Changepoint Tests Designed for the Analysis of Hiring Data Arising in Employment Discrimination Cases

Boris FREIDLIN
National Cancer Institute, Bethesda, MD 20892-7434 (*freidlinb@ctep.nci.nih.gov*)

Joseph L. GASTWIRTH
Department of Statistics, George Washington University, Washington, DC 20052 (*jlgast@research.circ.gwu.edu*)

When a complaint of discrimination is made, an employer may respond by hiring more minorities. From a legal viewpoint, the practices in effect during the time period prior to the complaint are more relevant for determining liability than those of the postcharge period. In *Gay v. Waiters,* the trial judge observed that the data suggested that a change occurred after the charge was filed. Because the data had not been subject to a formal statistical analysis, the court was reluctant to base its decision on this observation. Gastwirth and Freidlin and Gastwirth proposed cumulative-sum-based procedures for the analysis of hiring data following the binomial model. In this article, the procedures are extended to data following the hypergeometric model and to analysis of stratified data. Several datasets that were submitted to the courts in the United States are analyzed by the proposed methods. Because the data are usually reported by year, the ordinary large-sample theory is not sufficiently accurate. Therefore, we obtain the p values of the statistics by simulation. For binomial data, recent improvements in the Bonferroni inequality are used to derive a new upper bound.

KEY WORDS: Binomial data; Changepoint; Fair hiring practices; Hypergeometric data.

After a complaint of discrimination in hiring is filed, the employer may change its practices. By the time the trial occurs, the defendant may submit hiring data showing that a protected group received its "fair share" of hiring over a time frame that includes both the precharge and postcharge periods. From a legal viewpoint the most relevant time period is the one just prior to a complaint (see *O'Brien v. Sky Chefs. Inc.* 1982). The issue of whether there has been a change in employment practices has occurred in several cases (e.g., *Gay v. Waiters* 1982).

When a test procedure is being developed for a particular application, one can increase its power by directing the test at a specific alternative. Agresti (1990, pp. 97–102) presented several examples of such procedures; Levin and Kline (1985) developed a changepoint test to detect a sudden increase in the proportion of chromosomally abnormal abortions and a subsequent decline to the normal level. The pattern of concern in the legal setting is underrepresentation before the charge and a change to fair or even overhiring of minorities sometime after the charge but prior to the trial. This article modifies the tests based on the cumulative sums (CUSUM's) to have more power against this alternative. In statistical literature, changepoint methods were discussed by Page (1954), Chernoff and Zacks (1964), Hinkley and Hinkley (1970), Pettitt (1979, 1980), Worsley (1983), Siegmund (1986), James, James, and Siegmund (1987), Lombard (1987), Csorgo and Horvath (1988), Bhattacharya (1987, 1994), and Brostrom (1997). Econometric applications often involve dependent observations. Appropriate changepoint techniques were developed by Ploberger and Kramer (1990), Banerjee, Lumsdaine, and Stock (1992), Chu and White (1992), Jandhyala and MacNeill (1992), Andrews and Ploberger (1994), and Vogelsang (1997).

The data available for examining hiring practices typically follow either a hypergeometric or a binomial model (Baldus and Cole 1980; Gastwirth 1988; Finkelstein and Levin 1990; Paetzold and Willborn 1994). When data on the actual applicants are available and reliable, the hypergeometric model is appropriate. When the applicant data are unavailable or when fair recruitment is an issue, the minority share of the qualified labor force is determined from the census data on the members of the labor force in the area with the relevant skills. Then the hiring data are modeled by a binomial distribution. Section 1 describes procedures for the binomial model, including an upper bound for the p value of our statistics. The additional insight obtained from the CUSUM methods is illustrated on data from a case that settled. The p values yielded by the proposed tests were notably smaller, thereby providing more convincing evidence of discrimination than a test of the data over the entire period. Analogs of these CUSUM-type procedures appropriate for data following the hypergeometric model are given in Section 2. Their extension to the stratified setting is also presented, along with a reanalysis of a stratified dataset from an actual case. Section 3 discusses a recent case in which the hypergeometric model was rejected by a court because the applicant data were affected by underrecruiting. The female fraction of applicants was much lower than their availability in the labor market be-

Journal of Business & Economic Statistics
July 2000, Vol. 18, No. 3

cause the recruiting process was biased. Using the court's availability figure, the binomial changepoint techniques of Section 1 strongly support the judge's decision.

1. STATISTICAL PROCEDURES FOR ANALYZING BINOMIAL DATA FOR CHANGEPOINT

One approach courts use to assess hiring practices is to compare the fraction of minority hires over the period of time to their fraction, π, of the qualified labor force. Typically, the value of π is determined from external data—for example, census data on persons employed in similar jobs in the area (Baldus and Cole 1980; Gastwirth 1988, 1997). Technically speaking, π is an estimate because the occupational and related data in the Census are obtained from a large sample (about one in six households receives the long survey form). The resulting small "sampling error" in π is typically ignored in this application.

When an employer hires a relatively small number, n_i, of workers in year i from a large labor pool, the number x_i of minority workers can be regarded as a binomial random variable with parameters n_i and π_i. Thus, the yearly hires $x_1, \ldots, x_T$ may be modeled as independent binomial random variables with parameters π_i, n_i. We are testing

$$\begin{aligned} H_0\colon &\quad \pi_i = \pi, &&\text{for } i = 1, \ldots, T \quad \text{against} \\ H_1\colon &\quad \pi_i = \pi_{01} < \pi, &&\text{for } i = 1, \ldots, k \\ &\quad \pi_i = \pi_{02} \geq \pi, &&\text{for } i = k+1, \ldots, T, \end{aligned} \tag{1.1}$$

where $1 \leq k < T$ and n_i employees are hired in the ith year. The null hypothesis, H_0, corresponds to "fair hiring," and H_1 corresponds to underrepresentation in the first k years.

Gastwirth (1996) based a test for underhiring on the CUSUM's

$$S_k = \sum_{i=1}^{k} (x_i - n_i\pi), \qquad \text{for } k = 1, \ldots T, \tag{1.2}$$

which is the CUSUM of the observed minus expected hires. The graph of S_k for $k = 1, \ldots, T$ helps to identify the time of the change if one exists. The test statistic is

$$\mathrm{ST}_1 = \frac{\min_{1 \leq k \leq T} (\sum_{i=1}^{k} (x_i - n_i\pi))}{\sqrt{N\pi(1-\pi)}},$$

$$\text{where} \quad N = \sum_{i=1}^{T} n_i, \tag{1.3}$$

which rejects for low values of ST_1. Freidlin and Gastwirth (1998) proposed the following test for change in hiring:

$$\mathrm{ST}_2 = \frac{\min_{1 \leq k \leq T} (\sum_{i=1}^{k} (x_i - \pi n_i) - \sum_{i=k+1}^{T} (x_i - \pi n_i))}{\sqrt{N\pi(1-\pi)}}. \tag{1.4}$$

The test based on (1.4) rejects for low values of ST_2 and is similar to the maximally selected chi-squared statistic (Miller and Siegmund 1982). In equal-employment applications, the alternative of interest has two components, (1) change in hiring after the charge was filed and (2) underrepresentation before the change. Because ST_1 is based on tests used to detect underrepresentation, whereas ST_2 was designed to detect a changepoint, a combination of ST_1 and ST_2 should have increased power against the alternatives arising in hiring cases. An investigation of several approaches to combining ST_1 and ST_2 concluded that sum $\mathrm{ST}_1 + \mathrm{ST}_2$ has good properties (Freidlin 1998).

Because the data usually are only reported on a yearly basis, a confidence interval for the changepoint would be quite wide, especially since a charge must be filed within 180 days of the alleged discriminatory act. Thus, we recommend using the CUSUM chart. Although one might simply compare the proportions of minority hires before and after the charge is filed, there can be a wide variation in the time when an employer may change its practices. Some employers change their practices just after the charge is filed; others wait until the Equal Employment Opportunity Commission (EEOC) has given the plaintiff a "right to sue letter," and sometimes there is a change in the year before a trial. Because employers are only required to keep records for two years, unless a charge is filed, the amount of precharge data is often considerably less than the amount of postcharge data.

The standard asymptotic approximations for ST_1, ST_2, and $\mathrm{ST}_1 + \mathrm{ST}_2$ were developed for individual observations and are not accurate for data reported by year (Freidlin and Gastwirth 1998), so p values of the tests should be estimated by simulation. Furthermore, the cumulative nature of the statistics S_k combined with their known correlation structure can be used to obtain an upper bound for the p values of the proposed tests. This upper bound is based on a two-point approximation (described in the Appendix) and is an improvement on the Bonferroni approach (Worsley 1985; Efron 1997). Upper bounds are useful in the legal setting because issues of the accuracy of an asymptotic approximation or simulation can no longer be raised. For the datasets with the upper bound slightly higher than the prescribed significance level—for example, .06 or .07 for a .05-level test—simulations should be carried out. The proposed tests are compared to the likelihood ratio test (Hinkley and Hinkley 1970; Worsley 1983), which is modified for the model (1.1) as described by Freidlin and Gastwirth (1998).

We now demonstrate the use of the procedure on data from a case that was settled. The p values of the tests are estimated empirically based on 100,000 replications. Upper bounds for the p values based on the two-point approximations are also given.

The issue in the case was whether the employer's hiring practices were gender neutral. The charge was filed in May 1994. It was determined that females formed 3.43% of the qualified labor force (QLF); that is, $\pi = .0343$. The data are given in Table 1 and the CUSUM chart is given in Figure 1. Note that the graph declines from 1991 until 1994; then it levels off indicating fair hiring. In 1996 the graph jumps dramatically, reflecting the fact that the company hired far more females than expected that year.

The results of the various tests on the data in the table are given in Table 2. All the proposed tests are highly

Table 1. Yearly Hiring Data for the Case That Was Settled

Year	Total hires	Female hires	Expected	S_k	$S_T - S_k$
1991	429	2	14.71	−12.71	−.17
1992	86	0	2.95	−15.66	2.78
1993	104	0	3.57	−19.23	6.34
1994	180	0	6.17	−25.41	12.52
1995	116	5	3.98	−24.38	11.50
1996	73	14	2.50	−12.89	0

NOTE: The "Expected" column is obtained by multiplying the total number (n_i) of hires in each year by the female fraction (π = .0343) of the QLF; S_k is the CUSUM of the observed minus expected number of female hires; S_T is the difference between the observed and expected numbers of female hires over the entire period.

significant. A defendant might ignore the change and use the normal form of the test S_T; that is, $Z = \sum_{i=1}^{T}(x_i - n_i\pi)/\sqrt{N\pi(1-\pi)}$. This Z test for the entire period has p value .0116, which is not as convincing as the changepoint procedures. This is a result of overhiring in 1996. Because the change in Figure 1 occurred at the end of 1994, analyzing the data for the 1991–1994 period with the binomial distribution yields an exact p value of 3.3E-10.

The methods of this section assumed that the applicants in each year were independent from year to year. If some applicants who are not hired the first time reapply, there would be a slight dependence. In actual cases, the fraction of repeat applications is minimal; however, the problem can be resolved by introducing new strata for these applicants. One would need to use the hypergeometric model discussed in Section 2 for those strata. Under the null hypothesis of nondiscrimination by the defendant, assuming that other employers in the area were fair, the probability that a rejected applicant from a minority group would reapply should be the same as that of a similar majority group member. Thus, π remains minority share of all applicants.

The focus of the methods of this article is on the analysis of one employer's practices. The issue of the fairness of many alternative employers is related to the more general question of economywide discrimination, which was discussed by Becker (1971), Blau, Ferber, and Winkler (1997), and Goldin (1992).

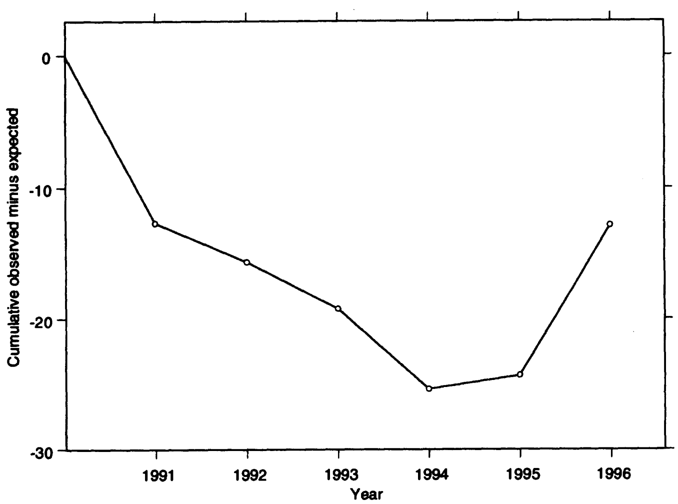

Figure 1. The CUSUM Chart for the Case That Was Settled.

Table 2. The p Values of the Tests Applied to the Case That Was Settled

Statistic	Empirical p value	Two-point p value
ST_1	<.0001	7.7E-8
ST_2	<.0001	1.8E-10
$ST_1 + ST_2$	<.0001	7.9E-13
LR	<.0001	
Z test	.0116	

NOTE: Empirical p values are based on 100,000 replications. The two-point p values were obtained using the two-point upper bound given in the Appendix.

2. STRATIFIED CHANGEPOINT TESTS FOR THE HYPERGEOMETRIC MODEL

Sometimes the numbers of applicants or employees eligible for promotion in both groups are known for year i. Then, for each year i conditional on the total number of hires (promotions), the number of minority hires (promotions) can be modeled by a hypergeometric random variable. Because employees in different occupations or at different qualification levels should only be compared to similar employees, the data should be appropriately stratified. In the sth stratum ($s = 1, \ldots, S$) of year i, let n_{1is} and n_{2is} denote the number of applicants or eligible employees in the two groups. The total number of hires (promotions) is denoted by m_{is}, and the number of minority hires (promotions) is x_{is}. Consider the following model, where $x_{is} \sim$ noncentral hypergeometric with $\theta_{is} = e^{\beta}$ for $i = 1, \ldots, k$ and $s = 1, \ldots, S$ and $x_{is} \sim$ noncentral hypergeometric with $\theta_{is} = e^{-\beta}$ for $i = k+1, \ldots, T$ and $s = 1, \ldots, S$. The likelihood is

$$L(\beta) = \prod_{i=1}^{k}\prod_{s=1}^{S} \frac{\binom{n_{1is}}{x_{is}}\binom{n_{2is}}{m_{is}-x_{is}} e^{\beta x_{is}}}{\sum_{j=m_{ij-}}^{m_{ij+}} \binom{n_{1is}}{j}\binom{n_{2is}}{m_{is}-j} e^{\beta j}} \times \prod_{i=k+1}^{T}\prod_{s=1}^{S} \frac{\binom{n_{1is}}{x_{is}}\binom{n_{2is}}{m_{is}-x_{is}} e^{-\beta x_{is}}}{\sum_{j=m_{ij-}}^{m_{ij+}} \binom{n_{1is}}{j}\binom{n_{2is}}{m_{is}-j} e^{-\beta j}}, \tag{2.1}$$

where $m_{is-} = \max(0, n_{1is} + m_{is} - n_{is})$ and $m_{is+} = \min(n_{1is}, m_{is})$. Here the null hypothesis H_0: $\beta = 0$ corresponds to a fair hiring or promotion practice. An alternative hypothesis H_1: $\beta < 0$ corresponds to underhiring (promoting) in the period $[1, k]$ with change to overhiring (promoting) after year k. The efficient score test for testing H_0 against H_1 is

$$\frac{\left[\frac{\partial \quad l}{\partial\beta}\right]_{\beta=0}}{\sqrt{-E\left[\frac{\partial^2 l}{\partial \quad \beta^2}\right]_{\beta=0}}} = \frac{\sum_{i=1}^{k}\sum_{s=1}^{S}\left(x_{is} - \frac{m_{is} \quad n_{1is}}{n_{is}}\right) - \sum_{i=k+1}^{T}\sum_{s=1}^{S}\left(x_{is} - \frac{m_{is} \quad n_{1is}}{n_{is}}\right)}{\sqrt{\sum_{i=1}^{T}\sum_{s=1}^{S}\frac{n_{1is}n_{2is}m_{is}(n_{is}-m_{is})}{n_{is}^2(n_{is}-1)}}}. \tag{2.2}$$

Table 3. Number of Employees/Promotions From Valentino v. U.S.P.S Case

Time period	Grade 17–19		Grade 20–22		Grade 23–25		Grade 26–28		Grade 29–31	
	M	*F*	*M*	*F*	*M*	*F*	*M*	*F*	*M*	*F*
06/74	229	73	360	48	703	33	236	7	82	1
−03/75	67	5	74	9	132	2	28	1	8	0
03/75	205	89	373	43	716	36	277	9	85	1
−01/76	40	6	39	5	41	1	19	0	7	0
01/76	233	101	396	52	727	36	271	9	85	2
−01/77	31	10	32	4	54	5	28	2	5	0
01/77	200	86	377	52	680	35	262	8	89	3
−01/78	43	18	80	9	57	6	18	1	14	0
01/78	196	90	325	50	685	37	252	9	78	3
−01/79	29	8	45	7	42	3	14	1	6	1

NOTE: Key to symbols: F = females; M = males; for any time period and grade group, the number of promotions is below the number of employees. For example, in grades 17–19 during the 06/74–03/75 period, 67 out of 229 eligible males were promoted compared to 5 out of 73 eligible females.

Analogs of the test statistics (1.3) and (1.4) are

$$\mathrm{ST}_{1sh} = \frac{\min_{1\le k<T} \sum_{i=1}^{k} \sum_{s=1}^{S} \left(x_{is} - \frac{m_{is} n_{1is}}{n_{is}} \right)}{\sqrt{\sum_{i=1}^{T} \sum_{s=1}^{S} \frac{n_{1is} n_{2is} m_{is} (n_{is}-m_{is})}{n_{is}^2 (n_{is}-1)}}} \tag{2.3}$$

and

$$\mathrm{ST}_{2sh} = \frac{\min_{1\le k<T} \left(\sum_{i=1}^{k} \sum_{s=1}^{S} \left(x_{is} - \frac{m_{is} n_{1is}}{n_{is}} \right) - \sum_{i=k+1}^{T} \sum_{s}^{S} \left(x_{is} - \frac{m_{is} n_{1is}}{n_{is}} \right) \right)}{\sqrt{\sum_{i=1}^{T} \sum_{s=1}^{S} \frac{n_{1is} n_{2is} m_{is} (n_{is}-m_{is})}{n_{is}^2 (n_{is}-1)}}}. \tag{2.4}$$

Again, we combine ST_{1sh} and ST_{2sh} in $\mathrm{ST}_{1sh}+\mathrm{ST}_{2sh}$. The generalization of the likelihood ratio test to the hypergeometric setting is routine and was discussed by Freidlin (1998).

We illustrate the procedures on data from *Valentino v. U.S.P.S* (1982) for which one of the authors (JLG) served as plaintiff's expert. The plaintiff filed a charge of sex discrimination in promotion at the Postal Service after she was denied a promotion in mid-1976. The judge certified the case as a class action concerning whether women employed at the grade 17 and higher were promoted fairly. The available data on employees, grouped into strata composed of several grades in accordance with the U.S. Civil Service Commission reports, is presented in Table 3. To show that the women had received fewer promotions than expected until the charge, the plaintiffs presented the Mantel–Haenszel analysis (Table 4). The Mantel–Haenszel statistic for each time period combines the differences between the observed and expected numbers of promotions across all the strata. Its standardized form is asymptotically normally distributed. For further details and examples, see Agresti (1990) and Gastwirth (1988). These results demonstrate that women did not receive their expected number of promotions during the two years prior to the charge, but afterward they did. The judge did not accept these data as showing discrimination because he preferred an occupational rather than a grade-level stratification. He did credit a regression submitted by the defendant, using grade level as a proxy for occupation, as showing that employees were fairly paid.

Table 4. The Mantel–Haenszel Analysis of the Valentino v. U.S.P.S Case Data

Year	Observed	Expected	p value (two-sided)
1974–1975	17	34.1	.0006
1975–1976	12	21.16	.020
1976–1977	21	20.32	.885
1977–1978	34	33.23	.869
1978–1979	20	21.66	.674

We now show how changepoint methods provide more insight. The CUSUM charts for the data in Table 3 are given in Figure 2; they show that first group (levels 17–19) dominates the overall statistic, although the CUSUM's of third group (levels 23–25) also decline in the early years. The p values of the tests are $\mathrm{ST}_{1sh}(.002)$, $\mathrm{ST}_{2sh}(.011)$, $\mathrm{ST}_{1sh} + \mathrm{ST}_{2sh}(.002)$, and $\mathrm{LR}_{sh}(< .001)$; all are significant at the .05 level.

As in other applications of combination methods, one needs to examine the data for homogeneity, especially for a qualitative interaction (Gail and Simon 1985). The paucity of females in the upper grade levels and the fact that no CUSUM chart was above its expected value until after 1976 indicate that combining the data is appropriate.

The remaining question is whether the Post Office changed it policy before or after the plaintiff was denied the promotion. Only half of the actual files for the appli-

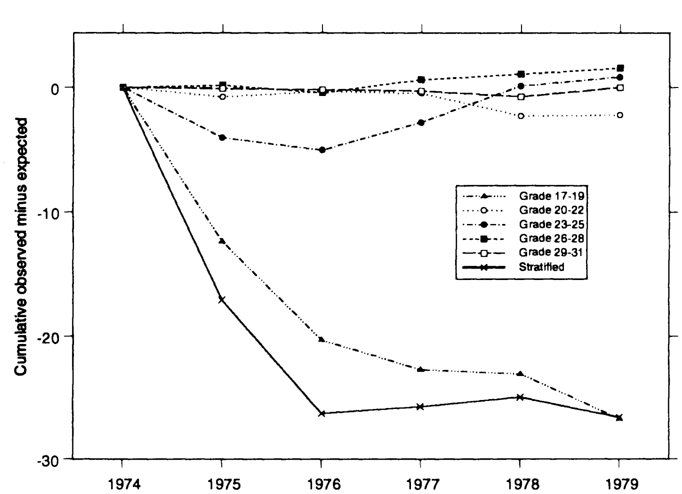

Figure 2. Individual and Stratified CUSUM Charts for Valentino Data.

Table 5. Number of Applicants and Hires for EEOC v. Joe's Stone Crab, Inc.

Season	Female applicants	Female hires	Male applicants	Male hires
1986–1987	5	0	95	35
1987–1988	5	0	95	19
1988–1989	5	0	95	19
1989–1990	5	0	95	14
1990–1991	5	0	95	21
1991–1992	21	3	118	12
1992–1993	28	5	100	17
1993–1994	21	2	70	17
1994–1995	42	6	115	11
1995–1996	30	3	99	12

NOTE: The data are taken from the report of the defendant's expert who assumed that in the early years there were about 100 applicants, 5% of whom were females. Thus, the data for those years were imputed.

Table 6. The p Values for EEOC v. Joe's Stone Crab, Inc.

Statistic	Empirical p value	Two-point p value
	Hypergeometric model	
ST_{1h}	.062	
ST_{2h}	.163	
$ST_{1h}+ST_{2h}$	.063	
LR_h	.024	
	Binomial model	
ST_1	<.0001	1.7E-13
ST_2	<.0001	9.5E-11
ST_1+ST_2	<.0001	1.8E-13
LR	<.0001	
Z test	<.0001	

NOTE: Empirical p values for the hypergeometric and binomial models are based on 10,000 and 100,000 replications, respectively. The two-point p values were obtained using the two-point upper bound given in the Appendix.

cants for promotions made in the period from January 1976 to the time of the complaint were available, creating the possibility of a sample-selection bias because the files were not randomly selected for retention. None of the promotions for which the data were available, however, went to a female.

3. CHOOSING BETWEEN THE BINOMIAL AND HYPERGEOMETRIC MODELS

When reliable data on actual applicants are available, courts prefer an analysis based on the hypergeometric model (Baldus and Cole 1980). When applicant data is unavailable or unreliable or when the fairness of the recruitment process itself is at issue, courts use the binomial model, where π is determined from external data.

We illustrate these issues on the data from *EEOC v. Joe's Stone Crab, Inc.* (1997), a case concerning alleged sex discrimination in the hiring of food servers at a Miami Beach restaurant. The EEOC charge was filed in June of 1991. The data are given in Table 5. The defendant asserted that an applicant flow analysis should be used, but the government argued that the applicant data reflected discriminatory recruitment because about 44% of the people employed as food servers in the Miami Beach area were female. The court ultimately determined the female availability, π, to be .319.

The CUSUM chart for the hypergeometric analysis of the data in Table 5 (defendant's approach) is given in Figure 3, and the test results are given in Table 6. The tests ST_{1h}, ST_{2h}, and $ST_{1h}+ST_{2h}$, which are the test statistics given in (2.3) and (2.4) when there is only one stratum, are not significant at the .05 level. The opinion stated that "Joe's historical practice of not hiring women as food servers resulted in the commensurate reputation"; that is, women did not bother applying, so the applicant data reflected the discriminatory practices at issue. Hence, we analyze the hiring data using the binomial model with $\pi = .319$. That CUSUM chart is given in Figure 4 and the test results in Table 6. All of the tests are extremely significant.

To assess the sensitivity of our inference to the choice of π, we found the lowest value of π for which ST_1 was significant, a technique approved of in *Capaci v. Katz & Besthoff* (1983). This was accomplished by testing various values of π until a borderline significant result was obtained. For the entire period, if $\pi \geq .065$, the one-sided p value of

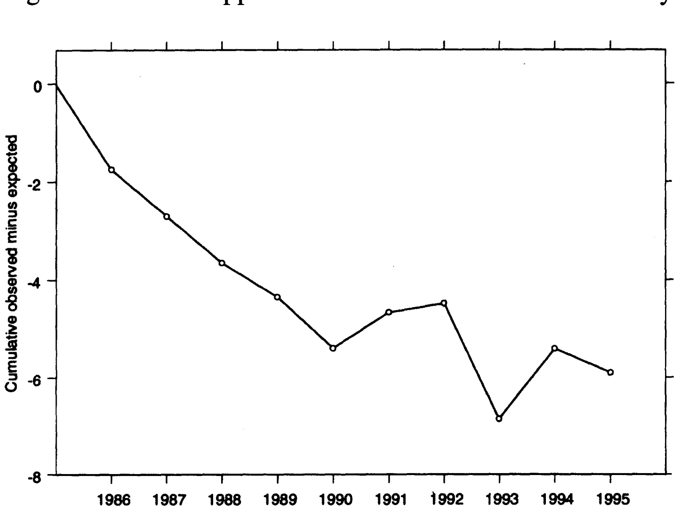

Figure 3. CUSUM Chart for the Hiring Data from EEOC v. Joe's Stone Crab, Inc. Using the Hypergeometric Model.

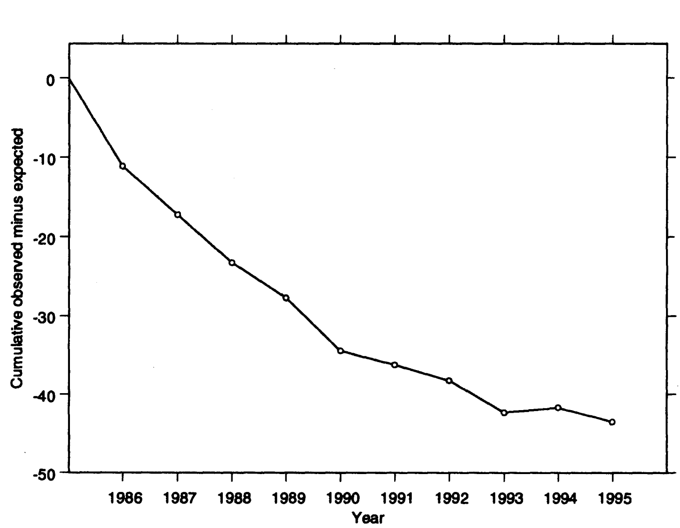

Figure 4. CUSUM Chart for the Hiring Data from EEOC v. Joe's Stone Crab, Inc. Using the Binomial Model With the Female Availability = .319.

ST_1 is $\leq .025$. If one focuses on the period, 1986–1991, when no female was hired, then if $\pi \geq .035$ the one-sided p value of ST_1 is $\leq .025$.

The reason for the apparent discrepancy between the hypergeometric and binomial analyses is that the defendant did not recruit females in proportion to their availability. To formally test whether females were represented in the applicant data in accordance to their share of the relevant labor force, we analyzed the applicant data in Table 5 using the binomial model ($\pi = .319$). All the tests gave highly significant results, and the CUSUM chart never leveled off.

4. DISCUSSION

This article presents CUSUM-based tests having desirable power characteristics under the alternative of interest in hiring cases. In conjunction with the CUSUM charts, the procedures illuminate the basic pattern of the hiring data. They enable the fact finder to examine the hiring process over time and to determine the time period when the pattern remained the same. This information can then be integrated with legally relevant time frames. If the original complaint was filed during a period of statistically significant underhiring *before* the change to fair hiring occurred, then the data are consistent with the plaintiff's claim. If the change to fair hiring occurred a legally sufficient amount of time (perhaps 180 days or more) *before* the plaintiff complained, then the data indicate that the employer was hiring fairly during the relevant time period.

An advantage of CUSUM methods is that they enable one to detect whether a change occurred and, if so, when it happened. Tests simply pooling all the data miss the changepoint. Thus, postcharge data may mask substantial underhiring in the important period just prior to the charge, or data much earlier than the charge might mask fair hiring in the period nearer and subsequent to the complaint.

ACKNOWLEDGMENTS

We thank Elvira Sisolak of the EEOC for providing the data used in Section 3 and Marc Rosenblum for helpful discussions. J. L. Gastwirth's research was partially supported by a grant from the National Science Foundation.

APPENDIX: TWO-POINT APPROXIMATIONS (UPPER BOUNDS) FOR THE *P VALUES*

Let $S_1, \ldots, S_T$ be statistics (possibly correlated). The minimum statistic is defined as $S_{\min} = \min(S_1, \ldots, S_T)$. Consider the problem of estimating $P(S_{\min} \leq C)$, where $C < 0$. Define $E_k = \{S_k \leq C\}$ and $\bar{E}_k = \{S_k > C\}$. Then $P(S_{\min} \leq C)$ can be expressed as

$$P(S_{\min} \leq C) = P(E_1) + P(\bar{E}_1 E_2) + P(\bar{E}_1 \bar{E}_2 E_3) + \cdots + P(\bar{E}_1 \ldots \bar{E}_{T-2} \bar{E}_{T-1} E_T). \quad \text{(A.1)}$$

Commonly used Bonferonni bound is based on

$$P(S_{\min} \leq C) \leq P(E_1) + P(E_2) + \cdots + P(E_k) + \cdots + P(E_T). \quad \text{(A.2)}$$

It ignores the correlation structure of the statistics $S_1, \ldots, S_T$. Because $\bar{E}_1 \bar{E}_2 \ldots \bar{E}_{k-1} E_k \subset \bar{E}_{k-1} E_k \subset E_k$, by replacing $P(\bar{E}_1 \bar{E}_2 \ldots \bar{E}_{k-1} E_k)$ with $P(\bar{E}_{k-1} E_k)$ instead of $P(E_k)$, an improved "two-point" upper bound is obtained (Worsley 1985; Efron 1997):

$$P(S_{\min} \leq C) \leq P(E_1) + P(\bar{E}_1 E_2) + \cdots + P(\bar{E}_{k-1} E_k) + \cdots + P(\bar{E}_{T-1} E_T). \quad \text{(A.3)}$$

The advantage of (A.3) over (A.2) is especially apparent when statistics S_k are strongly autocorrelated; that is, $\mathrm{corr}(S_{k-1}, S_k)$ is high. This is to be expected in the case of the proposed CUSUM-based tests.

Let us assume that x_i for $i = 1, \ldots, T$ are binomial random variable with parameters (n_i, π). The following notation is used in the derivations:

- $N_k = \sum_{i=1}^k n_i$, $N_k^* = \sum_{i=k+1}^T n_i$, $y_k = \sum_{j=1}^k x_j$, and $y_k^* = \sum_{j=k+1}^T x_j$.
- $B(n, u)$ denotes the probability that a binomial random variable with parameters (n, π) is less than or equal to u.
- $B_p(n, u)$ denotes the probability that a binomial random variable with parameters (n, π) is equal to u.
- $H(n_1, n_2, m, u)$ denotes the probability that a hypergeometric random variable with parameters (n_1, n_2, m) is less than or equal to u.

1. For ST_1,

$$S_k = \frac{\sum_{i=1}^k (x_i - n_i \pi)}{\sqrt{N\pi(1-\pi)}},$$

$$\begin{aligned} P(E_1) &= P(S_1 \leq C) = P[x_1 \leq C\sqrt{N\pi(1-\pi)} + n_1 \pi] \\ &= B(n_1, (C\sqrt{N\pi(1-\pi)} + n_1 \pi)), \end{aligned}$$

$$\begin{aligned} &P(\bar{E}_{k-1} E_k) \\ &= P(S_{k-1} > C, S_k \leq C) \\ &= \sum_{b \in \{b > C\sqrt{N\pi(1-\pi)} + N_{k-1}\pi\}} \\ &\quad \times P[y_{k-1} = b, y_k \leq (C\sqrt{N\pi(1-\pi)} + N_k \pi)] \\ &= \sum_{b \in \{b > C\sqrt{N\pi(1-\pi)} + N_{k-1}\pi\}} \\ &\quad \times P[y_k \leq (C\sqrt{N\pi(1-\pi)} + N_k \pi)/y_{k-1} = b] \\ &\quad \times P[y_{k-1} = b] \\ &= \sum_{b \in \{b > C\sqrt{N\pi(1-\pi)} + N_{k-1}\pi\}} \\ &\quad \times P[x_k \leq (C\sqrt{N\pi(1-\pi)} + N_k \pi - b)/y_{k-1} = b] \\ &\quad \times P[y_{k-1} = b] \end{aligned}$$

$$= \sum_{b \in \{b > C\sqrt{N\pi(1-\pi)} + N_{k-1}\pi\}} \times B(n_k, (C\sqrt{N\pi(1-\pi)} + N_k\pi - b)) B_p(N_{k-1}, b).$$

2. For ST_2,

$$S_k = \frac{\sum_{i=1}^{k}(x_i - n_i\pi) - \sum_{i=k+1}^{T}(x_i - n_i\pi)}{\sqrt{N\pi(1-\pi)}}$$

$$\begin{aligned}
P(E_1) &= P(S_1 \leq C) \\
&= P[x_1 - y_1^* \leq (C\sqrt{N\pi(1-\pi)} + (n_1 - N_1^*)\pi)] \\
&= \sum_{b_1, b_2 \in \{b_1 - b_2 \leq (C\sqrt{N\pi(1-\pi)} + (n_1 - N_1^*)\pi)\}} \\
&\quad \times B_p(n_1, b_1) B_p(N_1^*, b_2),
\end{aligned}$$

$$\begin{aligned}
&P(\bar{E}_{k-1}E_k) \\
&= P(S_{k-1} > C, S_k \leq C) \\
&= \sum_{b_1, b_2 \in \{b_1 - b_2 > C\sqrt{N\pi(1-\pi)} + (N_{k-1} - N_{k-1}^*)\pi\}} \\
&\quad \times P[y_{k-1} = b_1, y_{k-1}^* = b_2, y_k - y_k^* \\
&\qquad \leq (C\sqrt{N\pi(1-\pi)} + (N_k - N_k^*)\pi)] \\
&= \sum_{b_1, b_2 \in \{b_1 - b_2 > C\sqrt{N\pi(1-\pi)} + (N_{k-1} - N_{k-1}^*)\pi\}} \\
&\quad \times P[y_k - y_k^* \leq (C\sqrt{N\pi(1-\pi)} \\
&\qquad + (N_k - N_k^*)\pi)/y_{k-1} = b_1, y_{k-1}^* = b_2] \\
&\quad \times P[y_{k-1} = b_1, y_{k-1}^* = b_2] \\
&= \sum_{b_1, b_2 \in \{b_1 - b_2 > C\sqrt{N\pi(1-\pi)} + (N_{k-1} - N_{k-1}^*)\pi\}} \\
&\quad \times P[2x_k \leq (C\sqrt{N\pi(1-\pi)} + (N_k - N_k^*)\pi \\
&\qquad - (b_1 - b_2))/y_{k-1} = b_1, y_{k-1}^* = b_2] \\
&\quad \times P[y_{k-1} = b_1, y_{k-1}^* = b_2] \\
&= \sum_{b_1, b_2 \in \{b_1 - b_2 > C\sqrt{N\pi(1-\pi)} + (N_{k-1} - N_{k-1}^*)\pi\}} \\
&\quad \times P[x_k \leq (1/2)(C\sqrt{N\pi(1-\pi)} \\
&\qquad + (N_k - N_k^*)\pi - (b_1 - b_2))/y_{k-1}^* = b_2] \\
&\quad \times P[y_{k-1} = b_1, y_{k-1}^* = b_2] \\
&= \sum_{b_1, b_2 \in \{b_1 - b_2 > C\sqrt{N\pi(1-\pi)} + (N_{k-1} - N_{k-1}^*)\pi\}} \\
&\quad \times H(n_k, N_k^*, b_2, 1/2(C\sqrt{N\pi(1-\pi)} \\
&\qquad + (N_k - N_k^*)\pi - (b_1 - b_2))) \\
&\quad \times B_p(N_{k-1}, b_1) B_p(N_{k-1}^*, b_2).
\end{aligned}$$

3. For $ST_1 + ST_2$, we can write

$$S_k = \frac{2\sum_{i=1}^{k}(x_i - n_i\pi) - \sum_{i=k+1}^{T}(x_i - n_i\pi)}{\sqrt{N\pi(1-\pi)}}.$$

Thus the approximation of the significance level can be done in a way similar to ST_2.

[Received August 1998. Revised September 1999.]

REFERENCES

Agresti, A. (1990), *Categorical Data Analysis*, New York: Wiley.

Andrews, D. W. K., and Ploberger, W. (1994), "Optimal Tests When a Nuisance Parameter Is Present Only Under the Alternative," *Econometrica*, 62, 1383–1414.

Baldus, D. C., and Cole, J. W. L. (1980), *Statistical Proof of Discrimination*, Colorado Springs: Shepard's/McGraw-Hill.

Banerjee, A. A., Lumsdaine, R. L., and Stock, J. H. (1992), "Recursive and Sequential Tests of the Unit-root and Trend-break Hypotheses: Theory and International Evidence," *Journal of Business & Economic Statistics*, 10, 271–287.

Becker, G. S. (1971), *The Economics of Discrimination*, Chicago: University of Chicago Press.

Bhattacharya, P. K. (1987), "Maximum Likelihood Estimation of a Change-point in the Distribution of Independent Random Variables: General Multiparameter Case," *Journal of Multivariate Analysis*, 23, 183–208.

——— (1994), "Some Aspects of Change-point Analysis," *IMS Lecture Notes*, 23, 28–56.

Blau, F. D., Ferber, M. A., and Winkler, A. E. (1997), *The Economics of Women, Men, and Work*. Englewood Cliffs, NJ: Prentice Hall.

Brostrom, G. (1997), "A Martingale Approach to the Changepoint Problem," *Journal of the American Statistical Association*, 92, 1177–1183.

Capaci v. Katz & Besthoff, Inc., 525 F. Supp. 317 (E.D. La. 1981) *aff'd in part*, and *reversed in part*, 711 F.2d 647 (5th Cir. 1983).

Chernoff, H., and Zacks, S. (1964), "Estimating the Current Mean of a Normal Distribution Which Is Subject to Changes in Time," *The Annals of Mathematical Statistics*, 35, 999–1028.

Chu, C. J., and White, H. (1992), "A Direct Test for Changing Trend," *Journal of Business & Economic Statistics*, 10, 289–299.

Csorgo, M., and Horvath, L. (1988), "Nonparametric Methods for Changepoint Problems," in *Handbook of Statistics* (Vol. 7), eds. P. R. Krishnaiah and C. R. Rao, Amsterdam: North-Holland, pp. 403–425.

EEOC v. Joe's Stone Crab, Inc., 74 FEP Cases 491 (S.D. Fla. 1997).

Efron, B. (1997), "The Length Heuristic for Simultaneous Hypothesis Tests," *Biometrika*, 84, 143–157.

Finkelstein, M. O., and Levin, B. (1990), *Statistics for Lawyers*, New York: Springer-Verlag.

Freidlin, B. (1998), "Statistical Methods Common to Legal and Medical Applications," unpublished Ph. D. thesis, George Washington University, Dept. of Statistics.

Freidlin, B., and Gastwirth, J. L. (1998), "The Application of Change Point Tests to Data Occurring in Fair Hiring Cases," in *Asymptotic Methods in Probability and Statistics (ICAMPS'97)*, ed. B. Szyszkowicz, Amsterdam: Elsevier Science, pp. 551–562.

Gail, M., and Simon, R. (1985), "Testing for Qualitative Interactions Between Treatment Effects and Patient Subsets," *Biometrics*, 41, 361–372.

Gastwirth, J. L. (1988), *Statistical Reasoning in Law and Public Policy*, Orlando: Academic Press.

——— (1996), "Statistical Issues Arising in Equal Employment Litigation," *Jurimetrics*, 36, 353–370.

——— (1997), "Statistical Evidence in Discrimination Cases," *Journal of the Royal Statistical Society*, Ser. A, 160, 289–303.

Gay v. Waiters, 489 F. Supp. 282 (N.D. Cal. 1980) *aff'd* 694 F.2d 531 (9th Cir. 1982).

Goldin, C. (1992), *Understanding the Gender Gap: An Economic History of American Women*, New York: Oxford University Press.

Hinkley, D., and Hinkley, E. A. (1970), "Inference About the Change-Point in a Sequence of Binomial Random Variables," *Biometrika*, 57, 477–488.

James, B., James, K. L., and Siegmund, D. (1987), "Tests for a Change-point," *Biometrika*, 74, 71–83.

Jandhyala, V. K., and MacNeill, I. B. (1992), "On Testing for the Constancy of Regression Coefficients Under Random Walk and Change Point Alternatives," *Econometric Theory*, 8, 501–517.

Levin, B., and Kline, J. (1985), "The Cusum Test of Homogeneity With an Application in Spontaneous Abortion Epidemiology," *Statistics in Medicine*, 4, 469–488.

Lombard, F. (1987), "Rank Test for Change-point Problems," *Biometrika*, 74, 615–624.

Miller, R. G., and Siegmund, D. (1982), "Maximally Selected Chi Square Statistics," *Biometrics*, 38, 1011–1016.

O'Brien v. Sky Chefs, Inc., 670 F.2d 864, 867 (9th Cir. 1982).

Paetzold, R. L., and Willborn, S. L. (1994), *The Statistics of Discrimination*, Colorado Springs: Shepard's/McGraw-Hill.

Page, E. S. (1954), "Continuous Inspection Schemes," *Biometrika*, 41, 100–114.

Pettitt, A. N. (1979), "A Non-parametric Approach to Change-point Problem," *Applied Statistics*, 18, 126–135.

——— (1980), "A Simple Cumulative Type Statistic for the Change Point Problem With Zero–One Observations," *Biometrics*, 67, 70–84.

Ploberger, W., and Kramer, W. (1990), "The Local Power of Cusum and Cusum of Squares Tests," *Econometric Theory*, 6, 335–347.

Siegmund, D. (1986), "Boundary Crossing Probabilities and Statistical Applications," *The Annals of Statistics*, 14, 361–404.

Valentino v. USPS, 511 F. Supp. 917 (C.D.C. 1981) *aff'd* 674 F.2d 56 (D.C. Cir. 1982).

Vogelsang, T. J. (1997), "Wald-Type Test for Detecting Shifts in Trend Function of a Dynamic Time Series," *Econometric Theory*, 13, 818–849.

Worsley, K. J. (1983), "The Power of the Likelihood Ratio and Cumulative Sum Test for a Change in Binomial Probability," *Biometrika*, 70, 455–467.

——— (1985), "Bonferroni (Improved) Wins Again," *The American Statistician*, 39, 235.

LINKED CITATIONS

You have printed the following article:

Changepoint Tests Designed for the Analysis of Hiring Data Arising in Employment Discrimination Cases
Boris Freidlin; Joseph L. Gastwirth
Journal of Business & Economic Statistics, Vol. 18, No. 3. (Jul., 2000), pp. 315-322.
Stable URL:
http://links.jstor.org/sici?sici=0735-0015%28200007%2918%3A3%3C315%3ACTDFTA%3E2.0.CO%3B2-E

This article references the following linked citations. If you are trying to access articles from an off-campus location, you may be required to first logon via your library web site to access JSTOR. Please visit your library's website or contact a librarian to learn about options for remote access to JSTOR.

References

Optimal Tests when a Nuisance Parameter is Present Only Under the Alternative
Donald W. K. Andrews; Werner Ploberger
Econometrica, Vol. 62, No. 6. (Nov., 1994), pp. 1383-1414.
Stable URL:
http://links.jstor.org/sici?sici=0012-9682%28199411%2962%3A6%3C1383%3AOTWANP%3E2.0.CO%3B2-7

Recursive and Sequential Tests of the Unit-Root and Trend-Break Hypotheses: Theory and International Evidence
Anindya Banerjee; Robin L. Lumsdaine; James H. Stock
Journal of Business & Economic Statistics, Vol. 10, No. 3. (Jul., 1992), pp. 271-287.
Stable URL:
http://links.jstor.org/sici?sici=0735-0015%28199207%2910%3A3%3C271%3ARASTOT%3E2.0.CO%3B2-C

A Martingale Approach to the Changepoint Problem
Goran Brostrom
Journal of the American Statistical Association, Vol. 92, No. 439. (Sep., 1997), pp. 1177-1183.
Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28199709%2992%3A439%3C1177%3AAMATTC%3E2.0.CO%3B2-N

LINKED CITATIONS

Estimating the Current Mean of a Normal Distribution which is Subjected to Changes in Time
H. Chernoff; S. Zacks
The Annals of Mathematical Statistics, Vol. 35, No. 3. (Sep., 1964), pp. 999-1018.
Stable URL:
http://links.jstor.org/sici?sici=0003-4851%28196409%2935%3A3%3C999%3AETCMOA%3E2.0.CO%3B2-N

A Direct Test for Changing Trend
Chia-Shang James Chu; Halbert White
Journal of Business & Economic Statistics, Vol. 10, No. 3. (Jul., 1992), pp. 289-299.
Stable URL:
http://links.jstor.org/sici?sici=0735-0015%28199207%2910%3A3%3C289%3AADTFCT%3E2.0.CO%3B2-P

The Length Heuristic for Simultaneous Hypothesis Tests
Bradley Efron
Biometrika, Vol. 84, No. 1. (Mar., 1997), pp. 143-157.
Stable URL:
http://links.jstor.org/sici?sici=0006-3444%28199703%2984%3A1%3C143%3ATLHFSH%3E2.0.CO%3B2-7

Testing for Qualitative Interactions between Treatment Effects and Patient Subsets
M. Gail; R. Simon
Biometrics, Vol. 41, No. 2. (Jun., 1985), pp. 361-372.
Stable URL:
http://links.jstor.org/sici?sici=0006-341X%28198506%2941%3A2%3C361%3ATFQIBT%3E2.0.CO%3B2-B

Statistical Evidence in Discrimination Cases
Joseph L. Gastwirth
Journal of the Royal Statistical Society. Series A (Statistics in Society), Vol. 160, No. 2. (1997), pp. 289-303.
Stable URL:
http://links.jstor.org/sici?sici=0964-1998%281997%29160%3A2%3C289%3ASEIDC%3E2.0.CO%3B2-N

Inference About the Change-Point in a Sequence of Binomial Variables
David V. Hinkley; Elizabeth A. Hinkley
Biometrika, Vol. 57, No. 3. (Dec., 1970), pp. 477-488.
Stable URL:
http://links.jstor.org/sici?sici=0006-3444%28197012%2957%3A3%3C477%3AIATCIA%3E2.0.CO%3B2-I

LINKED CITATIONS

Tests for a Change-Point
Barry James; Kang Ling James; David Siegmund
Biometrika, Vol. 74, No. 1. (Mar., 1987), pp. 71-83.
Stable URL:
http://links.jstor.org/sici?sici=0006-3444%28198703%2974%3A1%3C71%3ATFAC%3E2.0.CO%3B2-2

Rank Tests for Changepoint Problems
F. Lombard
Biometrika, Vol. 74, No. 3. (Sep., 1987), pp. 615-624.
Stable URL:
http://links.jstor.org/sici?sici=0006-3444%28198709%2974%3A3%3C615%3ARTFCP%3E2.0.CO%3B2-2

Maximally Selected Chi Square Statistics
Rupert Miller; David Siegmund
Biometrics, Vol. 38, No. 4. (Dec., 1982), pp. 1011-1016.
Stable URL:
http://links.jstor.org/sici?sici=0006-341X%28198212%2938%3A4%3C1011%3AMSCSS%3E2.0.CO%3B2-D

Continuous Inspection Schemes
E. S. Page
Biometrika, Vol. 41, No. 1/2. (Jun., 1954), pp. 100-115.
Stable URL:
http://links.jstor.org/sici?sici=0006-3444%28195406%2941%3A1%2F2%3C100%3ACIS%3E2.0.CO%3B2-H

A Non-Parametric Approach to the Change-Point Problem
A. N. Pettitt
Applied Statistics, Vol. 28, No. 2. (1979), pp. 126-135.
Stable URL:
http://links.jstor.org/sici?sici=0035-9254%281979%2928%3A2%3C126%3AANATTC%3E2.0.CO%3B2-1

Boundary Crossing Probabilities and Statistical Applications
David Siegmund
The Annals of Statistics, Vol. 14, No. 2. (Jun., 1986), pp. 361-404.
Stable URL:
http://links.jstor.org/sici?sici=0090-5364%28198606%2914%3A2%3C361%3ABCPASA%3E2.0.CO%3B2-D

The Power of Likelihood Ratio and Cumulative Sum Tests for a Change in a Binomial Probability
K. J. Worsley
Biometrika, Vol. 70, No. 2. (Aug., 1983), pp. 455-464.
Stable URL:
http://links.jstor.org/sici?sici=0006-3444%28198308%2970%3A2%3C455%3ATPOLRA%3E2.0.CO%3B2-P

Letters to the Editor
Keith J. Worsley; Steven J. Schwager; J. Aleong; Ed Carlstein; Daniel Krewski; Mikelis Bickis; A. Arcese; H. Joseph Newton; A. Ian McLeod; Carlos Jimenez; Steven D. Stellman; William S. Cleveland; A. I. Dale; Adi Raveh; Norman L. Johnson; Walter W. Piegorsch; George Casella; R. W. Farebrother; S. R. Searle; Spyros Missiakoulis; Robin Darton; Oscar Kempthorne; Craig F. Ansley
The American Statistician, Vol. 39, No. 3. (Aug., 1985), pp. 235-241.
Stable URL:
http://links.jstor.org/sici?sici=0003-1305%28198508%2939%3A3%3C235%3ALTTE%3E2.0.CO%3B2-E