Inference About the Change-Point in a Sequence of Random Variables

Author(s): David V. Hinkley

Source: *Biometrika*, Apr., 1970, Vol. 57, No. 1 (Apr., 1970), pp. 1-17
Published by: Oxford University Press on behalf of Biometrika Trust
Stable URL: https://www.jstor.org/stable/2334932

Biometrika (1970), **57**, 1, *p.* 1
Printed in Great Britain

Inference about the change-point in a sequence of random variables

By DAVID V. HINKLEY
Imperial College

Summary

Inference is considered about the point in a sequence of random variables at which the probability distribution changes. In particular, we examine a normal distribution with changing mean. The asymptotic distribution of the maximum likelihood estimate is derived and also the asymptotic distribution of the likelihood ratio statistic for testing hypotheses about the change-point. These asymptotic distributions are compared with some finite sample empirical distributions.

1. Introduction

Given a sequence $(x_1, \ldots, x_T)$ of observations on the variables $(X_1, \ldots, X_T)$, there are available many techniques for detecting underlying patterns or relationships, most of which assume that a single model is valid for the whole range of the sample. Often the model is of the form

$$X_t = \theta(t) + \epsilon_t \quad (t = 1, \ldots, T),$$

where $\{\epsilon_t\}$ is a sequence of uncorrelated error terms with zero mean and $\theta(t)$ is a continuous mean function. In practice, however, a single model is sometimes inappropriate. It may be strongly suspected that the model valid near $t = 1$ is not valid near $t = T$. It is then relevant to consider models of the form

$$\left.\begin{aligned} X_t &= \theta_0(t) + \epsilon_t \quad (t = 1, \ldots, \tau), \\ X_t &= \theta_1(t) + \epsilon_t \quad (t = \tau+1, \ldots, T), \end{aligned}\right\} \tag{1·1}$$

where the change-point τ is unknown. Experimental and scientific grounds may force us to consider (1·1), or even its generalization to $(p+1)$ submodels with p unknown change-points.

Two simple but interesting and important special cases of (1·1) are (*a*) two constant means θ_0 and θ_1, and (*b*) two intersecting regressions $\theta_i(t) = \alpha + \beta_i(z_t - \gamma)$ $(i = 0, 1)$, where z_t is an independent variable and $z_\tau \leqslant \gamma < z_{\tau+1}$. In both cases it is usual to assume the error terms ϵ_t to be $N(0, \sigma^2)$. Maximum likelihood estimation and inference in the regression case (*b*) have been discussed in detail by Hudson (1966) and Hinkley (1969). In case (*a*) the emphasis of published work, in particular that of Page (1954, 1955, 1957) on cumulative sum schemes, has been on testing the null hypothesis H_0: $\theta_0 = \theta_1$ against the two-mean alternative. Chernoff & Zacks (1964) and Bhattacharyya & Johnson (1968) have discussed the same problem within a Bayesian framework. The present paper is concerned primarily with estimating and making inference about the change-point τ in this two-mean case.

We consider initially (§ 2) a generalization of case (*a*) where X_t has an arbitrary continuous probability density function $f(x, \theta)$ with θ changing from θ_0 to θ_1 after index τ. The asymptotic distribution of the maximum likelihood estimate $\hat{\tau}$ is derived for θ_0 and θ_1 known.

In § 3 we discuss the computation of the asymptotic distribution in the normal case, and show that the asymptotic distribution is unchanged when θ_0 and θ_1 are unknown. The asymptotic distribution of the likelihood ratio statistic for testing hypotheses about τ is also derived (§ 4). Tables of these distributions are given, and in § 5 we compare the asymptotic results with some empirical finite sample results.

2. Distribution of the maximum likelihood estimate: θ_0 and θ_1 known

In this section we derive the maximum likelihood estimate $\hat{\tau}$ for a general probability density function $f(x,\theta)$ of the random variable X_t and find its asymptotic distribution; that is, for both τ and $T-\tau$ indefinitely large. No explicit form for $\hat{\tau}$ exists, but it is conveniently defined by variables associated with random walks, whose properties we use to get the asymptotic distribution of $\hat{\tau}$ in a form suitable for computation.

2·1. *Exact results*

Let $(X_1, \ldots, X_T)$ be a sequence of independent continuous random variables such that X_i has probability density function $f(x,\theta_0)$ $(i = 1, \ldots, \tau)$ and X_i has probability density function $f(x,\theta_1)$ $(i = \tau+1, \ldots, T)$, where θ_0 and θ_1 are known $(\theta_0 \neq \theta_1)$ but τ is unknown. To obtain the maximum likelihood estimate $\hat{\tau}$ from a sample, $x_1, \ldots, x_T$, we have to maximize the log likelihood function

$$L(t) = \sum_{i=1}^{t} \log f(x_i, \theta_0) + \sum_{i=t+1}^{T} \log f(x_i, \theta_1) \tag{2·1}$$

over admissible values of t, namely $t = 1, \ldots, T-1$ since we assume at least one observation to come from each distribution.

A more convenient form for $L(t)$ is obtained by defining the log likelihood increments

$$U_j = \log f(X_j, \theta_0) - \log f(X_j, \theta_1). \tag{2·2}$$

Then (2·1) becomes

$$L(t) = \sum_{j=1}^{t} u_j + \sum_{j=1}^{T} \log f(x_j, \theta_1),$$

where u_j is the observed value of U_j, so that the maximum likelihood estimate $\hat{\tau}$ is the value of t which maximizes the sequence of partial sums

$$V_t = \sum_{j=1}^{t} U_j \quad (t = 1, \ldots, T-1).$$

The U_j's are independent because of the independence of the X_j's. It is not difficult to see that the differences $V_t - V_\tau$ define two independent random walks, namely

$$\left.\begin{aligned} \mathbf{W} &= \left\{0, -\sum_{j=0}^{k} U_{\tau-j} \quad (k = 0, 1, \ldots, \tau-1)\right\}, \\ \mathbf{W}' &= \left\{0, \sum_{j=0}^{k} U_{\tau+1+j} \quad (k = 0, 1, \ldots, T-\tau-2)\right\}, \end{aligned}\right\} \tag{2·3}$$

each random walk having independent, identically distributed increments with negative mean. Thus $\mathbf{W}$ represents the log likelihood for integers less than τ, and $\mathbf{W}'$ likewise for integers greater than τ, relative to the log likelihood of τ itself. To determine $\hat{\tau}$ we must find the larger of the two random walk maxima; if each maximum is zero, then $\hat{\tau} = \tau$.

Now let $Y_j = -U_{\tau-j+1}$ and $Y'_j = U_{\tau+j}$ $(j = 1, 2, \ldots)$, so that (2·3) becomes

$$\left.\begin{aligned} \mathbf{W} &= \left\{0, Y_1, Y_1+Y_2, \ldots, \sum_{j=1}^{\tau-1} Y_j\right\}, \\ \mathbf{W}' &= \left\{0, Y'_1, Y'_1+Y'_2, \ldots, \sum_{j=1}^{T-\tau-1} Y'_j\right\}. \end{aligned}\right\} \tag{2·4}$$

Also let M and M' be the respective maxima of $\mathbf{W}$ and $\mathbf{W}'$. The finite sample distribution of $(\hat{\tau}-\tau)$ depends implicitly on τ and $(T-\tau)$ because M and M' do, but we shall assume both τ and $(T-\tau)$ to be infinitely large and derive the asymptotic distribution of $(\hat{\tau}-\tau)$ using properties of the random walks $\mathbf{W}$ and $\mathbf{W}'$.

We can now express events involving $\hat{\tau}$ in terms of events involving M and M': we have seen already that $\hat{\tau} = \tau$ is equivalent to $M = M' = 0$. Further, $\hat{\tau} = \tau+k$ is equivalent to $M' = Y'_1+\ldots+Y'_k > 0$ and $M' > M$, and similarly for the event $\hat{\tau} = \tau-k$. The event $M = M' > 0$ has zero probability because the random variables are continuous. To simplify the analysis let

$$I = \inf\left(k\colon M = \sum_{j=1}^{k} Y_j\right), \quad I = 0 \quad \text{if} \quad M = 0,$$

$$I' = \inf\left(k\colon M' = \sum_{j=1}^{k} Y'_j\right), \quad I' = 0 \quad \text{if} \quad M' = 0.$$

Thus I and I' are the indices of the maxima of $\mathbf{W}$ and $\mathbf{W}'$. It follows by the independence of $\mathbf{W}$ and $\mathbf{W}'$ that

$$\left.\begin{aligned} \operatorname{pr}(\hat{\tau} = \tau) &= \operatorname{pr}(M = 0)\operatorname{pr}(M' = 0), \\ \operatorname{pr}(\hat{\tau} = \tau+k) &= \operatorname{pr}(I' = k, M' > M, M' > 0) \quad (k = 0, 1, \ldots), \\ \operatorname{pr}(\hat{\tau} = \tau-k) &= \operatorname{pr}(I = k, M > M', M > 0) \quad (k = 0, 1, \ldots). \end{aligned}\right\} \tag{2·5}$$

Now define, for $k = 0, 1, \ldots$ and $x \geqslant 0$, the probability measures

$$\beta_k(x)\,dx = \operatorname{pr}(I = k, x \leqslant M < x+dx), \quad \beta'_k(x)\,dx = \operatorname{pr}(I' = k, x \leqslant M' < x+dx), \tag{2·6}$$

$$\alpha(x) = \operatorname{pr}(M \leqslant x), \quad \alpha'(x) = \operatorname{pr}(M' \leqslant x), \tag{2·7}$$

where the prime distinguishes $\mathbf{W}$ and $\mathbf{W}'$ and never means differentiation. Note that, since $I = 0$ implies $M = 0$ and $I' = 0$ implies $M' = 0$, $\beta_0(x) = \alpha(0)\,\delta(x)$ and $\beta'_0(x) = \alpha'(0)\,\delta(x)$, where $\delta(0) = 1$ and $\delta(x) = 0$ $(x \neq 0)$. We omit the dependence of $\alpha(.), \ldots$ on θ_0, θ_1 and f for simplicity of notation.

With the definitions (2·6) and (2·7), the probabilities (2·5) become, again using the independence of $\mathbf{W}$ and $\mathbf{W}'$,

$$\operatorname{pr}(\hat{\tau} = \tau) = \alpha(0)\,\alpha'(0), \tag{2·8}$$

$$\operatorname{pr}(\hat{\tau} = \tau+k) = \int_0^\infty \beta'_k(x)\,\alpha(x)\,dx \quad (k = 0, 1, \ldots), \tag{2·9}$$

$$\operatorname{pr}(\hat{\tau} = \tau-k) = \int_0^\infty \beta_k(x)\,\alpha'(x)\,dx \quad (k = 0, 1, \ldots). \tag{2·10}$$

It is easy to verify that $\alpha(.)$ and $\alpha'(.)$ satisfy integral equations. For example, if the probability density function of Y_j is $g(.)$, then

$$\alpha(x) = \int_0^\infty \alpha(y)\,g(x-y)\,dy. \tag{2·11}$$

The densities $\beta_k(.)$ and $\beta'_k(.)$ also satisfy integral equations, but of a recursive type. However, the integral equations do not have explicit solutions, except in trivial cases. We therefore derive a method for calculating the integrals in (2·9) and (2·10) which does not involve an explicit solution for $\alpha(.), \ldots$.

An important result, deduced from Feller (1966, Chap. 18), is that if the sequence of independent identically distributed random variables $Y'_1, Y'_2, \ldots$ satisfies

$$\sum_{n=1}^{\infty} \frac{1}{n} \operatorname{pr}\left(\sum_{j=1}^{n} Y'_j > 0\right) < \infty, \tag{2·12}$$

then

$$E(z^{I'} e^{-\omega M'}) = \sum_{n=0}^{\infty} z^n \int_0^{\infty} e^{-\omega x} \beta'_n(x)\, dx$$

$$= \exp\left\{\sum_{n=1}^{\infty} \frac{z^n c'_n(\omega)}{n} - \sum_{n=1}^{\infty} \frac{c'_n(0)}{n}\right\}, \tag{2·13}$$

where

$$c'_n(\omega) = \int_{0+}^{\infty} e^{-\omega x} \operatorname{pr}\left(x \leqslant \sum_{j=1}^{n} Y'_j < x + dx\right) \quad (n = 1, 2, \ldots), \tag{2·14}$$

$|z| \leqslant 1$ and $\mathscr{R}(\omega) > 0$. A similar result holds for I and M with $c'_n(\omega)$ replaced by

$$c_n(\omega) = \int_{0+}^{\infty} e^{-\omega x} \operatorname{pr}\left(x \leqslant \sum_{j=1}^{n} Y_j < x + dx\right) \quad (n = 1, 2, \ldots).$$

The marginal distributions of I' and M' are derived by setting $\omega = 0$ and $z = 1$ respectively in (2·13). In particular we find that

$$\sum_{k=0}^{\infty} z^k \operatorname{pr}(I' = k) = \exp\left\{\sum_{n=1}^{\infty} \frac{(z^n - 1)\, c'_n(0)}{n}\right\} \quad (|z| \leqslant 1), \tag{2·15}$$

$$\alpha'(0) = \exp\left\{-\sum_{n=1}^{\infty} \frac{c'_n(0)}{n}\right\}, \tag{2·16}$$

with corresponding results for I and M.

Now consider the probabilities $\operatorname{pr}(\hat{\tau} = \tau + k)$ in (2·9). In principle, we can obtain the generating function of these probabilities explicitly. For (2·9) can be written as

$$\operatorname{pr}(\hat{\tau} = \tau + k) = \int_0^{\infty} \beta'_k(x)\, dx - \int_0^{\infty} \beta'_k(x)\{1 - \alpha(x)\}\, dx, \tag{2·17}$$

and hence

$$\sum_{k=0}^{\infty} z^k \operatorname{pr}(\hat{\tau} = \tau + k) = \sum_{k=0}^{\infty} z^k \operatorname{pr}(I' = k) - \int_0^{\infty} \sum_{k=0}^{\infty} z^k \beta'_k(x)\{1 - \alpha(x)\}\, dx. \tag{2·18}$$

The two functions in the integral of (2·18) are integrable, assuming (2·12), with Laplace transforms determined explicitly from (2·13). Hence we can apply the Parseval relation to transform the integral in (2·18) into the integral in the ω plane of the product of two Laplace transforms. Unfortunately, however, the resulting complex integral does not appear to have an explicit form. But the result (2·13) and its corollaries can be used to derive an approximation to (2·17) which is suitable for computation.

The corresponding results hold for $\operatorname{pr}(\hat{\tau} = \tau - k)$ by applying the same arguments to (2·10), simply interchanging $\alpha(.)$ and $\alpha'(.)$, $\beta'_k(.)$ and $\beta_k(.)$.

2·2. *Approximation to the distribution*

The approximation to (2·17) is based on approximating $\{1-\alpha(\,.\,)\}$ by a sum of exponential terms. For suppose that

$$1-\alpha(x) = \sum_{r=1}^{R} a_r e^{-\omega_r x} \quad (x \geqslant 0), \tag{2·19}$$

where R is possibly infinite. Then (2·17) becomes

$$\mathrm{pr}\,(\hat{\tau} = \tau + k) = q_k' - \sum_{r=1}^{R} a_r\, \tilde{\beta}_k'(\omega_r) \quad (k = 0, 1, \ldots), \tag{2·20}$$

where

$$q_k' = \int_0^\infty \beta_k'(x)\,dx = \mathrm{pr}\,(I' = k),$$

$$\tilde{\beta}_k'(\omega) = \int_0^\infty e^{-\omega x}\beta_k'(x)\,dx,$$

which are respectively the coefficients of z^k in (2·15) and (2·13). Computation of (2·20) is straightforward because q_k' and $\beta_k'(\omega)$ satisfy simple recurrence relations. In fact both (2·15) and (2·13) are equations of the type

$$\sum_{n=0}^{\infty} \rho_n z^n = A \exp\left(\sum_{n=1}^{\infty} \frac{z^n \sigma_n}{n}\right),$$

whose solution is

$$\rho_0 = A,$$

$$\rho_{n+1} = \frac{1}{n+1} \sum_{j=0}^{n} \sigma_{n+1-j}\rho_j \quad (n = 0, 1, \ldots). \tag{2·21}$$

In general, (2·19) with R finite will be an approximation, so that (2·20) will be an approximation also. The way in which we get the approximation (2·19) is to fit the series of exponential terms by least squares to numerical values of $\{1-\alpha(x)\}$, which are obtained by numerical solution of (2·11). For this solution it is convenient to consider (2·11) in the form

$$1-\alpha(x) = \int_x^\infty g(u)\,du + \int_0^\infty \{1-\alpha(u)\}\,g(x-u)\,du. \tag{2·22}$$

Then numerical values of $\{1-\alpha(x)\}$ are obtained by taking a discrete, finite version of (2·22). That is, we take a finite set of x values, $0 = x_0 < x_1 < \ldots < x_n$ with $x_{i+1} - x_i = d$, say, and approximate (2·22) by

$$1-\alpha(x_i) = \int_{x_i}^\infty g(u)\,du + d \sum_{j=0}^{n} h_j\{1-\alpha(x_j)\}\,g(x_i - x_j) \quad (i = 0, 1, \ldots, n), \tag{2·23}$$

where $h_0 = h_n = \frac{1}{2}$ and $h_j = 1$ $(j = 1, \ldots, n-1)$. This system of equations is linear and is easily solved. Both d and n can be varied until satisfactory accuracy is obtained. A detailed description of this numerical solution is given in § 3 for the normal case.

In some cases, for example, that of normal probability density functions discussed in § 3, a theoretical analysis of (2·22) will give useful information about one or more dominant exponential terms in the series (2·19).

2·3. *General remarks*

A necessary condition for the moments of $(\hat{\tau}-\tau)$ to exist is that (2·12) holds for both $\{Y_j\}$ and $\{Y'_j\}$. Expressed in terms of $f(x,\theta)$ this condition is that

$$\sum_{n=1}^{\infty} \frac{1}{n} \operatorname{pr}\left\{\prod_{j=1}^{n} \frac{f(X_j,\theta_1)}{f(X_j,\theta_0)} > 1 \,\middle|\, X_j \text{ has probability density function } f(x,\theta_0)\right\} < \infty,$$

and similarly with θ_0 and θ_1 interchanged. Roughly speaking, this means that whether or not the moments of $(\hat{\tau}-\tau)$ are finite depends on the power of the likelihood ratio test to distinguish between the alternatives $\theta = \theta_0$ and $\theta = \theta_1$ in samples from one population.

One awkward property of $\hat{\tau}$ is that it is not consistent, since as τ and $(T-\tau)$ tend to infinity all existing moments of $(\hat{\tau}-\tau)$ have nonzero limits. This property is to be expected because observations distant from the change-point give negligible information about τ. Consistent estimation of τ could only follow from consistent classification of observations.

For $\hat{\tau}$ to be symmetrically distributed about τ it is necessary and sufficient that Y_j and Y'_j be identically distributed. In terms of $f(x,\theta)$ this implies that

$$\operatorname{pr}\{f(X,\theta_0) \leqslant yf(X,\theta_1) \,|\, \theta = \theta_1\} = \operatorname{pr}\{f(X,\theta_1) \leqslant yf(X,\theta_0) \,|\, \theta = \theta_0\}. \tag{2·24}$$

It is easy to see that (2·24) holds when $f(x,\theta)$ is symmetric and θ is a location parameter only, or if any (1, 1) continuous transformation of X has such a probability density function. Thus, for example, (2·24) holds for the normal distribution with mean θ and the corresponding log normal distribution, but not for the exponential density $\theta e^{-\theta x}$.

We have assumed in this section that τ is the only unknown parameter; that is, that θ_0 and θ_1 are known. But it is not difficult to see that the asymptotic distribution of $\hat{\tau}$ will be unchanged when θ_0 and θ_1 are unknown, provided that the maximum likelihood estimates of θ_0 and θ_1 conditional on $\tau = t$ are consistent. In § 3·2 we show that this is so for the particular case of normally distributed random variables.

3. THE NORMAL CASE

We now apply the results of § 2 to the case where θ_0 and θ_1 are mean values of normal distributions with constant variance σ^2. The computation of the asymptotic distribution of $\hat{\tau}$ is discussed in detail and some numerical results are given. We also show that the same asymptotic distribution holds when θ_0 and θ_1 are unknown and when σ^2 is unknown.

3·1. *θ_0 and θ_1 known*

Let $f(x,\theta_\nu)$ be the probability density function of the $N(\theta_\nu,\sigma^2)$ $(\nu = 0, 1)$ distribution with θ_0 and θ_1 known. If σ^2 is also known we can follow the results of § 2 directly, where now the log likelihood increments U_i defined in (2·2) are given by

$$U_i = [(\theta_0-\theta_1)\{X_i-\tfrac{1}{2}(\theta_0+\theta_1)\}]/\sigma^2. \tag{3·1}$$

If σ^2 is unknown the likelihood of $(X_1,\ldots,X_T)$ maximized over σ^2 conditional on $\tau = t$ is proportional to

$$\left\{\sum_{i=1}^{T} (X_i-\theta_1)^2 - \sigma^2 \sum_{i=1}^{t} U_i\right\}^{-\frac{1}{2}T} \quad (t = 1,\ldots,T-1),$$

so that $\hat{\tau}$ is the value of t which maximizes $\sigma^2 \sum_{i=1}^{t} U_i$. Hence the results of § 2 also apply when σ^2 is unknown.

Now the random walk increments Y_i and Y_i' are, by (3·1), identically distributed $N(-2\Delta^2, 4\Delta^2)$, where $\Delta = |\theta_1 - \theta_0|/(2\sigma)$. The distribution of $\hat{\tau}$ will be unchanged if both Y_i and Y_i' are rescaled by a factor $(2\Delta)^{-1}$. Therefore, without loss of generality we can take Y_i and Y_i' to be $N(-\Delta, 1)$. Further, the distribution of $\hat{\tau}$ is symmetric about τ since Y_i and Y_i' are identically distributed, so we can omit the redundant prime superfix used in § 2 and consider only the distribution of $\hat{\tau}$ for $\hat{\tau} \geqslant \tau$. To emphasize their dependence on Δ, the functions $\alpha(.), \beta_k(.), \ldots$ will be denoted by $\alpha(., \Delta), \beta_k(., \Delta), \ldots$.

First we summarize the exact results of § 2·1 for this special case. Let

$$p_k(\Delta) = \mathrm{pr}\,(\hat{\tau} = \tau \pm k; \Delta).$$

Then we have immediately from (2·8) and (2·22) that

$$p_0(\Delta) = \{\alpha(0, \Delta)\}^2, \tag{3·2}$$

$$1 - \alpha(x, \Delta) = 1 - \Phi(x + \Delta) + \int_0^\infty \{1 - \alpha(u, \Delta)\}\, \phi(x - u + \Delta)\, du, \tag{3·3}$$

where ϕ and Φ are respectively the standard normal density function and integral. Also, by (2·17) and the symmetry of $\hat{\tau}$,

$$p_k(\Delta) = q_k(\Delta) - \int_0^\infty \{1 - \alpha(u, \Delta)\}\, \beta_k(u, \Delta)\, du, \tag{3·4}$$

where $q_k(\Delta) = \mathrm{pr}\,(I = k; \Delta)$. It is not difficult to see from (2·14), (2·15), (2·16) and (2·21) that when Y_j is $N(-\Delta, 1)$, $q_k(\Delta)$ satisfies the recurrence relation

$$q_{k+1}(\Delta) = \frac{1}{k+1} \sum_{j=0}^{k} q_j(\Delta)\, \Phi\{-\Delta \surd(k+1-j)\} \quad (k \geqslant 1) \tag{3·5}$$

with $q_0(\Delta) = \alpha(0, \Delta)$.

We use the technique outlined in § 2·2 for approximating to the distribution (3·4), first approximating $\{1 - \alpha(x, \Delta)\}$ by

$$\sum_{r=1}^{R} a_r(\Delta) \exp\{-\omega_r(\Delta) x\}, \tag{3·6}$$

as in (2·19). Then the approximation to (3·4) is, by (2·20),

$$p_k(\Delta) = q_k(\Delta) - \sum_{r=1}^{R} a_r(\Delta)\, \tilde{\beta}_k\{\omega_r(\Delta), \Delta\}. \tag{3·7}$$

The Laplace transforms $\tilde{\beta}_k(\omega, \Delta)$ satisfy a recurrence relation which we deduce from (2·13), (2·14), (2·16) and (2·21) to be

$$\tilde{\beta}_{k+1}(\omega, \Delta) = \frac{1}{k+1} \sum_{j=0}^{k} \tilde{\beta}_j(\omega, \Delta)\, c_{k+1-j}(\omega, \Delta) \quad (k \geqslant 1) \tag{3·8}$$

with $\tilde{\beta}_0(\omega, \Delta) = \alpha(0, \Delta)$ and

$$c_k(\omega, \Delta) = \exp\,(k\Delta\omega + \tfrac{1}{2}k\omega^2)\, \Phi\{-(\omega + \Delta)\surd k\}. \tag{3·9}$$

It remains only to derive the coefficients in the series approximation (3·6).

In this particular case it is easy to deduce from the integral equation (3·3) that as x tends to infinity

$$1 - \alpha(x, \Delta) \sim a_1(\Delta)\, e^{-2\Delta x},$$

as may be verified by substitution in (3·3). This asymptotic solution is the first term in (3·6).

The constant $a_1(\Delta)$ is determined numerically from the numerical solution of (3·3), which follows the outline given in § 2·2.

It is worth describing the numerical solution of (3·3) in some detail. The discrete version of the integral equation in (2·23) implicitly assumed $\{1-\alpha(x,\Delta)\}$ to be zero for $x > x_n$ by the necessary truncation, but here this assumption can be removed by substituting

$$1-\alpha(x,\Delta) = a_1(\Delta)\, e^{-2\Delta x} \quad (x > x_n)$$

in (3·3) before taking the discrete approximation to the integral. The discrete equation corresponding to (2·23) then becomes

$$1-\alpha(jd,\Delta) = \psi(jd+\Delta)+d\sum_{k=0}^{n} h_k\{1-\alpha(kd,\Delta)\}\,\phi\{(j-k)\,d+\Delta\}, \qquad (3\cdot 10)$$

where

$$\psi(jd+\Delta) = 1-\Phi(jd+\Delta)+a_1(\Delta)\,e^{-2jd\Delta}\,[1-\Phi\{(n-j)\,d+\Delta\}] \quad (j = 0, 1, \ldots, n),$$

$h_0 = h_n = \frac{1}{2}$ and $h_k = 1$ otherwise. Here $a_1(\Delta)$ is unknown, but using successive trial values for $a_1(\Delta)$ we can determine the correct value by iteration. As an example consider the case $\Delta = 1\cdot 0$ and let $d = 0\cdot 1$ and $n = 100$. The solutions to (3·10) for trial values $a_1(\Delta) = 0$ and 1 are given in the first two columns of Table 3·1 in terms of $\{1-\alpha(x,\Delta)\}\exp(2\Delta x)$. Both solutions become stable around $x = 4$, but then diverge toward the assumed value of $a_1(\Delta)$. The next trial value suggested by these solutions is $a_1(\Delta) = 0\cdot 34$, and in two more iterations we

Table 3·1. *Values of* $\{1-\alpha(x,\Delta)\}\exp(2\Delta x)$ *from* (3·10) *for three trial values of* $a_1(\Delta)$ *with* $\Delta = 1\cdot 0$

x \ $a_1(\Delta)$	0	1	0·32037
0	0·19946	0·29509	0·19939
2	·32396	·38415	·32403
4	·32039	·36452	·32035
6	·32035	·36429	·32036
8	·31848	·36799	·32037
10	·25638	·49121	·32037
>10	·00000	1·00000	·32037

Table 3·2. *Comparison of exact and approximate values of* $\{1-\alpha(x,\Delta)\}$ *using* (3·10) *and* (3·11): $\Delta = 1\cdot 0$

x	Exact	Approximate
0·0	0·1994	0·1988
0·5	·0969	·0978
1·0	·0414	·0400
1·5	·0161	·0154
2·0	·00593	·00578
3·0	·00079	·00079

get the final solution in Table 3·1 with $a_1(\Delta) = 0\cdot 32037$. Note that this solution reaches its asymptotic form before x_n. The value of x_n, in this case 10·0, is chosen so that this happens, since only then can we be satisfied that the final solution is the solution to (3·3). For advice on this method of solution I am indebted to Dr Morven Gentleman.

To obtain further terms in the series (3·6) it is convenient to remove the first term and work with

$$\gamma(x,\Delta) = e^{2\Delta x}\{1-\alpha(x,\Delta)\}-a_1(\Delta)$$

rather than $\{1-\alpha(x,\Delta)\}$. Numerical solution for $\gamma(x,\Delta)$ proceeds in the same way as for

Table 3·3. *Asymptotic distribution of $\hat{\tau}$ in the normal case*

$\Delta = 0{\cdot}5\,(0{\cdot}1)\,1{\cdot}5.$

k \ Δ	0·5		0·6		0·7		0·8		0·9		1·0		1·1		1·2		1·3		1·4		1·5	
	$p_k(\Delta)$	$P_k(\Delta)$	$p_k(\Delta)$	$P_k(\Delta)$	$p_k(\Delta)$	$P_k(\Delta)$	$p_k(\Delta)$	$P_k(\Delta)$	$p_k(\Delta)$	$P_k(\Delta)$	$p_k(\Delta)$	$P_k(\Delta)$	$p_k(\Delta)$	$P_k(\Delta)$	$p_k(\Delta)$	$P_k(\Delta)$	$p_k(\Delta)$	$P_k(\Delta)$	$p_k(\Delta)$	$P_k(\Delta)$	$p_k(\Delta)$	$P_k(\Delta)$
0	0·280	0·640	0·360	0·680	0·438	0·719	0·511	0·756	0·579	0·790	0·641	0·820	0·696	0·848	0·746	0·873	0·788	0·894	0·825	0·913	0·857	0·928
1	·114	·754	·124	·804	·128	·847	·127	·882	·121	·911	·113	·933	·103	·951	·092	·965	·081	·975	·070	·982	·059	·988
2	·067	·821	·066	·870	·061	·908	·054	·936	·046	·957	·038	·971	·030	·981	·024	·988	·018	·993	·013	·996	·010	·997
3	·044	·865	·040	·910	·034	·941	·027	·963	·021	·977	·015	·987	·010	·992	·007	·996	·005	·998				
4	·031	·896	·026	·936	·020	·962	·015	·978	·010	·988	·007	·993	·004	·997								
5	·023	·919	·018	·953	·013	·974	·009	·986	·005	·993	·003	·997										
6	·017	·935	·012	·965	·008	·982	·005	·991	·003	·996												
7	·013	·948	·009	·974	·005	·988	·003	·995														
8	·010	·958	·006	·981	·004	·991	·002	·996														
9	·008	·966	·005	·985	·003	·994																
10	·006	·973	·004	·989	·002	·996																
12	·004	·982	·002	·993																		
14	·003	·987	·001	·996																		
16	·002	·991																				
18	·001	·994																				
20	·001	·996																				

$\{1-\alpha(x,\Delta)\}$ with obvious changes. Now truncation at x_n has negligible effect if we have chosen x_n correctly. A detailed analysis of $\gamma(x,\Delta)$ is given by Dr Gentleman in an unpublished Bell Telephone Laboratories Memorandum, but in the present context a satisfactory approximation to $\{1-\alpha(x,\Delta)\}$ is achieved by fitting a single exponential term to numerical values of $\gamma(x,\Delta)$ using unweighted least squares. The approximation (3·6) is then

$$1-\alpha(x,\Delta) \simeq a_1(\Delta)\exp\{-\omega_1(\Delta)x\}+a_2(\Delta)\exp\{-\omega_2(\Delta)x\}, \tag{3·11}$$

where $\omega_1(\Delta) = 2\Delta$ and $a_1(\Delta)$, $a_2(\Delta)$ and $\omega_2(\Delta)$ are determined numerically. Table 3·2 compares the exact numerical solution of (3·10) with values computed from (3·11) in the case $\Delta = 1$. Looking back to (2·9) we can clearly see that the relative error of the approximation (3·7) is no greater than the maximum relative error in the approximation to $\alpha(x,\Delta)$, which is less than 0·3 % in this case.

Extensive calculations of (3·7) have been done by the above method for $0·5 \leqslant \Delta \leqslant 1·5$; that is, for a change in normal mean between one and three standard deviations. In each case the values $d = 0·1$ and $n = 100$ were used in (3·10), the relative error of the resulting approximation to $\alpha(x,\Delta)$ in (3·11) being consistently less than 0·3 %. The computed approximations to $p_k(\Delta)$ are given in Table 3·3 for $\Delta = 0·5\,(0·1)\,1·5$ together with the cumulative probabilities

$$P_k(\Delta) = \mathrm{pr}\,(\hat{\tau} \leqslant \tau+k;\Delta) = \tfrac{1}{2}+\tfrac{1}{2}p_0(\Delta)+\sum_{j=1}^{k} p_j(\Delta).$$

The error relative to the values given is estimated to be less than 0·2 %. For $\Delta > 1·5$ the distribution of $\hat{\tau}$ is heavily concentrated at τ: Table 3·4 gives some values of $p_0(\Delta)$ for $1·75 \leqslant \Delta \leqslant 3·00$. When Δ is very large, $\alpha(0,\Delta)$ is close to $\Phi(\Delta)$, as can be deduced from the integral equation (3·3); agreement to three figures occurs for $\Delta > 2·0$. Consequently

$$p_0(\Delta) \sim \{\Phi(\Delta)\}^2$$

for large Δ. For $\Delta < 0·5$ the numerical computation of the asymptotic distribution is difficult to handle. We have not derived the limiting distribution of $\hat{\tau}$ as $\Delta \to 0$.

Table 3·4. *Values of* $p_0(\Delta)$: $\Delta = 1·75\,(0·25)\,3·00$

Δ	1·75	2·00	2·25	2·50	2·75	3·00
$p_0(\Delta)$	0·916	0·953	0·975	0·988	0·994	0·997

3·2. θ_0 *and* θ_1 *unknown*

In many situations the two means θ_0 and θ_1 will be unknown. Here we show that the asymptotic distribution of § 3·1 remains valid. For convenience we assume the variance σ^2 to be known and equal to one. The log likelihood of the observed sequence $(x_1, \ldots, x_T)$ is, apart from a constant,

$$L(x_1, \ldots, x_T|\theta_0, \theta_1, \tau) = -\frac{1}{2}\left\{\sum_{i=1}^{\tau}(x_i-\theta_0)^2+\sum_{i=\tau+1}^{T}(x_i-\theta_1)^2\right\}. \tag{3·12}$$

The maximum likelihood estimators for θ_0 and θ_1 conditional on $\tau = t$ are

$$\tilde{\theta}_{0t} = \frac{1}{t}\sum_{i=1}^{t} x_i = \bar{x}_t,$$

$$\tilde{\theta}_{1t} = \frac{1}{T-t}\sum_{i=t+1}^{T} x_i = \bar{x}_t^*,$$

say. Then the log likelihood, (3·12), maximized over θ_0 and θ_1 conditional on $\tau = t$, becomes

$$L(t) = -\frac{1}{2}\left\{\sum_{i=1}^{T}(x_i-\bar{x}_T)^2 - t(T-t)(\bar{x}_t-\bar{x}_t^*)^2/T\right\},$$

so that $\hat{\tau}$ is the value of t which maximizes the observed value z_t^2 of Z_t^2, say, where

$$Z_t^2 = t(T-t)(\bar{X}_t-\bar{X}_t^*)^2/T \quad (t = 1, \ldots, T-1). \tag{3·13}$$

It is not difficult to see that this is also true when σ^2 is unknown. Note that $\hat{\tau}$ is the value of t giving the most significant difference between conditional estimates of θ_0 and θ_1.

Now suppose $\theta_0 > \theta_1$, then $(\bar{X}_t-\bar{X}_t^*)$ is positive with probability tending to 1 as t and $(T-t)$ increase indefinitely, so that we can take Z_t as asymptotically equivalent to Z_t^2 in order to find the distribution of $\hat{\tau}$. The sequence $\{Z_t\}$ is an autocorrelated sequence with complicated covariance and mean structure, but from (3·13) we obtain the autoregressive representation

$$Z_{t+1} = a_{t+1}Z_t + b_{t+1}(X_{t+1}-\bar{X}_T) \quad (t = 1, \ldots, T-2), \tag{3·14}$$

where

$$a_t = \left\{\frac{(t-1)(T-t+1)}{t(T-t)}\right\}^{\frac{1}{2}},$$

$$b_t = \left\{\frac{T}{t(T-t)}\right\}^{\frac{1}{2}}.$$

We shall show that asymptotically (3·14) defines a random walk with a negligible superimposed term. Let $\tau = \lambda T$ and assume τ and $(T-\tau)$ to be large. Then if $(t-\tau)$ is $o(T)$,

$$\begin{aligned} a_t &= 1 - \frac{(1-2\lambda)}{2\lambda(1-\lambda)T} + o(T^{-1}), \\ b_t &= \{\lambda(1-\lambda)T\}^{-\frac{1}{2}} + o(T^{-1}). \end{aligned} \tag{3·15}$$

It follows from (3·14) and (3·15) that

$$Z_{\tau+1} - Z_\tau = -\frac{(1-2\lambda)}{2\lambda(1-\lambda)T}Z_\tau + \frac{X_{\tau+1}-\bar{X}_T}{\{\lambda(1-\lambda)T\}^{\frac{1}{2}}} + \epsilon, \tag{3·16}$$

where $E(\epsilon) = 0$ and $\operatorname{var}(\epsilon) = O(T^{-2})$, since $\operatorname{var}(Z_\tau) = 1$ and $\operatorname{var}(X_{\tau+1}-\bar{X}_T) = O(1)$. Then writing

$$Z_\tau = E(Z_\tau) + \eta, \quad \bar{X}_T = E(\bar{X}_T) + \xi,$$

it is not difficult to see that (3·16) becomes

$$Z_{\tau+1} - Z_\tau = Y_1 + \epsilon',$$

where Y_1 is normally distributed with mean $-\Delta\{\lambda(1-\lambda)T\}^{-\frac{1}{2}}$ and variance $\{\lambda(1-\lambda)T\}^{-1}$, $E(\epsilon') = o(T^{-1})$ and $\operatorname{var}(\epsilon') = O(T^{-2})$. Similarly we find that, to order T^{-2},

$$Z_{\tau+k} - Z_\tau = \sum_{j=1}^{k} Y_j + k\epsilon' \quad (k = 1, 2, \ldots), \tag{3·17}$$

where the Y_j are identically distributed. Rescaling by a factor $\{\lambda(1-\lambda)T\}^{\frac{1}{2}}$, we find that (3·17) defines a random walk with $N(-\Delta, 1)$ increments plus a linear term whose coefficient has mean $o(T^{-\frac{1}{2}})$ and variance $O(T^{-1})$. A similar representation exists for $(Z_{\tau-k}-Z_\tau)$ $(k = 1, 2, \ldots)$ and we deduce that the asymptotic distribution of $\hat{\tau}$ is determined by the random walks alone; that is, the results of § 3·1 hold.

When θ_0 and θ_1 are unknown, we are also interested in estimating their difference. Since the distribution of $\hat{\tau}$ depends on $\Delta = |\theta_1 - \theta_0|/(2\sigma)$, we are specifically interested in estimating Δ. The maximum likelihood estimate $\hat{\Delta}$ is then given by

$$\hat{\Delta} = |\hat{\theta}_1 - \hat{\theta}_0|/(2\sigma),$$

assuming σ^2 to be known, where $\hat{\theta}_\nu = \tilde{\theta}_{\nu\hat{\tau}}$ ($\nu = 0, 1$). Now the distribution of $(\hat{\tau} - \tau)$ has moments of order 1, therefore it follows that asymptotically

$$\hat{\Delta} = |\tilde{\theta}_{0\tau} - \tilde{\theta}_{1\tau}|/(2\sigma),$$

and hence that $\hat{\Delta}$ is asymptotically $N(\Delta, \sigma_\Delta^2)$, with

$$\sigma_\Delta^2 = \frac{T}{4\tau(T-\tau)}.$$

For finite T, $\hat{\Delta}$ will have a positive bias because $\hat{\tau}$ is obtained by finding the most significant difference $(\tilde{\theta}_{0t} - \tilde{\theta}_{1t})$. In fact it is not difficult to see from the random-walk representation of $\{Z_t\}$ that $E(\hat{\theta}_0 - \hat{\theta}_1) = \theta_0 - \theta_1 + \{\lambda(1-\lambda)T\}^{-1} E\{\max(M, M')\} + o(T^{-1})$, M and M' both having the distribution $\alpha(x, \Delta)$ as in § 3·1. Thus the bias in $\hat{\Delta}$ to order T^{-1} can be computed.

3·3. θ_0 *known and* θ_1 *unknown*

To complete the analysis of the normal case we look at the situation where one of the means is known, but the other unknown. In practice it will usually happen that θ_0 is known rather than θ_1, therefore we assume that this is so. Then to find the maximum likelihood estimate $\hat{\tau}$ we note that the log likelihood of τ is (3·12) with $\tilde{\theta}_{1\tau} = \bar{x}_\tau^*$ replacing θ_1. Hence $\hat{\tau}$ is the value of t which maximizes

$$S_t^2 = t(\bar{X}_t^* - \theta_0)^2 \quad (t = 1, \ldots, T-1),$$

corresponding to Z_t^2 in § 3·2. For $\theta_1 > \theta_0$, $\hat{\tau}$ is asymptotically equivalent to the value of t which maximizes S_t, which we can write as

$$S_t = t^{-\frac{1}{2}} \sum_{j=T-t+1}^{T} (X_j - \theta_0).$$

Then by arguments similar to those of § 3·2 we again find that the asymptotic distribution of $\hat{\tau}$ is that of § 3·1.

In this case the maximum likelihood estimate $\hat{\Delta}$ is given by

$$|\theta_0 - \tilde{\theta}_{1\hat{\tau}}|/(2\sigma)$$

when σ is known, and $\hat{\Delta}$ is asymptotically $N(\Delta, \sigma_\Delta^2)$ with

$$\sigma_\Delta^2 = \frac{1}{4(T-\tau)}.$$

4. INFERENCE ABOUT τ

In most situations we not only want to estimate the change-point, but also to make inference about it in the form of a confidence interval or a significance test. For convenience assume that we want to test $H_0^*: \tau = \tau_0$ with either the one-sided alternative $H_1^*: \tau > \tau_0$ or the two-sided alternative $H_2^*: \tau \neq \tau_0$. If θ_0 and θ_1 are known, we can use the distribution of $\hat{\tau}$ to calculate the significance of a sample value of $\hat{\tau}$. The likelihood ratio test, however, is easier to apply and is more efficient because $\hat{\tau}$ is not sufficient, even asymptotically. Only the observations themselves are sufficient and the likelihood ratio test uses all the information.

4·1. *General results*

Consider the two-sided test of H_0^* with alternative H_2^*, and assume θ_0 and θ_1 to be known. Then in the notation of § 2 the likelihood ratio test statistic Λ_2, say, is given by

$$\Lambda_2 = L(\hat{\tau}) - L(\tau_0) = \sum_{j=1}^{\hat{\tau}} U_j - \sum_{j=1}^{\tau_0} U_j. \tag{4·1}$$

We reject H_0^* when the sample value of Λ_2 is greater than l, where l is determined by the required test size. Under the null hypothesis H_0^* we see by (4·1) and the definitions of M and M' in § 2 that

$$\Lambda_2 = \max(M, M')$$

and hence the asymptotic distribution of Λ_2 under H_0^* is

$$P_2(x) = \mathrm{pr}(\Lambda_2 \leqslant x) = \alpha(x)\,\alpha'(x), \tag{4·2}$$

since M and M' are independent. This distribution is easier to compute than that of $\hat{\tau}$ since it only requires solution of the integral equations for $\alpha(x)$ and $\alpha'(x)$.

For the one-sided test of H_0^* with alternative H_1^* the log likelihood ratio test statistic is

$$\Lambda_1 = \max_{t \geqslant \tau_0}\left(\sum_{i=1}^{t} U_i\right) - \sum_{i=1}^{\tau_0} U_i, \tag{4·3}$$

which under H_0^* becomes $\Lambda_1 = M'$. Hence the asymptotic distribution of Λ_1 is

$$P_1(x) = \mathrm{pr}(\Lambda_1 \leqslant x) = \alpha'(x). \tag{4·4}$$

4·2. *The normal case*

For the normal case discussed in § 3 with θ_0, θ_1 and σ^2 known the increments U_i are given by (3·1). Then using the notation of § 3 we find that (4·1) and (4·2) become

$$\Lambda_2 = \left((\theta_0 - \theta_1)\left[\sum_{i=1}^{\hat{\tau}} \{X_i - \tfrac{1}{2}(\theta_0 + \theta_1)\} - \sum_{i=1}^{\tau_0} \{X_i - \tfrac{1}{2}(\theta_0 + \theta_1)\}\right]\right)\Big/\sigma^2, \tag{4·5}$$

$$P_2(x, \Delta) = [\alpha\{x/(2\Delta), \Delta\}]^2.$$

Similarly (4·3) and (4·4) become

$$\Lambda_1 = \left((\theta_0 - \theta_1)\left[\max_{t \geqslant \tau_0} \sum_{i=1}^{t} \{X_i - \tfrac{1}{2}(\theta_0 + \theta_1)\} - \sum_{i=1}^{\tau_0} \{X_i - \tfrac{1}{2}(\theta_0 + \theta_1)\}\right]\right)\Big/\sigma^2, \tag{4·6}$$

$$P_1(x, \Delta) = \alpha\{x/(2\Delta), \Delta\}.$$

From computed values of $\alpha(x, \Delta)$, we can easily obtain quantiles of $P_1(x, \Delta)$ and $P_2(x, \Delta)$, which we have done for $\Delta = 0·5(0·1)\,1·5$. Table 4·1 gives the 95, 98 and 99 % quantiles of both $P_1(x, \Delta)$ and $P_2(x, \Delta)$, with the notation

$$P_\nu(l_{\nu,p}, \Delta) = p/100 \quad (\nu = 1, 2).$$

If either or both of θ_0 and θ_1 are unknown and the variance σ^2 is known, the asymptotic distributions $P_1(x, \Delta)$ and $P_2(x, \Delta)$ are still appropriate. The log likelihood ratio statistics for θ_0 and θ_1 both unknown are

$$\Lambda_1 = \frac{1}{2\sigma^2} \max_{t \geqslant \tau_0} (Z_t^2 - Z_{\tau_0}^2),$$

$$\Lambda_2 = \frac{1}{2\sigma^2} (Z_{\hat{\tau}}^2 - Z_{\tau_0}^2),$$

with

$$Z_t^2 = t(T-t)(\bar{X}_t - \bar{X}_t^*)^2/T$$

as in § 3·2. Similarly if θ_0 is known but θ_1 unknown, we hav

$$\Lambda_1 = \frac{1}{2\sigma^2} \max_{t \geqslant \tau_0} (S_t^2 - S_{\tau_0}^2),$$

$$\Lambda_2 = \frac{1}{2\sigma^2} (S_{\hat{\tau}}^2 - S_{\tau_0}^2),$$

with

$$S_t^2 = t(\bar{X}_t^* - \theta_0)^2$$

as in § 3·3. The asymptotic distributions $P_1(x, \Delta)$ and $P_2(x, \Delta)$ for Λ_1 and Λ_2 can be verified by arguments similar to those of § 3·2.

Table 4·1. *The* 95, 98 *and* 99 % *quantiles for statistics* Λ_1 *and* Λ_2; $\Delta = 0{\cdot}5\,(0{\cdot}1)\,1{\cdot}5$

Δ	$l_{1,95}$	$l_{1,98}$	$l_{1,99}$	$l_{2,95}$	$l_{2,98}$	$l_{2,99}$
0·5	2·42	3·32	4·02	3·09	4·02	4·72
0·6	2·32	3·21	3·91	2·98	3·91	4·60
0·7	2·20	3·10	3·80	2·88	3·79	4·48
0·8	2·08	2·99	3·69	2·77	3·69	4·37
0·9	1·94	2·88	3·58	2·66	3·58	4·27
1·0	1·79	2·76	3·47	2·53	3·48	4·17
1·1	1·62	2·63	3·35	2·38	3·37	4·06
1·2	1·41	2·48	3·23	2·22	3·25	3·95
1·3	1·18	2·33	3·11	2·04	3·10	3·84
1·4	0·92	2·13	2·97	1·82	2·94	3·71
1·5	0·62	1·89	2·78	1·59	2·74	3·56

When the variance σ^2 is unknown, the statistics Λ_1 and Λ_2 are defined as above but with σ^2 replaced by its maximum likelihood estimate under the relevant alternative hypothesis, H_1^* or H_2^*. For example, if θ_0 and θ_1 are known, Λ_2 is given by

$$\Lambda_2 = \left((\theta_0 - \theta_1)\left[\sum_{i=1}^{\hat{\tau}} \{X_i - \tfrac{1}{2}(\theta_0 + \theta_1)\} - \sum_{i=1}^{\tau_0} \{X_i - \tfrac{1}{2}(\theta_0 + \theta_1)\}\right]\right) \Big/ \hat{\sigma}^2,$$

where

$$\hat{\sigma}^2 = \frac{1}{T}\left[\sum_{i=1}^{T} (X_i - \theta_1)^2 - (\theta_0 - \theta_1) \sum_{i=1}^{\hat{\tau}} \{X_i - \tfrac{1}{2}(\theta_0 + \theta_1)\}\right].$$

Although we have not discussed $\hat{\sigma}^2$ before, it is clearly a consistent estimate and hence the asymptotic distributions of Λ_1 and Λ_2 are unchanged.

We have shown that the asymptotic distributions of Λ_1 and Λ_2 are always $P_1(x, \Delta)$ and $P_2(x, \Delta)$ as given by (4·5) and (4·6). But if any one of θ_0, θ_1 and σ^2 is unknown the parameter

Δ is unknown, so that we cannot compute the relevant significance probabilities exactly. It is possible to obtain bounds for $P_\nu(x,\Delta)$ $(\nu = 1,2)$ using confidence limits for Δ derived from $\hat{\Delta}$, but in general these bounds are unsatisfactory particularly for tail probabilities. One direct approach is to estimate $P_\nu(x,\Delta)$ using $\hat{\Delta}$. For example, if θ_0 and θ_1 are unknown and σ^2 is known, it is easy to show that an unbiased estimate of $P_2(x,\Delta)$ to order T^{-1} is

$$P_2(x,\hat{\Delta}) - \tfrac{1}{2}\sigma_\Delta^2 \frac{d^2 P_2(x,\hat{\Delta})}{d\Delta^2},$$

where σ_Δ is the variance of $\hat{\Delta}$ under the hypothesis H_0^*. This estimate should have a distribution closer to the uniform distribution on $(0,1)$ than $P_2(x,\hat{\Delta})$. The notion of uniformity stems from the Neyman–Pearson theory, and is a property we want our analogue of $P_2(x,\Delta)$ to have. The problem is similar to that of testing a normal mean with unknown variance, except that here we do not have a pair of sufficient statistics: only the observations themselves are sufficient.

The distributions of Λ_1 and Λ_2 under the alternative $\tau = \tau_1 \neq \tau_0$ have not been derived, although in the normal case it is easy to show that they depend on absorption probabilities in $\mathbf{W}$ and $\mathbf{W}'$ with nontrivial boundaries. We have not obtained a solution suitable for computation.

5. Empirical results

To see how well the asymptotic results of §§ 3 and 4 work in finite sample situations, we carried out an extensive empirical study of the distributions of $\hat{\tau}$, $\hat{\Delta}$ and Λ_2. A summary of these results is given here. For several values of T, τ and Δ, we generated 500 samples of observations using pseudo random normal deviates for the error terms ϵ_i in the model

$$X_i = \begin{cases} \theta_0 + \epsilon_i & (i = 1, \ldots, \tau), \\ \theta_1 + \epsilon_i & (i = \tau+1, \ldots, T). \end{cases}$$

To see how the finite sample distributions vary according as θ_0 and/or θ_1 are known or unknown, we calculated the empirical distributions under the three assumptions: (i) θ_0 and θ_1 both unknown; (ii) θ_0 known, θ_1 unknown; (iii) θ_0 and θ_1 both known, in the same samples. The estimate $\hat{\Delta}$ is redundant under assumption (iii).

Inspection of the empirical results showed that agreement between empirical and asymptotic distributions depends on the value of $D = \Delta/\sigma_\Delta$ when Δ is unknown. If D is large the two means are easily distinguished, in which case we would expect the asymptotic properties of $\hat{\tau}$ and Λ_2 to be good approximations. Table 5·1 contains the empirical and asymptotic means and variances of $\hat{\tau}$ and $\hat{\Delta}$ for four cases with $\Delta = 0{\cdot}5$. On the basis of all the empirical results it seems safe to take cases with D greater than 6 as well defined in the sense of practical validity of the asymptotic results. For example, the ill-defined case $T = 50$ and $\tau = 25$ with θ_1 unknown in Table 5·1 gives variances of $\hat{\tau}$ and $\hat{\Delta}$ which are appreciably larger than the asymptotic values. Probability plots of the asymptotic and empirical distributions reinforce these conclusions. When both θ_0 and θ_1 are known the agreement between empirical and asymptotic distributions is good provided that the asymptotic distribution of $\hat{\tau}$ is concentrated between 1 and T.

Table 5·1. *Comparison of empirical and asymptotic moments of $\hat{\tau}$ and $\hat{\Delta}$.*
(i) θ_0 *and* θ_1 *unknown*, (ii) θ_0 *known*, θ_1 *unknown, and* (iii) θ_0 *and* θ_1 *known.* $\Delta = 0·5$

Case	Empirical $E(\hat{\tau})$	Empirical Var $(\hat{\tau})$	Asymptotic Var $(\hat{\tau})$	Empirical $E(\hat{\Delta})$	Empirical Var $(\hat{\Delta})$	Asymptotic Var $(\hat{\Delta})$
$T = 50, \tau = 15$						
(i) $D = 3·2$	20·3	83·6	24·1	0·56	0·054	0·024
(ii) $D = 5·9$	14·8	19·5	24·1	·50	·0095	·0072
(iii)	15·0	23·2	24·1	—	—	—
$T = 50, \tau = 25$						
(i) $D = 3·5$	25·5	41·7	24·1	·59	·037	·020
(ii) $D = 5·0$	24·8	31·4	24·1	·51	·015	·010
(iii)	25·4	25·5	24·1	—	—	—
$T = 100, \tau = 25$						
(i) $D = 4·3$	25·4	60·5	24·1	·56	·019	·013
(ii) $D = 8·6$	25·3	25·0	24·1	·51	·0035	·0033
(iii)	25·3	21·6	24·1	—	—	—
$T = 200, \tau = 50$						
(i) $D = 6·1$	50·4	31·1	24·1	·51	·0065	·0067
(ii) $D = 12·3$	50·0	26·3	24·1	·50	·0018	·0017
(iii)	50·0	24·3	24·1	—	—	—

6. FURTHER DEVELOPMENTS

The empirical results described in § 5 show that the asymptotic distributions of $\hat{\tau}$, Λ_1, Λ_2 and $\hat{\Delta}$ are poor approximations in small samples, or more precisely in ill-defined cases. Analysis of the finite sample distributions is needed, but appears to be difficult. For θ_0 and θ_1 known, a finite sample version of (2·13) exists but is very complicated; see Feller (1966, Chap. 18). In the normal case with unknown means (§ 3·2) the exact finite sample distributions involve extremely complicated multivariate normal integrals.

It is not difficult to verify that the results of §§ 3 and 4 also apply to inference about τ in the parallel line regression model

$$X_t = \alpha_0 + \beta v_t + \epsilon_t \quad (t = 1, \ldots, \tau),$$
$$X_t = \alpha_1 + \beta v_t + \epsilon_t \quad (t = \tau + 1, \ldots, T),$$

where z is an independent variable and the ϵ_t are independent $N(0, \sigma^2)$. For if α_0, α_1, β and σ^2 are known, the log likelihood increments U_i corresponding to (3·1) are given by

$$U_i = [(\alpha_1 - \alpha_0)\{X_i - \tfrac{1}{2}(\alpha_0 + \alpha_1) - \beta v_i\}]/\sigma_2$$

and the asymptotic results of §§ 3·1 and 4·2 hold with $\Delta = |\alpha_1 - \alpha_0|/(2\sigma)$. If α_0, α_1 and β are unknown, the results of §§ 3·2 and 4·2 hold with

$$Z_t^2 = [t(T-t)\{\bar{X}_t - \bar{X}_t^* - \tilde{\beta}_t(\bar{v}_t - \bar{v}_t^*)\}^2]/T,$$

where $\tilde{\beta}_t$ is the maximum likelihood estimate of β conditional on $\tau = t$; see, for example, (3·13).

The arguments of §§ 2 and 3 are directly relevant to the asymptotic distribution of the cumulative sum scheme estimate of τ, which is the significant turning-point in the cumulative sum plot (Page, 1957). A detailed account of this problem is given in the author's unpublished Ph.D. thesis and will be published elsewhere.

The results of § 2 are, of course, applicable to nonnormal distributions. In fact they can be generalized to include discrete distributions, the simplest case being that of binary data with changing binomial parameter. Nor are the arguments of § 2 restricted to problems with one distribution $f(x, \theta)$. It is just as easy to deal with the case of a change in distribution from $f_0(x, \theta)$ to $f_1(x, \phi)$.

Thanks are particularly due to Professor D. R. Cox for suggesting this research as part of my Ph.D. thesis, and for his constant encouragement. The computational results owe much to the excellent work of Jane Gentleman and Elizabeth Chambers. The research was financially supported by the Science Research Council and Shell International Petroleum Company.

References

Bhattacharyya, G. K. & Johnson, R. A. (1968). Non-parametric tests for shift at an unknown time point. *Ann. Math. Statist.* **39**, 1731–43.

Chernoff, H. & Zacks, S. (1964). Estimating the current mean of a normal distribution which is subjected to changes in time. *Ann. Math. Statist.* **35**, 999–1018.

Feller, W. (1966). *An Introduction to Probability Theory and its Applications*, vol. II. New York: Wiley.

Hinkley, D. V. (1969). Inference about the intersection in two-phase regression. *Biometrika* **56**, 495–504.

Hudson, D. J. (1966). Fitting segmented curves whose joint points have to be estimated. *J. Am. Statist. Ass.* **61**, 1097–129.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika* **41**, 100–14.

Page, E. S. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika* **42**, 523–7.

Page, E. S. (1957). On problems in which a change in parameter occurs at an unknown point. *Biometrika* **44**, 248–52.

[*Received June* 1969]