The Annals of Statistics
1995, Vol. 23, No. 4, 1350–1375

A ROBUST SURVEILLANCE SCHEME FOR STOCHASTICALLY ORDERED ALTERNATIVES

BY LOUIS GORDON[1,2] AND MOSHE POLLAK[2]

University of Southern California and Hebrew University of Jerusalem

We sequentially observe independent observations $X_1, X_2, \ldots$ such that initially they have distribution G_0; at some unknown time ν they become stochastically larger, having distribution G_1. Neither G_0 nor G_1 is fully specified. We wish to detect that a change has taken place as soon as possible after its occurrence, subject to a constraint on the rate of false alarms. We derive a family of nonparametric sequential procedures based on ranks, with noncontiguous alternatives in mind. Large-sample approximations to the operating characteristics are obtained analytically.

The proposed procedures all possess robustness of validity, because they are based on ranks. Near-optimal sensitivity can be obtained for specific alternatives by choosing an appropriate procedure. For example, when observations are normally distributed with one standard deviation shift in mean postchange, an appropriate nonparametric surveillance scheme yields 97% asymptotic relative efficiency, as compared to the optimal procedure when all distributional parameters are known.

Our procedures are computationally feasible. Monte Carlo experiments confirm the applicability of the asymptotic calculations, including high levels of efficiency, for sample sizes met in practice.

1. Introduction and summary. We consider observations $X_1, X_2, \ldots$ taken sequentially. Under the measure $\mathbb{P}_\infty^G$, the observations are i.i.d. with continuous distribution G_0. Under the measure $\mathbb{P}_k^G$, the prechange observations $X_1, \ldots, X_{k-1}$ are i.i.d. with distribution G_0 which are independent of the postchange observations $X_k, \ldots$; the latter are i.i.d. with continuous distribution G_1.

The change-point problem (in the Russian literature, the problem of disruption) is to determine a sequential detection scheme which, on the basis of the first n observations, will raise an alarm when it is plausible that the last few observations have the postchange distribution G_1. The cusum procedures introduced by Page (1954) are known to be optimal in a very strong sense when the prechange and postchange distributions are completely specified,

Received September 1991; revised July 1994.

[1]Research supported in part by the John Simon Guggenheim Memorial Foundation and the United States–Israel Educational Foundation.

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

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

Key words and phrases. Cusum, change-point, disruption, nonparametric.

and are indeed the distributions from which prechange and postchange observations are drawn.

In this paper, we study the detection problem when the prechange distribution is not completely known. One reason such a problem is of substantial theoretical and practical import is shown in Table 1, in which we give Siegmund's [(1985), Theorem (10.17) and page 225] approximation to the average run length (ARL) to false alarm of a cusum procedure appropriate for detecting a shift from $\mathcal{N}(0,1)$ to $\mathcal{N}(1,1)$, when the prechange distribution is actually $\mathcal{N}(0,\sigma^2)$. [We denote the normal distribution with mean μ and variance σ^2 by $\mathcal{N}(\mu,\sigma^2)$.] To facilitate comparison with published simulations, the cusum we select has nominal ARL to false alarm equal to 792. The procedure is extremely sensitive to misspecification of the prechange standard deviation, which causes a substantial nonrobustness of validity. The derivative $(d/d\sigma)$ of the ratio of actual ARL to false alarm as a fraction of nominal ARL to false alarm is about -10 when $\sigma = 1$; a 5% underspecification in prechange standard deviation results in a procedure having actual ARL to false alarm 30% less than nominal. Van Dobben de Bruyn [(1968), Section 2.4] remarks on the nonrobustness of two-sided cusum procedures to small misspecifications of variance.

An extensive study of robustness appears in Lucas and Crosier (1982). They present approximations to ARL to false alarm and expected lag to detection for a number of robust alternatives to the standard cusum process, with emphasis on contamination models for normal shift alternatives. Their computational method is to approximate the random process of cusums as a finite-state Markov chain with one or two absorbing states. Results again show a substantial nonrobustness of validity (i.e., a decrease in ARL to false alarm) when the parametric cusum tuned to detect a shift from $\mathcal{N}(0,1)$ to $\mathcal{N}(1,1)$ is truly confronted with i.i.d. observations from a contaminated normal distribution having greater variability.

To remedy this nonrobustness, we look for a family of nonparametric procedures with the hope that some member of the family might have acceptably high postchange sensitivity. Work of Pollak and Siegmund (1975) suggests that average lag to detection is relatively insensitive to misspecification of postchange distribution, depending mainly on the postchange expectation of the log-likelihood. Inspired by the robustness of the Wilcoxon statistic, we consider procedures based on ranks. Because the procedures are based on ranks, they all possess robustness of validity. Because our problem is sequential, we use the sequential ranks (formally defined in the next

TABLE 1
ARL to false alarm for a cusum $\mathcal{N}(0,1)$ versus $\mathcal{N}(1,1)$ with nominal ARL to false alarm 792 when the prechange distribution is really $\mathcal{N}(0,\sigma^2)$

True σ	**0.98**	**0.99**	**1.00**	**1.01**	**1.02**	**1.03**	**1.04**	**1.05**	**1.10**
$\mathbb{E}_\infty\{N_{\text{cusum}}\}/790$	1.20	1.09	1.00	0.92	0.84	0.77	0.71	0.66	0.46
$\mathbb{E}_\infty\{N_{\text{cusum}}\}$	951	865	790	723	663	610	563	521	364

section, where we present our main results). Our statistic (a sum of likelihood ratios computed under various hypotheses of time of change) and stopping rule (the first level crossing of the statistic) follow the original treatments of Shiryayev (1963) and Roberts (1966). Our work is couched in a frequentist context and does not use a prior distribution for the time of change.

2. Main results. In this section we first give the likelihood ratio upon which we base the Shiryayev–Roberts detection procedure. Next we present the major theorems which yield asymptotic approximations for the procedure's operating characteristics. Proofs are deferred until the end of the paper.

It is natural from invariance considerations to choose the first n of the sequential ranks $\rho_n = \sum_{j=1}^n I_{\{X_j \le X_n\}}$ as the basis for a detection procedure for $n = 1, 2, \dots$. The vector of the first n sequential ranks is denoted $\boldsymbol{\rho}_n$.

It is sometimes more convenient to describe our results in terms of the ranked observations relative to data seen by time n, denoted $\rho(i, n) = \sum_{j=1}^n I_{\{X_j \le X_i\}}$. Note that $\rho_n = \rho(n, n)$. We define the σ-fields based on the ranks of the first n observations to be $\mathscr{F}_n = \sigma\{\boldsymbol{\rho}_n\} = \sigma\{\rho(i, j) | 1 \le i \le j \le n\}$, where the equality of the two σ-fields follows by induction on n. Because we work with continuous distributions, we may take $\rho(1, n), \dots, \rho(n, n)$ to determine a permutation on the first n integers. Its inverse permutation has elements $\tau(1, n), \dots, \tau(n, n)$, so that $\tau(\rho(j, n), n) = j$. In a slight abuse of notation, we will refer to both nonrandom and random permutations by the same notation; the context will make the distinction clear.

Our procedure is based on nonparametric likelihood ratios for sequential ranks generated by distinguished prechange and postchange distributions F_0 and F_1. We study the operating characteristics of the procedure when the actual prechange and postchange distributions are G_0 and G_1. Note that the invariance of the ranks under strictly increasing transformations causes the ARL to false alarm to be identical for any continuous G_0.

Let $f_0(x)$ denote the density $\exp(-|x|)/2$. Let $f_1(x)$ denote the density $p\alpha \exp(-\alpha x) I_{\{x \ge 0\}} + q\beta \exp(\beta x) I_{\{x < 0\}}$, where $p = 1 - q$. The distributions F_0 and F_1 correspond to f_0 and f_1. We choose positive parameters $p \in [1/2, 1)$, $\alpha \le 1$ and $\beta \ge 1$. Under the measure $\mathbb{P}_\infty^F$, the observations are i.i.d. with density f_0. Under the measure $\mathbb{P}_k^F$, the prechange observations $X_1, \dots, X_{k-1}$ are i.i.d. with density f_0 independent of the postchange observations $X_k, \dots,$ which are i.i.d. with density f_1. (We exclude from consideration the special case $p = 1/2$ and $\alpha = \beta = 1$ in which prechange and postchange distributions are identical.)

We use $\mathbb{E}_\infty^G$, $\mathbb{E}_k^G$ and so on for expectations computed with respect to the correspondingly indexed measures. When it is convenient and unambiguous, we sometimes write $\mathbb{P}_k$ and $\mathbb{E}_k$ for $\mathbb{P}_k^G$ and $\mathbb{E}_k^G$. The validity of our procedures turns on the observation that $\mathbb{P}_\infty^F(A) = \mathbb{P}_\infty^G(A)$ for all $A \in \mathscr{F}_n$.

The choice of representative distribution f_1 as the postchange distribution is important. We argue in Section 3 that appropriate choices of tuning parameters α, β and p lead to appealing alternatives to parametric procedures.

Of fundamental importance is the following explicit computation of a likelihood ratio based on ranks. The theorem generalizes an identity of Savage (1956). This identity was used as the basis for a nonparametric sequential test of hypothesis by Savage and Sethuraman (1966).

THEOREM 1. *Let prechange and postchange densities be hypothesized to be f_0 and f_1 as defined above. Given the permutation $\rho(j, n)$ with inverse permutation $\tau(j, n)$, the corresponding likelihood ratio function is*

$$\Lambda_k^n(\boldsymbol{\rho}_n) = \frac{P_k^F\{X_{\tau(1,n)} < \cdots < X_{\tau(n,n)}\}}{P_\infty^F\{X_{\tau(1,n)} < \cdots < X_{\tau(n,n)}\}} = \sum_{m=0}^{n} \lambda_{k,m}^n(\boldsymbol{\rho}_n),$$

where, for $1 \le k \le n$, we define $U_k(m, n) = \sum_{j=k}^n I_{\{\rho(j,n) > m\}}$, let $V_k(m, n) = (n+1-k) - U_k(m, n)$ and write

$$\lambda_{k,m}^n(\boldsymbol{\rho}_n) = \binom{n}{m}\left(\frac{1}{2}\right)^n \left(\frac{p\alpha}{q\beta}\right)^{U_k(m,n)} (2q\beta)^{n+1-k}$$

$$(1) \qquad \times \prod_{i=1}^{m}\left(1 + \frac{V_k(i,n)}{i}(\beta - 1)\right)^{-1}$$

$$\times \prod_{i=m+1}^{n}\left(1 + \frac{U_k(i-1,n)}{n+1-i}(\alpha - 1)\right)^{-1}.$$

Our procedure is to compute the Shiryayev–Roberts statistic $R_n = \sum_{k=1}^n \Lambda_k^n$ and to stop at that time N_A when R_n first achieves or exceeds the critical level A. Because $R_n - n$ is a zero expectation $\mathbb{P}_\infty$-martingale, we can always bound the ARL to false alarm by $A \le \mathbb{E}_\infty\{N_A\}$. In fact, the ARL to false alarm grows asymptotically linearly in A as the critical level becomes large. The following theorem identifies the constant of proportionality.

THEOREM 2. *Let $p \in (1/2, 1)$, $\alpha \in (0,1)$ and $\beta \in [1, \infty)$ with $p\alpha \ge q\beta$ determine the non parametric stopping rule $N_A = \inf\{n \mid \sum_{j=1}^n \Lambda_k^n \ge A\}$, which is based on the likelihood ratio of ranks determined by f_0 and f_1 and is used to detect a change from G_0 to G_1. The ARL to false alarm has limit $\lim_{A\to\infty} \mathbb{E}_\infty^G\{N_A/A\} = \Delta(f_0, f_1)$, where $\Delta(f_0, f_1)$ has the same value as the limit computed in Gordon and Pollak* (1994a). *Specifically, if $2p\alpha \le 1$, then $\Delta(f_0, f_1) = 1/\alpha$.*

That $\mathbb{E}_\infty\{N_A\}/A$ is finite and has a limit expressible in renewal-theoretic terms is a consequence of the general theorem proved in Gordon and Pollak (1994b). We verify its hypotheses and so prove Theorem 2 in Section 4.4.

The expected lag to (true) detection is the subject of Theorem 3, proved in Section 4.5. We say the distribution G_1 is stochastically greater than G_0 and write $G_1 \succ_{st} G_0$ when $1 - G_1(t) \ge 1 - G_0(t)$ for all t.

THEOREM 3. *Assume $p \in (1/2, 1)$, $\alpha \in (0,1)$ and $\beta \in [1, \infty)$ with $p\alpha \geq q\beta$. Define*

$$D(G_0, G_1; f_0, f_1) = \mathbb{E}_1^G\left\{\log\left(\frac{f_1(F_0^{-1}(G_0(X_1)))}{f_0(F_0^{-1}(G_0(X_1)))}\right)\right\}.$$

Let B_ε denote the set $(\log(6\varepsilon), -4\varepsilon) \cup (4\varepsilon, -\log(6\varepsilon))$. Let $\nu(A)$ and $\varepsilon(A)$ be functions such that, as $A \to \infty$, the following hold:

(a) $$(\log A) P_1\{F_0^{-1}(G_0(X_1)) \notin B_{\varepsilon(A)}\} \to 0;$$

(b) $$(\log A)\varepsilon(A)\log(\varepsilon(A)) \to 0;$$

(c) $$\frac{\nu(A)\varepsilon^4(A)}{\log A} \to \infty.$$

If $\infty > D(G_0, G_1; f_0, f_1) > 0$ and $G_1 \succ_{st} G_0$, then

$$\limsup_{A\to\infty} \sup_{\nu \geq \nu(A)} \frac{\mathbb{E}_\nu^G\{N_A - \nu | N_A \geq \nu\}}{\log A} \leq \frac{1}{D(G_0, G_1; f_0, f_1)}.$$

We believe the new mixing technique used to prove Theorem 3 (see Lemmas 12 and 14) is of technical interest. A common difficulty in proving results like Theorem 3 stems from the need to analyze conditional expectations in which the conditioning event can have vanishingly small probability. Hence it is not enough to prove convergence in distribution; one must also have some form of uniform integrability. By suitably randomizing the change-point with a binomial prior and by making use of the hypothesized stochastic ordering, we are able to bound certain $\mathbb{E}_\nu^G$-expectations by $\mathbb{E}_\infty$-expectations, thereby controlling a number of vexing remainder terms. We believe the technique will be useful in other similar situations.

We have only proved an asymptotic upper bound on the average lag to detection. We conjecture that the upper bound is in fact a limit, but have only been able to verify the conjecture in special cases. We discuss the technical difficulties in proving equality at the end of Section 4.5.

Theorem 3 may be specialized to the case typically analyzed in the literature: the problem of detecting a shift in normal mean among observations with constant variance.

COROLLARY 4. *Assume $p \in (1/2, 1)$, $\alpha \in (0,1)$ and $\beta \in [1, \infty)$ with $p\alpha \geq q\beta$. Let the prechange distribution G_0 be standard normal, and let the postchange distribution G_1 be $\mathcal{N}(\mu, 1)$ with $\mu > 0$. For any $\xi > 0$,*

$$\limsup_{A\to\infty} \sup_{\nu \geq \log^{5+\xi} A} \frac{\mathbb{E}_\nu^G\{N_A - \nu | N_A \geq \nu\}}{\log A} \leq \frac{1}{D(G_0, G_1; f_0, f_1)},$$

whenever $\infty > D(G_0, G_1; f_0, f_1) > 0$.

Numerical integration yields $D(\mathcal{N}(0,1), \mathcal{N}(1,1), f_0, f_1) = 0.4854$ for the procedure with $\alpha = 0.53$, $\beta = 1.7$ and $p = 0.8413$. The corresponding quantity for the parametric cusum statistic (the postchange expected value of the log-likelihood ratio) is 0.5. The ratio $0.4854/0.5 = 97\%$ provides a measure of the relative speed with which the nonparametric and parametric procedures detect a change, and it is the basis of our claim of high asymptotic relative efficiency for our nonparametric Shiryayev–Roberts procedure.

Much stronger conclusions can be had with stronger hypotheses. Lorden (1971) shows that if G_0 and G_1 are completely known and if the corresponding cusum detection scheme is chosen, then the conditional expectation of the ARL to detection *given the entire past history before the change* is asymptotically best possible. Moustakides (1986) and Ritov (1990) strengthen this characterization of cusum procedures as minimax.

Under less stringent criteria (conditioning only on the event of no false alarm and not on the entire prechange history), Pollak (1985) shows that Shiryayev–Roberts procedures possess asymptotic minimax properties. Yakir (1994) shows that Shiryayev–Roberts rules are minimax in a plausible decision-theoretic setting. Our Theorem 3 is stronger in that we hypothesize different prechange and postchange distributions from those used to generate the procedure; it is similar to Pollak's (1985) result in that we condition only on the event of no false alarm and not on the entire history before the change-point.

Pollak and Siegmund (1991) consider the problem of Corollary 4 in which one wishes to detect a change from $\mathscr{N}(\mu, 1)$ prechange to $\mathscr{N}(\mu + \delta, 1)$ postchange when the initial mean μ is unknown. They analyze the behavior of a number of procedures, including the parametric version of the Shiryayev–Roberts procedure for unknown initial mean. They obtain rates of growth beyond the leading constant of proportionality.

There are two other approaches available to the problem we treat. McDonald (1990) uses the prechange uniform distribution of the sequential ranks to specify a cusum procedure for detection to a stochastically larger alternative; ARL to false alarm for this procedure is obtained by numerical methods. We compare expected lag to detection by Monte Carlo experimentation in Section 3.

In contrast to our procedures, whose large-sample analysis is based on renewal theory and large-deviation inequalities, alternative procedures can be based on the theories of weak convergence and contiguity. An example of such is Lombard (1983), where a stopping time (say, N^*) is suggested. Weak convergence there yields approximations to probabilities $P_\nu\{N^* > (1 + \delta)\nu\}$ for various values of $\delta > 0$, valid as $\nu \to \infty$ and as the postchange measure G_1 converges at a suitable rate to the prechange measure. For these procedures, results giving approximate expected lag to detection like those of Lorden (1971), Pollak and Siegmund (1991) and our Theorem 3 do not appear to be available in the literature.

3. Implementation and simulation. In this section we present results of several Monte Carlo experiments. All relate to the performance of our nonparametric Shiryayev–Roberts procedures in the case of a positive jump in mean for normal distributions with the same variance prechange and postchange. These suggest that the large-sample results given in the previous section yield insight useful in predicting the behavior of our procedure in situations of practical interest. We refer to the procedure as NPSRI, for invariant nonparametric Shiryayev–Roberts procedure. All simulations were done using the MATLAB programming language, version 3.5h, running on Sun Sparcstation computers [see MathWorks (1989)].

In Table 2, we present simulation results related to our Theorem 2. Specifically, we give a summary of 1000 realizations of ARL to false alarm under the $\mathbb{P}_\infty$-measure in which a change never occurs. A detailed description of the simulation procedures is deferred to the end of this section. Sample means for ARL to false alarm are presented for various critical levels A using NPSRI with parameters $p = 0.8413$, $\alpha = 0.53$ and $\beta = 1.7$. These values are derived in Gordon and Pollak (1994a) as the optimal choices for detecting a shift from $\mathscr{N}(0,1)$ to $\mathscr{N}(1,1)$ and are applicable here as well. Because $2p\alpha < 1$, Theorem 2 yields $\Delta = 1/\alpha = 1.89$.

The simulation results yield ratios for $\mathbb{E}_\infty\{N_A\}/A$ that are consistently lower than the limiting value 1.89; the observed values do appear, however, to be increasing as the critical level A increases. For $A < 500$, use of the limit Δ leads to setting a critical level 5–15% too low, yielding a procedure which tends to give false alarms somewhat more often than desired. More simulation work clearly needs to be done.

In Table 3, we compare the performance of NPSRI with that of McDonald's (1990) nonparametric cusum procedure, denoted NPCusum. The procedure is also based on sequential ranks and so is invariant to strictly increasing transformations of the data. We base our comparisons on his Table C. Both NPSRI and NPCusum are tuned for detecting a shift from $\mathscr{N}(\mu, \sigma^2)$ to $\mathscr{N}(\mu + \sigma, \sigma^2)$, a positive shift of one standard deviation from the unknown baseline μ.

NPCusum is tuned with parameters $k = 0.6428$ and $h = 1.203$, yielding a procedure with ARL to false alarm approximately 500. Because of the results presented in Table 2, we choose an NPSRI with $A = 300$, tuned with parameters $\alpha = 0.53$, $\beta = 1.7$ and $p = 0.8413$, to provide us with procedures having

TABLE 2

Average run length to false alarm for various critical levels A of the NPSRI procedure with parameters $\alpha = 0.53$, $\beta = 1.7$ *and* $p = 0.8413$ (*sample mean plus or minus standard error*, 1000 *simulations*)

A	**200**	**266**	**275**	**300**	**400**	**419.8**	**425**	**450**	**475**	**500**
$\mathbb{E}_\infty\{N_A\}$	328.9	449.1	462.5	512.4	687.9	726.2	733.8	791.9	856.5	896.0
± s.e.	6.2	9.1	9.2	10.4	13.6	14.2	14.4	15.3	19.0	19.6
$\mathbb{E}_\infty\{N_A\}/A$	1.64	1.69	1.68	1.71	1.72	1.73	1.73	1.76	1.80	1.79
± s.e.	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.04	0.04
Truncations at $N_A > 3500$	0	0	0	2	5	7	7	10	15	22

TABLE 3
Expected lag to detection for various stopping rules and change-points by simulation (sample mean plus or minus standard error); ARL to false alarm greater than or equal to 500

Actual distributions prechange and postchange	Procedure	$\mathbb{E}_\nu\{N-(\nu-1) \mid N\ge\nu\}$					$\mathbb{E}_\infty\{N\}$
		$\nu = 21$	51	101	201	501	
$G_0 = \mathscr{N}(\mu\,\sigma^2)$, $G_1 = \mathscr{N}(\mu + 0.5\sigma, \sigma^2)$	NPCusum	—	—	—	—	34.0	530
	±s.e.					0.6	5
	NPSRI_{300}	154.6	88.3	49.9	34.0	28.9	512
	±s.e.	9.7	7.4	3.1	1.7	1.4	10
	NPSRI_{500}	329.3[a]	165.7[b]	87.4	47.9	34.9	896
	±s.e.	21.0	14.0	6.9	2.6	1.3	20
$G_0 = \mathscr{N}(\mu, \sigma^2)$, $G_1 = \mathscr{N}(\mu + \sigma, \sigma^2)$	NPCusum	—	—	—	—	9.5	530
	±s.e.					0.1	5
	NPSRI_{300}	21.8	10.6	9.1	9.1	8.7	512
	±s.e.	1.8	0.3	0.2	0.2	0.3	10
	NPSIR_{500}	36.1	12.6	10.5	10.5	9.9	896
	±s.e.	3.7	0.3	0.2	0.2	0.2	20
$G_0 = \mathscr{N}(\mu, \sigma^2)$, $G_1 = \mathscr{N}(\mu + 2\sigma, \sigma^2)$	NPCusum	—	—	—	—	4.0	530
	±s.e.					0.0	5
	NPSRI_{300}	5.9	4.4	4.0	3.8	3.6	512
	±s.e.	0.1	0.1	0.1	0.1	0.1	10
	NPSRI_{500}	7.0	5.0	4.5	4.1	3.9	896
	±s.e.	0.1	0.1	0.1	0.1	0.1	20

[a]Average includes eight truncations at 3500, each imputed to contribute 4500 to average.
[b]Average includes four truncations at 3500, each imputed to contribute 4500 to average.

comparable ARL to false alarm and comparable optimization for sensitivity to prechange and postchange distributions. In light of the remarks following the statement of Theorem 1, we could also conservatively specify $A = 500$, guaranteeing ARL to false alarm above 500.

Performance under three different circumstances is presented in Table 3. The table's upper third illustrates performance when a shift of half a standard deviation occurs at times 21, 51, 101, 201 and 501. Because the baseline is unknown, there is no discrimination possible between $\mathbb{P}_\infty^G$ and $\mathbb{P}_1^G$. Included are McDonald's (1990) simulation results for ARL to false alarm and for average lag to detection, available only for $\nu = 501$.

Following the NPCusum results are averages for the NPSRI with the two critical levels $A = 300$ and $A = 500$. Each column summarizes the results of applying NPSRI to 1000 sequences of observations with specified change-point and postchange distribution. Included in each average are only those simulations which did not signal a false alarm prior to the change-point, thus yielding an estimate of the average lag to detection, conditional on a false alarm's not having been signaled. The same format is repeated in the middle third of the table, in which the postchange distribution is shifted right by one standard deviation, and the bottom third of the table, in which the postchange distribution is shifted right by two standard deviations. Each third of the table represents simulation results independent of the other two thirds. Simulations resulting in no signal after 3500 observations were truncated, and an average run length of an additional 1000 observations was imputed. (The correction ought to be conservative, if one believes approximate exponentiality of the stopping times.)

When a long history of prechange observations has been built up, NPCusum tends to have slightly higher expected lag to detection relative to NPSRI_{300} and slightly lower expected lag to detection relative to NPSRI_{500}. The advantage of NPSRI_{300} over NPSRI_{500} is apparent when a moderate-sized change occurs early in the sequence of observations; the expected lag is greatly reduced in the former case relative to the latter. The reduction in lag to detection is striking, for example, when $\nu = 21$ or $\nu = 51$, indicating the utility of Theorem 2 in approximating the appropriate critical level A when using NPSRI.

The middle third of the table represents the situation for which both NPCusum and NPSRI are tuned: normally distributed observations with a postchange shift upward in mean by a single standard deviation. The NPSRI does a little better than the NPCusum, perhaps because the family of shifts allows us three parameters with which to tune the likelihood ratio, as opposed to a single parameter for the NPCusum. Note that average lag to detection is fairly stable for as few as 50 observations prechange, but there is substantial deterioration in sensitivity for $\nu = 21$. These general conclusions remain the same for both the one-half and two standard deviation shifts, save that sensitivity deteriorates more rapidly with more recent change-points when the true shift has small magnitude.

In Table 4, we present comparisons of NPSRI versus several parametric procedures. The table is structured similarly to Table 3. As in Table 3, all procedures are tuned for detecting a change from standard normal to a postchange normal distribution with unit positive shift in mean. Comparison values for the parametric procedures are taken from Pollak and Siegmund (1990). As with Table 3, this table is divided into three sections representing first a small shift, then the shift for which the procedures were all tuned and, finally, a shift of larger magnitude than anticipated by tuning. All procedures have comparable specificity, yielding nominal ARL to false alarm approximately 800. For the NPSRI, we use Table 2 to choose $A = 450$, with tuning parameters as before.

The first line in each segment of the table corresponds to the cusum procedure tuned for $\mathscr{N}(0,1)$ changing to $\mathscr{N}(1,1)$, with all distributional parameters fully specified. The second and third entries in each segment are, respectively, to cusum and parametric Shiryayev–Roberts procedures for detecting a shift from $\mathscr{N}(\mu, 1)$ to $\mathscr{N}(\mu + 1, 1)$ with unknown baseline mean and known standard deviation. These procedures are studied intensively in Pollak and Siegmund (1991). They here are labelled respectively CusumI and S-RI for invariant cusum and Shiryayev–Roberts procedures. Finally, the NPSRI is tuned to detect shifts from $\mathscr{N}(\mu, \sigma^2)$ to $\mathscr{N}(\mu + \sigma, \sigma^2)$, for unknown

TABLE 4

Expected lag to detection for various stopping rules and change-points by simulation (sample mean plus or minus standard error); ARL to false alarm approximately 800

Actual distributions prechange and postchange	Procedure	$\mathbb{E}_\nu\{N-(\nu-1)\mid N\geq\nu\}$ $\nu=21$	51	101	201	$\mathbb{E}_\infty\{N\}$
$G_0=\mathcal{N}(0,1)$,	Cusum	15.6	15.2	15.9	16.0	781.0
$G_1=\mathcal{N}(0.75,1)$	±s.e.	0.2	0.2	0.2	0.2	15.1
	CusumI	142.5	34.6	19.7	17.3	780.8
	±s.e.	7.8	2.6	0.4	0.3	15.3
	S-RI	98.5	26.5	17.3	15.7	772.7
	±s.e.	5.8	1.4	0.3	0.2	8.8
	NPSRI_{450}	94.3	31.0	17.9	16.4	791.9
	±s.e.	8.6	3.7	0.6	0.5	15.3
$G_0=\mathcal{N}(0,1)$,	Cusum	9.3	9.3	9.2	9.3	781.0
$G_1=\mathcal{N}(1,1)$	±s.e.	0.1	0.1	0.1	0.1	15.1
	CusumI	38.0	12.1	10.5	9.8	780.8
	±s.e.	3.1	0.3	0.1	0.1	15.3
	S-RI	26.7	11.4	10.1	9.7	772.7
	±s.e.	2.6	0.4	0.1	0.1	8.8
	NPSRI_{450}	33.8	12.2	10.2	10.2	791.9
	±s.e.	3.4	0.3	0.2	0.2	15.3
$G_0=\mathcal{N}(0,1)$,	Cusum	5.1	5.2	5.2	5.1	781.0
$G_1=\mathcal{N}(1.5,1)$	±s.e.	0.0	0.1	0.1	0.1	15.1
	CusumI	7.2	5.7	5.5	5.2	780.8
	±s.e.	0.1	0.1	0.1	0.1	15.3
	S-RI	7.1	5.7	5.4	5.3	772.7
	±s.e.	0.1	0.1	0.1	0.1	8.8
	NPSRI_{450}	9.2	6.6	5.9	5.8	791.9
	±s.e.	0.2	0.1	0.1	0.1	15.3

baseline level of mean, and unknown variance. We have again chosen a conservative comparison for the NPSRI, because the parametric procedures against which we compare all require that the standard deviation is known.

In the upper third of the table we consider the case of a smaller than anticipated change in distribution from $\mathcal{N}(0,1)$ to $\mathcal{N}(0.75,1)$. Presented in the middle third of the table are average lags to detection for $\mathcal{N}(0,1)$ to $\mathcal{N}(1,1)$. Finally, we present averages corresponding to the same procedures when the shift is larger than anticipated [from $\mathcal{N}(0,1)$ to $\mathcal{N}(1.5,1)$].

Note that the fully parameterized cusum statistic makes full use of the known prechange baseline. Its average lag to detection is effectively constant over all tabulated values of change-point. The three invariant procedures, CusumI, S-RI and NPSRI, however, are all incapable of distinguishing between $\mathbb{P}_\infty^G$ and $\mathbb{P}_1^G$. This is reflected in the average run lengths to detection which decrease with increasing change-point.

For change-points occurring after 100 observations, one pays little for using the NPSRI instead of CusumI and S-RI. For shifts 0.75 and 1, average lags to detection are higher by roughly 5% or less of average lag to detection possessed by the invariant parametric schemes. For shift 1.5, added lag may be 10%. One pays 10–15% of average lag to detection relative to the fully specified cusum procedure, whose use requires knowledge of variance and prechange mean. One buys with this additional lag full protection against the cusum's high sensitivity of ARL to false alarm, exemplified in Table 1.

For small values 21 and 51 of the change-point and smaller to moderate values of shift, NPSRI is again comparable to (although a little worse than) the better of the two invariant parametric procedures. Relative performance is worst for the large-shift case, although speed of detection is still quite rapid, giving up two or fewer postchange observations at early change-points. We attribute the poorer performance of NPSRI relative to S-RI to the need to estimate the prechange distribution function from the empirical distribution and by the bounded influence a single observation has on Λ_k^n.

A related sequential procedure is studied in Gordon and Pollak (1994a), in which the prechange distribution is known to be symmetric with known center of symmetry. Comparison with simulations done there show comparable average run lengths to detection with large numbers of prechange observations, but show that the known center of symmetry allows for much more rapid detection in the case of early change.

Our procedure is easily implemented in the Matlab programming language. Figure 1 presents the subroutine which is used to compute the NPSRI statistics used in the simulations we have just discussed. The choice of variable names closely follows the computation of the nonparametric likelihood ratio presented in Theorem 1. Unfortunately the computation requires $O(n)$ operations to compute each λ_k^n. Hence Λ_k^n requires $O(n^2)$ operations, so that a simulation stopping at N_A requires $O(N_A^4)$ time, which can be substantial for large N_A.

A remedy is available in Lemma 9, because the sum of likelihood ratios Λ_k^n for k of intermediate magnitude contribute vanishingly little to R_n. In performing the simulations reported in Table 2, we used this reduced form for most of the simulation and then estimated the effect of this reduction in a smaller simulation. The sample correction for using the reduced form of the statistic was of negligible magnitude, suggesting that such a scheme might be useful in implementation.

Of greater effect on the results of Table 2 was the truncation of the simulations at $n = 3500$. If a simulation did not stop after 3500 observations, it was truncated and both N_A and R_{N_A} were imputed to have the value $3500 + 1.5A$. In addition, because $R_n - n$ is a $\mathbb{P}_\infty$-martingale, it was possible to estimate $\mathbb{E}_\infty\{N_A\}$ by the sample mean of realizations of $(N_A + R_{N_A})/2$, effecting a reduction in variance. The adjustments described above had relatively small effect. For example, the unadjusted sample means corresponding to the entries $A = 475$ and $A = 500$ are 1.78 and 1.76 with respective standard errors 0.04 and 0.04. The vast majority of adjustment was attributable to adjustment for truncation.

```
% Inputs:   data     row of n entries
%           lnbnml   row of logs of binomial(n,1/2) density, n+1 entries
%           selectk  row vector containing indices of k values for
%                    which lambdakn is to be computed.
%           procparm row of procedure parameters: [alpha, beta, p]
% Outputs:  lambdakn row of n entries, one for each non-parametric
%                    likelihood ratio computed; some will be zero
%                    unless selectk==(1:1:n)
function lambdakn = umkrnlb(data,lnbnml,selectk,procparm);
%---- get parameters ----
n     = length(data);
alpha = procparm(1); beta = procparm(2); p = procparm(3); q = 1-p;
ln2palpha = log(2*p*alpha); ln2qbeta  = log(2*q*beta);
lnratio   = ln2palpha - ln2qbeta;
%----  initialize and allocate vectors -----
incr              =  1:1:n;          % 1 to n row vector
decr              = [(n-1:-1:1) 1];  % n-1 to 1 then 1
reverse           =  (n+1:-1:1);     % row vector to reverse order
lambdakn          =  zeros(1,n);     % to hold little lambda sub kn's
[dummy,invrankt]  = sort(data');     % index of smallest in invrankt(1)
invrank           = invrankt' ;      % row of inverse ranks
%----  compute vector of SR statistics -----
for k = selectk, %--- compute little lambdakn for selected k values ---
   timegek  =(invrank>=k)    ; % time not less than k
   vsubk    =cumsum(timegek) ; % for i=1 to n , watch out for i=0
   usubk    =(n+1-k) - vsubk ;
   lnvdenom =log([ 1                       (1+(vsubk./incr)*( beta-1))]);
   lnudenom =log([(1+(n+1-k)*(alpha-1)/n)  (1+(usubk./decr)*(alpha-1))]);
      % lnvdenom and lnudenom are n+1 vectors containing log-denominators
      % index of vectors is (one plus) number of putative negatives m
   lnprodneg = cumsum(lnvdenom);
   dummy1    = lnudenom(reverse);
   dummy2    = cumsum(dummy1);
   lnprodpos = dummy2(reverse);
   usubkmod = [(n+1-k) usubk];
   lambdakn(k) ...
       = sum(exp(  lnbnml + (n+1-k)*ln2qbeta + usubkmod*lnratio  ...
                             - lnprodneg - lnprodpos ));
end;
```

FIG. 1. *MATLAB program for computing the statistic R_n.*

4. Proofs. Throughout we follow the convention that summation over a null set of indices is 0. Consistent with the former convention, the product over a null set of indices is by convention 1. As usual, we take $\binom{n}{m} = 0$ when $n < m$.

Because the proofs of Theorems 1 and 2 are similar to those of Gordon and Pollak (1994a), we omit a number of details. In particular, we sketch the proof of Theorem 2 by stating a number of intermediate results as lemmas, without proof.

4.1. *The nonparametric likelihood ratio function: proof of Theorem* 1. Throughout Sections 4.1 and 4.2 we treat $\rho(\cdot, n)$ as a given nonrandom permutation on the first n positive integers, induced by the n scalars $x_1, \dots, x_n$. Recall that $\tau(\cdot, n)$ is the permutation inverse to $\rho(\cdot, n)$. Set $\tau(j, n) = 0$ if $j \le 0$ or if $j > n$. Define the following for all $n \ge k \ge 1$, for $j \ge 1$ and for all m: $\sigma(j, m, n) = I_{\{\rho(j, n) > m\}}$ and $U_k(m, n) = \sum_{j=k}^{n} \sigma(j, m, n)$. By convention, $U_k(m, n) = 0$ if $k > n$. Let $V_k(m, n) = \max\{0, (n + 1 - k) - U_k(m, n)\}$. When we wish to emphasize the dependence of the various quantities upon the scalars, we write $\sigma(j, m, n | x_1, \dots, x_n)$, $U_k(m, n | x_1, \dots, x_n)$ and similarly for other quantities.

Informally, think of $\sigma(j, m, n)$ as the indicator that X_j is positive, given that we knew that there were exactly m negative observations among the first n observations and that X_j had rank $\rho(j, n)$. For example, $\sigma(j, 0, n) = 1$ and $\sigma(j, n, n) = 0$, for all $1 \le j \le n$. Similarly $U_k(m, n)$ and $V_k(m, n)$, respectively, count the number of hypothetically "positive" and "negative" observations among the observations taken at times on or after the hypothetical change-point k, through time n.

PROOF OF THEOREM 1. We need to compute

$$\Lambda_k^n(\boldsymbol{\rho}_n) = \mathbb{P}_k^F\{X_{\tau(1, n)} < X_{\tau(2, n)} < \cdots < X_{\tau(n, n)}\}/(1/n!).$$

Let D_m denote the event $\{\max_{1 \le j \le m} X_{\tau(j, n)} < 0 < \min_{m < j \le n} X_{\tau(j, n)}\}$ for $1 \le m < n$. Let D_0 denote the event $\{0 < \min_{1 \le j \le n} X_{\tau(j, n)}\}$, and let D_n denote $\{\max_{1 \le j \le n} X_{\tau(j, n)} < 0\}$. In words, D_m is the event that there are exactly m negative observations among $\{X_1, \dots, X_n\}$; necessarily, the negative observations must have the m smallest ranks. Write

$$\lambda_{k, m}^n = (n!)\mathbb{P}\{X_{\tau(1, n)} < \cdots < X_{\tau(n, n)} | D_m\}\mathbb{P}_k^F\{D_m\}.$$

Because the events D_m partition the event of interest, we have $\Lambda_k^n = \sum_{m=0}^{n} \lambda_{k, m}^n$. Define functions

$$\gamma_k(j, m, n) = \begin{cases} 1, & \text{if } j < k, \\ \alpha, & \text{if } j \ge k \text{ and } \rho(j, n) > m, \\ \beta, & \text{if } j \ge k \text{ and } \rho(j, n) \le m. \end{cases}$$

We first show that, for $1 \le k \le n$,

$$\begin{aligned} \lambda_{k, m}^n(\boldsymbol{\rho}_n) = \binom{n}{m}\left(\frac{1}{2}\right)^{k-1} & p^{U_k(m, n)} q^{V_k(m, n)} \\ & \times \prod_{i=1}^{m} \frac{\gamma_k(\tau(i, n), m, n)}{(1/i)\sum_{j=1}^{i} \gamma_k(\tau(j, n), m, n)} \\ & \times \prod_{i=m+1}^{n} \frac{\gamma_k(\tau(i, n), m, n)}{[1/(n + 1 - i)]\sum_{j=i}^{n} \gamma_k(\tau(j, n), m, n)}. \end{aligned} \tag{2}$$

Because $\rho(\cdot, n)$ and $\tau(\cdot, n)$ are inverse permutations,

$$\{j | n \geq j > m \text{ and } \tau(j, n) \geq k\} = \{\rho(i, n) | \rho(i, n) > m \text{ and } n \geq i \geq k\}.$$

Hence

$$\mathbb{P}_k^F\{D_m\} = \left(\tfrac{1}{2}\right)^{k-1} p^{U_k(m, n)} q^{(n+1-k) - U_k(m, n)}. \tag{3}$$

Conditional on D_m, the random variables $X_{\tau(1, n)}, \dots, X_{\tau(m, n)}$ are distributed as -1 times exponential variates, respectively, with means $1/\gamma_k(\tau(j, n), m, n)$ for $1 \leq j \leq m$. By Savage's (1956) identity,

$$\mathbb{P}_k^F\{X_{\tau(1, n)} < \cdots < X_{\tau(m, n)} | D_m\} = \frac{1}{m!} \prod_{i=1}^{m} \frac{\gamma_k(\tau(i, n), m, n)}{\sum_{j=1}^{i} \gamma_k(\tau(j, n), m, n)/i}. \tag{4}$$

Conditional independently of the ordering of the m negative observations, the ordering of the $n - m$ positive observations has conditional probability

$$\begin{aligned} &\mathbb{P}_k^F\{X_{\tau(m+1, n)} < \cdots < X_{\tau(n, n)} | D_m\} \\ &\quad = \frac{1}{(n-m)!} \prod_{i=m+1}^{n} \frac{\gamma_k(\tau(i, n), m, n)}{\sum_{j=i}^{n} \gamma_k(\tau(j, n), m, n)/(n+1-i)}. \end{aligned} \tag{5}$$

Combining (3), (4) and (5) yields (2).

To prove (1), first consider the numerators in the product terms of (2). In the first product, as the index i runs from 1 through m, there are $V_k(m, n)$ numerator terms β; the other numerators are 1. Similarly, in the second product with $m + 1 \leq i \leq n$, there are $U_k(m, n)$ numerator terms α; the other numerators are again 1. Consolidate these numerator terms with the terms $p^{U_k(m, n)}$ and $q^{V_k(m, n)}$ in (3).

Now consider the denominator terms in (4) and (5). A denominator in (4) is an average of i terms among which $V_k(i, n)$ are β; the rest are 1. Similarly, a denominator in (5) is an average of $n + 1 - i$ terms among which $U_k(i - 1, n)$ are α; the rest are 1. Hence (1) follows from (2). □

4.2. *The likelihood ratio function: identities and bounds.* We shall need a number of relations among the $U_k(m, n)$. These are collected in the following lemma. We have already used (6) in proving Theorem 1.

LEMMA 1. *Assume $1 \leq k \leq n + 1$. If $m \leq 0$, then $U_k(m, n) = n - k + 1$. If $m \geq n$, then $U_k(m, n) = 0$. In addition, the following hold:*

$$U_k(m, n) = \sum_{j=m+1}^{n} I_{\{\tau(j, n) \geq k\}}; \tag{6}$$

$$U_k(m, n) - U_k(m+1, n) = I_{\{\tau(m+1, n) \geq k\}}; \tag{7}$$

$$U_k(m, n+1) - U_k(m, n) = I_{\{\tau(m, n) \geq k\}} + I_{\{\tau(m, n) < k\}} I_{\{\rho_{n+1} > m\}}; \tag{8}$$

$$U_k(m+1, n+1) - U_k(m, n) = I_{\{\tau(m+1, n) < k\}} I_{\{\rho_{n+1} > m+1\}}; \tag{9}$$

$$V_k(m, n+1) - V_k(m, n) = I_{\{\rho_{n+1} \leq m\}} I_{\{\tau(m, n) < k\}}. \tag{10}$$

We finish the section with a number of properties of the nonparametric likelihood ratio (1). The values $\lambda_{k,m}^n$ were defined in (1). Consistent with our conventions and with (1), we define $\lambda_{k,m}^n(\boldsymbol{\rho}_n) = \binom{n}{m}2^{-n}$ for $n < k \le \infty$. We use (1) and (7) to prove Lemma 2. It is then used to prove Lemma 3.

Lemma 2 tells us that very little contribution is made to Λ_k^n by those $\lambda_{k,m}^n$ having m sufficiently far from $n/2$. Specifically, for these m, the contributions decay uniformly geometrically rapidly in m.

LEMMA 2. *Let $p\alpha \ge q\beta$. If $n > m \ge 2n/3$ and $n + 1 \ge k \ge 1$, then $\lambda_{k,m}^n \ge 2\lambda_{k,m+1}^n$. If $0 < m \le qn/3$, then $\lambda_{k,m}^n \ge 2\lambda_{k,m-1}^n$.*

If the prechange density were known to be symmetric and if the putative change-point k were large but smaller than the true disruption time, then we would expect that the center of symmetry of the prechange distribution would be close to the median of the first k observations. If k is also close to n, then the median of the first k observation should be little different from the median of all n observations. The next lemma makes this heuristic precise in terms of the likelihood function.

LEMMA 3. *Assume $p\alpha \ge q\beta$ and $n > 6/q$. Let $\theta = (n - k + 1)/n$. There exist positive constants $b_i^{(3)} = b_i^{(3)}(\alpha, \beta, p)$, for $i = 1, 2$, which depend only on the tuning parameters such that, for all $\theta \in (0, 1/4)$ and $\delta \in (0, \sqrt{q/8})$, we have $\sum_{0 \le m \le (1/2-\delta)n} \lambda_{k,m}^n(\boldsymbol{\rho}_n) < b_1^{(3)} \exp[n(b_2^{(3)}\theta - \delta^2)]\Lambda_k^n(\boldsymbol{\rho}_n)$. For all $0 \le k \le n$ and all $\delta > 0$, we have $\sum_{(1/2+\delta)n \le m \le n} \lambda_{k,m}^n(\boldsymbol{\rho}_n) \le b_1^{(3)} \exp(-n\delta^2)\Lambda_k^n(\boldsymbol{\rho}_n)$. The bounds are uniform over all realizations of $\boldsymbol{\rho}_n$.*

We write $\lambda_{k,m}^{n+1}(\boldsymbol{\rho}_n, \rho_{n+1})$ for $\lambda_{k,m}^{n+1}(\boldsymbol{\rho}_{n+1})$. The proof of the next lemma involves the identities established in Lemma 1.

LEMMA 4. *If $1 \le k \le n + 1$ and $0 \le m \le n$, then*

$$
\begin{aligned}
&\frac{\lambda_{k,m}^{n+1}(\boldsymbol{\rho}_n, \rho_{n+1})}{\lambda_{k,m}^n(\boldsymbol{\rho}_n)} \\
&\quad = \frac{1}{2}\frac{n+1}{n+1-m}\left(\frac{p\alpha}{q\beta}\right)^{I_{\{\tau(m,n)\ge k\}} + I_{\{\tau(m,n)<k\}}I_{\{\rho_{n+1}>m\}}}(2q\beta) \\
&\quad \times \prod_{i=\rho_{n+1}}^{m}\left(1 + \frac{I_{\{\tau(i,n)<k\}}}{1 + [V_k(i,n)/i](\beta-1)} \cdot \frac{\beta-1}{i}\right)^{-1} \\
&\quad \times \prod_{i=m+1}^{\rho_{n+1}-1}\left(1 + \frac{I_{\{\tau(i,n)<k\}}}{1 + [U_k(i,n)/(n+1-i)](\alpha-1)} \cdot \frac{\alpha-1}{n+1-i}\right)^{-1} \\
&\quad \times \left(1 + \frac{U_k(m, n+1)}{n+1-m}(\alpha-1)\right)^{-1}.
\end{aligned}
\tag{11}
$$

In addition,

$$\frac{\lambda^{n+1}_{k,m+1}(\boldsymbol{\rho}_n, \rho_{n+1})}{\lambda^n_{k,m}(\boldsymbol{\rho}_n)}$$

$$= \frac{1}{2}\frac{n+1}{m+1}\left(\frac{p\alpha}{q\beta}\right)^{I_{\{\tau(m+1,n)<k\}}I_{\{\rho_{n+1}>m+1\}}}(2q\beta)$$

$$\times \prod_{i=\rho_{n+1}}^{m}\left(1+\frac{I_{\{\tau(i,n)<k\}}}{1+[V_k(i,n)/i](\beta-1)}\cdot\frac{\beta-1}{i}\right)^{-1} \tag{12}$$

$$\times\left(1+\frac{V_k(m+1,n+1)}{m+1}(\beta-1)\right)^{-1}$$

$$\times \prod_{i=m+1}^{\rho_{n+1}-1}\left(1+\frac{I_{\{\tau(i,n)<k\}}}{1+[U_k(i,n)/(n+1-i)](\alpha-1)}\cdot\frac{\alpha-1}{n+1-i}\right)^{-1}.$$

4.3. *The likelihood ratio function: monotonicity.* A crucial observation is that the nonparametric likelihood ratio is a nondecreasing function of the most recent observations. The simplest manifestation of this monotonicity is Lemma 5, which is used to prove Theorem 2. The exact contribution of stochastic ordering is made precise in Lemma 6, which generalizes the former lemma. Lemma 5 follows from (8); Lemma 6 is proved from first principles by a counting argument.

LEMMA 5. *If $p\alpha \geq q\beta$, then the likelihood ratio $\Lambda^n_k(\boldsymbol{\rho}_{n-1}, \rho_n)$ is a monotone nondecreasing function of ρ_n when the first $n-1$ sequential ranks $\boldsymbol{\rho}_{n-1}$ are held fixed.*

Assume that the true prechange and postchange distributions are stochastically ordered. Write $\psi(x) = G_0^{-1}(G_1(x))$, so that $X \sim G_1$ implies $\psi(X) \sim G_0$. Observe that $G_1 \succ_{st} G_0$ implies ψ is nondecreasing and $x \geq \psi(x)$.

LEMMA 6. *Let $\psi(\cdot)$ be a strictly increasing function such that $\psi(x) \leq x$ for all x. Assume $\{x_1, \dots, x_n\}$ and $\{x_1, \dots, x_{j-1}, \psi(x_j), \dots, \psi(x_n)\}$ are both sets of n distinct scalars. Then*

$$U_k(m, n | x_1, \dots, x_n) \geq U_k\big(m, n | x_1, \dots, x_{j-1}, \psi(x_j), \dots, \psi(x_n)\big),$$

for all $1 \leq j \leq n$, all $1 \leq k \leq n+1$, and all $0 \leq m \leq n$.

Lemma 7 allows us to bound the contributions of integrals over sets of small $\mathbb{P}^G_k$ probablity by using stochastic ordering to allow integration with respect to $\mathbb{P}^G_\infty$. Its proof depends on Theorem 1, Lemma 5 and Lemma 6.

LEMMA 7. *Assume* $G_1 \succ_{st} G_0$ *and let* $\psi(x) = G_0^{-1}(G_1(x))$. *If* $p\alpha \ge q\beta$ *and* $\alpha \le 1 \le \beta$, *then, for all* $k \le n + 1$ *and all* $j \ge 1$,

$$\mathbb{P}_j^G\Big\{\Lambda_k^n(X_1,\dots,X_n) \ge \Lambda_k^n\big(X_1,\dots,X_{j-1},\psi(X_j),\dots,\psi(X_n)\big)\Big\} = 1$$

and $\mathbb{P}_j^G\{N_A \ge n\} \le \mathbb{P}_n^G\{N_A \ge n\} = \mathbb{P}_\infty\{N_A \ge n\}$.

Note that we do not claim $\mathbb{P}_j\{N_A \ge n\}$ is monotone in $j \le n$. Indeed such an assertion cannot be true because $\mathbb{P}_\infty = \mathbb{P}_1$.

4.4. *ARL to false alarm*: *Proof of Theorem* 2.

LEMMA 8. *Let* $\theta = (n - k + 1)/n$. *For all* $\delta > 0$, *all* $1 \le k \le n$ *and all* $0 \le i \le n$,

$$\mathbb{E}_\infty\{U_k(i,n)\} = (n-i)\theta,$$
$$\mathbb{P}_\infty\{U_k(i,n) - (n-i)\theta > \delta(n-i)\} < \exp\big[-2(n-i)\delta^2\big],$$
$$\mathbb{E}_\infty\{V_k(i,n)\} = i\theta,$$
$$\mathbb{P}_\infty\{V_k(i,n) - i\theta < -\delta i\} < \exp(-2i\delta^2).$$

PROOF. Under $\mathbb{P}_\infty$, the random variables $X_1,\dots$ are exchangeable, and so $U_k(i,n)$ and $V_k(i,n)$ have hypergeometric distributions. From Kemperman (1973) the binomial distribution is a dilation of the hypergeometric. Therefore, the usual exponential bounds for the tails of the binomial distribution apply also to the hypergeometric. □

The next lemma tells us that likelihood ratios for hypothetical changepoints a small distrance from the most recent observation time make negligible contribution to the NPSRI statistic. The proof uses Lemma 2 and Lemma 8.

LEMMA 9. *Assume* $p\alpha \ge q\beta$ *and* $n \ge 3(q+3)/q^2$. *Let* $t < n/3$ *be an arbitrary positive integer, and let* η *be positive. Write* $\theta = (n + 1 - k)/n$. *If* $3\eta/q < \theta < 1 - \eta$, *then with* $\mathbb{P}_\infty$ *probability at least*

$$1 - 2n\frac{\exp(-t\eta^2/8)}{1 - \exp(-\eta^2/8)}, \tag{13}$$

we have the inequality

$$\Lambda_k^n < 2\left(\frac{\beta}{\alpha}\right)^t\left(\frac{p}{q}\right)^{\eta n}\exp\left(-2n\theta(1-\theta)\left(p - \frac{1}{2}\right)^2\right), \tag{14}$$

holding simultaneously for all k in the specified interval.

Given scalar x and nonnegative integers h and l, we define the modified rank function as

$$r_n(x,h,l) = \left[1 + h + \sum_{j=1}^{n} I_{\{X_j \le x\}}\right] \wedge (n + l + 1). \tag{15}$$

Note that $\rho_{n+1} = r_n(X_{n+1}, 0, 0)$, generalizing the notion of sequential rank. Given the n-vector of integers $\boldsymbol{\rho}_n$, the l-vector of integers $\mathbf{s}_l$ and the integer t, we write $(\boldsymbol{\rho}_n, \mathbf{s}_l, t)$ for the concatenated $(n+l+1)$-vector of the elements of $\boldsymbol{\rho}_n$, $\mathbf{s}_l$ and t, with similar definition of the $(n+l)$-vector $(\boldsymbol{\rho}_n, \mathbf{s}_l)$. In a slight abuse of notation, we write $\lambda_{k,m}^{n+l+1}(\boldsymbol{\rho}_n, \mathbf{s}_l, t)$ for $\lambda_{k,m}^{n+l+1}((\boldsymbol{\rho}_n, \mathbf{s}_l, t))$, with similar conventions for $\lambda_{k,m}^{n+l}(\boldsymbol{\rho}_n, \mathbf{s}_l)$, $\Lambda_k^{n+l+1}(\boldsymbol{\rho}_n, \mathbf{s}_l, t)$ and $\Lambda_k^{n+l+1}(\boldsymbol{\rho}_n, \mathbf{s}_l)$.

Lemma 10 tells us that the change in nonparametric likelihood ratio is a good approximation to the change in parametric likelihood. The approximation is robust, in that it is insensitive to the values of the last few sequential ranks. The proof uses Lemmas 3 and 4.

LEMMA 10. *Given $\varepsilon \in (0, \alpha \wedge \frac{1}{6})$, let $B_\varepsilon = (\log(6\varepsilon), -4\varepsilon) \cup (4\varepsilon, -\log(6\varepsilon))$. Let S_l be the set of l-vectors of integers $\mathbf{s}_l = (s_{l1}, \dots, s_{ll})$ such that $s_{lj} \in [1, n+j]$ for $j = 1, \dots, l$, write*

$$L_k^n(\mathbf{s}_l, x, h) = \log\left(\frac{\Lambda_k^{n+l+1}(\boldsymbol{\rho}_n, \mathbf{s}_l, r_n(x, h, l))/\Lambda_k^{n+l}(\boldsymbol{\rho}_n, \mathbf{s}_l)}{f_1(x)/f_0(x)} \right).$$

There exist positive constants $b_1^{(10)}$, $b_2^{(10)}$ and $b_3^{(10)}$ such that

$$\mathbb{P}_\infty^F\left\{ \sup_{x \in B_\varepsilon} \sup_{h < n\varepsilon^2} \sup_{\mathbf{s}_l \in S_l} \sup_{0 \le n+l+1-k < n\varepsilon^2} |L_k^n(\mathbf{s}_l, x, h)| > b_1^{(10)}\left(-\varepsilon \log(\varepsilon) + \frac{1}{n\varepsilon} \right) \Big| \mathscr{F}_n \right\} < b_2^{(10)} \exp\left(-b_3^{(10)} \varepsilon^4 n \right),$$

whenever $l \vee 3\beta \vee 6/q < n\varepsilon^2$ and $p\alpha \ge q\beta$.

The proof of the next lemma requires Lemmas 1, 2, 4 and 5.

LEMMA 11. *Let M be a stopping time adapted to the σ-fields $\mathscr{F}_n$. Let $Q_n = \sum_{k=M+1}^n \Lambda_k^n$, and let N_A^M be the first index at which $Q_n \ge A$.*

If $p\alpha \ge q\beta$, there exists a constant $\kappa > 1$ such that

$$\mathbb{E}_\infty\left\{ Q_{n+1}^{1+\alpha/2} \big| \boldsymbol{\rho}_n \text{ and } N_A^M = n+1 \right\} \le \kappa A^{1+\alpha/2},$$

for all $A > 1$.

We are now ready to prove Theorem 2. To do so, we verify the hypotheses of the following theorem proved in Gordon and Pollak (1994b).

THEOREM A. *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 $n \ge 1$,*

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

(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*

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

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

Condition (a) is obtained from Lemma 9 by setting $t = \delta n$ for some small constant $\delta > 0$, and choosing η small relative to the desired range of values θ. Condition (b) follows from Lemma 10 by setting $l = h = 0$. Condition (c) in strengthened form is exactly the statement of Lemma 11. The form required by Theorem A is obtained from Hölder's inequality. See Gordon and Pollak (1994b) for details. The form of the constant is that computed in Gordon and Pollak (1994a), where the case $2p\alpha \wedge 2\theta\beta \leq 1$ is treated.

4.5. *Conditional bounds for postchange ARL: Proof of Theorem* 3. In this section, we study the postchange behavior of our nonparametric Shiryayev–Roberts statistic. The chief tools are Lemma 10, used previously to prove Theorem 2, and a new mixing technique, which will enable us to use Lemma 11 in bounding the postchange average run length.

Throughout this section we assume $G_1 \succ_{st} G_0$. Given $\xi \in (0,1)$, we write $G_\xi = (1-\xi)G_0 + \xi G_1$. Note that $G_1 \succ_{st} G_\xi \succ_{st} G_0$.

Let ν and $\nu^+ > \nu$ be given. Define the binomial change-point model BCP(ν^+, ν/ν^+) as follows. Choose J distributed as binomial(ν^+, ν/ν^+). Conditional on J, let $X_1, \dots, X_J$ be distributed i.i.d. with prechange distribution G_0 independent of $X_{J+1}, \dots$ which are distributed i.i.d. with postchange measure G_1. Denote the associated probability measure and expectation by $\mathbb{P}_{\mathrm{BCP}}$ and $\mathbb{E}_{\mathrm{BCP}}$, respectively. The crucial observation is that, under BCP(ν^+, ν/ν^+), the empirical distribution of $X_1, \dots, X_{\nu^+}$ is that of ν^+ observations i.i.d. as G_{ν/ν^+}.

LEMMA 12. *Let $\nu^+ > \nu > 0$ be given, and let $G_1 \succ_{st} G_0$. Then*

$$\mathbb{E}_\nu^G\{N_A - \nu^+; N_A > \nu^+\} < 3\kappa(1)A\sqrt{\nu^+ - \nu}\,\mathbb{P}_{BCP}\{N_A > \nu^+\}.$$

PROOF. Write $\xi = \nu/\nu^+$, and assume $\xi > \frac{1}{2}$. From Stirling's formula,

$$\mathbb{P}\{\text{bionomial}(n, k/n) = k\} \geq \tfrac{1}{3}k^{-1/2},$$

for integers $0 < k < n$. Use the bound to show

$$\begin{aligned}
&\mathbb{E}_\nu^G\{N_A - \nu^+; N_A > \nu^+\}\\
&\quad< 3\sqrt{\nu^+ - \nu}\sum_{k=0}^{\nu^+}\binom{\nu^+}{k}\xi^k(1-\xi)^{\nu^+-k}\mathbb{E}_k\{N_A - \nu^+; N_A > \nu^+\}\\
&\quad= 3\sqrt{\nu^+ - \nu}\,\mathbb{E}_{\text{BCP}}\{\mathbb{E}_{\text{BCP}}\{N_A - \nu^+ | \mathscr{F}_{\nu^+}\}; N_A > \nu^+\}\\
&\quad\leq 3\sqrt{\nu^+ - \nu}\,\mathbb{E}_{\text{BCP}}\{\mathbb{E}_{\text{BCP}}\{T_A | \mathscr{F}_{\nu^+}\}; N_A > \nu^+\},
\end{aligned}$$

where $T_A = \inf\{n \geq 1 | \sum_{k=1+\nu^+}^{n+\nu^+}\Lambda_k^{n+\nu^+} \geq A\}$. Note that T_A depends on the first ν^+ observations only through their empirical distribution. Let $\psi_\xi(x) = G_\xi^{-1}(G_1(x))$. By hypothesis, $\psi_\xi(\cdot)$ is nondecreasing and $\psi_\xi(x) \leq x$. Let $(\pi(1), \ldots, \pi(\nu^+))$ be a uniformly distributed random permutation of the first ν^+ indices chosen independent of all other random variables. Let

$$\hat{X}_j = \begin{cases} X_{\pi(j)}, & \text{if } j \leq \nu^+,\\ \psi_\xi(X_j), & \text{if } j > \nu^+.\end{cases}$$

Observe that under $\mathbb{P}_{\text{BCP}}$ the sequence of $\hat{X}_j$ is i.i.d. as G_ξ and that the empirical distribution of $\{X_1, \ldots, X_{\nu^+}\}$ is identical to that of $\{\hat{X}_1, \ldots, \hat{X}_{\nu^+}\}$. Let $\hat{\boldsymbol{\rho}}_n$ and $\hat{T}_A$ be sequential ranks and stopping times which are analogous to $\boldsymbol{\rho}_n$ and T_A, computed using $\hat{X}_1, \ldots$ instead of $X_1, \ldots$. From Lemma 7, $\Lambda_k^n(\boldsymbol{\rho}_n) \geq \Lambda_k^n(\hat{\boldsymbol{\rho}}_n)$ pointwise for all $\nu^+ < k \leq n$. Hence we obtain the coupling $\hat{T}_A \geq T_A$, from which follows

$$\begin{aligned}
\mathbb{E}_\nu^G\{N_A - \nu^+; N_A > \nu^+\} &\leq 3\sqrt{\nu^+ - \nu}\,\mathbb{E}_{\text{BCP}}\left\{\mathbb{E}_{\text{BCP}}\{\hat{T}_A | \mathscr{F}_{\nu^+}\}; N_A > \nu^+\right\},\\
&\leq 3\sqrt{\nu^+ - \nu}\,\mathbb{E}_{\text{BCP}}\left\{\mathbb{E}_\infty\{\hat{T}_A | \mathscr{F}_{\nu^+}\}; N_A > \nu^+\right\},\\
&\leq 3\kappa(1)A\sqrt{\nu^+ - \nu}\,\mathbb{P}_{\text{BCP}}\{N_A > \nu^+\},
\end{aligned}$$

proving the lemma. □

LEMMA 13. *Let $p \in (0, \frac{1}{2})$ and $|\varepsilon| < 1$. Then*

$$H((1+\varepsilon)p, p) > \tfrac{1}{4}\varepsilon^2 p,$$

where $H(a, p) = a\log(a/p) + (1-a)\log((1-a)/(1-p))$ is the relative entropy of the Bernoulli(a) *distribution relative to the* Bernoulli(p) *distribution.*

PROOF. Bound the second partial derivative of $H((1+\varepsilon)p, p)$ with respect to ε and apply a Taylor expansion with remainder. □

The function D was defined in the statement of Theorem 3.

LEMMA 14. *Let $G_1 \succ_{st} G_0$ be, respectively, the postchange and prechange distributions. Let B_ε denote the set* $(\log(6\varepsilon), -4\varepsilon) \cup (4\varepsilon, -\log(6\varepsilon))$. *Choose $\varepsilon(A)$ and $\nu(A)$ such that conditions* (a)–(c) *of Theorem* 3 *are satisfied. Given $\infty > D(G_0, G_1; f_0, f_1) > 0$, there exists a constant $A^{(14)}$ such that $A > A^{(14)}$ and $\nu^+ = \nu + \lceil 8(\log^2 A)/D(G_0, G_1; f_0, f_1)\rceil$ implies*

$$\sup_{\nu \ge \nu(A)} \mathbb{P}_{\mathrm{BCP}}\{N_A > \nu^+\} \le A^{-1}(\log^{-2} A)\mathbb{P}_\nu\{N_A \ge \nu\}.$$

PROOF. Write $\xi = \nu^+/\nu$ and $\delta^+ = \nu^+ - \nu = \lceil 8(\log^2 A)/D\rceil$. From Lemmas 7 and 13,

$$\begin{aligned}\mathbb{P}_{\mathrm{BCP}}\{N_A > \nu^+\} &< 2\exp\left(-\frac{\delta^+}{16}\right)\mathbb{P}_\nu\{N_A \ge \nu\} \\ &\quad + \sum_{k=\lceil \nu - \delta^+/2\rceil}^{\lfloor \nu + \delta^+/2\rfloor} \binom{\nu^+}{k}\xi^k(1-\epsilon)^{\nu^+ - k}\mathbb{P}_k\{N_A > \nu^+\}.\end{aligned} \tag{16}$$

Write $D = D(G_0, G_1; f_0, f_1)$; by hypothesis, $D > 0$. Now let $u(x) = F_0^{-1}(G_0(x))$, so that $u(X_j) \sim F_0$ under the $\mathbb{P}_k^G$-measure, when $j < k$. Let

$$Z_j = \log\left(\frac{f_1\big(F_0^{-1}(G_0(X_j))\big)}{f_0\big(F_0^{-1}(G_0(X_j))\big)}\right) = \log\left(\frac{f_1\big(u(X_j)\big)}{f_0\big(u(X_j)\big)}\right)$$

denote the parametric log-likelihood of the transformed data after applying $u(\cdot)$ to all the observations.

Write $\nu_* = \nu - \delta^+$ and $l_n = n - \nu_*$. Let S_n denote the set of positive integer l_n-sequences $\mathbf{s}_{l_n} = (s_{l_n 1}, \ldots, s_{l_n l_n})$ such that $s_{l_n j} \le \nu_* + j$. In preparation for applying Lemma 10, define the random variable

$$\dot{L}_k^n(h, \mathbf{s}_{l_n}) = \left|\log\left(\frac{\Lambda_k^{n+1}(\boldsymbol{\rho}_{\nu_*}, \mathbf{s}_{l_n}, r_{\nu_*}(X_{n+1}, h, l_n))}{\Lambda_k^n(\boldsymbol{\rho}_{\nu_*}, \mathbf{s}_{l_n})}\right) - Z_{n+1}\right| I_{\{u(X_{n+1}) \in B_{\varepsilon(A)}\}}.$$

Note that $\dot{L}_k^n(h, \mathbf{s}_{l_n})$ is measurable with respect to the σ-field generated by $\mathscr{F}_{\nu_*}$ and X_{n+1}. For $j \ge 0$, let $\nu_j^* = \lceil(\nu + \nu^+)/2\rceil + j\lfloor 2\log(A)/D\rfloor$. Now define the events

$$B_1(j) = \left\{\log\left(\Lambda_{1+\nu_j^*}^{\nu_{j+1}^*}\right) > \log A\right\},$$

$$B_2(j) = \left\{\sum_{i=1+\nu_j^*}^{\nu_{j+1}^*} Z_i > \frac{3}{2}\log(A) \text{ and } \bigcap_{i=1+\nu_j^*}^{\nu_{j+1}^*}\left\{u(X_i) \in B_{\varepsilon(A)}\right\}\right\},$$

$$B_3(j) = \bigcap_{i=0}^{\lceil 2\log(A)/D\rceil - 1}\left\{\sup_{0 \le h \le l_{\nu_j^*+i}}\ \sup_{\mathbf{s}_{l_{\nu_j^*+i}} \in S_{\nu_j^*+i}} \dot{L}_{\nu_j^*+1}^{\nu_j^*+i}\left(h, \mathbf{s}_{l_{\nu_j^*}+i}\right) < \frac{D}{4}\right\},$$

with respective complements $\overline{B}_1(j)$ through $\overline{B}_3(j)$. Define also the event $B_0 = \{N_A \geq \nu\}$. Recall that $\Lambda^n_{n+1} = 1$ identically. Hence,

$$\begin{aligned} \{N_A > \nu^+\} &\subset \bigcap_{j=0}^{\lceil 2\log A\rceil - 1} B_0 \overline{B}_1(j) \\ &\subset B_0 \left[\bigcap_{j=0}^{\lceil 2\log A\rceil - 1} \overline{B}_2(j) \right] \cup \left[\bigcup_{j=0}^{\lceil 2\log A\rceil - 1} B_0 \overline{B}_3(j) \right]. \end{aligned} \tag{17}$$

For the rest of the proof we only consider $k \in (\nu - \delta^+/2, \nu + \delta^+/2)$. From the strong law of large numbers and hypothesis (a), $\mathbb{P}_k\{B_2(j)\} > 1 - e^{-1}$ whenever A exceeds some A^*. Because $k < \nu_0^*$, the events B_0 and $B_2(j)$ for $j = 0, 1, \ldots$ are independent under any measure $\mathbb{P}_k$. Hence

$$\mathbb{P}_k\left\{ B_0 \bigcap_{j=0}^{\lceil 2\log A\rceil - 1} \overline{B}_2(j) \right\} < \mathbb{P}_k\{B_0\}\exp(-2\log A) < A^{-2}\mathbb{P}_\infty\{N_A \geq \nu\}, \tag{18}$$

the last inequality from Lemma 7.

From hypotheses (b) and (c), there exists $A^{**} > A^*$ such that

$$b_1^{(10)}\left(-\epsilon(A)\log(\varepsilon(A)) + \frac{1}{\nu(A)\varepsilon(A)} \right) < \frac{D}{4},$$

for all $A > A^{**}$.

Let $\psi(x) = G_0^{-1}(G_1(x))$; by hypothesis $\psi(x) \leq x$. Consider now the sequence of random variables

$$\hat{X}_j = \begin{cases} X_j, & \text{if } j < k, \\ \psi(X_j), & \text{if } k \leq j < \nu, \\ X_j, & \text{if } \nu \leq j. \end{cases} \tag{19}$$

Note that $X_j = \hat{X}_j$ identically if $k \geq \nu$, that $\hat{X}_j \leq X_j$ always and that under $\mathbb{P}_k$ the observations $\hat{X}_1, \ldots, \hat{X}_{\nu-1}$ are i.i.d. with distribution G_0. Let $\hat{\Lambda}^n_k$, $\hat{N}_A$ and the events $\hat{B}_3(j)$ denote the nonparametric likelihood ratio, stopping time and events computed using the $\{\hat{X}_j\}$ instead of the $\{X_j\}$. By construction, $\hat{B}_3(j) = B_3(j)$. Because the likelihood ratios are invariant to (almost surely) strictly increasing transformations, we may use Lemma 7 to conclude $B_0 \subset \hat{B}_0$. Use the coupling (19) and apply Lemma 10 $\delta^+/2$ times to conclude that

$$\begin{aligned} \mathbb{P}_k\left\{ \bigcup_{j=1}^{\lceil 2\log A\rceil} B_0 \overline{B}_3(j) \right\} &\leq \mathbb{P}_k\left\{ \bigcup_{j=1}^{\lceil 2\log A\rceil}{}' \hat{B}_0 \overline{B}_3(j) \right\} \\ &= \mathbb{P}_\nu\left\{ \bigcup_{j=1}^{\lceil 2\log A\rceil} B_0 \overline{B}_3(j) \right\} \\ &= \mathbb{E}_\nu\left\{ \mathbb{P}_\nu\left\{ \bigcup_{j=1}^{\lceil 2\log A\rceil} \overline{B}_3(j) \middle| \mathscr{F}_{\nu-1} \right\}; B_0 \right\} \\ &\leq b_2^{(10)}\delta^+ \exp\bigl(-b_3^{(10)}(\nu(A) - \delta^+)\varepsilon^4(A)\bigr)\mathbb{P}_\nu\{B_0\}. \end{aligned} \tag{20}$$

Finally, combine (16), (17), (18) and (20) using hypothesis (c) to choose $A^{(14)}$ so that $A > A^{(14)}$ implies $\mathbb{P}_k\{N_A > \nu^+\} \le A^{-1}(\log^{-2} A)\mathbb{P}_\nu\{N_A \ge \nu\}$. □

LEMMA 15. *Let $\nu(A)$ and $\varepsilon(A)$ satisfy hypotheses* (a)–(c) *used in Lemma* 14. *Let $\infty > D = D(G_0, G_1; f_0, f_1) > 0$ and $p\alpha \ge q\beta$. Given $\zeta \in (0, \frac{1}{3})$, there exists $A^{(15)}(\zeta)$ such that $A > A^{(15)}(\zeta)$ and $\nu > \nu(A)$ imply*

$$\mathbb{E}_\nu^G\left\{N_A - \nu; \nu \le N_A \le \nu + \left\lceil \frac{8(\log^2 A)}{D} \right\rceil\right\} \le (1 + \zeta)\frac{\log A}{D}\mathbb{P}_\nu^G\{\nu \le N_A\}.$$

PROOF. Let $\delta \in (0, \frac{1}{3})$ be some constant whose value is to be specified later. Let $u(x)$ and Z_n be as in the proof of Lemma 14. In preparation for applying Lemma 10 to the sequence $u(X_1), \dots$, let $l_n = n - (\nu - 1)$ and let S_n denote the set of all positive integer sequences $\mathbf{s}_{l_n} = (s_{l_n 1}, \dots, s_{l_n l_n})$, where $s_{l_n j} \le \nu + j - 1$. Define the random variable $\ddot{L}_k^n$ to be

$$\ddot{L}_k^n(h, \mathbf{s}_{l_n}) = \left|\log\left(\frac{\Lambda_k^{n+1}(\boldsymbol{\rho}_{\nu-1}, \mathbf{s}_{l_n}, r_{\nu-1}(X_{n+1}, h, l_n))}{\Lambda_k^n(\boldsymbol{\rho}_{\nu-1}, \mathbf{s}_{l_n})}\right) - Z_{n+1}\right| I_{\{u(X_{n+1}) \in B_{\varepsilon(A)}\}}.$$

Note that $\ddot{L}_k^n(h, \mathbf{s}_{l_n})$ is measurable with respect to the σ-field generated by $\mathscr{F}_{\nu-1}$ and X_{n+1}. For $j \ge 0$, let $\nu_j = \nu + j\lfloor(1 + 3\delta)(\log A)/D\rfloor$. Let

$$B_0 = \{N_A \ge \nu\},$$

$$B_1(j) = \left\{\log\left(\Lambda_{\nu_j}^{\nu_{1+j}-1}\right) > \log A\right\},$$

$$B_2(j) = \left\{\sum_{i=\nu_j}^{\nu_{1+j}-1} Z_i > (1 + 2\delta)\log A \text{ and } \bigcap_{i=\nu_j}^{\nu_{1+j}-1} \{u(X_i) \in B_{\varepsilon(A)}\}\right\},$$

$$B_3(j) = \bigcap_{i=0}^{\lfloor(1+3\delta)(\log A)/D\rfloor - 1} \left\{\sup_{0 \le h \le l_{\nu_j+i-1}} \sup_{\mathbf{s}_{l_{\nu_j+i-1}} \in S_{\nu_j+i-1}} \ddot{L}_{\nu_j}^{\nu_j+i-1}\left(h, \mathbf{s}_{l_{\nu_j+i-l}}\right) < \delta D\right\},$$

with corresponding complements $\overline{B}_1(j)$ through $\overline{B}_3(j)$. Recall that $\Lambda_{n+1}^n = 1$ identically.

Choose $A^*(\delta)$ so that $A > A^*(\delta)$ implies that

$$\mathbb{P}_\nu^G\{\overline{B}_2(j)\} < \delta,$$

$$b_1^{(10)}\left(-\varepsilon(A)\log(\varepsilon(A)) + \frac{1}{\varepsilon(A)\nu_A}\right) < \delta,$$

$$\mathbb{P}_\nu^G\left\{\bigcup_{j=0}^{\lceil 2\log(A)\rceil - 1} \overline{B}_3(j) \middle| \mathscr{F}_{\nu-1}\right\} < A^{-2},$$

the last inequality from Lemma 10, applied once for each observation taken between times ν and $\nu + \lceil 8(\log^2 A)/D\rceil$.

Note that $\overline{B}_1(j) \subset \overline{B}_2(j) \cup \overline{B}_3(j)$ and that the events $B_2(j)$ are independent under the $\mathbb{P}_\nu^G$-probability. Hence, when $A > A^*(\delta)$,

$$\mathbb{E}_\nu^G\left\{N_A - \nu; \nu \le N_A \le \nu + \left\lceil 8\frac{\log^2(A)}{D}\right\rceil\right\}$$

$$\le (1+3\delta)\frac{\log A}{D}\sum_{j=0}^{\lceil 2\log A\rceil - 1}(j+1)\mathbb{P}_\nu^G\left\{B_0 B_1(j) \cap \bigcap_{i=0}^{j-1}\overline{B}_1(i)\right\}$$

$$+\left\lceil 8\frac{\log^2(A)}{D}\right\rceil \mathbb{P}_\nu^G\left\{B_0 \cap \bigcap_{i=0}^{\lceil 2\log A\rceil - 1}\overline{B}_1(i)\right\}$$

$$< (1+3\delta)\frac{\log A}{D}\sum_{j=0}^{\lceil 2\log A\rceil - 1}\left[\mathbb{P}_\nu^G\left\{B_0 \cap \bigcap_{i=0}^{j-1}\overline{B}_2(i)\right\} + \mathbb{P}_\nu^G\left\{\bigcup_{i=0}^{\lceil 2\log A\rceil - 1} B_0\overline{B}_3(i)\right\}\right]$$

$$+\left\lceil 8\frac{\log^2(A)}{D}\right\rceil\left[\mathbb{P}_\nu^G\left\{B_0 \bigcap_{i=0}^{\lceil 2\log A\rceil - 1}\overline{B}_2(i)\right\} + \mathbb{P}_\nu^G\left\{\bigcup_{i=0}^{\lceil 2\log A\rceil - 1} B_0\overline{B}_3(i)\right\}\right]$$

$$< \left[\frac{1+3\delta}{1-\delta}\frac{\log A}{D} + 2\frac{\lceil 8(\log^2 A)/D\rceil}{A^2}\right]\mathbb{P}_\nu^G\{\nu \le N_A\}.$$

Choose and fix δ sufficiently small, then choose A sufficiently large to complete the proof. □

To prove Theorem 3, combine Lemmas 12–15.

In a similar problem studied in Gordon and Pollak (1994a), we did establish equality in an analog of Theorem 3. The source of the additional technical difficulties in the current problem is that $\mathbb{P}_\infty^G = \mathbb{P}_1^G$, making it much harder to show the negligibility of contributions attributable to an hypothesized early change.

In attempting to strengthen Theorem 3, we must condition on $\{N_A > \nu\}$, whose probability tends to 0 as $\nu \to \infty$. One can show for $0 < c < \infty$ that $\liminf_{A\to\infty} \mathbb{P}_\infty\{N_A > cA\} > 0$. Therefore, if $\nu(A)$ can be chosen to be of the order of A or less and still be consistent with hypothesis (c) of Theorem 3, one may then obtain equality in that theorem. [For example, $\nu(A) = \log^{2+\omega} A$ for $\omega > 0$ suffices to establish the lower bound when $G_0 = \mathscr{N}(0,1)$ and $G_1 = \mathscr{N}(1,1)$.] One needs an analog of Lemma 3 valid for small values of k. This is possible if one observes that since most observations are hypothesized to behave like F_1 when k is close to 1, the pivot for m should be nq instead of $n/2$. The rest of the proof is similar in concept to proofs given above.

4.6. *Sensitivity to a shift in normal mean*: *Proof of Corollary* 4. We prove Corollary 4 by establishing the conditions of Theorem 3. We first find a candidate for $\varepsilon(A)$.

LEMMA 16. *Let $V \sim \mathcal{N}(0,1)$ have the standard normal distribution $\Phi(\cdot)$, and let $\mu > 0$. Let B_ε be the set* $(\log(6\varepsilon), -4\varepsilon) \cup (4\varepsilon, -\log(6\varepsilon))$.
Choose $\xi > 0$, and let $\varepsilon(A) = \log^{-(1+\zeta)} A$. As $A \to \infty$,

$$(\log A)\mathbb{P}\{F_0^{-1}(\Phi(\mu + V)) \notin B_{\varepsilon(A)}\} \to 0.$$

PROOF. Write $\phi(\cdot)$ for the standard normal density. Because the family $\phi(x - \mu)$ has monotone likelihood ratio, and because $\phi(\cdot)$ is unimodal and symmetric,

$$\begin{aligned}
&\mathbb{P}\{F_0^{-1}(\Phi(\mu + V)) \notin B_\varepsilon\} \\
&\quad < \mathbb{P}\{\Phi(V) \le 3\varepsilon\} + \mathbb{P}\{V \in [\Phi^{-1}(\tfrac{1}{2}\exp(-4\varepsilon)), \Phi^{-1}(1 - \tfrac{1}{2}\exp(-4\varepsilon))]\} \\
&\quad\quad + \mathbb{P}\{\Phi(\mu + V) \ge 1 - 3\varepsilon\}.
\end{aligned}$$

Hence $\mathbb{P}\{F_0^{-1}(\Phi(\mu + V)) \notin B_{\varepsilon(A)}\} < 3\varepsilon + 4\varepsilon + \mathbb{P}\{3\varepsilon \ge 1 - \Phi(\mu + V)\}$. We now use Birnbaum's (1942) inequality $1 - \Phi(x) > 2(x + \sqrt{x^2 + 4})^{-1}\phi(x)$ for $x \ge 0$ to obtain

$$\begin{aligned}
&\mathbb{P}\{3\varepsilon \ge 1 - \Phi(\mu + V)\} \\
&\quad \le \mathbb{P}\left\{3\varepsilon \ge \frac{1}{\sqrt{(\mu + V)^2 + 2}} \exp\left(-\frac{\mu^2}{2}\right)\exp(-\mu V)\phi(V)\right\}
\end{aligned}$$

and then use that $(2 + (\mu + v)^2)^{-1/2}\exp(-\mu v)\phi(v)$ is decreasing in $v > 0$ to show

$$\mathbb{P}\{3\varepsilon \ge 1 - \Phi(\mu + V)\} \le \mathbb{P}\left\{V > \sqrt{-2\log(3\varepsilon) - 4\mu\sqrt{-\log(3\varepsilon)}}\right\},$$

for ε sufficiently small. Hence, for A sufficiently large,

$$\begin{aligned}
&\log(A)\mathbb{P}\{F_0^{-1}(\Phi(\mu + V)) \notin B_{\varepsilon(A)}\} \\
&\quad < 7\log^{-\xi}(A) + 3\log^{-\xi}(A)\exp\left(2\mu\sqrt{-\log(3\log^{-(1+\xi)}(A))}\right) \\
&\quad < \log^{-\xi/2} A,
\end{aligned}$$

proving the lemma. □

To complete the proof of Corollary 4, observe that $\varepsilon(A)$ (as given in the preceding lemma) and $\nu(A) = \log^{5+5\xi} A$ satisfy the conditions required by Theorem 3.

REFERENCES

BIRNBAUM, Z. W. (1942). An inequality for Mill's ratio. *Ann. Math. Statist.* **13** 245–246.

GORDON, L. and POLLAK, M. (1994a). An efficient sequential nonparametric scheme for detecting a change of distribution. *Ann. Statist.* **22** 763–804.

GORDON, L. and POLLAK, M. (1994b). Average run length to false alarm for surveillance schemes designed with partially specified prechange distribution. Unpublished manuscript.

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

LOMBARD, F. (1983). Asymptotic distributions of rank statistics in the change-point problem. *South African Statist. J.* **17** 83–105.

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). Robust cusum: a robustness study for cusum quality control schemes. *Comm. Statist. Theory Methods* **11** 2669–2687.

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.

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. and SIEGMUND, D. (1975). Approximations to the expected sample size of certain sequential tests. *Ann. Statist.* **3** 1267–1282.

POLLAK, M. and SIEGMUND, D. (1990). Personal communication.

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

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-samples 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.

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.

VAN DOBBEN DE BRUYN, C. S. (1968). *Cumulative Sum Tests.* Hafner, New York.

YAKIR, B. (1994). Optimal detection of a change in distribution, when the observations are independent. Unpublished manuscript.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SOUTHERN CALIFORNIA
1042 WEST 36TH PLACE, DRB #155
LOS ANGELES, CALIFORNIA 90089-1113

DEPARTMENT OF STATISTICS
HEBREW UNIVERSITY OF JERUSALEM
91905 JERUSALEM
ISRAEL