PROBLEMS ASSOCIATED WITH POINTS OF DISCONTINUITY

by

DAVID VICTOR HINKLEY

Thesis submitted for the degree of
Ph.D. in the University of London

Imperial College, 1969.

ABSTRACT

The thesis, which is in two parts, is concerned with two particular statistical models in which a change occurs at an unknown point.

The first model is one for a sequence of independent random variables $x_1, x_2, \ldots, x_T$ with a distributional parameter θ which changes from θ_0 to θ_1 at an unknown point in the sequence. The asymptotic distribution of the maximum likelihood estimate of the change-point is derived and discussed in detail for sequences of normally distributed variables with mean θ. The asymptotic distribution of the likelihood ratio test statistic for testing hypotheses about the change-point is also derived. We also discuss the estimation of the change-point in Cumulative Sum schemes. The asymptotic results are compared with finite sample simulation results.

In the second part of the thesis we look at a simple linear regression model where the slope changes at an unknown value of the independent variable. A procedure for calculating the maximum likelihood estimate of the change-point is given, and an asymptotic distribution of the estimate is derived which is a good approximation in finite samples. Inference about the change-point is also discussed. The results are compared with finite sample simulation results.

ACKNOWLEDGEMENTS

I am most grateful to my supervisor, Professor David R. Cox, for suggesting the research described in this thesis, and for his constant interest and guidance. I also wish to thank Mrs. Betty Chambers and Mrs. Jane Gentleman for their advice on computational problems, and Miss Helen MacKenzie for her excellent typing of the manuscript.

The work was financially supported by the Science Research Council and Shell International Petroleum Company Ltd.

CONTENTS

1. Introduction

Given an observed sequence of independent random variables $(x_1, x_2, \ldots, x_T)$, say, there are available many techniques for detecting underlying patterns or relationships. Such patterns often fall into the categories of linear regression, polynomial regression and cyclic trends, all of which are useful descriptively and have simple statistical techniques associated with them. Common to nearly all of these, however, is an assumption that a single model holds for the whole range of the sample. More specifically it is usual to fit a model of the form

$$x_t = \theta(t) + \varepsilon_t \qquad (t = 1, 2, \ldots, T)$$

where $\{\varepsilon_t\}$ is a sequence of uncorrelated error terms with zero mean, often normally distributed. We can often think of these observations as being a discrete sample from a continuous process $x(t)$, so that $\theta(t)$ is a continuous function of t.

In application, however, a single trend or mean is sometimes inappropriate. It may be strongly suspected from prior experimentation or study that the model valid near t = 1 is invalid in the neighbourhood of t = T. This leads us to consider fitting models such as

$$x_t = \theta_0(t) + \varepsilon_t \qquad (t = 1, 2, \ldots, \tau)$$

$$x_t = \theta_1(t) + \varepsilon_t \qquad (t = \tau + 1, \ldots, T), \qquad (1.1)$$

where τ is unknown and $\{\epsilon_t\}$ is a sequence of uncorrelated zero-mean error terms. Experimental and scientific grounds may force us to consider (1.1) or even its generalization to p + 1 sub-models with mean functions $\theta_o(t), \ldots, \theta_p(t)$ and change-points $\tau_1, \tau_2, \ldots, \tau_p$. Scientific literature contains several such cases, for example the analysis of discontinuities in intermolecular activation energy by Ahsanullah and Qurashi (1965) where the energy level apparently exhibited distinct jumps at specific temperatures for various liquids.

Two simple but interesting and important special cases of (1.1) are

Model A:

$$\begin{aligned} x_t &= \theta_o + \epsilon_t \qquad (t = 1, \ldots, \tau), \\ x_t &= \theta_1 + \epsilon_t \qquad (t = \tau + 1, \ldots, T), \end{aligned} \tag{1.2}$$

and the regression model,

Model B:

$$\begin{aligned} x_t &= \alpha + \beta_o(u_t - \gamma) + \epsilon_t \quad (t = 1, \ldots, \tau) \\ x_t &= \alpha + \beta_1(u_t - \gamma) + \epsilon_t \quad (t = \tau + 1, \ldots, T), \end{aligned} \tag{1.3}$$

where $u_\tau \leq \gamma < u_{\tau+1}$.

In both Models A and B it is usual to assume the error terms ϵ_t to be $N(0, \sigma^2)$.

Model A would apply in a continuous inspection scheme where the purpose is to detect a change in the mean θ of the process and, further, to estimate where the change took place. The emphasis of work in this field, in particular that of Page (1954, 1955 and 1957) on cumulative sum techniques, has been on testing the null hypothesis $\theta_o = \theta_1$. Chernoff and Zacks (1965) and Bhattacharyya and Johnson (1968) have discussed the same problem within a Bayesian framework.

The regression model B is relevant in some scientific situations where a change in the relationship between two quantities X and U occurs at some threshold $U = \gamma$. Often β_o or β_1 will be zero, corresponding to a notion of inactivity or saturation. This situation has been analyzed in some detail by Hudson (1966) and recently by Feder and Sylvester (1968).

In Part I we consider initially a generalization of Model A to arbitrary error distributions. We are concerned primarily with the estimation of τ, not with the null hypothesis $\theta_o = \theta_1$. This has some justification in that the hypothesis $\theta_o = \theta_1$ will often be replaced by a more specific hypothesis $\tau = \tau_o$, a test of which involves the estimate $\hat{\tau}$ of τ, which in our case is the m.l.e. (maximum likelihood estimate).

Since we are dealing with constant means θ_o and θ_1, it is advantageous to consider general probability distributions for the random variables $x_1, \ldots, x_T$. At first we assume θ_o and θ_1 to be known in order to establish a clear approach to the problem of the distribution of $\hat{\tau}$, but later we drop the assumption and show that the asymptotic distribution of $\hat{\tau}$ is unchanged. (All the theoretical results concern asymptotic distributions, that is distributions as τ and $T - \tau$ increase indefinitely). In Section 3 we discuss problems arising in the computation of the asymptotic distribution in the normal case, which has no explicit form. The remaining sections of Part I also deal with the special case of normally distributed random variables, including a discussion on the application of the results to cumulative sum schemes (Section 5). In Section 4 we look at the problem of inference about τ when θ_o and θ_1 are known or unknown.

In Part II of the thesis we examine the estimation of γ in Model B by maximum likelihood. A concise procedure for computing the m.l.e. $\hat{\gamma}$ is constructed in Section 9, which is basically a likelihood search: the procedure is similar to that of Hudson (1966). We find that unconstrained least squares estimates of γ conditional on $\tau = t$ provide a great deal of information about $\hat{\gamma}$ which is useful in finding an asymptotic distribution for $\hat{\gamma}$ (Section 10). An empirical study of the finite sample use of this asymptotic distribution is described in Section 11. We then discuss the problems of inference about $\beta = (\beta_1 - \beta_o)/\sigma$ and γ (Sections 12 and 13). This is followed by a brief look at further developments. Section 15 includes a detailed examination of the distribution of ratios of normal variates and an approximation to this distribution used in Section 10.

PART I

ESTIMATION OF THE CHANGE-POINT IN THE DISTRIBUTION OF A SEQUENCE OF RANDOM VARIABLES

2. Distribution of the m.l.e.; θ_o and θ_1 Known

In this section we derive the m.l.e. $\hat{\tau}$ for a general p.d.f. $f(x, \theta)$ of the random variables x_t and find its asymptotic distribution (that is for τ and $T - \tau$ indefinitely large). No explicit form for $\hat{\tau}$ exists, but it is conveniently defined by variables associated with two random walks which represent the log likelihood. Random walk results are then used to obtain the asymptotic distribution of $\hat{\tau}$ in a form suitable for computation.

Let $x_1, x_2, \ldots, x_T$ be a sequence of independent continuous random variables such that

x_i has p.d.f. $f(x, \theta_o)$ for $i = 1, \ldots, \tau$

and x_i has p.d.f. $f(x, \theta_1)$ for $i = \tau + 1, \ldots, T$, $(\theta_o \neq \theta_1)$,

where θ_o and θ_1 are known but τ is unknown.

To obtain the m.l.e. $\hat{\tau}$ from a sample $(x_1, \ldots, x_T)$ we have to maximize the log likelihood

$$L(t) = L(x_1, \ldots, x_T | \theta_o, \theta_1, t) = \sum_{j=1}^{t} \log f(x_j, \theta_o) + \sum_{j=t+1}^{T} \log f(x_j, \theta_1) \tag{2.1}$$

over admissible values of t (i.e. $t = 1, 2, \ldots, T - 1$ since we assume that at least one observation comes from each distribution).

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

$$u_i = \log f(x_i, \theta_o) - \log f(x_i, \theta_1).$$

Then (2.1) becomes

$$L(t) = \sum_{i=1}^{t} u_i + \sum_{i=1}^{T} \log f(x_i, \theta_1).$$

Since the second term on the right hand side is independent of t, the m.l.e. $\hat{\tau}$ will be the value of t which maximizes the sequence of partial sums

$$V_t = \sum_{i=1}^{t} u_i \qquad (t = 1, \ldots, T - 1) .$$

Now the u_i's are independent, because of the independence of the x_i's, so that the sequence $\{V_t\}$ defines two independent random walks, namely

$$W = \{0, V_{\tau-1} - V_\tau, \ldots, V_{\tau-k} - V_\tau, \ldots, V_1 - V_\tau\}$$

and

$$W' = \{0, V_{\tau+1} - V_\tau, \ldots, V_{\tau+k} - V_\tau, \ldots, V_{T-1} - V_\tau\} \quad ,$$

or alternatively

$$W = \{0, -\sum_{j=0}^{k} u_{\tau-j} : k = 0, 1, \ldots, \tau-1\}$$

and

$$W' = \{0, \sum_{j=0}^{k} u_{\tau+1+j} : k = 0, 1, \ldots, T-\tau-2\} . \qquad (2.2)$$

Each random walk has independent identically distributed increments with negative means. Both random walks stem from the point (τ, V_τ):W represents the log likelihood for integers less than τ, and W′ likewise for integers greater than τ relative to the log likelihood of τ. To maximize L(t) we must find the larger of the two random walk maxima. If each maximum is zero, clearly $\hat{\tau} = \tau$.

We introduce the following notation:

Let $y_j = -u_{\tau-j+1}$ and $z_j = u_{\tau+j}$ $(j = 1, 2, \ldots)$

so that (2.2) becomes

$$W = \{0, y_1, y_1 + y_2, \ldots, \sum_{k=1}^{\tau-1} y_k\}$$

and
$$W' = \{0, z_1, z_1 + z_2, \ldots, \sum_{k=1}^{T-\tau-1} z_k\} . \tag{2.3}$$

Also, let

$$S = \max[y_1, y_1 + y_2, \ldots, \sum_{k=1}^{\tau-1} y_k]$$

and
$$S' = \max[z_1, z_1 + z_2, \ldots, \sum_{k=1}^{T-\tau-1} z_k].$$

Note that these are not quite the random walk maxima, which we shall denote by

$$M = \max (0, S) \quad \text{and} \quad M' = \max (0, S').$$

The finite sample distribution of $\hat{\tau} - \tau$ depends implicitly on τ and $T - \tau$ because S and S' do, but here we shall assume both τ and $T - \tau$ to be infinitely large and derive the asymptotic distribution of $\hat{\tau}$.

We can now express events involving $\hat{\tau}$ in terms of events involving S and S': we have seen already that $\hat{\tau} = \tau$ is equivalent to $S < 0$ and $S' < 0$. Further $\hat{\tau} = \tau + k$ is equivalent to $S' = z_1 + \ldots + z_k > 0$ and $S' > S$; and $\hat{\tau} = \tau - k$ is equivalent to $S = y_1 + \ldots + y_k > 0$ and $S > S'$. (The events $S = 0$, $S' = 0$

and S = S′ have zero probability because the random variables are continuous). To simplify the analysis let

$$I = \inf \{ j : S = \sum_{i=1}^{j} y_i \} , \quad \text{where } I = 0 \text{ if } S < 0$$

and

$$I' = \inf \{ j : S' = \sum_{i=1}^{j} z_i \} , \quad \text{where } I' = 0 \text{ if } S' < 0.$$

Then it follows that

$$\begin{aligned} pr(\hat{\tau} = \tau) &= pr(I = I' = 0) = pr(S < 0)\ pr\ (S' < 0) \\ &= \alpha(0)\ \alpha'(0) , \end{aligned} \qquad (2.4)$$

say, and also that

$$pr(\hat{\tau} = \tau + k) = pr(I' = k,\ S' > S,\ S' > 0)$$

and $$pr(\hat{\tau} = \tau - k) = pr(I = k,\ S > S',\ S > 0). \qquad (2.5)$$

Now define

$$\left.\begin{aligned} \beta_k(x)dx &= pr(I = k,\ x \le S < x + dx) \\ \beta_k'(x)dx &= pr(I' = k,\ x \le S' < x + dx) \end{aligned}\right\} \quad (k \ge 1,\ x > 0) \qquad (2.6)$$

and $$\begin{aligned} \alpha\ (x) &= pr(S < x) \\ \alpha'(x) &= pr(S' < x) . \end{aligned} \qquad (2.7)$$

The use of primes here distinguishes properties of W and W′ and does not mean derivatives.

Then (2.5) becomes

$$pr(\hat{\tau} = \tau + k) = \int_0^{\infty} \beta_k'(x)\alpha(x)dx \qquad (2.8a)$$

$$\mathrm{pr}(\hat{\tau} = \tau - k) = \int_0^\infty \beta_k(x)\alpha'(x)dx \; . \tag{2.8b}$$

It is easy to verify by probabilistic arguments that $\alpha(\cdot)$, $\alpha'(\cdot)$, $\beta_k(\cdot)$ and $\beta_k'(\cdot)$ satisfy integral equations: let $g(\cdot)$ and $g'(\cdot)$ be the p.d.f.'s of y_i and z_j respectively, then

$$\alpha(x) = \int_{-\infty}^{x} \alpha(x - y)g(y)dy \quad ,$$

$$\alpha'(x) = \int_{-\infty}^{x} \alpha'(x - y)g'(y)dy \quad , \tag{2.9}$$

$$\beta_{k+1}(x) = \int_0^\infty \beta_k(y)\; g(x - y)dy \qquad (k \geq 0,\; x > 0)$$

and $$\beta_{k+1}'(x) = \int_0^\infty \beta_k'(y)g'(x - y)dy \qquad (k \geq 0,\; x > 0), \tag{2.10}$$

where we define $\beta_o(x) = \alpha(0)\varepsilon(x)$ and $\beta_o'(x) = \alpha'(0)\varepsilon(x)$ with $\varepsilon(0) = 1$, $\varepsilon(x) = 0$ $(x > 0)$.

Now (2.9) and (2.10) do not have explicit solutions, so we need to find a method of calculating the integrals in (2.8a, b) that does not involve explicit solution for $\alpha(x)$, $\beta_k(x)$, $\alpha'(x)$ and $\beta_k'(x)$. We make use of the following theorem, given by Feller (1966, chapter 18) and proved originally by Spitzer.

<u>Theorem</u> Let $y_1, y_2, \ldots$ be a sequence of independent, identically distributed random variables with probability distribution function $G(y)$ and define $M = \max\{0, \sum_{j=1}^{n} y_j, n \geq 1\}$ and

$I = \inf\{k : M = \sum_{j=1}^{k} y_j\}$, where $I = 0$ if and only if $M = 0$.

If
$$\sum_{k=1}^{\infty} k^{-1} \, pr\left(\sum_{j=1}^{k} y_j > 0\right) < \infty,$$

then
$$E\left(s^I e^{-\omega M}\right) = \exp\left\{\sum_{k=1}^{\infty} \frac{s^k}{k} c_k(\omega) - \sum_{k=1}^{\infty} \frac{1}{k} c_k(0)\right\} , \quad (2.11)$$

where
$$c_k(\omega) = \int_0^{\infty} e^{-\omega x} \, pr\left(x \leq \sum_{j=1}^{k} y_j < x + dx\right) \qquad (k \geq 1), \quad (2.12)$$

and $Re(\omega) > 0$.

It follows from (2.11) that

$$\int_0^{\infty} e^{-\omega x} \, d\alpha(x) = E(e^{-\omega M}) = \exp\left\{\sum_{k=1}^{\infty} \frac{c_k(\omega) - c_k(0)}{k}\right\} , \quad (2.13)$$

$$\sum_{k=0}^{\infty} s^k \, pr\,(I=k) = \exp\left\{\sum_{k=1}^{\infty} \frac{s^k - 1}{k} c_k(0)\right\} \quad (2.14)$$

and
$$\alpha(0) = \exp\left\{-\sum_{k=1}^{\infty} \frac{c_k(0)}{k}\right\} . \quad (2.15)$$

Note that the Laplace transform of $\beta_k(x)$ is the coefficient of s^k

in (2.11). Corresponding to the results (2.11) - (2.15) are similar results for $\alpha'(\cdot)$, $\beta_k'(\cdot)$ with

$$c_k'(\omega) = \int_0^\infty e^{-\omega x}\, pr(x \le \sum_{j=1}^{k} z_j < x + dx)$$

replacing $c_k(\omega)$; we shall refer to these as (2.11′) - (2.15′).

Now from (2.8a) and (2.4) we have

$$\begin{aligned}\sum_{k=o}^{\infty} s^k pr(\hat{\tau} = \tau + k) &= \int_0^\infty \left\{ \sum_{k=o}^{\infty} s^k\, \beta_k'(x) \right\} \alpha(x)dx \\ &= \sum_{k=o}^{\infty} s^k \int_0^\infty \beta_k'(x)dx - \int_0^\infty \left\{ \sum_{k=o}^{\infty} s^k \beta_k'(x) \right\} \{1-\alpha(x)\}dx \\ &= \sum_{k=o}^{\infty} s^k pr(I' = k) - \int_0^\infty \left\{ \sum_{k=o}^{\infty} s^k \beta_k'(x) \right\} \{1-\alpha(x)\}dx \,. \end{aligned} \tag{2.16}$$

A similar expression can be derived from (2.8b) for the generating function of $pr(\hat{\tau} = \tau - k)$, $(k \ge 0)$. Now the first term on the right hand side of (2.16) is given by (2.14′). The Laplace transforms of the functions in the integrand of the second term are given by (2.11′) and (2.13), so that we can rewrite the integral by using Parseval's relation for the integral of the product of two integrable functions.

Then (2.16) becomes, after a little calculation,

$$\sum_{k=o}^{\infty} s^k \mathrm{pr}(\hat{\tau} = \tau + k) = \alpha'(0) \exp\left\{ \sum_{k=1}^{\infty} \frac{s^k}{k} c_k'(0) \right\}$$

$$- \frac{\alpha'(0)}{2\pi} \int_{-i\infty}^{i\infty} \exp\left\{ \sum_{k=1}^{\infty} \frac{s^k}{k} c_k'(\omega)\right\} \left[1-\alpha(0)\exp\left\{ \sum_{k=1}^{\infty} \frac{c_k(\omega)}{k} \right\} \right] \frac{d\omega}{\omega} \cdot \tag{2.17}$$

Unfortunately we have been unable to calculate the integral in (2.17); one difficulty is that the integrand has no poles in the ω plane. We therefore derive a suitable method of approximating to the distribution of $\hat{\tau}$ using the results prior to (2.16). Essentially we want an approximation which gives good numerical results in specific cases, and which is not excessively difficult to compute.

First, we shall find it more convenient to work with $1 - \alpha(x)$ rather than $\alpha(x)$; $1 - \alpha(x)$ is integrable and tends to zero, making asymptotic and numerical computation easier. So we rewrite (2.8a) and (2.9) as

$$\mathrm{pr}(\hat{\tau} = \tau + k) = \int_0^\infty \beta_k'(x)\,dx - \int_0^\infty \{1-\alpha(x)\}\beta_k'(x)\,dx$$

$$= q_k' - \int_0^\infty \{1-\alpha(x)\}\beta_k'(x)\,dx \ , \tag{2.18}$$

with $q_k' = \mathrm{pr}(I' = k)$,

and

$$1-\alpha(x) = \int_x^\infty g(u)du + \int_0^\infty \{1-\alpha(u)\}g(x-u)du \, . \qquad (2.19)$$

Now suppose that we can express the solution of (2.19) as an expansion in terms of exponentials, i.e.

$$1-\alpha(x) = \sum_{r=1}^{R} h_r \exp(-\omega_r x) \qquad (x \geq 0) \, . \qquad (2.20)$$

Then (2.18) becomes

$$\text{pr}(\hat{\tau} = \tau + k) = q_k' - \sum_{r=1}^{R} h_r \tilde{\beta}_k'(\omega_r) \qquad (k \geq 0), \qquad (2.21)$$

where $\tilde{\beta}_k'(\omega) = \int_0^\infty e^{-\omega x} \beta_k'(x)dx \qquad (k \geq 0)$

is the coefficient of s^k in (2.11′). Exact computation of (2.21) is straightforward, in principle, because q_k' and $\tilde{\beta}_k'(\omega)$ both satisfy recurrence relations. Both (2.11′) and (2.14′) are equations of the type

$$\sum_{n=o}^{\infty} \rho_n s^n = A \exp\left(\sum_{m=1}^{\infty} \frac{s^m \sigma_m}{m} \right) ,$$

whose solution is

$$\rho_{n+1} = \frac{1}{n+1} (\sigma_1 \rho_n + \ldots + \sigma_{n+1} \rho_o) \qquad (n \geq 0),$$

with $\rho_o = A$.

Therefore q_k' satisfies

$$q_{k+1}' = \frac{1}{k+1}\{c_{k+1}'(0)q_o' + \dots + c_1'(0)\, q_k'\} \quad (k \geq 0)$$

$$q_o' = \alpha'(0) , \tag{2.22}$$

and $\tilde{\beta}_k'(\omega)$ satisfies

$$\tilde{\beta}_{k+1}'(\omega) = \frac{1}{k+1}\{c_{k+1}'(\omega)\tilde{\beta}_o'(\omega) + \dots + c_1'(\omega)\tilde{\beta}_k'(\omega)\} \quad (k \geq 0)$$

$$\tilde{\beta}_o'(\omega) = \alpha'(0). \tag{2.23}$$

To derive the expansion (2.20) in practice, we first solve a discretized, finite version of (2.19) for numerical values of $1 - \alpha(x)$. That is we take a finite set of x values, $0 = x_o < x_1 < \dots < x_n$, say, with $x_{i+1} - x_i = d$, and rewrite (2.19) as

$$1-\alpha(x_j) = \int_{x_j}^{\infty} g(u)du + d\sum_{i=o}^{n} e_i \{1-\alpha(x_i)\}g(x_j - x_i) \quad (j = 0, 1, \dots n) \tag{2.24}$$

where $e_i = 1$ $(1 \leq i < n)$ and $e_i = \frac{1}{2}$ $(i = 0, n)$. Then (2.24) is a matrix equation of the form $(I - B)\,\xi = \eta$, whose solution is the values $1 - \alpha(x_j)$ $(j = 0, 1, \dots, n)$. The accuracy of the solution of (2.19) increases as d decreases and as x_n increases; both d and x_n should be varied to obtain a reasonably stable solution. Having got numerical values of $1 - \alpha(x)$, we can in general fit a finite number of terms of (2.20) by least squares; in special cases a detailed examination of the values of $1 - \alpha(x)$

may reveal the dominant exponential terms without resort to the least squares method. It is also possible that an asymptotic study of $1 - \alpha(x)$ will reveal useful information. Rather than discuss the generalities of this approximation problem here, we study the particular case of the normal distribution at length in Section 3, where we obtain numerical results for the asymptotic distribution of $\hat{\tau}$.

We end this section by making some general remarks about the distribution of $\hat{\tau}$. First, the condition for (2.11) to hold implies that the moments of $\hat{\tau} - \tau$ are finite only if

$$\sum_{k=1}^{\infty} k^{-1} \operatorname{pr}\left(\sum_{j=1}^{k} y_j > 0 \right) < \infty$$

and similarly for $\{z_j\}$. In terms of $f(x,\theta)$ this condition is

$$\sum_{k=1}^{\infty} k^{-1} \operatorname{pr}\left\{ \prod_{j=1}^{k} \frac{f(X_j,\ \theta_1)}{f(X_j,\ \theta_o)} > 1 \mid X_j \text{ has p.d.f. } f(x,\theta_o) \right\} < \infty \tag{2.25}$$

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

For $\hat{\tau}$ to be symmetrically distributed about τ it is necessary and sufficient that y_j and z_j have the same distribution.

This implies that

$$\text{pr}\{f(X,\theta_o) \le y\, f(X,\theta_1) |\; \theta = \theta_1\} = \text{pr}\{f(X,\theta_1) \le y\, f(X,\theta_o) | \theta = \theta_o\} \tag{2.26}$$

for all $y \ge 0$. It is easy to verify that (2.26) holds when $f(x,\theta)$ is symmetric and θ is a location parameter only, or if any one-one continuous transformation of X has such a distribution. For example, it holds for normal and log normal distributions with θ as the mean of the normal distribution, but not for the exponential density $\theta \exp(-\theta x)$.

3. Distribution of the m.l.e. in the Normal Case

In this section we study the distribution of $\hat{\tau}$ when θ_o and θ_1 are the mean values of a sequence of normally distributed random variables. First we assume θ_o and θ_1 to be known, and derive a method for computing the asymptotic distribution of $\hat{\tau}$. Then we show that the asymptotic distribution is unchanged when θ_o and θ_1 are unknown.

3.1. θ_o and θ_1 known

Let the observations $(x_1, \ldots, x_T)$ have the probability density function

$$f(x,\theta) = \frac{1}{\sigma\sqrt{2\pi}} \exp\{-(x - \theta)^2/(2\sigma^2)\} ,$$

and assume that θ_o, θ_1 and σ^2 are known. Then following the results of Section 2, we see first that the log likelihood increments u_i are given by

$$u_i \;=\; \{x_i - \tfrac{1}{2}(\theta_o + \theta_1)\}\,(\theta_o - \theta_1)/\sigma^2 \;.$$

Define $\Delta = |\theta_1 - \theta_o|/(2\sigma)$, then it follows that the increments y_i and z_i of the random walks W and W$'$ are $N(-2\Delta^2, 4\Delta^2)$. Clearly the distribution of $\hat{\tau}$ is unchanged if y_i and z_i are rescaled to have unit variance, so we can take these increments as being $N(-\Delta, 1)$ without loss of generality. Since y_i and z_i have the same distribution, $\hat{\tau}$ will be symmetrically distributed about τ. We need then consider the distribution of $\hat{\tau}$ only for $\hat{\tau} \geq \tau$.

To emphasize the dependence of the distribution on Δ, we denote $\alpha(\cdot)$, $\beta_k(\cdot)$, ... by $\alpha(\cdot,\Delta)$, $\beta_k(\cdot,\Delta)$, ...; this is particularly important when we come to Section 5. We can omit the now redundant prime superfix used in Section 2. The equations defining the distribution of $\hat{\tau}$ in this case are, from (2.18), (2.4), (2.19), (2.22), (2.23) and (2.15) respectively,

$$\mathrm{pr}(\hat{\tau} = \tau \pm k) = q_k(\Delta) - \int_0^\infty \{1-\alpha(u,\Delta)\}\, \beta_k(u,\Delta)du \qquad (k \geq 1)$$

$$\mathrm{pr}(\hat{\tau} = \tau) \;=\; \{\alpha(0,\Delta)\}^2 \;, \qquad (3.1)$$

$$1-\alpha(x,\Delta) = 1 - \Phi(x+\Delta) + \int_0^\infty \{1-\alpha(u,\Delta)\}\varphi(x-u+\Delta)du, \qquad (3.2)$$

$$\tilde{\beta}_{k+1}(w,\Delta) = \frac{1}{k+1}\sum_{j=o}^{k}\tilde{\beta}_j(w,\Delta)c_{k+1-j}(w,\Delta) \qquad (k \geq 1)$$

$$\tilde{\beta}_o(w,\Delta) = \alpha(0,\Delta), \qquad (3.3)$$

$$q_{k+1}(\Delta) = \frac{1}{k+1}\sum_{j=o}^{k} q_j(\Delta)c_{k+1-j}(0,\Delta) \qquad (k \geq 1)$$

$$q_o(\Delta) = \alpha(0,\Delta), \qquad (3.4)$$

$$\alpha(0,\Delta) = \exp\left\{-\sum_{n=1}^{\infty}\frac{1}{n}c_n(0,\Delta)\right\} \qquad (3.5)$$

and
$$c_k(w,\Delta) = \exp(k\Delta w + \tfrac{1}{2}kw^2)\,\Phi\{-(w+\Delta)\sqrt{k}\} \quad (k \geq 1) \qquad (3.6)$$

In this particular case we can get some useful information about $1-\alpha(x,\Delta)$ by examining the asymptotic behaviour. It is possible to show that $1-\alpha(x,\Delta) \sim a_\Delta e^{-b_\Delta x}$, for large x. Supposing this to be true and substituting in (3.2), we get

$$a_\Delta e^{-b_\Delta x} \sim 1 - \Phi(x+\Delta) + a_\Delta \int_0^\infty e^{-b_\Delta u}\varphi(x-u+\Delta)du. \qquad (3.7)$$

The expression for the integral here is valid for large x since the mass of the kernel $\varphi(x-u+\Delta)$ is heavily concentrated about $u = x+\Delta$. Now $\Phi(y)$ satisfies the asymptotic relation

$$1 - \Phi(x+\Delta) \sim e^{-b_\Delta x}\{1 - \Phi(x+\Delta-b_\Delta)\},$$

so that (3.7) becomes

$$a_\Delta e^{-b_\Delta x} \sim e^{-b_\Delta x} + \Phi(x+\Delta-b_\Delta)\{a_\Delta \exp(\tfrac{1}{2}b_\Delta^2 - \Delta b_\Delta) - 1\}.$$

This implies that $b_\Delta = 2\Delta$, and we then deduce that

$$1-\alpha(x,\Delta) = c(x,\Delta)e^{-2\Delta x} \tag{3.8}$$

say, where $c(x,\Delta)$ has a finite limit as $x \to \infty$.

To evaluate the probabilities (3.1) by the method discussed in Section 2 we need to express $1 - \alpha(x,\Delta)$ in terms of exponentials of the form e^{-wx}; (3.8) represents the first step in this direction. We need next to examine $c(x,\Delta)$ for finite x, and this we do by computing numerical values.

Following the outline in Section 2 for numerical solution of $1 - \alpha(x,\Delta)$, we truncate the integral in (3.2) at x_n and approximate the integral by Simpson's Rule using interval width d, say. Note that the truncation forces $1 - \alpha(x,\Delta)$ to be zero for $x > x_n$, leading to error in the computed values of $1 - \alpha(x,\Delta)$. The discrete version of (3.2), corresponding to (2.24), is then

$$1-\alpha(jd,\Delta) = \psi(jd + \Delta) + d\sum_{k=o}^{n} e_k \{1-\alpha(kd,\Delta)\}\varphi\{(j-k)d+\Delta\}$$

$$(j = 0, 1, \ldots, n), \tag{3.9}$$

where $\psi(y) = 1 - \Phi(y)$, $e_o = e_n = 0.5$ and $e_1 = \ldots = e_{n-1} = 1.0$. A more accurate solution is obtained if we substitute

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

in (3.2), especially if x_n is large enough for the asymptotic form of $1 - \alpha(x,\Delta)$ to be valid for $x \geq x_n$. Doing this, the equation (3.9) then holds with

$$\psi(jd + \Delta) = 1 - \Phi(jd + \Delta) + c(\infty,\Delta)e^{-2jd\Delta}[1-\Phi\{(j-n)d+\Delta\}].$$

The value of $c(\infty,\Delta)$ is unknown, but we can take a trial value, then see how well the solution matches up, and hence obtain a better value. Specifically we assume that $c(x_n,\Delta)$ is close to $c(\infty,\Delta)$, so that a comparison between the trial value of $c(\infty,\Delta)$ and the computed value of $c(x_n,\Delta)$ determines a second approximation; this procedure can be iterated until the agreement between $c(\infty,\Delta)$ and $c(x_n,\Delta)$ is satisfactory. As an illustration, consider the case $\Delta = 1.0$; for this and other cases when $\Delta \geq 0.5$ we have used the values $d = 0.1$ and $n = 100$ in (3.9). The numerical solutions of $1 - \alpha(x,\Delta)$ for trial values $c(\infty,\Delta) = 0$ and $c(\infty,\Delta) = 1$ are given in Table 3.1 in terms of $c(x,\Delta)$. Notice that in both cases $c(x,\Delta)$ becomes stable around $x = 5$, but then moves off towards the assigned value of $c(\infty,\Delta)$. From these two solutions we might take the next trial value of $c(\infty,\Delta)$ as 0.34, say. Iterating further, we quickly obtain the solution given in the final column of Table 3.1 with $c(\infty,\Delta) = 0.32037$. This solution is very stable and clearly represents the best solution to (3.9).

x	$c(\infty,1) =$ 0	1	0.32037
0	0.19946	0.29509	0.19939
1	.30588	.37371	.30586
2	.32396	.38415	.32403
3	.31991	.36609	.31991
4	.32039	.36452	.32035
6	.32035	.36429	.32036
8	.31848	.36799	.32037
10	.25638	.49121	.32037
>10	.00000	1.00000	.32037

Table 3.1 Numerical values of $c(x,\Delta)$ from (3.9) using different trial values of $c(\infty,\Delta)$ in the case $\Delta = 1.0$

The next stage in getting a suitable expression for $1 - \alpha(x,\Delta)$ is to explain the behaviour of $c(x, \Delta)$ before it settles on the asymptotic value $c(\infty, \Delta)$, in this case for $x \le 4$. Let

$$c^*(x,\Delta) = c(x,\Delta) - c(\infty,\Delta), \tag{3.10}$$

and write the integral equation (3.2) as an integral equation for $c^*(x,\Delta)$. Then we have

$$c^*(x,\Delta) = \psi^*(x,\Delta) + \int_0^\infty c^*(u,\Delta)\varphi(x-u-\Delta)du, \tag{3.11}$$

by virtue of (3.8), where

$$\psi^*(x,\Delta) = e^{2\Delta x}\{1-\Phi(x+\Delta)\} - c(\infty,\Delta)\{1-\Phi(x-\Delta)\}. \qquad (3.12)$$

The numerical solution of (3.11) is carried out in the same way as for (3.9), that is by truncation and discretization; truncation now has negligible effect because $c^*(x,\Delta)$ approaches zero rapidly. The same values of d and n are used. A detailed study of the numerical solution of (3.11) led Gentleman, in unpublished work, to an expression for $c^*(x,\Delta)$ which matches numerical values within their estimated accuracy, namely

$$c^*(x,\Delta) = \psi^*(x,\Delta) + \gamma_\Delta\,\varphi(x-\mu_\Delta); \qquad (3.13)$$

the values of γ_Δ and μ_Δ can be determined graphically. The last term in (3.13) is typically quite small, but not negligible. However if we are to approximate to (3.13) with exponentials e^{-wx}, we must omit this term to avoid considerable difficulty. This would lead to the expression

$$1-\alpha(x,\Delta) \doteq 1-\Phi(x+\Delta) + c(\infty,\Delta)\Phi(x-\Delta)e^{-2\Delta x}. \qquad (3.14)$$

A comparison of the numerical values derived by solving (3.9) and numerical values of (3.14) is given in Table 3.2 for the case $\Delta = 1.0$. Now in order to represent $1-\alpha(x,\Delta)$ in terms of exponentials using (3.14), we must represent $\Phi(y)$ in terms of exponentials for y in a region including $y = 0$. In fact this is difficult and requires a large number of terms for a satisfactory approximation. In the present context we can get as much accuracy by working directly with $c^*(x,\Delta)$ itself.

x	$1-\alpha(x,1.0)$ from (3.9)	$1-\alpha(x,1.0)$ from (3.14)	$1-\alpha(x,1.0)$ from (3.15)
0	0.1994	0.2095	0.1988
0.2	.1520	.1606	.1556
0.4	.1134	.1202	.1152
0.6	.0825	.0881	.0825
0.8	.0590	.0632	.0579
1.0	.0414	.0444	.0400
1.5	.0161	.0172	.0154
2.0	.00593	.00628	.00578
3.0	.00079	.00082	.00079
4.0	.00011	.00011	.00011

Table 3.2 Comparison of approximations (3.14) and (3.15) with values of $1-\alpha(x,1.0)$ derived from (3.9)

For the values of Δ which are of interest here (between 0.5 and 1.5), $c^*(x,\Delta)$ is very small when x is greater than about 1.5. We therefore consider fitting exponential terms to numerical values of $c^*(x,\Delta)$ for $x < 1.5$. Fitting a single term $a(\Delta)\exp\{-w(\Delta)x\}$ we use values of $c^*(x,\Delta)$ at $x = 0(0.2)1.4$.

If these are denoted by $y_1, \ldots, y_8$, then the fitted value of $w(\Delta)$ is given by

$$e^{-0.2w(\Delta)} = \sum_{i=1}^{7} y_i y_{i+1} \Big/ \sum_{i=1}^{7} y_i^2 ,$$

and $a(\Delta)$ is determined by least squares regression of y on $e^{-w(\Delta)x}$. The resulting approximation to $1 - \alpha(x,\Delta)$ is, by virtue of (3.8) and (3.10),

$$c(\infty,\Delta)\ \exp(-2\Delta x) + a(\Delta)\ \exp[-\{2\Delta + w(\Delta)\}x]$$

$$= h_1(\Delta)\ \exp\{-w_1(\Delta)x\} + h_2(\Delta)\ \exp\{-w_2(\Delta)x\}, \qquad (3.15)$$

say. An illustration of the accuracy of (3.15) is given in Table 3.2 for the case $\Delta = 1.0$, when $a(\Delta) = -0.1216$ and $w(\Delta) = 1.601$. Note that the error compares favourably with that of (3.14), being 5% or less relative to the actual value of $1 - \alpha(x,\Delta)$. To see how this will affect the probabilities (3.1), note that by (2.8a, b) we can write

$$\mathrm{pr}(\hat{\tau} = \tau \pm k) = \int_0^\infty \alpha(x,\Delta)\ \beta_k(x,\Delta)\,dx\ ;$$

the error in $\alpha(x,\Delta)$ is 0.3 per cent or less, and hence the error in these probabilities using (3.15) will be 0.3 per cent or less. The same magnitude of error applies for all the values of Δ considered here. Thus the approximation (3.15) is good enough for us to use in calculating the asymptotic distribution of $\hat{\tau}$

given by (3.1). Substituting from (3.14) in (3.1) we get

$$\mathrm{pr}(\hat{\tau} = \tau \pm k;\ \Delta) \doteq q_k(\Delta) - \sum_{i=1}^{2} h_i(\Delta)\, \tilde{\beta}_k\{w_i(\Delta),\ \Delta\} \quad (k \geq 1)$$

and

$$\mathrm{pr}(\hat{\tau} = \tau;\Delta) = \{\alpha(0,\Delta)\}^2 \ ; \qquad \text{~~(3.15)~~}$$

the latter, from (2.4), involves no approximation. It is then a simple matter to calculate the asymptotic distribution of $\hat{\tau}$ using (3.3) - (3.6). Numerical values of the distribution are given in Table 3.3 for the cases $\Delta = 0.5(0.1)1.5$. The notation used is

$$p(k,\Delta) = \mathrm{pr}(\hat{\tau} = \tau \pm k;\ \Delta) \qquad (k = 0,\ 1,\ \ldots)$$

and $$P(k,\Delta) = \sum_{-\infty}^{k} p(k,\Delta) = \mathrm{pr}(\hat{\tau} \leq \tau + k;\ \Delta) \qquad (k = 0,\ 1,\ \ldots).$$

Entries in the table have estimated accuracy within 0.2 per cent of the values given.

For values of Δ greater than 1.5 the asymptotic distribution of $\hat{\tau}$ will clearly be heavily concentrated at τ. Extrapolation from Table 3.3 for $p(0,\Delta)$, $p(1,\Delta)$ and $p(2,\Delta)$ should be satisfactory for practical purposes.

	Δ = 0.5		Δ = 0.6		Δ = 0.7	
k	p(k,Δ)	P(k,Δ)	p(k,Δ)	P(k,Δ)	p(k,Δ)	P(k,Δ)
0	0.2802	0.6401	0.3600	0.6800	0.4376	0.7188
1	.1139	.7540	.1241	.8041	.1280	.8468
2	.0668	.8208	.0657	.8698	.0609	.9077
3	.0441	.8650	.0398	.9096	.0337	.9414
4	.0310	.8959	.0259	.9355	.0202	.9615
5	.0226	.9185	.0176	.9531	.0126	.9742
6	.0169	.9354	.0123	.9653	.0082	.9823
7	.0129	.9484	.0088	.9741	.0054	.9878
8	.0100	.9584	.0064	.9805	.0037	.9914
9	.0079	.9663	.0047	.9852	.0025	.9939
10	.0062	.9725	.0035	.9887	.0017	.9957
12	.0040	.9815	.0020	.9933		
14	.0026	.9874	.0012	.9959		
16	.0018	.9913				
18	.0012	.9939				
20	.0008	.9958				

Table 3.3 Asymptotic distribution of $\hat{\tau}$ in the Normal case for $\Delta = 0.5(0.1)1.5$

k	Δ = 0.8 p(k,Δ)	Δ = 0.8 P(k,Δ)	Δ = 0.9 p(k,Δ)	Δ = 0.9 P(k,Δ)	Δ = 1.0 p(k,Δ)	Δ = 1.0 P(k,Δ)
0	0.5110	0.7555	0.5790	0.7895	0.6409	0.8204
1	.1266	.8821	.1212	.9107	.1130	.9334
2	.0539	.9360	.0459	.9566	.0378	.9713
3	.0271	.9631	.0208	.9774	.0153	.9866
4	.0148	.9779	.0103	.9877	.0068	.9934
5	.0085	.9863	.0054	.9931	.0032	.9966
6	.0051	.9914	.0029	.9960		
7	.0031	.9945				
8	.0019	.9964				

k	Δ = 1.1 p(k,Δ)	Δ = 1.1 P(k,Δ)	Δ = 1.2 p(k,Δ)	Δ = 1.2 P(k,Δ)	Δ = 1.3 p(k,Δ)	Δ = 1.3 P(k,Δ)
0	0.6963	0.8481	0.7458	0.8726	0.7881	0.8940
1	.1030	.9511	.0920	.9646	.0807	.9748
2	.0303	.9813	.0235	.9882	.0179	.9926
3	.0109	.9922	.0074	.9956	.0049	.9976
4	.0043	.9965				

Table 3.3 (continued) Asymptotic distribution of $\hat{\tau}$ in the Normal case for Δ = 0.5(0.1)1.5

k	$\Delta = 1.4$ $p(k,\Delta)$	$P(k,\Delta)$	$\Delta = 1.5$ $p(k,\Delta)$	$P(k,\Delta)$
0	0.8251	0.9126	0.8568	0.9284
1	.0697	.9823	.0594	.9877
2	.0132	.9955	.0096	.9973

Table 3.3 (continued) Asymptotic distribution of $\hat{\tau}$ in the Normal case for $\Delta = 0.5(0.1)1.5$

Table 3.4 gives a few values of $p(0,\Delta)$ for Δ between 1.5 and 3.0.

Δ	1.75	2.00	2.25	2.50	2.75	3.00
$p(0,\Delta)$	0.9160	0.9531	0.9751	0.9875	0.9940	0.9973

Table 3.4 Values of $p(0,\Delta)$ for $\Delta = 1.75(0.25)3.00$

For very large Δ, $1 - \alpha(0,\Delta)$ is very close to $1 - \Phi(\Delta)$, as can be deduced from the integral equation (3.2); agreement to four figures is achieved for $\Delta \geq 2.75$. Consequently we have the asymptotic expression

$$p(0,\Delta) \sim \{\Phi(\Delta)\}^2.$$

For values of Δ less than 0.5 it is difficult to handle the numerical computation involved. We have not satisfactorily explored the limiting distribution of $\hat{\tau}$ as $\Delta \to 0$.

Note that the asymptotic distribution of $\hat{\tau}$ remains the same when σ^2 is unknown. For then the likelihood of $(x_1, \ldots, x_T)$ maximized over σ^2 conditional on $\hat{\tau} = t$ is proportional to

$$\left\{ \sum_{i=1}^{t} (x_i - \theta_o)^2 + \sum_{i=t+1}^{T} (x_i - \theta_1)^2 \right\}^{-\frac{1}{2}T}$$

$$= \left\{ \sum_{i=1}^{T} (x_i - \theta_1)^2 - \sigma^2 \sum_{i=1}^{t} u_i \right\}^{-\frac{1}{2}T} ,$$

so that $\hat{\tau}$ is still the value of t which maximizes $\sum_{i=1}^{t} u_i$ over $t = 1, \ldots, T - 1$.

3.2 θ_o and θ_1 unknown

In many situations the two means θ_o and θ_1 will be unknown. Here we show, without formal proof, that the asymptotic distribution of Section 3.1 remains valid. To do so we again use the random walk representation of the log likelihood, and the argument of Section 2.

Assume, then, that $(x_1, x_2, \ldots, x_T)$ is a sequence of independent normally distributed random variables whose mean is θ_o for the first τ variables and θ_1 for the remainder; for convenience we let the variance σ^2 be 1. Then the log likelihood of the sample corresponding to (2.1) is

$$L(x_1, \ldots, x_T | \theta_o, \theta_1, \tau) = -\tfrac{1}{2} \left\{ \sum_{i=1}^{\tau} (x_i - \theta_o)^2 + \sum_{i=\tau+1}^{T} (x_i - \theta_1)^2 \right\} \tag{3.16}$$

Now the m.l.e.'s for θ_o and θ_1 conditional on $\tau = t$ are

$$\tilde{\theta}_{ot} = \frac{1}{t} \sum_{j=1}^{t} x_j = \bar{x}_t$$

and $$\tilde{\theta}_{1t} = \frac{1}{T-t} \sum_{j=t+1}^{T} x_j = \bar{x}_t^* \text{, say.} \tag{3.17}$$

Hence the log likelihood (3.16) maximized over θ_o and θ_1 conditional on $\tau = t$ is

$$L(t) = -\tfrac{1}{2}\left\{\sum_{i=1}^{t}(x_i - \bar{x}_t)^2 + \sum_{i=t+1}^{T}(x_i - \bar{x}_t^*)^2\right\}$$

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

$$(t = 1, 2, \ldots, T-1),$$

using the definitions (3.17). Therefore $\hat{\tau}$ is the value of t which maximises U_t^2, say, where

$$U_t^2 = \frac{t(T-t)}{T}(\bar{x}_t - \bar{x}_t^*)^2 \qquad (t = 1, \ldots, T-1). \qquad (3.18)$$

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

Now suppose $\theta_o > \theta_1$, then $(\bar{x}_t - \bar{x}_t^*)$ is positive with probability tending to one as both t and T - t increase indefinitely, so that we can take U_t as asymptotically equivalent to U_t^2 in order to find the asymptotic distribution of $\hat{\tau}$. Strictly speaking we should consider only values of t and T - t such that both tend to infinity as τ and $T - \tau$ do, but this would cause no difficulty. The sequence $\{U_t\}$ is an auto-correlated sequence with known but complicated covariance matrix and mean. However from (3.18) we deduce the following autoregressive representation

$$U_{t+1} = a_{t+1} U_t + b_{t+1} (x_{t+1} - \bar{x}_T) \quad (t = 1, \ldots, T-1), \tag{3.19}$$

where

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

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

This relation is close to that defining a random walk, since $a_t \sim 1$. We shall show that (3.19) does represent a random walk with a negligible superimposed deterministic function. Let $\tau = \lambda T$ and consider τ and $T - \tau$ large. Then we have

$$a_t \sim 1 - \frac{1-2\lambda}{2\lambda(1-\lambda)T}$$

and $$b_t \sim \{\lambda(1-\lambda)T\}^{-\frac{1}{2}} ; \tag{3.20}$$

here we assume that $t - \tau$ is $o(\tau)$. We also have

$$E(\bar{x}_T) = \lambda\theta_o + (1-\lambda)\theta_1 \tag{3.21}$$

$$\text{var}(\bar{x}_T) = T^{-1}$$

and

$$E(U_\tau) = \{\lambda(1-\lambda)T\}^{\frac{1}{2}}(\theta_o - \theta_1) \tag{3.22}$$

$$\text{var}(U_t) = 1 \quad \text{for all } t$$

Consider the increment $U_{\tau+1} - U_\tau$, which by (3.19) and (3.20) can be written

$$U_{\tau+1} - U_\tau = -\frac{(1-2\lambda)}{2\lambda(1-\lambda)T} U_\tau + \frac{(x_{t+1} - \bar{x}_T)}{\{\lambda(1-\lambda)T\}^{\frac{1}{2}}} + \varepsilon ,$$

where $\mathrm{var}(\varepsilon) = o(T^{-2})$.

Substituting $U_\tau = E(U_\tau) + \eta$ and $\bar{x}_T = E(\bar{x}_T) + \xi$ and using (3.18) and (3.19) we get

$$U_{\tau+1} - U_\tau = \frac{\left\{x_{\tau+1} - \lambda\theta_o - (1-\lambda)\theta_1 - \left(\frac{1-2\lambda}{2}\right)(\theta_o - \theta_1)\right\}}{\{\lambda(1-\lambda)T\}^{\frac{1}{2}}} + \varepsilon + \eta' + \xi'$$

$$= z_1 + \varepsilon + \eta' + \xi', \qquad (3.23)$$

say, where $E(\varepsilon) = o(T^{-1})$, $E(\eta') = E(\xi') = 0$, $\mathrm{var}(\varepsilon) = o(T^{-2})$, $\mathrm{var}(\eta') = O(T^{-2})$ and $\mathrm{var}(\xi') = O(T^{-2})$, and z_1 is

$$N\left[-\frac{(\theta_o - \theta_1)}{2\{\lambda(1-\lambda)T\}^{\frac{1}{2}}}, \quad \frac{1}{\lambda(1-\lambda)T}\right].$$

Similarly we find that

$$U_{\tau+k} - U_\tau = \sum_{j=1}^{k} z_j + k(\varepsilon + \eta' + \xi') \quad (k = 1, 2, \ldots), \qquad (3.24)$$

where z_1, z_2, ... are identically and independently distributed. Rescaling by a factor $\{\lambda(1-\lambda)T\}^{\frac{1}{2}}$ we see that (3.24) defines a random walk whose increments are $N(-\Delta,1)$ with a superimposed linear term whose coefficient has variance $O(T^{-1})$. A similar representation exists for $U_{\tau-k} - U_\tau$ $(k = 1, 2, \ldots)$, and we deduce that the asymptotic distribution of $\hat{\tau}$ is determined by the random walks alone. That is, the results of Section 3.1

hold. Note that the distribution of $\hat{\tau}$ does not depend on λ, the reason being that only observations 'near' τ influence $\hat{\tau}$.

In general it should be possible to prove similar results for arbitrary distributions $f(x,\theta)$, presumably under conditions on the asymptotic behaviour of $\tilde{\theta}_{ot}$ and $\tilde{\theta}_{1t}$; for example, (2.25) is probably a necessary condition.

Since the moments of $\hat{\tau}$ are $O(1)$, i.e. the distribution does not spread out as T increases indefinitely, it follows that the m.l.e.'s $\hat{\theta}_o$ and $\hat{\theta}_1$ are asymptotically normally distributed and unbiased, with variances $(\lambda T)^{-1}$ and $\{(1-\lambda)T\}^{-1}$ respectively. However it is clear that in finite samples $\hat{\theta}_o - \hat{\theta}_1$ will have a positive bias (assuming $\theta_o > \theta_1$), since $\hat{\tau}$ is determined by finding the most significant difference $\tilde{\theta}_{ot} - \tilde{\theta}_{1t}$. This is important in making inference about τ when θ_o and θ_1 are unknown (Section 4). In principle we can get an asymptotic expression for the bias, since by (3.17), (3.18) and the random walk representation of U_t we have

$$\hat{\theta}_o - \hat{\theta}_1 = \tilde{\theta}_{o\hat{\tau}} - \tilde{\theta}_{1\hat{\tau}} = (\bar{x}_\tau - \bar{x}^*_\tau) + \{\lambda(1-\lambda)T\}^{-1}M^* + o(T^{-1}),$$

where $M^* = \max(M, M')$ is the overall maximum of W and W'.

Hence

$$E(\hat{\theta}_o - \hat{\theta}_1) = (\theta_o - \theta_1) + \{\lambda(1-\lambda)T\}^{-1}E(M^*) + o(T^{-1}).$$

The distribution of M^* depends only on $\alpha(x,\Delta)$, so that a numerical evaluation of $E(M^*)$ is possible; there is no explicit formula for $E(M^*)$, although it is bounded by $E(M)$ and $2E(M)$, and from (2.13) and (3.6)

$$E(M) = \sum_{k=1}^{\infty} k^{\frac{1}{2}} \varphi(-\Delta\sqrt{k})\, \Phi(-\Delta\sqrt{k}) - \Delta \sum_{k=1}^{\infty} k\, \Phi(-\Delta\sqrt{k}) \quad . \tag{3.25}$$

Some comment on the magnitude of the bias in $\hat{\theta}_o - \hat{\theta}_1$ is given in Section 6.

4. Inference about τ in the Normal Case

In most situations we not only want to estimate τ, but also to make inference about τ in the form of a confidence interval or a test of significance. For convenience suppose that we want to test $H_o^* : \tau = \tau_o$ with either a one- or two-sided alternative. If θ_o and θ_1 are known, we can use the distribution of $\hat{\tau}$ derived in Section 3.1 and calculate the significance of $\hat{\tau}$ computed from $x_1, \ldots, x_T$. The likelihood ratio test, however, is at least as efficient asymptotically and easier to apply. It may be more efficient, because $\hat{\tau}$ is not asymptotically sufficient: only the observations themselves are sufficient and the likelihood ratio test uses all the information.

Consider the two-sided test of H_o^* with alternative $H_1^* : \tau \neq \tau_o$. Then, using the notation of Section 2, the log likelihood ratio test statistic $\log \Lambda_2$, say, is given by

$$\log \Lambda_2 = \frac{L(\hat{\tau})}{L(\tau_o)} = \sum_{j=1}^{\hat{\tau}} u_j - \sum_{j=1}^{\tau_o} u_j$$

$$= \max(M, M'). \qquad (4.1)$$

We reject H_o^* when $\log \Lambda_2 > \ell$, say, where ℓ is determined from the required test size. The asymptotic distribution of $\log \Lambda_2$ under H_o^* is, then,

$$\mathrm{pr}(\log \Lambda_2 \leq x) = \mathrm{pr}(M \leq x)\mathrm{pr}(M' \leq x) = \alpha(x)\alpha'(x) \qquad (4.2)$$

by the independence of M and M'. This distribution is easier to calculate than that of $\hat{\tau}$ in general. In the normal case (4.1) becomes

$$\log \Lambda_2 = \frac{\theta_o - \theta_1}{\sigma^2}\left\{\sum_{i=1}^{\hat{\tau}}\left(x_i - \frac{\theta_o + \theta_1}{2}\right) - \sum_{i=1}^{\tau_o}\left(x_i - \frac{\theta_o + \theta_1}{2}\right)\right\},$$

and with the notation of Section 3 the distribution (4.2) becomes

$$\mathrm{pr}(\log \Lambda_2 = 0) = \{\alpha(0,\Delta)\}^2$$

$$\mathrm{pr}(\log \Lambda_2 \leq x) = \{\alpha(\tfrac{x}{2\Delta}; \Delta)\}^2 . \qquad (4.3)$$

For the one-sided test of H_o^* with alternative $H_2^* : \tau > \tau_o$, the log likelihood ratio test statistic is

$$\log \Lambda_1 = \max_{t \geq \tau_o} \sum_{i=1}^{t} u_i - \sum_{i=1}^{\tau_o} u_i$$

$$= \frac{\theta_o - \theta_1}{\sigma^2} \left\{ \max_{t \geq \tau_o} \sum_{j=1}^{t} \left(x_i - \frac{\theta_o + \theta_1}{2} \right) - \sum_{j=1}^{\tau_o} \left(x_i - \frac{\theta_o + \theta_1}{2} \right) \right\}$$

$$= M'.$$

Then corresponding to (4.3) we have the null distributions

$$\mathrm{pr}(\log \Lambda_1 = 0) = \alpha(0, \Delta)$$

$$\mathrm{pr}(\log \Lambda_1 \leq x) = \alpha\left(\frac{x}{2\Delta}, \Delta\right) . \qquad (4.4)$$

Values of the null distributions (4.3) and (4.4) are given in Table 4.1 for the cases $\Delta = 0.5$ and 1.0 as an illustration. Here

$$P_1(x,\Delta) = \mathrm{pr}(\log \Lambda_1 \leq x;\ \Delta)$$

and $$P_2(x,\Delta) = \mathrm{pr}(\log \Lambda_2 \leq x;\ \Delta)$$

x	$P_1(x,0.5)$	$P_2(x,0.5)$	$P_1(x,1.0)$	$P_2(x,1.0)$
0.4	0.6522	0.4254	0.8480	0.7191
0.8	.7543	.5689	.8868	.7863
1.2	.8316	.6916	.9175	.8418
1.6	.8866	.7860	.9410	.8855
2.0	.9240	.8538	.9586	.9189
2.4	.9492	.9009	.9714	.9436
2.8	.9659	.9330	.9805	.9613
3.2	.9772	.9549	.9868	.9738
3.6	.9847	.9696	.9911	.9823
4.0	.9897	.9796	.9941	.9882
4.4	.9931	.9863	.9960	.9921
4.8	.9954	.9908	.9974	.9947

Table 4.1 Asymptotic null distributions of the one- and two-sided log likelihood ratio test statistics; Δ = 0.5 and 1.0

The values were calculated from the numerical solution of $\alpha(x,\Delta)$ described in Section 3.1. Note that the distributions depend on Δ. Quantiles of the distributions can be obtained by inverse interpolation from extended versions of Table 4.1. This we have done for the 95%, 98% and 99% quantiles for Δ ranging from

0.5 to 1.5. The results are given in Table 4.2 to three significant figures. Here $\mathrm{pr}(\log \Lambda_i \le \ell_{i,p}) = \frac{p}{100}$ $(i = 1,2)$.

Δ	$\ell_{1,95}$	$\ell_{1,98}$	$\ell_{1,99}$	$\ell_{2,95}$	$\ell_{2,98}$	$\ell_{2,99}$
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

Table 4.2 One- and two-sided 95, 98 and 99 per cent quantiles for the log likelihood ratio test; $\Delta = 0.5(0.1)1.5$

When θ_o and θ_1 are unknown, the log likelihood ratio (4.1) becomes

$$\log \Lambda_2 = (U^2_{\hat{\tau}} - U^2_{\tau_o})/(2\sigma^2).$$

By arguments similar to those of Section 3.2 it is not difficult to see that $\log \Lambda_2$ still has the asymptotic distribution (4.3), since $\hat{\theta}_o$ and $\hat{\theta}_1$ are consistent. That is

$$\log \Lambda_2 = (U_{\hat{\tau}} + U_{\tau_o})(U_{\hat{\tau}} - U_{\tau_o})/(2\sigma^2)$$

$$\sim E(U_{\tau_o})\ (U_{\hat{\tau}} - U_{\tau_o})/\sigma^2$$

$$= \sqrt{\frac{\tau_o(T-\tau_o)}{T}}\left(\frac{\theta_o - \theta_1}{\sigma^2}\right)\left(U_{\tau} - U_{\tau_o}\right)$$

$$\sim \frac{(\theta_o - \theta_1)}{\sigma} \max\ (M_1, M_2)\ ,$$

where M_1 and M_2 are the maxima of two independent random walks with independent $N(-\Delta,1)$ increments, and so have distribution function $\alpha(x,\Delta)$. The corresponding result holds for

$$\log \Lambda_1 = \left(\max_{t\geq\tau_o} U_t^2 - U^2_{\tau_o}\right)\Big/(2\sigma^2)\ .$$

Now with θ_o and θ_1 unknown, we have to test H_o^* with Δ as a nuisance parameter. The m.l.e. $\hat{\Delta}$ is given by

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

and is asymptotically normally distirbuted $N(\Delta, \sigma^2_{\Delta})$, where

$$\sigma^2_{\Delta} = \frac{T}{4\tau(T-\tau)}\quad ;$$

this follows from the asymptotic normality of $\hat{\theta}_o$ and $\hat{\theta}_1$. What we want is a function corresponding to $P_2(x,\Delta)$ for Δ unknown. The problem has a certain resemblance to that of testing a hypothetical normal mean μ_o when the variance σ^2 is unknown.

We can obtain bounds for $P_2(x,\Delta)$ by deriving upper and lower confidence limits for Δ. If $\Phi(u_\beta) = 1 - \frac{1}{2}\beta$, the $100(1 - \beta)\%$ confidence limits for Δ are $\hat{\Delta} \pm u_\beta \sigma_\Delta$, where

$\sigma_\Delta^2 = \dfrac{T}{4\tau_o(T-\tau_o)}$ under H_o^*. Then we have

$$P_2(x,\Delta) \geq (1 - \tfrac{1}{2}\beta)\, P_2(x,\hat{\Delta} - u_\beta\sigma_\Delta)$$

and

$$P_2(x,\Delta) \leq (1 - \tfrac{1}{2}\beta)\, P_2(x,\hat{\Delta} + u_\beta\sigma_\Delta) + \tfrac{1}{2}\beta. \qquad (4.5)$$

These bounds are not sharp. For example when $\Delta = 1.0$, $\tau_o = 50$, $T = 100$ and $\hat{\Delta} = 1.1$, the bounds for $P_2(3,\Delta)$ using $\beta = 0.05$ differ by approximately 0.04 where the true value is 0.97.

One possible approach is to estimate $P_2(x,\Delta)$ using $\hat{\Delta}$. It is easy to show that an unbiased estimate of $P_2(x,\Delta)$ to order σ_Δ^2 (i.e. order T^{-1}) is

$$P_2(x,\hat{\Delta}) - \tfrac{1}{2}\sigma_\Delta^2\, P_2''(x,\hat{\Delta}) , \qquad (4.6)$$

where $P_2''(x,\Delta)$ is the second derivative with respect to Δ. This estimate should have a distribution closer to the uniform distribution on (0,1) than $P_2(x,\hat{\Delta})$. This notion of uniformity stems from the Neyman-Pearson theory of hypothesis testing and is a property we want our analogue of $P_2(x,\Delta)$ to have. One source of difficulty in obtaining a solution analogous to the Student t distribution is that the only sufficient statistics here are the observations themselves. The general problem of inference in the presence of nuisance parameters is a difficult one and needs further investigation.

5. Cumulative Sum Techniques

A standard practical method for detecting a change in the mean value θ of a sequence of observed random variables $(x_1, x_2, \ldots, x_T)$ is the Cusum (cumulative sum) technique developed by Page (1957). In this section we derive the asymptotic distribution of the Cusum estimate of the change-point τ. We make use of the results of Section 3. A direct comparison of the Cusum estimate and the m.l.e. is given.

First we outline the Cusum testing procedure for fixed sample size T. It is assumed that the initial mean value θ_o is known and that the x_i are independent and normally distributed with known variance σ^2. For simplicity, we consider the one-sided test for an increase in mean θ. The procedure is to plot the cumulative sums

$$S_t = \sum_{j=1}^{t} (x_j - \theta_o - \delta\sigma) \qquad (t = 1, 2, \ldots, T) \qquad (5.1)$$

and to reject the hypothesis of constant mean if

$$S_T - \min_{t<T} S_t > h , \qquad (5.2)$$

where both δ and h are chosen to give the test required size and required power under a specific alternative $\theta_o^* > \theta_o$, say, which are determined by the experimenter. Then the estimate of the change-point τ is the index of the Cusum minimum, that is

$$\tilde{\tau} = \inf\{t : S_t \leq S_k, k = 1, \ldots, T\} .$$

Note that if the mean is known to change to θ_o^*, so that the Cusum is used only to estimate τ, then we should choose $\delta = (\theta_o^* - \theta_o)/(2\sigma)$ and hence have $\tilde{\tau} = \hat{\tau}$ by the definition of the m.l.e. $\hat{\tau}$ in Section 3. In general, however, the magnitude of the change is unknown, and $\tilde{\tau}$ is asymptotically less efficient than $\hat{\tau}$. We show this by deriving the asymptotic distributions of both $\tilde{\tau}$ and $\hat{\tau}$, assuming both τ and $T - \tau$ to be indefinitely large.

Let the actual increase in mean be from θ_o to θ_1. Then the sequence $S_1, S_2, \ldots, S_T$ defined by (5.1) is equivalent to the sequence $-V_1, -V_2, \ldots, -V_T$ with V_t as defined in Section 2. The Cusum estimate $\tilde{\tau}$ is the index of the maximum of $\{V_t\}$, and we can apply the results of Sections 2 and 3. In this case it is easy to see that the increments y_j and z_j of the two associated random walks W and W' are respectively $N(-\delta\sigma, \sigma^2)$ and $N(\theta_o - \theta_1 + \delta\sigma, \sigma^2)$. Then it follows without difficulty that the asymptotic distribution

$$p(k; \delta, \Delta) = pr(\tilde{\tau} = \tau + k; \delta, \Delta)$$

is given by

$$p(k;\ \delta,\ \Delta) = q_k(2\Delta-\delta) - \int_0^\infty \{1-\alpha(u,\delta)\}\beta_k(u,\ 2\Delta-\delta)du \quad (k \geq 1)$$

$$p(0;\ \delta,\ \Delta) = \alpha(0,\delta)\ \alpha(0,2\Delta-\delta)$$

$$p(-k;\ \delta,\ \Delta) = q_k(\delta) - \int_0^\infty \{1-\alpha(u,\ 2\Delta-\delta)\}\beta_k(u,\delta)du \quad (k \geq 1)\ . \quad (5.3)$$

The notation here is that of Section 3.1; in particular $\Delta = (\theta_1 - \theta_0)/(2\sigma)$. Numerical calculation of the distribution (5.3) is done by the same method as in Section 3.1, that is using the two-term approximation to $1 - \alpha(x,\Delta)$ and the Laplace transforms $\tilde{\beta}_k(w,\Delta)$. It is not difficult to see from (5.3) that $\tilde{\tau}$ is biased. In fact both the bias and the variance increase as $|\Delta - \delta|$ increases. We can explain this as follows: a change ϵ in the value of δ is equivalent to rotating $\{S_0 = 0,\ S_1,\ \ldots,\ S_T\}$ about the origin (0,0) through an angle $\tan^{-1}(\epsilon)$. Now if Δ were known, the best choice of δ for estimating τ is $\delta = \Delta$ which gives the unbiased m.l.e. $\hat{\tau}$. Not to choose $\delta = \Delta$, or a consistent estimate of Δ when Δ is unknown, leads to a loss of efficiency and introduces a bias as a result of the rotation away from the optimum position. This is what happens when θ_1 is unknown.

The question we have to answer is: how well does $\tilde{\tau}$ compare with $\hat{\tau}$ in efficiency, in possible bias and in simplicity of use?

If the distributional difference between $\hat{\tau}$ and $\tilde{\tau}$ is small, it may be that the Cusum estimate is to be preferred because of its simplicity. To find the m.l.e. $\hat{\tau}$ when θ_1 is unknown, we note first that the log likelihood of τ is (3.16) with $\tilde{\theta}_{1\tau} = \bar{x}^*_\tau$ replacing θ_1. It follows that $\hat{\tau}$ is the value of t which maximizes

$$Z_t^2 = t(\bar{x}_t^* - \theta_o)^2 \qquad (t = 1, 2, \ldots, T - 1);$$

cf. (3.18). For $\theta_1 > \theta_o$, $\hat{\tau}$ is asymptotically equivalent to the value of t which maximizes Z_t, which we can write as

$$Z_t = t^{-\frac{1}{2}} \sum_{j=T-t+1}^{T} (x_j - \theta_o). \tag{5.4}$$

The connection between $\tilde{\tau}$ and $\hat{\tau}$ is seen by noticing from (5.1) and (5.4) that

$$Z_t = (S_T - S_{T-t} + t\delta\sigma)t^{-\frac{1}{2}} . \tag{5.5}$$

This implies that $\hat{\tau}$ can be obtained directly from the Cusum plot by superimposing the family of parabolas

$$y = - At^{\frac{1}{2}} + t\delta\sigma \qquad (A > 0) \tag{5.6}$$

on the plot with origin at (T, S_T), and adjusting A until only one Cusum point lies below the curve (5.6). The index of that point is $\hat{\tau}$. The use of parabolic curves in Cusum procedures was proposed by Barnard (1959). Likelihood ratio tests of the constant mean hypothesis correspond to selecting particular values of A in (5.6) and rejecting if a Cusum point falls below the curve.

From (5.4) it is easy to show that the asymptotic distribution of $\hat{\tau}$ is the same as when θ_o and θ_1 are known. The argument follows closely that of Section 3.2 and we omit the details here. In particular, $\hat{\tau}$ is asymptotically unbiased. However, in finite samples we might expect $\hat{\tau}$ to be biased because of the indeterminacy of θ_1. It might be possible to derive the finite sample distribution of $\hat{\tau}$ by considering the properties of a random walk with a parabolic boundary, but we have not been able to do this.

To illustrate the difference between $\tilde{\tau}$ and $\hat{\tau}$ for various values of δ and Δ, we have calculated the asymptotic distribution of $\tilde{\tau}$ from (5.3) for the cases $\Delta = 1.0$, $\delta = 0.5(0.1)1.0$. Table 5.1 gives the asymptotic bias and variance for each case, and Figure 5.1 is a plot of the cumulative distribution of $\tilde{\tau}$ for $\delta = 0.5$, 0.7 and 1.0. Note that $\delta = 1.0$ corresponds to the m.l.e. distribution; note also that

$$\text{var}(\tilde{\tau};\ \delta,\ \Delta) \quad = \quad \text{var}(\tilde{\tau};\ 2\Delta-\delta,\ \Delta)$$

and $$E(\tilde{\tau} - \tau;\ \delta,\ \Delta) = -E(\tilde{\tau} - \tau;\ 2\Delta-\delta,\ \Delta).$$

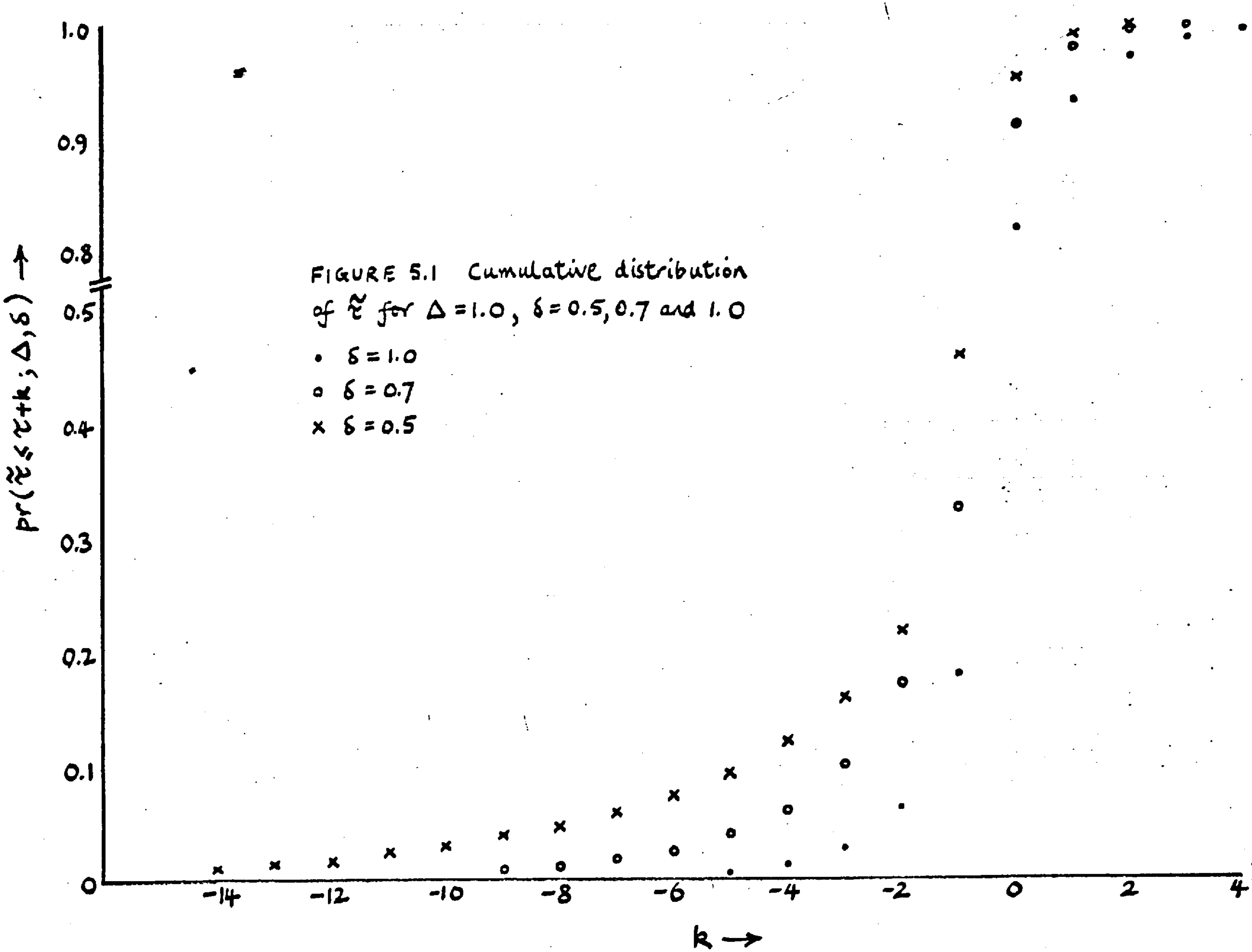


FIGURE 5.1 Cumulative distribution of $\tilde{\tau}$ for $\Delta = 1.0$, $\delta = 0.5, 0.7$ and 1.0
• $\delta = 1.0$
o $\delta = 0.7$
× $\delta = 0.5$
$pr(\tilde{\tau} \le \tau + k; \Delta, \delta)$
k
1.0
0.9
0.8
0.5
0.4
0.3
0.2
0.1
0
-14
-12
-10
-8
-6
-4
-2
0
2
4

δ	0.5	0.6	0.7	0.8	0.9	1.0
$E(\tilde{\tau}-\tau;\delta,1.0)$	-1.70	-1.07	-0.68	-0.40	-0.19	0.00
$\mathrm{var}(\tilde{\tau};\delta,1.0)$	12.00	5.99	3.43	2.24	1.71	1.56

Table 5.1 Asymptotic bias and variance of $\tilde{\tau}$ for the cases $\Delta = 1.0$, $\delta = 0.5(0.1)1.0$

It is clear from these numerical results that $\tilde{\tau}$ can be very inferior to $\hat{\tau}$ even for small differences between δ and Δ. One practical interpretation of the results is that if the mean change is larger (smaller) than was anticipated, then the estimate of the change-point will tend to be too small (too large).

So far we have assumed that the Cusum procedure is used for fixed sample size testing. But often the procedure is used as a method of sequential control for a process $\{x_t\}$. Then the Cusum (5.1) is plotted until for some T (5.2) is satisfied. The criterion for choosing δ and h is then average sample number before a decision is reached rather than power; the average sample numbers for mean values θ_o and θ_o^* determine the values of δ and h. In this situation T will be a random variable, but the effect of the stopping rule (5.2) on the likelihood of the sample is independent of τ, so that the m.l.e. remains unaltered. Now $T - \tau$ becomes a random variable whose distribution depends on Δ and δ, and whose mean value is chosen to be large when no change occurs and small when the change in mean is to θ_o^*. It is not clear what effect this will have on the distributions of $\tilde{\tau}$ and $\hat{\tau}$. The first reaction is to suppose that the asymptotic distributions will not be reasonable approximations in this situation; but if the asymptotic distributions are concentrated on the integers less than $E(T|\delta,\Delta)$ then they should be adequate approximations. We have not looked at the theoretical aspect of this in any detail. The empirical results for fixed sample size described in Section 6 indicate how large $T - \tau$ must be for asymptotic results to give reasonable agreement.

The corresponding test for a decrease in mean is to plot

$$S_t' = \sum_{j=1}^{t} (x_i - \theta_o + \delta\sigma) \qquad (t = 1, \ldots, T)$$

and to reject the hypothesis of constant mean if

$$S_T' - \max_{t<T} S_t' < -h.$$

Then $\tilde{\tau}$ is the index of the Cusum maximum, and the distribution (5.3) applies. The two-sided test combines both one-sided tests, and it can be shown that rejection in both tests simultaneously is impossible. The probability of rejection in favour of the wrong one-sided alternative will usually be negligible, so that the asymptotic distribution of $\tilde{\tau}$ (a maximum or minimum point) will again be (5.3).

A Cusum procedure is also defined when both θ_0 and θ_1 are unknown. The Cusums (5.1) are then replaced by

$$S_0 = S_T = 0, \quad S_t = \sum_{j=1}^{t} (x_j - \bar{x}_T) \quad (t = 1, \ldots, T - 1), \tag{5.7}$$

with decision rules as before. The asymptotic distribution of $\tilde{\tau}$ can be derived by following the arguments of Section 3, and is given by (5.3) with the substitution $\delta = 2(1-\lambda)\Delta$, where $\tau = \lambda T$. Note that when $\lambda \neq \frac{1}{2}$ more weight is given to observations on one side of the change-point. This is because the mean slope of S_t is different on each side of $t = \tau$ due to the end conditions $S_0 = S_T = 0$; this makes $\tilde{\tau}$ biased. To compare $\tilde{\tau}$ with the m.l.e., we note that the m.l. estimating statistics U_t of (3.18) can be expressed as

$$U_t = \left\{\frac{T}{t(T-t)}\right\}^{\frac{1}{2}} S_t \quad (t = 1, \ldots, T - 1)$$

by the definition (5.7). Hence $\hat{\tau}$ can be obtained from theCusum plot by superimposing curves

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

and varying A until one point lies outside the curve, whose index is $\hat{\tau}$. The distributional comparisons of $\hat{\tau}$ and $\tilde{\tau}$ for the case θ_o known apply here.

6. Monte Carlo Results

To see how well the asymptotic results of Sections 3, 4 and 5 work in finite sample situations, we carried out an extensive simulation study of the distributions of $\hat{\tau}$, $\tilde{\tau}$, $\hat{\Delta}$ and log Λ_2. In this section we give a summary of the conclusions from this study.

For each set of values of T, τ and Δ we generated 500 samples of observations on an electronic computer, using pseudo random normal deviates for the error terms ε_i in the model

$$x_i = \begin{cases} \theta_o + \varepsilon_i & (i = 1, \ldots, \tau) \\ \theta_1 + \varepsilon_i & (i = \tau + 1, \ldots, T) \,. \end{cases}$$

It is of particular interest to see how the finite sample distributions of $\hat{\tau}$, $\hat{\Delta}$ and log Λ_2 vary according as θ_o and/or θ_1 are known or unknown. To eliminate "between samples" error when making comparisons of this type we calculated the empirical distributions under the three assumptions (a) θ_o, θ_1 both unknown, (b) θ_o known, θ_1 unknown, (c) θ_o and θ_1 both known, in the same samples; $\hat{\Delta}$ is redundant under assumption (c). Also in the same samples we derived the empirical distributions of $\tilde{\tau}$, the Cusum estimate, for several values of δ.

Before looking at specific results, we remark that in general the agreement between empirical and asymptotic results seems to depend critically on the value of

$$D_o = \Delta\sqrt{\frac{\tau(T-\tau)}{T}}$$

when both θ_o and θ_1 are unknown, and on $D_1 = \Delta\sqrt{T-\tau}$ when only θ_1 is unknown. Note that small values of D_o and D_1 will arise in cases where we would have difficulty in determining the presence of two means rather than one. Intuitively we would expect the asymptotic properties of $\hat{\tau}$ and log Λ_2 to be good approximations only if the two means are easily distinguished. It appears from the empirical results described here that if D_o (or D_1) is greater than 3 then a case is well-defined, that is the asymptotic distributions agree well enough for practical use. The cases described here are $\Delta = 0.5$ and 1.0, T = 50, 100 and 200, and various τ not necessarily equal to $\frac{1}{2}T$. To illustrate the general remarks made in the last paragraph, Table 6.1 was compiled for some of the cases studied. This table gives the empirical means and variances of $\hat{\tau}$ and $\hat{\Delta}$, and the corresponding asymptotic variances, for four cases where $\Delta = 0.5$.

	Empirical		Asymptotic	Empirical		Asymptotic
	$E(\hat{\tau})$	$var(\hat{\tau})$	$var(\hat{\tau})$	$E(\hat{\Delta})$	$var(\hat{\Delta})$	$var(\hat{\Delta})$
$T=50,\ \tau=15$						
(a) $D_o=1.6$	20.30	83.62	24.10	.564	.0544	.0238
(b) $D_1=3.0$	14.84	19.53	24.10	.501	.0095	.0072
(c)	15.02	23.16	24.10	-	-	-
$T=50,\ \tau=25$						
(a) $D_o=1.8$	25.46	41.65	24.10	.593	.037	.020
(b) $D_1=2.5$	24.79	31.36	24.10	.514	.015	.010
(c)	25.37	25.49	24.10	-	-	-
$T=100,\ \tau=25$						
(a) $D_o=2.2$	25.37	60.52	24.10	.560	.0185	.0133
(b) $D_1=4.3$	25.26	25.01	24.10	.512	.0035	.0033
(c)	25.29	21.62	24.10	-	-	-
$T=200,\ \tau=50$						
(a) $D_o=3.06$	50.36	31.13	24.10	.509	.0065	.0067
(b) $D_1=6.13$	49.97	26.34	24.10	.496	.0018	.0017
(c)	49.50	24.30	24.10	-	-	-

Table 6.1 Comparison of empirical and asymptotic moments of $\hat{\tau}$ and $\hat{\Delta}$, $\Delta = 0.5$. (a) θ_o, θ_1 unknown; (b) θ_o known, θ_1 unknown; (c) θ_o, θ_1 known.

The respective values of D_o and D_1 are given in the table. Similar comparisons hold in cases where $\Delta = 1.0$. Note that when Δ is known the mean and variance of $\hat{\tau}$ agree with the asymptotic values even in the worst case, namely $T = 50$, $\tau = 15$, $\Delta = 0.5$. In general the asymptotic distribution of $\hat{\tau}$ agrees well with the empirical distribution for known Δ provided that $\hat{\tau}$ lies between 1 and T with high probability. These remarks do not depend on τ being equal to $\frac{1}{2}T$.

Examination of the empirical distributions of $\hat{\tau}$ confirms the above remarks. Figures 6.1 and 6.2 are plots of cumulative empirical distributions against the corresponding cumulative asymptotic distribution for the case $T = 50$, $\tau = 15$, $\Delta = 0.5$ with θ_o and θ_1 known and unknown, respectively. (We shall refer to these and similar plots as percentage plots). Clearly the asymptotic distribution of $\hat{\tau}$ is not a good approximation in the situation where Δ is unknown. The chi square values in the usual goodness-of-fit tests for these two examples are respectively 20.8 and 86.1; the 95 percentage point of the null distribution is 31.4. Behaviour in a well-defined case is illustrated by the percentage plots in Figures 6.3 and 6.4. They indicate good agreement with the asymptotic distribution irrespective of whether or not Δ is known. In these and other well-defined cases the finite sample distributions are not noticeably affected by non-centrality of τ.

FIGURE 6.1 Percentage plot of empirical and asymptotic distributions of $\hat{\tau}$, θ_0 and θ_1 known. $T=50$, $\tau=15$, $\Delta=0.5$

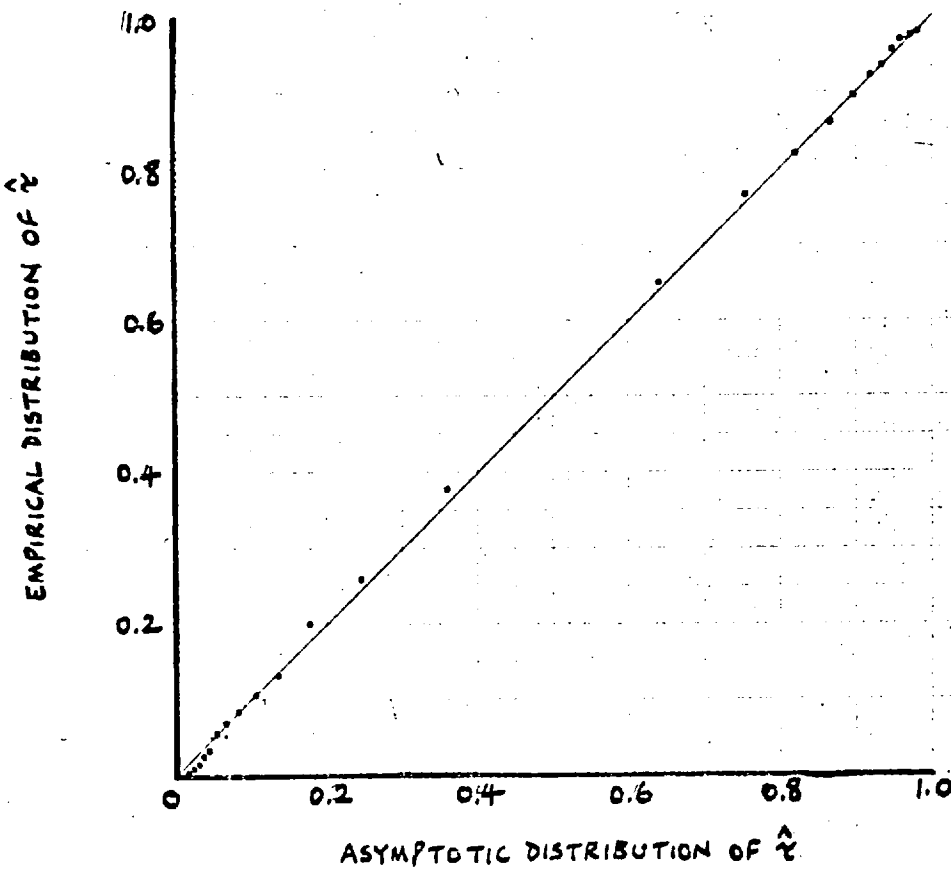


FIGURE 6.2 Percentage plot of empirical and asymptotic distributions of $\hat{\tau}$, θ_0 and θ_1 unknown. $T = 50$, $\tau = 15$, $\Delta = 0.5$

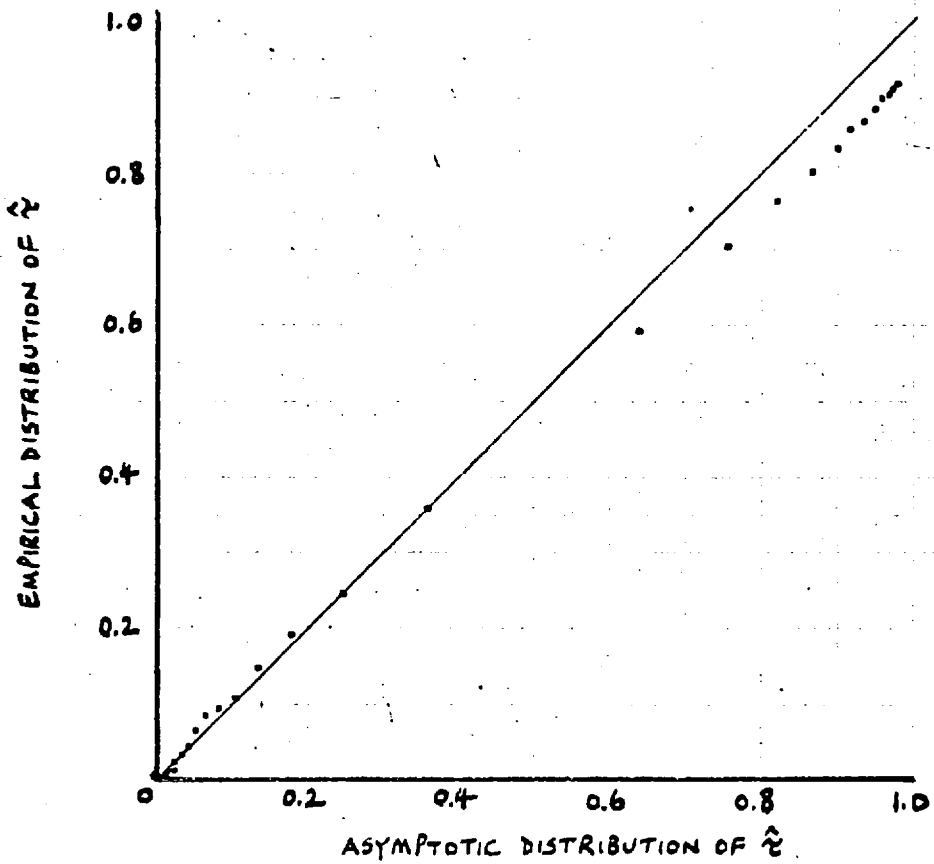


FIGURE 6.3 Percentage plot of empirical and asymptotic distributions of $\hat{\tau}$, θ_0 and θ_1 known
$T=100$, $\tau=25$, $\Delta=1.0$

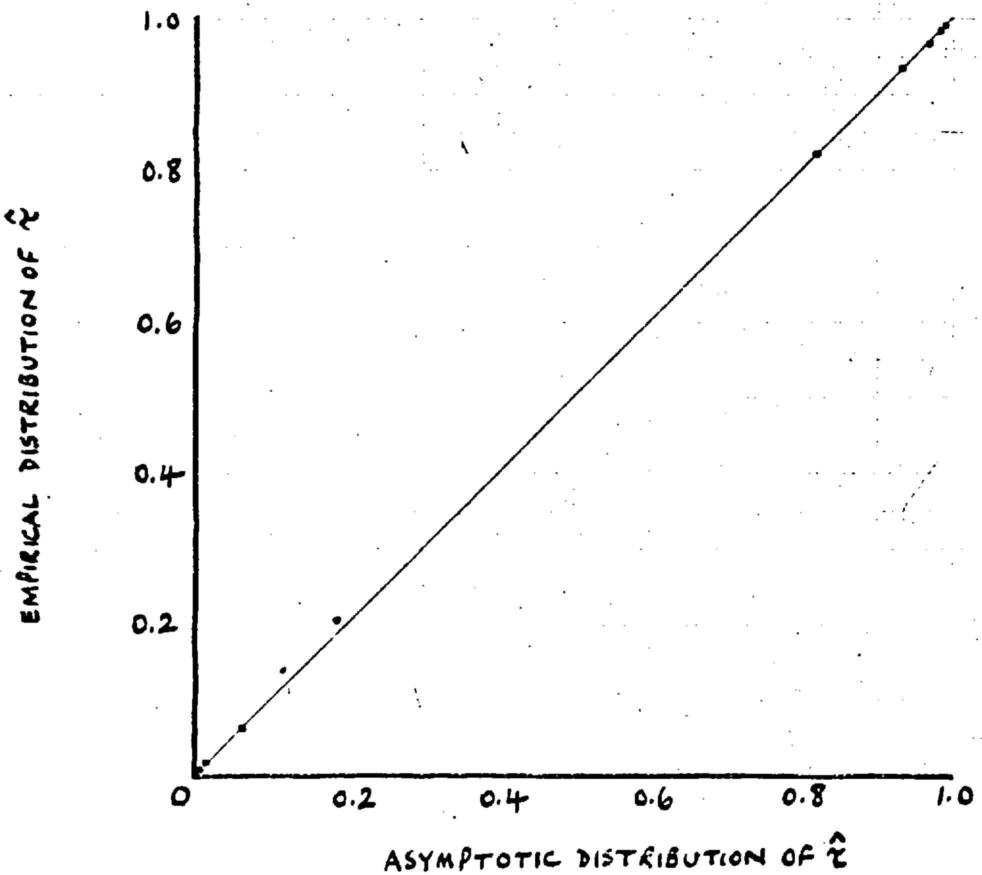
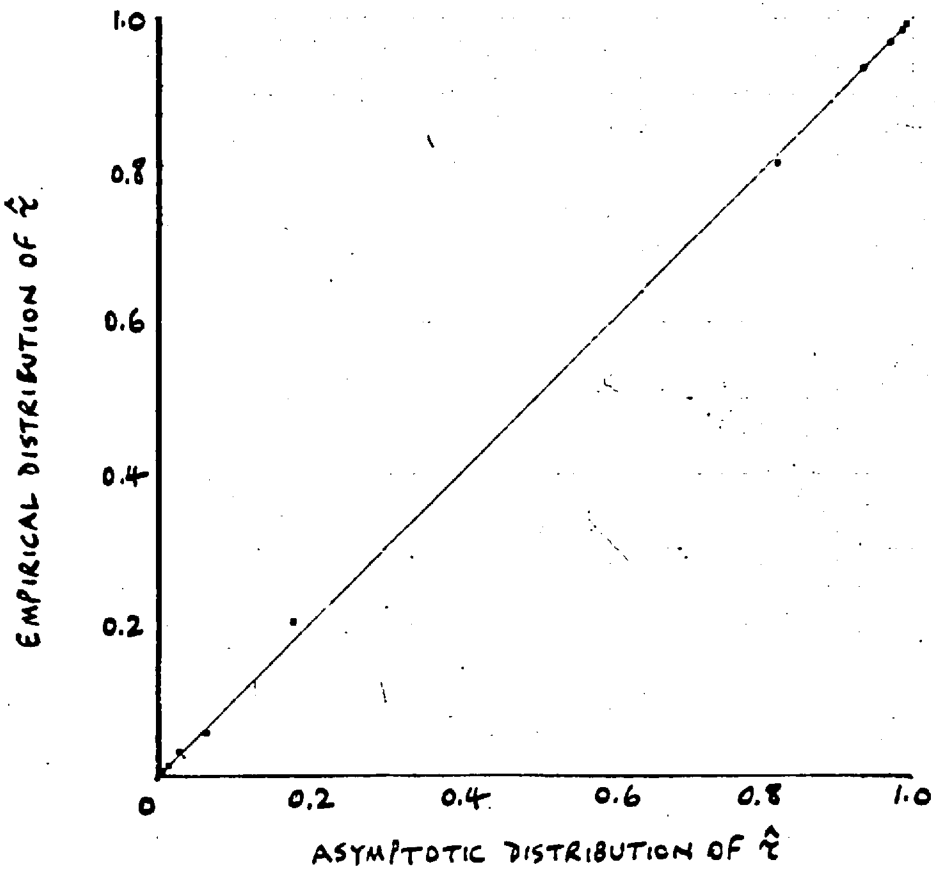


FIGURE 6.4 Percentage plot of empirical and asymptotic distributions of $\hat{\tau}$, θ_0 and θ_1 unknown. $T = 100$, $\tau = 25$, $\Delta = 1.0$



The empirical distributions of $\log \Lambda_2$ agree well with the asymptotic distribution derived in Section 4. When D_o or D_1 is greater than 3 the agreement is good over the whole distribution. The non-centrality of τ has no visible effect. Figure 6.5 for the case $T = 100$, $\tau = 25$, $\Delta = 1.0$ is a typical example of a percentage plot of empirical against asymptotic distributions in a well-defined case. For smaller values of D_o and D_1, the agreement with the asymptotic distribution is still quite good for cumulative probabilities greater than 0.90, which is the region of interest in significance tests, but not for smaller probabilities; this is true for D_o or D_1 as low as 2. When Δ is known the agreement is good even in the case $T = 50$, $\tau = 15$, $\Delta = 0.5$. Figures 6.6 and 6.7 are percentage plots for the cases $T = 50$, $\tau = 15$, $\Delta = 0.5$, Δ known and unknown.

Looking at the empirical distributions of $\hat{\Delta}$, it is clear that the asymptotic normal distribution is a good approximation when D_o or D_1 is greater than 3, except for the small positive bias mentioned in Section 3.2 and illustrated in Table 6.1. In practice the bias is probably negligible in most well-defined cases, but it might be worth examining the asymptotic form of the bias for marginally well-defined cases. Figure 6.8 is a normal plot of the empirical distribution of $\hat{\Delta}$ in the well-defined case $T = 200$, $\tau = 50$, $\Delta = 0.5$, and Figure 6.9 a corresponding

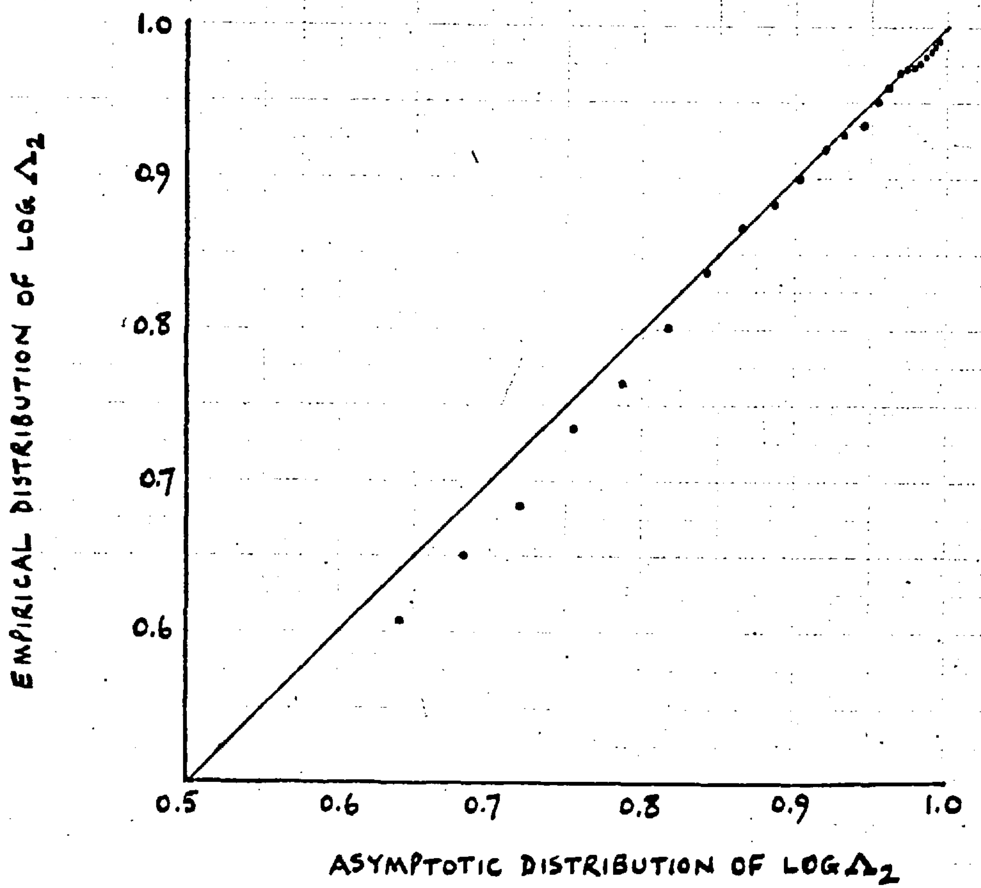


FIGURE 6.5 Percentage plot of empirical and asymptotic distributions of $\log \Delta_2$, θ_0 and θ_1 unknown. $T=100$, $\tau=25$, $\Delta=1.0$

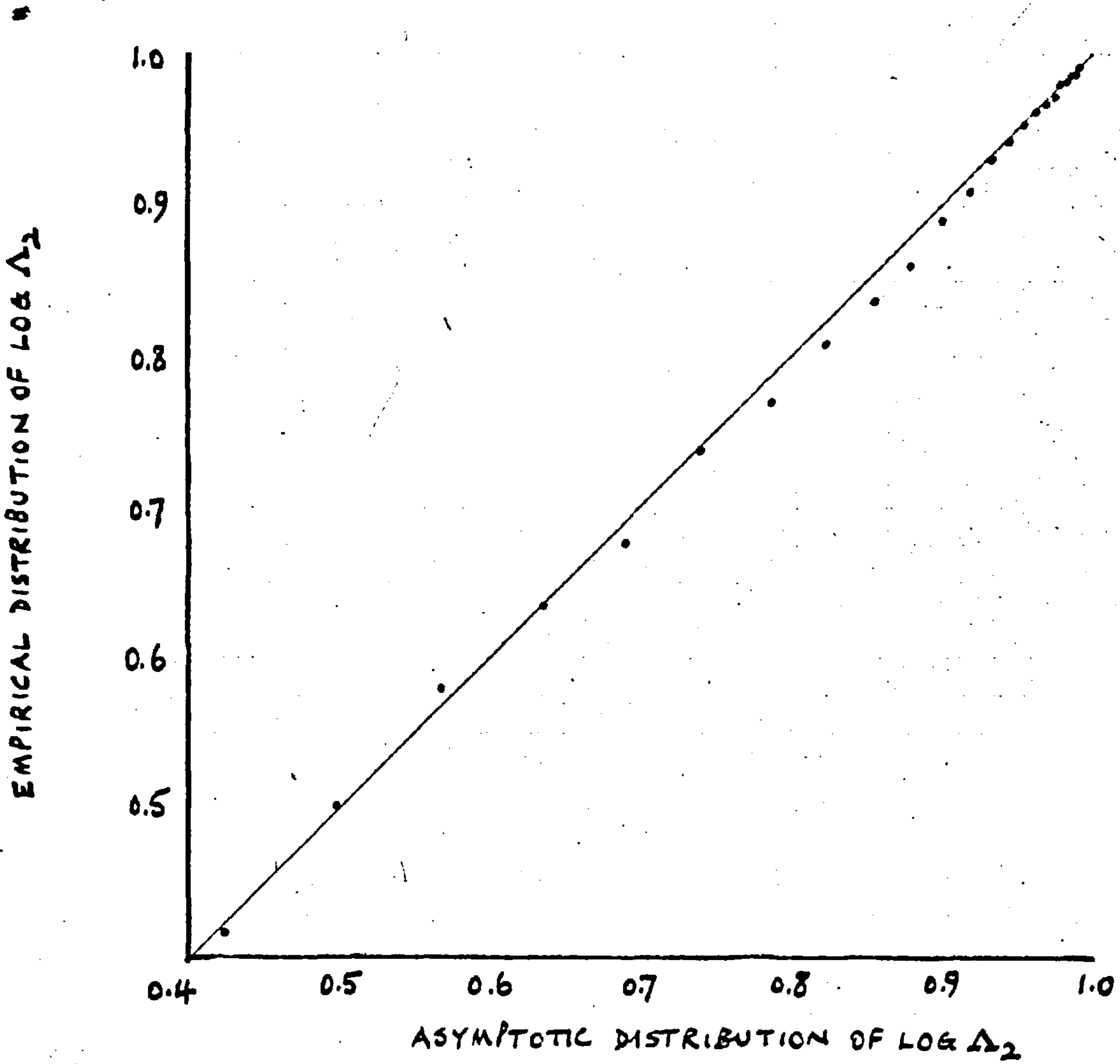


FIGURE 6.6 Percentage plot of empirical and asymptotic distributions of $\log \Lambda_2$, θ_0 and θ_1 known. $T = 50$, $\tau = 15$, $\Delta = 0.5$

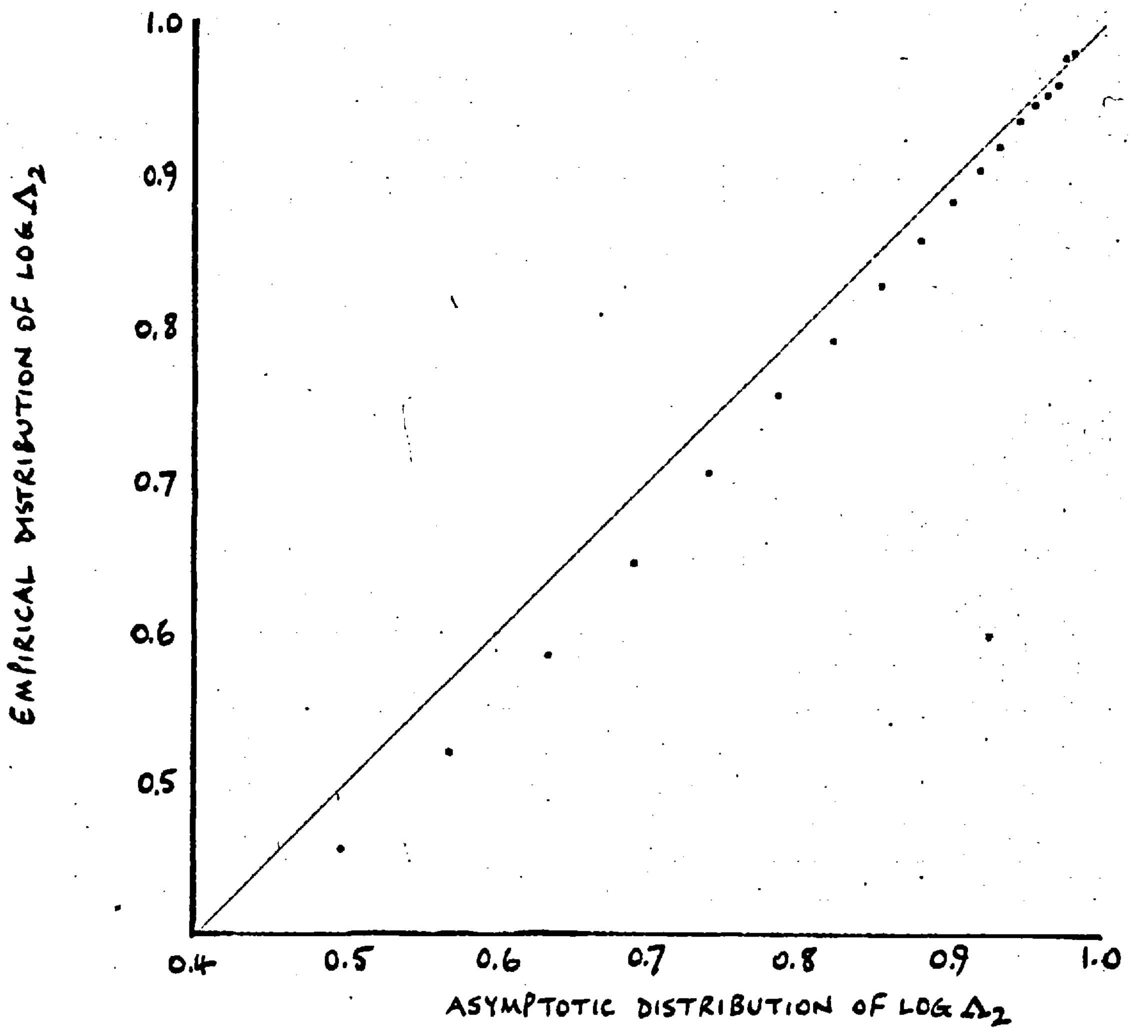


FIGURE 6.7 Percentage plot of empirical and asymptotic distributions of $\log \Lambda_2$, θ_0 and θ_1 unknown. $T=50$, $\tau=15$, $\Delta=0.5$

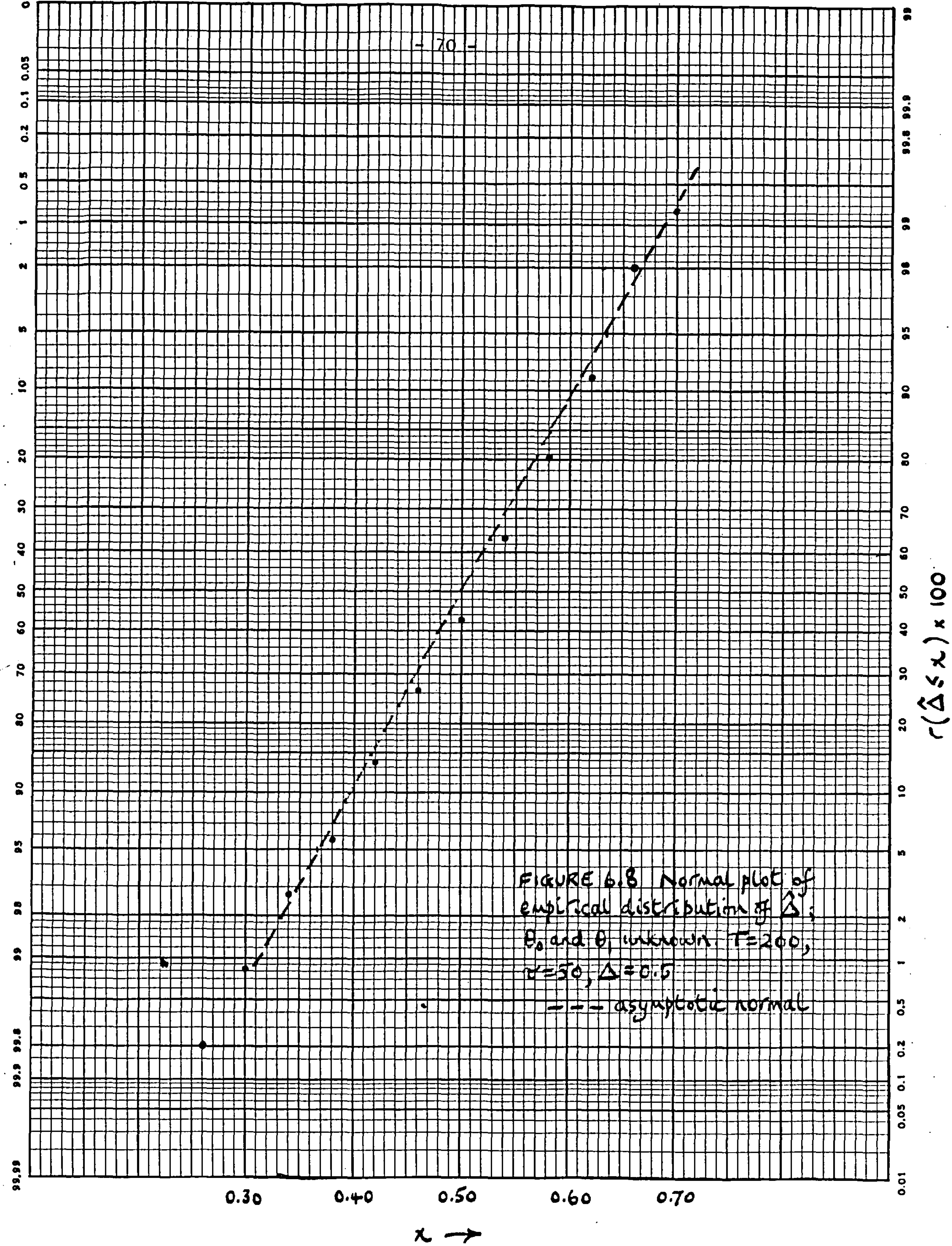


FIGURE 6.8 Normal plot of empirical distribution of $\hat{\Delta}$; θ_0 and θ_1 unknown. $T=200$, $\tau=50$, $\Delta=0.5$

— — — asymptotic normal

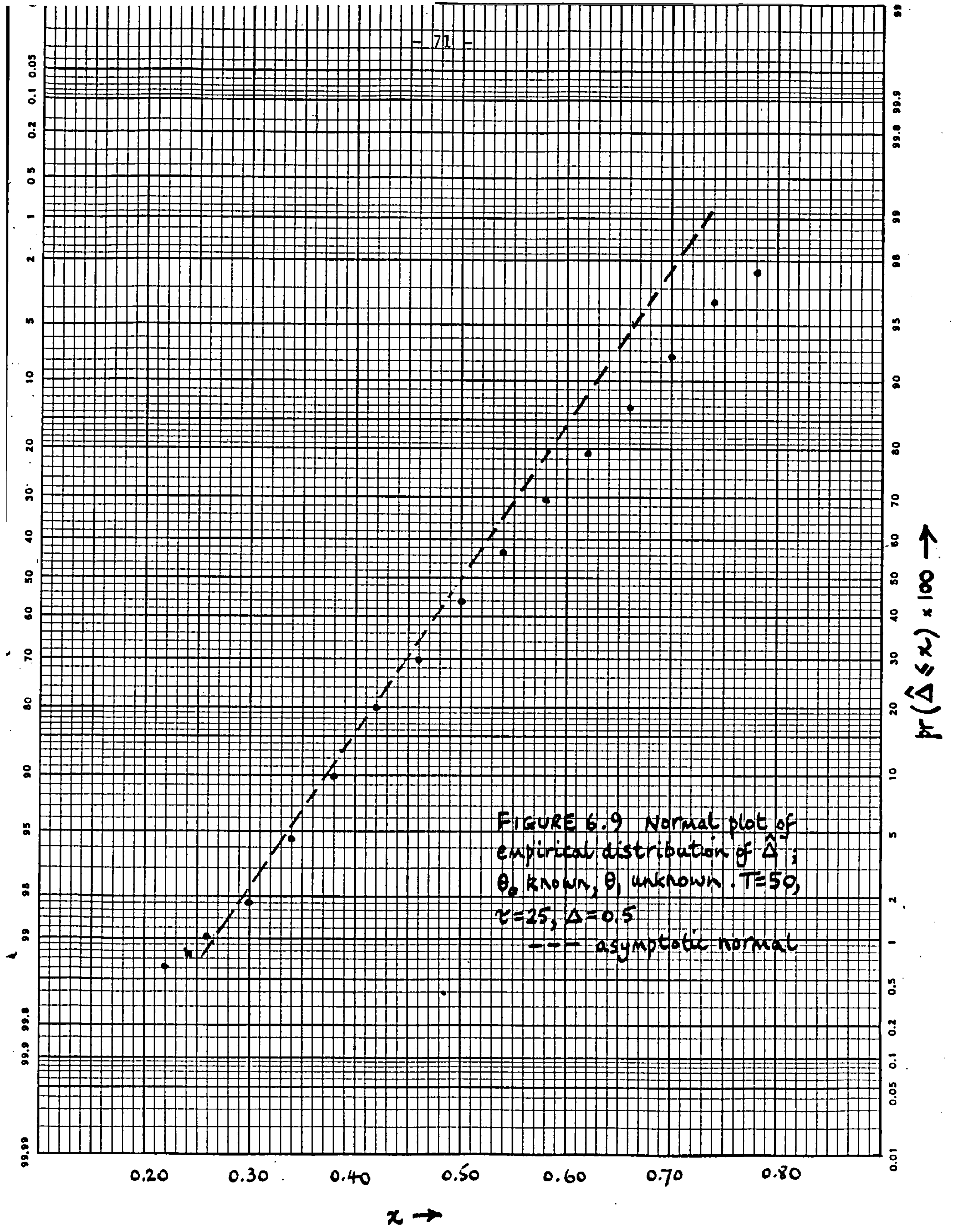


FIGURE 6.9 Normal plot of empirical distribution of $\hat{\Delta}$; θ_0 known, θ_1 unknown. T=50, τ=25, Δ=0.5
--- asymptotic normal

plot for the ill-defined case $T = 50$, $\tau = 25$, $\Delta = 0.5$. In the latter case the bad fit is due to both significant bias and excessive spread.

The general conclusions we draw from the above discussion are, first, that when Δ is unknown the asymptotic results should be good approximations if D_o or D_1 is greater than 3; it is important that $\hat{\Delta}$ should be distributed close to the asymptotic normal distribution because methods of inference about τ involve $\hat{\Delta}$ as the estimate of a nuisance parameter. Second, when Δ is known the asymptotic distribution of log Λ_2 may be used in the tails for quite small samples, say those for which D_o is as low as 1.5. Third, the asymptotic distribution of $\hat{\tau}$ is valid when Δ is known provided $\hat{\tau}$ lies between 1 and T with high probability.

The empirical analysis of the distribution of the Cusum estimate $\tilde{\tau}$ confirms that there is good agreement between the finite sample distribution and the asymptotic distribution when the asymptotic distribution is concentrated in the interval (1,T). Figure 6.10 is a typical percentage plot of the two cumulative distributions, using the same simulated samples as in Figures 6.3 and 6.4

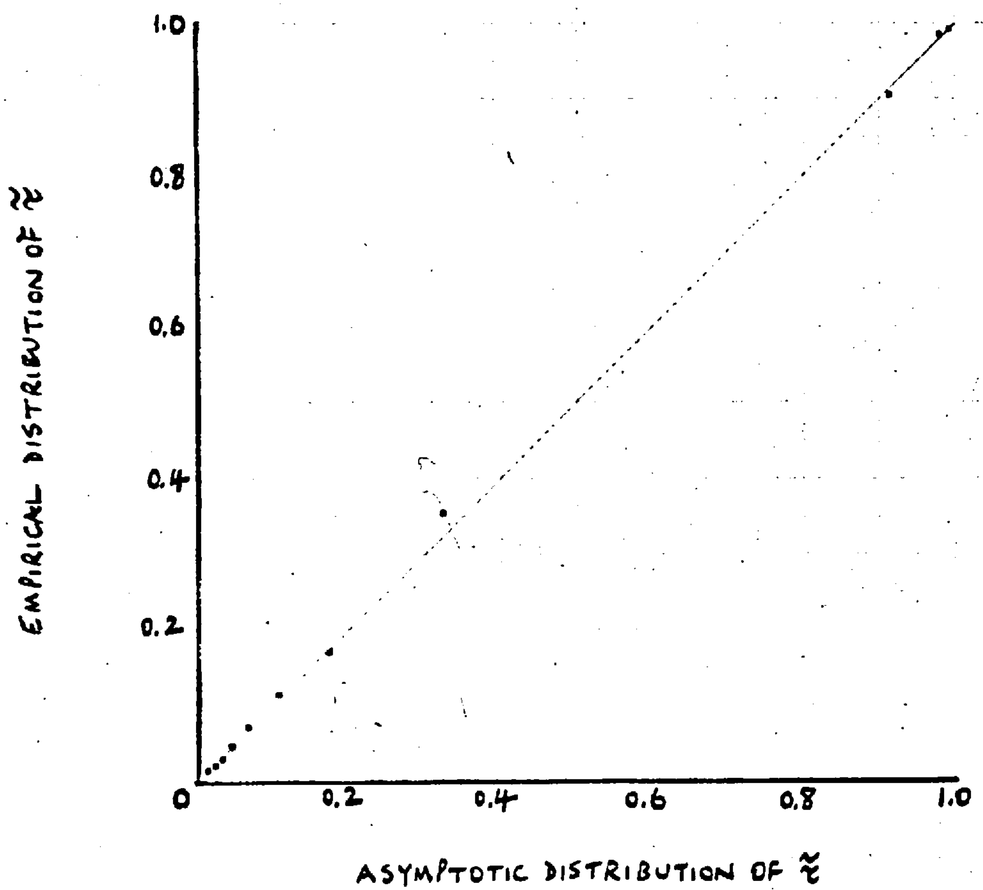


FIGURE 6/10 Percentage plot of empirical and asymptotic distributions of the Cusum estimate $\tilde{\tau}$. $T=100$, $\tau=25$, $\Delta=1.0$, $\delta=0.7$

7. Further Developments

In the preceding sections we have discussed the m.l.e. $\hat{\tau}$ and its asymptotic distribution in the normal case, also the properties of the likelihood ratio test. Empirical results show that in small samples the asymptotic distributions can be poor approximations, particularly when θ_0 and θ_1 are unknown. Exact finite sample results, however, appear very difficult to obtain. For going back to (3.18), the distribution of $\hat{\tau}$ involves the probabilities

$$pr(U_k^2 > U_j^2, \quad 1 \le j \ne k \le T - 1; \tau,\Delta) \quad (k = 1, \ldots, T - 1),$$

where $\underline{U} = (U_1, \ldots, U_{T-1})$ has a multivariate normal distribution with non-zero mean and covariance matrix $((\sigma_{ij}))$ given by

$$\sigma_{ij} = \frac{i'(T - i')}{j'(T - j')}, \qquad i' = \min(i,j),\ j' = \max(i,j).$$

The only method for calculating the relevant multivariate integrals is simulation. Even with exact results about $\hat{\tau}$, inference would be difficult because of the non-normality of $\hat{\Delta}$ in small samples.

On a more hopeful note, the results of Section 2 apply to distributions other than the normal distribution discussed in Section 3, for example the exponential distribution. Nor are the results restricted to problems with one density $f(x,\theta)$, for it is not difficult to see that the arguments of Section 2 go through for the general model

$$x_i \text{ has p.d.f. } f_o(x) \quad (i = 1, \ldots, \tau)$$
$$x_i \text{ has p.d.f. } f_1(x) \quad (i = \tau + 1, \ldots, T).$$

It should also be possible to extend the theory of Section 2 to discrete distributions, for example the case of binary data.

The asymptotic results we have obtained for $\hat{\tau}$ are relevant to the problem of testing the hypothesis

$$H_1 : x_j = \begin{cases} \theta_o + \epsilon_j & (j = 1, \ldots, \tau) \\ \theta_1 + \epsilon_j & (j = \tau + 1, \ldots, T) \end{cases}$$

against

$$H_{reg} : x_j = \alpha + \beta u_j + \epsilon_j \quad (j = 1, \ldots, T),$$

where u_j is an independent variable (possibly $u_j = j$) and the ϵ_j are independent $N(0,\sigma^2)$ error terms. The hypotheses H_1 and H_{reg} are separate in the sense of Cox's (1961) definition.

A different aspect is the testing of the hypothesis of homogeneity $H_o : \theta_o = \theta_1$, i.e. no change-point. Apart from the work by Page (1954, 1955, 1957), there have recently been several published accounts of related work by Chernoff and Zacks (1964), who use a Bayesian framework, and Bhattacharyya and Johnson (1968). The discussion of Section 5 would be relevant to the likelihood ratio test of H_o, for by the definition of the random walks W and W′

in (2.2), the log likelihood ratio for testing H_o against the presence of a change-point is $\sum_{i=1}^{\hat{\tau}} u_i$. We have not looked at the properties of $\hat{\tau}$ in the null case.

PART II

ESTIMATION OF THE CHANGE-POINT IN TWO-PHASE REGRESSION

8. Previous Results

In this second part of the thesis we look at Model B of the introduction, given in (1.3). It is more convenient to express the model in the form

$$x_t = \alpha_o + \beta_o u_t + \epsilon_t \quad (t = 1, \ldots, \tau)$$

$$x_t = \alpha_1 + \beta_1 u_t + \epsilon_t \quad (t = \tau + 1, \ldots, T) , \qquad (8.1)$$

where $\alpha_o + \beta_o\gamma = \alpha_1 + \beta_1\gamma$ and $u_\tau \leq \gamma < u_{\tau+1}$. The independent variables are ordered $u_1 < u_2 < \ldots < u_T$, and the error terms ϵ_t are independently distributed $N(0, \sigma^2)$.

The parameters α_o, β_o, α_1, β_1 and γ are unknown, and so also is τ. Our object is to estimate γ, the abcissa of intersection of the two regression lines of (8.1), and to make inference about γ. Statements about τ are derived from statements about γ, so that estimation of τ is not discussed explicitly.

Quandt (1958) suggested the following method for locating $\hat{\tau}$, the m.l.e. (maximum likelihood estimate) of τ. Take any trial value t for τ and fit the two linear regression equations of (8.1) by least squares, ignoring the constraint $u_t \leq \gamma < u_{t+1}$. Calculate the value of the likelihood maximised over α_o, β_o, α_1 and β_1 given t. Repeat this for all values of t in the range 2, ..., T-2. The value of t giving the overall maximum is then $\hat{\tau}$. For the more restricted case of our model, introduction of the constraint on γ would involve much more calculation; the procedure for finding $\hat{\tau}$ would be nearly as complicated as that for finding $\hat{\gamma}$ (Section 9). Quandt does not analyze the distributional properties of $\hat{\tau}$, the majority of his work being concerned with tests of the null hypothesis $(\alpha_o, \beta_o) = (\alpha_1, \beta_1)$. He has not gone further with estimating τ (or γ) in subsequent work.

Sprent (1961) considers estimation of γ when τ is known, which might be the case if the u_t's are very dispersed and/or some prior knowledge is available. An important assumption is made, which we also make in following sections, namely that the two line model (8.1) provides an adequate description of the data. The error variance σ^2 is unknown in general. Sprent takes a trial value γ_o in the known interval $[u_\tau, u_{\tau+1})$ and fits the least squares two line model constrained to intersect at $u = \gamma_o$. This is done by minimizing the Lagrange expression

$$\sum_{i=1}^{\tau} (x_i - \alpha_o - \beta_o u_i)^2 + \sum_{i=\tau+1}^{T} (x_i - \alpha_1 - \beta_1 u_i)^2$$
$$+ 2\lambda\{\alpha_1 - \alpha_o + \gamma_o(\beta_1 - \beta_o)\},$$

where λ is the Lagrange multiplier.

Now the residual sum of squares $S^2(\gamma_o)$, say, is distributed as $\sigma^2\chi^2_{T-3}$ under the hypothesis $\gamma = \gamma_o$. However, we could fit two unconstrained regression lines and obtain a residual sum of squares S^2 distributed as $\sigma^2\chi^2_{T-4}$ whatever value γ takes. Therefore the statistic

$$F(\gamma_o) = \frac{S^2(\gamma_o) - S^2}{S^2} \times (T - 4)$$

will have the F distribution with 1 and T - 4 degrees of freedom under the hypothesis $\gamma = \gamma_o$. A confidence region for γ is then defined by the values of γ satisfying

$$F(\gamma) \leq F_{1,T-4}(\xi) ,$$

where $F_{1,T-4}(\xi)$ is the upper $100\xi\%$ point of the F distribution. This region can alternatively be derived by the use of Fieller's Theorem in the following way.

The random variable

$$(\hat{\alpha}_1 - \hat{\alpha}_o) + \gamma_o(\hat{\beta}_1 - \hat{\beta}_o)$$

is, under the hypothesis $\gamma = \gamma_o$, normally distributed with mean zero and variance $\sigma^2 V(\gamma_o)$ with $V(\gamma_o)$ known. Here $\hat{\alpha}_o$, $\hat{\alpha}_1$, $\hat{\beta}_o$ and $\hat{\beta}_1$ are the unconstrained least squares estimates.

The confidence region for γ is, then, the set of values of γ satisfying

$$\left\{(\hat{\alpha}_1 - \hat{\alpha}_o) + \gamma(\hat{\beta}_1 - \hat{\beta}_o)\right\}^2 - K^2_{T-4}(\xi)\, \hat{\sigma}^2\, V(\gamma) \leq 0 ,$$

where $K_{T-4}(\xi)$ is the two-sided $100\xi\%$ point of Student's t-distribution, and $\hat{\sigma}^2 = S^2/(T-4)$. A little calculation shows that

$$V(\gamma) = \frac{1}{\tau} + \frac{1}{T-\tau} + \frac{(\bar{u}_\tau - \gamma)^2}{C_{uu,\tau}} + \frac{(\bar{u}^*_\tau - \gamma)^2}{C^*_{uu,\tau}} ,$$

where $\bar{u}_t = \frac{1}{t}\sum_{i=1}^{t} u_i$, $\bar{u}_t^* = \frac{1}{T-t}\sum_{i=t+1}^{T} u_i$,

$$C_{uu,t} = \sum_{i=1}^{t} (u_i - \bar{u}_t)^2$$

and $$C^*_{uu,t} = \sum_{i=t+1}^{T} (u_i - \bar{u}^*_t)^2 \quad . \qquad (8.2)$$

Because τ is known we are interested solely in the intersection of the confidence interval and $[u_\tau, u_{\tau+1})$. The point estimate of γ is

$$\hat{\gamma} = (\hat{\alpha}_o - \hat{\alpha}_1)/(\hat{\beta}_1 - \hat{\beta}_o).$$

No indication is given as to the procedure to be adopted when τ is unknown, although it is fairly clear that one would calculate

$$\min_{u_t \le \gamma < u_{t+1}} S^2(\gamma) \qquad (t = 2, \ldots, T-2)$$

and take the estimate $\hat{\gamma}$ to be the value of γ giving the overall minimum. No analysis of the distribution of $\hat{\gamma}$ is attempted.

Sprent's results are generalized by Robison (1964), who extends the model (8.1) to two polynomial regression curves. For the two-line special case the results are the same, τ being assumed known throughout. It is, however, pointed out that $\hat{\gamma}$ may not fall inside $[u_\tau, u_{\tau+1})$. If this should happen, Robison suggests that a $100\xi\%$ confidence interval for γ should be constructed and ξ increased until there is an intersection with $[u_\tau, u_{\tau+1})$. The revised estimate of γ is this intersection, that is either u_τ or $u_{\tau+1}$. Such an estimate seems questionable in derivation and usefulness. It would be better to say either (a) the assumption "τ is known" was wrong, or (b) select as the estimate whichever of u_τ and $u_{\tau+1}$ is closer to $\hat{\gamma}$.

A recent paper by Hudson (1966) on point estimation of γ drops the assumption that τ is known and gives a clear step-by-step estimation procedure. The derivation of the procedure is similar to our own, given in the next section, and we shall refer to Hudson's work there. Hudson does not consider the problem of finding the distribution of $\hat{\gamma}$; he suggests that "approximate likelihood regions" would be satisfactory for inference. Several numerical illustrations of the estimation procedure are given.

Aspects not dealt with in the work summarized above are

(i) the distribution properties of the m.l.e. $\hat{\gamma}$

(ii) the information (if any) obtained when starting with an incorrect value of τ and getting a point estimate outside $[u_\tau, u_{\tau+1})$

(iii) tests for the null hypothesis $(\alpha_o, \beta_o) = (\alpha_1, \beta_1)$ and power functions of such tests.

We are concerned in this thesis with (i) and (ii). We shall assume that evidence exists for the two-line model (8.1). The object is to estimate γ by maximum likelihood and to examine the distribution of the estimate. It will turn out that (ii) plays an important role in our discussion.

On the subject of testing $(\alpha_o, \beta_o) = (\alpha_1, \beta_1)$, there has recently appeared some work by Brown and Durbin (1968) based on residuals examination. Some discussion based on an empirical investigation is given in Section 13.

9. Maximum Likelihood Estimation of γ

We assume that the data (u_i, x_i) $(i = 1, \ldots, T)$ are described by (8.1), with all parameters unknown except for σ^2. First consider estimation of γ by an analysis of residuals from two regression lines in the following way. Select a trial value of γ, γ_o say, in some interval $[u_t, u_{t+1})$, and transform the independent variable to $V = u - \gamma_o$. Then fit by the method of least squares the two lines $\hat{\alpha} + \hat{\beta}_o V$ and $\hat{\alpha} + \hat{\beta}_1 V$ for non-positive and positive values of V respectively. Next define the set of differences $d_1, d_2, \ldots, d_T$ by

$$d_s = x_s - \hat{\alpha} - \hat{\beta}_1 V_s \qquad (s = 1, \ldots, t)$$

and
$$d_s = x_s - \hat{\alpha} - \hat{\beta}_o V_s \qquad (s = t + 1, \ldots, T) .$$

These quantities are like residuals, except that they are differences between observations following one model and the estimated values following another model. Assuming that $\beta_1 \neq \beta_o$, the mean value of these differences should increase as $|V|$ increases, since the vertical distance between two lines increases as one moves away from the intersection. Therefore a linear combination of $d_1, \ldots, d_T$ will give a test for the null hypothesis $\beta_1 = \beta_o$. More importantly, the linear combination will, in the mean, increase in magnitude as γ_o approaches the true value of γ; this is because the difference $\hat{\beta}_1 - \hat{\beta}_o$ increases in mean as γ_o tends to the true value. Since the differences d_j increase in mean as $|V_j|$ increases, an obvious choice for the linear combination of $d_1 \ldots, d_T$ is $\sum_{s=1}^{T} \{A + B|V_s|\} d_s$ for some A, B. The trial value γ_o can be varied to achieve the maximum of this linear combination. In fact it is easy to see that the combination is a linear combination of sufficient statistics conditional on $\tau = t$, and is, further, proportional to $(\hat{\beta}_1 - \hat{\beta}_o)$ for all A, B. Here $\hat{\beta}_1$ and $\hat{\beta}_o$ are constrained least squares estimates conditional on $\tau = t$. Transforming back to the independent variable u, the combination $\sum_1^T d_s$ is explicitly a function of γ_o and t. As one would expect, because of the sufficiency referred to, the same combination of $d_1, \ldots, d_T$ arises in the maximum likelihood estimation of γ.

It is important to emphasize that (u_τ, x_τ) is the last sample point which belongs to the regression $E(x) = \alpha_o + \beta_o u$ and that the intersection of the two lines is between $u = u_\tau$ and $u = u_{\tau+1}$. That is we do not consider situations exemplified by Figure 9.0, where $\gamma < u_\tau$.

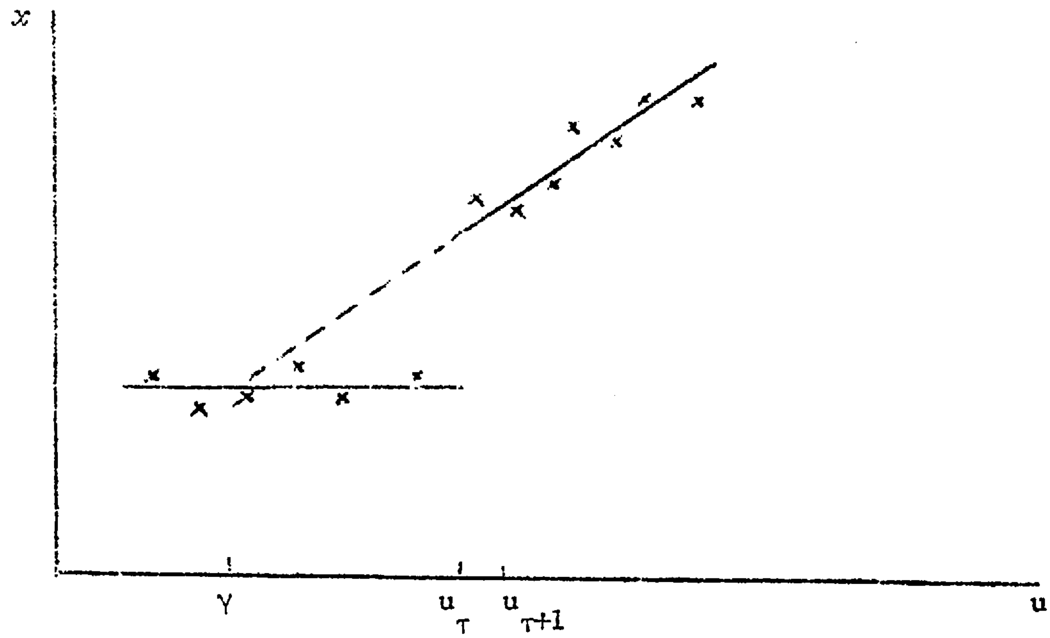


Figure 9.0. Discontinuous change in regression model.

The likelihood of the sample $S = \{(u_1, x_1), \ldots, (u_T, x_T)\}$ is, by (8.1),

$$L(S; \alpha_o, \beta_o, \alpha_1, \beta_1, \gamma, \tau, \sigma)$$

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{T}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{\tau}(x_i - \alpha_o - \beta_o u_i)^2 - \frac{1}{2\sigma^2}\sum_{i=\tau+1}^{T}(x_i - \alpha_1 - \beta_1 u_i)^2\right\} \tag{9.1}$$

where $(\alpha_1 - \alpha_o) + \gamma(\beta_1 - \beta_o) = 0$ and $u_\tau \leq \gamma < u_{\tau+1}$. We regard the u_t's as fixed quantities.

Conditionally on $\tau = t$, we can maximize the likelihood over α_o, β_o, α_1 and β_1 subject to the linear constraint on them, and obtain a function $L_t(\gamma)$, say, which is the marginal likelihood function of γ for $u_t \leq \gamma < u_{t+1}$. Doing the same thing for all admissible values of t, i.e. t = 2, 3, ..., T-2, we obtain a sequence of such functions which are pieces of the marginal likelihood function of γ. That is, if $L(\gamma)$ is the marginal likelihood function of γ then

$$L(\gamma) = L_t(\gamma), \ u_t \leq \gamma < u_{t+1} \qquad (t = 2, 3, \ldots, T-2).$$

It is important to note that $L_t(\gamma)$ is mathematically defined for all γ, but because of the restriction $u_\tau \leq \gamma < u_{\tau+1}$, here it has a statistical interpretation only for $u_t \leq \gamma < u_{t+1}$. We can write $L_t(\gamma)$ in the form

$$L_t(\gamma) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{T}{2}} \exp\left\{-\frac{1}{2\sigma^2} \ S_t^2(\gamma)\right\},$$

where $S_t^2(\gamma)$ is the residual sum of squares for two regression lines constrained to meet at $u = \gamma$. The formula for $S_t^2(\gamma)$ is given by Sprent (1961). If σ^2 is unknown, then

$$L_t(\gamma) = \left\{\frac{T}{2\pi S_t^2(\gamma)}\right\}^{\frac{T}{2}} .$$

Now under the null hypothesis $\beta_1 = \beta_0$ we get the corresponding maximised likelihood

$$L_0 = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{T}{2}} \exp\left(-\frac{1}{2\sigma^2} S_0^2\right) \text{, say,}$$

when σ^2 is known, or

$$L_o = \left(\frac{T}{2\pi S_o^2} \right)^{\frac{T}{2}}$$

when σ^2 is unknown.

Since S_o^2 is independent of γ, the m.l.e. $\hat{\gamma}$ will be the value of γ which maximises $L(\gamma)/L_o$, i.e. the value of γ which maximises $S_o^2 - S_t^2(\gamma)$, taking $u_t \leq \gamma < u_{t+1}$ and letting t run from 2, 3, ..., T-2. This is true for σ^2 known or unknown; we assume from now on that σ^2 is known.

Let the difference $S_o^2 - S_t^2(\gamma)$ be denoted by $Z_t^2(\gamma)$, then we deduce from Sprent (1961, p. 638) that

$$Z_t^2(\gamma) = \frac{(A_t - B_t\gamma)^2}{C_{uu,T}\,(C_t - 2D_t\gamma + E_t\gamma^2)}, \qquad (9.2)$$

where

$$A_t = C_{ux,t}\,C^*_{uu,t} - C^*_{ux,t}\,C_{uu,t} - (\bar{x}^*_t - \bar{x}_t)\,D_t$$

$$+ \frac{tt^*}{T}(\bar{u}^*_t - \bar{u}_t)(\bar{u}^*_t\,C_{ux,t} + \bar{u}_t C^*_{ux,t}),$$

$$B_t = \frac{tt^*}{T}\left\{(\bar{x}_t - \bar{x}^*_t)\,C_{uu,T} - (\bar{u}_t - \bar{u}^*_t)\,C_{ux,T}\right\},$$

$$C_t = C_{uu,t}\,C^*_{uu,t} + \frac{tt^*}{T}(\bar{u}^2_t\,C^*_{uu,t} + \bar{u}^{*2}_t\,C_{uu,t}),$$

$$D_t = \frac{tt^*}{T}(\bar{u}_t\,C^*_{uu,t} + \bar{u}^*_t\,C_{uu,t}),$$

and

$$E_t = \frac{tt^*}{T}(C_{uu,t} + C^*_{uu,t}). \qquad (9.3)$$

These involve the sample moments conditional on t,

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

$$C_{ux,t} = \sum_{i=1}^{t} u_i x_i - t\bar{u}_t \bar{x}_t$$

$$C^*_{ux,t} = \sum_{i=t+1}^{T} u_i x_i - (T-t)\, \bar{u}^*_t \bar{x}^*_t$$

and $t^* = T-t$.

The other expressions $C_{uu,t}$, $C^*_{uu,t}$, $\bar{u}_t$ and $\bar{u}^*_t$ were defined in (8.2)

Since A_t and B_t are defined in terms of means and cross-products, we can simplify their expression by introducing the following definitions.

Local maximum likelihood estimates

Let $\tilde{\gamma}_t$ be the value of γ at which $L_t(\gamma)$ reaches its (unique) maximum. Then conditional on $\tau = t$, $\tilde{\gamma}_t$ is the m.l.e. of γ not constrained to lie in $[u_t, u_{t+1})$. The corresponding estimates of α_o β_o, α_1 and β_1 are the standard least squares estimates

$$\begin{aligned} \tilde{\beta}_{ot} &= C_{ux,t}/C_{uu,t} , & \tilde{\beta}_{1t} &= C^*_{ux,t}/C^*_{uu,t} , \\ \tilde{\alpha}_{ot} &= \bar{x}_t - \tilde{\beta}_{ot}\,\bar{u}_t , & \tilde{\alpha}_{1t} &= \bar{x}^*_t - \tilde{\beta}_{1t}\,\bar{u}^*_t \end{aligned} \qquad (9.4)$$

The relation between these and $\tilde{\gamma}_t$ is

$$\tilde{\gamma}_t = \frac{\tilde{\alpha}_{ot} - \tilde{\alpha}_{1t}}{\tilde{\beta}_{1t} - \tilde{\beta}_{ot}} . \qquad (9.5)$$

All of these estimates will be referred to as local m.l.e.'s The overall m.l.e's, which maximize (9.1), will be denoted by $\hat{\alpha}_o$, $\hat{\beta}_o$, $\hat{\alpha}_1$, $\hat{\beta}_1$, $\hat{\gamma}$ and $\hat{\tau}$ and satisfy the relations

$$\hat{\gamma} = \frac{\hat{\alpha}_o - \hat{\alpha}_1}{\hat{\beta}_1 - \hat{\beta}_o} \quad \text{and } u_{\hat{\tau}} \leq \hat{\gamma} < u_{\hat{\tau}+1} \ .$$

Substituting in (9.3) from (9.4) and (9.5) we get

$$A_t = (\tilde{\beta}_{ot} - \tilde{\beta}_{1t})(C_t - D_t \tilde{\gamma}_t)$$

and
$$B_t = (\tilde{\beta}_{ot} - \tilde{\beta}_{1t})(D_t - E_t \tilde{\gamma}_t) \ ,$$

so that (9.2) becomes

$$Z_t^2 (\gamma) = \frac{(\tilde{\beta}_{1t} - \tilde{\beta}_{ot})^2 \left\{ C_t - D_t(\tilde{\gamma}_t + \gamma) + E_t \tilde{\gamma}_t \gamma \right\}^2}{C_{uu,T} (C_t - 2D_t \gamma + E_t \gamma^2)}$$

$$(t = 2, 3, \ldots, T - 2) \qquad (9.6)$$

We shall study $Z_t(\gamma)$ in some detail in order to obtain useful results about $\hat{\gamma}$ and the sequence $\{\tilde{\gamma}_t\}$. The function $Z_t(\gamma)$ is the most convenient form equivalent to $L_t(\gamma)$; Hudson (1966) worked with $S_t^2(\gamma)$ and leaned more toward geometrical proof of the same results.

First, the likelihood $L(\gamma)$ is continuous. By inspection of (9.6) it is clear that continuity holds for $u_t < \gamma < u_{t+1}$ and for all t. It remains to show that $Z_t(u_{t+1}) = Z_{t+1}(u_{t+1})$ $(t = 2, \ldots, T - 3)$. We deduce that this is so after some lengthy calculation has shown that

$$A_t - B_t \; u_{t+1} = A_{t+1} - B_{t+1} \; u_{t+1}$$

and

$$C_t - 2D_t u_{t+1} + E_t u_{t+1}^2 = C_{t+1} - 2D_{t+1} u_{t+1} + E_{t+1} u_{t+1}^2 \qquad (9.7)$$

Secondly, $L_t(\gamma)$ has the following property:
If $\overline{u}_t \leq \tilde{\gamma}_t \leq \overline{u}_t^*$ then $L_t(\gamma)$ decreases as $|\gamma - \tilde{\gamma}_t|$ increases provided that $\overline{u}_t \leq \gamma \leq \overline{u}_t^*$. (9.8)

A geometrical proof of this, in terms of $S_t^2(\gamma)$, by McLaren (1965) was used by Hudson (1966). We give below an analytic proof which results in a relaxation of the constraint on γ (or $\tilde{\gamma}_t$).

Let the minimum of $Z_t(\gamma)$ occur at $\gamma = \delta_t$. Then from (9.6) we see that δ_t is unique and given by

$$\delta_t = \frac{C_t - D_t \tilde{\gamma}_t}{D_t - E_t \tilde{\gamma}_t} \qquad (t = 2, \ldots, T-2). \qquad (9.9)$$

Incidentally $L_t(\delta_t) = L_o$. Now if $\delta_t < \tilde{\gamma}_t$, then $Z_t(\gamma)$ decreases as $|\gamma - \tilde{\gamma}_t|$ increases, so long as $\gamma \geq \delta_t$; if $\delta_t > \tilde{\gamma}_t$ then $Z_t(\gamma)$ decreases as $|\gamma - \tilde{\gamma}_t|$ increases, so long as $\gamma \leq \delta_t$. We can now rephrase (9.8):

If $\overline{u}_t \leq \tilde{\gamma}_t \leq \overline{u}_t^*$ then $\delta_t \leq \overline{u}_t$ or $\delta_t \geq \overline{u}_t^*$. (9.10)

To prove this, first define

$$h_t(y) = \frac{C_t - D_t y}{D_t - E_t y} \qquad (t = 2, \ldots, T-2)$$

which has derivative

$$h_t'(y) = \frac{(C_t E_t - D_t^2)}{(D_t - E_t y)^2} \quad ,$$

whenever this exists.

Using the definitions of C_t, D_t and E_t it is easy to see that

$$C_t E_t - D_t^2 = \frac{tt^*}{T} C_{uu,t} C^*_{uu,t} \left\{ C_{uu,t} + C^*_{uu,t} + \frac{tt^*}{T} (\overline{u}_t - \overline{u}^*_t)^2 \right\} > 0,$$

so that h_t (y) is everywhere increasing except at the point of discontinuity $y_t = \frac{D_t}{E_t}$.

Clearly $h_t(y_t -) = \infty$ and $h_t(y_t +) = -\infty$.

By definition

$$y_t = (\overline{u}_t C^*_{uu,t} + \overline{u}^*_t C_{uu,t})/(C_{uu,t} + C^*_{uu,t})$$

and hence $\overline{u}_t < y_t < \overline{u}^*_t$.

We have now to establish that

$$h_t(\overline{u}_t) > \overline{u}^*_t \quad \text{and} \quad h_t(\overline{u}^*_t) < \overline{u}_t \ ,$$

for then

$$\overline{u}_t \le y \le \overline{u}^*_t \quad \text{implies} \quad h_t(y) < \overline{u}_t \ \text{or} \ h_t(y) > \overline{u}^*_t \ ,$$

which is (9.10) with y written instead of $\tilde{y}_t$.

A little calculation gives

$$h_t(\overline{u}_t) = \overline{u}^*_t + \frac{TC^*_{uu,t}}{tt^*(\overline{u}^*_t - \overline{u}_t)} > \overline{u}^*_t$$

and $h_t(\bar{u}_t^*) = \bar{u}_t - \frac{TC_{uu,t}}{tt^*(\bar{u}_t^* - \bar{u}_t)} < \bar{u}_t$,

the required result. In fact this shows that (9.8) may be re-written either

(i) If $\bar{u}_t \le \tilde{\gamma}_t \le \bar{u}_t^*$ then $L_t(\gamma)$ decreases as $|\gamma - \tilde{\gamma}_t|$ increases for

$$\bar{u}_t - \frac{TC_{uu,t}}{tt^*(\bar{u}_t^* - \bar{u}_t)} \le \gamma \le \bar{u}_t^* + \frac{TC^*_{uu,t}}{tt^*(\bar{u}_t^* - \bar{u}_t)} \qquad (9.11)$$

or

(ii) If $\bar{u}_t - \frac{TC_{uu,t}}{tt^*(\bar{u}_t^* - \bar{u}_t)} \le \tilde{\gamma}_t \le \bar{u}_t^* + \frac{TC^*_{uu,t}}{tt^*(\bar{u}_t^* - \bar{u}_t)}$

then $L_t(\gamma)$ decreases as $|\gamma - \tilde{\gamma}_t|$ increases for $\bar{u}_t \le \gamma \le \bar{u}_t^*$. (9.12)

The more useful form for our present purpose is (9.12).

We now set out some useful results about $\hat{\gamma}$.

First, $u_t < \hat{\gamma} < u_{t+1}$ implies $u_t < \tilde{\gamma}_t < u_{t+1}$ (9.13)

since $L_t(\tilde{\gamma}_t) > L_t(\gamma)$ for $\gamma \ne \tilde{\gamma}_t$ and no other local maximum of $L_t(\gamma)$ exists.

Second, if $\bar{u}_t^* + \frac{TC^*_{uu,t}}{tt^*(\bar{u}_t^* - \bar{u}_t)} \ge \tilde{\gamma}_t > u_{t+1}$,

then $L_t(u_t) < L_t(u_{t+1})$ by virtue of (9.12).

Third, if $\bar{u}_t - \frac{TC_{uu,t}}{tt^*(\bar{u}_t^* - \bar{u}_t)} \le \tilde{\gamma}_t < u_t$

then $L_t(u_t) > L_t(u_{t+1})$,

again by virtue of (9.12). The last two results indicate which of u_t and u_{t+1} could possibly be $\hat{\gamma}$, thus eliminating unnecessary calculation of both $L_t(u_t)$ and $L_t(u_{t+1})$.

We can now set down the procedure for calculating $\hat{\gamma}$ from the data $\left\{(u_1, x_1), \ldots, (u_T, x_T)\right\}$. It is the same as that derived independently by Hudson (1966) with the slight improvement gained by using (9.12) instead of (9.8).

Procedure for finding $\hat{\gamma}$

To facilitate the use of the estimation procedure we set it out in a logical step-by-step form which might easily be translated into a computer program.

The following symbols are employed throughout:

M the maximum over s and γ of all previous admissible values of $Z_s^2(\gamma)$

G the value of γ at which M was attained

I the interval enclosing G - i.e. $u_I \le G < u_{I+1}$.

By previous admissible values of $Z_s^2(\gamma)$ we mean, when examining $Z_t^2(\gamma)$, the set $Z_s^2(\gamma)$; $u_s \le \gamma < u_{s+1}$, $2 \le s \le t - 1$.

The procedure is, then,

A (i) calculate $\tilde{\alpha}_{o2}$, $\tilde{\beta}_{o2}$, $\tilde{\alpha}_{12}$, $\tilde{\beta}_{12}$ and hence $\tilde{\gamma}_2$

A (ii) if $u_2 \le \tilde{\gamma}_2 < u_3$, set $M = Z_2^2(\tilde{\gamma}_2)$, $G = \tilde{\gamma}_2$, $I = 2$ and go to step B (i)

A (iii) if $\bar{u}_2 - \frac{TC_{uu,2}}{2(T-2)(\bar{u}_2^*-\bar{u}_2)} \leq \check{\gamma}_2 < u_2$, set $M = Z_2^2(u_2)$, $G = u_2$, $I = 2$ and go to B (i)

A (iv) if $u_3 \leq \check{\gamma}_2 < \bar{u}_2^* + \frac{TC^*_{uu,2}}{2(T-2)(\bar{u}_2^*-\bar{u}_2)}$, set $M = Z_2^2(u_3)$, $G = u_3$, $I = 3$ and go to B (i)

A (v) if $\check{\gamma}_2 < \bar{u}_2 - \frac{TC_{uu,2}}{2(T-2)(\bar{u}_2^*-\bar{u}_2)}$ or $\check{\gamma}_2 > \bar{u}_2^* + \frac{TC^*_{uu,2}}{2(T-2)(\bar{u}_2^*-\bar{u}_2)}$,

set $M = \max\left\{Z_2^2(u_2),\ Z_2^2(u_3)\right\}$,

$$G = \begin{cases} u_2 & \text{if } M = Z_2^2(u_2) \\ u_3 & \text{if } M = Z_2^2(u_3) \end{cases}, \qquad I = \begin{cases} 2 & \text{if } M = Z_2^2(u_2) \\ 3 & \text{if } M = Z_2^2(u_3) \end{cases}$$

and go to B (i)

B (i) set $t = t + 1$ and go to END if $t = T - 1$, otherwise calculate $\check{\alpha}_{jt}$, $\tilde{\beta}_{jt}$ $(j = 0, 1)$ and $\check{\gamma}_t$

B (ii) calculate $Z_t^2(\check{\gamma}_t) = \frac{(\check{\beta}_{1t} - \tilde{\beta}_{ot})^2 (C_t - 2D_t\check{\gamma}_t + E_t\,\check{\gamma}_t^2)}{C_{uu,T}}$

B (iii) if $Z_t^2(\check{\gamma}_t) \leq M$ go to B (i)

B (iv) if $u_t \leq \check{\gamma}_t < u_{t+1}$ set $M = Z_t^2(\check{\gamma}_t)$, $G = \check{\gamma}_t$, $I = t$ and go to B (i)

B (v) if $\bar{u}_t - \frac{TC_{uu,T}}{tt^*(\bar{u}_t^* - \bar{u}_t)} \leq \check{\gamma}_t < u_t$

calculate $Z_t^2(u_t)$; then if $Z_t^2(u_t) \leq M$ go to B (i), if $Z_t^2(u_t) > M$ set $M = Z_t^2(u_t)$, $G = u_t$, $I = t$ and go to B (i)

B (vi) if $u_{t+1} \leq \tilde{\gamma}_t \leq \overline{u}^*_t + \dfrac{TC^*_{uu,t}}{tt^*(\overline{u}^*_t - \overline{u}_t)}$

calculate $Z^2_t(u_{t+1})$; then if $Z^2_t(u_{t+1}) \leq M$ go to B (i),

if $Z^2_t(u_{t+1}) > M$ set $M = Z^2_t(u_{t+1})$, $G = u_{t+1}$, $I = t + 1$

and go to B (i)

B (vii) if $\tilde{\gamma}_t$ is anywhere else calculate both $Z^2_t(u_t)$ and $Z^2_t(u_{t+1})$;

then if $\max\left\{Z^2_t(u_t),\ Z^2_t(u_{t+1})\right\} \leq M$ go to B (i);

if not set $M = \max\left\{Z^2_t(u_t),\ Z^2_t(u_{t+1})\right\}$,

$$(G,I) = \begin{cases} (u_t, t) & \text{if } M = Z^2_t(u_t) \\ (u_{t+1}, t+1) & \text{if } M = Z^2_t(u_{t+1}) \end{cases}$$

and go to B (i).

END The search for the likelihood maximum is complete: the m.l.e.'s are $\hat{\gamma} = G$ and $\hat{\tau} = I$.

One small point to note is that if $G = u_{t+1}$ as a result of calculations in B (vi) or B (vii), the calculation of $Z^2_{t+1}(u_{t+1})$ is unnecessary in the next cycle B (i) - B (vii) since $Z^2_t(u_{t+1}) = Z^2_{t+1}(u_{t+1})$.

Usually we also wish to determine $\hat{\alpha}_o$, $\hat{\alpha}_1$, $\hat{\beta}_o$ and $\hat{\beta}_1$ in addition to $\hat{\gamma}$. Although this is not included in the above procedure, the necessary steps can be incorporated with little difficulty. If $u_I < G < u_{I+1}$ at the END stage,

$$\hat{\alpha}_j = \tilde{\alpha}_{jI} \text{ and } \hat{\beta}_j = \tilde{\beta}_{jI} \qquad (j = 0, 1).$$

If $G = u_I$, a constrained least squares fit is necessary. Running m.l.e.'s of $\hat{\alpha}_o$, $\hat{\alpha}_1$, $\hat{\beta}_o$ and $\hat{\beta}_1$ could be used, analogous to G for $\hat{\gamma}$.

For large samples the calculation involved in using this procedure could be considerable. If the approximate position of $\hat{\gamma}$ can be determined prior to the calculation (e.g. graphically), some reduction in the number of values of t examined might be achieved. However if $(\beta_1 - \beta_o)/\sigma$ is small such prior location could be difficult.

It is intuitively reasonable that there is more information in each $\tilde{\gamma}_t$ than we have used so far, since successive values are very much dependent on one another. For example, if $\hat{\gamma} = \tilde{\gamma}_t$ then $\tilde{\gamma}_{t+1}$ should be close to $\hat{\gamma}$, more so the larger t and T - t become. To illustrate what we mean we look at two examples, in both of which the m.l. estimation procedure is used. In each case a sample of points from the model (8.1) was generated using pseudo random normal deviates for the errors and having $u_t = t$, i.e. equally spaced observations.

Example 1

Sample size T = 20, with $\beta_1 - \beta_o = \sigma$ and $\gamma = 10.5$. A plot of the data is given in Figure 9.1, and in Figure 9.2 we have a plot of $Z_t^2(\gamma)$, over t and γ, at intervals 0.25. From the data

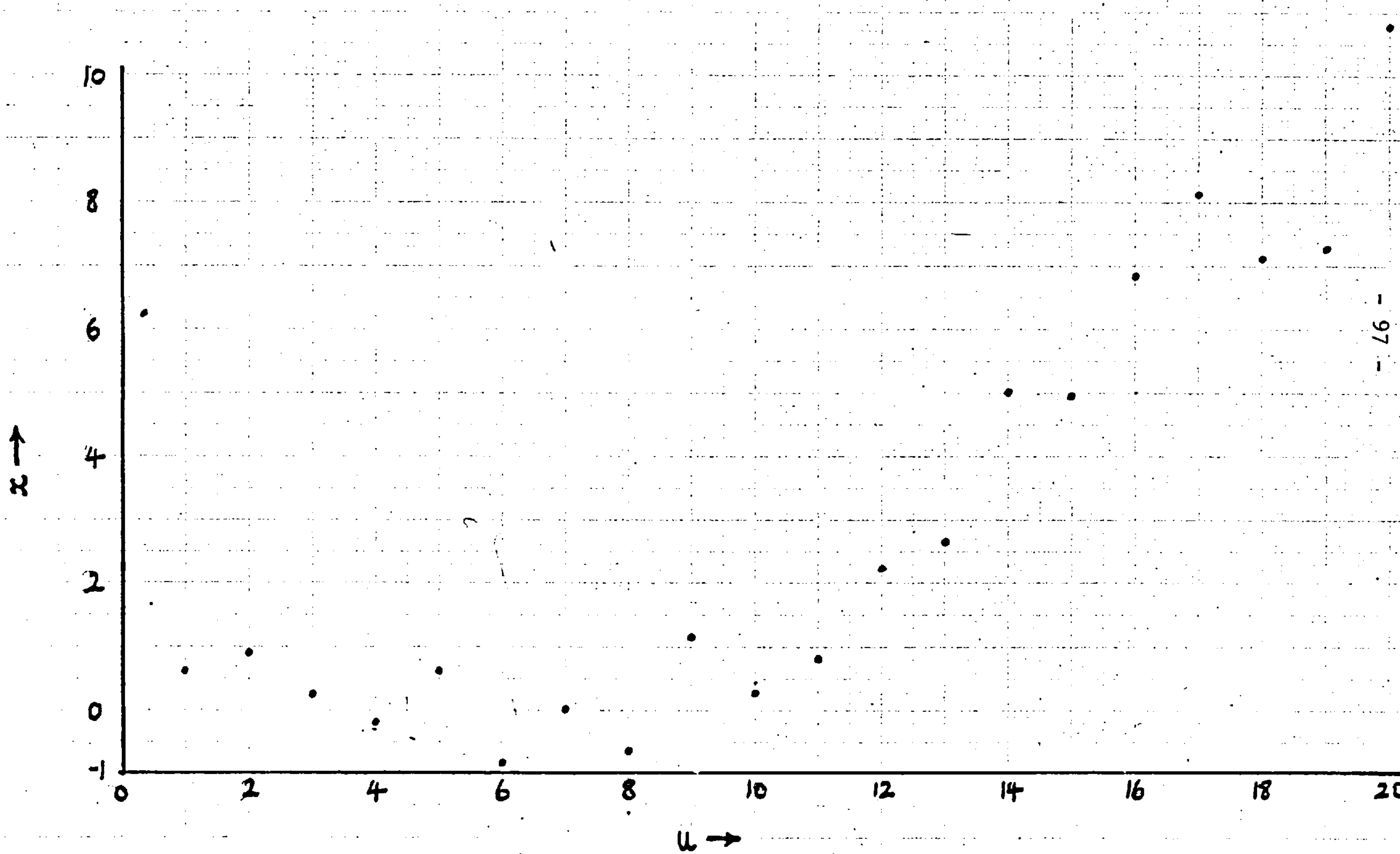


FIGURE 9.1 Scatter diagram for Example 1

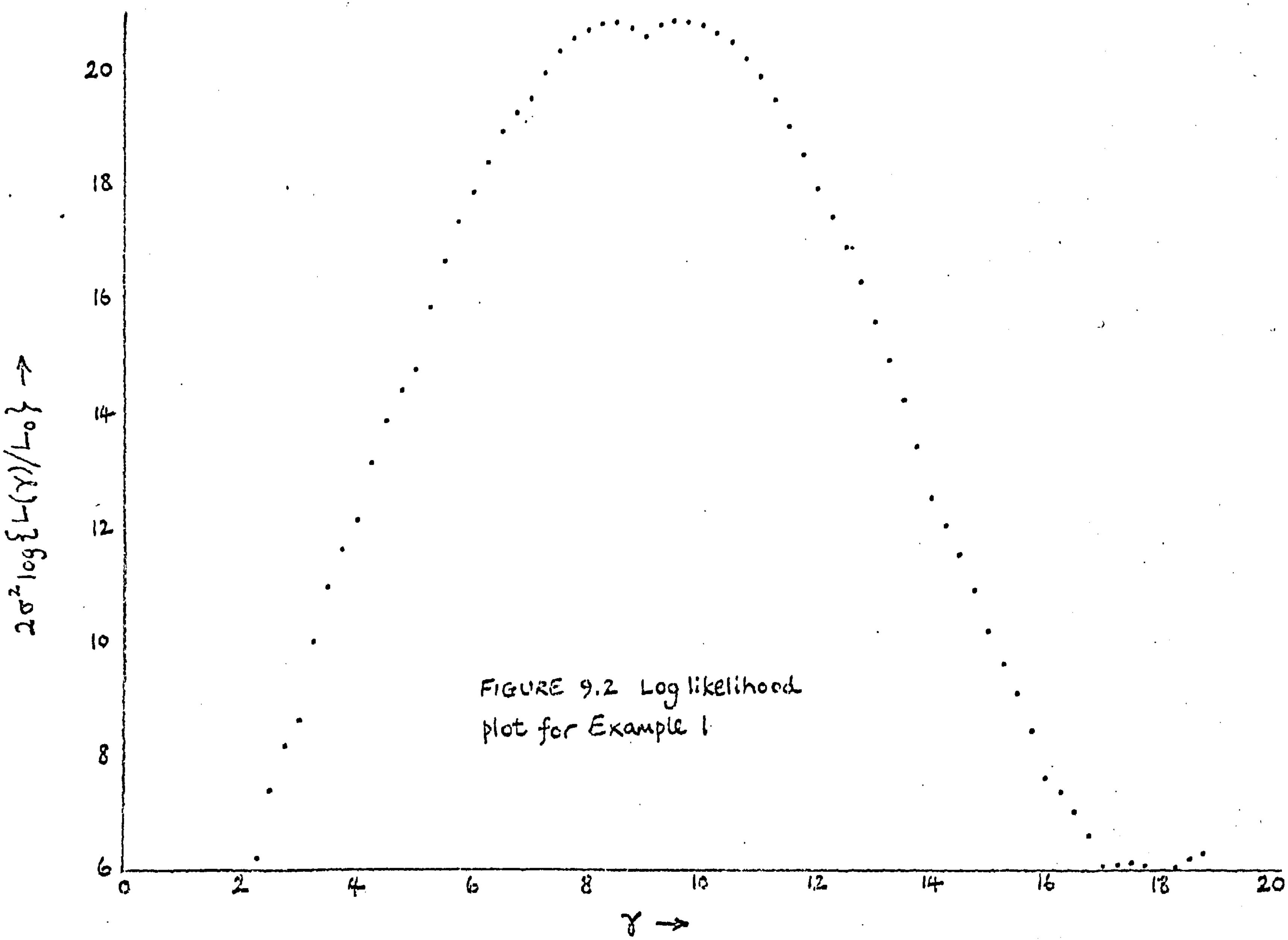


FIGURE 9.2 Log likelihood plot for Example 1

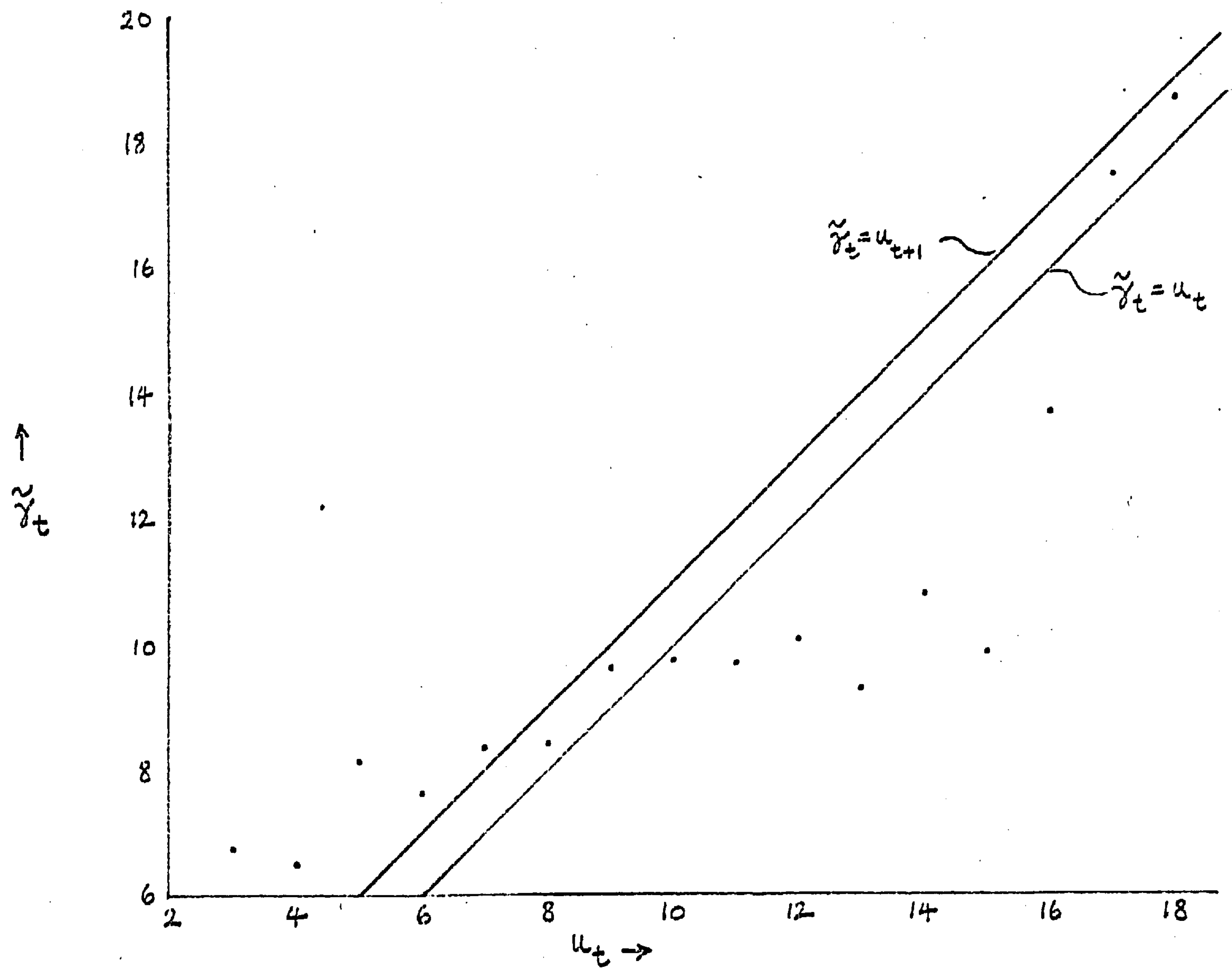


FIGURE 9.3 Plot of $\tilde{\gamma}_t$ against u_t for Example 1

plot it is clear that $8 \le \hat{\gamma} \le 12$, so that a complete likelihood search is unnecessarily lengthy. The value of $\hat{\gamma}$ is 9.5 although from Figure 9.2 it is apparent that any value of γ between 8.5 and 10.0 would come close to maximising the log likelihood; a confidence region for γ would therefore probably be quite large (cf. Section 13).

To see what information is contained in the local estimates $\tilde{\gamma}_t$ we have plotted $\tilde{\gamma}_t$ against t in Figure 9.3. The two diagonal lines on the graph correspond to $\tilde{\gamma}_t = u_t$ and $\tilde{\gamma}_t = u_{t+1}$ as indicated. An outstanding feature of the graph is the tendency for $\tilde{\gamma}_t$ to be larger than u_t when $\hat{\gamma} > u_t$, and similarly for $\tilde{\gamma}_t$ to be smaller than u_{t+1} when $\hat{\gamma} < u_{t+1}$. The principal exceptions are $\tilde{\gamma}_8$ and $\tilde{\gamma}_9$, either of which could be taken as $\hat{\gamma}$ so close are the corresponding log likelihood values. The other exceptions are $\tilde{\gamma}_{17}$ and $\tilde{\gamma}_{18}$, and here we notice that in the scatter diagram (Figure 9.1) a second change-point γ' is indicated, although the log likelihood is relatively small in that neighbourhood. A further point of interest in Figure 9.3 is that in the neighbourhood of u = 9 the local estimates $\tilde{\gamma}_t$ fluctuate near $\hat{\gamma}$; the inter-dependence between successive $\tilde{\gamma}_t$'s would account for this, and might be expected to have a stronger effect in larger samples.

Example 2

In this second example we take a larger sample size $T = 50$, with $\beta_1 - \beta_o = 0.4\sigma$ and $\gamma = 25$. Figures 9.4, 9.5 and 9.6 give the scatter diagram of the generated data, the log likelihood plot and the plot of $\tilde{\gamma}_t$ versus t, respectively. The log likelihood is smoother than in Example 1 as we would expect with a larger sample. This smoothness is reflected in the plot of $\tilde{\gamma}_t$ against t, where the $\tilde{\gamma}_t$'s fluctuate close to $\hat{\gamma}$. Note that our remarks about the information contained in $\tilde{\gamma}_t$ are well substantiated, i.e. $\tilde{\gamma}_t > u_t$ when $\hat{\gamma} > u_t$ and $\tilde{\gamma}_t < u_{t+1}$ when $\hat{\gamma} < u_{t+1}$.

The conclusions that we draw from these and other examples are

(i) as T increases the likelihood becomes smoother

(ii) as T increases the $\tilde{\gamma}_t$'s fluctuate less around $\hat{\gamma}$, but not relatively less in terms of $1/\sqrt{T}$

(iii) the statements "$\tilde{\gamma}_t > u_t$ implies $\hat{\gamma} > u_t$" and "$\tilde{\gamma}_t < u_{t+1}$ implies $\hat{\gamma} < u_{t+1}$" hold with more regularity as T increases, or rather as t and T - t both increase.

From previous remarks it is clear that (i), (ii) and (iii) are intimately connected.

Conclusion (iii) has a direct application in the search for $\hat{\gamma}$. For large τ and $T - \tau$, a procedure converging on $\hat{\gamma}$ from a trial value could be set up:

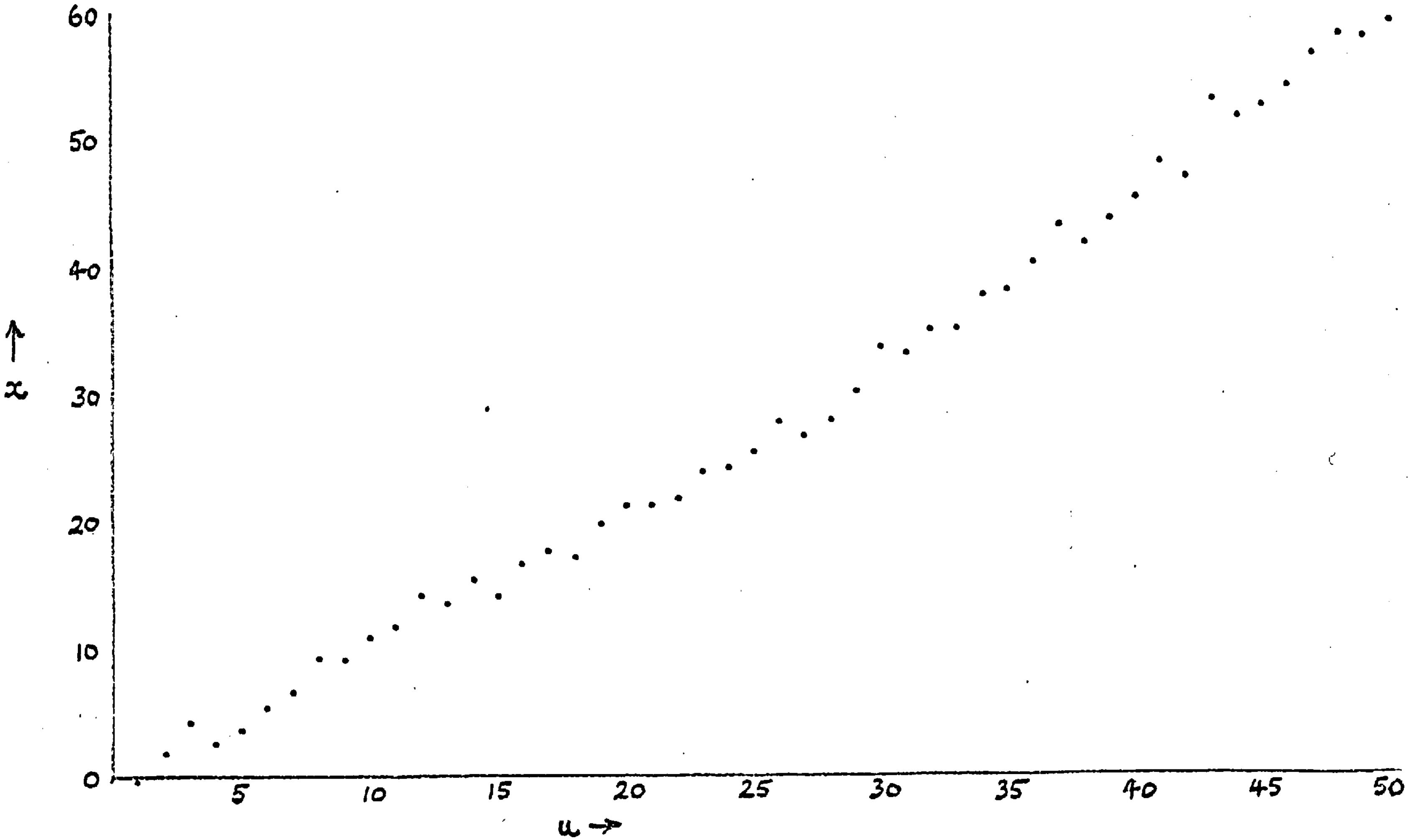


FIGURE 9.4 Scatter diagram for Example 2.

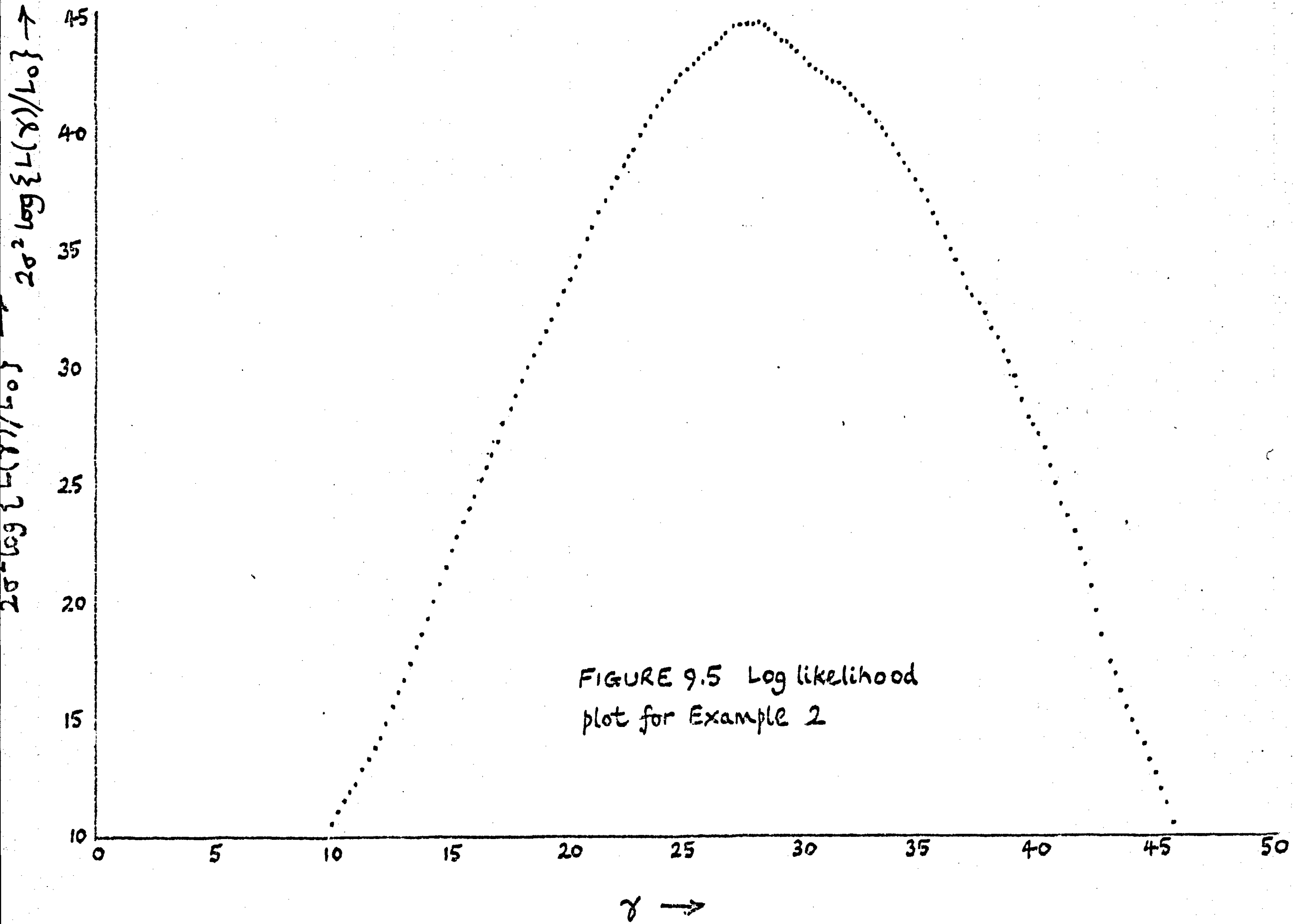


FIGURE 9.5 Log likelihood plot for Example 2

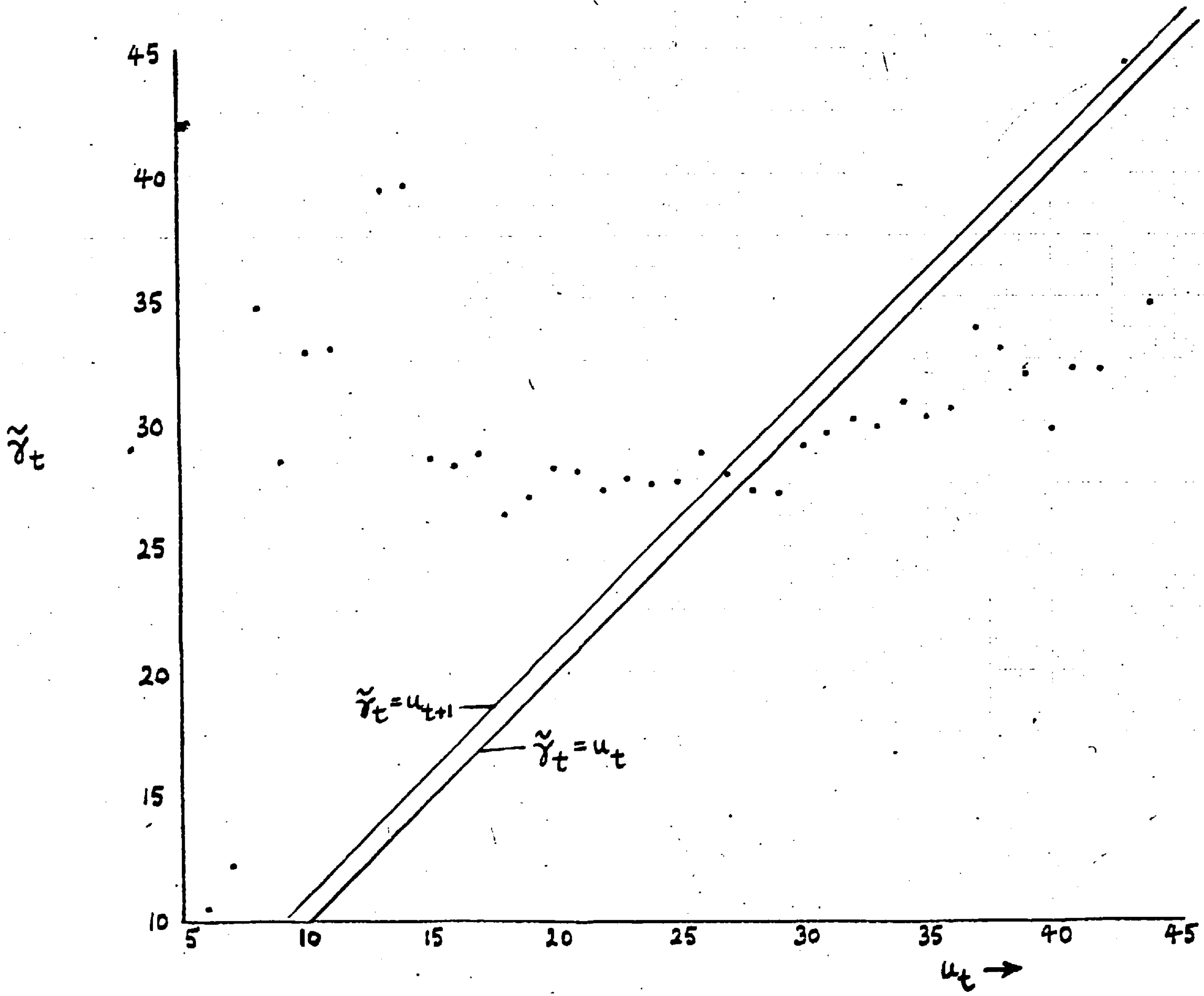


FIGURE 9.6 Plot of $\tilde{\gamma}_t$ against u_t for Example 2

Select an initial value $t = t_o$ (not near 1 or T to avoid possible end effects) and calculate $\tilde{\gamma}_{t_o}$. If $u_{t_o} \leq \tilde{\gamma}_{t_o} < u_{t_o+1}$ calculate $\tilde{\gamma}_t$, $Z_t^2(u_t)$, $Z_t^2(u_{t+1})$ and $Z_t^2(\tilde{\gamma}_t)$ for $t = t_o-1, t_o, t_o+1$. Then let $\hat{\gamma}$ be the point at which $Z_t^2(\gamma)$ achieves its maximum in $[u_{t_o-1}, u_{t_o+1}]$, where γ and t range over the appropriate values. If, however, $\tilde{\gamma}_{t_o}$ is not in $[u_{t_o}, u_{t_o+1})$ it will be in some interval $[u_{t_1}, u_{t_1+1})$, say; so calculate $\tilde{\gamma}_{t_1}$. Repeat the process described for $\tilde{\gamma}_{t_o}$ until a local estimate $\tilde{\gamma}_{t_k}$ is in the interval $[u_{t_k}, u_{t_k+1})$. Finally calculate $\tilde{\gamma}_t$, $Z_t^2(u_t)$, $Z_t^2(u_{t+1})$ and $Z_t^2(\tilde{\gamma}_t)$ for $t = t_k-1, t_k, t_k+1$ and let $\hat{\gamma}$ be the point at which the maximum of these values occurs.

This would be only the outline for a working procedure. We shall not discuss the efficiency of this or more sophisticated "convergent" procedures. However, it is worth remarking that in Example 2 with t_o anywhere between 5 and 45 (with the exception of 7 and 12), the above procedure converges in not more than 6 steps to $\hat{\gamma}$: usually 4 or less steps are needed.

We look next at another use of conclusion (iii), that of approximating the distribution of $\hat{\gamma}$.

10. Approximating to the Distribution of the m.l.e.

Feder and Sylwester (1968) have shown in some unpublished work that $\hat{\gamma}$ is asymptotically normally distributed, but the empirical study described in Section 11 suggests that the normal distribution is inadequate as a finite sample approximation. However the asymptotic normality of $\hat{\gamma}$ does indicate a degree of asymptotic smoothness of $L(\gamma)$ that is useful in the following discussion which concerns alternative asymptotic results for the distribution of $\hat{\gamma}$. The central result establishes a connection between $\hat{\gamma}$ and the $\tilde{\gamma}_t$'s.

Now $\hat{\gamma}$ is difficult to work with because it has no explicit definition. But the asymptotic normality of $\hat{\gamma}$ indicates that $\hat{\gamma} = u_s$ with zero probability (at least asymptotically), hence the statement

$$\tilde{\gamma}_t > u_t \quad \text{implies} \quad \tilde{\gamma}_{t-1} > u_t \quad (t < t_o) \tag{10.1}$$

is equivalent to

$$\tilde{\gamma}_t > u_t \quad \text{implies} \quad \hat{\gamma} > u_t \qquad (t < t_o)$$

by the definition of $\hat{\gamma}$. We show that (10.1) holds with high probability. Without loss of generality we assume that $\beta_1 > \beta_o$, also that the u_t's are equally spaced.

Since the numerator and denominator of $\log L(\gamma)$ are continuous by (9.7), we see from (9.6) that

$$\left\{C_t - D_t\,(\tilde{\gamma}_t + u_{t+1}) + E_t\,\tilde{\gamma}_t\,u_{t+1}\right\}\tilde{\beta}_t$$
$$= \left\{C_{t+1} - D_{t+1}\,(\tilde{\gamma}_{t+1} + u_{t+1}) + E_{t+1}\,\tilde{\gamma}_{t+1}\,u_{t+1}\right\}\tilde{\beta}_{t+1} \tag{10.2}$$

where $\tilde{\beta}_s = \tilde{\beta}_{1s} - \tilde{\beta}_{os}$ $(s = 2, 3, \ldots, T - 2)$;

also

$$C_t - 2D_t\,u_{t+1} + E_t\,u^2_{t+1} = C_{t+1} - 2D_{t+1}\,u_{t+1} + E_{t+1}\,u^2_{t+1}\,. \tag{10.3}$$

Combining (10.2) and (10.3) we then have

$$\left\{Q_t + (D_t - E_t\,u_{t+1})(u_{t+1} - \tilde{\gamma}_t)\right\}\tilde{\beta}_t$$
$$= \left\{Q_t + (D_{t+1} - E_{t+1}\,u_{t+1})(u_{t+1} - \tilde{\gamma}_{t+1})\right\}\tilde{\beta}_{t+1}, \tag{10.4}$$

where $Q_t = C_t - 2D_t\,u_{t+1} + E_t\,u^2_{t+1}$.

Now when the u_t's are equally spaced, as we assume, simple calculation shows that both $D_t - E_t\,u_{t+1}$ and $D_{t+1} - E_{t+1}\,u_{t+1}$ are negative or positive according as $t < \frac{T-3}{2}$ or $t > \frac{T+1}{2}$, respectively. Also $Q_t > 0$ for all t.

Then for $t < \frac{T-1}{2}$, we have by (10.4) that

$$\text{pr}(\tilde{\gamma}_{t-1} > u_t \mid \tilde{\gamma}_t = u_t + \epsilon\,)$$
$$= \text{pr}\left\{\frac{\tilde{\beta}_{t-1}}{\tilde{\beta}_t} < 1 + \frac{\epsilon(E_t\,u_t - D_t)}{Q_{t-1}}\right\}. \tag{10.5}$$

Evaluating C_t, D_t and E_t for the case $u_t = t$, we find that (10.5) becomes

$$\mathrm{pr}(\tilde{\gamma}_{t-1} > u_t | \tilde{\gamma}_t = u_t + \epsilon) \sim \mathrm{pr}\left\{\frac{\tilde{\beta}_{t-1}}{\tilde{\beta}_t} < 1 + \frac{3(t^* - t)}{2tt^*}\epsilon\right\},$$

and hence

$$\mathrm{pr}(\tilde{\gamma}_{t-1} > u_t | \tilde{\gamma}_t = u_t + \epsilon)$$
$$\sim \mathrm{pr}(\tilde{\beta}_t > 0)\ \mathrm{pr}\left\{\tilde{\beta}_{t-1} - \tilde{\beta}_t\left[1 + \frac{3(t^* - t)}{2tt^*}\epsilon\right] > 0\right\}, \qquad (10.6)$$

since $\beta_1 > \beta_0$. Now the mean and variance of $\tilde{\beta}_t$ are $O(1)$ and $O(T^{-3})$, and the mean and variance of

$$\tilde{\beta}_{t-1} - \tilde{\beta}_t\left[1 + \frac{3(t^* - t)}{2tt^*}\epsilon\right]$$

are $O(T^{-1})$ and $O(T^{-5})$, so that (10.6) gives

$$\mathrm{pr}(\tilde{\gamma}_{t-1} > u_t | \tilde{\gamma}_t = u_t + \epsilon) \sim 1 - \Phi\left\{-O(T^{3/2})\right\}$$

for all $\epsilon > 0$ and $t < \frac{T-1}{2}$, by the normality of $\tilde{\beta}_t$ and $\tilde{\beta}_{t-1}$.

Since the distribution of $\tilde{\gamma}_t$ is continuous and bounded at u_t it is easy to deduce that

$$\mathrm{pr}(\tilde{\gamma}_{t-1} > u_t | \tilde{\gamma}_t > u_t) \sim 1 - \Phi\left\{-O(T^{3/2})\right\} \qquad (10.7)$$

It is easy to see from the definition of $\hat{\gamma}$ that

$$\mathrm{pr}(\hat{\gamma} > u_t) \geq \mathrm{pr}(\tilde{\gamma}_t > u_t, \tilde{\gamma}_{t-1} > u_t, \tilde{\gamma}_{t-2} > u_{t-1}, \ldots, \tilde{\gamma}_2 > u_3)$$
$$= \mathrm{pr}(\tilde{\gamma}_t > u_t)\ \mathrm{pr}(\tilde{\gamma}_{t-1} > u_t | \tilde{\gamma}_t > u_t) \ldots$$
$$\ldots \mathrm{pr}(\tilde{\gamma}_2 > u_3 | \tilde{\gamma}_3 > u_4, \ldots, \tilde{\gamma}_t > u_t) \qquad (10.8)$$

The possible inequality in (10.8) is due to the fact that $\hat{\gamma}$ may be larger than u_t even though $u_s \leq \tilde{\gamma}_s < u_{s+1}$ for some $s < t$; the probability of such an event will decrease as $T \to \infty$, and we shall ignore it for the purposes of getting our approximation. Now looking at the conditional probabilities in (10.8), it is intuitively clear that the more complicated of these satisfy inequalities like

$$\text{pr}(\tilde{\gamma}_s > u_{s+1} \mid \tilde{\gamma}_{s+1} > u_{s+2}, \tilde{\gamma}_{s+2} > u_{s+2}) \geq \text{pr}(\tilde{\gamma}_s > u_{s+1} \mid \tilde{\gamma}_{s+1} > u_{s+2}),$$

since the imposition of extra conditioning events (such as $\tilde{\gamma}_{s+2} > u_{s+2}$ here) restricts the sample space in favour of the event $\tilde{\gamma}_s > u_{s+1}$. This argument would be difficult to verify formally. But accepting its validity we deduce from (10.7) and (10.8) that

$$\begin{aligned} \text{pr}(\hat{\gamma} > u_t) &\sim \text{pr}(\tilde{\gamma}_t > u_t)[1 - \Phi\{-0(T^{3/2})\}]^t \\ &\sim \text{pr}(\tilde{\gamma}_t > u_t), \quad (t < \tfrac{T-1}{2}) . \end{aligned} \tag{10.9}$$

Note that we cannot prove the reverse statement "$\hat{\gamma} > u_t$ implies $\tilde{\gamma}_t > u_t$ with high probability" without proving difficult results about asymptotic smoothness of $L(\gamma)$. This smoothness may, however, be inferred by examining the log likelihood

$$\log L = -\tfrac{1}{2} \sum_{i=1}^{\tau} \{x_i - \alpha - \beta_0(u_i - \gamma)\}^2 - \tfrac{1}{2} \sum_{i=\tau+1}^{T} \{x_i - \alpha - \beta_1(u_i - \gamma)\}^2,$$

where $\sigma^2 = 1$ for convenience.

For $u_\tau < \gamma < u_{\tau+1}$ we have

$$\frac{\partial(\log L)}{\partial \gamma} = -\beta_0 \sum_{i=1}^{\tau} \{x_i - \alpha - \beta_0(u_i - \gamma)\} - \beta_1 \sum_{i=\tau+1}^{T} \{x_i - \alpha - \beta_1(u_i - \gamma)\} . \tag{10.10}$$

But log L is not differentiable with respect to γ at $\gamma = u_\tau$: the difference between the right- and left-hand derivatives at $\gamma = u_\tau$ is $(\beta_1 - \beta_0)(x_\tau - \alpha)$, so that the kink in slope of the log likelihood is of order $\frac{1}{T}$ relative to the slope. We infer that the log likelihood, and hence $L(\gamma)$, are asymptotically smooth in the obvious sense.

Corresponding to (10.9) is a similar expression for $t > \frac{T+1}{2}$, derived in the same way from (10.4), namely

$$\mathrm{pr}(\hat{\gamma} < u_t) \sim \mathrm{pr}(\check{\gamma}_{t-1} < u_t), \quad (t > \frac{T+1}{2}). \tag{10.11}$$

Both (10.9) and (10.11) will be taken as our approximation to the distribution of $\hat{\gamma}$. It remains to account for the missing values of t: so far we have covered $t < \frac{T-1}{2}$ and $t > \frac{T+1}{2}$. Consider the symmetric case, T even and $\gamma = \frac{T+1}{2}$. Then it is clear that

$$\mathrm{pr}(\check{\gamma}_{\frac{1}{2}T-r} < \tfrac{1}{2}T-r) = \mathrm{pr}(\check{\gamma}_{\frac{1}{2}T+r} > \tfrac{1}{2}T+r)$$

by symmetry. Therefore, corresponding to (10.11) with $t = \frac{T}{2} + 1$ we should use (10.9) with $t = \frac{T}{2} - 1$, since $\hat{\gamma}$ will have a symmetric distribution. Lastly, when T is odd, either of (10.9) and (10.11) may be used for $t = \frac{1}{2}(T + 1)$: asymptotically they are equivalent.

We denote our approximation by $G_1(\cdot)$, which by the above discussion is given by

$$G_1(u_t) = \begin{cases} pr(\tilde{\gamma}_t < u_t) & t \le \frac{T+1}{2} \\ pr(\tilde{\gamma}_{t-1} < u_t) & t > \frac{T+1}{2}. \end{cases} \qquad (10.12)$$

To get a continuous version $G_1(w)$ of (10.12) valid for all w an interpolation is necessary. For example

$$G_1(w) = \frac{(w - u_t)}{(u_{t+1} - u_t)} \; pr(\tilde{\gamma}_{t+1} < w) + \frac{(u_{t+1} - w)}{(u_{t+1} - u_t)} \; pr(\tilde{\gamma}_t < w)$$

$$(u_t \le w \le u_{t+1}),$$

which reduces to (10.12) for $w = u_t$ and $w = u_{t+1}$.

An alternative to $G_1(\cdot)$ can be derived from the fact that the asymptotic smoothness of $L(\gamma)$ implies asymptotic equivalence of $\ldots, \tilde{\gamma}_{\tau-1}, \tilde{\gamma}_\tau, \tilde{\gamma}_{\tau+1}, \ldots$. Without giving a more formal argument we deduce that

$$pr(\hat{\gamma} < w) \sim G_2(w) = pr(\tilde{\gamma}_\tau < w) . \qquad (10.13)$$

This, incidentally, leads to the asymptotic normality of $\hat{\gamma}$. We shall discuss this briefly later in this section.

It is reasonable to assume, under mild conditions on the u_t's, that (10.12) holds for general u_t's but with modification to the dichotomy $t \le \frac{T+1}{2}$ and $t > \frac{T+1}{2}$. To avoid this difficulty we assume in what follows that the u_t's are equally spaced.

Now (10.12) represents only the first stage of our approximation. In principle the probabilities $pr(\tilde{\gamma}_t < w)$ can be calculated explicitly using the bivariate normal distribution (see Section 15), but in practice such calculation is often difficult and usually laborious. We use a standard method of approximating the distribution of a ratio of normal variates to approximate the probabilities $pr(\tilde{\gamma}_t < w)$.

By definition $\tilde{\gamma}_t = (\tilde{\alpha}_{ot} - \tilde{\alpha}_{1t})/(\tilde{\beta}_{1t} - \tilde{\beta}_{ot})$ $(t = 2, \ldots, T - 2)$; we shall assume, without loss of generality, that $\beta_1 > \beta_o$ so that the denominator is positive with probability tending to one as t , $t^* \to \infty$. Both numerator and denominator are normally distributed with non-zero means, so that $\tilde{\gamma}_t$ has a "generalized" Cauchy distribution. We use the approximation

$$pr(\tilde{\gamma}_t < w) \doteq pr\left\{\tilde{\alpha}_{ot} - \tilde{\alpha}_{1t} < w(\tilde{\beta}_{1t} - \tilde{\beta}_{ot})\right\} \quad . \qquad (10.14)$$

A detailed discussion of this approximation is held over to Section 15 to avoid loss of continuity in the present section.

First we have to calculate the means, variances and correlation coefficient appearing in the distribution of $\tilde{\gamma}_t$. Both numerator and denominator of $\tilde{\gamma}_t$ are unbiased for $t = \tau$ but not otherwise. Suppose $t < \tau$, then clearly

$$E(\tilde{\alpha}_{ot}) = \alpha_o \quad \text{and} \quad E(\tilde{\beta}_{ot}) = \beta_o .$$

A little calculation shows that

$$E(\tilde{\alpha}_{1t}) = \alpha_1 + (\beta_o - \beta_1)\,\theta_t$$

and

$$E(\tilde{\beta}_{1t}) = \beta_1 + (\beta_o - \beta_1)\,\phi_t, \quad (t < \tau), \tag{10.15}$$

where

$$\theta_t = \frac{1}{t^*}\left\{ \sum_{i=t+1}^{\tau} (u_i - \gamma) - t^*\overline{u}_t^*\,\phi_t \right\}$$

and
$$\phi_t = \sum_{i=t+1}^{\tau} (u_i - \overline{u}_t^*)(u_i - \gamma)/C^*_{uu,t} \,. \tag{10.16}$$

When $t > \tau$, $E(\tilde{\alpha}_{1t}) = \alpha_1$ and $E(\tilde{\beta}_{1t}) = \beta_1$.

A further calculation shows that

$$E(\tilde{\alpha}_{ot}) = \alpha_o - (\beta_o - \beta_1)\,\theta_t$$

and
$$E(\tilde{\beta}_{ot}) = \beta_o - (\beta_o - \beta_1)\,\phi_t, \quad (t > \tau), \tag{10.17}$$

where
$$\theta_t = \frac{1}{t}\left\{ \sum_{i=\tau+1}^{t} (u_i - \gamma) - t\overline{u}_t\phi_t \right\}$$

and
$$\phi_t = \sum_{i=\tau+1}^{t} (u_i - \overline{u}_t)(u_i - \gamma)/C_{uu,t} \,. \tag{10.18}$$

If we define $\theta_\tau = \phi_\tau = 0$, we can write

$$\begin{aligned} E(\tilde{\alpha}_{ot} - \tilde{\alpha}_{1t}) &= \alpha_o - \alpha_1 + (\beta_1 - \beta_o)\,\theta_t \\ &= (\beta_1 - \beta_o)(\gamma + \theta_t) \end{aligned} \tag{10.19}$$

and
$$E(\tilde{\beta}_{1t} - \tilde{\beta}_{ot}) = (\beta_1 - \beta_o)(1 - \phi_t), \quad (t = 2, 3, \ldots, T - 2), \tag{10.20}$$

with θ_t, ϕ_t given by (10.16) and (10.18) for $t < \tau$ and $t > \tau$ respectively.

For the variances and correlation coefficient we use the standard expressions for covariance matrices in least squares regression, namely

$$\text{cov}(\tilde{\alpha}_{ot}, \tilde{\beta}_{ot}, \tilde{\alpha}_{1t}, \tilde{\beta}_{1t}) = \begin{pmatrix} V_t & 0 \\ 0 & V_t^* \end{pmatrix},$$

where

$$tC_{uu,t}\, V_t = \sigma^2 \begin{pmatrix} \sum_{i=1}^{t} u_i^2 & -\sum_{i=1}^{t} u_i \\ -\sum_{i=1}^{t} u_i & t \end{pmatrix}$$

and

$$t^*C^*_{uu,t}\, V_t = \sigma^2 \begin{pmatrix} \sum_{i=t+1}^{T} u_i^2 & -\sum_{i=t+1}^{T} u_i \\ -\sum_{i=t+1}^{T} u_i & t^* \end{pmatrix}.$$

Then $\sigma_t^2(\alpha) = \text{var}(\tilde{\alpha}_{ot} - \tilde{\alpha}_{1t}) = \sigma^2\left(\dfrac{C_{uu,t} + t\bar{u}_t^2}{tC_{uu,t}} + \dfrac{C^*_{uu,t} + t^*\bar{u}_t^{*2}}{t^*C^*_{uu,t}} \right)$,

$$\sigma_t^2(\beta) = \text{var}(\tilde{\beta}_{1t} - \tilde{\beta}_{ot}) = \sigma^2\left(\frac{1}{C_{uu,t}} + \frac{1}{C^*_{uu,t}} \right)$$

and $\sigma_t(\alpha,\beta) = \text{cov}(\tilde{\alpha}_{ot} - \tilde{\alpha}_{1t}, \tilde{\beta}_{1t} - \tilde{\beta}_{ot}) = \sigma^2\left(\dfrac{\bar{u}_t}{C_{uu,t}} + \dfrac{\bar{u}_t^*}{C^*_{uu,t}} \right)$

$(t = 2, 3, \ldots, T - 2)\ldots$ (10.21)

Employing the definitions of C_t, D_t and E_t in (9.3), (10.21) becomes

$$\sigma_t^2(\alpha) = \frac{\sigma^2 T}{tt^* C_{uu,t} C^*_{uu,t}} C_t$$

$$\sigma_t^2(\beta) = \frac{\sigma^2 T}{tt^* C_{uu,t} C^*_{uu,t}} E_t$$

and $$\sigma_t(\alpha,\beta) = \frac{\sigma^2 T}{tt^* C_{uu,t} C^*_{uu,t}} D_t ,$$

whence $$\rho_t = \frac{D_t}{\sqrt{C_t E_t}}$$

is the coefficient of correlation between $\tilde{\alpha}_{ot} - \tilde{\alpha}_{1t}$ and $\tilde{\beta}_{1t} - \tilde{\beta}_{ot}$. Now define also

$$\Sigma_t^2(w) = \frac{(E_t w^2 - 2D_t w + C_t) T}{tt^* C_{uu,t} C^*_{uu,t}} , \tag{10.22}$$

the variance of $(\tilde{\alpha}_{ot} - \tilde{\alpha}_{1t}) - w(\tilde{\beta}_{1t} - \tilde{\beta}_{ot})$.

Then (10.14) becomes

$$\mathrm{pr}(\tilde{\gamma}_t < w) \doteq \Phi\left\{-\beta(\gamma - w + \theta_t + w\phi_t)/\Sigma_t(w)\right\} , \tag{10.23}$$

where $\beta = (\beta_1 - \beta_o)/\sigma$.

The quantity $\Sigma_t^2(w)$ has the alternative representation

$$\Sigma_t^2(w) = \frac{1}{t} + \frac{1}{t^*} + \frac{(\bar{u}_t - w)^2}{C_{uu,t}} + \frac{(\bar{u}_t^* - w)^2}{C^*_{uu,t}}$$

which is easier for calculation; cf. V (·) in (8.2)

Together with (10.23), (10.12) and (10.13) give approximations to the distribution of $\hat{\gamma}$ which are easy to calculate. The accuracy of both approximations is studied in detail in Section 11.

We mentioned earlier that $G_2(w)$ in fact tends to the asymptotic normal distribution derived by Feder and Sylwester (1968). This follows from (10.23) and the definition $G_2(w) = \text{pr}(\breve{\gamma}_\tau < w)$ by letting $u_\tau = 0$ for convenience and noting that

$$\Sigma_\tau^2(w) \sim 4\left(\frac{1}{\tau} + \frac{1}{\tau^*}\right) + o\left(\frac{1}{\tau}\right) + o\left(\frac{1}{\tau^*}\right) \quad , \quad w = O(T^{\frac{1}{2}}).$$

Then since $\phi_\tau = \theta_\tau = 0$, (10.23) gives

$$\text{pr}(\hat{\gamma} < w) \sim \Phi\left\{\beta\sqrt{\left(\frac{\tau\tau^*}{4T}\right)}\ (w - \gamma)\right\} \tag{10.24}$$

We can proceed to this result directly from the joint likelihood of α, β_0, β_1 and γ by evaluating the required derivatives and assuming $\gamma \neq u_\tau$. Then the asymptotic covariance matrix J is given by

$$J = \sigma^2 \begin{pmatrix} T & \tau(\bar{u}_\tau - \gamma) & \tau^*(\bar{u}^*_\tau - \gamma) & -\tau\beta_o - \tau^*\beta_1 \\ \cdot & \sum_{j=1}^{\tau}(u_i - \gamma)^2 & 0 & -\beta_o\tau(\bar{u}_\tau - \gamma) \\ \cdot & \cdot & \sum_{j=\tau+1}^{T}(u_i - \gamma)^2 & -\beta_1\tau^*(\bar{u}^*_\tau - \gamma) \\ \cdot & \cdot & \cdot & \tau\beta_o^2 + \tau^*\beta_1^2 \end{pmatrix}^{-1}$$

It follows that

$$\text{var}(\hat{\beta}_o) \sim \frac{\sigma^2}{C_{uu,\tau}}, \qquad \text{var}(\hat{\beta}_1) \sim \frac{\sigma^2}{C^*_{uu,\tau}}$$

$$\text{cov}(\hat{\beta},\hat{\gamma}) \sim \frac{1}{\beta}\left\{\frac{(\bar{u}^*_\tau - \gamma)}{C^*_{uu,\tau}} + \frac{(\bar{u}_\tau - \gamma)}{C_{uu,\tau}}\right\}$$

and $$\text{var}(\hat{\gamma}) \sim \frac{1}{\beta^2}\left\{\frac{\sum_{i=1}^{\tau}(u_i - \gamma)^2}{\tau\, C_{uu,\tau}} + \frac{\sum_{i=\tau+1}^{T}(u_i - \gamma)^2}{\tau^* C^*_{uu,\tau}}\right\}$$

$$\sim \frac{4}{\beta^2}\left(\frac{1}{\tau} + \frac{1}{\tau^*}\right) + o\left(\frac{1}{\tau}\right) + o\left(\frac{1}{\tau^*}\right)$$

for equally spaced u_t's. We emphasise that these calculations assume the stated asymptotic normality: no proof is offered here.

11. An Empirical Study of the Model

To study the distribution of $\hat{\gamma}$ empirically, and hence make comparisons with the approximations of Section 10, we generated realizations of the model (8.1) for varying β,γ and T. For each combination of β,γ and T we took 500 realizations. The error terms ε_t of the model were simulated in a standard way, using pseudo random numbers and a power series approximation to the inverse of the normal probability integral.

In the summary of results presented here we take three values of T, namely 25, 50 and 100, with two values of γ in each case and β = 0.2 (0.2) 1.0. As well as analyzing the empirical distributions of $\hat{\gamma}$, we also calculated $E(\hat{\beta})$, var $(\hat{\beta})$ and corr $(\hat{\beta},\hat{\gamma})$ empirically. Examination of these is useful in deciding which combinations of β, γ and T give "well-defined" situations, by which we mean situations where inference about β and γ is straightforward. In this analysis the explicit definition of "well-defined" is that $\hat{\beta}$ has a bias of less than 5%, the variance of $\hat{\beta}$ is stable (independent of β), and the distribution of $\hat{\gamma}$ is well contained in the interval (1, T). We shall see later that the well-defined cases possess a useful property concerning the log likelihood ratio test of hypotheses about γ (Section 13).

The empirical moments of $\hat{\beta}$ and $\hat{\gamma}$ are contained in Table 11.1, in which the well-defined cases are marked with an asterisk. For example in the case T = 25, β = 0.2 and γ = 8.5 we see that the mean of $\hat{\beta}$ is approximately 2β, and the size of the variance implies that β cannot be determined with any useful precision. Also the spread of the distribution of $\hat{\gamma}$ is large enough to make inference about γ difficult even for known β.

T	γ	β	mean of $\hat{\gamma}$	variance of $\hat{\gamma}$	correlation corr($\hat{\beta},\hat{\gamma}$)	mean of $\hat{\beta}$	variance of $\hat{\beta}$
25	8.5	0.2	11.23	49.13	-0.13	0.405	1.09
		0.4	9.30	20.91	- .22	.651	0.459
		0.6	8.79	5.98	- .46	.678	.107
		0.8	8.65	2.54	- .61	.844	.059
		1.0*	8.56	1.28	- .58	1.026	.034
	12.5	0.2	12.80	38.68	- .07	0.427	.790
		0.4	12.43	12.46	- .08	.542	.174
		0.6	12.43	4.38	+ .04	.642	.025
		0.8*	12.54	1.70	- .15	.803	.014
		1.0*	12.54	0.94	- .10	1.013	.014

Table 11.1 Empirical moments of $\hat{\beta}$ and $\hat{\gamma}$ based on 500 realizations of model (8.1) in each case.

T	γ	β	mean of $\hat{\gamma}$	variance of $\hat{\gamma}$	correlation corr($\hat{\beta}$,$\hat{\gamma}$)	mean of $\hat{\beta}$	variance of $\hat{\beta}$
50	15.5	0.2	16.55	37.98	+0.01	0.257	0.056
		0.4*	15.64	3.79	- .69	.408	.0047
		0.6*	15.48	1.41	- .62	.604	.0041
		0.8*	15.54	0.80	- .67	.803	.0045
		1.0*	15.50	0.47	- .59	1.000	.0042
	25.5	0.2	25.25	20.19	- .01	0.226	.0152
		0.4*	25.52	2.98	- .04	.404	.0014
		0.6*	25.56	1.24	+ .01	.600	.0016
		0.8*	25.53	0.52	- .09	.803	.0015
		1.0*	25.51	0.36	+ .01	1.002	.0015
100	25.5	0.2*	25.59	9.10	- .71	0.204	.0011
		0.4*	25.52	1.54	- .67	.402	.0009
		0.6*	25.48	0.67	- .67	.602	.0009
		0.8*	25.52	0.38	- .70	.800	.0008
		1.0*	25.50	0.26	- .68	1.000	.0009
	50.5	0.2*	50.40	4.82	+ .04	0.201	.0002
		0.4*	50.42	1.07	+ .10	.400	.0002
		0.6*	50.50	0.51	+ .08	.601	.0002
		0.8*	50.49	0.25	- .01	.800	.0002
		1.0*	50.51	0.17	- .01	1.000	.0002

Table 11.1 (continued) Empirical moments of $\hat{\beta}$ and $\hat{\gamma}$ based on 500 realizations of model (8.1) in each case.

Examination of the cases asterisked in Table 11.1 suggests a rough method for determining whether or not a particular case is well-defined, namely to calculate $\frac{\beta\tau(T - \tau)}{T}$ which is greater than about 5 for well-defined cases.

To illustrate the comparisons between the empirical distribution of $\hat{\gamma}$ and the approximations $G_1(\cdot)$ and $G_2(\cdot)$ we drew several normal plots as in Figures 11.1 to 11.6. These are all for well-defined cases. From these we conclude that $G_1(\cdot)$ is a good approximation except when τ or $T - \tau$ is less than 10. The approximation $G_2(\cdot)$ is under-dispersed even when $T = 100$. The evident non-linearity of these plots indicates the non-normality of $\hat{\gamma}$ in finite samples, although for $T = 100$ the discrepancy is apparently quite small.

All the calculations of $G_1(\cdot)$ and $G_2(\cdot)$ used the approximation (10.23). The error involved was negligible.

It is worth noting here that in the null case $(\beta_o = \beta_1)$ the distribution of $\hat{\gamma}$ is apparently uniform over its range. We have not investigated this theoretically, although it would not appear to be a difficult problem. Figure 11.7 is a histogram for $\hat{\tau}$ in the null case when $T = 50$.

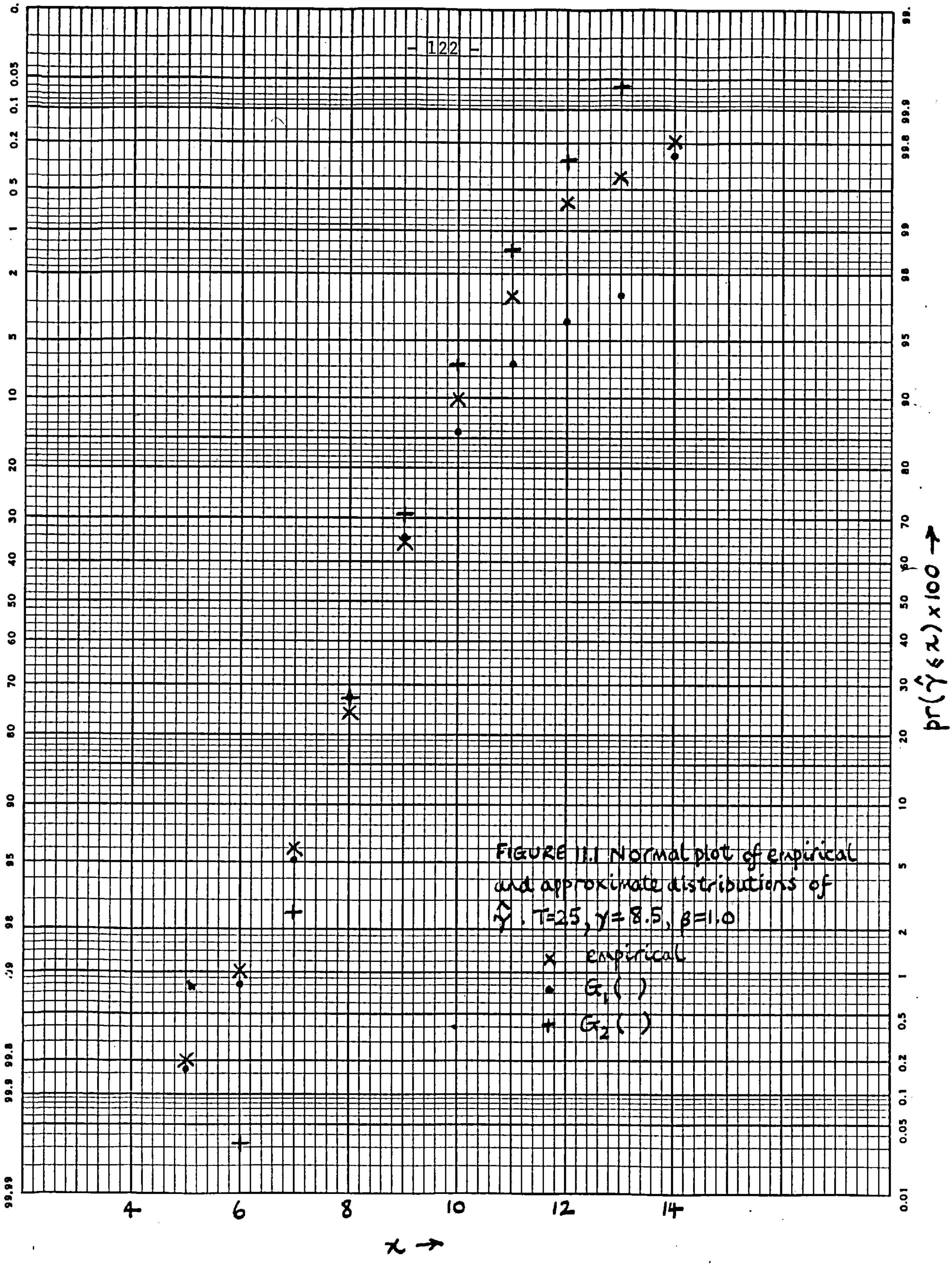


FIGURE 11.1 Normal plot of empirical and approximate distributions of $\hat{\gamma}$. T=25, γ=8.5, β=1.0

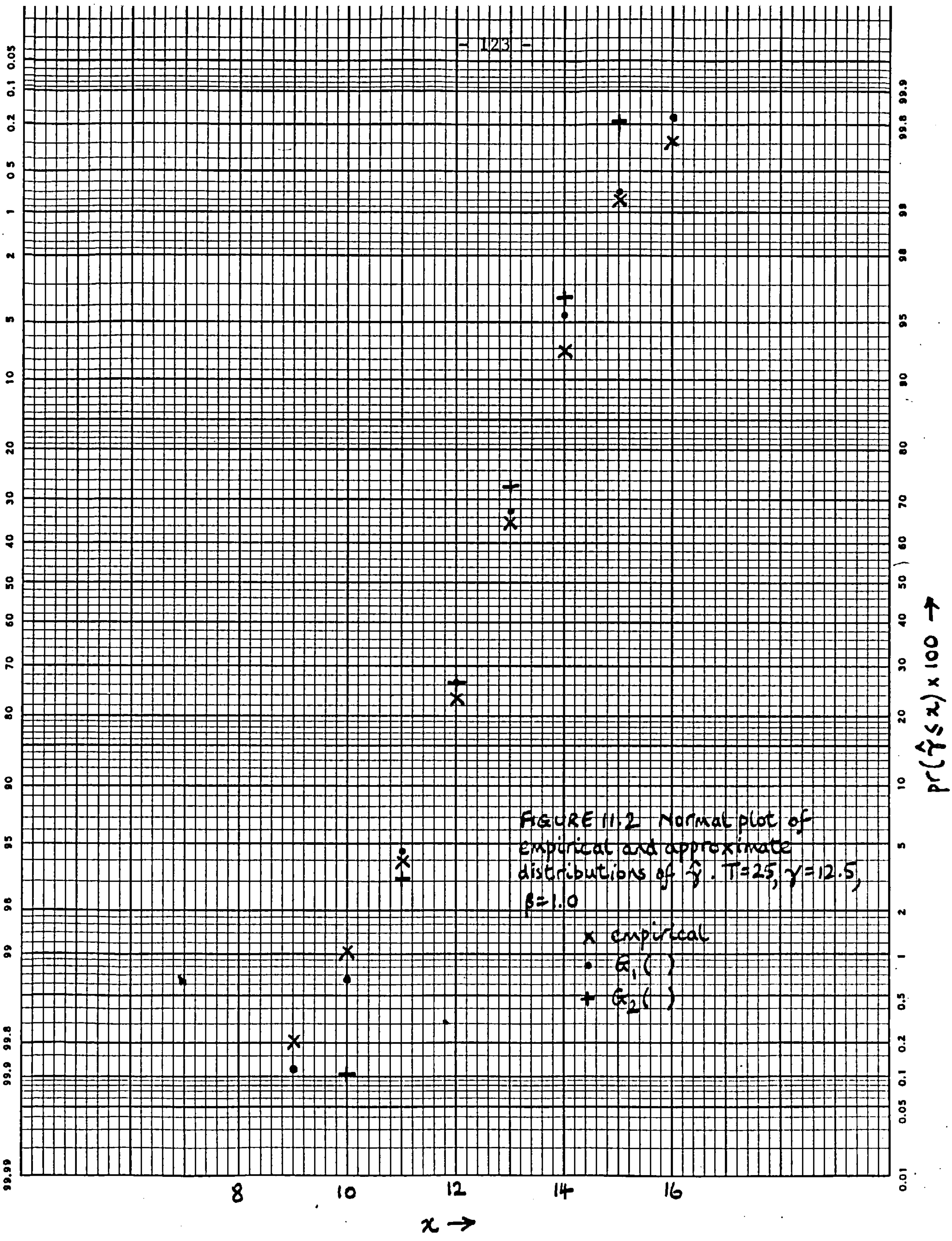


FIGURE 11.2 Normal plot of empirical and approximate distributions of $\hat{\gamma}$. $T=25$, $\gamma=12.5$, $\beta=1.0$

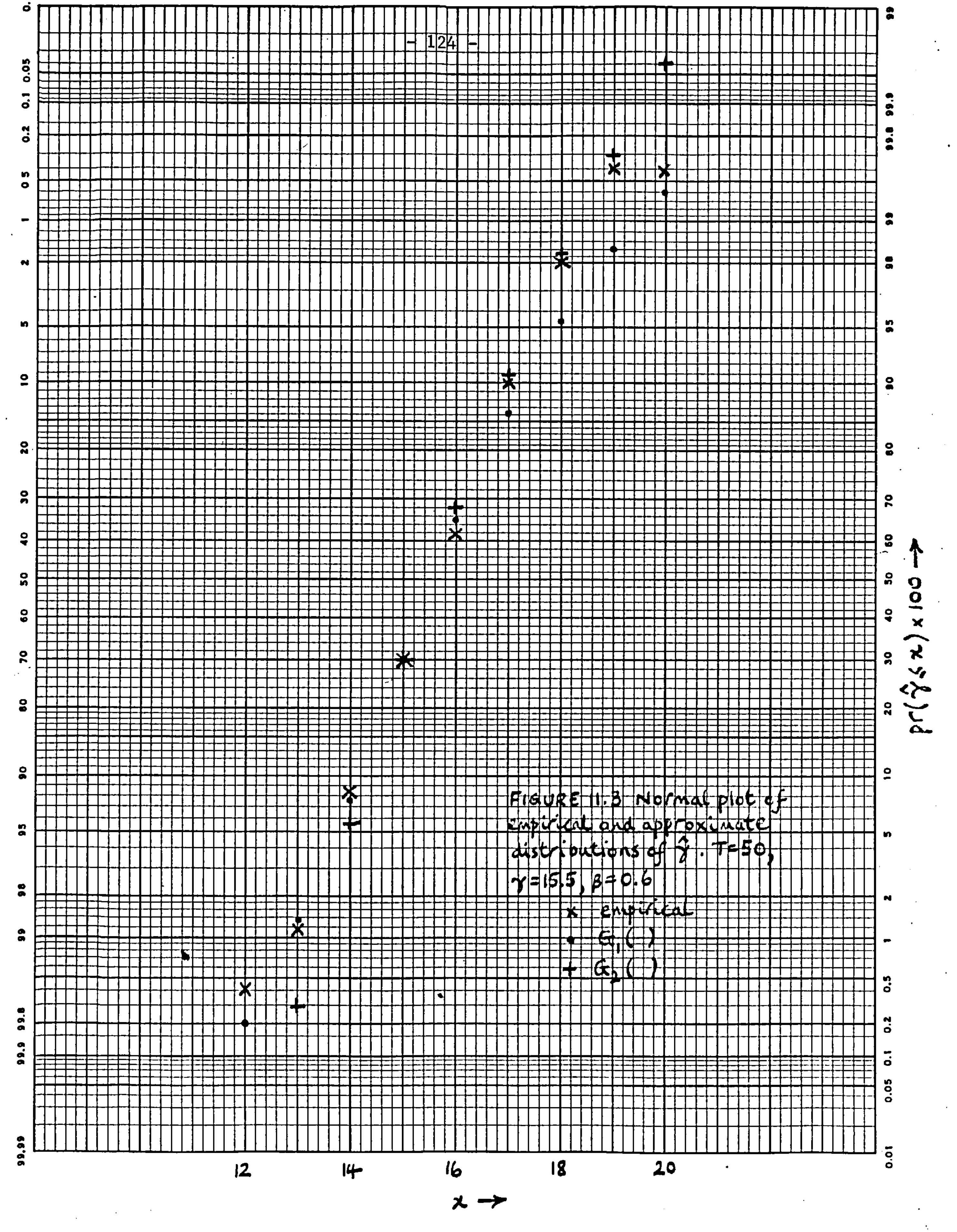


FIGURE 11.3 Normal plot of empirical and approximate distributions of $\hat{\gamma}$. T=50, γ=15.5, β=0.6

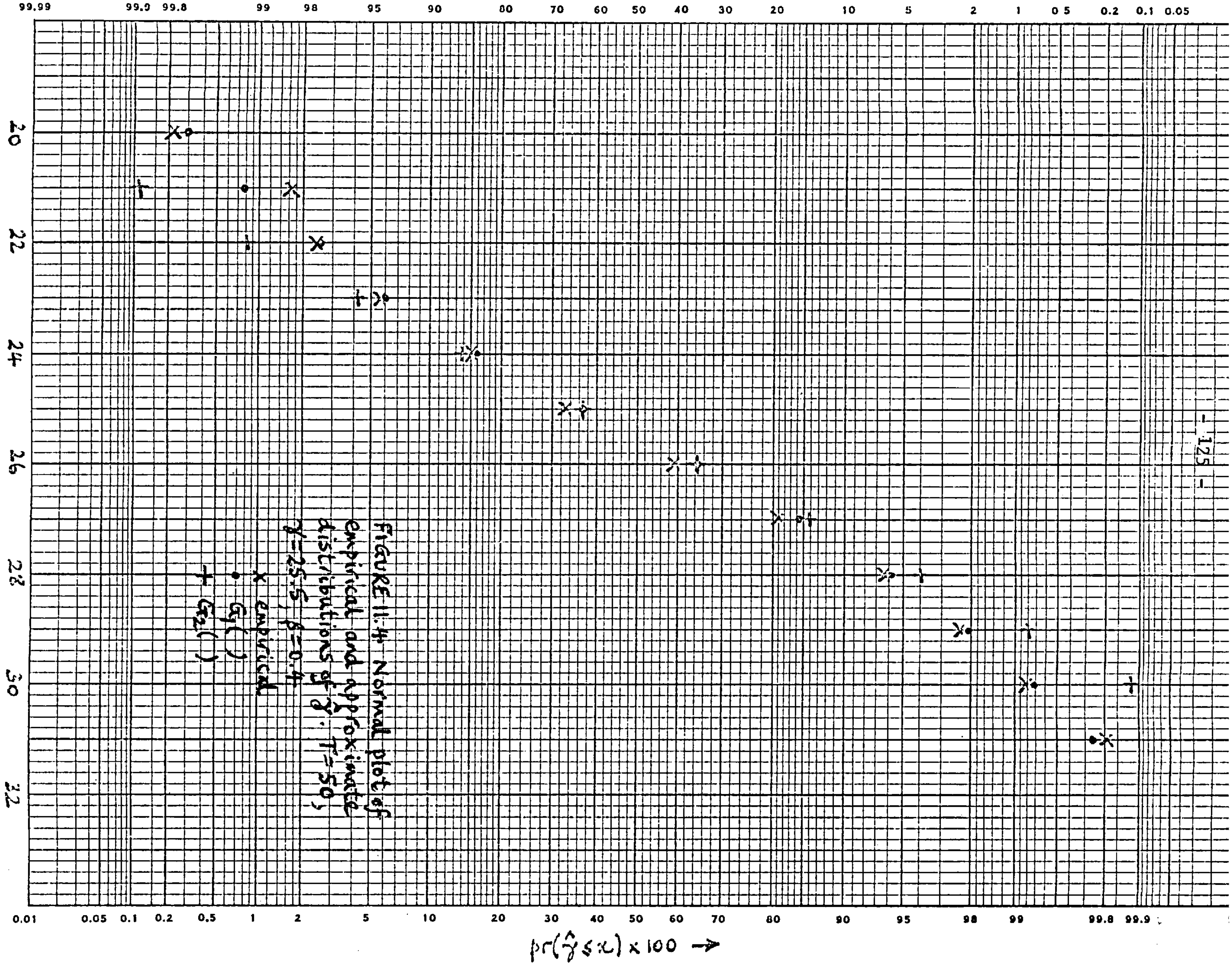


FIGURE 11.4 Normal plot of empirical and approximate distributions of $\hat{\gamma}$. T=50, γ=25.5, β=0.4

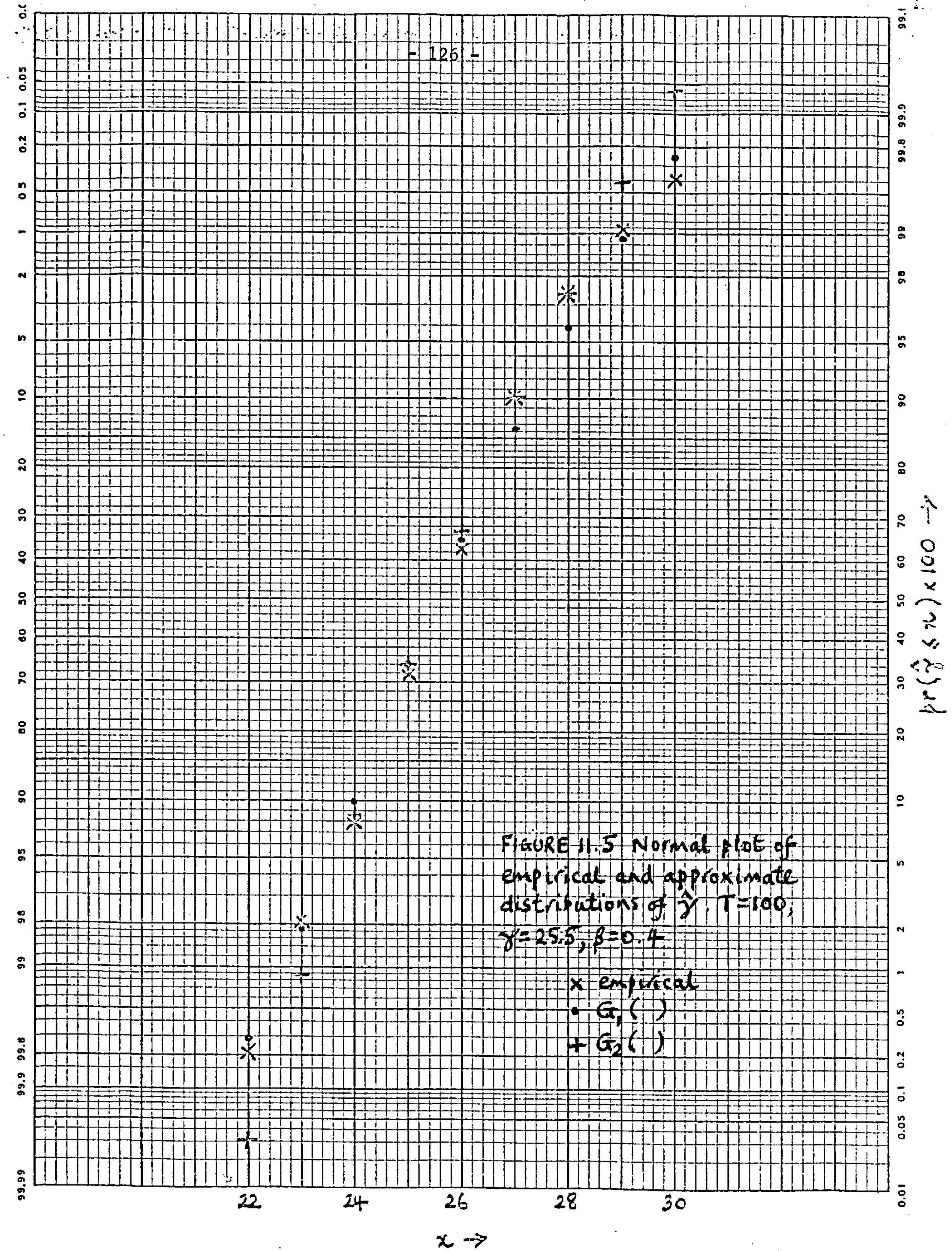


FIGURE 11.5 Normal plot of empirical and approximate distributions of $\hat{\gamma}$. T=100, γ=25.5, β=0.4

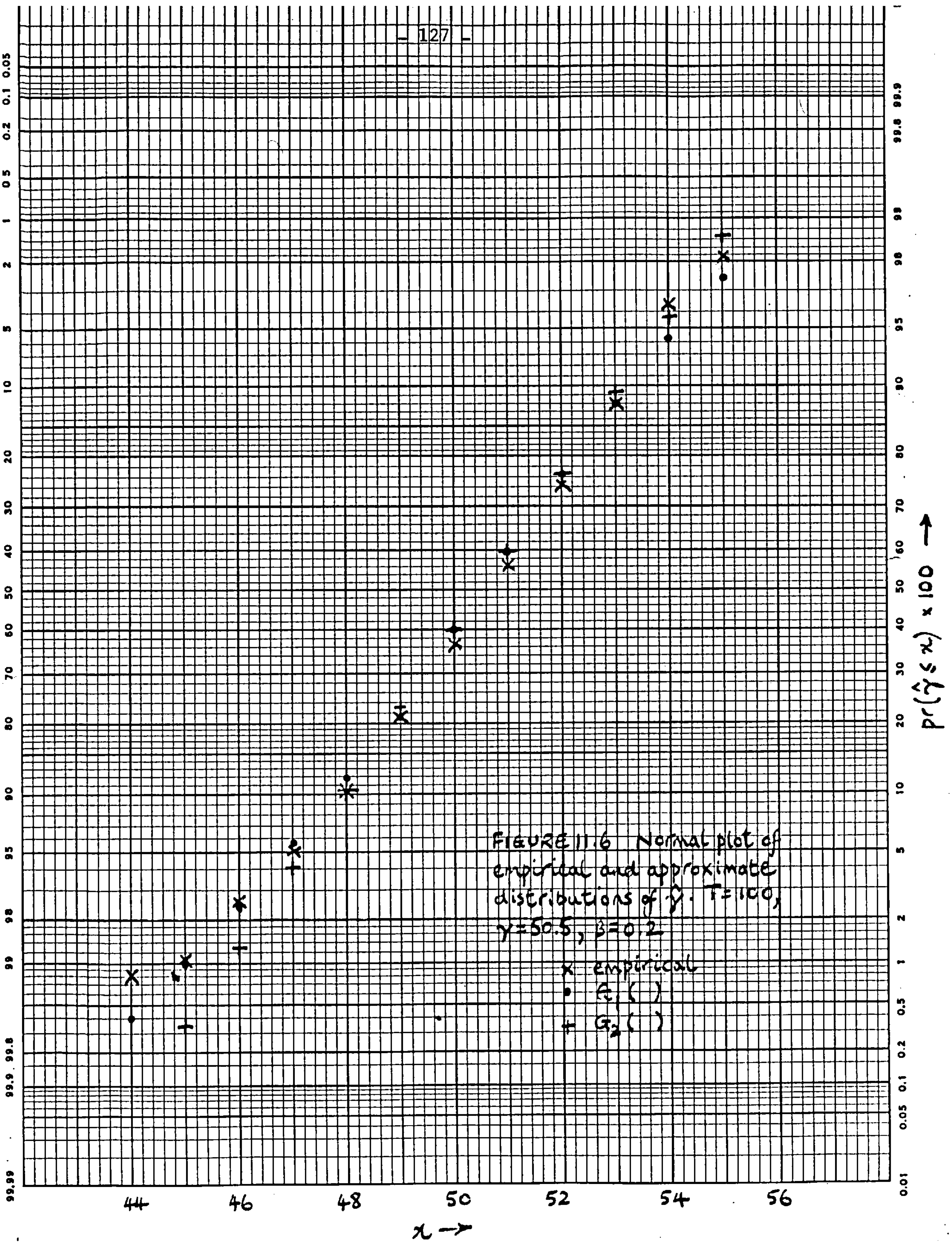


FIGURE 11.6 Normal plot of empirical and approximate distributions of $\hat{\gamma}$. T=100, γ=50.5, β=0.2

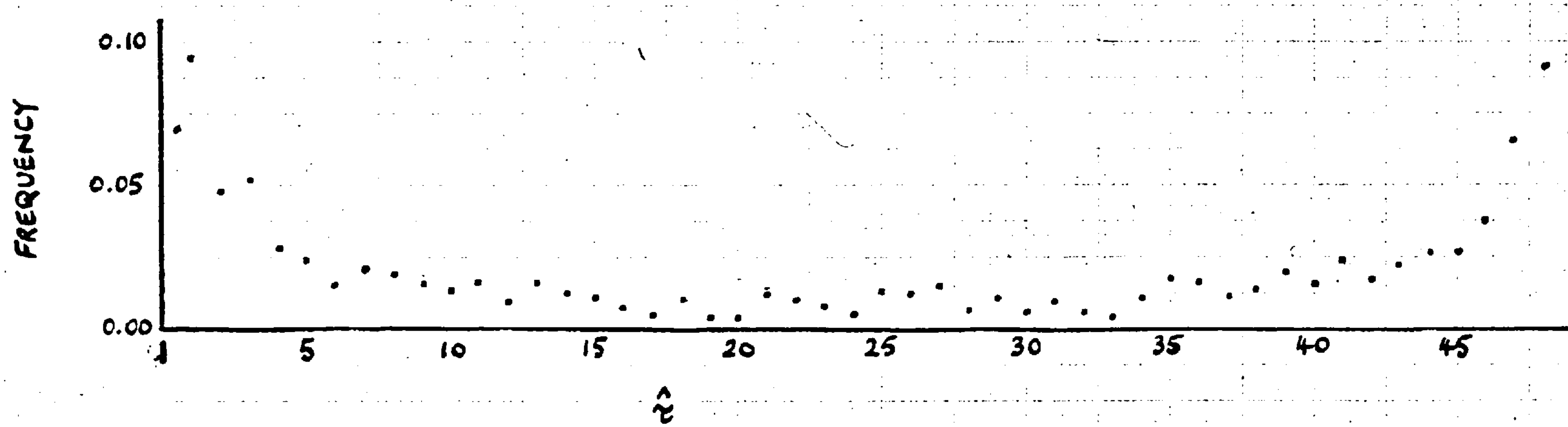


FIGURE 11.7 Empirical frequencies of $\hat{\tau}$ in the null case. T=50

12. Distribution of $\hat{\beta}$.

Although γ is the parameter of primary interest in our discussion, in most practical applications one would wish also to make inference about β. For any t we know that $\tilde{\beta}_t$ is normally distributed, and the asymptotic normality discussed in Section 10 shows that $\hat{\beta} \sim \tilde{\beta}_T$. This is a useful result, but it is not always a good approximation. For the well-defined cases of the previous section, the empirical distribution of $\hat{\beta}$ is very close to the asymptotic normal distribution. But clearly in other cases $\hat{\beta}$ has a non-negligible bias and inflated variance; see, e.g., the case $T = 50$, $\gamma = 25.5$, $\beta = 0.2$ in Table 11.1. The normal plots of the empirical distribution of $\hat{\beta}$ in such cases illustrate clear non-normality. Figures 12.1 and 12.2 are typical examples of well-defined and ill-defined cases, respectively. We remark that in the null case the non-normality is even more marked.

For completeness we note briefly that the sample correlations between $\hat{\beta}$ and $\hat{\gamma}$ in Table 11.1 increase as $|\frac{\tau}{T}-\frac{1}{2}|$ increases. A further investigation shows a clear relationship between the correlation and the ratio $\frac{\tau}{T}$, independent of β and T. However there seems little merit in pursuing the point further.

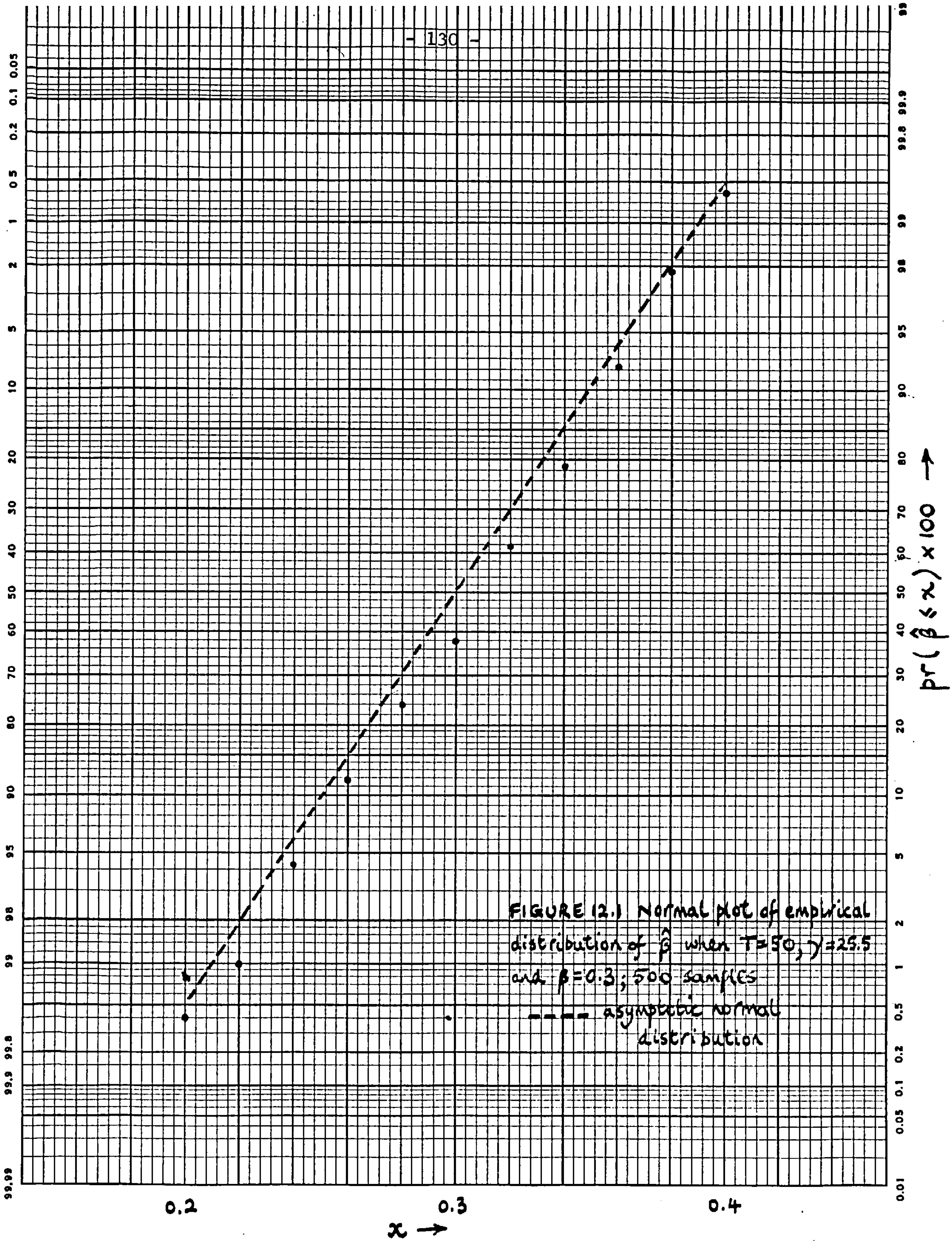


FIGURE 12.1 Normal plot of empirical distribution of $\hat{\beta}$ when $T=50$, $\gamma=25.5$ and $\beta=0.3$; 500 samples

- - - - asymptotic normal distribution

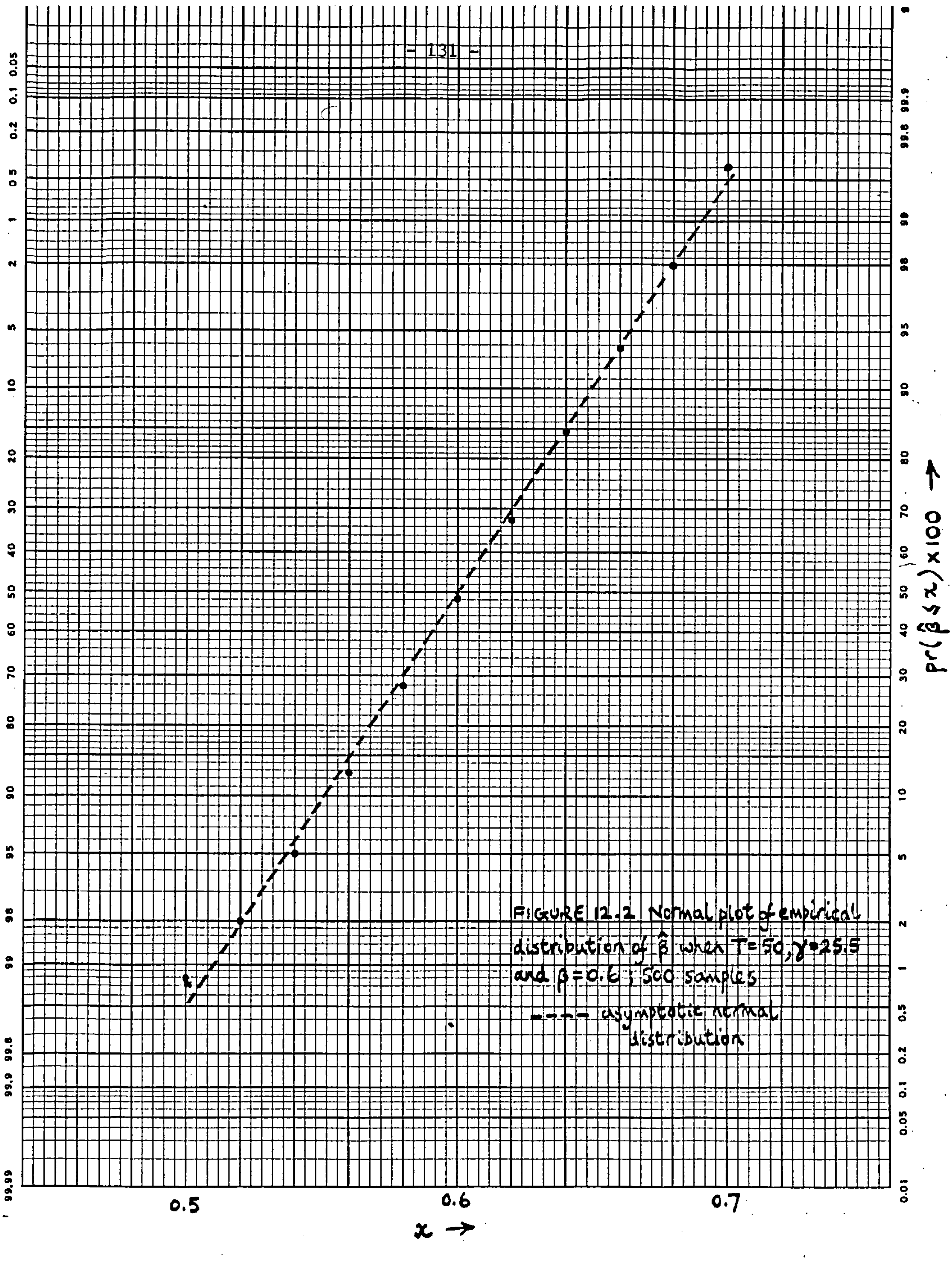


FIGURE 12.2 Normal plot of empirical distribution of $\hat{\beta}$ when T=50, γ=25.5 and β=0.6 ; 500 samples

- - - - asymptotic normal distribution

13. Inference about γ

When $\beta = (\beta_1 - \beta_o)/\sigma$ is known, there is in principle no difficulty about performing significance tests of hypotheses about γ using the distribution function approximation $G_1(\cdot)$. When β is unknown, however, we encounter the same problem as in Section 4 of Part I, namely inference with nuisance parameters. The remarks of that section apply here; again the nuisance parameter is estimated by a statistic which is asymptotically normally distributed with known mean and variance dependent on τ (we assume the case to be well-defined). The significant correlation between $\hat{\beta}$ and $\hat{\gamma}$, mentioned in Section 12, must of course be taken into account.

The continuity of $L(\gamma)$, and the discussion of asymptotic normality in Section 10, indicate a method of avoiding the nuisance parameter difficulty. Under the conditions for asymptotic normality of $(\hat{\alpha}, \hat{\beta}_o, \hat{\beta}_1, \hat{\gamma})$, the log likelihood ratio test statistic

$$\lambda = -2 \log\left\{\frac{L(\gamma_o)}{L(\hat{\gamma})}\right\}$$

is asymptotically distributed as χ_1^2, the distribution being central chi square under the hypothesis H_o^*: $\gamma = \gamma_o$. For a discussion of this result see Kendall and Stuart (1961, chapter 24). The test based on λ is asymptotically consistent and asymptotically more stringent than any other test of H_o^*. By the definitions in

Section 9 we have

$$\lambda = \frac{1}{\sigma^2}\left\{ Z^2_{\hat{\tau}}(\hat{\gamma}) - Z^2_{\tau_o}(\gamma_o)\right\} , \qquad (13.1)$$

where $u_{\tau_o} \le \gamma_o < u_{\tau_o+1}$ and $u_{\hat{\tau}} \le \hat{\gamma} < u_{\hat{\tau}+1}$.

This may be used to carry out a significance test of H_o^* without any knowledge of β. Methods based on $G_1(\cdot)$ and β may have certain advantages, but without further study it is difficult to be precise on this point. The important question to answer here is: when is the χ_1^2 distribution a valid approximation to the distribution of (13.1)? For the empirical study of Section 11 we also examined the empirical distributions of (13.1) and found the use of the chi square approximation to be accurate in the well-defined cases. Figures 13.1, 13.2 and 13.3 illustrate the comparison between empirical and chi square distributions by percentage plots; (that is by plotting cumulative empirical frequency up to x against $pr(\chi_1^2 \le x)$). Figure 13.1, for an ill-defined case, shows how over-dispersed is the distribution of λ relative to that of χ_1^2. In Figure 13.2 the comparison in the upper tail (probabilities greater than 0.9) is good, as it is for all the well-defined cases. As the sample size T increases, so the fit improves over the whole range: the plot in Figure 13.3 is linear for cumulative probabilities greater than 0.20. Since the upper tail of the distribution is of most importance, we

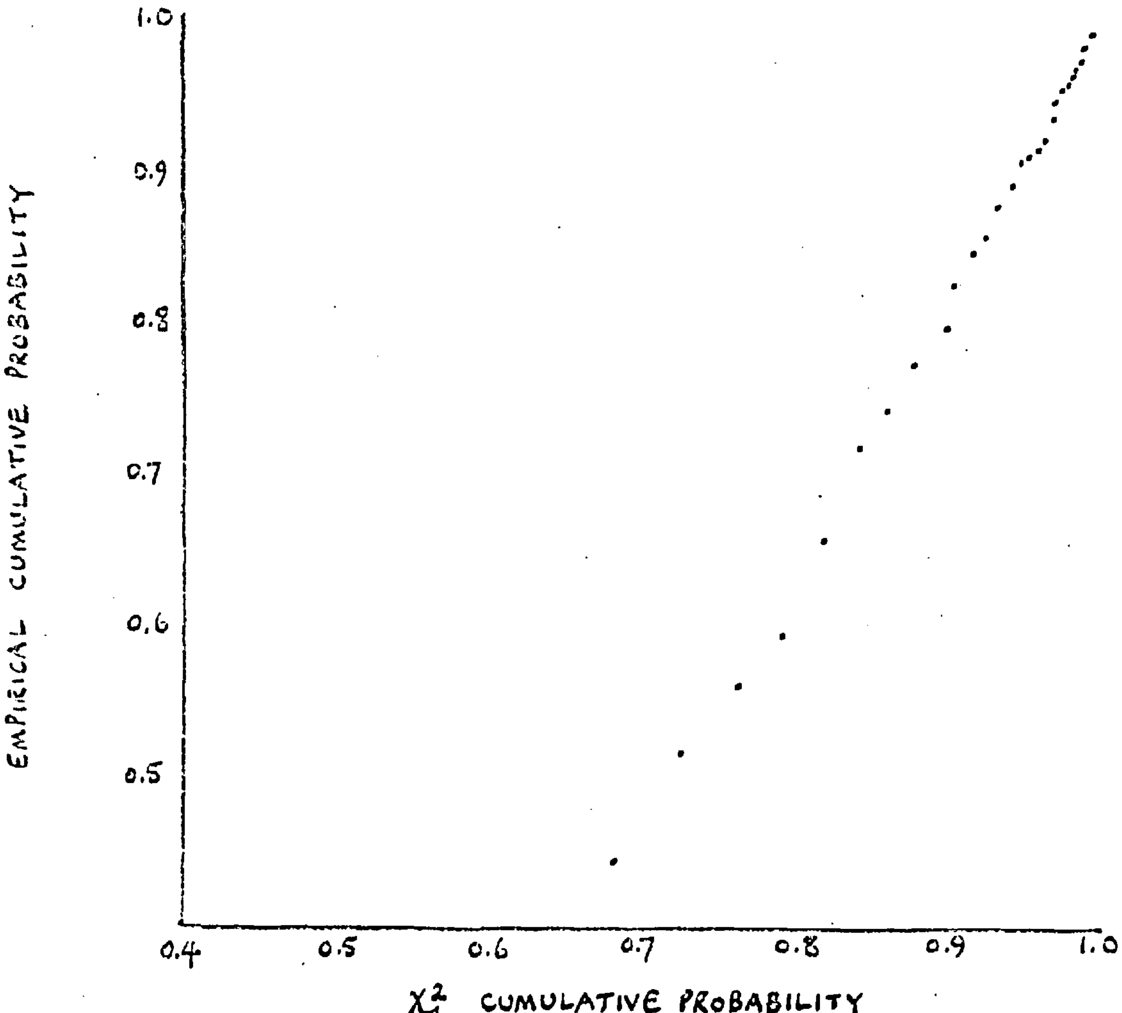


FIGURE 13.1 Percentage plot of empirical distribution of λ against the χ^2_1 distribution. $T=25$, $\gamma=8.5$, $\beta=0.2$; 500 samples

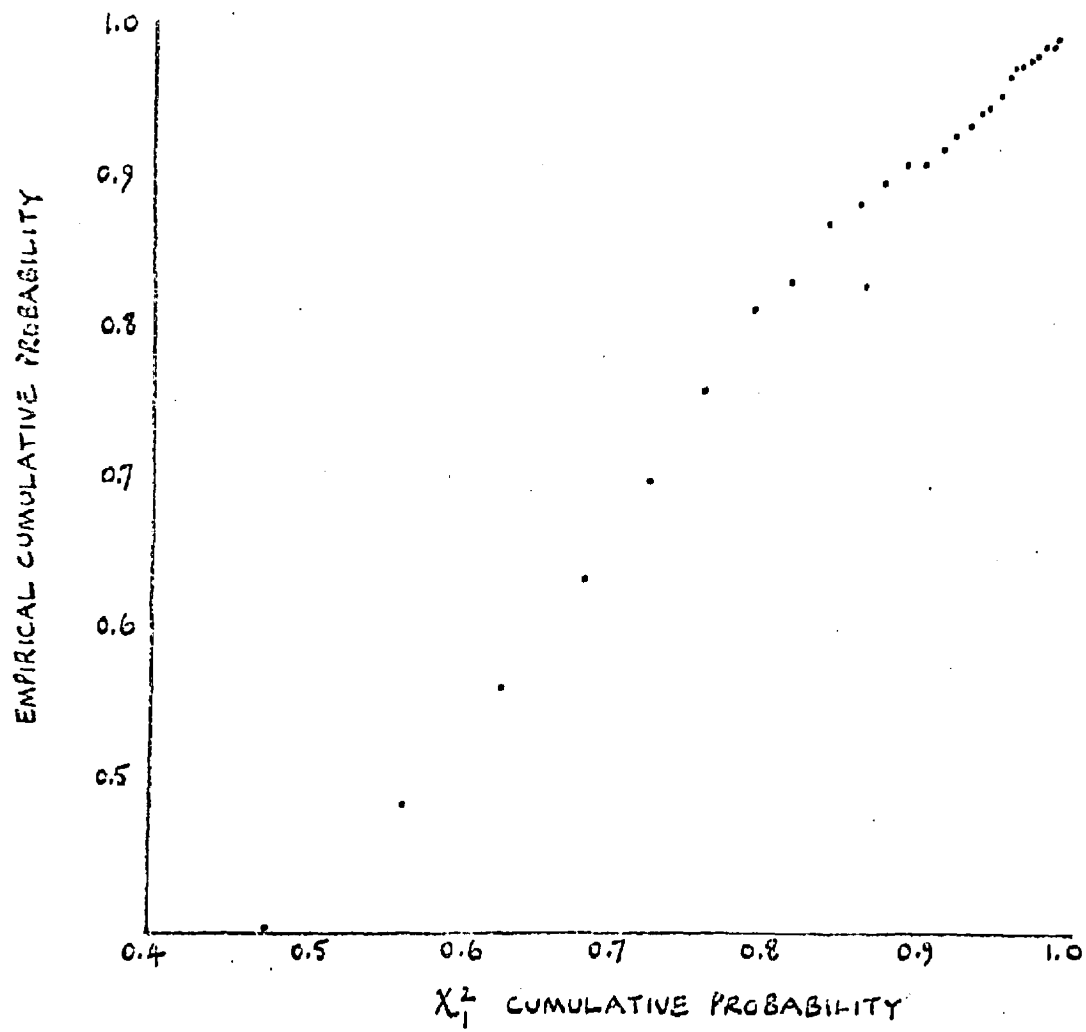


FIGURE 13.2 Percentage plot of empirical distribution of λ against the χ_1^2 distribution. $T=25$, $\gamma=8.5$ and $\beta=1.0$; 500 samples

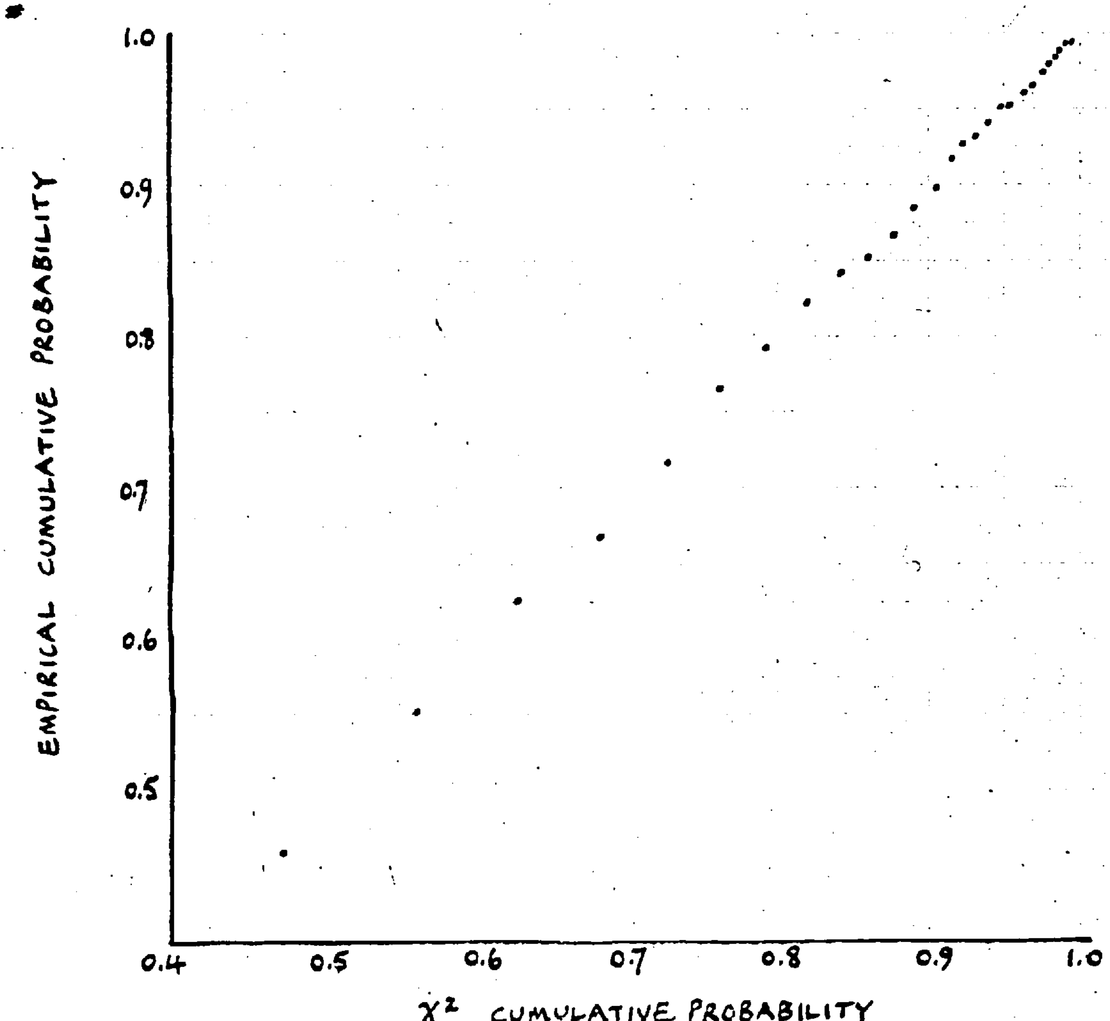


FIGURE 13.3 Percentage plot of empirical distribution of λ against the χ_1^2 distribution. $T=100$, $\gamma=25.5$, $\beta=0.6$; 500 samples.

conclude that the chi square approximation is good for well-defined cases.

(Further empirical study showed the accuracy of this approximation to be independent of β and $\frac{I}{T}$, except in so far as these determine whether or not a case is well-defined.)

In any particular example the l.r. test using (13.1) is simple to use. The values $Z^2_{\tau_o}(\gamma_o)$ and $Z^2_{\hat{\tau}}(\hat{\gamma})$ are calculated during the derivation of $\hat{\gamma}$ set out in Section 9; the value of λ is then referred to tables. In Example 2 of Section 9, where $T = 50$, $\gamma = 25$ and $\beta = 0.4$, the 95% confidence interval using the chi square distribution is (23.5, 32.8), which can be read off from Figure 9.5 by solving

$$2 \log\left\{ L(\hat{\gamma})/L_o \right\} - 3.84 = 2 \log\left\{ L(\gamma)/L_o \right\} .$$

This compares favourably with the distance between upper and lower 2½% points of the actual distribution of $\hat{\gamma}$ when β is known, which is approximately 7.5.

In the null case ($\beta_o = \beta_1$) we found that the empirical distribution of $Z^2_{\hat{\tau}}(\hat{\gamma})$, the log likelihood ratio test statistic for testing H_o: $\beta_o = \beta_1$, is very close to the chi square distribution with 3 degrees of freedom; Figure 13.4 gives a typical percentage plot, for the case $T = 25$. The explanation of this has not been found, although it is clearly an important result in testing H_o. The statistic $Z^2_{\hat{\tau}}(\hat{\gamma})$ is automatically computed in the procedure for calculating $\hat{\gamma}$.

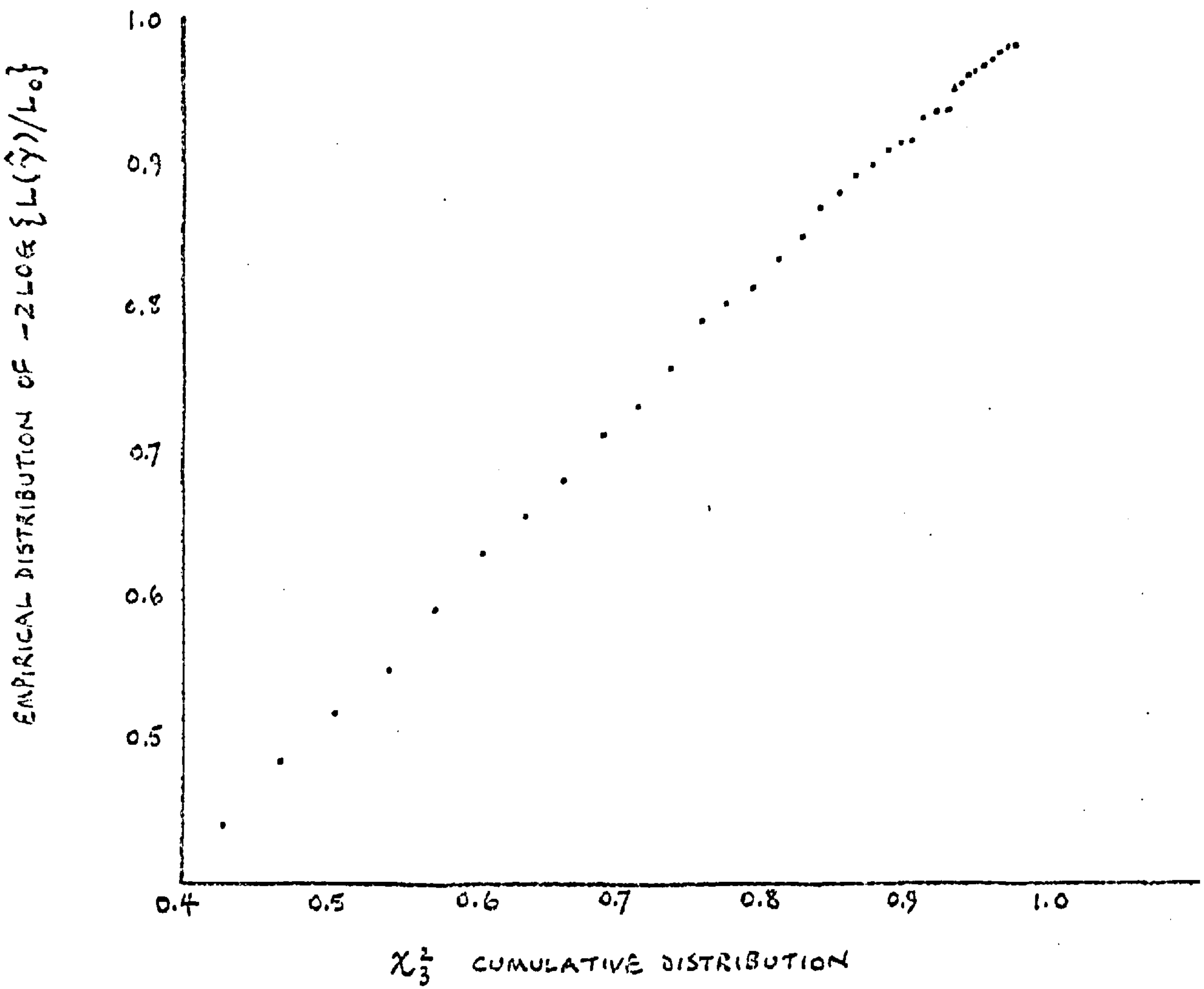


FIGURE 13.4 Percentage plot of empirical distribution of $-2\log\{L(\hat{\gamma})/L_0\}$ against the χ^2_3 distribution. T=25

14. Further Developments

The empirical investigation of Section 11 shows that the approximation $G_1(\cdot)$ is good for well-defined cases. For unknown β the likelihood ratio test statistic is the principal instrument for testing hypotheses of the form $\gamma = \gamma_0$. Further investigation of inference about γ will include some study of significance testing in the presence of a nuisance parameter, which we discussed briefly in Section 4. The problem needs closer examination for ill-defined cases; as with the problem in Part I, theoretical results in ill-defined cases (equivalently small sample cases in general) seem difficult to obtain.

The results we have obtained concerning $\hat{\gamma}$ and the log likelihood $L(\gamma)$ are relevant to the problem of testing the hypothesis

$$H_1 : \begin{cases} x_i = \alpha + \beta_0(u_i - \gamma) + \epsilon_i & (i = 1, \ldots, \tau) \\ x_i = \alpha + \beta_1(u_i - \gamma) + \epsilon_i & (i = \tau + 1, \ldots, T) \end{cases}$$

against the quadratic alternative

$$H_2: \quad x_i = a + bu_i + cu_i^2 + \epsilon_i \quad (i = 1, \ldots, T).$$

These hypotheses are separate in the sense of Cox (1961) who mentioned this problem. Also relevant to this is the work on residual examination by Brown and Durbin (1968).

Another possible extension of our work is to the estimation of planar intersections, that is to take a multiple regression model of the form

$$x_t = \alpha_o + \beta_{1o} u_{1t} + \sum_{j=2}^{p} \beta_j u_{jt} + \epsilon_t \qquad (t = 1, \ldots, \tau)$$

$$x_t = \alpha_1 + \beta_{11} u_{1t} + \sum_{j=2}^{p} \beta_j u_{jt} + \epsilon_t \qquad (t = \tau + 1, \ldots, T),$$

where $\alpha_o + \beta_{1o}\gamma = \alpha_1 + \beta_{11}\gamma$ and $u_{1\tau} \leq \gamma < u_{1,\tau+1}$. Again the ϵ_t are $N(0, \sigma^2)$, independent, and the independent variable u_{1t} increases with t. The difficulty with this generalization is that estimation of β_{1o}, β_{11} and γ is never independent of estimation of $\beta_2, \ldots, \beta_p$. (The independence would, of course, guarantee that the results of Sections 9 and 10 hold). The cause of the difficulty lies in the fact that when τ is unknown, no design of $U = ((u_{jt}))$ is possible to achieve the necessary orthogonality.

It should, however, be possible to attack the problem in the same way as Section 9. A generalization of Sprent's (1961) results would be necessary for this.

Two further problems that might be studied are

(i) multiple intersection, that is k-phase regression systems (see Hudson, 1966);

(ii) shifted intersection , that is removing the restriction $u_\tau \leq \gamma < u_{\tau+1}$.

As to the problem of testing H_o: $\beta_o = \beta_1$, which we mentioned very briefly in Section 13, it would be worth investigating formally the empirically-derived result about the log likelihood ratio $Z^2_{\hat{\tau}}(\hat{\gamma})$ being distributed approximately as a χ^2_3 variate.

15. Generalised Cauchy Distribution

Let X_1 and X_2 have a bivariate normal distribution with means θ_i, variances σ_i^2 $(i = 1, 2)$ and correlation coefficient ρ, and let $W = X_1/X_2$. In Section 10 we were concerned with the case $X_1 = \tilde{\alpha}_{ot} - \tilde{\alpha}_{1t}$, $X_2 = \tilde{\beta}_{1t} - \tilde{\beta}_{ot}$ and $W = \tilde{\gamma}_t$. Here we examine, in a general setting, the standard approximation to the distribution of W based on assuming $X_2 > 0$; see (10.14). First we derive the exact distribution of W.

The distribution of W in the case $\theta_1 = \theta_2 = 0$ was discussed by Geary (1930). Fieller (1932), and more recently Marsaglia (1965) considered the general problem with non-zero means. In fact the latter studied the ratio $Z = \dfrac{a + Y_1}{b + Y_2}$ (our notation), where Y_1 and Y_2 are independent standard normal variables. The connection between Z and W is said (incorrectly) to be that "It suffices to study [Z]; translations and changes of scale will provide the general ratio [W]". This assumes there to be four parameters, whereas there are five (θ_1, θ_2, σ_1^2, σ_2^2 and ρ). In fact Z has no great advantage over W anyway, since the distributions of both involve the bivariate normal distribution.

If the bivariate normal density of (X_1, X_2) is $\varphi(x, y)$, and the density of W is $f(w)$, then

$$f(w) = \int_{-\infty}^{\infty} |y| \, \varphi(yw, y)\,dy \,.$$

Substituting for $\varphi(x, y)$ and carrying out a simple integration gives

$$f(w) = \frac{b(w)\,d(w)}{\sqrt{(2\pi)}\,\sigma_1\sigma_2 a^3(w)} \left[\Phi\left\{\frac{b(w)}{\sqrt{(1-\rho^2)}\,a(w)}\right\} - \Phi\left\{-\frac{b(w)}{\sqrt{(1-\rho^2)}\,a(w)}\right\}\right]$$

$$+ \frac{\sqrt{(1-\rho^2)}}{\pi\sigma_1\sigma_2 a^2(w)} \exp\left\{-\frac{c}{2(1-\rho^2)}\right\} \,, \qquad (15.1)$$

where

$$a(w) = \left(\frac{w^2}{\sigma_1^2} - \frac{2\rho w}{\sigma_1\sigma_2} + \frac{1}{\sigma_2^2}\right)^{\frac{1}{2}} \,,$$

$$b(w) = \frac{\theta_1 w}{\sigma_1^2} - \frac{\rho(\theta_1 + \theta_2 w)}{\sigma_1\sigma_2} + \frac{\theta_2}{\sigma_2^2} \,,$$

$$c = \frac{\theta_1^2}{\sigma_1^2} - \frac{2\rho\theta_1\theta_2}{\sigma_1\sigma_2} + \frac{\theta_2^2}{\sigma_2^2}$$

and $$d(w) = \exp\left\{\frac{b^2(w)/a^2(w) - c}{2(1-\rho^2)}\right\} \qquad (15.2)$$

This result was obtained by Fieller (1932).

To obtain the cumulative distribution function $F(w)$ we introduce the familiar notation

$$L(h, k; \rho) = \frac{1}{2\pi\sqrt{(1-\rho^2)}} \int_h^\infty \int_k^\infty \exp\left\{-\frac{1}{2(1-\rho^2)} (x^2 - 2\rho xy + y^2)\right\} dxdy,$$

Then

$$F(w) = L\left\{\frac{\theta_1 - \theta_2 w}{\sigma_1\sigma_2 a(w)}, -\frac{\theta_2}{\sigma_2}; \frac{\sigma_2 w - \rho\sigma_1}{\sigma_1\sigma_2 a(w)}\right\}$$

$$+ L\left\{\frac{\theta_2 w - \theta_1}{\sigma_1\sigma_2 a(w)}, \frac{\theta_2}{\sigma_2}; \frac{\sigma_2 w - \rho\sigma_1}{\sigma_1\sigma_2 a(w)}\right\}. \qquad (15.3)$$

As $\frac{\theta_2}{\sigma_2} \to \infty$, $F(w) \to \Phi\left\{\frac{\theta_2 w - \theta_1}{\sigma_1\sigma_2 a(w)}\right\}$. This may be seen another way, for if X_2 were positive with probability one, then

$$F(w) = \text{pr}(X_1 - wX_2 \le 0) = \Phi\left\{\frac{\theta_2 w - \theta_1}{\sigma_1\sigma_2 a(w)}\right\}.$$

This suggests that, if $0 < \sigma_2 \ll \theta_2$, a good approximation to $F(w)$ is

$$F^*(w) = \Phi\left\{\frac{\theta_2 w - \theta_1}{\sigma_1\sigma_2 a(w)}\right\},$$

which has derivative

$$f^*(w) = \frac{b(w)d(w)}{\sqrt{(2\pi)}\sigma_1\sigma_2 a^3(w)} \qquad (15.4)$$

Now

$$\begin{aligned} F(w) &= \text{pr}(X_1 - wX_2 \le 0, X_2 > 0) + \text{pr}(X_1 - wX_2 \ge 0, X_2 < 0) \\ &= \text{pr}(X_1 - wX_2 \le 0) + \text{pr}(X_2 < 0) - 2\text{pr}(X_1 - wX_2 \le 0, X_2 < 0) \\ &= F^*(w) + \text{pr}(X_2 < 0)\left\{1 - 2\text{pr}(X_1 - wX_2 \le 0 \mid X_2 < 0)\right\}, \end{aligned} \qquad (15.5)$$

and hence

$$|F(w) - F^*(w)| \le \mathrm{pr}(X_2 < 0) = \Phi\left(-\frac{\theta_2}{\sigma_2}\right) \qquad (15.6)$$

This bound is attained at $w = \pm\infty$, for by (15.5) or the definition of $F^*(w)$ we have

$$F^*(-\infty) = \Phi\left(-\frac{\theta_2}{\sigma_2}\right) \quad \text{and} \quad F^*(+\infty) = 1 - \Phi\left(-\frac{\theta_2}{\sigma_2}\right); \qquad (15.7)$$

the c.d.f. $F(w)$ is proper. If θ_2 were negative, we would use $\mathrm{pr}(X_1 - wX_2 \ge 0)$ as the approximation. In either case $F^* \to F$ uniformly as $\sigma_2 \to 0$.

Note from (15.4) that the approximate density $f^*(w)$ is negative when $b(w) < 0$, i.e. for w in the region

$$w < -\psi_1/\psi_2 \qquad (\psi_2 < 0)$$

or

$$w > -\psi_1/\psi_2 \qquad (\psi_2 > 0), \qquad (15.8)$$

where $\psi_1 = \rho\theta_1\sigma_1\sigma_2 - \theta_2\sigma_1^2$ and $\psi_2 = \rho\theta_2\sigma_1\sigma_2 - \theta_1\sigma_2^2$.

Further, when $b(w) > 0$ it is easy to show that $f(w) > f^*(w)$. For by (15.1) and (15.4) we have

$$\frac{f(w)}{f^*(w)} = \Phi\left\{\frac{b(w)}{a(w)\sqrt{(1-\rho^2)}}\right\} - \Phi\left\{-\frac{b(w)}{a(w)\sqrt{(1-\rho^2)}}\right\}$$

$$+ \frac{2\sqrt{(1-\rho^2)}\,a(w)}{b(w)} \varphi\left\{\frac{b(w)}{a(w)\sqrt{(1-\rho^2)}}\right\}$$

The result follows on using the inequality $\Phi(-u) < \varphi(u)/u$ for $u > 0$. Therefore $F(w) - F^*(w)$ is monotonically increasing from

$$- \Phi\left(-\frac{\theta_2}{\sigma_2}\right) \text{ at } w = -\infty \text{ to } +\Phi\left(-\frac{\theta_2}{\sigma_2}\right) \text{ at } w = +\infty.$$

By (15.5) we have

$$F(w) = F^*(w) + \text{pr}(X_2 < 0) - 2\int_{-\infty}^{0}\int_{-\infty}^{wy} \varphi(x,y)\,dx\,dy, \qquad (15.9)$$

where

$$\varphi(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{(1-\rho^2)}} \times$$

$$\exp\left[-\frac{1}{2(1-\rho^2)}\left\{\frac{(x-\theta_1)^2}{\sigma_1^2} - \frac{2\rho(x-\theta_1)(y-\theta_2)}{\sigma_1\sigma_2} + \frac{(y-\theta_2)^2}{\sigma_2^2}\right\}\right].$$

The double integral in (15.9) is easily reduced to

$$\int_{-\infty}^{0}\int_{-\infty}^{wy} \varphi(x,y)\,dx\,dy = \int_{-\infty}^{-\frac{\theta_2}{\sigma_2}} \varphi(u)\,\Phi\left\{\frac{(w\sigma_2 - \rho\sigma_1)u + (w\theta_2 - \theta_1)}{\sigma_1\sqrt{(1-\rho^2)}}\right\} du, \qquad (15.10)$$

where $\varphi(u) = (2\pi)^{-\frac{1}{2}} \exp(-\frac{1}{2}u^2)$.

Hence the point at which $F(w) = F^*(w)$ is the value of w satisfying

$$\int_{-\infty}^{-\frac{\theta_2}{\sigma_2}} \varphi(u)\,du = 2\int_{-\infty}^{-\frac{\theta_2}{\sigma_2}} \varphi(u)\,\Phi\left\{\frac{(w\sigma_2 - \rho\sigma_1)u + (w\theta_2 - \theta_1)}{\sigma_1\sqrt{(1-\rho^2)}}\right\} du. \qquad (15.11)$$

The simplest case is when $\psi_2 = 0$, the solution being $w = \rho\sigma_1/\sigma_2$; this is also the only time when $f^*(w) > 0$ for all w. In general it is easy to show that the solution w_o of (15.11) satisfies

$$w_o < \frac{\rho\sigma_1}{\sigma_2} \qquad \text{if} \quad \psi_2 < 0$$

and

$$w_o > \frac{\rho\sigma_1}{\sigma_2} \qquad \text{if} \quad \psi_2 > 0. \qquad (15.12)$$

Since $|F(w) - F^*(w)|$ only attains its bound at $w = \pm\infty$, we would like to obtain sharper bounds on the difference for finite w. Integration by parts in (15.10) gives

$$\int_{-\infty}^{0}\int_{-\infty}^{-\frac{\theta_2}{\sigma_2}} \varphi(x,y)\,dxdy = \Phi\left(-\frac{\theta_2}{\sigma_2}\right)\Phi\left\{\frac{\rho\sigma_1\theta_2 - \sigma_2\theta_1}{\sigma_1\sigma_2\sqrt{(1-\rho^2)}}\right\}$$

$$-\frac{(w\sigma_2 - \rho\sigma_1)}{\sigma_1\sqrt{(1-\rho^2)}}\int_{-\infty}^{-\frac{\theta_2}{\sigma_2}} \Phi(u)\varphi\left\{\frac{(w\sigma_2 - \rho\sigma_1)u + (w\theta_2 - \theta_1)}{\sigma_1\sqrt{(1-\rho^2)}}\right\} du$$

and since $\theta_2 > 0$ we may use the inequality $\Phi(u) \le \varphi(u)/(-u)$, giving

$$I_o \le \int_{-\infty}^{0}\int_{-\infty}^{wy} \varphi(x,y)\,dx\,dy \le I_o + I_1(w) \qquad (w < \rho\sigma_1/\sigma_2)$$

and $$I_o + I_1(w) \le \int_{-\infty}^{0}\int_{-\infty}^{wy} \varphi(x,y)\,dx\,dy \le I_o \qquad (w > \rho\sigma_1/\sigma_2),$$

(15.13)

where $I_o = \Phi\left(-\frac{\theta_2}{\sigma_2}\right)\Phi\left\{\frac{\rho\sigma_1\theta_2 - \sigma_2\theta_1}{\sigma_1\sigma_2\sqrt{(1-\rho^2)}}\right\}$

and $I_1(w) = \frac{\rho\sigma_1 - w\sigma_2}{\sigma_1\theta_2 a(w)} \quad \varphi\left\{\frac{w\theta_2 - \theta_1}{\sigma_1\sigma_2 a(w)}\right\}\Phi\left\{-\frac{b(w)}{a(w)\sqrt{(1-\rho^2)}}\right\}$.

Therefore, by (15.9), the difference $F(w) - F^*(w)$ has the following bounds:

$$J \le F(w) - F^*(w) \le J + K(w) \qquad (w > \rho\sigma_1/\sigma_2)$$

and $$J + K(w) \le F(w) - F^*(w) \le J \qquad (w \le \rho\sigma_1/\sigma_2) \; , \qquad (15.14)$$

where $J = \Phi\left(-\frac{\theta_2}{\sigma_2}\right)\left[1 - 2\,\Phi\left\{\frac{\psi_2}{\sigma_1\sigma_2^2\sqrt{(1-\rho^2)}}\right\}\right]$

and $K(w) = \frac{2(w\sigma_2 - \rho\sigma_1)}{\sigma_1\theta_2 a(w)} \quad \varphi\left\{\frac{w\theta_2 - \theta_1}{\sigma_1\sigma_2 a(w)}\right\}\Phi\left\{-\frac{b(w)}{a(w)\sqrt{(1-\rho^2)}}\right\}$.

(15.15)

When $w = \rho\sigma_1/\sigma_2$, $F(w) - F^*(w) = J$.

Together with (15.6), (15.14) gives improved bounds on the difference $F(w) - F^*(w)$. Also the solution to (15.11), bounded in (15.12), indicates where $F(w)$ overtakes $F^*(w)$.

To see more clearly the implication of these results, consider first the simple example of a single linear regression

$$y_t = \alpha + \beta u_t + \varepsilon_t \qquad (t = 1, \ldots, T)$$

where $\varepsilon_1, \ldots, \varepsilon_T$ are independent $N(0, 1)$ and $X_1 = \hat{\alpha}$, $X_2 = \hat{\beta}$ are the m.l.e.'s of α, β. In particular let $\alpha = 0$, $\beta = 0.2$ and $T = 10$. Table 15.1 contains exact values of $F(w)$, $F^*(w)$, the overall bound in (15.6) and the bounds of (15.14) for seven values of w.

w	F(w)	$F^*(w)$	overall bound	bounds from (15.14) lower	upper
-15.0	0.05940	0.09403	0.03464	-0.03464	-0.03462
-10.0	.08994	.12458	03464	- .03464	- .03462
- 5.0	.16740	.20203	.03464	- .03464	- .03462
0.0	.46536	.50000	.03464	- .03464	- .03462
5.0	.96445	.99908	.03464	- .03463	- .03462
10.0	.96568	.99967	.03464	- .03462	- .03352
15.0	.96838	.99698	.03464	- .03462	- .02495

Table 15.1 Comparison of $F(w)$, $F^*(w)$ and bounds on their difference for $\alpha = 0$, $\beta = 0.2$, $T = 10$.

In this example the difference $F(w) - F^*(w)$ is approximately constant over a wide range, in fact in all but the extreme tails. The closeness of the lower and upper bounds confirms this. The maximum of $F^*(w)$ is at $w = 7.0$, after which $F^*(w)$ decreases albeit slowly. The practical significance of the results is that if the overall bound is less than 0.05, say, the exact probability $F(w)$ may be approximated accurately by $F^*(w) + pr(X_2 < 0)$. These remarks are, of course, not necessarily true outside the regression context. One use of these precise bounds is in checking computed values of $F(w)$, which involves bivariate integrals; computation of $F^*(w)$ and the bounds is less liable to error. The method used for computing values of $F(w)$ was an iterative two-variate Simpson's Rule.

For the double regression relationship, similar results hold. In the examples illustrated by normal plots in Section 11, the overall bound is very small indeed. For example in Figure 11.1, where the approximation is probably least accurate, the difference between exact and approximate values of $G_1(13)$ is less than $\Phi(-9)$, which is negligible.

REFERENCES

Ahsanullah, A.K.M. and Qurashi, M.M. (1965). Fine-structure discontinuities in the inter-molecular activity energy flow of 2 to 30% aqueous ethyl alcohol. Proc. Roy. Soc. A, 285, 480-500.

Barnard, G.A. (1959). Control charts and stochastic processes. J. R. Statist. Soc. B, 21, 239-57.

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

Brown, R.L. and Durbin, J. (1968). Methods of investigating whether a regression relationship is constant over time. European Meeting 1968. Selected Statistical Papers I, 37-45. Amsterdam: Mathematisch Centrum.

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

Cox, D.R. (1961). Tests of separate families of hypotheses. Proc. 4th Berkeley Symposium 1, 105-123.

Feder, P.I. and Sylwester, D.L. (1968). On the asymptotic theory of least squares estimation in segmented regression: identified case. (Abstract) Ann. Math. Statist. 39, 1362.

Feller, W. (1966). An Introduction to Probability Theory and Its Applications, 2. New York: Wiley.

Fieller, E.C. (1932). The distribution of the index in a normal bivariate population. Biometrika 24, 428-40.

Geary, R.C. (1930). The frequency distribution of the quotient of two normal variates. J. R. Statist. Soc. 93, 442-6.

Hudson, D.J. (1966). Fitting segmented curves whose join points have to be estimated. J. Am. Statist. Ass. 61, 1097-1129.

Kendall, M.G. and Stuart, A. (1961). The Advanced Theory of Statistics 2. London: Griffin.

Marsaglia, G. (1965). Ratios of normal variables and ratios of sums of uniform variables. J. Am. Statist. Ass. 60, 193-204.

McLaren, A.D. (1965). The fit of two straight lines when their intersection has a specified abscissa. Unpublished paper in Statistical Laboratory, Cambridge University.

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

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.

Quandt, R.E. (1958). The estimation of the parameters of a linear regression system obeying two separate regimes. J. Am. Statist. Ass. 53, 873-80.

Robison, D.E. (1964). Estimates for the points of intersection of two polynomial regressions. J. Am. Statist. Ass. 59, 214-24.

Sprent, P. (1961). Some hypotheses concerning two-phase regression lines. Biometrics 17, 634-45.