Then the two essential columns are the binomial coefficients for $n = 9$, forming column a_x, and those for $n = 11$, forming column b_x. Thus

x	a_x	b_x
0	1	165
1	9	330
2	36	462
3	84	462
4	126	330
5	126	165
6	84	55
7	36	11
8	9	1

Then, the sum of the product column, $\Sigma a_x b_x = 125{,}970$, which can be obtained directly on the machine without writing down the individual terms. Since the observed number, $x = 7$, in the cell with the smallest expectation is greater than expectation, we would require in this case the right-hand tail of the distribution. Thus

$$(36 \times 11) + (9 \times 1) = 405,$$

and

$$\sum_{7}^{8} P(x) = 405/125970 = 0{\cdot}003215.$$

A test for a change in a parameter occurring at an unknown point

By E. S. PAGE

Department of Mathematics, University of Durham

1. Introduction and summary

Consider a sample of independent observations in the order in which they were obtained, $x_1 \ldots x_n$; it is sometimes required to test the null hypothesis that all the observations are drawn from the same population with distribution function $F(x \mid \theta)$ against the alternative that $x_1, \ldots, x_m$ come from $F(x \mid \theta)$, and $x_{m+1}, \ldots, x_n$ from $F(x \mid \theta')$ $(\theta' \neq \theta)$. If m is known this is a straightforward problem of comparing two samples. In this paper we suppose that m is unknown; this raises new problems. A test is proposed for a case where θ is known and some comments are made on the problems presented by other cases.

2. One-sided case: θ known

Suppose that the initial value, θ, is known, One possibility for a test is to regard all the observations as a single sample and to use the best test that all the observations are from $F(x \mid \theta)$ against the alternative that all are from $F(x \mid \theta')$ for some $\theta' \neq \theta$. Such a test cannot be expected to be very powerful if the change occurs late in the sample; the few observations on the new parameter value would be obscured by the many on the old parameter. Since the problem of detecting a change in a parameter is important in controlling the quality of the output from a continuous production process, it is reasonable to investigate whether the methods of process inspection schemes can provide useful tests for the case in which we are interested.

A process inspection scheme for detecting a change in one direction in the parameter was given by the author in an earlier paper (Page, 1954). We suppose throughout that the parameter under consideration is the mean of the distribution unless the contrary is stated, and in this section we further assume that the value, θ, at the start of the observations is known. The scheme consisted of recording the cumulative sums $S_r = \sum_{i=1}^{r} (x_i - \theta)$, $S_0 = 0$, and taking action to rectify a possible change in the parameter when $S_r - \min_{0 \leqslant i < r} S_i \geqslant h$, i.e. when the sample path rises a height h above its previous minimum value; this procedure can be displayed clearly on a chart. If there is no change in θ, the mean path of the cumulative sum is horizontal, while if an increase in θ occurs the new mean path has positive slope so that the above criterion would be satisfied without too much delay. The significance test suggested by this procedure is as follows:

I. *Given the observations* $x_1 \dots x_n$. *It is required to test the hypothesis that the mean is constantly* θ. *Use as the test statistic* $m = \max_{0 \leqslant r \leqslant n} \{S_r - \min_{0 \leqslant i < r} S_i\}$, *where* $S_r = \sum_{j=1}^{r} (x_j - \theta)$, $S_0 = 0$, *taking large values as significant, i.e. reject the hypothesis if* $m \geqslant h$.

It was shown in the paper cited that the properties of the corresponding process inspection scheme depended upon the characteristics of linear sequential tests; as the test I is a truncated form of the process inspection scheme it is to be expected that in general the properties of the test will be difficult to evaluate. A special case that is tractable is where the observations are nought-or-one binomial variables. Accordingly, we consider a test for the general case using binomial variables.

II. *Given the observations* $x_1 \dots x_n$. *It is required to test the hypothesis that the mean is constantly* θ. *Let* $y_i = a$ *if* $x_i - \theta \geqslant 0$ *and* $y_i = -b$ *if* $x_i - \theta < 0$, *and choose* $a, b\ (>0)$ *so that* $E(y_i \mid \theta) = 0\ (i = 1, \dots, n)$. *Use as the test statistic* $m = \max_{0 \leqslant r \leqslant n} \{S_r - \min_{0 \leqslant i < r} S_i\}$, *where* $S_r = \sum_{j=1}^{r} y_j$, $S_0 = 0$, *taking large values as significant, i.e. reject the hypothesis if* $m \geqslant h$.

For simplicity we shall consider only the case where the distribution of the x_i is symmetrical, so that we can take $a = b = 1$; hence $y_i = \operatorname{sgn}(x_i - \theta)$. In order to evaluate the properties of the test let $m_r = S_r - \min_{0 \leqslant i < r} S_i$ and let $p_{r,i}$ be the probability that $m_r = i\ (i = 0, 1, \dots, h-1)$ and that $m_s < h$ for all s, $1 \leqslant s < r$. Then

$$1 - \sum_{i=1}^{h-1} p_{n,i} \tag{1}$$

is the probability that the null hypothesis is rejected. Let prob $(y_i = 1)$ be $p = 1 - q$. By considering the result of the next observation we have the relations

$$\left.\begin{aligned} p_{r+1,0} &= q(p_{r,0} + p_{r,1}), \\ p_{r+1,i} &= p \cdot p_{r,i-1} + q \cdot p_{r,i+1} \quad (1 \leqslant i < h-1), \\ p_{r+1,h-1} &= p \cdot p_{r,h-2}. \end{aligned}\right\} \tag{2}$$

In matrix notation we have

$$\mathbf{p}_{r+1} = \mathbf{P} \cdot \mathbf{p}_r, \tag{3}$$

where $\mathbf{P}$ is the square matrix,

$$\mathbf{P} = \begin{pmatrix} q & q & 0 & 0 & \dots & 0 & 0 \\ p & 0 & q & 0 & \dots & 0 & 0 \\ 0 & p & 0 & q & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & p & 0 \end{pmatrix}. \tag{4}$$

Initially $p_{0,0} = 1$, $p_{0,i} = 0\ (i \neq 0)$. In this formulation we are implicitly using the fact that the m_r are variables in a Markov chain with h live states. Clearly

$$\mathbf{p}_r = \mathbf{P}^r \cdot \mathbf{p}_0. \tag{5}$$

The expression for $\mathbf{p}_r$ may alternatively be written in terms of the latent roots and vectors, but the simplicity of the matrix $\mathbf{P}$ makes it quite convenient to use (4) for calculation. On the null hypothesis $p = \frac{1}{2}$ constantly. Table 1 shows values of h and the sample size n for which the probabilities of errors of the first kind are at most α, where $\alpha = 0{\cdot}05$ and $0{\cdot}01$. In order to ensure that the Type I errors are at most α for non-tabular values of n the larger value of h should be taken. For larger values of n rough interpolation will provide a sufficiently accurate value of h.

The power of the test II depends both on the value of p after the change and the position of the change. If prob $(y_i = +1)$ is constantly equal to p so that the change can be considered as having occurred immediately, the probability that the null hypothesis is rejected in a sample of n observations is given by equations (1) and (5) with $r = n$. If the change occurs after the kth observation the value of $\mathbf{p}_n$ to be used in (1) is given by

$$\mathbf{p}_n = \mathbf{P}_2^{n-k} \cdot \mathbf{P}_1^k \cdot \mathbf{p}_0,$$

where $\mathbf{P}_1$, $\mathbf{P}_2$ are the matrix $\mathbf{P}$ with $p = \frac{1}{2}$, $p = p$, respectively.

Table 1. *Values of n and h*

$\alpha = 0{\cdot}05$				$\alpha = 0{\cdot}01$	
n	h	n	h	n	h
21	10	75	19	20	12
26	11	83	20	27	14
31	12	91	21	35	16
36	13	100	22	43	18
41	14	119	24	53	20
47	15	139	26	64	22
54	16	161	28	76	24
60	17	185	30	89	26
67	18			103	28
				118	30

In Table 2 the power of the test II is compared with the power of the simple binomial test with approximately the same probability of Type I errors for a sample of 50 observations. For test II to have probability of Type I errors just less than 0·05 we need $h = 16$. The corresponding single sample test is 'Reject the null hypothesis if more than 31 of the y_i are positive', i.e. if $\sum_{i=1}^{50} y_i > 12$. The loss of power from using Test II instead of the single sample test is remarkably small.

Table 2. *Powers of the tests for different p*

p	Test II	Single sample test
0·50	0·039	0·032
0·55	0·136	0·127
0·60	0·336	0·336
0·65	0·609	0·622
0·70	0·844	0·859
0·75	0·964	0·971
0·80	0·996	0·997

The same two tests are contrasted in Table 3, where their powers are shown for different positions of the change from $p = 0{\cdot}5$ to $p = 0{\cdot}75$. Here, however, the test II has an appreciably greater power than the single-sample test when the change occurs near the middle of the set of observations. Also shown in Table 3 are the powers of the single-sample test on the last $50-m$ observations, when it is known that the change has occurred immediately before the mth observation. The differences between the power of this test and that of test II gives an indication of what is lost from the ignorance of the position of change.

In order to illustrate the test we give an example constructed from tables of random normal deviates. A sample of forty observations was constructed, the first twenty having mean 5 and unit variance, and the last twenty having mean 6 and unit variance; these are shown in Table 4. Suppose that it is required to test the hypothesis that the mean is constantly 5 against the alternatives that an increase in the mean has occurred within the sample. The observations, x_i, are shown in Table 4 together with $y_i = \operatorname{sgn}(x_i - 5)$, and the value taken by $S_r - \min S_i$.

The greatest value, h, of $S_r - \min S_i$ in the sample of 40 is 17, which approaches the 1 % significance point given in Table 1 (for $n = 40$, the approximate 5 % point is $h = 14$, the approximate 1 % point is $h = 18$). This significance level can be compared with that obtained from other tests applied to the

Table 3. *Powers of the tests for different positions of the change*

m	Test II	Single-sample test on whole sample	Single-sample test, m known
0	0·964	0·971	0·971
10	0·906	0·864	0·946
20	0·733	0·625	0·894
30	0·398	0·330	0·618
40	0·122	0·122	0·244
50	0·039	0·032	0·032

Table 4. *Artificial sampling experiment*

Observation no.	1	2	3	4	5	6	7	8	9	10
Value of x_i	3·95	5·96	6·22	5·58	4·02	4·97	3·46	4·29	4·65	5·66
$y_i = \operatorname{sgn}(x_i - 5)$	−1	+1	+1	+1	−1	−1	−1	−1	−1	+1
$S_r - \min S_i$	0	1	2	3	2	1	0	0	0	1
Observation no.	11	12	13	14	15	16	17	18	19	20
Value of x_i	5·44	5·91	4·98	3·58	5·26	3·98	4·19	6·66	6·05	5·97
$y_i = \operatorname{sgn}(x_i - 5)$	+1	+1	−1	−1	+1	−1	−1	+1	+1	+1
$S_r - \min S_i$	2	3	2	1	2	1	0	1	2	3
Observation no.	21	22	23	24	25	26	27	28	29	30
Value of x_i	7·14	6·22	4·76	6·60	5·72	4·88	5·44	5·03	5·66	5·56
$y_i = \operatorname{sgn}(x_i - 5)$	+1	+1	−1	+1	+1	−1	+1	+1	+1	+1
$S_r - \min S_i$	4	5	4	5	6	5	6	7	8	9
Observation no.	31	32	33	34	35	36	37	38	39	40
Value of x_i	6·37	6·66	5·10	5·80	6·29	5·49	4·93	6·18	8·29	6·84
$y_i = \operatorname{sgn}(x_i - 5)$	+1	+1	+1	+1	+1	+1	−1	+1	+1	+1
$S_r - \min S_i$	10	11	12	13	14	15	14	15	16	17

sample. The single-sample binomial test on the y's has 26 positives, 14 negatives; on the null hypothesis. the probability of this or a larger number of positives is 0·04. The change in the mean causes the estimate of variance of the x's to be inflated, and a t-test fails to give significance. The computation required by test II is so simple that it is unnecessary to record the y's, or even the $S_r - \min S_i$. An additional advantage of the test is that it gives an indication where the change took place; the position of the last zero of $S_r - \min S_i$ is an estimate (of course, biased) of the position of change. Thus in the example we would suspect that the change had occurred near observation 17.

3. General remarks

In this section we comment briefly on some other possible methods for the problem of § 2 and related problems without investigating their properties.

Another test for a change in one direction of a parameter from a known specified value may be obtained by analogy with the standard control chart process inspection scheme. The sample is divided into a number of subsamples of equal size and a statistic calculated from each subsample; the hypothesis of no change is rejected unless all the statistics fall within a certain range. The properties of this test are easy to evaluate and the number of subsamples and the permissible interval for the statistics can be chosen to control the errors. The test is also easy to apply and it is frequently useful in rough work. However the temporal ordering of the observations enters only into the division into subsamples, and it is of interest to examine whether it is advantageous to employ a slightly more complicated test of the form 'Reject the null hypothesis if any k of the statistics calculated from m consecutive subsamples fall outside an interval I, or if any one falls outside a wider interval I'' (cf. Wilkinson, 1951; Tippett, 1931).

The control chart procedure can also provide a test for the two-sided case where the change from the known value can be in either direction. A test based on the mean path of a cumulative sum similar to test I is a truncated sequential test (Rao, 1950). Another case that needs to be considered is where the initial value of the parameter is unknown.

I wish to thank Dr D. R. Cox for a number of discussions on the subject of this paper, and the Director, Mathematical Laboratory, Cambridge, for permission to use the EDSAC.

REFERENCES

PAGE, E. S. (1954). *Biometrika*, **41**, 100.
RAO, C. R. (1950). *Sankhyā*, **10**, 361.
TIPPETT, L. H. C. (1931). *The Methods of Statistics*. London: Williams and Norgate.
WILKINSON, B. (1951). *Psychol. Bull.* **48**, 156.

A paradox in statistical estimation

BY ALAN STUART

Division of Research Techniques, London School of Economics

1. Sundrum (1954) has recently shown to be incorrect the intuitive idea that the more efficient of two estimators of a parameter necessarily provides the more powerful test of a hypothesis concerning that parameter. This note discusses a similar paradox which arises in a problem concerned purely with estimation: given an estimator u of a parameter θ in a multiparameter distribution, one does not necessarily improve its efficiency by substituting true parameter values into u to replace estimators of them.

2. We shall consider the case where there is only one other parameter, say μ, and where we have a consistent estimator of θ

$$t = f(s, \mu), \tag{1}$$

where s is a function of the n observations only. If μ is unknown, we reduce t to a function of the observations only by substituting for μ a consistent estimator of it, say m, giving

$$u = f(s, m). \tag{2}$$

From (1) we have, to order n^{-1},

$$V(t) = \left(\frac{\partial t}{\partial s}\right)^2 V(s). \tag{3}$$

From (2) we have the corresponding result for a function of two random variables

$$V(u) = \left(\frac{\partial u}{\partial s}\right)^2 V(s) + \left(\frac{\partial u}{\partial m}\right)^2 V(m) + 2\frac{\partial u}{\partial s}\frac{\partial u}{\partial m} C(s, m). \tag{4}$$

Since all the derivatives in (3) and (4) are to be taken at the true parameter point (θ, μ), the first term on the right of (4) is equal to (3). Thus

$$V(u) - V(t) = \left(\frac{\partial u}{\partial m}\right)^2 V(m) + 2\frac{\partial u}{\partial s}\frac{\partial u}{\partial m} C(s, m). \tag{5}$$

(5) is not generally positive, although it must be so if s and m are uncorrelated. In general, their correlation must be taken into account before the effect of substituting parameters into u can be assessed. It is the correlation term in (5) which resolves the paradox.