symmetric about 0 and continuous. Suppose also that $G_1(x)$ is the c.d.f. of the observations after the change has occurred and that the statistician uses the stopping rule N_A. Without changing ranks of absolute values or their associated signs, one can transform the observations to make their distribution prior to change double exponential. The transformation is

$$Q(x) = -(2\sigma(x) - 1)\log\Big(2 - 2G_0(|x|)\Big), \tag{10}$$

where $\sigma(x) = I_{\{x>0\}}$. We write $G_1 \geq_{\mathrm{stoch}} G_0$ if $1 - G_1(t) \geq 1 - G_0(t)$ for all t, the usual definition of stochastic ordering between distributions. Note that stochastic ordering is preserved by increasing transformations.

THEOREM 3.1. *Let $G_0(\cdot)$ and $G_1(\cdot)$ be the true prechange and postchange distributions. Let both be continuous. Define*

$$\begin{aligned} D_Q &= (1 - G_1(0))\log(2p\alpha) + G_1(0)\log(2q\beta) \\ &\quad + (1-\alpha)\int_0^\infty Q(x)G_1(dx) + (\beta - 1)\int_{-\infty}^0 Q(x)G_1(dx) \qquad (11) \\ &= \int_{-\infty}^\infty \log\left(\frac{f_1(Q(x))}{f_0(Q(x))}\right)G_1(dx), \qquad (12) \end{aligned}$$

where the transformation $Q(x)$ is given in (10). *If $\infty > D_Q > 0$, if $G_1 \geq_{\mathrm{stoch}} G_0$, if $\alpha < 1 < \beta$ and if $p\alpha \geq q\beta$, then*

$$\lim_{A\to\infty} \limsup_{k\to\infty} \frac{\mathbb{E}_k\{N_A - k \mid N_A \geq k\}}{D_Q^{-1}\log(A)} = 1 \tag{13}$$

and

$$\lim_{A\to\infty} \liminf_{k\to\infty} \frac{\mathbb{E}\{N_A - k \mid N_A \geq k\}}{D_Q^{-1}\log(A)} = 1. \tag{14}$$

A sketch of the proof is given in Sections 6.4 and 6.5.

If G_0 and G_1 are the suspected prechange and postchange distributions, then one should choose α, β and p to maximize D_Q. Differentiate (11) to obtain the critical values

$$\begin{aligned} p &= 1 - G_1(0), \\ \alpha &= \frac{1 - G_1(0)}{\int_0^\infty Q(x)G_1(dx)}, \qquad (15) \\ \beta &= \frac{-G_1(0)}{\int_{-\infty}^0 Q(x)G_1(dx)}. \end{aligned}$$

For example, if $G_0 = \mathcal{N}(0,1)$ and $G_1 = \mathcal{N}(1,1)$, then numerical evaluation of the integrals yields

$$\begin{aligned} p &= 0.8413, \\ \alpha &= 0.531, \qquad (16) \\ \beta &= 1.703. \end{aligned}$$

We use (15) to compute optimal tunings for various normal shift alternatives. These are displayed in the first four columns of Table 1. Specifically, if we anticipate that prechange distributions are approximately standard normal and the postchange distribution is approximately $\mathcal{N}(\mu, 1)$, then the tabled parameters give an NPSR particularly sensitive to that combination.

Note, for example, that the parameters shown in (16) appear in the $\mu = 1$ line of the table. We have slightly understated the utility of NPSR. In actuality, because NPSR is invariant to scale changes, the tunings presented in Table 1 are optimal for detecting shifts in normal mean with unknown scale.

The asymptotic efficiency relative to optimum ($\mathrm{ARE}_{\mathrm{opt}}$) as $A \to \infty$ is obtained by comparing $D_Q^{-1}\log(A)$ to $\lim_{k\to\infty} \mathbb{E}_k\{N^p - k \mid N^p \geq k\}$, where N^p is a stopping time which would have been used in the parametric case that G_0 and G_1 were known to be the true prechange and postchange distributions, and N^p is selected to have comparable false alarm rate $\mathbb{E}_\infty\{N^p\} = \Delta A$, with Δ the constant of Theorem 2.2. Lorden (1971) evaluates the limit as $k \to \infty$ for the Cusum procedure. Pollak (1985) evaluates the limit as $k \to \infty$ for the Shiryayev–Roberts procedure. In both cases, the limiting value is of the form

$$\frac{\log(A)}{\int_{-\infty}^{\infty} \log\bigl(dG_1(x)/dG_0(x)\bigr)G_1(dx)} + O(1),$$

where the $O(1)$ remainder is bounded as $A \to \infty$. Because of the optimality and approximate optimality results of Lorden (1971), Moustakides (1986), Pollak (1985) and Ritov (1990), we therefore compute

$$\begin{aligned}\mathrm{ARE}_{\mathrm{opt}} &= \lim_{A\to\infty} \frac{\log(A)/\int_{-\infty}^{\infty}\log\bigl(dG_1(x)/dG_0(x)\bigr)G_1(dx) + O(1)}{D_Q^{-1}\log(A) + O(1)} \\ &= \frac{D_Q}{\int_{-\infty}^{\infty}\log\bigl(dG_1(x)/dG_0(x)\bigr)G_1(dx)},\end{aligned} \tag{17}$$

the asymptotic efficiency of detection for NPSR relative to both Cusum and parametric Shiryayev–Roberts procedures. In the case of normal shift alternatives $G_0 = \mathcal{N}(0,1)$ versus $G_1 = \mathcal{N}(1,1)$ we obtain $\mathrm{ARE}_{\mathrm{opt}} = 0.971$, when the procedure parameters α, β and p are chosen as in (16).

In Table 1, we present in column 5 the $\mathrm{ARE}_{\mathrm{opt}}$ for detecting a change from $\mathcal{N}(0,1)$ prechange to $\mathcal{N}(\mu, 1)$ postchange, for selected values of μ. NPSR and Cusum are comparable since both are optimally tuned for the size change which is truly realized. $\mathrm{ARE}_{\mathrm{opt}}$ is quite high, never below 94% for the values tabled.

In column 6 of Table 1 we give the other natural comparison of NPSR with Cusum. It is possible that one tunes a procedure to behave optimally for a specific postchange distribution other than the one which is observed. In column 6, we give the ARE of NPSR to that of the Cusum procedure, both tuned for optimal sensitivity to prechange standard normal distribution, and postchange distribution normal with the unit variance and unit mean. ARE's are computed in situations in which the true postchange distribution is $\mathcal{N}(\mu, 1)$. Specifically,

TABLE 1
Parameters for various optimal tunings of NPSR—detection of $\mathcal{N}(0, \sigma^2)$ prechange to $\mathcal{N}(\mu\sigma, \sigma^2)$ postchange

True mean after change	Optimal tunings (15) to detect $\mathcal{N}(\mu\sigma, \sigma^2)$ postchange			$\mathrm{ARE_{opt}}$ (17) and ARE (18) at true μ; NPSR vs. Cusum both tuned for		$\lim_{A\to\infty} E_\infty\{N_A\}/A$ from Section 4.1	
μ	$p(\mu)$	$\alpha(\mu)$	$\beta(\mu)$	$G_0 = \mathcal{N}(0,1)$, $G_1 = \mathcal{N}(\mu,1)$	$\mathcal{N}(0,1)$, $\mathcal{N}(1,1)$	$\underline{\Delta}(\mu)$ lower bound*	$\frac{1}{\alpha(\mu)}$
0.35	0.637	0.808	1.221	0.983	—	1.2383	1.2371
0.45	0.674	0.759	1.289	0.982	—	1.3180	1.3174
0.50	0.691	0.735	1.324	0.981	∞	1.3602	1.3599
0.51	0.695	0.731	1.331	0.981	1.880	1.3689	1.3686
0.53	0.702	0.721	1.345	0.980	1.264	1.3864	1.3862
0.55	0.709	0.712	1.360	0.980	1.140	1.4041	1.4040
0.60	0.726	0.690	1.396	0.979	1.046	1.4499	1.4499
0.70	0.758	0.646	1.469	0.978	0.998	1.5467	1.5468
0.80	0.788	0.606	1.545	0.975	0.982	1.6511	1.6512
0.90	0.816	0.567	1.623	0.973	0.975	1.7633	1.7634
1.00	0.841	0.531	1.703	0.971	0.971	1.8837	1.8838
1.10	0.864	0.497	1.784	0.968	0.970	2.0128	2.0129
1.20	0.885	0.465	1.868	0.966	0.970	2.1510	2.1511
1.30	0.903	0.435	1.953	0.963	0.972	2.2985	2.2986
1.40	0.919	0.407	2.040	0.960	0.976	2.4558	2.4558
1.50	0.933	0.381	2.128	0.958	0.980	2.6230	2.6230
1.75	0.960	0.324	2.355	0.952	0.997	3.0860	3.0860
2.00	0.977	0.277	2.591	0.946	1.019	3.6150	3.6150
2.25	0.988	0.238	2.835	0.942	1.046	4.2104	4.2105
2.50	0.994	0.205	3.085	0.938	1.078	4.8718	4.8718
2.75	0.997	0.179	3.342	0.936	1.114	5.5973	5.5974
3.00	0.999	0.157	3.604	0.936	1.153	6.3855	6.3856
3.25	0.999	0.138	3.870	0.936	1.194	7.2349	7.2350
3.50	1.000	0.123	4.141	0.937	1.237	8.1445	8.1446
3.75	1.000	0.110	4.416	0.938	1.283	9.1138	9.1139
4.00	1.000	0.099	4.694	0.940	1.329	10.1426	10.1426

* Lower bounds are computed by truncating the series in (21). Lower bounds are within 0.001 of Δ.

we compute and table

$$\text{ARE} = \frac{\int_{-\infty}^{\infty} \log\Big(f_1(Q(x))/f_0(Q(x))\Big)H(dx)}{\int_{-\infty}^{\infty} \log\big(dG_1(x)/dG_0(x)\big)H(dx)}, \tag{18}$$

when the denominator is positive and finite, H is the true postchange distribution and G_0 and G_1 are the distributions used to choose p, α and β in the numerator's NPSR. In our case G_0 is $\mathcal{N}(0,1)$, the nominal postchange distribution G_1 is $\mathcal{N}(1,1)$ and the true postchange distribution H is $\mathcal{N}(\mu, 1)$. The results are surprisingly favorable to NPSR.

The ARE of NPSR relative to Cusum as a function of the true postchange mean μ is 97% or above. Lowest efficiency obtains near the value $\mu = 1$, the region in which both procedures are optimally tuned for sensitivity. In situations where the true postchange mean is far from μ, the NPSR does better than its parametric competitor. It can be shown that ARE grows arbitrarily large as $\mu \to \infty$.

Note that (18) is appropriate when both NPSR and Cusum have average lags to detection which grow linearly in the logarithm of ARL to false alarm. In the case $\mu = \frac{1}{2}$, the slope corresponding to NPSR is positive (but quite small), while the slope corresponding to Cusum is 0. We have indicated the ratio by ∞. For the other values $\mu < \frac{1}{2}$ which we have tabled, slopes corresponding to neither NPSR nor Cusum are positive and so we have indicated that our calculations do not apply.

In Section 5, we complement our large-sample results by simulations that suggest that column 6 gives a reasonable approximation to the behavior encountered for a practically interesting range of changepoints k and means μ when attempting to detect a shift from $\mathcal{N}(0, 1)$ to $\mathcal{N}(1, 1)$.

4. Computing Δ. In this section we compute the constant Δ necessary to specify a NPSR procedure for low false alarm rate. In Section 6 we shall verify hypotheses that allow us to use general results of Gordon and Pollak (1990). Specifically, we use Theorem 1 of Gordon and Pollak (1990). The constant Δ is there shown to be

$$\Delta = \frac{1}{\lim_{b\to\infty} \mathbb{E}_1\left\{\exp\left[-\left(\sum_{j=1}^{\tau} W_j - b\right)\right]\right\}},$$

where the random variables $W_j = \log(f_1(X_j)/f_0(X_j))$ are the log-likelihood ratios of the observations, and τ is the first time the random walk of partial sums exceeds the barrier b, computed under the $\mathbb{P}_1$-measure corresponding to immediate change.

In Section 4.1, we provide a simple proof showing that $\Delta = 1/\alpha$ whenever $2p\alpha \vee 2q\beta \le 1$. This is not only useful in its own right, but provides an easy check on the formulas we next derive for Δ when the special conditions are not satisfied.

In the remaining sections, we specialize known renewal theoretic results to the particular f_0 and f_1 of (4) and (5). The expression we obtain involves the distributions of sums of independent gamma variates with differing scale parameters. These probabilities are explicitly evaluated in Section 4.3. Finally, in Section 4.4 we show how to compute Δ with arbitrary accuracy.

4.1. $\Delta = 1/\alpha$ *when* $2p\alpha \vee 2q\beta \le 1$. Under the immediate postchange probability $\mathbb{P}_1$, the log-likelihood ratios W_j are distributed i.i.d. as

$$\left[\log(2p\alpha) + \frac{1-\alpha}{\alpha} Y_j\right] B_j + \left[\log(2q\beta) + \frac{1-\beta}{\beta} Y_j\right](1 - B_j),$$

where the Y_j are i.i.d. unit exponential, independent of the B_j, which are i.i.d. Bernoulli with success probability p. Because of Jensen's inequality, $\mathbb{E}_1\{W_j\} >$

0, which implies that $\tau < \infty$ almost surely. If both $\log(2p\alpha)$ and $\log(2q\beta)$ are negative, the amount of overshoot above b when the walk stops is exponentially distributed with mean $\alpha^{-1} - 1$, because of the lack of memory of exponential variates. Evaluation of the expected overshoot yields the result.

4.2. *The parametric case.* For the remainder of this section, we assume that $X_1, X_2, \ldots$ are independent with $X_k \sim f_0$ for $k < \nu$ and $X_k \sim f_1$ for $k \geq \nu$. We write the following:

$$\begin{aligned}
W_k &= \log\left(\frac{f_1(X_k)}{f_0(X_k)}\right);\\
S_n &= \sum_{k=1}^{n} W_k;\\
\tau_+ &= \min\{n \mid S_n > 0,\, n \geq 1\};\\
\tau_- &= \min\{n \mid S_n \leq 0,\, n \geq 1\};\\
F_+(x) &= \mathbb{P}_1\{S_{\tau_+} \leq x\};\\
\mu_+ &= \mathbb{E}_1\{S_{\tau_+}\};\\
\mu &= \mathbb{E}_1\{W_1\};\\
\tau &= \tau_b = \min\{n \mid S_n \geq b,\, n \geq 1\}.
\end{aligned}$$

Each of the stopping times τ_+, τ_- and τ_b are set equal to ∞ if its associated level is never attained. Note that

$$\mathbb{P}_\infty\{\tau < \infty\} = \mathbb{E}_1\{\exp(-S_\tau)\}$$

and that

$$\mathbb{E}_1\{\exp(-S_{\tau_+})\} = \mathbb{P}_\infty(\tau_+ < \infty) = 1 - \mathbb{P}_\infty(\tau_+ = \infty). \tag{19}$$

From renewal theory and (19), we take limits as $b \to \infty$ to obtain

$$\begin{aligned}
\mathbb{E}_1\Big\{\exp\big[-(S_\tau - b)\big]\Big\} &\to \frac{1}{\mu_+}\int_0^\infty \exp(-y)\big(1 - F_+(y)\big)\,dy\\
&= \frac{1}{\mu_+}\Big(1 - \mathbb{E}_1\{\exp(-S_{\tau_+})\}\Big)\\
&= \frac{1}{\mu_+}\mathbb{P}_\infty(\tau_+ = \infty).
\end{aligned} \tag{20}$$

By Wald's lemma,

$$\mu_+ = \mathbb{E}_1\{S_{\tau_+}\} = \mu\mathbb{E}_1\{\tau_+\},$$

where

$$\mu = p\left(\log(2p\alpha) + \frac{1-\alpha}{\alpha}\right) + q\left(\log(2q\beta) + \frac{1-\beta}{\beta}\right).$$

Next use Corollaries 8.39 and 8.44 of Siegmund (1985) to obtain

$$\mathbb{E}_1\{\tau_+\} = \frac{1}{\mathbb{P}_1\{\tau_- = \infty\}} = \exp\left(\sum_{n=1}^{\infty} \frac{\mathbb{P}_1\{S_n \le 0\}}{n}\right)$$

and

$$\mathbb{P}_\infty\{\tau_+ = \infty\} = \exp\left(-\sum_{n=1}^{\infty} \frac{\mathbb{P}_\infty\{S_n > 0\}}{n}\right).$$

Substitute in (20) to obtain

$$\begin{aligned}\Delta = &\left[p\left(\log(2p\alpha) + \frac{(1-\alpha)}{\alpha}\right) + q\left(\log(2q\beta) + \frac{1-\beta}{\beta}\right)\right] \\ &\times \exp\left(\sum_{n=1}^{\infty} \frac{\mathbb{P}_1\{S_n \le 0\}}{n}\right) \exp\left(\sum_{n=1}^{\infty} \frac{\mathbb{P}_\infty\{S_n > 0\}}{n}\right).\end{aligned} \tag{21}$$

It remains to compute the probabilities and the exponentiated sums required in (21). We write

$$\begin{aligned}W_i = &\left(\log(2p\alpha) + (1-\alpha)\,|\,X_i|\right)I_{\{X_i > 0\}} \\ &+ \left(\log(2q\beta) + (1-\beta)|X_i|\right)I_{\{X_i < 0\}}.\end{aligned} \tag{22}$$

Under $\mathbb{P}_\infty$, we represent X_i as $(-1)^{1+B_i}Y_i$, where Y_i is unit exponential and B_i is Bernoulli $(\frac{1}{2})$, independent of Y_i. Hence we may write

$$W_i = \left(\log(2p\alpha) + (1-\alpha)Y_i\right)B_i + \left(\log(2q\beta) + (1-\beta)Y_i\right)(1-B_i).$$

Similarly, under $\mathbb{P}_1$, we represent X_i as $(-1/\beta)^{1-B_i}(1/\alpha)^{B_i}Y_i$ where Y_i is again unit exponential, and B_i is Bernoulli(p), independent of Y_i, so that

$$W_i = \left(\log(2p\alpha) + \frac{1-\alpha}{\alpha}Y_i\right)B_i + \left(\log(2q\beta) + \frac{1-\beta}{\beta}Y_i\right)(1-B_i).$$

Note therefore that the conditional distribution of S_n given the binomially distributed number of positive contributions is the distribution of a constant plus the sum or the difference of two independent gamma variates, under either $\mathbb{P}_1$- or $\mathbb{P}_\infty$-measures.

4.3. *Sums and differences of gamma variates.* Let $Y_1, Y_2, \ldots$ be i.i.d unit exponential variates. Let $G_n = \Sigma_{i=1}^n Y_i$ when $n > 0$, and write $G_0 = 0$. Let G'_m be gamma$(m, 1)$ variates independent of G_n.

Throughout we let ξ and ζ be positive constants and let $g_{n,m}$ be the density of $G_n/\xi + G'_m/\zeta$. We define $h_{n,m}$ to be the density of $G_n/\xi - G'_m/\zeta$.

LEMMA 4.1. *Let $\xi \neq \zeta$ be positive. Write $x = \xi/(\xi - \zeta)$ and let $z = \zeta/(\zeta - \xi) = 1 - x$. Then, for $y > 0$, and $m, n \geq 1$,*

$$g_{n,0}(y) = \exp(-\xi y)\xi^n y^{n-1}/(n-1)!$$
$$g_{0,m}(y) = \exp(-\zeta y)\zeta^m y^{m-1}/(m-1)!$$
$$g_{1,1}(y) = x g_{0,1}(y) + z g_{1,0}(y).$$

PROOF. The first two assertions are only notational. The third assertion follows by directly computing

$$\mathbb{P}\{Y_1/\xi + Y_2/\zeta > y\} = \left(\xi e^{-\zeta y} - \zeta e^{-\xi y}\right)/(\xi - \zeta).$$

Now differentiate. □

PROPOSITION 4.2. *Let $\xi \neq \zeta$. Write $x = \xi/(\xi - \zeta)$ and let $z = \zeta/(\zeta - \xi) = 1 - x$. Then, for $y > 0$ and $m, n \geq 1$,*

$$\begin{aligned} g_{n,m}(y) = &\sum_{j=0}^{n-1} \binom{m+j-1}{j} x^j z^m g_{n-j,0}(y) \\ &+ \sum_{j=0}^{m-1} \binom{n+j-1}{j} x^n z^j g_{0,m-j}(y). \end{aligned} \tag{23}$$

PROOF. Denote the convolution operation by "$*$". Arguing probabilistically, use Lemma 4.1 to obtain

$$g_{n,m} = g_{n-1,m-1} * (x g_{0,1} + z g_{1,0}) = x g_{n-1,m} + z g_{n,m-1}. \tag{24}$$

Iterate (24) until only $g_{j,0}$ or $g_{0,j}$ terms remain. Call a sequence of x's and z's a *path*. The coefficient of $g_{n-j,0}$ is determined by the number of paths ending in z, having exactly j x's and m z's. Each such path contributes $x^j z^m$ to the coefficient. The same argument holds for the coefficient of $g_{0,m-j}$. □

We now treat the distribution of the difference of two independent gamma variates.

LEMMA 4.3. *Let ξ and ζ be positive. Write $x = \xi/(\xi + \zeta)$, and let $z = \zeta/(\zeta + \xi) = 1 - x$. Then, for $m, n \geq 1$,*

$$h_{n,0}(y) = \exp(-\xi y)\xi^n y^{n-1}/(n-1)!, \qquad \text{if } y > 0,$$
$$h_{0,m}(y) = \exp\left(-\zeta|y|\right)\zeta^m |y|^{m-1}/(m-1)!, \quad \text{if } y < 0,$$
$$h_{1,1}(y) = x h_{0,1}(y) + z h_{1,0}(y).$$

PROOF. The first two assertions are notational. The third assertion follows from the exponential distribution's lack of memory. □

PROPOSITION 4.4. *Let ξ and ζ be positive. Write $x = \xi/(\xi+\zeta)$, and let $z = \zeta/(\zeta+\xi) = 1 - x$. Then, for $m, n \geq 1$,*

$$
\begin{aligned}
h_{n,m}(y) &= \sum_{j=0}^{n-1} \binom{m+j-1}{j} x^j z^m h_{n-j,0}(y) \\
&\quad + \sum_{j=0}^{m-1} \binom{n+j-1}{j} x^n z^j h_{0,m-j}(y).
\end{aligned}
\tag{25}
$$

PROOF. The proof is formally identical to the proof of (23). □

The similarity of (23) and (25) is of course no accident, as can be seen from the characteristic functions of the sum and difference of the variates. In effect, we have evaluated the coeffcients in a partial fraction expansion of two formally identical characteristic functions.

4.4. *A numerical expression for* Δ. Let G_n and G'_m be as in Section 4.3. From (22) we see that

$$
\mathbb{P}_1\{S_n \leq 0\} = \mathbb{P}\left\{\frac{G_n}{\alpha/(1-\alpha)} + \frac{G'_{n-N}}{\beta/(1-\beta)} \leq -\left[N\log\left(\frac{p\alpha}{q\beta}\right) + n\log(2q\beta)\right]\right\},
$$

where $N \sim \text{binomial}(n, p)$, and that

$$
\mathbb{P}_\infty\{S_n > 0\} = \mathbb{P}\left\{\frac{G_N}{1/(1-\alpha)} + \frac{G'_{n-N}}{1/(1-\beta)} > -\left[N\log\left(\frac{p\alpha}{q\beta}\right) + n\log(2q\beta)\right]\right\},
$$

where $N \sim \text{binomial}(n, \frac{1}{2})$. Hence we may use standard large-deviation bounds and (23) or (25) to compute the infinite sums in (21) to prescribed accuracy.

The seventh column of Table 1 presents values of Δ calculated to within precision 0.001 by the methods described. Specifically, we truncate the series (21) and then make use of (23) and (25) to compute the individual terms. Because all summands are positive, truncation yields a lower bound on the actual value of Δ. As a reminder that we table lower bounds, we use the notation $\underline{\Delta}$ in Table 1.

Over the tabled range, $2p\alpha \vee 2q\beta \leq 1$ for $\mu \geq 0.70$, and $2p\alpha > 1 > 2q\beta$ for $p \leq 0.60$. Note that the lower bound for the limit $\underline{\Delta}$ exceeds $1/\alpha$ when $\mu \leq 0.60$ and $2p\alpha \vee 2q\beta > 1$. The lower bound $\underline{\Delta}$ is less than or equal to $1/\alpha$ when $\mu \geq 0.70$ and $2p\alpha \vee 2q\beta < 1$. These inequalities are consistent with the results of Section 4.1.

5. Monte Carlo simulation results. We discussed in Section 3 the choice of a sensitivity index of efficiency in detecting that the change in distribution

TABLE 2
Expected time to false alarm when $\nu = \infty$, by simulation

A	**100**	**200**	**300**	**400**	**450**	**500**
$\mathbb{E}_\infty\{N_A/A\} \pm$ s.e.	1.68 ± 0.03	1.72 ± 0.03	1.76 ± 0.04	1.77 ± 0.04	1.78 ± 0.04	1.79 ± 0.04
Truncations at $n = 4500$	0	0	1	2	2	6

has occurred. That index, $\lim_{k\to\infty} \mathbb{E}_k\{N - k \mid N \geq k\}$, implicitly depends on the parameter A which controls the specificity of the NPSR procedure by controlling the rate of false alarms when it is known that the prechange density is symmetric about 0. Theorem 2.2 tells us how to choose A to yield low rates of false alarms by evaluating the constant Δ, which depends on the initial density only through its symmetry.

In order to assess the NPSR procedure's suitability, we need to know how well the asymptotic approximations for low false alarm rates (i.e., for large A) actually behave for finite values of the parameters. In this section we provide evidence which suggests that the asymptotic results are useful in selecting the parameters α, β, p and A which in turn control the NPSR procedure's specificity and sensitivity.

In particular, we study the procedure for $\alpha = 0.53, \beta = 1.7$ and $p = 0.8413$. The choice of these values is discussed in Section 3. They are chosen to be appropriate when the prechange density is standard normal and the postchange density is shifted to the right by one unit. Because $2p\alpha \wedge 2q\beta < 1$, we use Section 4.1 to evaluate $\Delta = 1.887 = 1/0.53$. Use of Theorem 2.2 therefore suggests the use of the NPSR stopping time N_{419} so that ARL to false alarm equals 792.

In Table 2, we give the results of a simulation experiment which approximates $\mathbb{E}_\infty\{N_A/A\}$ for $A \in \{100, 200, 300, 400, 450, 500\}$, using 1000 realizations of the NPSR procedure. For all choices of A but the smallest, our asymptotic approximation is about 7% higher than that observed in the simulation experiment. Use of the limiting values over the range of our experiment would therefore result in the choice of threshold A about 7% too low.

To compute (6) for a single choice of k requires $O(n)$ time. Note that our procedure requires $O(n^2)$ time to compute (7) for n observations. In light of Theorem 2.2 and Jensen's inequality, the computer time required for one simulation experiment when $\nu = \infty$ will grow at least as rapidly as $O(A^3)$, imposing a limit on the size of simulation we could undertake.

In actuality, the tabled values are biased estimates. We really present means of realizations of $(t_0 \wedge N_A + A \vee R_{t_0 \wedge N_A})/2$, where we have truncated the stopping time N_A at $t_0 = 4500$. Because the sequence $R_n - n$ is a martingale under the $\nu = \infty$ measure, the downward bias is due to truncation.

We next give some indication of the rapidity with which NPSR detects a change in distribution. In Table 3, we provide a comparison of detection lags for the normal parametric Cusum and Shiryayev–Roberts procedures with the NPSR procedure having $\alpha = 0.53, \beta = 1.70$ and $p = 0.8413$. All three procedures

TABLE 3

Expected lag of Cusum and Shiryayev–Roberts parametric procedures and NPSR in detecting normal shifts, by simulation

True drift	Rule	Changepoint $\nu = 1$	$\nu = 21$	$\nu = 51$	$\nu = 101$	$\nu = 201$	$\nu = \infty$
0.75	Cusum	16.05 ± 0.21	15.59 ± 0.23	15.18 ± 0.22	15.92 ± 0.22	15.95 ± 0.23	781.0 ± 15.1
	N_A^p	16.78 ± 0.20	14.57 ± 0.18	14.98 ± 0.20	14.46 ± 0.19	14.55 ± 0.19	805.5 ± 16.2
	N_{450}	19.84 ± 0.21	15.86 ± 0.23	15.18 ± 0.24	15.67 ± 0.25	14.94 ± 0.24	802.4 ± 16.7
1.00	Cusum	10.10 ± 0.11	9.25 ± 0.10	9.26 ± 0.10	9.21 ± 0.11	9.27 ± 0.11	781.0 ± 15.1
	N_A^p	10.76 ± 0.10	9.21 ± 0.10	9.27 ± 0.10	9.09 ± 0.10	9.19 ± 0.10	805.5 ± 16.2
	N_{450}	14.92 ± 0.11	10.34 ± 0.12	9.68 ± 0.12	9.63 ± 0.13	9.73 ± 0.13	802.4 ± 16.7
1.50	Cusum	5.61 ± 0.04	5.11 ± 0.04	5.21 ± 0.05	5.18 ± 0.05	5.12 ± 0.05	781.0 ± 15.1
	N_A^p	6.20 ± 0.04	5.21 ± 0.04	5.17 ± 0.04	5.19 ± 0.04	5.21 ± 0.04	805.5 ± 16.2
	N_{450}	11.77 ± 0.04	6.33 ± 0.05	5.87 ± 0.06	5.60 ± 0.06	5.51 ± 0.06	802.4 ± 16.7

are tuned for the normal shift case of detecting a change from $\mathcal{N}(0,1)$ to $\mathcal{N}(1,1)$.

Tabled are lags in detection for various true drifts, $\mu \in \{0.75, 1.0, 1.5\}$, which occur at various times $\nu \in \{1, 21, 51, 101, 201\}$. All procedures have critical value selected to yield nominal ARL to false alarm equal to 792 under the $\nu = \infty$ measure. The values for the normal parametric Cusum and Shiryayev–Roberts procedures are taken from Pollak and Siegmund (1989). The asymptotic approximation of Theorem 2.2 would lead to a choice of $A = 792/1.89 = 419$. In light of Table 2, we have chosen to use the nonparametric rule N_{450}, a 7% adjustment over the asymptotic approximation.

Reported are sample means of $N_A - (\nu - 1)$ for those of an original 2000 simulations for which $N_A \geq \nu$. Analogous quantities are reported for the parametric procedures. In the case $\nu = 1$, the parametric procedures are clearly preferred to the nonparametric, for detecting an immediate change. On average, the cost of using our nonparametric scheme is between four and six additional observations taken postchange and before detection.

The situation when $\nu > 50$ is much more favorable to the nonparametric procedure. The average cost of using the nonparametric procedure is about $\frac{1}{2}$ additional observation before detection. On a relative basis, in the situation $\mu = 1$ for which all procedures are tuned, our nonparametric procedure takes about 5% longer on average than the parametric procedures in detectng the change. These small sample values are consistent with our claimed 97% ARE.

The situation is intermediate when $\nu = 21$ with the nonparametric procedure taking about one additional observation on average to detect the change, relative to the parametric procedures.

6. Proofs.

6.1. *Proof of* (8), *Theorem* 2.2. We show $\mathbb{E}_\infty\{N_A\} = \mathbb{E}_\infty\{R_{N_A}\} \geq A$. Let $\mathcal{F}_n$ be the sigma field generated by the signs and ranked absolute values of the first n observations. Observe that $(R_n - n, \mathcal{F}_n)$ is a martingale with zero expectation under $\mathbb{P}_\infty$, and so $\mathbb{E}_\infty\{N_A \wedge m\} = \mathbb{E}_\infty\{R_{N_A \wedge m}\}$ for any finite m. Hence, by monotone convergence,

$$\begin{aligned}(26)\qquad &\mathbb{E}_\infty\{R_{N_A} I_{\{N_A < \infty\}} + A I_{\{N_A = \infty\}}\}\\ &\quad = \lim_{m\to\infty} \mathbb{E}_\infty\{A \vee R_{N_A \wedge m}\} \geq \lim_{m\to\infty} \mathbb{E}_\infty\{N_A \wedge m\} = \mathbb{E}_\infty\{N_A\}.\end{aligned}$$

From Lemma 6.7, proved below, there exists a constant κ for which (26) is bounded above by $\kappa A + A$. Hence $\mathbb{P}_\infty\{N_A < \infty\} = 1$, and $\mathbb{E}_\infty\{R_{N_A}\} \geq \mathbb{E}_\infty\{N_A\}$. The reverse inequality follows from Fatou's lemma applied to the truncated times $N_A \wedge m$. □

6.2. *Outline of the proof of* (9), *Theorem* 2.2. Because $\mathbb{E}_\infty\{N_A\} = \mathbb{E}_\infty\{R_{N_A}\}$, we study $\mathbb{E}_\infty\{N_A\}$ by analyzing $\mathbb{E}_\infty\{R_{N_A}\}$. We make precise below that under the $\mathbb{P}_\infty$-measure, the nonparametric likelihood ratios Λ_k^{n+1} through Λ_k^{n+j} are well approximated by $\Lambda_k^n f_1(X_{n+1})/f_0(X_{n+1})$ through $\Lambda_k^n \Pi_{i=1}^j [f_1(X_{n+i})/f_0(X_{n+i})]$ when k is close to n. We then appeal to our knowledge of the overshoot in the parametric case.

Here is the heuristic reasoning underlying the assertion. If A is large, the nonparametric procedure probably will not stop until a substantial number of observations have accrued. Hence, for fairly large n, the empirical distribution of the first $n-j$ observations should be close to the true prechange distribution. If j is small relative to n, we should practically know the values of X_{n-j} through X_{n+j} from their corresponding signs and ranks of absolute values. Hence the relative magnitudes of the nonparametric likelihood ratios $\Lambda_{n-i'}^{n+i}$ should therefore be close to the analogous relative magnitudes of the parametric likelihood ratios for $0 \leq i, i' \leq j$.

By analogy with the parametric problem, we hope the influence of $\Sigma_{k \ll n} \Lambda_k^{n+j}$ upon R_{n+j} will be negligible, as will be the contribution by Λ_k^{n+j} for $k > n$ to R_{n+j} when R_n is very large. If we can establish these approximations, we would be led to believe that the $\mathbb{P}_\infty$-behavior of R_{n+1} through R_{n+j} should be much like the behavior of $R_n[f_1(X_{n+1})/f_0(X_{n+1})]$ through $R_n \Pi_{i=1}^j [f_1(X_{n+i})/f_0(X_{n+i})]$, when the base R_n is large. It will then follow that the level of overshoot in the nonparametric problem will be the same as in the parametric problem, as $A \to \infty$.

This reasoning is formalized in [Gordon and Pollak (1990), Theorem 1]. What needs to be shown is that conditions (a), (b) and (c) of that theorem are satisfied. For the sake of convenience we state the theorem here.

THEOREM 6.1. *Suppose that the following three conditions hold*:

(a) *Let* $0 < \varepsilon_1, \varepsilon_2 < 1$ *be given. There then exist positive constants* a_1, a_2 *and* a_3 *depending on* ε_1 *and* ε_2 *such that, for all* $n \geq 1$,

$$\mathbb{P}_\infty\left\{\sup_{n\varepsilon_1 \leq k \leq n(1-\varepsilon_2)} \Lambda_k^n > \exp(-a_1 n)\right\} < a_2 \exp(-a_3 n).$$

(b) *Let $0 < \varepsilon < 1$ be given. There then exist positive constants $\theta < 1, b_1, b_2$ and a set B_ε, all depending only on ε such that, for all $n \geq 1$,*

$$\mathbb{P}_\infty\left\{X_{n+1} \in B_\varepsilon \text{ and } \max_{(1-\theta)n \leq k \leq n+1}\left|1 - \frac{\Lambda_k^{n+1}}{\Lambda_k^n}\Big/\frac{f_1(X_{n+1})}{f_0(X_{n+1})}\right| > \varepsilon\right\} \leq b_1 \exp(-b_2 n)$$

and

$$\mathbb{P}_\infty\{X_{n+1} \notin B_\varepsilon\} < \varepsilon.$$

(c) *For $t \geq 1$ there exist finite functions $A_0(t)$ and $\kappa(t)$ such that $\kappa(t) \to 0$ as $t \to \infty$ and such that if $M \geq 0$ is any stopping time adapted to $\mathcal{F}_n$ and $N_A^M = N = \min\{n|\Sigma_{k=M+1}^n \Lambda_k^n \geq A\}$, then*

$$\mathbb{E}_\infty\left\{\left[\sum_{k=M+1}^{N} \Lambda_k^N\right] I_{\{\Sigma_{k=M+1}^N \Lambda_k^N \geq At\}}\Big|\mathcal{F}_M\right\} \leq \kappa(t)A,$$

uniformly for all $A > A_0(t)$, all $t \geq 1$ and all stopping times M.

Suppose also that the log-likelihood ratio $\log(f_1(X)/f_0(X))$ *has a continuous distribution when the prechange density is $f_0(\cdot)$. We may then conclude that*

$$\lim_{A \to \infty} \mathbb{E}_\infty\{N_A/A\} = \Delta(f_0, f_1).$$

An alternative condition which implies condition (c) is that, for some positive constants δ, κ' and A_0,

$$A^{-(1+\delta)}\mathbb{E}_\infty\left\{\left\{\sum_{k=M+1}^{N} \Lambda_k^N\right\}^{1+\delta}\Big|\mathcal{F}_M\right\} < \kappa',$$

for stopping times N, M and $A > A_0$ as in the statement of the theorem. Sufficiency of this condition for (c) follows from the Hölder and Markov inequalities.

We now proceed to demonstrate the validity of conditions (a)–(c). We shall repeatedly use the following elementary large-deviation inequality, where Bin (j, θ) denotes a binomial variate of j trials and θ success probability:

$$\mathbb{P}\{\mathrm{Bin}(j,\theta) - j\theta \geq j\varepsilon\} < \exp(-2j\varepsilon^2),$$

valid for all $\varepsilon > 0$.

6.3. *Proof of* (9), *Theorem* 2.2.

6.3.1. *Theorem* 6.1, *condition* (a) *holds.* Rewrite (6) as

$$\Lambda_k^n = \frac{(2p\alpha)^{U(k,n)}(2q\beta)^{V(k,n)}}{\prod_{i=1}^n \left[\big(1/(n+1-i)\big)\sum_{j=i}^n \gamma\big(\tau,(j,n),k\big)\right]}.$$

Let

$$\mu(n,k) = \frac{n+1-k}{n}\frac{\alpha+\beta}{2} + \frac{k-1}{n},$$

for $1 \le k \le n$. Consistent with this definition, let $\mu(k-1,\ k) = 1$. Now define

$$\eta(n,i,k) = \frac{1}{n+1-i}\left[\sum_{j=i}^{n} \gamma(\tau(j,n),k)\right] - \mu(n,k) \tag{27}$$

for $1 \le i,k \le n$. Note that $\mathbb{E}_\infty\{\eta(n,i,k)\} = 0$ because the average term in (27) is a sample mean, where the uniformly distributed permutation generated by the ranked absolute values has been used to select a simple random sample without replacement from a finite population consisting of the following: a binomial$(n-k+1,\ \frac{1}{2})$ number of individuals with characteristic α; the complementary binomial number with characteristic β; and $k-1$ with characteristic 1. Hence

$$\begin{aligned}
\Lambda_k^n &= \frac{(2p\alpha)^{U(k,n)}(2q\beta)^{V(k,n)}}{\prod_{i=1}^n \left[\eta(n,i,k)+\mu(n,k)\right]} \\
&= \frac{\left[p\alpha/q\beta\right]^{U(k,n)}(2q\beta)^{n+1-k}}{\big(\mu(n,k)\big)^n \prod_{i=1}^n \left[1+\eta(n,i,k)/\mu(n,k)\right]} \\
&= \frac{\left[\left(\sqrt{\alpha\beta}/\left[(\alpha+\beta)/2\right]\right)\sqrt{4pq}\right]^{n+1-k}\left[p\alpha/q\beta\right]^{U(k,n)-(n+1-k)/2}}{\prod_{i=1}^n \left[1+\eta(n,i,k)/\mu(n,k)\right]} \\
&\quad\times \frac{\left[\mu(k-1,k)/\left[(\alpha+\beta)/2\right]\right]^{k-1}}{\left[\mu(n,k)/\left[(\alpha+\beta)/2\right]\right]^{n}}.
\end{aligned}$$

Note for any $-\infty < y < \infty$, that $(1+y/x)^x$ increases in $x > 0 \vee -y$. Hence $(2\mu(n,k)/(\alpha+\beta))^n$ is increasing in n. Write

$$\lambda = \frac{\sqrt{\alpha\beta}}{(\alpha+\beta)/2}\sqrt{4pq} \tag{28}$$

and observe that $\lambda < 1$. Hence,

$$\Lambda_k^n < \frac{\lambda^{n+1-k} big\left[p\alpha/q\beta\right]^{U(k,n)-(n+1-k)/2}}{\prod_{i=1}^n \left[1+\eta(n,i,k)/\mu(n,k)\right]}. \tag{29}$$

We now need a large deviation result for simple random sampling without replacement. The inequality will be used in showing that Λ_k^n is negligible when k is far from n with high $\mathbb{P}_\infty$-probability. The connection with simple random sampling is through the uniform permutation distribution by which the ranks are assigned independently to observation order under the $\mathbb{P}_\infty$-measure.

LEMMA 6.2. *Consider a finite population with scalar characteristics* $\{\zeta_1, \dots, \zeta_m\}$. *Assume* $\Sigma_{i=1}^m \zeta_i = 0$ *and that* $\max_{i \le m} |\zeta_i| < b$. *Let* $\overline{Z}_n$ *be the mean of a simple random sample of n individuals, taken without replacement. Then there exists a constant* $\theta = \theta(b)$ *independent of n and m such that*

$$\mathbb{P}\{|\overline{Z}_n| > \varepsilon\} < 2\exp(-\theta\varepsilon^2 n).$$

PROOF. The lemma is an immediate corollary of the remark following Theorem 6 in Kemperman (1973). For the sake of completeness we sketch the proof. Write $\overline{\overline{Z}}_n$ for the mean of a size n simple random sample taken with replacement from the same population. Kemperman (1973) proves that sampling with replacement is a dilatation of sampling without replacement, hence for all t,

$$\mathbb{E}\{\exp(t\overline{Z}_n)\} \le \mathbb{E}\{\exp(t\overline{\overline{Z}}_n)\}.$$

Consider the moment generating function $f(t) = m^{-1}\Sigma_{i=1}^m \exp(t\zeta_i)$. By hypothesis, $f'(0) = 0$. Use of a Taylor series expansion yields $f(t) < 1 + b^2t^2\exp(b)/2$ for $|t| < 1$. Also, $f(t) < \exp(b|t|)$ for all t. Hence there exists a constant $K = K(b)$ for which $f(t) < \exp(Kt^2)$, for all t. Now use the Markov inequality and dilatation to establish

$$\mathbb{P}\{\overline{Z}_n > \varepsilon\} < \exp(-\varepsilon t)f^n(t/n) < \exp(Kt^2/n - \varepsilon t).$$

The inequality follows by setting $t = n\varepsilon/2K$ and repeating argument for the lower bound. □

In the next two lemmas, we bound the denominator of the right-hand side of (29).

LEMMA 6.3. *There exist positive constants* $b_1 = b_1^{(6.3)}$ *and* $b_2 = b_2^{(6.3)}$ *such that, for all* $n \ge 1$ *and all positive* $\varepsilon < 1$,

$$\mathbb{P}_\infty\left\{\prod_{i=1}^n \left[1 + \frac{\eta(n,i,k)}{\mu(n,k)}\right] < \exp(-\varepsilon n) \textit{ for some } 1 \le k \le n\right\} < \frac{b_1}{\varepsilon^5}\exp(-b_2 n\varepsilon^3).$$

PROOF Set $\delta = \min\{-(\varepsilon/2)/\log(\alpha \wedge 2\alpha/(\alpha+\beta)), 1/2\}$. For all $1 \le k \le n$,

$$\prod_{n \ge i > n(1-\delta)} \left[1 + \frac{\eta(n,i,k)}{\mu(n,k)}\right] \ge \left[\frac{\alpha}{\mu(n,k)}\right]^{n\delta} \ge \left[\alpha \wedge \frac{\alpha}{(\alpha+\beta)/2}\right]^{n\delta} \ge \exp\left(\frac{-n\varepsilon}{2}\right).$$

Hence,

$$\mathbb{P}_\infty\left\{\prod_{i=1}^{n}\left[1+\frac{\eta(n,i,k)}{\mu(n,k)}\right] < \exp(-\varepsilon n) \text{ for some } 1 \le k \le n\right\}$$
$$< \mathbb{P}_\infty\left\{\prod_{1\le i<(1-\delta)_n}\left[1+\frac{\eta(n,i,k)}{\mu(n,k)}\right] < \exp\left(\frac{-n\varepsilon}{2}\right) \text{ for some } 1 \le k \le n\right\}.$$

Write $M(n,k) = (\alpha U(k,n) + \beta V(k,n) + k - 1)/n$, a linear function of a binomial $(n+1-k, \frac{1}{2})$ variate. Also, note that $E_\infty\{M(n,k)\} = \mu(n,k)$ and that $M(n,k)$ is the mean of $\eta(n,i,k) + \mu(n,k)$ over the uniform permutation distribution induced under $\mathbb{P}_\infty$ by the locations of the ranked absolute values of observations. Hence we may apply Lemma 6.2 to $\eta(n,i,k) + \mu(n,k) - M(n,k)$. Because $1 + \eta(n,i,k)/\mu(n,k) \ge \alpha(1 + (\alpha+\beta)/2)^{-1}$, there exists a constant ξ for which $\log(1+x) \ge \xi x$, for all $0 \ge x \ge \alpha(1+(\alpha+\beta)/2)^{-1} - 1$. By Lemma 6.2, there exists a constant θ such that

$$\mathbb{P}_\infty\left\{\prod_{i=1}^{n}\left[1+\frac{\eta(n,i,k)}{\mu(n,k)}\right] < \exp(-\varepsilon n) \text{ for some } 1 \le k \le n\right\}$$
$$\le P_\infty\left\{\frac{\sum_{1\le i<n(1-\delta)}\log\big(1+\eta(n,i,k)/\mu(n,k)\big)}{\lceil n(1-\delta)\rceil - 1} < -\frac{\varepsilon/2}{1-\delta} \text{ for some } 1 \le k \le n\right\}$$
$$\le \sum_{i=1}^{\lfloor n(1-\delta)\rfloor} \mathbb{P}_\infty\left\{\xi\frac{\eta(n,i,k)}{\mu(n,k)} < -\frac{\varepsilon/2}{1-\delta} \text{ for some } 1 \le k \le n\right\}$$
$$\le \sum_{i=1}^{\lfloor n(1-\delta)\rfloor} \mathbb{P}_\infty\left\{\eta(n,i,k) + \mu(n,k) - M(n,k) < -\frac{\alpha\varepsilon/4}{\xi(1-\delta)} \text{ for some } 1 \le k \le n\right\}$$
$$+ \mathbb{P}_\infty\left\{M(n,k) - \mu(n,k) < -\frac{\alpha\varepsilon/4}{\xi(1-\delta)} \text{ for some } 1 \le k \le n\right\}$$
$$< \sum_{i=1}^{\lfloor n(1-\delta)\rfloor} \sum_{1\le k\le n} 2\exp\left(-\theta\left[\frac{\alpha\varepsilon}{4\xi(1-\delta)}\right]^2 (n+1-i)\right)$$
$$+ \sum_{1\le k\le n} \exp\left(-2\left[\frac{\alpha\varepsilon}{4\xi(1-\delta)(\beta-\alpha)}\right]^2 \frac{n^2}{n+1-k}\right)$$
$$< 3n\frac{\exp(-c\varepsilon^2\delta n)}{1-\exp(-c\varepsilon^2)},$$

for some constant c, from which the result follows. □

We now complete the proof that condition (a) holds. We prove a stronger version of the condition than is immediately needed. The extra strength is used in Section 6.4.

LEMMA 6.4. *There exist positive constants $b_i = b_i^{(6.4)}$ such that, for all positive $\varepsilon < 1$ and all $n > b_1/\varepsilon$,*

$$\mathbb{P}_\infty\left\{ \sum_{1 \le k \le n(1-\varepsilon)} \Lambda_k^n > \exp(-\varepsilon n) \right\} < \frac{b_2}{\varepsilon^5} \exp(-b_3 \varepsilon^3 n).$$

PROOF. Let λ be as defined in (28), and recall $\lambda < 1$. Choose $\delta > 0$ such that

$$\lambda \left[\frac{p\alpha}{q\beta} \vee \frac{q\beta}{p\alpha} \right]^\delta = \frac{1+\lambda}{2},$$

and write $\lambda' = (1+\lambda)/2$. Set $\varepsilon' = \varepsilon(1 + \frac{1}{2}\log\lambda')$, so that $\varepsilon' < \varepsilon$. Finally, choose $b_1 = 2\log(1-\lambda')/\log(\lambda')$. These choices imply

$$\frac{(\lambda')^{n\varepsilon}}{1-\lambda'} < \exp\big[-(\varepsilon - \varepsilon')n\big],$$

for all $n \ge b_1/\varepsilon$.

Define the event

$$B_0 = \left\{ \prod_{i=1}^{n} \left[1 + \frac{\eta(n,i,k)}{\mu(n,k)} \right] \ge \exp(-\varepsilon' n) \text{ for all } 1 \le k \le n(1-\varepsilon) \right\},$$

with complementary event $\overline{B}_0$. We now use Lemma 6.3 to show

$$\begin{aligned}
&\mathbb{P}_\infty\left\{ \sum_{1 \le k \le n(1-\varepsilon)} \Lambda_k^n > \exp(-\varepsilon n) \right\} \\
&\quad \le \mathbb{P}_\infty\left\{ \sum_{k=1}^{\lfloor n(1-\varepsilon)\rfloor} \frac{\lambda^{n+1-k} (p\alpha/q\beta)^{U(k,n)-(n+1-k)/2}}{\prod_{i=1}^n [1+\eta(n,i,k)/\mu(n,k)]} > \exp(-\varepsilon n) \right\} \\
&\quad \le P_\infty\left\{ \sum_{k=1}^{\lfloor n(1-\varepsilon)\rfloor} \frac{\lambda^{n+1-k} (p\alpha/q\beta)^{U(k,n)-(n+1-k)/2}}{\exp(-\varepsilon' n)} > \exp(-\varepsilon n) \text{ and } B_0 \right\} \\
&\qquad + \mathbb{P}_\infty\{\overline{B}_0\} \\
&\quad \le \mathbb{P}_\infty\left\{ \sum_{k=1}^{\lfloor n(1-\varepsilon)\rfloor} \lambda^{n+1-k} \left(\frac{p\alpha}{q\beta} \vee \frac{q\beta}{p\alpha} \right)^{\delta(n+1-k)} > \exp\big[-(\varepsilon-\varepsilon')n\big] \right\} + \mathbb{P}_\infty\{\overline{B}_0\} \\
&\qquad + \sum_{k=1}^{\lfloor n(1-\varepsilon)\rfloor} \mathbb{P}_\infty\left\{ \left| U(k,n) - \frac{n+1-k}{2} \right| > \delta(n+1-k) \right\} \\
&\quad < 0 + \frac{b_1^{(6.3)}}{(\varepsilon')^5} \exp\big(-b_2^{(6.3)} n(\varepsilon')^3\big) + \frac{2\exp(-\delta^2 n\varepsilon)}{1-\exp(-\delta^2\varepsilon)},
\end{aligned}$$

from which the lemma follows. □

6.3.2. *Theorem* 6.1, *condition* (b) *holds.* Define the function

$$r_n(x) = 1 + \sum_{j=1}^{n} I_{\{|X_j| \leq |x|\}}.$$

Note that $\rho(n+1, n+1) = r_n(X_{n+1})$. Recall that the vectors $Z_j^n = (\rho(j,n), \sigma_j)$ of ranks of absolute values and signs were used in Section 2 to obtain the nonparametric likelihoods $h_k(Z_1^n, \dots, Z_n^n)$ and subsequently in (6), the nonparametric likelihood ratio Λ_k^n. So long as there are no ties among the observations, there is a function ζ^n such that

$$\zeta^n\left(Z_1^n, \dots, Z_n^n, Z_{n+1}^{n+1}\right) = \left(Z_1^{n+1}, \dots, Z_{n+1}^{n+1}\right).$$

The next lemma makes precise the intuition that, with enough data, we can reconstruct the parametric log-likelihood ratio of the next observation with arbitrarily high precision. We prove a stronger form of the lemma than is needed to prove condition (b). The extra strength, provided by the conditioning in the statement of the lemma, is intended as a prototype of the proof of Lemma 6.8.

LEMMA 6.5. *Let the true prechange distribution be the double exponential. Given* $0 < \varepsilon < \frac{1}{3}$, *let* $B_\varepsilon = (\log(3\varepsilon), -\log(3\varepsilon))$. *Define, for* $k \leq n+1$,

$$L_k^n(x) = \log\left(\frac{\Lambda_k^{n+1}\Big(\zeta^n\Big(Z_1^n, \dots, Z_n^n, \big(r_n(x), I_{\{x>0\}}\big)\Big)\Big)}{\Lambda_k^n\left(Z_1^n, \dots, Z_n^n\right)} \Bigg/ \frac{f_1(x)}{f_0(x)}\right).$$

There exist positive constants $b_1 = b_1^{(6.5)}$, $b_2 = b_2^{(6.5)}$ *and* $b_3 = b_3^{(6.5)}$, *which do not depend on* ε, *such that*

$$\mathbb{P}_\infty\left\{\sup_{x \in B_\varepsilon} \sup_{n(1-\varepsilon^2) \leq k \leq n+1} \left|L_k^n(x)\right| > b_1\left(\varepsilon + \frac{1}{n\varepsilon}\right) \Bigg| \mathcal{F}_n\right\} < b_2 \exp\left(-b_3\varepsilon^4 n\right),$$

for all $n \geq 1$.

PROOF. Define $H(y) = 1 - e^{-y}$ for $y > 0$. Define the events

$$B_1 = \left\{\frac{r_n(x)}{n} \leq 1 - \varepsilon \text{ for all } x \in B_\varepsilon\right\},$$

$$B_2 = \left\{\sup_{-\infty < x < \infty}\left|\frac{r_n(x)}{n+1} - H(|x|)\right| < 3\varepsilon^2\right\},$$

with respective complements $\overline{B}_1$ and $\overline{B}_2$.

Because $|X_1|, \dots, |X_n|$ are i.i.d. as H under $\mathbb{P}_\infty$, because the empirical distribution function of their absolute values is independent of $\mathcal{F}_n$ under $\mathbb{P}_\infty$ and because $\mathbb{P}_\infty\{X_1 \notin B_\varepsilon\} = 3\varepsilon$ by construction, the result of Hu (1985) implies that there exist positive constants b_2^* and b_3^* such that, for all $n \geq 1$,

$$\mathbb{P}_\infty\{\overline{B}_1 \cup \overline{B}_2 \mid \mathcal{F}_n\} < b_2^* \exp(-b_3^* \varepsilon^4 n).$$

Write $\tau_x(j, n+1)$ for the values of the inverse rank $\tau(j, n+1)$ induced by observations $X_1, \dots, X_n, x$. Write

$$\gamma_x(j,k) = \begin{cases} \gamma(j,k), & \text{if } j \leq n, \\ I_{\{k > n+1\}} + I_{\{k \leq n+1\}}\left[\alpha I_{\{x>0\}} + \beta I_{\{x \leq 0\}}\right], & \text{if } j = n+1. \end{cases}$$

From (6), we obtain

$$\begin{aligned}
&L_k^n(x) + \log\left(\frac{f_1(x)}{f_0(x)}\right) \\
&\quad = L_k^n(x) + \log\left((2p\alpha)^{I\{x>0\}}(2q\beta)^{I\{x\leq 0\}}\right) \\
&\qquad - \left(1 - \gamma_x(n+1,k)\right)\log\left(1 - H(|x|)\right) \\
&\quad = \log\left((2p)^{I\{x>0\}}(2q)^{I\{x\leq 0\}}\right) + \log \gamma_x(n+1,k) \\
&\qquad - \log\left(\frac{1}{n+2-r_n(x)} \sum_{j=r_n(x)}^{n+1} \gamma_x\left(\tau_x(j,n+1),k\right)\right) \\
&\qquad - \sum_{i=1}^{r_n(x)-1}\left[\log\left(\frac{\gamma_x(n+1,k) + \sum_{j=i}^n \gamma\left(\tau(j,n),k\right)}{n+2-i}\right)\right. \\
&\qquad\qquad \left. - \log\left(\frac{\sum_{j=i}^n \gamma\left(\tau(j,n),k\right)}{n+1-i}\right)\right] \\
&\quad = \log\left((2p\alpha)^{I\{x>0\}}(2q\beta)^{I\{x\leq 0\}}\right) \\
&\qquad - \log\left(\frac{1}{n+2-r_n(x)} \sum_{j=r_n(x)}^{n+1} \gamma_x\left(\tau_x(j,n+1),k\right)\right) \qquad (30)\\
&\qquad - \sum_{i=1}^{r_n(x)-1} \log\left(1 + \frac{\gamma_x(n+1,k)}{\sum_{j=i}^n \gamma\left(\tau,(j,n),k\right)}\right) \qquad (31)\\
&\qquad - \sum_{i=1}^{r_n(x)-1} \log\left(\frac{n+1-i}{n+2-i}\right). \qquad (32)
\end{aligned}$$

Now consider k such that $(1-\varepsilon^2)n \leq k \leq n+1$. Recall $\alpha \leq 1 \leq \beta$. There exists a constant $b > 0$ for which we bound (30) on the event B_1 by

$$\left|1 - \frac{1}{n+2-r_n(x)} \sum_{j=r_n(x)}^{n} \gamma\left(\tau(j,n),k\right)\right| + \frac{\beta}{n+2-r_n(x)} < b\left(\varepsilon + \frac{1}{n\varepsilon}\right),$$

because $k \geq r_n(x)$ and only $n+2-k$ of the $\gamma_x(\tau_x(j, n+1), k)$ do not equal 1. Similarly, on the event B_1, (31) is bounded below by

$$-\sum_{i=1}^{r_n(x)-1} \frac{\gamma_x(n+1,k)}{n+1-i-n\varepsilon^2}$$

and above by

$$-\sum_{i=1}^{r_n(x)-1} \left[\frac{\gamma_x(n+1,k)}{n+1-i+\beta n\varepsilon^2} - \frac{\beta^2}{\left(n+1-i-n\varepsilon^2\right)^2}\right].$$

Hence, on B_1, (31) is approximable uniformly in x using

$$\left|-\sum_{i=1}^{r_n(x)-1} \log\left(1+\frac{\gamma_x(n+1,k)}{\sum_{j=i}^{n} \gamma\left(\tau(j,n),k\right)}\right) - \gamma_x(n+1,k)\log\left(1-\frac{r_n(x)}{n+1}\right)\right| < b\left(\varepsilon+\frac{1}{n\varepsilon}\right)$$

for some (possibly different) b. Because (32) is a telescoping series,

$$\left|L_k^n(x) - (1-\gamma_x(n+1,k))\log\left(1-H(\,|\,x\,|\,)\right) + (1-\gamma_x(n+1,k))\log\left(1-\frac{r_n(x)}{n+1}\right)\right|$$
$$< b\left(\varepsilon+\frac{1}{n\varepsilon}\right),$$

for some (possibly different) constant b. Finally, on the event $B_1 \cap B_2$, one obtains by a Taylor series argument that

$$\left|\log\left(1-\frac{r_n(x)}{n+1}\right) - \log\left(1-H(\,|\,x\,|\,)\right)\right| < 6\varepsilon.$$

Hence for some b we have shown $|L_k^n(x)| < b(\varepsilon+(n\varepsilon)^{-1})$ on the event $B_1 \cap B_2$, having sufficiently high conditional probability. □

Given $\varepsilon > 0$ we now establish condition (b). Let $b_1^{(6.5)}$, $b_2^{(6.5)}$ and $b_3^{(6.5)}$ be the three constants found in Lemma 6.5. Choose integer n_1 and positive ε_1 such that $\varepsilon > 3\varepsilon_1$ and so that $b_1^{(6.5)}(\varepsilon_1 + 1/(n_1\varepsilon_1)) < \varepsilon$. Set $\theta = \varepsilon_1^2$. Hence, for $n \geq n_1$, the probability required by the condition is bounded by $b_2^{(6.5)}\exp(-b_3^{(6.5)}\varepsilon_1^4 n)$. Now set $b_2 = b_3^{(6.5)}\varepsilon_1^4$ and $b_1 = b_2^{(6.5)}\exp(b_2 n_1)$ to make the condition hold for all choices of n.

6.3.3. *Theorem* 6.1, *condition* (c) *holds.* The next lemma tells us that the most recent observation—if positive—influences the nonparametric likelihood ratio in a monotone fashion.

LEMMA 6.6. *Suppose* $0 < \alpha \leq 1 \leq \beta < \infty$. *If* $X_n > 0$, *then* Λ_k^n *(and therefore* R_n*) are nondecreasing functions of* X_n.

PROOF. Rewrite (6) as

$$\Lambda_k^n = \frac{(2p\alpha)^{U(k,n)}(2q\beta)^{V(k,n)}}{\prod_{i=1}^n \left[\left(1/(n-i+1)\right)\sum_{j=i}^n \gamma\left(\tau(j,n),k\right)\right]}.$$

By hypothesis, $\gamma(n,k) = \alpha = \min\{\alpha,\beta,1\}$ for $k \le n$. Clearly, $\rho(n,n)$ is nondecreasing in X_n. As X_n progressively increases, it contributes the minimum possible value, α, to increasingly many sums in the denominator. Hence, the denominator decreases, increasing Λ_k^n. $\square$

The following lemma now establishes the validity of condition (c).

LEMMA 6.7. *Suppose $0 < \alpha \le 1 \le \beta < \infty$ and $\alpha < \beta$. Given $M \ge 0$, a stopping time adapted to $\mathcal{F}_n$, define the new stopping time*

$$N = \min\left\{n \,\middle|\, n > M \text{ and } \sum_{k=M+1}^{n} \Lambda_k^n \ge A\right\}.$$

There exist constants $\delta > 0$ and $\kappa = \kappa^{(6.7)} > 1$ such that

$$A^{-(1+\delta)}\mathbb{E}_\infty\left\{\left[\sum_{k=M+1}^{n+1} \Lambda_k^{n+1}\right]^{1+\delta} \,\middle|\, N = n+1 \text{ and } \mathcal{F}_n\right\} < \kappa,$$

uniformly for all $A > 1$, all stopping times $M \ge 0$ and all n.

PROOF. Condition first on $\{X_{n+1} < 0 \text{ and } N = n+1 \text{ and } \mathcal{F}_n\}$. Note that $A > 1$ implies that $n \ge 1$. We use (6) with $\sigma_{n+1} = 0$ to obtain

$$\begin{aligned}
&\log\left(\Lambda_k^{n+1}\right)\\
&\quad = \log\left(\Lambda_k^n\right) + \log(2q) + \log(\beta)\\
&\qquad - \log\left(\frac{1}{n - \rho(n+1,n+1) + 2}\sum_{j=\rho(n+1,n+1)}^{n+1} \gamma\left(\tau(j,n+1),k\right)\right)\\
&\qquad - \sum_{i=1}^{\rho(n+1,n+1)-1}\left[\log\left(\frac{\sum_{j=i}^n \gamma\left(\tau(j,n),k\right) + \beta}{n-i+2}\right)\right.\\
&\qquad\qquad\qquad \left. - \log\left(\frac{\sum_{j=i}^n \gamma\left(\tau(j,n),k\right)}{n-i+1}\right)\right].
\end{aligned}$$

By assumption, $\alpha \le 1 \le \beta$, so the last sum of differences is nonnegative in each term. Hence $\log(\Lambda_k^{n+1}) - \log(\Lambda_k^n) < \log(2) + \log(\beta) - \log(\alpha)$, so that $\Lambda_k^{n+1} < 2\beta\Lambda_k^n/\alpha$, on the set under consideration. Because $\Lambda_{n+1}^n = 1$, we obtain $R_{n+1} < 2\beta(R_n + 1)/\alpha < 2\beta(A+1)/\alpha$ on the event in question.

Now condition on $\{X_{n+1} > 0 \text{ and } N = n+1 \text{ and } \mathcal{F}_n\}$. We again use (6). Recall $\beta = \max\{\alpha, 1, \beta\}$, so that, for $k \le n+1$,

$$\begin{aligned}
\log(\Lambda_k^{n+1}) &\\
&= \log(\Lambda_k^n) + \log(2p) + \log(\alpha)\\
&\quad - \log\left(\frac{1}{n - \rho(n+1,n+1)+2}\sum_{j=\rho(n+1,n+1)}^{n+1}\gamma\big(\tau(j,n+1),k\big)\right)\\
&\quad - \sum_{i=1}^{\rho(n+1,n+1)-1}\left[\log\left(\frac{\sum_{j=i}^n\gamma\big(\tau(j,n),k\big)+\alpha}{n-i+2}\right) - \log\left(\frac{\sum_{j=i}^n\gamma\big(\tau(j,n),k\big)}{n-i+1}\right)\right]
\end{aligned}\tag{33}$$

$$\begin{aligned}
&\ge \log(\Lambda_k^n) + \log(2p) + \log(\alpha) - \log(\beta)\\
&\quad - \sum_{i=1}^{\rho(n+1,n+1)-1}\left[\log\left(\frac{\sum_{j=i}^n\gamma\big(\tau(j,n),k\big)+\alpha}{n-i+2}\right) - \log\left(\frac{\sum_{j=i}^n\gamma\big(\tau(j,n),k\big)}{n-i+1}\right)\right].
\end{aligned}\tag{34}$$

We continue to follow the convention that summation over a null set of indices is zero. Write

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

Note that $L(n,i,k) \le 0$.

Recall that $\mathcal{F}_n$ is the σ-algebra generated by the sequence of signs and ranks of absolute values in observation order. By virtue of the monotonicity established in Lemma 6.6, there is a random threshold $\rho_0 \le n+1$ depending on the conditioning events only through $\mathcal{F}_n$ and σ_{n+1}. The threshold is the least rank which $\rho(n+1,\ n+1)$ could attain and still permit $\Sigma_{k=M+1}^{n+1}\Lambda_k^{n+1}$ to exceed the level A.

We substitute $\rho_0 - 1$ for $\rho(n+1,n+1)$ in (34) to obtain

$$A > \frac{2p\alpha}{\beta}\sum_{k=M+1}^{n}\Lambda_k^n\exp\left[-\sum_{i=1}^{\rho_0-2}L(n,i,k)\right]\tag{35}$$

and

$$A > \frac{2p\alpha}{\beta}\exp\left[-\sum_{i=1}^{\rho_0-2}L(n,i,n+1)\right].\tag{36}$$

Write $\rho = \rho(n+1,n+1)$. Apply the identity (33) to each Λ_k^{n+1} for $k \le n+1$

to obtain

$$
\begin{aligned}
\sum_{k=M+1}^{n+1} \Lambda_k^{n+1} \le 2p \sum_{k=M+1}^{n} \Lambda_k^n \exp\left[-\sum_{i=1}^{\rho_0-2} L(n,i,k)\right] \exp\left[-\sum_{i=\rho_0-1}^{\rho-1} L(n,i,k)\right] \\
+ 2p \exp\left[-\sum_{i=1}^{\rho_0-2} L(n,i,n+1)\right] \exp\left[-\sum_{i=\rho_0-1}^{\rho-1} L(n,i,n+1)\right].
\end{aligned} \tag{37}
$$

Now choose and fix K and integer D such that $1-\alpha/\beta < K < 1$ and $-\log(1-(1-\alpha/\beta)/m) < K/m$ whenever $m \ge D$. On the conditioning events, $\gamma(n+1,k) = \alpha$ for $k \le n+1$. Hence,

$$
\begin{aligned}
&-\sum_{i=\rho_0-1}^{\rho-1} L(n,i,k) \\
&\quad = -\sum_{i=\rho_0-1}^{\rho-1} \log\left(\frac{\sum_{j=i}^n \gamma(\tau(j,n),k)+\alpha}{n-i+2}\right) - \log\left(\frac{\sum_{j=i}^n \gamma(\tau(j,n),k)}{n-i+1}\right) \\
&\quad = -\sum_{i=\rho_0-1}^{\rho-1} \log\left(1 - \frac{1}{n+2-i}\left[1 - \frac{\alpha}{\left[\sum_{j=i}^n \gamma(\tau(j,n),k)\right]\big/(n-i+1)}\right]\right) \\
&\quad \le -\sum_{i=\rho_0-1}^{\rho-1} \log\left(1 - \frac{1}{n+2-i}\left(1-\frac{\alpha}{\beta}\right)\right) \\
&\quad < -D\log(1-K) + K \sum_{i=\rho_0-1}^{(\rho-1)\wedge(n-D)} \frac{1}{n+2-i} \\
&\quad < -D\log(1-K) + K\log\left(\frac{n+3-\rho_0}{n+2-\rho}\right),
\end{aligned} \tag{38}
$$

yielding a common upper bound for all $k \le n+1$.

Now combine (35), (36), (37) and (38), yielding the inequality

$$
\sum_{k=M+1}^{n+1} \Lambda_k^{n+1} < 2\frac{\beta}{\alpha}(1-K)^{-D}\left(\frac{n+3-\rho_0}{n+2-\rho}\right)^K A \tag{39}
$$

on the sets in $\{\mathcal{F}_n$ and $X_{n+1} > 0$ and $N = n+1\}$.

The crucial observation is that, under the $\mathbb{P}_\infty$-measure, $\mathcal{F}_n$ carries information about a random permutation, but not about the unsigned magnitudes of the underlying variates which generated the permutation. Hence the rank of the $(n+1)$th absolute value is uniformly distributed over $\{1,\dots,n+1\}$. Hence, given $\mathcal{F}_n$ and $X_{n+1} > 0$ and $\rho(n+1,n+1) \ge \rho_0$, the rank $\rho(n+1,n+1)$ has uniform conditional distribution over $\{\rho_0,\dots,n+1\}$, enabling us to compute, for $0 < \delta < (1-K)/K$,

$$\mathbb{E}_\infty\left\{\left(\frac{n+3-\rho_0}{n+2-\rho}\right)^{K(1+\delta)} \middle| \mathcal{F}_n \text{ and } X_{n+1} > 0 \text{ and } N = n+1\right\}$$

$$= \frac{1}{n+2-\rho_0}\sum_{i=\rho_0}^{n+1}\left(\frac{n+3-\rho_0}{n+2-i}\right)^{K(1+\delta)}$$

$$= \frac{(n+3-\rho_0)^{K(1+\delta)}}{n+2-\rho_0}\sum_{j=1}^{n+2-\rho_0} j^{-K(1+\delta)}$$

$$< 2\big(1-K(1+\delta)\big)^{-1}.$$

Finally, apply (39) and the latter inequality to obtain

$$\mathbb{E}_\infty\left\{\left[\sum_{k=M+1}^{n+1}\Lambda_k^{n+1}\right]^{1+\delta} \middle| N = n+1 \text{ and } \mathcal{F}_n \text{ and } X_{n+1} > 0\right\}$$

$$< 2\frac{\left[2(\beta/\alpha)(1-K)^{-D}\right]^{1+\delta}}{1-K(1+\delta)}A^{1+\delta},$$

proving the lemma. □

6.4. *Sketch of proof of* (13), *Theorem* 3.1. We here show that, for arbitrary positive δ,

$$\limsup_{\nu\to\infty}\ \mathbb{E}_\nu\{N_A - \nu \mid N_A \ge \nu\} < (1+\delta)D_Q^{-1}\log(A), \tag{40}$$

for all $A > A_0$ sufficiently large. This inequality, in combination with that of Section 6.5, establishes the equalities (13) and (14).

Our procedure depends only on the signs of the observations and the relative ranks of their absolute values in the order in which they were observed. Therefore, we assume without loss of generality that the data have been transformed symmetrically about 0 to make the prechange density truly double exponential. Denote this symmetric transformation by $Q(\cdot)$.

By assumption, the true postchange measure is absolutely continuous with respect to the prechange measure. Denote the postchange density of the transformed observations by $g_Q(\cdot)$, determined by the original postchange density and the transformation $Q(\cdot)$. We continue to write $f_1(\cdot)$ for the density (5) used to define the NPSR procedure. We emphasize that in the following discussion we do not assume that $f_1(\cdot) = g_Q(\cdot)$. We remind the reader that $\mathbb{P}_1$-probability refers to the distribution in which all observations follow the true postchange distribution with density g_Q, and that the true prechange distribution is assumed to have density f_0.

By assumption, $\alpha < 1 < \beta$ and $p\alpha \ge q\beta$, implying $q < \frac{1}{2}$. Use (6) to show $\Lambda_k^n \le (2q)^{n+1-k}$ on $\{X_n < X_{n-1} < \cdots < X_1 < 0\}$. Hence on that event $R_n \le$

$2q(1-2q)^{-1}$, and so $\mathbb{P}_\infty\{N_A > n\} > 0$ for all n and all A large enough. We are therefore free to compute $\mathbb{P}_\infty$-conditional probabilities given $\{N_A > \nu\}$, for ν arbitrarily large, without fear of conditioning on null events.

We require the following straightforward extension of Lemma 6.5, stated without proof. In a slight abuse of notation, we write $\Lambda_k^n(x_1,\dots,x_n)$ for the nonparametric likelihood ratio based on the ranks induced by the data sequence $x_1,\dots,x_n$. Observe that $\Lambda_k^n = 1$ for $k > n$.

LEMMA 6.8. *Let s be a fixed positive integer and let $\varepsilon \in (0, \frac{1}{3})$. Define the set B_ε as in Lemma 6.5. Define*

$$L_k^{n,t}(x_1,\dots,x_t) = \log\left(\frac{\Lambda_k^{n+t}(X_1,\dots,X_n,\, x_1,\dots,x_t)}{\Lambda_k^n(X_1,\dots,X_n)} \Big/ \prod_{j=1+(k-(n+1))^+}^{t} \frac{f_1(x_j)}{f_0(x_j)}\right).$$

There exist positive functions $\theta = \theta(\varepsilon, s)$ and $b_i = b_i(\varepsilon, s)$ such that

$$\mathbb{P}_\infty\left\{ \sup_{1\le t\le s}\ \sup_{(x_1,\dots,x_t)\in B_\varepsilon^t}\ \sup_{(1-\theta)n\le k\le n+s} \left|L_k^{n,t}(x_1,\dots,x_t)\right| > \varepsilon \,\middle|\, \mathcal{F}_n \right\} < b_1\exp(-b_2 n).$$

Lemma 6.8 allows us to approximate the behavior of changes in nonparametric likelihood ratio by their corresponding parametric values. Such approximation is only good for probabilities. Our plan is to prove good approximations in probability, and then to bound the contribution to expectations on exceptional sets of small probability. Lemma 6.7 gives us a tool for proving such bounds. It is exploited in Lemmas 6.9 and 6.10, stated and proved next.

The parametric likelihood ratio $f_1(x)/f_0(x)$ is increasing over the full range of its argument, so long as $p\alpha \ge q\beta$. The following lemma gives an analogous monotonicity property for the nonparametric likelihood ratio. Note that the additional hypothesis $p\alpha \ge q\beta$ is needed to obtain results stronger than those given in Lemma 6.6. We refer to the ranks of the absolute values of the observations as their "absolute ranks."

LEMMA 6.9. *Let $n \ge k > m$ be given. Let $\psi(\cdot)$ be a function such that $\psi(x) \le x$ for all x. Denote by $\Lambda_j^n(x_1,\dots,x_n)$ the nonparametric likelihood ratio determined by the signs and absolute ranks induced by the sequence of measurements $x_1,\dots,x_n$. If $p\alpha \ge q\beta$, then*

$$\Lambda_k^n(x_1,\dots,x_n) \ge \Lambda_k^n\big(x_1,\dots,x_m, \psi(x_{m+1}),\dots,\psi(x_n)\big),$$

for all choices of $x_1,\dots,x_n$.

PROOF. From (6),

$$\Lambda_k^n = \frac{(p\alpha/q\beta)^{U(k,n)}(2q\beta)^{n+1-k}}{\prod_{i=1}^n [1/(n-i+1)] \sum_{j=i}^n \gamma(\tau(j,n),k)},$$

where $U(m, n)$ counts the number of positive observations seen on or after time m. By hypothesis, there are at least as many positive observations among $x_m, \dots, x_n$ as among $x_1, \dots, x_m$ $\psi(x_{m+1}), \dots, \psi(x_n)$, so that the numerator contribution to the nonparametric likelihood ratio belonging to the unmodified data is at least as large as that of the modified data.

We now show inequality in the same direction for corresponding factors in the denominator. Consider, say, the ith term in the denominator. It is an average of $n + 1 - i$ contributions, all of which are either α or 1 or β. The absolute rank of negative observations seen on or after time k can only increase, so all those observations contributing β to the sum before transformation continue to contribute β after transformation. Hence the number of contributions of magnitude β cannot be less in the ith term belonging to the transformed data.

Similarly, if an observation contributes α to the ith term in the denominator for the transformed data, then it must have contributed α to the corresponding sum in the untransformed data, because the absolute ranks of observations which remain positive can only decrease after transformation. Hence the number of contributions α in the untransformed data cannot increase after transformation, and so the total of the $n + 1 - i$ summands in the original ith term must be no greater than the $n + 1 - i$ summands making up the ith term for the transformed data. □

The next lemma uses the coupling we have just proved. The hypothesized stochastic ordering of prechange and postchange distributions gives us a function $\psi(x) = G_0^{-1}(G_1(x))$ satisfying the conditions of Lemma 6.9.

LEMMA 6.10. *Assume $G_1 \geq_{\text{stoch}} G_0$ are both continuous and that $p\alpha \geq q\beta$. There exist a constant A_0 and functions $y(A)$ and $\nu_0(A)$ such that $A > A_0$ implies*

$$\mathbb{E}_\nu\{N_A - \nu; N_A > \nu + y(A)\} < \mathbb{P}_\nu\{N_A > \nu\},$$

for all positive integers $\nu > \nu_0(A)$.

PROOF. Let A be given. We implicitly determine A_0 and $\nu_0(\cdot)$ during the course of the argument. Recall that, from invariance considerations, we assume without loss of generality that the true prechange distribution G_0 equals the double exponential distribution F_0. We continue to write $W_i = \log(f_1(X_i)/f_0(X_i))$. Let $\delta = \delta(A) = \mathbb{P}_\infty\{\exp(W_i) > 2A\}$. Let

$$y(A) = -\frac{2\log(A)}{\log\left(1 - \delta(A)/2\right)},$$

so that $(1 - \delta(A)/2)^{y(A)} \leq A^{-2}$.

Define the stopping times $T_m = \inf\{j \mid \Sigma_{k=m+1}^{m+j} \Lambda_k^{m+j} \geq A\}$. Because $G_1 \geq_{\text{stoch}} G_0$, the function $\psi(x) = G_0^{-1}(G_1(x))$ satisfies the conditions of Lemma 6.9. Let $X_j^* = X_j$ if $j < \nu$, and $X_j^* = \psi(X_j)$ if $j \geq \nu$. By continuity of G_0 and G_1, the X_j^* are i.i.d. as G_0 under the $\mathbb{P}_\infty$-measure. Use Lemma 6.9 to write

$$\begin{aligned}
&\mathbb{E}_\nu\{N_A - \nu; N_A > \nu + y\}\\
&\quad \le \mathbb{E}_\nu\{T_\nu - \nu; N_A > \nu + y\}\\
&\quad \le \mathbb{E}_\nu\Big\{T_\nu\big(X_1^*, X_2^*, \dots\big) - \nu; T_\nu\big(X_1^*, X_2^*, \dots\big) > \nu + y \text{ and } N_A > \nu\Big\}\\
&\quad = \mathbb{E}_\infty\big\{T_\nu(X_1, X_2, \dots) - \nu;\ T_\nu(X_1, X_2, \dots) > \nu + y \text{ and } N_A > \nu\big\}.
\end{aligned}$$

Now condition on $\mathcal{F}_\nu$ and apply Lemma 6.7 to show

$$\begin{aligned}
&\mathbb{E}_\nu\{N_A - \nu; N_A > \nu + y\}\\
&\quad \le (\kappa^{(6.7)}A + y)\mathbb{P}_\infty\{T_\nu > n + y \mid N_A > \nu\}\mathbb{P}_\infty\{N_A > \nu\}.
\end{aligned} \tag{41}$$

Use Lemma 6.5 to choose A_0 and $\nu_0(A)$ sufficiently large that

$$\mathbb{P}_\infty\big\{\Lambda_k^{k+1} > A \mid \mathcal{F}_k\big\} > \delta/2,$$

for all $A > A_0$ and all $k > \nu_0(A)$. Finally, observe that

$$\mathbb{P}_\infty\big\{T_\nu > \nu + y \mid \mathcal{F}_{\nu-1}\big\} \le \mathbb{P}_\infty\left\{\bigcap_{k=\nu+1}^{\nu+y} \{\Lambda_k^{k+1} < A\} \,\middle|\, \mathcal{F}_{\nu-1}\right\}.$$

By successively conditioning on earlier and earlier σ-fields, we obtain from (41) that

$$\begin{aligned}
\mathbb{E}_\nu\{N_A - \nu; N_A > \nu + y\} &\le \big(\kappa^{(6.7)}A + y\big)\big(1 - \delta/2\big)^y\mathbb{P}_\infty\{N_A > \nu\}\\
&\le \big(\kappa^{6.7}A + y\big)A^{-2}\mathbb{P}_\infty\{N_A > \nu\},
\end{aligned}$$

which can be made as small as desired by making A large. □

We now can show (40) for all $A > A_0$ sufficiently large. Let $0 < \delta < \frac{1}{2}$ be given. Let A_0, n_0 and ε be constants whose exact values will be chosen and fixed in the course of the argument. By hypothesis, $D_Q = \mathbb{E}_1\{\log\big(f_1(X_1)/f_0(X_1)\big)\} \in (0, \infty)$. Hence there exists $A_0' > 1$ such that $A > A_0'$ implies

$$\mathbb{P}_1\left\{\sum_{j=1}^{\lfloor(1+\delta)\log(A)/D_Q\rfloor} \log\left(\frac{f_1(X_j)}{f_0(X_j)}\right) < \left(1 + \frac{\delta}{2}\right)\log A\right\} < \delta.$$

We now do the following:

1. write $t = \lfloor(1+\delta)\log(A)/D_Q\rfloor$ and $s = \lceil\log(\delta/A)/\log(2\delta)\rceil$, so that $A(2\delta)^s < \delta$;
2. choose positive $\varepsilon < \log(1 + \delta/2)$ such that $\mathbb{P}_1\{X_1 \notin B_\varepsilon\} < \delta/t$ for B_ε as in Lemma 6.8;

3. define events

$$A_j = \left\{ \sum_{i=\nu+(j-1)t}^{\nu+jt-1} \log\left(\frac{f_1(X_i)}{f_0(X_i)}\right) > \left(1+\frac{\delta}{2}\right)\log A \right\},$$

$$B_j = \{X_i \in B_\varepsilon \text{ for all } \nu+(j-1)t \le i \le \nu+jt-1\},$$

$$C_j = \left\{ \sup_{(x_1,\dots,x_{tj})\in B_\varepsilon^{tj}} \left| \log\left(\frac{\Lambda_{\nu+t(j-1)}^{\nu+tj-1}(X_1,\dots,X_{\nu-1},x_1,\dots,x_{tj})}{\prod_{j=\nu+(j-1)t}^{\nu+jt-1}[f_1(x_j)/f_0(x_j)]} \right) \right| < \varepsilon \right\},$$

with respective complements $\overline{A}_j, \overline{B}_j$ and $\overline{C}_j$; note that $N_A \le \nu + tj - 1$ on $A_j \cap B_j \cap C_j$, because then $\Lambda_{\nu+(j-1)t}^{\nu+jt-1} > A$;

4. use Lemma 6.8 to choose n_0 such that $\nu > n_0$ implies

$$\mathbb{P}_\infty\{\overline{C}_j \mid \mathcal{F}_{\nu-1}\} < \frac{\delta}{As^2},$$

for any $1 \le j \le s$.

Given $1 \le r \le s$, observe that the events C_j are measurable with respect to the first $\nu - 1$ observations, so that

$$\begin{aligned}
&\mathbb{P}_\nu\{N_A > \nu + tr - 1\} \\
&\quad \le \mathbb{P}_\nu\left\{N_A \ge \nu \text{ and } \bigcap_{j=1}^{r}\{\overline{A}_j \cup \overline{B}_j \cup \overline{C}_j\}\right\} \\
&\quad \le \mathbb{P}_\nu\left\{N_A \ge \nu \text{ and } \bigcup_{j=1}^{r}\overline{C}_j\right\} + \left(\delta + t\left(\frac{\delta}{t}\right)\right)^r \mathbb{P}_\nu\{N_A \ge \nu\} \qquad (42)\\
&\quad \le \mathbb{P}_\infty\left\{N_A \ge \nu \text{ and } \bigcup_{j=1}^{r}\overline{C}_j\right\} + (2\delta)^r \mathbb{P}_\infty\{N_A \ge \nu\} \\
&\quad \le \left[(2\delta)^r + \frac{\delta}{As}\right]\mathbb{P}_\infty\{N_A \ge \nu\}.
\end{aligned}$$

Finally, we use (42) to bound probabilities and Lemmas 6.9 and 6.10 to bound the expectation taken over an event of sufficiently small probability, thereby showing

$$\begin{aligned}
&\mathbb{E}_\nu\{N_A - \nu; N_A \ge \nu\} \\
&\quad \le t\left[\mathbb{P}_\nu\{N_A \ge \nu\} + \sum_{r=1}^{s}\mathbb{P}_\nu\{N_A \ge \nu + tr - 1\}\right] \\
&\quad\quad + \mathbb{E}_\nu\{N_A - (\nu + st); N_A \ge \nu + st\} \\
&\quad < \left[t\left(1 + \frac{2\delta}{1-2\delta} + \frac{\delta}{A}\right) + (b^{(6.7)}\sqrt{st}A + st)\left[(2\delta)^s + \frac{\delta}{As}\right]\right]\mathbb{P}_\infty\{N_A \ge \nu\}.
\end{aligned}$$

Because the inequality holds for ν sufficiently large, and because $st = O(\log^2 A)$, we have established (40).

6.5. *Sketch of proof of* (14), *Theorem* 3.1. We finally show that, for arbitrary positive ε,

$$\liminf_{\nu \to \infty} \mathbb{E}_\nu\{N_A - \nu \mid N_A \ge \nu\} > (1-\varepsilon) D_Q^{-1} \log(A),$$

for all $A > A_0$ sufficiently large. We continue to assume that the data have been transformed by Q to make the prechange density truly f_0. We need a lower bound; hence we show $\mathbb{P}_\nu\{N_A - \nu > (1-\varepsilon)D_Q^{-1}\log(A) \mid N_A \ge \nu\}$ can be made arbitrarily close to 1 for sufficiently large ν. The chief technical difficulty comes from our taking limits in which the probability of the conditioning event tends to 0. Fortunately, Lemma 6.8 is stated conditionally, so the assertion will be true if R_ν is small when the change occurs. This intuition is formalized in Lemmas 6.11 and 6.12.

We continue to denote the parametric log-likelihood ratio by W_i. We assume throughout this subsection that $\alpha < 1$, so that $\mathbb{E}_\infty\{\exp[(1+\alpha)W_1]\} < \infty$. Because W_1 has a moment generating function in a neighborhood of 1, there exist positive constants δ and ω such that

$$\mathbb{P}_\infty\left\{\exp\left(\sum_{i=1}^{j} W_i\right) < (1-\delta)\exp(-2\omega j) \text{ for all } j \ge 1\right\} > \delta. \tag{43}$$

Because a false alarm is raised when a very few adjacent observations apparently behave as if they came from the F_1 postchange distribution, it is often useful to think of N_A under $\mathbb{P}_\infty$ as if it had an exponential distribution with mean ΔA. The following lemma says that this heuristic device is not too bad, even in the extreme tails of the distribution of N_A.

LEMMA 6.11. *There exist positive constants A_0 and $b_i = b_i^{(6.11)}$ and a function $n_0(A)$ such that, for all $A > A_0$,*

$$\mathbb{P}_\infty\{N_A \ge \nu\} > b_1 \exp\left[-\frac{b_2 \nu(\log^4 A)}{A}\right], \tag{44}$$

for all $\nu > 0$, and

$$\mathbb{P}_\infty\{N_A \ge \nu\} < b_3 \mathbb{P}_\infty\left\{N_A \ge \nu + b_2 \frac{A}{\log^4 A}\right\}, \tag{45}$$

for all $\nu > n_0(A)$.

PROOF. Let A_0 be a positive constant whose value is only implicitly determined. We assume throughout that $A > A_0$. The values of $\varepsilon = O(\log^{-2} A)$ and $n_0 = O(\log^{13} A)$ will be specified immediately before (47).

Let δ and ω be as in (43), and let $s = \lceil 2\log(A)/\omega \rceil$, so that $\exp(-\omega s) \leq A^{-2}$. Now define sums for $m \geq n$ by the following:

$$\begin{aligned} R_m^{\text{top}} &= \sum_{n<k\leq m} \Lambda_k^m; \\ R_m^{\text{mid}} &= \sum_{(1-\varepsilon^2)n<k\leq n} \Lambda_k^m; \\ R_m^{\text{bot}} &= \sum_{1\leq k\leq (1-\varepsilon^2)n} \Lambda_k^m; \\ R_m^{\text{low}} &= R_m^{\text{bot}} + R_m^{\text{mid}}. \end{aligned}$$

Note that, for $m \geq n$, R_m^{low} is an $\mathcal{F}_m$-martingale and that R_m^{top} is a submartingale.

The set B_ε is defined in Lemma 6.5. Recall that $B_{2\varepsilon} \subset B_\varepsilon$. We require $\varepsilon < \frac{1}{12}$ and $n\varepsilon > 1$. For t a positive integer, define the events

$$B_0 = \left\{ 4\varepsilon < \frac{\rho(n+j, n+j)}{n+j} < 1 - 4\varepsilon \text{ for all } 1 \leq j \leq s \right\},$$

$$B_0' = \{X_{n+j} \in B_{4\varepsilon} \text{ for all } 1 \leq j \leq s\},$$

$$B_1 = \{R_{n+j}^{\text{mid}} < (1-\delta) R_n \exp(-j\omega) \text{ for all } 1 \leq j \leq s\},$$

$$B_1' = \left\{ \max_{1\leq j\leq s} \max_{(1-\varepsilon^2)n\leq k\leq n} \log\left(\frac{\Lambda_k^{n+j}/\Lambda_k^{n+j-1}}{f_1(X_{n+j})/f_0(X_{n+j})} \right) < \omega \right\},$$

$$B_1'' = \left\{ \exp\left(\sum_{i=n+1}^{n+j} W_i \right) < (1-\delta)\exp(-2j\omega) \text{ for all } 1 \leq j \leq s \right\},$$

$$B_2 = \left\{ R_{n+j}^{\text{top}} < \frac{\delta A}{2} \text{ for all } 1 \leq j \leq st \right\},$$

$$B_3 = \{R_{n+j}^{\text{bot}} < 1 \text{ for all } 1 \leq j \leq st\},$$

with respective complements $\overline{B}_i$.

We first show that B_0 almost contains B_0'. By the independence of the empirical distribution and the order in which the observations appear,

$$\begin{aligned} &\mathbb{P}_\infty \left\{ X_{n+j} \in B_{4\varepsilon} \text{ and } \frac{\rho(n+j, n+j)}{n+j} \notin (4\varepsilon, 1-4\varepsilon) \,\middle|\, \mathcal{F}_n \right\} \\ &\quad < \mathbb{P}\{1 + \text{Bin}(n+j-1, 1-6\varepsilon) \geq (n+j-1)(1-4\varepsilon)\} \\ &\qquad + \mathbb{P}\{1 + \text{Bin}(n+j-1, 6\varepsilon) \leq (n+j-1)4\varepsilon\} \\ &\quad < 2\exp(-2\varepsilon^2 n), \end{aligned}$$

valid for all $n\varepsilon > 1$ and all $j \geq 1$. Hence,

$$\mathbb{P}_\infty\{\overline{B}_0 B_0' \mid \mathcal{F}_n\} < 2s\exp(-2\varepsilon^2 n).$$

By the independence of the observations $\mathbb{P}_\infty\{B_0' \mid \mathcal{F}_n\} > 1 - 12s\varepsilon$, so that

$$\mathbb{P}_\infty\{B_0 \mid \mathcal{F}_n\} \geq \mathbb{P}_\infty\{B_0 B_0' \mid \mathcal{F}_n\} > 1 - s\big(12\varepsilon + 2\exp(-2\varepsilon^2 n)\big). \tag{46}$$

We now show that $B_0 B_1$ almost contains $B_0' B_1' B_1''$.Choose n to make $s < n\varepsilon^2$. Hence $n(1-\varepsilon^2) > (n+s)(1-4\varepsilon^2)$. Use (46) and Lemma 6.5 to conclude

$$\mathbb{P}_\infty\{B_0 B_0' B_1' \mid \mathcal{F}_n\} > 1 - s\big(12\varepsilon + 2\exp(-2\varepsilon^2 n)\big) - b_2^{(6.5)} s \exp\big(-b_3^{(6.5)}\varepsilon^4 n\big),$$

valid for all ε and n such that $b_1^{(6.5)}(2\varepsilon + (2\varepsilon n)^{-1}) < \omega$. Now use (43) to obtain that

$$\mathbb{P}_\infty\{B_0 B_0' B_1' B_1'' \mid \mathcal{F}_n\} \geq \delta - s\Big(12\varepsilon + 2\exp\big(-2\varepsilon^2 n\big)\Big) - b_2^{(6.5)} s \exp\big(-b_3^{(6.5)}\varepsilon^4 n\big).$$

Because $B_1' B_1'' \subset B_1$, we may choose and fix $\varepsilon(A) = c_1/\log^2 A$ and $n_0 = n_0(A) = \lfloor c_2 \log^{13} A \rfloor$ such that

$$\mathbb{P}_\infty\{B_0 B_1 \mid \mathcal{F}_n\} \geq \mathbb{P}_\infty\{B_0 B_0' B_1' B_1'' \mid \mathcal{F}_n\} \geq \frac{\delta}{2}, \tag{47}$$

for all $n > n_0$. Set $t = \lceil \delta A \varepsilon^2/(28s) \rceil$ and write $\alpha = st = O(A\log^{-4} A)$.

Because $B_1 B_2 B_3 \cap \{N_A > n\} \subset \{N_A > n+s\}$, we use (46) and (47) and apply the submartingale maximum inequality to both R_{n+j}^{top} and R_{n+j}^{mid}, thus concluding that

$$\begin{aligned}
&\mathbb{P}_\infty\{N_A > n+\alpha\} \\
&\quad \geq \mathbb{P}_\infty\{N_A > n+\alpha \text{ and } B_0 B_1 B_2 B_3\} \\
&\quad > \mathbb{P}_\infty\{N_A > n \text{ and } B_0 B_1\}\left(1 - \frac{A\exp(-\omega s)}{1}\right) \\
&\qquad - \mathbb{P}_\infty\{N_A > n\}\frac{\alpha}{\delta A/2} - \mathbb{P}_\infty\{\overline{B}_3\} \\
&\quad > \mathbb{P}_\infty\{N_A > n\}\left[\frac{\delta}{2} - \frac{2\alpha+1}{\delta A}\right] - c_3\varepsilon^{-10}\exp\big(-c_4\varepsilon^6 n\big) - \exp\big(-\varepsilon^2 n\big),
\end{aligned}$$

where we use Lemma 6.4 and $(c_3, c_4) = (b_2^{(6.4)}, b_3^{(6.4)})$ for the last inequality. Hence

$$\mathbb{P}_\infty\{N_A > n+\alpha\} > \frac{\delta}{4}\mathbb{P}_\infty\{N_A > n\} - 2c_3\varepsilon^{-10}\exp\big(-c_4\varepsilon^6 n\big), \tag{48}$$

whenever $A > A_0$ and $n > n_0$. Write $\theta = \delta/4$. Iterate the inequality to show

$$\begin{aligned}
&\mathbb{P}_\infty\{N_A > n + j\alpha\} + \sum_{i=1}^{j} 2c_3\varepsilon^{-10}\theta^{j-i}\exp\Big[-c_4\varepsilon^6\big(n + (i-1)\alpha\big)\Big] \\
&\qquad > \theta^j \mathbb{P}_\infty\{N_A > n\}.
\end{aligned}$$

Now choose A_0 to ensure as well that $\theta^{-1}\exp(-c_4\varepsilon^6\alpha) < \frac{1}{2}$, and set $n = 2n_0$ so that

$$\begin{aligned}\mathbb{P}_\infty\{N_A > 2n_0 + j\alpha\} &> \theta^j\mathbb{P}_\infty\{N_A > 2n_0\} - 4c_3\varepsilon^{-10}\theta^{j-1}\exp\big(-2c_4\varepsilon^6 n_0\big)\\ &> \theta^j\left[1 - \frac{2n_0}{A} - 4c_3\theta^{-1}\varepsilon^{-10}\exp\big(-2c_4\varepsilon^6 n_0\big)\right].\end{aligned}$$

Because $n_0 = O(\log^{13} A)$, the lower bound is itself bounded below by $\theta^j/2$ for A sufficiently large, from which (44) follows.

Assertion (45) follows immediately from (44) and (48) because the latter's remainder term decays more rapidly than the lower bound for $\mathbb{P}\{N_A > n\}$. □

The next lemma formalizes the intuition that if the nonparametric Shiryayev–Roberts statistic R_n is not very large, then it is very small.

LEMMA 6.12. *There exist constants $b_i = b_i^{(6.12)}$ and A_0 and a function $n_0(A)$ such that both $A > A_0$ and $n > n_0(A)$ implies that*

$$\mathbb{E}_\infty\{R_n;\, N_A > n\} < b_1\log^2(A)\mathbb{P}_\infty\{N_A > n - b_2\log^2 A\}$$

and that

$$\mathbb{P}_\infty\{N_A > n \mid N_A > n - b_2\log^2 A\} > 1 - \frac{b_1\log^2 A}{A}.$$

PROOF. We implicitly determine A_0 and $n_0(A)$ in the course of the argument. Throughout we assume $n > n_0(A)$ and $A > A_0$.

Let $s = \lceil\omega^{-1}\log A\rceil$ and $t = \lceil-\log(A)/\log(1-\delta/2)\rceil$, where ω and δ are defined in (43). Let $\varepsilon = 1/\log A$. Write $n^* = n - st$ and $n^*_{(i)} = n^* + is$. We shall define below an auxiliary $\mathcal{F}_n$-stopping time N^* taking values $n^*_{(i)}$ for $1 \le i \le t$. For any such stopping time,

$$\mathbb{E}_\infty\{R_n;\, N_A > n\} = \sum_{i=1}^{t}\mathbb{E}_\infty\{R_n;\, N_A > n \text{ and } N^* = n^*_{(i)}\}$$

$$< A\mathbb{P}_\infty\{N^* = n \text{ and } N_A > n^*\} \tag{49}$$

$$+\,\mathbb{E}_\infty\left\{\sum_{n^* < k \le n}\Lambda_k^n;\, N_A > n^*\right\} \tag{50}$$

$$+\sum_{i=1}^{t-1}\mathbb{E}_\infty\left\{\sum_{(1-\varepsilon^2)n^* < k \le n^*}\Lambda_k^n;\, N_A > N^* = n^*_{(i)}\right\} \tag{51}$$

$$+\,A\mathbb{P}_\infty\left\{\sum_{1 \le k \le (1-\varepsilon^2)n^*}\Lambda_k^{n^*} > \exp(-\varepsilon^2 n^*)\right\} + \exp(-\varepsilon^2 n^*). \tag{52}$$

We now proceed to construct a stopping time N^* that, with high probability, will stop before time n and then will leave R_{N^*} small. This is accomplished by a construction similar to that used in the proof of Lemma 6.11. Given $1 \le i \le t$, define events

$$B_0(i) = \left\{ 4\varepsilon < \frac{\rho\left(n^*_{(i-1)} + j, n^*_{(i-1)} + j\right)}{n^*_{(i-1)} + j} < 1 - 4\varepsilon \text{ for all } 1 \le j \le s \right\},$$

$$B_1(i) = \left\{ \sum_{(1-\varepsilon^2)n^* < k \le n^*} \Lambda_k^{n^*_{(i-1)}+j} < (1-\delta) R_{n^*_{(i-1)}} \exp(-j\omega) \text{ for all } 1 \le j \le s \right\}.$$

Now define the $\mathcal{F}_n$-stopping time

$$N^* = n \wedge \min\left\{ n^*_{(i)} \mid 1 \le i \le t \text{ and } B_0(i) \cap B_1(i) \text{ is true} \right\},$$

following the usual convention that an infimum over an empty set is infinite. The identical argument leading to (47) shows that

$$\mathbb{P}_\infty\left\{ B_0(i) \cap B_1(i) \mid \mathcal{F}_{n^*_{(i-1)}} \right\} \ge \frac{\delta}{2},$$

for all $1 \le i \le t$ and all A and n sufficiently large.

By successive conditioning and repeated application, the last inequality implies that (49) is bounded above by $A(1 - \delta/2)^{t-1} \mathbb{P}_\infty\{N_A > n^*\}$. Recall that $\Lambda_k^{n^*} = 1$ for $n^* < k$. Because it is composed of likelihood ratios, (50) is equal to $st\mathbb{P}_\infty\{N_A > n^*\}$. The martingale property of the individual likelihood ratios allows us to bound (51) by

$$\begin{aligned} \sum_{i=1}^{t-1} \mathbb{E}_\infty &\left\{ \sum_{(1-\varepsilon^2)n^* < k \le n^*} \Lambda_k^{n^*_{(i)}}; N_A > N_A \wedge N^* = n^*_{(i)} \right\} \\ &< A \exp(-s\omega) \mathbb{P}_\infty\{N_A > n^*\}. \end{aligned}$$

Finally, we use Lemmas 6.4 and 6.11 to bound (52) by $\mathbb{P}_\infty\{N_A > n^*\}$ for all A and n sufficiently large. Hence,

$$\mathbb{E}_\infty\{R_n; N_A > n\} < [2 + st + 1 + 1] \mathbb{P}_\infty\{N_A > n^*\},$$

proving the first assertion and implicitly determining the constant b_i.

Finally, we prove the second assertion using the first. Because R_n is a submartingale and the Λ_k^n are martingales,

$$\begin{aligned} \mathbb{P}_\infty\{n > N_A > n^*\} &< A^{-1} \mathbb{E}_\infty\{R_n; n > N_A > n^*\} \\ &< A^{-1} \mathbb{E}_\infty\{R_n; N_A > n^*\} \\ &< \frac{b_1 \log^2 A}{A} \mathbb{P}_\infty\{N_A > n^* - b_2 \log^2 A\} \\ &< \frac{b_1 b_3^{(6.11)} \log^2 A}{A} \mathbb{P}_\infty\{N_A > n^*\}, \end{aligned}$$

for some constant b_1, where the last inequality follows from (45) for A and n large enough. □

The rest of the proof of (14) is rather straightforward. Use (6) to show that, for all $k \leq n$,

$$\frac{\Lambda_k^{n+1}}{\Lambda_k^n} \leq 2p(n+1).$$

Recall that D_Q was defined in the statement of Theorem 3.1. Choose and fix A sufficiently large, $s = O(\log A)$, $\delta > 0$ and $\varepsilon = O(\log^{-2} A)$. (The last is for use with Lemmas 6.8 and 6.4.) Conditional on the event $\{N_A \geq n\}$, an $\mathcal{F}_{n-1}$-event, Lemmas 6.8, 6.4, 6.11 and 6.12 and the strong law of large numbers imply there exist $n_0(A)$ and $\zeta = \zeta(A) > 0$ such that $n > n_0(A)$ implies

$$R_{n-1+s} < O(\log^2 A)\Big[\exp(s(D_Q + \delta)) + \big(2p(n+s)\big)^s \exp(-n\zeta)\Big]$$

with arbitrarily high $\mathbb{P}_n\{\cdot \mid N_A \geq n\}$-probability, proving the required lower bound.

7. Remarks.

1. A similar analysis could be done when the postchange distribution is a more complicated mixture of positive and negative exponentials. However, the technical difficulties are formidable and—given the high relative efficiency of the simple schemes—do not currently promise rewards commensurate with the effort necessary to analyze them.
2. It is conceivable that the NPSR scheme might be substantially robust when faced with symmetric contamination.
3. It seems intuitively clear from the proof of Theorem 2.2 that the nonparametric likelihood ratios Λ_k^n should be almost the same as the parametric likelihood ratios when $\nu = \infty$ and k is close to n. Because, when $\nu = \infty$, only these Λ_k^n contribute substantially to R_n, it seems plausible that $R_n = R_n^p + o_p(1)$. Such an argument might provide an alternative proof to Theorem 2.2. This approach might also facilitate an analysis of a nonparametric Cusum scheme.
4. One might be tempted to use the methods of this paper for hypothesis testing. Efficiency against contiguous alternatives will be low. Difficulties arise because the sign test is the locally most powerful rank test for shift alternatives with the double exponential density. Similarly, when $\nu = 1$, all the information about the shift is essentially in the signs of the observations, again resulting in the relatively unsatisfactory lags seen in Table 3.
5. Our approach can be used to detect a change in distribution from $G_0(\cdot)$ prechange to $G_1(\cdot)$ postchange in the case when it is known that G_1 is stochastically larger than G_0. See Gordon and Pollak (1991) for details.

Acknowledgments. The authors thank David Siegmund and Charles Bell for a number of enlightening conversations and remarks, and are very grateful to an Associate Editor and a referee for their careful reading and insightful suggestions.

REFERENCES

BHATTACHARYA, P. K. and FRIERSON, D. (1981). A nonparametric control chart for detecting small disorders. *Ann Statist.* **9** 544–554.

GORDON, L. and POLLAK, M. (1990). Average run length to false alarm for surveillance schemes designed with partially specified prechange distribution. Unpublished manuscript.

GORDON, L. and POLLAK, M. (1991). A robust surveillance scheme for stochastically ordered alternatives. Unpublished manuscript.

HU, I. (1985). A uniform bound for the tail probability of Kolmogorov–Smirnov statistics. *Ann. Statist.* **13** 821–826.

KEMPERMAN, J. H. B. (1973). Moment problems for sampling without replacement I; II; III. *Indag. Math.* **35** 149–164; 165–180; 181–188.

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

LUCAS, J. M. and CROSIER, R. B. (1982). Fast initial response for CUSUM quality control schemes: give your CUSUM a head start. *Technometrics* **24** 199–205.

MATHWORKS (1989). *Pro-Matlab User's Guide.* The Mathworks, South Natick, MA.

MCDONALD, D. (1990). A Cusum procedure based on sequential ranks. *Naval Res. Logist.* **37** 627–646.

MILLER, R. G. (1972). Sequential rank tests—one sample case. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **1** 97–108. Univ. California Press, Berkeley.

MOUSTAKIDES, G. V. (1986). Optimal stopping times for detecting changes in distributions. *Ann. Statist.* **14** 1379–1387.

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. (1985). A diffusion process and its application to detecting a change in the drift of Brownian motion. *Biometrika* **72** 267–280.

POLLAK, M. and SIEGMUND, D. (1989). Personal communication.

RITOV, Y. (1990). Decision theoretic optimality of the CUSUM procedure. *Ann. Statist.* **18** 1464–1469.

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.

SAVAGE, I. R. and SETHURAMAN, J. (1966). Stopping time of a rank order sequential test based on Lehmann alternatives. *Ann. Math. Statist.* **37** 1154–1160.

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.

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

WOODROOFE, M. (1983). On sequential rank tests. In *Recent Advances in Statistics* (M. H., Rizvi, J. S. Rustagi and D. Siegmund, eds.) 115–140. Academic, New York.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SOUTHERN CALIFORNIA
UNIVERSITY PARK
LOS ANGELES, CALIFORNIA 90089–1113

DEPARTMENT OF STATISTICS
HEBREW UNIVERSITY OF JERUSALEM
91905 JERUSALEM
ISRAEL