Appl. Statist. (1979),
28, *No.* 2, *pp.* 126–135

A Non-parametric Approach to the Change-point Problem

By A. N. Pettitt

University of Technology, Loughborough, Leics., England

[Received March 1978. Final revision November 1978]

SUMMARY

Non-parametric techniques are introduced for the change-point problem. Exact and approximate results are obtained for testing the null hypothesis of no change. The methods are illustrated by the analysis of three sets of data illustrating the techniques for zero–one observations, Binomial observations and continuous observations. Some comparisons are made with methods based on CUSUMS.

Keywords: CHANGE-POINT PROBLEM; CUSUMS; KOLMOGOROV–SMIRNOV D STATISTIC; MANN–WHITNEY U STATISTIC; NON-PARAMETRIC TESTS

1. Introduction

CONSIDER a sequence of random variables $X_1, X_2, \ldots, X_T$ then the sequence is said to have a change-point at τ if X_t for $t = 1, \ldots, \tau$ have a common distribution function $F_1(x)$ and X_t for $t = \tau+1, \ldots, T$ have a common distribution function $F_2(x)$, and $F_1(x) \neq F_2(x)$. We consider the problem of testing the null hypothesis of "no-change", H: $\tau = T$, against the alternative of "change", A: $1 \leqslant \tau < T$, using a non-parametric statistic. We make no assumptions about the functional forms of F_1 and F_2 except that they are continuous, and then we relax this condition to allow the X_t's to be Bernoulli and Binomial variables. In the paper we describe how a version of the Mann–Whitney statistic can be utilized for the problem in question and derive approximate significance probabilities for testing "no-change" against "change". We also give exact results for the case where F_1 and F_2 refer to Bernoulli distributions. We give three numerical examples of how the technique can be used extending the theory to Binomial random variables. Estimation of the change-point can be made by simple use of the test statistics.

The change-point problem has been considered before by various authors. Page (1954, 1955, 1957) considered the problem by introducing cumulative sums (CUSUMS). Sen and Srivastava (1975a, b) consider tests for a change in mean level assuming a normal model. Hinkley (1970), among other things, derives likelihood ratio tests for a specified value of τ (not equal to T) and estimation of τ. Smith (1975) considers a Bayesian approach to making inferences about the change-point. McGilchrist and Woodyer (1975) consider a distribution-free CUSUM and Sen and Srivastava (1975a) suggest distribution-free techniques. Many of these methods consider the initial distribution to be known, but our methods make no such assumptions.

In the main, except for the last two papers, the problem of testing for no change against change when the initial distribution is unknown, is not tackled. The work of McGilchrist and Woodyer (1975) becomes a special case of our Section 2.2 and that of Sen and Srivastava (1975a), who consider the problem using maximum likelihood statistics, find the problem intractable and use Monte Carlo methods. Smith (1975) derives posterior distributions for τ which can be used to estimate τ and test hypotheses concerning values of τ, including $\tau = T$.

Copyright © 2001. All Rights Reserved.

2. Non-parametric Approach

2.1. *Introduction*

An appealing non-parametric test to detect a change would be to use a version of the Mann–Whitney two-sample test. Let

$$D_{ij} = \operatorname{sgn}(X_i - X_j)$$

where $\operatorname{sgn}(x) = 1$ if $x > 0$, 0 if $x = 0$, -1 if $x < 0$, then consider

$$U_{t,T} = \sum_{i=1}^{t} \sum_{j=t+1}^{T} D_{ij}. \tag{2.1}$$

The statistic $U_{t,T}$ is equivalent to a Mann–Whitney statistic for testing that the two samples $X_1, \ldots, X_t$ and $X_{t+1}, \ldots, X_T$ come from the same population. The statistic $U_{t,T}$ is then considered for values of t with $1 \leqslant t < T$. We propose for the test of H: no change against A: change, the use of the statistic

$$K_T = \max_{1 \leqslant t < T} |U_{t,T}| \tag{2.2}$$

and for changes in one direction, the statistics

$$K_T^+ = \max_{1 \leqslant t < T} U_{t,T}, \tag{2.3}$$

$$K_T^- = -\min_{1 \leqslant t < T} U_{t,T}. \tag{2.4}$$

Obviously $K_T = \max(K_T^+, K_T^-)$. It should be noted that, on the null hypothesis H, $E(D_{ij}) = 0$ and the distribution of $U_{t,T}$ is symmetric about zero for each t. Thus K_T^+ and K_T^- have the same null distributions.

The statistics K_T^+ and K_T^- are one-sided, and using the theory of the Mann–Whitney statistic one would expect to find K_T^+ large if there has been a shift down in level from the beginning of the series. Using the notation of Section 1, K_T^+ would be large if $F_1(x) \leqslant F_2(x)$, with the inequality strict for at least some x. Similarly, we would expect K_T^- to be large if there had been a shift up in level or $F_1(x) \geqslant F_2(x)$.

The statistic K_T has not been considered before but Sen and Srivastava (1975a, Section 4) briefly reported some Monte Carlo power results of a standardized version of K_T, which was essentially $\max|U_{t,T}/(Tt - t^2)^{\frac{1}{2}}|$.

2.2. *An Exact Result for Bernoulli Random Variables*

Let us suppose the X_t are independent Bernoulli random variables taking values 0 or 1 only with unknown probability, θ, of obtaining the value 1. This is a useful initial approach to data analysis since any observations can be reduced to Bernoulli observations by introducing some dichotomy. In this case, after some algebra, $U_{t,T}$ is simplified to give

$$U_{t,T} = T(S_t - tS_T/T), \tag{2.5}$$

where $S_t = \Sigma X_j$ and $S_T = \Sigma X_j$, summing over j from 1 to t and 1 to T, respectively. On the null hypothesis of no-change, H, we have a sequence of identical independent Bernoulli random variables so that S_T is sufficient for θ. On the alternative, where some change has taken place, then no sufficient statistics exist, except the complete sample. If we make inferences conditional on knowing S_T, then it is possible to find the null distributions of K_T, K_T^+ and K_T^-.

Conditional on knowing $S_T = m$, say, consider two independent random samples $Y_1, \ldots, Y_m$ and $Y_1^*, \ldots, Y_n^*$ and their sample distribution functions $F_{1m}(y)$ and $F_{2n}(y)$ where

Copyright © 2001. All Rights Reserved.

$mF_{1m}(y)$ ($nF_{2n}(y)$) is the number of the Y_j's (Y_j^*'s) less than or equal to y. Let $(Z_1, Z_2, \ldots, Z_{n+m})$ be the indicator random variables for the combined sample of Y's and Y^*'s so that

$$Z_j = \begin{cases} 1 & \text{if the } j\text{th smallest in the combined sample is a } Y_j, \\ 0 & \text{if the } j\text{th smallest in the combined sample is a } Y_j^*. \end{cases}$$

Then at the jth smallest observation in the combined sample

$$F_{1m}(y) = \sum_{i=1}^{j} Z_i/m, \quad F_{1n}(y) = \sum_{i=1}^{j} (1-Z_i)/n,$$

so that

$$F_{1m}(y) - F_{1n}(y) = \frac{m+n}{m.n}\left(\sum_{i=1}^{j} Z_i - \frac{im}{m+n}\right), \tag{2.6}$$

after some algebra.

Now writing $t = j$, $T = m+n$, $S_T = m$, then

$$\frac{U_{j,T}}{m.n} = \frac{(m+n)}{m.n}\left(S_j - \frac{jm}{m+n}\right).$$

We see the connection between $U_{j,T}$ and (2.5). Now conditional on knowing $S_T = m$, the $(X_1, \ldots, X_T)$ have the same distribution as the $(Z_1, \ldots, Z_{n+m})$ with $T = n+m$, when the Y_j's and Y_j^*'s are continuous and come from the same population. So that, for example, $(K_T/m.n)$, conditional on $S_T = m$, $(n = T-m)$, has the same distribution as

$$\max_{1 \leqslant j < m+n} \left| \frac{m+n}{m.n}\left(\sum_{i=1}^{j} Z_i - \frac{jm}{m+n}\right) \right|$$

or

$$\sup_{-\infty < y < \infty} |F_{1m}(y) - F_{1n}(y)|.$$

But this latter statistic is the Kolmogorov–Smirnov two-sample statistic $D_{n,m}$, for which the distribution has been extensively tabulated under the assumption that the Y_j's and Y_j^*'s are continuous and come from the same population. Similarly, the conditional null distributions of K_T^+/mn and K_T^-/mn are the same as D_{mn}^+ and D_{mn}^-, the one-sided two-sample Kolmogorov–Smirnov statistics. Extensive tables of the distribution of $D_{m,n}$ for $m, n \leqslant 25$ are given in Pearson and Hartley (1972, Table 55) and tables of $D_{m,n}^+$ are given in Gail and Green (1976) for $m, n \leqslant 30$.

A similar idea to that given here has been discussed by McGilchrist and Woodyer (1975) who consider continuous observations which are reduced to Bernoulli observations by considering whether a particular observation is greater or less than the sample median. In their case they have $S_T = [\frac{1}{2}(T+1)]$ always. Our development is somewhat more general.

For approximate significance probabilities associated with K_T, etc. we can use the well-known approximations to the distributions of $D_{m,n}$, etc. If we have

$$\Pr(mnD_{m,n}^+ \geqslant c_\alpha) = \alpha,$$

then, using the asymptotic distribution, we have that c_α is approximated by

$$c_\alpha^* = \{nm(n+m)(-\ln\alpha)/2\}^{\frac{1}{2}}.$$

Equivalently the approximate significance probability p_{OA} associated with the value k^+ of K_T^+ is given by

$$p_{OA} = \exp[-2(k^+)^2\{S_T(T^2 - TS_T)\}^{-1}]. \tag{2.7}$$

Copyright © 2001. All Rights Reserved.

This approximation is conservative in that $p_{OA} \geqslant p$, where p is the exact significance probability; see Gail and Green (1976), for example.

2.3. *Approximations for Continuous Observations*

Since the theory of ranks is well known it is easier to re-write $U_{t,T}$ as a rank statistic. Suppose $R_1, \ldots, R_t$ are the ranks of the t observations $X_1, \ldots, X_t$ in the complete sample of T observations, then

$$U_{t,T} = 2W_t - t(T+1),$$

where $W_t = \Sigma R_j$, summing over j from 1 to t.

Now from standard rank theory, on the null hypothesis of no change, we have

$$E(W_t) = \frac{t(T+1)}{2}, \quad \operatorname{var}(W_t) = \frac{t(T-t)(T+1)}{12};$$

for $t \leqslant s$, we can show

$$\operatorname{cov}(W_t, W_s) = \frac{t(T+1)(T-s)}{12}.$$

When there are ties in the data, then the usual correction to the variance of W_t can be made by multiplying it by the factor

$$1 - \sum_{j=1}^{r} q_j(q_j^2 - 1) T^{-1}(T^2-1)^{-1}, \tag{2.8}$$

where r is the distinct number of values in the T observations and the jth value occurs with frequency q_j ($\Sigma q_j = T$, summing over j from 1 to r).

Now W_t is asymptotically normally distributed, so putting $x = t/T$, we see that

$$y_T(x) = T^{-1}\{3/(T+1)\}^{\frac{1}{2}} U_{t,T}$$

has a limiting distribution as $T \to \infty$ equal to that of the Brownian Bridge, $y(x)$, say (see Billingsley, 1968, p. 64, for example). The limiting distribution of $T^{-1}\{3/(T+1)\}^{\frac{1}{2}} K_T$ is the same as $\sup|y(x)|$, which is well known to be given by

$$\Pr(\sup|y(x)| \leqslant a) = 1 + 2\sum_{r=1}^{\infty}(-1)^r \exp(-2r^2a^2). \tag{2.9}$$

This is the limiting distribution of the Kolmogorov–Smirnov goodness of fit statistic $\sqrt{(n)}\, D_n$. The function (2.9) is tabulated by Smirnov (1948) for various values of a. The limiting distribution of $T^{-1}\{3/(T+1)\}^{\frac{1}{2}} K_T^+$ is similarly given by $\sup y(x)$, where

$$\Pr(\sup y(x) \leqslant a) = 1 - \exp(-2a^2).$$

Thus the significance probabilities associated with the values k^+ (or k^-) of K_T^+ (or K_T^-) and the value of k of K_T are given approximately by

$$p_{OA} = \exp\{-6(k^+)^2/(T^3+T^2)\} \tag{2.10}$$

for K_T^+, and

$$p_{OA} = 2\sum_{r=1}^{\infty}(-1)^{r+1} \exp\{-6kr^2/(T^3+T^2)\} \tag{2.11}$$

$$\simeq 2\exp\{-6k^2/(T^3+T^2)\} \tag{2.12}$$

for K_T, where the approximation holds good, accurate to two decimal places, for $p_{OA} \leqslant 0{\cdot}5$.

Copyright © 2001. All Rights Reserved.

3. Calculation of Statistics—Application to Page's Data

We illustrate the techniques by applying the methods of Section 2 to the simulated data of Page (1955) given in Table 1. For the 40 observations the hypothesis of no change can be tested. The data were considered in two ways; first as 40 observations (with 5 subtracted from Page's values) from a continuous population (the first 20 are from a normal distribution with mean 0 and variance 1, the second 20 have a change in mean, which equals 1), giving the X_t in row 2 of Table 1 and then, secondly, as Bernoulli observations, setting an observation equal to 1 if greater than 0 and set equal to 0 otherwise, giving the X_t' in row 5 of Table 1.

For *continuous data* the $U_{t,T}$ can be calculated using the formula

$$U_{t,T} = U_{t-1,T} + V_{t,T}$$

for $t = 2, \ldots, T$, where

$$V_{t,T} = \sum_{j=1}^{T} \operatorname{sgn}(X_t - X_j) \tag{3.1}$$

Table 1

Page's data

In rows 1–4 we give the continuous data and the statistics required to calculate K_T in rows 3 and 4. In rows 5 and 6 we give the Bernoulli data and the relevant statistics

t	1	2	3	4	5	6	7	8	9	10
X_t	−1·05	0·96	1·22	0·58	−0·98	−0·03	−1·54	−0·71	−0·35	0·66
$V_{t,T}$ (3.1)	−35	11	20	−1	−31	−17	−39	−27	−25	2
$U_{t,T}$ (3.2)	−35	−24	−4	−5	−36	−53	−92	−119	−144	−142
X_t'	0	1	1	1	0	0	0	0	0	1
$U_{t,T}$ (2.5)	−27	−14	−1	12	−15	−42	−69	−96	−123	−110
t	11	12	13	14	15	16	17	18	19	20
X_t	0·44	0·91	−0·02	−1·42	1·26	−1·02	−0·81	1·66	1·05	0·97
$V_{t,T}$ (3.1)	−8	9	−15	−37	23	−33	−29	32	15	13
$U_{t,T}$ (3.2)	−150	−141	−156	−193	−170	−203	−232	−200	−185	−172
X_t'	1	1	0	0	1	0	0	1	1	1
$U_{t,T}$ (2.5)	−97	−84	−111	−138	−125	−152	−179	−166	−153	−140
t	21	22	23	24	25	26	27	28	29	30
X_t	2·14	1·22	−0·24	1·60	0·72	−0·12	0·44	0·03	0·66	0·56
$V_{t,T}$ (3.1)	37	20	−23	29	5	−21	−8	−13	2	−3
$U_{t,T}$ (3.2)	−135	−115	−138	−109	−104	−125	−133	−146	−144	−147
X_t'	1	1	0	1	1	0	1	1	1	1
$U_{t,T}$ (2.5)	−127	−114	−141	−128	−115	−142	−129	−116	−103	−90
t	31	32	33	34	35	36	37	38	39	40
X_t	1·37	1·66	0·10	0·80	1·29	0·49	−0·07	1·18	3·29	1·84
$V_{t,T}$ (3.1)	27	32	−11	7	25	−5	−19	17	39	35
$U_{t,T}$ (3.2)	−120	−88	−99	−92	−67	−72	−91	−74	−35	0
X_t'	1	1	1	1	1	1	0	1	1	1
$U_{t,T}$ (2.5)	−77	−64	−51	−38	−25	−12	−39	−26	−13	0

and $U_{1,T} = V_{1,T}$. In Table 1, row 3, we give the values of $V_{t,T}$. When there are no ties in the data, $V_{t,T} = T+1-2R_t$, where R_t is the rank of X_t in the sample of T observations. The statistic $U_{t,T}$ is then the cumulative sum of the $V_{t,T}$ and the values are given in Row 4. Note

Copyright © 2001. All Rights Reserved.

$U_{T,T} \equiv 0$ always. For the *continuous data* we find $K_T^+ = 0$, $K_T^- = 232$ and so $K_T = 232$. From Section 2.3 the standardized value of K_T or K_T^- is $T^{-1}\{3/(T+1)\}^{\frac{1}{2}} K_T$ or 1·569 with $T = 40$. The approximate significance probability p_{OA} associated with this value of K_T^- is 0·007, given by (2.10), and that associated with K_T is 0·014, given by (2.12), so indicating strong evidence of a change.

For the *Bernoulli data*, we find $S_T = 27$ and $K_T = 179$, $K_T^+ = 12$ and $K_T^- = 179$. The $U_{t,T}$ are calculated from formula (2.5) or more easily from the relationship

$$U_{t,T} = U_{t-1,T} + \begin{cases} -S_T & \text{if the } t\text{th observation is a zero,} \\ T - S_T & \text{if the } t\text{th observation is a one,} \end{cases}$$

for $t = 1, \ldots, T-1$, with $U_{0,T} = 0$. We know the exact null distribution of $K_T^-(K_T)$ conditional on knowing $S_T = 27$ (see Section 2.2) is given by the distribution of the Kolmogorov–Smirnov statistic $mnD_{mn}^-(mnD_{mn})$ with $m = S_T$, $n = T - S_T$. From the Tables of Gail and Green (1976) we find, with $m = 27$, $n = 13$, $\Pr(mnD_{mn}^- \geqslant 165) \simeq 0{\cdot}01$, but no smaller significance points are given. From Pearson and Hartley (1972, Table 55), we find $\Pr(mnD_{mn} \geqslant 172) \simeq 0{\cdot}01$ with $m = 25$, $n = 13$. The approximate significance probability, p_{OA}, given by (2.7) associated with this value of K_T^- is given by 0·0104. All this suggests the significance probability of the value of K_T^- is about 0·01, so indicating again strong evidence of a change.

For both the continuous and Bernoulli data the maximum value of the statistic was obtained at $t = 17$. These findings should be compared with those of Page (1955), who found significance at the 1 per cent level for a one-sided test of no change against change, analysing the Bernoulli data, and suggested the same value, 17, for the change-point.

There is, however, a difference between the approach of Page (1955) and the approach here, and that is that Page assumed the initial value for the Bernoulli distribution was known (and provided percentage points for it equal to one half), whereas we make no such assumption here.

The results of this section confirm that there are no real differences between the methods of analysis when the data are well behaved. In Section 5, we give an example which shows the sensitivity (to "wild" observations) of Page's CUSUM scheme when applied to continuous data.

4. Lindisfarne Scribes' Data: Binomial Data

The Lindisfarne Scribes' data refer to data given originally by Ross (1950) and subsequently analysed by Silvey (1956). The data refer to the number of occurrences of present indicative third person singular endings "-s" and "-ð" for different sections of Lindisfarne. It is believed different scribes used the endings "-s" and "-ð" in different proportions. The data are given in Table 2. Here we illustrate the techniques described in Section 2.2 modified for the Binomial data now presented in Table 2.

Table 2

Lindisfarne Scribes' data (the $U_{t_i,T}$ were calculated using (4.2))

Section, i	1	2	3	4	5	6	7	8	9
Number of "-s"	12	26	31	17	7	28	34	10	29
Number of "-ð"	9	10	13	4	2	24	11	1	8
t_i	21	57	101	122	131	183	228	239	276
$-U_{t_i,T}$	1,782	2,318	3,334	2,796	2,678	7,906	7,880	7,090	6,584
Section, i	10	11	12	13	14	15	16	17	18
Number of "-s"	30	16	17	24	14	5	17	17	16
Number of "-ð"	9	2	0	7	2	1	3	4	4
t_i	315	333	350	381	397	403	423	444	464
$-U_{t_i,T}$	6,314	5,190	3,552	2,966	2,070	1,850	962	424	0

Copyright © 2001. All Rights Reserved.

Suppose we have a sequence of independent Binomial random variables $Z_1, Z_2, \ldots, Z_N$ where Z_i is $\text{Bi}(n_i, \theta_i)$, $i = 1, \ldots, N$. Then the $Z_1, Z_2, \ldots, Z_N$ could be thought of as arising from a sequence of independent Bernoulli random variables. Let $t_i = \Sigma n_j$, summing over j from 1 to i, with $t_0 = 0$ then the Bernoulli random variables $X_1, X_2, \ldots, X_{t_N}$ where $\Pr(X_j = 1) = \theta_i$ for $t_{i-1}+1 \leqslant j \leqslant t_i$, define the Z_i by

$$Z_i = \sum_{j=t_{i-1}+1}^{t_i}.$$

Thus the Binomial random variables can be thought of as the sequence of Bernoulli random variables only seen at the points $t_1, t_2, \ldots, t_N$ and then the sums ΣX_j are observed, which are over j from $t_{i-1}+1$ to t_i. Bearing this in mind, we can return to Section 2.2 and define the $U_{t,T}$ (2.5) with $T = t_N$ using the X_j's and

$$S_t = \sum_{i=1}^{j} Z_i = \sum_{i=1}^{t_j} X_i$$

where $t = t_j$, $j = 1, \ldots, N$.

Now we consider $\tilde{K}_T$ defined by

$$\tilde{K}_T = \max_{t=t_i:\, i=1,\ldots,N} |U_{t,T}|.$$

Obviously if all the X_j were known then $\tilde{K}_T \leqslant K_T$ where

$$K_T = \max_{1 \leqslant t \leqslant T} |U_{t,T}|.$$

The U_{t_i} can be easily calculated using the following method. Let S_T be the total number of ones, i.e. $S_T = \Sigma Z_i$, summing over i from 1 to N, and define

$$V_{t_i} = Z_i(t_N - S_T) - (n_i - Z_i) S_T = Z_i t_N - n_i S_T; \tag{4.1}$$

then

$$U_{t_i} = U_{t_{i-1}} + V_{t_i}, \quad i = 2, \ldots, N-1, \tag{4.2}$$

with $U_{t_1} = V_{t_1}$ and $U_{t_N} \equiv 0$.

Using this technique the Lindisfarne data were analysed. Associating "-s" with one and "-ð" with zero we found S_T (the total number of "-s" endings) = 350, $t_N = 464$ with $N = 18$. The V_{t_i} can then be calculated easily. For example, using (4.1)

$$V_{t_1} = 12 \times 464 - 21 \times 350 = -1{,}782,$$

taking Z_i to be the number of "-s"s in the ith section and n_i the number of "-s"s and "-ð"s in the ith section. The results are given in Table 2 and the value of $\tilde{K}_T$ was found to be 7,906 and when standardized (using the results of Section 2.2) gave the value 1·83. Using the table of Smirnov (1948) this gives a significance probability of 0·25 per cent. Of course this value is conservative in that the true significance probability is smaller than this. The suggested change point is after the sixth section.

This analysis should be compared with an unpublished analysis by A. F. M. Smith (1977, personal communication), who allowed more than one change-point, and found "overwhelming evidence that two changes have taken place" and that these were (taking into account his groupings) after the sixth section and the seventh section. His analysis follows a somewhat more complicated Bayesian procedure.

5. Some Industrial Data

In Table 3 we present some industrial data which give the percentage of a particular material in 27 batches taken from a given source. The observations were taken in the given order in

Copyright © 2001. All Rights Reserved.

time and it is assumed nothing is known about the observations. It is wished to see whether there has been any change in distribution. In Table 3 we give the results of the analysis as

TABLE 3

Industrial data (Row 3 gives CUSUM *of* $\bar{x}-x_t$*)*

t	1	2	3	4	5	6	7	8	9
X_t	7·1	8·1	8·2	11·1	6·6	4·9	4·0	17.7	6·5
$\bar{X}-X_t$	1·33	1·66	1·89	−0·78	1·05	5·58	9·01	−0·26	1·67
$U_{t,T}$	12	16	15	−7	11	33	59	33	53
t	10	11	12	13	14	15	16	17	18
X_t	4·6	8·8	11·6	6·8	7·5	6·9	8·1	9·3	7·5
$\bar{X}-X_t$	5·70	5·33	2·16	3·79	4·70	6·24	6·57	5·70	5·63
$U_{t,T}$	77	71	47	63	72	86	90	76	85
t	19	20	21	22	23	24	25	26	27
X_t	10·0	8·7	9·1	8·9	9·1	9·6	8·1	9·8	8·2
$\bar{X}-X_t$	4·06	3·79	3·12	2·65	1·98	0·81	1·14	−0·23	0·00
$U_{t,T}$	65	61	50	42	31	15	19	1	0

described in Section 2 and following the methods of Section 3, giving the values of $U_{t,T}$ in row 4. The value of K_T is 90 and $U_{t,T}$ obtains this value at $t = 16$. The approximate one-sided significance probability of this value, using equation (2.9), is 0·092, so indicating a probable change. The correction for ties, (2.8), makes a negligible difference in this case. In Fig. 1 we give a stem-and-leaf plot of the first 16 observations and the remaining 11 observations; Fig. 1 plainly displays the difference in distribution before and after the estimated change-point.

First sixteen observations (before estimated change-point)

4	0 6 9
5	
6	5 6 8 9
7	1 5
8	1 1 2 8
9	
10	
11	1 6
12	
13	
14	
15	
16	
17	7

Remaining eleven observations (after estimated change-point)

7	5
8	1 2 7 9
9	1 1 3 6 8
10	0

FIG. 1. Stem-and-leaf display of the observations before and after $t = 16$.

If we were to use the CUSUM technique then we could consider the CUSUM of the $\bar{X}-X_t$; these values are given in row 3 of Table 3. There is no procedure to test for a change using these values, but the value of t which maximizes $|\bar{X}-\bar{X}_t|\{t(T-t)\}^{-\frac{1}{2}}$, where $\bar{X}_t = t^{-1}\Sigma X_j$, summing over the range 1 to t, is the maximum likelihood estimate of the change-point assuming the normal model and a location change only at the change-point (Hinckley, 1970).

Copyright © 2001. All Rights Reserved.

This value occurs at $t = 7$, which reflects the effect of the "wild" observation, 17·7, at $t = 8$, and is not a satisfactory estimate of the change-point.

Alternatively a CUSUM of the values $X_i - a$, where a is some target value, could be made and the change-point estimated by considering the deviations between the CUSUM, $\Sigma(X_j - a)$ for j from 1 to t, and the straight line between the points (0, 0) and $(T, \Sigma(X_j - a))$ for j from 1 to T, the first and last plotted points. A little algebra shows this is equivalent to considering the differences $t(\bar{X}_t - \bar{X})$, or the CUSUM of the $X_t - \bar{X}$. Techniques based on the CUSUM test for a deviation from the target value, a.

In Fig. 2 we have plotted the CUSUM of the $\bar{X} - X_t$ (row 3 of Table 3) and the $U_{t,T}$. We note that the range of the $U_{t,T}$'s is approximately ten times that of the CUSUM of the $\bar{X} - X_t$ so that multiplying the latter by ten makes the two roughly compatible. The plots in Fig. 2 show the effects of the wild observation 17.7, at $t = 8$, in the CUSUM of $\bar{X} - X_t$, and the smaller effect it has in the $U_{t,T}$. The plot of $U_{t,T}$ is somewhat smoother than the plot of the CUSUM of $\bar{X} - X_t$, which is erratic in the first part of the series.

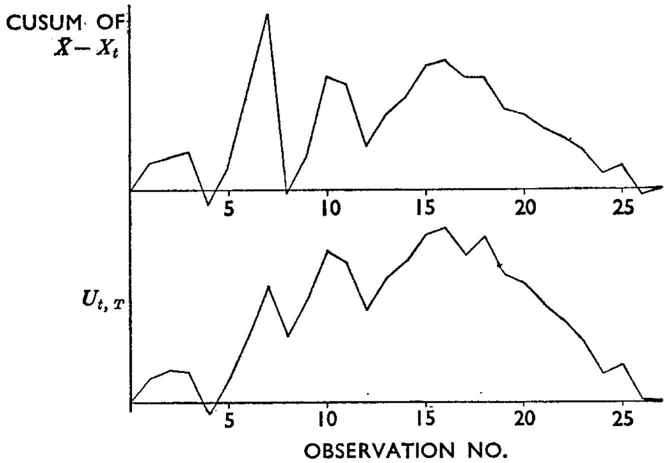

FIG. 2. The upper graph gives the CUSUM of $\bar{X} - X_t$, and the lower graph gives the plot of the $U_{t,T}$ from Table 3. The vertical scale of the lower graph is one-tenth that of the upper graph.

Using the CUSUM of the $\bar{X} - X_t$, we could estimate the change-point as being at $t = 7$ (where the maximum deviation of the plot of the CUSUM of the $\bar{X} - X_t$ occurs in Fig. 2). Using the $U_{t,T}$, we found above that the suggested change-point is at $t = 16$, a much more reasonable estimate considering the stem-and-leaf plot of Fig. 1.

However, the series is so erratic in the initial part, that determining the change-point numerically, using the maximum deviation of the $t(\bar{X}_t - \bar{X})$, is highly questionable. However, *the plot of the* CUSUM does, of course, offer the experienced analyst the information that the Series is highly variable before $t = 16$, it is much less variable after $t = 16$ and the mean has probably changed at about $t = 16$.

This example and that considered in Section 3 plainly show the robustness of the non-parametric technique and, on the other hand, the sensitivity of techniques based the differences $\bar{X}_t - \bar{X}$ to wild observations.

6. CONCLUSIONS

In this paper we have presented some simple techniques for testing for a change of distribution in a sequence of observations when the initial distribution is unknown.

For discrete Bernoulli and Binomial data, exact and conservative tests have been developed which are simple to use. For continuous data, approximate tests are developed which are both

Copyright © 2001. All Rights Reserved.

simple and robust against changes in distributional form. We propose to investigate the efficiency of these methods in a future paper, however, compared to normal theory CUSUM methods based on differences of means, the techniques for continuous data given here should be highly efficient when used with normal or near normal samples. This should follow from the known results on asymptotic relative efficiency of the two-sample t-test and the Mann–Whitney–Wilcoxon test, which show the latter to be highly efficient.

Acknowledgements

I thank Dr Derek Bissell for providing and allowing me to use the Industrial data of Section 5. Also the Lindisfarne data were brought to my attention by Professor Adrian Smith, for which I am grateful.

I am also indebted to the Editor and referees for their constructive criticisms on an original draft which led to improvements in the paper.

References

Billingsley, P. (1968). *Convergence of Probability Measures.* London: Wiley.

Gail, M. H. and Green, S. B. (1976). Critical values for the one-sided two-sample Kolmogorov–Smirnov statistic. *J. Amer. Statist. Ass.*, **71**, 757–760.

Hinkley, D. V. (1970). Inference about the change-point in a sequence of random variables. *Biometrika*, **57**, 1–17.

McGilchrist, C. A. and Woodyer, K. D. (1975). Note on a distribution-free CUSUM technique. *Technometrics*, **17**, 321–325.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, **41**, 100–114.

—— (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika*, **42**, 523–527.

—— (1957). On problems in which a change in parameter occurs at an unknown point. *Biometrika*, **44**, 248–252.

Pearson, E. S. and Hartley, H. O. (1972). *Biometrika Tables for Statisticians*, Vol. 2. London: Cambridge University Press.

Ross, A. S. C. (1950). Philological probability problems. *J. R. Statist. Soc.* B, **12**, 19–40.

Sen, A. and Srivastava, M. S. (1975a). On tests for detecting change in mean. *Ann. Statist.*, **3**, 98–108.

—— (1975b). Some one-sided tests for changes in level. *Technometrics*, **17**, 61–64.

Silvey, S. D. (1958). The Lindisfarne Scribes' problem. *J. R. Statist. Soc.* B, **20**, 93–101.

Smirnov, N. (1948). Table for estimating the goodness of fit of empirical distributions. *Ann. Math. Statist.*, **19**, 279–281.

Smith, A. F. M. (1975). A Bayesian approach to inference about a change-point in a sequence of random variables. *Biometrika*, **62**, 407–416.

Copyright © 2001. All Rights Reserved.

Copyright of Journal of the Royal Statistical Society: Series C (Applied Statistics) is the property of Wiley-Blackwell and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.