STATISTICAL ESTIMATES *and* TRANSFORMED BETA-VARIABLES

By GUNNAR BLOM

JOHN WILEY & SONS, INC. / NEW YORK

ALMQVIST & WIKSELL / STOCKHOLM

Almqvist & Wiksell / Gebers förlag AB
STOCKHOLM

John Wiley & Sons, Inc.
NEW YORK

PRINTED IN SWEDEN BY

Almqvist & Wiksells

BOKTRYCKERI AKTIEBOLAG

UPPSALA 1959

ACKNOWLEDGEMENTS

I wish to convey my sincere gratitude to

my teacher, Professor Harald Cramér, for his encouragement and interest in my work, and for valuable advice concerning the presentation of the results,

Ivar Weibull, Civil Engineer, who posed the plotting problem connected with the normal probability paper, and thus gave me the impulse to embark upon the present investigation,

Martin Sandelius, Fil. Lic., who read a preliminary draft of Parts II and III and made many valuable comments,

Bertil Matérn, Fil. Lic., for several helpful suggestions,

my wife for typing the manuscript and reading the proofs,

the Swedish Natural Science Research Council and the Henrik Granholm Stiftelse for financial grants.

Malmö, March 1958.

Gunnar Blom

PREFACE

The work leading up to this thesis was begun in 1954 at a time when I had become interested in the problem of how to plot points on a normal probability paper. This research involved a study of the properties of order statistics and led to the introduction of the α, β-correction, described in Chapter 6 of the present work.

When, in the same year, in a paper in the Journal of the American Statistical Association, Chernoff and Lieberman demonstrated the existence of a close relationship between the plotting problem and the much more general problem of estimating parameters by means of linear functions of order statistics, my interest in the first of these problems was extended also to the second. Another source of inspiration was Jung's paper in Arkiv för matematik in 1955, in which, for the first time, the laborious procedure for finding linear estimates with minimum variance was replaced by an approximate one. The nearly best linear estimates discussed in the present thesis owe their existence partly to Jung's work.

It soon became desirable to study the asymptotic properties of the nearly best estimates. Fisher's classical theory of estimation proved too specialized for this purpose, and, as a third and final stage in the research reported in this thesis, a more general theory was developed covering all possible continuous probability distributions.

The three stages in my work outlined above are reflected in the division of the thesis into three parts, which, however, are presented in an order differing from the chronological one.

Part I contains a generalization of that part of Fisher's theory of estimation which concerns the asymptotic minimum variance of an estimate. The results are summarized in Sections 1.1 and 2.1.

In Part II, general properties of transformed beta-variables[1] are investigated. Special attention is paid to the problem of finding approximations to the first two moments of single transformed beta-variables and of linear combinations of such variables. Further, the asymptotic distribution of a linear combination is studied.

[1] For the definition of a transformed beta-variable, which is a more general concept than the familiar term order statistic, the reader is referred to Section 3.1.

The results obtained in this part are applied in Part III to linear estimation problems. Apart from these applications, Part II is deemed to have some interest of its own, being a contribution to the general theory of transformed beta-variables.

Part III is devoted to a study of the problem of constructing linear estimates of location and scale parameters in a continuous distribution of any form. A method is presented for constructing nearly best linear estimates of such parameters, the name being suggested by the fact that these estimates are often highly efficient as compared to minimum variance estimates.

A more detailed summary of the results obtained in each chapter in Part II and Part III is given in Section 3.3.

It may be appropriate to finish this preface by giving some advice to different categories of readers concerning the selection of material from the thesis.

The reader who wants a general survey of the results is advised to read the introduction to each chapter and to glance, at least, at Chapter 4. Furthermore, he might read Sections 5.3–5.5, 6.7–6.10, 8.2–8.4, 8.7, 10.3, 10.8, 12.2, 12.4, 12.6, 14.5, and consult other sections when necessary.

Readers with a special interest in the practical problem of constructing estimates of location and scale parameters are advised to read Chapters 9–12 and 14, and take, at least, a glance at Chapter 13.

The specialist in the general theory of estimation is recommended to study Chapters 1–2, 10, 13–14, and look up references to other chapters when these are needed for a complete understanding of the results.

CONTENTS

PART I

CONTRIBUTIONS TO THE LARGE SAMPLE THEORY OF ESTIMATION

CHAPTER 1

ESTIMATION OF A SINGLE UNKNOWN PARAMETER

1.1. Introduction

The theory of estimation was put forward by R. A. Fisher (1921, 1925). A useful review of this theory is given in Fisher (1935). Among his many contributions to the theory of estimation, we shall here only mention Fisher's well-known result concerning the asymptotic behaviour of an estimate α^* of an unknown parameter α in a frequency function (fr. f.) $f(z;\alpha)$.

Let us suppose that α^* is a function of the observations in a random sample of size n from the population with the fr. f. $f(z;\alpha)$. If α^* is asymptotically unbiased, its variance has a lower limit so that, asymptotically,

$$\text{var}\ \alpha^* \geqslant \frac{1}{I}, \tag{1.1.1}$$

where I is the amount of information in the sample, defined by

$$I = nE\left(\frac{\partial \log f}{\partial \alpha}\right)^2 = n\int_{-\infty}^{\infty}\left(\frac{\partial \log f}{\partial \alpha}\right)^2 f(z;\alpha)\,dz. \tag{1.1.2}$$

Fisher partly used heuristic methods when proving this result and gave no exact conditions. He realized that there are also other possibilities, as shown by a passage on p. 350 in his paper from 1921, where, in an analysis of the rectangular distribution, he demonstrated that the variance of the mean of the extreme values in the sample has the order of magnitude n^{-2}.

In later years, Fisher's inequality (1.1.1) has been studied by Rao (1945), Cramér (1946 *a*; Ch. 32), and others, who have shown that it holds good also for unbiased estimates, based upon samples of *any* size, provided that certain specified conditions are fulfilled. When the formula is interpreted in this way, it is generally called the Cramér–Rao inequality.

Returning to the case of large samples, we observe that, apart from Fisher's results, our knowledge concerning the behaviour of the variance of an estimate is still incomplete. For example, very little is known about the situations arising when $f(z)$ has discontinuities, or when the integral appearing in (1.1.2) is zero.

In this chapter we shall attempt to give a general solution to the question of the asymptotic behaviour of the variance of an estimate, constructed by means of a sample from a continuous cumulative distribution function (cdf.). We shall replace Fisher's result (1.1.1) by a more general inequality, which is valid for a very wide class of asymptotically unbiased estimates. The proof of this inequality is given in 1.2. In 1.3 we shall introduce a classification of distribution functions into three categories, viz. Type 1, Type 2, and Type 3 distributions, which together embrace all possible continuous distributions. Type 1 distributions fall within the scope of Fisher's theory, while the other types do not seem to have been systematically studied before.

1.2. A general theorem

We shall describe the situation considered in the introduction in more detail. Suppose that the random variable z has the continuous cdf. $F(z;\alpha)$, where α is an unknown parameter belonging to an open interval Ω. The range of variation of z is denoted by (A, B) and may depend upon α. Let $z = G(u;\alpha)$ denote the inverse function of $u = F(z;\alpha)$. We shall suppose that the partial derivatives $\dfrac{\partial F}{\partial z} = f(z;\alpha)$ and $\dfrac{\partial F}{\partial \alpha}$ exist for any z and for any $\alpha \in \Omega$.

Let α^* be an estimate of α, based upon n independent observations of z. This estimate is a random variable defined over the n-dimensional Euclidean sample-space corresponding to the n observations. Let $\xi_1, \dots, \xi_n$ denote any point in this space and denote by P the probability mass corresponding to the observations. We shall suppose that the mean

$$E\,\alpha^* = \int \alpha^*\,dP \tag{1.2.1}$$

exists and tends to α when n approaches infinity; in other words, we assume that α^* is an asymptotically unbiased estimate of α.

We shall derive a lower bound for the variance of α^* which is valid for large n. The exact meaning of this statement will become clear later. In order to construct this bound, we proceed as follows. Let $\alpha_0 \in \Omega$ be some fixed value of the parameter. Define a sequence

$$\lambda_0 < \lambda_1 < \cdots < \lambda_{n+1}$$

by

$$F(\lambda_i;\, \alpha_0) = \frac{i}{n+1}, \quad (i = 0, 1, \ldots, n+1). \tag{1.2.2}$$

Clearly $$\lambda_0 = A, \quad \lambda_{n+1} = B.$$

Alternatively, we may write

$$\lambda_i = G\left(\frac{i}{n+1};\, \alpha_0\right).$$

Further, define for any $\alpha \in \Omega$

$$\Delta_i = F(\lambda_{i+1};\, \alpha) - F(\lambda_i;\, \alpha), \quad (i = 0, 1, \ldots, n). \tag{1.2.3}$$

We observe that, in the special case where $\alpha = \alpha_0$,

$$\Delta_i = \frac{1}{n+1}. \tag{1.2.4}$$

Divide the sample-space into $(n+1)^n$ parts by means of the planes $\xi_i = \lambda_j$, $(i = 1, \ldots, n;\ j = 0, \ldots, n+1)$, and assign numbers $1, 2, \ldots$ to these parts according to any convenient principle. Denote by α_ν^* the value of the estimate in some arbitrarily chosen point in the νth part. Further, let P_ν be the probability mass corresponding to this part. Evidently, the quantities P_ν for the different parts are terms in the series obtained by expanding the expression

$$\left(\sum_{i=0}^{n} \Delta_i\right)^n$$

into a sum of products of the Δ_i's. It should be observed that Δ_i and P_ν depend both upon α and upon the given value α_0.

The sum $$\sum_{\nu=1}^{(n+1)^n} \alpha_\nu^* P_\nu$$

is, by the definition of a Riemann–Stieltjes integral, an approximation to the integral appearing in (1.2.1). As the estimate is assumed to be asymptotically unbiased, we have

$$\sum_{\nu} \alpha_{\nu}^{*} P_{\nu} = \alpha + R_{n}^{(1)}, \tag{1.2.5}$$

where the error term tends to zero when n approaches infinity. Further, we write

$$\sum_{\nu} (\alpha_{\nu}^{*} - \alpha)^{2} P_{\nu} = \text{var}\ \alpha^{*} + R_{n}^{(2)}, \tag{1.2.6}$$

where var α^{*} denotes the variance of α^{*}. Generally, both terms in the right member of this relation tend to zero when n tends to infinity. We shall, however, make the somewhat more restrictive assumption that the error term converges to zero more rapidly than the leading term; in other words, var α^{*} and the sum in the left member are assumed to be asymptotically equivalent.[1]

Returning to (1.2.5), let us suppose that this relation may be differentiated term by term with respect to α. Observing that

$$\sum_{\nu} \frac{d P_{\nu}}{d \alpha} = 0,$$

we obtain

$$\sum_{\nu} (\alpha_{\nu}^{*} - \alpha) \frac{d P_{\nu}}{d \alpha} = 1 + \frac{d R_{n}^{(1)}}{d \alpha},$$

where the remainder term converges to zero when n tends to infinity. Using Cauchy's inequality, we obtain

$$\sum_{\nu} (\alpha_{\nu}^{*} - \alpha)^{2} P_{\nu} \cdot \sum_{\nu} \frac{1}{P_{\nu}} \left(\frac{d P_{\nu}}{d \alpha}\right)^{2} \geqslant \left(1 + \frac{d R_{n}^{(1)}}{d \alpha}\right)^{2}, \tag{1.2.7}$$

where the summation extends over all ν with non-zero P_{ν}. The second factor in the left member can after some easy calculation be reduced to

$$n \sum_{i} \frac{1}{\Delta_{i}} \left(\frac{d \Delta_{i}}{d \alpha}\right)^{2}, \tag{1.2.8}$$

where the summation should be made over all $i = 0, 1, \ldots, n$ with non-zero Δ_{i}.

Now take $\alpha = \alpha_{0}$ in (1.2.7). Using (1.2.4), we see that the expression (1.2.8) is then, apart from a multiplicative factor $1 + 1/n$, equal to the value of the function

$$I = n^{2} \sum_{i=0}^{n} \left(\frac{d \Delta_{i}}{d \alpha}\right)^{2}$$

[1] For a formal definition of this term, which will be encountered many times in the sequel, see p. 35.

for $\alpha = \alpha_0$. By this observation, and using equation (1.2.6) with $\alpha = \alpha_0$, we conclude that

$$(\text{var}\ \alpha^* + R_n^{(2)})\, I \geqslant \left(1 + \frac{d\, R_n^{(1)}}{d\, \alpha}\right)^2.$$

Since we have assumed that $R_n^{(2)}$ converges more rapidly than var α^*, we infer that, when $n \to \infty$, the product $I \cdot \text{var}\ \alpha^*$ remains greater than or equal to unity. Since α_0 is arbitrary, this result holds good for any $\alpha \in \Omega$. Thus we have proved the following theorem.

THEOREM. *Let $F(z;\alpha)$ be a continuous cdf., $G(u;\alpha)$ the inverse function of F, and α^* an asymptotically unbiased estimate of α, based upon a random sample of n observations. Define, further,*

$$I = n^2 \sum_{i=0}^{n} \left(\frac{\partial F(\lambda_{i+1})}{\partial \alpha} - \frac{\partial F(\lambda_i)}{\partial \alpha}\right)^2, \tag{1.2.9}$$

where $\frac{\partial F(\lambda_i)}{\partial \alpha}$ *is the value of the partial derivative* $\frac{\partial F}{\partial \alpha}$ *for*

$$z = \lambda_i = G\left(\frac{i}{n+1};\ \alpha\right).$$

If

(a) $\frac{\partial F}{\partial \alpha}$ *exists for any α belonging to an open interval Ω,*

(b) *the series appearing in* (1.2.5) *may be differentiated term by term for any $\alpha \in \Omega$,*

(c) *the sum of the series appearing in* (1.2.6) *is asymptotically equivalent to* var α^*,

then, for any $\alpha \in \Omega$,

$$\liminf_{n \to \infty}\ (I\ \text{var}\ \alpha^*) \geqslant 1. \tag{1.2.10}$$

It should be noted that, as $F(\lambda_0) = 0$ and $F(\lambda_{n+1}) = 1$, we have $\frac{\partial F(\lambda_i)}{\partial \alpha} = 0$ for $i = 0$ and $i = n+1$.

The relation (1.2.10) is equivalent to stating that

$$\text{var}\ \alpha^* \geqslant \frac{1}{I} \tag{1.2.11}$$

when n is large enough. Following R. A. Fisher, we shall call I the information available in a large sample of n values. The quantity $1/I$

will be called the *asymptotic minimum variance* of the estimate. If an estimate has a variance, which is asymptotically equivalent to $1/I$, it will be called *asymptotically efficient.* As will follow from the discussion in 1.3, this definition is in accordance with the definition of asymptotic efficiency used by Cramér (1946 *a*, p. 489) in all cases where the latter is applicable.

Rectangular distribution.

To get a first, very simple application of the theorem, we shall suppose that z is rectangularly distributed in the range $\mu \pm \frac{1}{2}$, where μ is unknown. We then have

$$F(z;\mu)=\begin{cases} 0 & \text{for } z-\mu \leqslant -\frac{1}{2} \\ z-\mu+\frac{1}{2} & \text{for } |z-\mu|<\frac{1}{2} \\ 1 & \text{for } z-\mu \geqslant \frac{1}{2}. \end{cases}$$

We get $I=2n^2$, and (1.2.11) reduces to $\text{var}\ \mu^* \geqslant 1/(2n^2)$. The mean of the extreme values of the sample has variance $1/[2(n+1)(n+2)]$ (cf. 9.2), and is thus asymptotically efficient.

The main application of the theorem will be given in Chapter 13, where it will be shown that, under very general conditions, asymptotically efficient estimates exist when α is a location or scale parameter in a continuous distribution of any form. This general result contains the example just given as a very special case.

It may be supposed that the conditions of the theorem can be stated in a simpler form. In particular, it is believed that condition (c) is on the whole superfluous. It should also be pointed out that the inequality (1.2.11) can easily be modified so as to apply also to asymptotically biased estimates. Finally, it may be mentioned that the improvement of the Cramér–Rao inequality performed by Bhattacharya (1946) can also be extended to the general situation discussed in this chapter.

1.3. Classification of distributions

The inequality (1.2.11) is very general, as it covers widely different situations. It is convenient to classify these situations according to the asymptotic behaviour of I. Properly speaking, this classification concerns the behaviour of the underlying cdf. For this purpose, we need two conditions.

Condition C 1. *The function* $\dfrac{\partial F(z;\alpha)}{\partial \alpha}$ *is a continuous function of z in the open interval* $A < z < B$, *and tends to zero when* $z \to A$ *or* $z \to B$.

Condition C 2. *The integral*

$$E\left(\frac{\partial \log f}{\partial \alpha}\right)^2 = \int_{-\infty}^{\infty} \left(\frac{\partial \log f}{\partial \alpha}\right)^2 f(z;\alpha)\, dz \qquad (1.3.1)$$

exists and is greater than zero.

Note that, when the interval (A, B) is finite, Condition C 1 may be given the simple formulation: $\dfrac{\partial F}{\partial \alpha}$ *is continuous for any z in the interval* $-\infty < z < \infty$. For $\dfrac{\partial F}{\partial \alpha}$ is then continuous, in particular, for $z = A$ and $z = B$; observing that $F = 0$ for $z \leqslant A$, and $F = 1$ for $z \geqslant B$, we conclude that $\dfrac{\partial F}{\partial \alpha}$ equals zero for $z = A$ and $z = B$.

We now introduce the following classification, which will play an important rôle in the sequel.

Type 1 distribution	Conditions C 1 and C 2 are satisfied.
Type 2 distribution	Condition C 1 is not satisfied, Condition C 2 is satisfied.
Type 3 distribution	Condition C 2 is not satisfied.

When, in either of these cases, an asymptotically efficient estimate exists (cf. 1.2), it may be termed an *asymptotically efficient estimate of Type 1*, etc., or, shorter, a *Type 1 estimate*, etc.

In one special case it will prove convenient to modify the rules of classification: If C 1 does not hold and the integral in (1.3.1) is zero, we shall classify the distribution as Type 2. This rule affects the rectangular distribution.

We shall make some comments concerning the classification. Suppose, first, that the cdf. is of Type 1. By an analysis of I, defined by (1.2.9), we easily see that this quantity is then asymptotically equivalent to

$$n\, E\left(\frac{\partial \log f}{\partial \alpha}\right)^2.$$

Thus the inequality (1.2.11) reduces to Fisher's inequality (1.1.1). If an asymptotically efficient estimate exists, its variance has the order of magnitude n^{-1}.

Secondly, consider a Type 2 situation. Suppose that $\frac{\partial F}{\partial \alpha}$, considered as a function of z, has at most a denumerable number of discontinuity points $w_1, w_2, \ldots, w_l$ with 'jumps' $\Delta_1, \Delta_2, \ldots, \Delta_l$. These points may be situated inside or at the ends of the range of variation of z. It is realized that I is then asymptotically equivalent to

$$n^2 \sum_{i=1}^{l} \Delta_i^2.$$

Thus the inequality (1.2.11) is specialized to

$$\text{var } \alpha^* \geqslant \frac{1}{n^2 \sum \Delta_i^2}.$$

Provided that the sum in the denominator is finite, we see that an asymptotically efficient estimate (if existent) has a variance of order n^{-2}.

The rectangular distribution, already mentioned in 1.2, and the exponential distribution are the best known examples of Type 2 distributions. We shall discuss these and other similar cases in 13.7.

If, finally, the cdf. belongs to Type 3, a special analysis has to be made of the limit in (1.2.11) in each particular case, and no general statements are possible with respect to the order of $1/I$. The reader is referred to 13.8 and also to Table 8, p. 165, where examples of Type 3 distributions are given. It follows from these examples that the order of an asymptotically efficient estimate may, for example, equal any power $n^{-\varkappa}$, where $1 \leqslant \varkappa \leqslant 2$. However, other situations are also possible.

It has always been maintained that the information I is additive, i. e. the amounts of information available in each of two independent samples are added when the samples are combined into one single sample. We now see that this statement is true when the samples are taken from a Type 1 distribution, but not necessarily otherwise. For instance, in the case of a Type 2 distribution, it follows from the above that, when two large samples of equal size are united into one sample, the total information becomes four times greater than the information inherent in each subsample.

CHAPTER 2

ESTIMATION OF SEVERAL UNKNOWN PARAMETERS

2.1. Introduction

The theorem concerning the asymptotic minimum variance and the classification of distributions discussed in the preceding chapter are extended to the case of several unknown parameters. The results are related to but more general than the multi-parameter version of the Cramér–Rao inequality considered by Cramér (1946 *b*) and Rao (1947).

2.2. A general theorem for several unknown parameters

Let z have the continuous cdf. $F(z; \alpha_1, \dots, \alpha_k)$, where $\alpha_1, \dots, \alpha_k$ are k unknown parameters, and each parameter α_r, $(r=1, \dots, k)$, is situated in an open interval Ω_r. Let us suppose that the partial derivatives $\dfrac{\partial F}{\partial z} = f(z; \alpha_1, \dots, \alpha_k)$ and $\dfrac{\partial F}{\partial \alpha_r}$ exist for any value z and for any particular values $\alpha_r \in \Omega_r$, $(r=1, \dots, k)$, of the parameters.

To estimate the parameters, we shall suppose that we have at our disposal a random sample of n independent observations of z. Let $\alpha_1^*, \dots, \alpha_k^*$ be the estimates of the parameters, based upon these observations. We shall assume that the means

$$E\,\alpha_r^* = \int \alpha_r^*\, dP, \quad (r=1, \dots, k), \tag{2.2.1}$$

exist and tend to α_r when n tends to infinity.

Denote, further, by V the variance–covariance matrix with the elements[1] $\operatorname{cov}(\alpha_r^*, \alpha_s^*)$, $(r, s=1, \dots, k)$. This matrix is the main object of our investigation.

In the first part of the investigation we proceed in the same way as in 1.2. Let α_{r0}, $(r=1, \dots, k)$, be any fixed values of the parameters. Define quantities

$$\lambda_0 < \lambda_1 < \cdots < \lambda_{n+1}$$

by the relations

---

[1] To avoid duplication of formulae, we shall in the sequel often write $\operatorname{cov}(x, y)$ even when $x=y$, instead of $\operatorname{var} x$.

$$F(\lambda_i; \alpha_{10}, \ldots, \alpha_{k0}) = \frac{i}{n+1}, \quad (i = 0, 1, \ldots, n+1), \tag{2.2.2}$$

and Δ_i by

$$\Delta_i = F(\lambda_{i+1}; \alpha_1, \ldots, \alpha_k) - F(\lambda_i; \alpha_1, \ldots, \alpha_k), \quad (i = 0, 1, \ldots, n). \tag{2.2.3}$$

The sample-space is divided into $(n+1)^n$ parts exactly as in 1.2, and the probability mass P_ν is defined as before. Note that Δ_i and P_ν are functions both of α_{r0} and α_r, $(r = 1, \ldots, k)$.

Let us designate by $\alpha^*_{r\nu}$ the value of the estimate α^*_r in some arbitrarily chosen point in the νth part. Since each estimate α^*_r is assumed to be an asymptotically unbiased estimate of the corresponding parameter, we obtain by analogy to (1.2.5)

$$\sum_\nu \alpha^*_{r\nu} P_\nu = \alpha_r + R^{(1)}_{rn}, \quad (r = 1, \ldots, k), \tag{2.2.4}$$

where $R^{(1)}_{rn}$ tends to zero when n tends to infinity.

Further, we write by analogy to (1.2.6)

$$\sum_\nu (\alpha^*_{r\nu} - \alpha_r)(\alpha^*_{s\nu} - \alpha_s) P_\nu = \operatorname{cov}(\alpha^*_r, \alpha^*_s) + R^{(2)}_{rsn}, \quad (r, s = 1, \ldots, k). \tag{2.2.5}$$

Both terms in the right member generally tend to zero when n tends to infinity. We shall, however, introduce a more stringent condition. Denoting by W the $k \cdot k$ matrix which has as its elements the expressions in the left member of (2.2.5), we shall assume that W and V are asymptotically equivalent (cf. p. 36).

Moreover, let us assume that the relations (2.2.4) may be differentiated term by term with respect to any parameter α_s. Observing that

$$\sum_\nu \frac{\partial P_\nu}{\partial \alpha_s} = 0, \quad (s = 1, \ldots, k),$$

we then find

$$\sum_\nu (\alpha^*_{r\nu} - \alpha_r) \frac{\partial P_\nu}{\partial \alpha_s} = \begin{cases} 1 + \dfrac{\partial R^{(1)}_{rn}}{\partial \alpha_r} & \text{for } r = s \\[2ex] \dfrac{\partial R^{(1)}_{rn}}{\partial \alpha_s} & \text{for } r \neq s. \end{cases} \tag{2.2.6}$$

Now consider the quadratic form

$$\sum_\nu \left[(\alpha^*_{1\nu} - \alpha_1) t_1 + \cdots + (\alpha^*_{k\nu} - \alpha_k) t_k + \frac{1}{P_\nu} \frac{\partial P_\nu}{\partial \alpha_1} t_{k+1} + \cdots + \frac{1}{P_\nu} \frac{\partial P_\nu}{\partial \alpha_k} t_{2k} \right]^2 P_\nu,$$

where the summation extends over all parts of the sample-space for which $P_\nu \neq 0$. Since this form is positive definite or semi-definite, the same is by (2.2.6) true of the $2k \cdot 2k$ matrix

$$\left\{\begin{array}{c|c} W & R \\ \hline R' & J \end{array}\right\},$$

where R is the $k \cdot k$ matrix with the expressions in the right member of (2.2.6) as elements, R' the transpose of R, and J the $k \cdot k$ matrix

$$J = \left\{\sum_\nu \frac{1}{P_\nu} \frac{\partial P_\nu}{\partial \alpha_r} \frac{\partial P_\nu}{\partial \alpha_s}\right\}.$$

By analogy to (1.2.8), we find after some calculation that J can be given the more convenient form

$$J = \left\{n \sum_i \frac{1}{\Delta_i} \frac{\partial \Delta_i}{\partial \alpha_r} \frac{\partial \Delta_i}{\partial \alpha_s}\right\}, \quad (r, s = 1, \ldots, k), \tag{2.2.7}$$

summing over all $i = 0, 1, \ldots, n$ for which $\Delta_i \neq 0$.

In the remaining part of our investigation we shall turn off from the path followed in 1.2 and make use of a device due to Rao (1947 or 1952, p. 144 ff.). Suppose that the reciprocal J^{-1} of J exists, and let δ_k denote the unit $k \cdot k$ matrix. The product[1]

$$\left|\begin{array}{c|c} \delta_k & -RJ^{-1} \\ \hline 0 & J^{-1} \end{array}\right| \cdot \left|\begin{array}{c|c} W & R \\ \hline R' & J \end{array}\right| = \left|\begin{array}{c|c} W - RJ^{-1}R' & 0 \\ \hline J^{-1}R' & \delta_k \end{array}\right|$$

is evidently non-negative. Hence

$$|W - RJ^{-1}R'| \geqslant 0.$$

We can repeat this argument for any subset of the estimates $\alpha_1^*, \ldots, \alpha_k^*$, and thus we infer that the matrix

$$W - RJ^{-1}R'$$

is positive definite or semi-definite.

Now let n approach infinity. Then R tends to δ_k and, as stated earlier, W is asymptotically equivalent to V. Consequently, the matrix

$$V - J^{-1}$$

is positive definite or semi-definite if n is large enough.

---

[1] We denote here and in the sequel by $|A|$ the determinant corresponding to the matrix A.

We now take $\alpha_r = \alpha_{r0}$, $(r = 1, \ldots, k)$. Let us add a subscript 0 to V and J in order to denote the values of the matrices for these particular values of the parameters. Since by (2.2.2) and (2.2.3) we then have $\Delta_i = 1/(n+1)$, we see that J_0 is equivalent to the value I_0 which the matrix

$$I = \left\{ n^2 \sum_{i=0}^{n} \frac{\partial \Delta_i}{\partial \alpha_r} \frac{\partial \Delta_i}{\partial \alpha_s} \right\}, \quad (r, s = 1, \ldots, k),$$

assumes for $\alpha_r = \alpha_{r0}$.

Denoting the reciprocal of I_0 by I_0^{-1}, we conclude from what has just been said that the matrix

$$V_0 - I_0^{-1}$$

is positive definite or semi-definite for large n. Since α_{r0} is an arbitrary point in Ω_r, the result holds true for any values of the parameters. Thus we have proved the following theorem.

THEOREM. *Let $F(z; \alpha_1, \ldots, \alpha_k)$ be a continuous cdf. with the inverse function $G(u; \alpha_1, \ldots, \alpha_k)$ and α_r^*, $(r = 1, \ldots, k)$, asymptotically unbiased estimates of the unknown parameters α_r, based upon a random sample of n observations. Let, further, V denote the variance–covariance matrix of the estimates and I the $k \cdot k$ matrix with the elements*

$$I_{rs} = n^2 \sum_{i=0}^{n} \left(\frac{\partial F(\lambda_{i+1})}{\partial \alpha_r} - \frac{\partial F(\lambda_i)}{\partial \alpha_r} \right) \left(\frac{\partial F(\lambda_{i+1})}{\partial \alpha_s} - \frac{\partial F(\lambda_i)}{\partial \alpha_s} \right), \tag{2.2.8}$$

where $\dfrac{\partial F(\lambda_i)}{\partial \alpha_r}$ is the value of the partial derivative $\dfrac{\partial F}{\partial \alpha_r}$ for

$$z = \lambda_i = G\left(\frac{i}{n+1}; \alpha_1, \ldots, \alpha_k \right). \tag{2.2.9}$$

If

(a) *the derivatives $\dfrac{\partial F}{\partial \alpha_r}$, $(r = 1, \ldots, k)$, exist for any α_r belonging to an open interval Ω_r and for any z,*

(b) *the series appearing in (2.2.4) may be differentiated term by term with respect to any parameter α_s, $(s = 1, \ldots, k)$,*

(c) *the matrix W defined by the left member of (2.2.5) is asymptotically equivalent to V,*

(d) *the reciprocal* I^{-1} *of* I *exists,*

then, for any $\alpha_r \in \Omega_r$, $(r=1, \ldots, k)$, *and for* n *large enough, the matrix*

$$V - I^{-1}$$

is positive definite or semi-definite.

Denote the elements of the matrix I^{-1} by I^{rs}. The theorem implies that we have, asymptotically,

$$\text{var}\, \alpha_r^* \geqslant I^{rr}, \quad (r=1, \ldots, k), \tag{2.2.10}$$

which may be regarded as a generalization of (1.2.11).

Further, when n is large enough, the generalized variance $|V|$ of the estimates satisfies the inequality

$$|V| \geqslant |I^{-1}|. \tag{2.2.11}$$

If the equality sign holds in this formula, we shall call the estimates *asymptotically jointly efficient.* It will ensue from the discussion in 2.3 that this definition coincides with that used by Cramér (1946 *a*, p. 494 ff.), whenever the latter is applicable.

When there are only two unknown parameters α_1 and α_2, the inequalities (2.2.10) reduce to

$$\text{var}\, \alpha_1^* \geqslant \frac{I_{22}}{|I|}, \qquad \text{var}\, \alpha_2^* \geqslant \frac{I_{11}}{|I|}, \tag{2.2.12}$$

where

$$|I| = I_{11} I_{22} - I_{12}^2. \tag{2.2.13}$$

Furthermore, (2.2.11) is specialized to

$$\text{var}\, \alpha_1^* \;\text{var}\, \alpha_2^* - [\text{cov}\,(\alpha_1^*, \alpha_2^*)]^2 \geqslant \frac{1}{|I|}\cdot \tag{2.2.14}$$

We shall demonstrate in Chapter 13 that asymptotically jointly efficient estimates of α_1 and α_2 generally exist in the important special case where α_1 and α_2 are identified as the location parameter μ and the scale parameter σ in a continuous cdf. $F[(z-\mu)/\sigma]$.

2.3. Extension of the classification

The classification of distributions with one unknown parameter performed in 1.3 can be extended so as to apply also to the more general situation considered in this chapter. Conditions C 1 and C 2 then assume a more general form. Set

$$e_{rs} = E\left(\frac{\partial \log f}{\partial \alpha_r}\,\frac{\partial \log f}{\partial \alpha_s}\right) = \int_{-\infty}^{\infty} \frac{\partial \log f}{\partial \alpha_r}\,\frac{\partial \log f}{\partial \alpha_s}\, f(z;\, \alpha_1, \ldots, \alpha_k)\, dz,$$

$$(r,\, s = 1, \ldots, k). \quad (2.3.1)$$

Let E be the $k \cdot k$ matrix with these quantities as elements and E^{-1} its reciprocal (if existent). Now consider any of the unknown parameters α_r. We have two conditions, the latter of which is the same for all parameters.

CONDITION C 1. *The function* $\frac{\partial F}{\partial \alpha_r}$ *is a continuous function of* z *in the open interval* $A < z < B$, *and tends to zero when* $z \to A$ *or* $z \to B$.

CONDITION C 2. *The quantities* e_{rs} *exist, and the determinant* $|E|$ *is greater than zero.*

Using these conditions, we are able to classify the cdf. into one of three types according to exactly the same scheme as that introduced in 1.3.[1]

The procedure of classification should be repeated for any other parameter, and thus should be performed k times in all. Generally, the cdf. belongs to the same type, regardless of which parameter is considered. It may, however, well happen that the classification leads to different results for different parameters. For instance, it is not uncommon that the partial derivative appearing in Condition C 1 is zero at the ends for certain of the parameters, but different from zero for others. For this reason, it is not sufficient to state that a cdf. belongs, for example, to Type 1. Instead, it should be stated that it belongs to Type 1 *with respect to the parameter* α_r.

It follows from these remarks that the classification of a cdf. depending upon several unknown parameters may be rather complicated, since many alternatives are possible. Of special interest is the case where the cdf. belongs to the same type with respect to all parameters. We shall say that the cdf. is then *uniformly of Type 1, 2, or 3*, or, alternatively, that we have a *uniform Type 1 distribution*, etc. Since Condition C 2 is the same for all parameters, Type 3 distributions are always

[1] We also use a similar modification of the rules as on p. 19: When Condition C 1 breaks down and $|E| = 0$, we classify the distribution as Type 2.

uniform. In the following two sections we shall make some general comments upon the properties of estimates based upon samples from uniform Type 1 and Type 2 distributions. For examples of other distributions, the reader is referred to Chapter 13.

2.4. Uniform Type 1 distributions

Suppose that the distribution belongs uniformly to Type 1. Condition C 1 is then satisfied for each parameter α_r, and Condition C 2 is fulfilled. We find after a simple calculation $n \cdot I_{rs} \rightarrow e_{rs}$, where e_{rs} is defined by (2.3.1). Thus we infer from the theorem in 2.2 that *the matrix*

$$V - \frac{1}{n} E^{-1}$$

is positive definite or semi-definite if n is large enough.

This result is equivalent to the large-sample version of a theorem proved by Cramér (1946 *b*; see also 1946 *a*, p. 490 ff.). In particular, we see that Cramér's definition of asymptotic joint efficiency agrees with the more general definition introduced in 2.2, p. 25. We notice that, when asymptotically jointly efficient estimates exist, they have variances and, generally, covariances of order n^{-1}.

Examples of uniform Type 1 distributions will be given in Chapter 13.

2.5. Uniform Type 2 distributions

We shall consider the interesting case of a distribution belonging uniformly to Type 2. Then each partial derivative $\frac{\partial F}{\partial \alpha_r}$, $(r = 1, \ldots, k)$, has at least one discontinuity point, which may be situated inside or at the ends of the range of variation of z.

Let us suppose that the discontinuity points are denumerable, and let us order all such points, corresponding to the different parameters, into a single sequence $w_1, \ldots, w_l$. The number l of these points may be finite or infinite. Denote the increase[1] of the function $\frac{\partial F}{\partial \alpha_r}$ in w_ν by $\Delta_{r\nu}$, and let Δ be the $k \cdot l$ matrix with these quantities as elements. Note that certain of the quantities $\Delta_{r\nu}$ may be zero, but, corresponding to each parameter α_r, there is at least one $\Delta_{r\nu} \neq 0$. When there is a single discontinuity point, Δ is a column vector with non-zero elements.

[1] The increase ('jump') of a function $g(x)$ for $x = w$ is defined as the difference $g(w+) - g(w-)$.

2.5

Let us now examine the expressions I_{rs} defined by (2.2.8). It is realized after some calculation that

$$I \sim n^2 \Delta \Delta' + n(A + E), \tag{2.5.1}$$

where the sign $\sim$ denotes asymptotic equivalence, and Δ' is the transpose of Δ. Further, A is a $k \cdot k$ matrix, the elements of which will not be given here (cf. 13.6).

Denote the $k \cdot k$ matrix $\Delta\Delta'$ by $S = \{S_{rs}\}$, and by $S^{-1} = \{S^{rs}\}$ its reciprocal (if existent). Now consider the theorem in 2.2. We must distinguish between two cases.

(a) *The matrix S is non-singular.*

The assertion made in the theorem is equivalent to stating that the matrix

$$V - \frac{1}{n^2} S^{-1}$$

is positive definite or semi-definite. In particular, (2.2.10) reduces to

$$\operatorname{var} \alpha_r^* \geqslant \frac{1}{n^2} S^{rr}, \quad (r, = 1, \ldots, k). \tag{2.5.2}$$

If asymptotically jointly efficient estimates exist, they have variances of order n^{-2}. We shall give an example of this situation.

Rectangular distribution.

Let z have two unknown parameters, viz. the mean μ and the range of variation σ, so that

$$F(z; \mu, \sigma) = \begin{cases} 0 & \text{for } z - \mu \leqslant -\frac{\sigma}{2} \\ \frac{z-\mu}{\sigma} + \frac{1}{2} & \text{for } |z - \mu| < \frac{\sigma}{2} \\ 1 & \text{for } z - \mu \geqslant \frac{\sigma}{2}. \end{cases}$$

$\frac{\partial F}{\partial \mu}$ and $\frac{\partial F}{\partial \sigma}$ are discontinuous at the ends of the range of variation, and we find

$$\Delta = \begin{Bmatrix} -\frac{1}{\sigma} & \frac{1}{\sigma} \\ \frac{1}{2\sigma} & \frac{1}{2\sigma} \end{Bmatrix}, \qquad S = \begin{Bmatrix} \frac{2}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^2} \end{Bmatrix}.$$

Denoting the estimates of μ and σ by μ^* and σ^*, we infer from (2.5.2) that

$$\text{var } \mu^* \geqslant \frac{\sigma^2}{2\,n^2}, \qquad \text{var } \sigma^* \geqslant \frac{2\,\sigma^2}{n^2}.$$

This result can, of course, also be obtained directly from the theorem in 2.2. The limits are attained, for example, by, respectively, the mean of the extreme values (cf. p. 18) and the range of the sample, which, consequently, are asymptotically jointly efficient. We shall return to this example several times in the sequel (cf. 9.2 and 13.7).

Further examples will be given in Chapter 13.

(b) *The matrix S is singular.*

This situation occurs, for example, when the derivatives $\dfrac{\partial F}{\partial \alpha_r}$ possess a single common discontinuity point.

When S is singular, the elements of I^{-1} have for the most part the order n^{-1}. Hence *the variances of the asymptotically efficient estimates generally converge as n^{-1}.* This result is very remarkable, as it differs radically from that obtained in (a). In fact, the situation has much more in common with that encountered in the case of Type 1 distributions.

We shall verify the truth of the above statement in the important special case where *S has the rank $k-1$.*

As seen from (2.5.1), the cofactor of any element I_{rs} is, apart from a factor n^{2k-2}, asymptotically equivalent to the cofactor of S_{rs} in the matrix S provided, however, that $S_{rs} \neq 0$. Since the rank of S is $k-1$, all these cofactors are not zero. Moreover, the determinant $|I|$ is, apart from a factor n^{2k-1}, asymptotically equivalent to the sum Σ of the k determinants obtained by replacing in the determinant $|S|$ one column of elements by the corresponding elements in $|A+E|$. Thus, if $\Sigma \neq 0$ one at least of the elements of I^{-1} is of order $n^{2k-2}/n^{2k-1} = n^{-1}$. Generally, all the elements have this order.

The case $\Sigma = 0$ is exceptional and will not be discussed here.

Particularly simple results are obtained when there are only two unknown parameters α_1 and α_2 and the corresponding partial derivatives have a single common discontinuity. Then the rank of S is $k-1=1$, and the above results are applicable. We shall discuss this situation more fully in 13.6 in the special case where α_1 and α_2 are location and scale parameters.

Returning to the general case of k unknown parameters, we shall, finally, mention that the above discussion may without difficulty be extended also to the situation when the rank of S is less than $k-1$. Since nothing essentially new occurs in this case, we shall not consider this alternative.

PART II

TRANSFORMED BETA-VARIABLES
MOMENTS AND PROBABILITY DISTRIBUTIONS

CHAPTER 3

GENERAL INTRODUCTION

3.1. Definition of a transformed beta-variable

Take n points at random in the unit interval $0 \leqslant u \leqslant 1$. Denote by

$$u_{1n}, u_{2n}, \ldots, u_{nn}$$

the distances of the points from the origin in ascending order of magnitude. In the language of statistics, u_{in} may be termed the ith order statistic in a sample of n independent observations of a rectangularly distributed variable η. It is a random variable, which is distributed according to the beta-distribution (see further 4.2).

Let $x = G(u)$ be a B-measurable function of u, defined over the interval $0 \leqslant u \leqslant 1$, and let $F(x) = G^{-1}(x)$ be its inverse. Any such function $G(u)$ will be called a *transform*. Let $u_{1n}, u_{2n}, \ldots, u_{nn}$ be transformed to

$$x_{1n}, x_{2n}, \ldots, x_{nn},$$

where $x_{in} = G(u_{in})$. The one-dimensional random variable x_{in} will be called a *transformed beta-variable* or, in abbreviated form, a *TRB-variable*. The n-dimensional variable $(x_{1n}, x_{2n}, \ldots, x_{nn})$ will be termed a *set of transformed beta-variables*. For simplicity, we shall mostly write u_i and x_i instead of u_{in} and x_{in}.

In applications, $G(u)$ is generally non-decreasing; then $F(x)$ is uniquely determined.[1] Since this condition is not essential for the validity of all results in Part II, it will not be introduced at this stage.

The random variable $G(\eta)$ will be denoted by ξ. When $G(u)$ is non-decreasing, ξ has the cdf. $F(x)$. Of special importance is the case when the probability mass corresponding to ξ is zero in every point, i.e. when $F(x)$ is everywhere continuous; then $G(u)$ is an increasing function of u.

[1] If $G(u)$ is constant between two points of increase, set $F(x) =$ the right point.

3.2. Place of transformed beta-variables in statistical theory

Two applications of transformed beta-variables will be mentioned. The first of these is of great importance in statistics. In both cases $G(u)$ is assumed to be non-decreasing.

First, let

$$\xi_1 \leqslant \xi_2 \leqslant \cdots \leqslant \xi_n$$

be an ordered random sample (set of order statistics) of n independent observations of ξ. It is known (cf., e.g., Scheffé & Tukey, 1945) that the variable $(\xi_1, \xi_2, \ldots, \xi_n)$ has the same distribution as $(x_1, x_2, \ldots, x_n)$. In particular, any order statistic ξ_i is distributed as a *TRB*-variable x_i. (Note, however, that, as pointed out by Scheffé & Tukey, the variable $F(\xi_i)$ is not distributed in the same way as $F(x_i) = u_i$, unless $F(x)$ is continuous.) Thus, when $G(u)$ is non-decreasing, the abstract notion of a set of *TRB*-variables can be materialized as a set of order statistics. An excellent review of the huge literature concerning order statistics can be obtained from the lists of references given by Wilks (1948) and David & Johnson (1956). Some further references are given at the end of this thesis.

The second application of *TRB*-variables concerns the problem of transforming binomial and related distributions. Denote generally by $P(E)$ the probability that the event E will occur. Now the event $\nu \geqslant i$ is equivalent to the event $u_i \leqslant p$, where ν is a binomial variable with parameters n and p. Hence

$$P(\nu \geqslant i) = P(u_i \leqslant p). \tag{3.2.1}$$

Since the distance u_i is beta-distributed, this formula expresses the well-known relation between the binomial distribution and the beta-distribution (cf., e.g., Deming, 1950, p. 480 ff.).

It is sometimes of importance to transform the binomial variable to a new variable whose distribution is easier to handle. Using the transform $x = G(u)$, we obtain from (3.2.1) the relation

$$P(\nu \geqslant i) = P(x_i \leqslant G(p)).$$

Thus, computing the binomial sum is equivalent to determining the value of the cdf. of x_i in the point $G(p)$. If x_i is approximately normally distributed, this procedure is particularly convenient. The transform $G(u)$ ought therefore to be chosen such that the cdf. of x_i is as close to the normal cdf. as possible. In a previous paper (Blom, 1954) the author has discussed this problem. He demonstrated that B-transforms (cf. 3.5) are particularly useful for the purpose of normalizing binomial, negative binomial, Poisson, and χ^2 variables.

A fundamental difference between the two above-mentioned applications of TRB-variables should be emphasized. In the former of these, $G(u)$ is given in advance, being the inverse of the cdf. of the sampled population. For, if $u=F(x)$, then $x=F^{-1}(u)=G(u)$. In the latter case, on the contrary, the transform has to be chosen in such a way that the resulting TRB-variable will possess certain desired properties.

As a result of this brief account of the place of TRB-variables in statistics, we infer that these variables occur in two fields, which may at first sight seem to have very little in common (a relationship which, to the best of the author's knowledge, has not been pointed out before). In the present work, the theory of TRB-variables will be applied only to the former of the two fields.

3.3. Summary of Parts II and III

The contents of Part II and Part III may be summarized as follows.

In Part II, which consists of Chapters 3–8, some general properties of TRB-variables are studied by means of the theory of probability.

In Chapter 4, a review is given of several properties of such variables, which are valid for any sample-size. Chapters 5 and 6 contain a detailed discussion of the problem of finding approximations to the two first moments of non-singular TRB-variables.[1] In Chapter 6, a special tool is designed for improving the approximations given in Chapter 5, namely the so-called α, β-correction. Chapters 5 and 6 also contain a study of a special type of weighted difference between consecutive TRB-variables. The use of such differences results in a substantial simplification of the derivations made in subsequent chapters.

In Chapter 7, some properties of singular[1] TRB-variables are derived, which are related to Fisher & Tippett's results concerning the extreme values in a sample.

In Chapter 8, the results obtained in Chapters 4–7 are applied to the main problem in Part II, which concerns an analysis of linear combinations of TRB-variables. Several asymptotical results concerning moments and distributions of such combinations are given. The results depend largely upon the properties of the transform, and the investigation

[1] For the definition of non-singular and singular TRB-variables, the reader is referred to 3.4.

concerns three general families of transforms, which cover most cases occurring in practical situations.

In Part III, which consists of Chapters 9–14, the theoretical results deduced in Part II are applied to the problem of estimating location and scale parameters by means of linear combinations of ordered sample-values (linear estimates).

Chapter 9 has a preliminary character. In Chapter 10, the minimum variance problem for linear estimates is studied, i.e. the problem of constructing an estimate with the least possible variance. The exact solution is known but is for the most part very tedious to apply in practice. Instead, an approximative method is introduced, by which so-called nearly best linear estimates can be constructed.

In Chapter 11, some remarks are made concerning the problem arising when, in the above-mentioned problem, the variance is replaced by the mean square deviation about the true value of the parameter.

A modification of the method presented in Chapter 10 is discussed in Chapter 12, which also contains an application of nearly best linear estimates to probability papers.

In Chapter 13, asymptotic properties of nearly best linear estimates are subject to investigation. It is proved that, given certain general conditions, such estimates are asymptotically efficient (in the general sense used in Part I). Furthermore, the estimates are divided into three types, viz. Type 1, Type 2, and Type 3 estimates according to the classification of distributions introduced in Part I. A detailed study is made of each type of estimate, and examples are given illustrating the general theory expounded in Part I. The asymptotic distribution of a nearly best estimate is also investigated, and it is proved that Type 1 estimates (but not generally Type 2 estimates) are asymptotically normally distributed.

In Chapter 14, nearly best linear estimates are compared with estimates obtained by some other methods.

The author has aimed at proving the results given in Part II and Part III for as general classes of transforms as possible. The detailed assumptions concerning the transform are stated in any special problem. Nevertheless, to facilitate the reading, we shall give a survey of the two parts from this point of view.

Generally speaking, we pass from more general to more special results. The relations given in Chapter 4 apply to any transform or, in certain situations, to non-decreasing and/or differentiable transforms.

In Chapters 5 and 6, the transform is for the most part assumed to be continuous and to have continuous derivatives of low order in some small portion of the unit interval. In Chapter 7, we study transforms which either tend to infinity in either of two specified ways, or which have a first derivative with the same property.

From Chapter 8 and onwards, the transform is assumed to be bounded and differentiable over the closed interval $0 \leqslant u \leqslant 1$, or sometimes, more generally, over the open interval $0 < u < 1$. (An exception is afforded by the general formulae in 8.1–8.6 containing the coefficients g_i explicitly, which apply to any transform.) When asymptotical problems are studied, some additional conditions are introduced.

In Part III we shall suppose that, besides the conditions just stated, the transform is an increasing function of u.

For the benefit of the reader who is interested in particular distributions, we shall, finally, present a list of the distributions discussed in Part II and Part III.

Distribution	Section
Cauchy	4.3, 6.12
Exponential	4.5, 6.15, 10.8, 13.7
Extreme-value	6.13, 10.8, 13.5
Laplace (double exponential)	6.11, 10.8, 13.5
Normal	6.10, 7.3, 7.4, 10.8, 12.4, 12.5, 12.6, 13.5
Rectangular	4.2, 9.2, 10.8, 11.2, 12.6, 13.7
Right triangular	10.8, 13.8
Triangular	10.8, 13.8
Weibull	6.14, 12.4, 13.5

3.4. Notations

Besides the symbols introduced in 3.1, we shall need several other terms and notations connected with *TRB*-variables. Before presenting these, we shall quote a few general symbols and expressions, most of which are commonly used.

By M we shall designate a general positive fixed quantity.

By $x_n = O(\psi(n))$ we denote that the quotient $x_n/\psi(n)$ is bounded when $n \to \infty$. Then x_n is said to be *at most of the order* $\psi(n)$.

In particular, if $x_n/\psi(n)$ tends to zero when $n \to \infty$, we shall say that x_n has *a smaller order than* $\psi(n)$; we write this $x_n = o(\psi(n))$.

On the other hand, if $|x_n/\psi(n)|$ is bounded and greater than zero when $n \to \infty$, we shall say that x_n has *the same order as* $\psi(n)$. If the quotient x_n/y_n of two quantities x_n and y_n, which depend upon n, tends to unity when $n \to \infty$, we shall call x_n and y_n *asymptotically equivalent*. We write this $x_n \sim y_n$. Two matrices A_n and B_n of the same size, which depend on n and are positive definite for any finite n, are said to be asymptotically equivalent if any two corresponding principal minors are asymptotically equivalent. We write this $A_n \sim B_n$.

Set $A = G(0)$ and $B = G(1)$. When $G(u)$ is non-decreasing, (A, B) is the range of variation of ξ. A and B may be finite or infinite. The derivatives of $G(u)$ (when existent) are denoted by $G'(u)$, $G''(u)$, . . .,. The first derivative of $F(x)$ is designated by $f(x)$ and the higher derivatives by $F''(x) = f'(x)$, and so on. Since $F(x)$ is the inverse of $G(u)$, we have $f(x) = [G'(u)]^{-1}$. When $G(u)$ is non-decreasing, $f(x)$ (if existent) is the fr.f. of ξ. In many applications $f(x)$ is a symmetrical function. This situation will be referred to as the symmetrical case.

The index i of the *TRB*-variable x_{in} will be called the rank of the variable. *TRB*-variables with ranks i and $i+1$ will be termed *consecutive.*

We shall often consider sequences $\{x_{i_\nu n_\nu}\}$, $(\nu = 1, 2, \ldots)$, of *TRB*-variables, where $n_\nu < n_{\nu+1}$ and $i_\nu / n_\nu \to c$ when $\nu \to \infty$. It will be assumed that the same transform $G(u)$ is used in the definition of all variables. c is a constant which satisfies $0 \leqslant c \leqslant 1$. A sequence defined in this way will be called a *c-sequence of TRB-variables.*

Some particular types of c-sequence will be considered. A c-sequence will be called *at least $n^{-\frac{1}{2}}$-convergent,* if

$$\frac{i_\nu}{n_\nu} - c = O(n_\nu^{-\frac{1}{2}}).$$

On the other hand, if positive quantities M and ϱ can be found such that for all but a finite number of ν's

$$\left| \frac{i_\nu}{n_\nu} - c \right| > M n_\nu^{-\frac{1}{2}+\varrho},$$

then the c-sequence will be called *less than $n^{-\frac{1}{2}}$-convergent.*

Further, we shall call the c-sequence *singular,* if (a) $G(u)$ or $G'(u)$ is unbounded for $u = c$, and (b) the c-sequence is at least $n^{-\frac{1}{2}}$-convergent. All other sequences will be called *non-singular.*

If, in the definition of a c-sequence, we take $i_\nu = i$ or $i_\nu = n - i + 1$, we obtain a special type of sequence of great importance. In other words, we consider a *TRB*-variable of a given rank, counted from the left or from the right, and allow n to grow to infinity. In this case c obviously equals zero or unity. Such sequences will be termed i-sequences of *TRB*-variables. In accordance with the definition given above, an i-sequence may be singular or non-singular. Singular i-sequences occur, for example, when A or B is infinite, and are therefore common in application.

3.5. Classes of transforms

We shall consider classes of transforms $G(u)$ of different generality.

(1) *H-bounded transforms.*

Let $H(u; a, b)$ be defined by

$$H(u; a, b) = u^a (1-u)^b, \tag{3.5.1}$$

where a and b are two constants. We shall often write $H(u)$ instead of $H(u; a, b)$.

A transform $G(u)$ will be called *H-bounded,* if we can find quantities a and b such that the product $G(u)H(u)$ is bounded for $0 \leqslant u \leqslant 1$.

The following two definitions concern the behaviour of the transform near $u=0$ and $u=1$.

(2) *AL-transforms.*

$G(u)$ is an *AL-transform*[1] at $u=0$ or $u=1$, if

$$G(u) = -c_0 \left(\log \frac{1}{u}\right)^k \left[1 + O\left(\left[\log \frac{1}{u}\right]^{-1}\right)\right]$$

or

$$G(u) = c_0 \left(\log \frac{1}{1-u}\right)^k \left[1 + O\left(\left[\log \frac{1}{1-u}\right]^{-1}\right)\right] \tag{3.5.2}$$

respectively, where $k>0$ and $c_0>0$.

If $G(u)$ is an AL-transform at both ends, it is obviously H-bounded with $a, b>0$. The inverse functions of many common cdf:s are AL-transforms. The most important example is the inverse of the normal cdf. A simple calculation shows that in this case $c_0 = \sqrt{2}$, $k=\frac{1}{2}$. The inverses of the gamma distribution, the exponential distribution, and Laplace's distribution are other examples of AL-transforms.

(3) *AP-transforms.*

$G(u)$ is an *AP-transform*[1] at $u=0$ or $u=1$, if (apart, possibly, from an additive constant)

$$G(u) = -c_0 u^k [1 + O(u)]$$

or

$$G(u) = c_0 (1-u)^k [1 + O(1-u)] \tag{3.5.3}$$

respectively, where $k<0$ or $0<k<1$, and $c_0>0$.

[1] The abbreviations AL- and AP-transform are suggested by the first letters in the words Asymptotic, Logarithmic, and Power.

If $G(u)$ is an AP-transform at both ends with the same value of k, it is H-bounded with $a=b\geqslant -k$. The inverse of Cauchy's distribution affords an example of such a transform, k being in this case equal to -1.

When necessary for the validity of the results concerning AL- and AP-transforms given in the sequel, we shall assume that the defining relations may be differentiated a suitable number of times.

(4) *B-transforms.*

A *B-transform* (beta-transform) is defined as a solution of the differential equation

$$G'(u)=c_0 u^{-\tau_1}(1-u)^{-\tau_2}, \quad (0\leqslant\tau_1\leqslant 1;\ 0\leqslant\tau_2\leqslant 1). \tag{3.5.4}$$

B-transforms are either AL- or AP-transforms (except for $\tau_1=\tau_2=0$).

It is convenient to divide the family of B-transforms into two subclasses: B_0-transforms with $\tau_1=0$ or $\tau_2=0$ (or both) and B_1-transforms with $\tau_1, \tau_2\neq 0$. The class of B_0-transforms has three members, viz.

(a) *Rectangular case.*

$$\tau_1=\tau_2=0,\ G(u)=u,\ F(x)=x,\ (0\leqslant x\leqslant 1).$$

(b) *Exponential case.*

$$\tau_1=0,\ \tau_2=1,\ G(u)=\log\frac{1}{1-u},\ F(x)=1-e^{-x},\ (x\geqslant 0).$$

(c) *Generalized geometric case.*

$$0<\tau_1<1,\ \tau_2=0,\ G(u)=u^{1-\tau_1},\ F(x)=x^{1/(1-\tau_1)},\ (0\leqslant x\leqslant 1).$$

It should be observed that each of these cases contains in reality a family of distributions, obtained by shifting the location and the scale of the cdf. given above as an example. Further, τ_1 and τ_2 may be interchanged.

Among the B_1-transforms we may mention in passing the inverse sine transformation $2\arcsin\sqrt{u}$, obtained for $\tau_1=\tau_2=\frac{1}{2}$, and the logarithmic transformation $\log[u/(1-u)]$, obtained for $\tau_1=\tau_2=1$. Both these functions may be used for solving the transformation problem for binomial variables outlined in 3.2 (cf. Blom, 1954).

For easy reference we shall also give two general sets of conditions concerning the transform, which will play an important rôle in various parts of the thesis.

Condition A. *In a given set of points $\mathfrak{C}$ in the interval $0\leqslant u\leqslant 1$ the transform $G(u)$ and its first two derivatives are bounded and continuous, and its third derivative is bounded.*

Condition B. *In a given set of points* $\mathfrak{C}$ *in the interval* $0 \leqslant u \leqslant 1$ *the transform* $G(u)$ *and its first four derivatives are bounded and continuous. Further,* $G'(u) \neq 0$ *in* $\mathfrak{C}$.

When the points $u=0$ and/or $u=1$ belong to $\mathfrak{C}$, continuity in these points should be interpreted as continuity to the right and continuity to the left, respectively.

CHAPTER 4

FUNDAMENTAL PROPERTIES OF TRANSFORMED BETA-VARIABLES

4.1. Introduction

In this chapter, a review is given of several fundamental properties of *TRB*-variables, the majority of which have been given before in the literature. Only exact relations will be considered, approximate relations being studied in the succeeding chapters.

We shall begin the review by considering certain properties of points chosen at random in the unit interval.

4.2. Points chosen at random in the unit interval

For a more detailed discussion of some of the basic formulae given in this section, the reader is referred to Wilks (1948).

Consider the set of n randomly chosen points $u_1, u_2, \ldots, u_n$ described in 3.1. This set is an n-dimensional random variable with the frequency element

$$n!\, d u_1 d u_2 \ldots d u_n, \quad (0 < u_1 < u_2 < \cdots < u_n < 1). \tag{4.2.1}$$

Any distance u_i is distributed as a beta-variable with the fr.f.

$$\beta_i(u; n) = \frac{n!}{(i-1)!\,(n-i)!}\, u^{i-1}(1-u)^{n-i}. \tag{4.2.2}$$

The mean of u_i will be denoted by p_i. We have

$$p_i = \frac{i}{n+1}. \tag{4.2.3}$$

The central moment $E(u_i - p_i)^r$ of order r will be denoted by μ_r. The variance of u_i is given by

$$\mu_2 = \text{var}\ u_i = \frac{p_i q_i}{n+2}, \tag{4.2.4}$$

where $q_i = 1 - p_i$.

The joint fr.f. of a pair of variables u_i and u_j is

$$\beta_{ij}(u, v; n) = \frac{n!}{(i-1)!\,(j-i-1)!\,(n-j)!} u^{i-1} (v-u)^{j-i-1} (1-v)^{n-j}, \quad (i<j), \tag{4.2.5}$$

where u and v correspond to u_i and u_j respectively.

The covariance of u_i and u_j is given by (cf. note on p. 21)

$$\operatorname{cov}(u_i, u_j) = \frac{p_i q_j}{n+2}, \quad (i \leqslant j). \tag{4.2.6}$$

We shall sometimes need more general central moments and denote any such moment $E[(u_i - p_i)^r (u_j - p_j)^s]$ by μ_{rs}. Explicit expressions for several μ_{rs} of low order have been given by David & Johnson (1954). It is interesting to observe that any μ_{rs} can be expressed in terms of two non-central moments of order $r+s-2$ by means of either of the relations

$$\mu_{rs} = \frac{(r-1)\,p_i q_i}{n+2} E[(u_{i+1\,n+2} - p_i)^{r-2} (u_{j+2\,n+2} - p_j)^s] +$$

$$+ \frac{s\,p_i q_j}{n+2} E[(u_{i\,n+2} - p_i)^{r-1} (u_{j+1\,n+2} - p_j)^{s-1}], \quad (r>0), \tag{4.2.6 a}$$

$$\mu_{rs} = \frac{(s-1)\,p_j q_j}{n+2} E[(u_{i\,n+2} - p_i)^r (u_{j+1\,n+2} - p_j)^{s-2}] +$$

$$+ \frac{r\,p_i q_j}{n+2} E[(u_{i+1\,n+2} - p_i)^{r-1} (u_{j+2\,n+2} - p_j)^{s-1}], \quad (s>0). \tag{4.2.6 b}$$

The formulae hold good, of course, also for single moments of u_i and u_j; then the second terms disappear. The relations, which seem not to have been given before in the literature, can be proved in various ways. For example, formula (4.2.6 a) can be proved by expanding μ_{rs} according to powers of u_i and $1 - u_j$ and using the auxiliary formula

$$E[u_{i\,n}^r (1-u_{j\,n})^s] - p_i E[u_{i\,n}^{r-1} (1-u_{j\,n})^s] =$$

$$= \frac{(r-1)\,p_i q_i}{n+2} E[u_{i+1\,n+2}^{r-2} (1-u_{j+2\,n+2})^s] -$$

$$- \frac{s\,p_i q_j}{n+2} E[u_{i\,n+2}^{r-1} (1-u_{j+1\,n+2})^{s-1}].$$

This relation may be verified e.g. by a straightforward evaluation of the moments involved.

By an inductive argument, the details of which we omit, we conclude from (4.2.6a) and (4.2.6b) that the moments μ_{rs} have the following general property. *For any non-negative integers r and s, it is possible to find a quantity M, which does not depend upon i, j, and n, such that*

$$|\mu_{rs}| < \frac{M}{n^k}, \tag{4.2.7}$$

where $k = (r+s)/2$ *if* $r+s$ *is even, and* $k = (r+s+1)/2$ *if* $r+s$ *is odd.*

Moreover, by application of Schwarz's inequality it follows from (4.2.7) that the absolute central moments $\gamma_{rs} = E[|u_i - p_i|^r \, |u_j - p_j|^s]$ satisfy a similar inequality, viz.

$$\gamma_{rs} < \frac{M}{n^{(r+s)/2}}. \tag{4.2.8}$$

Both inequalities hold, of course, also when $r = 0$ or $s = 0$; then μ_{rs} and γ_{rs} are specialized to moments of the single variables u_i and u_j.

It is sometimes of interest to consider the intervals

$$\delta_i = u_{i+1} - u_i, \quad (i = 0, 1, \ldots, n; \; u_0 = 0; \; u_{n+1} = 1) \tag{4.2.9}$$

between adjacent points. Any n of these $n+1$ intervals, say $\delta_1, \ldots, \delta_n$, constitute an n-dimensional random variable with the frequency element

$$n! \, d\delta_1 \, d\delta_2 \cdots d\delta_n, \qquad \left(\sum_{i=1}^{n} \delta_i < 1\right). \tag{4.2.10}$$

More generally, any k intervals, for example $\delta_1, \ldots, \delta_k$, have the joint fr.f.

$$n(n-1) \cdots (n-k+1) \, (1 - \delta_1 - \delta_2 - \cdots - \delta_k)^{n-k}. \tag{4.2.11}$$

It follows from (4.2.3), (4.2.4), (4.2.6), and the definition of δ_i that the mean and the variance of δ_i are given by

$$E\,\delta_i = \frac{1}{n+1}, \qquad \operatorname{var} \delta_i = \frac{n}{(n+1)^2 (n+2)}. \tag{4.2.12}$$

Further, it follows from (4.2.6) that

$$\operatorname{cov}(\delta_i, \delta_j) = -\frac{1}{(n+1)^2 (n+2)}, \quad (i \neq j). \tag{4.2.13}$$

4.3. Distribution of transformed beta-variables

In this section, $G(u)$ is assumed to be non-decreasing. However, the formulae (4.3.1) and (4.3.2) given below are valid even without this restriction.

The probability distribution of a set of *TRB*-variables can be obtained by applying the transformation $u_i = F(x_i)$, $(i = 1, \ldots, n)$, to (4.2.1). If, in particular, $f(x)$ exists and is continuous nearly everywhere, the frequency element of the set is found to be

$$n!\, f(x_1) \ldots f(x_n)\, d x_1 \ldots d x_n.$$

By the same transformation, the fr.f. of a single variable x_i or of a pair of variables x_i and x_j can be obtained from (4.2.2) and (4.2.5) respectively. The resulting distributions are not very simple to handle, and numerical integration has to be resorted to in most cases.

The mean of x_i is by (4.2.2) and the definition of a *TRB*-variable

$$E\, x_i = \int_0^1 G(u)\, \beta_i(u; n)\, d u. \tag{4.3.1}$$

The variance can be calculated from the usual expression

$$\operatorname{var} x_i = E\, x_i^2 - (E\, x_i)^2,$$

where

$$E\, x_i^2 = \int_0^1 [G(u)]^2\, \beta_i(u; n)\, d u.$$

For the covariance of a pair of variables $x_i, x_j, (i < j)$, we have

$$\operatorname{cov}(x_i, x_j) = E\, x_i x_j - (E\, x_i)(E\, x_j),$$

where

$$E\, x_i x_j = \iint\limits_{0<u<v<1} G(u)\, G(v)\, \beta_{ij}(u, v; n)\, d u\, d v \tag{4.3.2}$$

with $\beta_{ij}(u, v; n)$ given by (4.2.5).

The integrals which appear in these formulae do not always exist. We shall give two conditions of different generality, by which the question of finiteness can be decided.

First, we observe that the mean of x_i is finite for any i if $E|\xi|$ is finite. For by (4.2.2) we then have

$$|E\, x_i| < M \int^1 |G(u)|\, d u = M E |\xi| < \infty.$$

Similarly, it is shown that the variances and covariances are finite for any i and j if $E\,\xi^2$ is finite.

Secondly, assume that $G(u)$ is H-bounded (cf. 3.5). The mean of x_i is then finite if

$$a < i < n - b + 1.$$

Further, the variances and covariances are finite if

$$2a < i \leqslant j < n - 2b + 1.$$

If, for example, ξ is a Cauchy variable, we can take $a = b = 1$. Thus the means are finite if $2 \leqslant i \leqslant n-1$, and the variances and covariances if $3 \leqslant i \leqslant j \leqslant n-2$.

4.4. Differences between consecutive transformed beta-variables

Throughout this section, we shall suppose that $G(u)$ is H-bounded and differentiable. We shall generalize (4.2.9) by defining δ_i as

$$\delta_i = x_{i+1} - x_i, \quad (i = 0, 1, \ldots, n;\ x_0 = A;\ x_{n+1} = B).$$

When $G(u)$ is non-decreasing, the differences δ_i are the lengths of the intervals, into which the range of variation (A, B) of ξ is divided by the points $x_1, \ldots, x_n$. We shall give expressions for the first two moments of the variables δ_i. When

$$a < i < n - b,$$

we have from (4.3.1) by a partial integration

$$E\,\delta_i = \frac{1}{n+1}\, E\,G'(u), \tag{4.4.1}$$

where

$$E\,G'(u) = \int_0^1 G'(u)\,\beta_{i+1}(u;\, n+1)\,du. \tag{4.4.2}$$

Here a and b are the constants which appear in the definition of a H-bounded transform (cf. 3.5). Moreover, when

$$a < i \leqslant j < n - b$$

we find by a series of partial integrations

$$E\,\delta_i\,\delta_j = \frac{\lambda}{(n+1)\,(n+2)}\, E\,G'(u)\,G'(v), \tag{4.4.3}$$

where

$$E\,G'(u)\,G'(v) = \iint_{0<u<v<1} G'(u)\,G'(v)\,\beta_{i+1\,j+2}(u, v;\, n+2)\,du\,dv, \tag{4.4.4}$$

and $\lambda=1$ when $i<j$, and $\lambda=2$ when $i=j$. Irwin (1925) has given a general expression for the mean $E\,\delta_i^k$ of the kth power of δ_i, which contains (4.4.1) and (4.4.3) (with $i=j$) as special cases.

We shall give an application of these expressions. If $G(u)$ is a B-transform (see 3.5), then (4.4.1) is specialized to[1]

$$E\,\delta_i=\frac{c_0}{n'+1}\cdot\frac{\binom{n}{i}}{\binom{n'}{i'}}, \tag{4.4.5}$$

where

$$i'=i-\tau_1,$$
$$n'=n-\tau_1-\tau_2.$$

Further, if $G(u)$ is a B_0-transform, the double integral appearing in (4.4.3) can be evaluated explicitly. Suppose, for instance, that $\tau_1=0$. Using the notation

$$k(x)=\frac{n!}{(n-2\,x+2)!\,(x!)^2},$$

we find after some calculation, which will not be reproduced here,

$$E\,\delta_i\,\delta_j=\frac{\lambda\,c_0^2\,k(\tau_2)}{\binom{n-j}{\tau_2}\binom{n-i-\tau_2+1}{\tau_2}}, \tag{4.4.6}$$

where $\lambda=1$ if $i<j$, and $\lambda=2$ if $i=j$. If $\tau_2=0$, the same expression holds with τ_2 replaced by τ_1, and i and j replaced by $n-j$ and $n-i$ respectively.

The interest of the formula (4.4.6) lies in the fact that, apart from the constants in the numerator, it is a product of two simple factors, which depend only upon i and j, respectively.

4.5. Exponential distribution

The distribution of the differences δ_i is for the most part difficult to handle. In one special case it is, however, quite simple. Suppose that $G(u)$ is the inverse of the exponential distribution, i.e.

$$G(u)=\log\frac{1}{1-u}.$$

[1] For any $x>-1$, we define $x!=\Gamma(x+1)$.

Then $$x_i = \log \frac{1}{1-u_i} = \delta_0 + \delta_1 + \cdots + \delta_{i-1}, \quad (i=1, \ldots, n). \tag{4.5.1}$$

Introduce the quantities δ_i, $(i=0, 1, \ldots, n-1)$, as variables in (4.2.1) instead of the variables u_i. The Jacobian of the transformation is

$$J = (1-u_1)(1-u_2) \cdots (1-u_n).$$

Hence the δ_i's are distributed according to the joint fr.f.

$$n!\, e^{-n\delta_0} \cdot e^{-(n-1)\delta_1} \cdots e^{-\delta_{n-1}}.$$

Thus these variables are independent and, furthermore, δ_i is exponentially distributed with the fr.f. $e^{-(n-i)\delta_i}$. The mean and the variance of δ_i are

$$E\,\delta_i = \frac{1}{n-i}, \quad \text{var}\ \delta_i = \frac{1}{(n-i)^2}. \tag{4.5.2}$$

These values can also be obtained by setting $\tau_1=0$, $\tau_2=1$ in (4.4.5) and (4.4.6). Using the latter formula, we can also verify that

$$\text{cov}\,(\delta_i, \delta_j) = 0,$$

as it should be. Returning to the variables x_i, we infer that

$$E\,x_i = \sum_{\nu=n-i+1}^{n} \frac{1}{\nu}, \quad \text{cov}\,(x_i, x_j) = \sum_{\nu=n-i+1}^{n} \frac{1}{\nu^2}, \quad (i \leqslant j) \tag{4.5.3}$$

Similar results are obtained when $G(u) = \log u$, the inverse of the negatively directed exponential distribution with the fr.f. $e^x, (x \leqq 0)$. The variable $-x_i$ can be written as a sum of $n-i+1$ exponentially distributed variables. Further,

$$E\,x_i = -\sum_{\nu=i}^{n} \frac{1}{\nu}, \quad \text{cov}\,(x_i, x_j) = \sum_{\nu=j}^{n} \frac{1}{\nu^2}, \quad (i \leqslant j). \tag{4.5.4}$$

Formulae (4.5.3) were proved by Pearson & Pearson (1932) by a complicated method. A simpler proof was given by Gumbel (1937), who used moment generating functions. A detailed discussion of the exponential case is found in a paper by Rényi (1953). Malmquist (1950) also obtained results which are closely connected with those reviewed in this section.

CHAPTER 5

NON-SINGULAR TRANSFORMED BETA-VARIABLES

5.1. Introduction

The asymptotic properties of non-singular *TRB*-variables have been studied by many authors. We shall review the most important work done in this field.[1]

First, we shall consider results which concern the behaviour of the first two moments as n tends to infinity. It follows from Hoeffding (1953, Lemma 5) that, given certain general conditions,

$$\lim E\, x_{in} = G(c), \quad (i/n \to c). \tag{5.1.1}$$

There is a corresponding result for the second moments, viz.

$$\lim_{n\to\infty} n \operatorname{cov}(x_{in}, x_{jn}) = c_1(1-c_2)\, G'(c_1)\, G'(c_2), \tag{5.1.2}$$
$$(i/n \to c_1;\ j/n \to c_2;\ c_1 \leqslant c_2).$$

If we take $c_1 = c_2$ and $i = j$, the formula also holds good for the variance of x_i. This relation has been used in various formulations by several authors (see, e.g., Pearson, 1920), but it is uncertain to whom it should be ascribed. More elaborate approximation formulae have been derived by Pearson & Pearson (1931) and, in the case of the sample median, by Chu & Hotelling (1955).

Secondly, some results concerning the asymptotic distribution of non-singular *TRB*-variables will be mentioned. Smirnoff (1935) has proved that, if x_i belongs to a c-sequence with $0 < c < 1$ and $G'(u)$ is bounded and continuous in $u = c$, then x_i is asymptotically normally distributed. He has also demonstrated that, if x_i belongs to a c_1-sequence and x_j to a c_2-sequence $(0 < c_1 < c_2 < 1)$, the variables x_i and x_j are asymptotically jointly normally distributed. This result has been extended to more than two *TRB*-variables by Mosteller (1946).

---

[1] It might be remarked that the work reviewed in the section concerns order statistics. It can, however, easily be extended to the more general case of *TRB*-variables.

More accurate results concerning the distribution can be obtained by means of Cornish–Fisher expansions (Cornish & Fisher, 1937) of the *TRB*-variables. Blom (1954) has derived such an expansion for a transformed binomial variable. This result is also valid for a *TRB*-variable because of the relationship pointed out in 3.2.

The contents of the present chapter can be summarized as follows. Sections 5.3 and 5.4 will be devoted to an investigation of two approximation formulae, which are closely related to (5.1.1) and (5.1.2). The proofs found in the literature of relations similar to those given here often lack rigour in certain respects. We have aimed at proving the results as completely as possible, using some general lemmas, proved in 5.2, which are believed to be of some value also in other problems. On the other hand, this makes the exposition rather long and detailed.

In 5.5 we shall study some properties of weighted differences between consecutive *TRB*-variables. Such differences will play an important rôle in the investigation of linear combinations performed in Chapter 8.

5.2. Three lemmas

We need three lemmas, which will be proved for the case of a function of a two-dimensional random variable, but which can be generalized to any number of dimensions.

Let $\{X_\nu\}$, $(\nu=1, 2, \ldots)$, be a finite or infinite sequence of two-dimensional random variables $X_\nu=(\xi_\nu, \eta_\nu)$. It will be assumed that the ranges of variation of the variables can be enclosed within a finite region Ω. Denote by $F_\nu(\xi, \eta)$ the cdf. of X_ν, by $\xi_{0\nu}$ and $\eta_{0\nu}$ the means of the components ξ_ν and η_ν respectively, by $\mu_{rs\nu}$ their central moment

$$E\,[(\xi_\nu-\xi_{0\nu})^r\,(\eta_\nu-\eta_{0\nu})^s],$$

and by $\gamma_{rs\nu}$ their absolute central moment

$$E\,[\,|\xi_\nu-\xi_{0\nu}|^r\,|\eta_\nu-\eta_{0\nu}|^s].$$

Set
$$\psi_{k\nu}=\sum_{r+s=k}\gamma_{rs\nu},$$

where the summation extends over all absolute central moments with $r+s=k$, $(r, s\geqslant 0)$.

Let, further, $g(\xi, \eta)$ be a real-valued function defined over Ω. Denote by $g^{(rs)}(\xi, \eta)$ the partial derivative obtained by differentiating the function r times with respect to ξ and s times with respect to η.

By an *ε-neighbourhood* of a sequence of points $\{\xi'_\nu, \eta'_\nu\}$, $(\nu = 1, 2, \ldots)$, we shall denote all points ξ, η in Ω which satisfy the inequalities

$$|\xi - \xi'_\nu| \leqslant \varepsilon \text{ and } |\eta - \eta'_\nu| \leqslant \varepsilon, \ (\nu = 1, 2, \ldots).$$

The first lemma concerns functions which are bounded in Ω.

LEMMA 1. *If*

(a) *the function* $g(\xi, \eta)$ *is bounded in* Ω,

(b) *a quantity* $\varepsilon > 0$ *can be found such that in the ε-neighbourhood of the sequence* $\{\xi_{0\nu}, \eta_{0\nu}\}$, $(\nu = 1, 2, \ldots)$, *the function* $g(\xi, \eta)$ *is continuous, and, further, the derivatives* $g^{(rs)}(\xi, \eta)$ *are continuous for all non-negative integers* r, s *such that* $r + s \leqslant k - 1$ *and bounded for all non-negative integers such that* $r + s = k$,

(c) *the integral*

$$I_\nu = \int_\Omega g(\xi, \eta)\, dF_\nu(\xi, \eta) \tag{5.2.1}$$

exists for any $\nu = 1, 2, \ldots$,

then
$$I_\nu = \sum_{r+s \leqslant k-1} \frac{1}{r!\, s!} \mu_{rs\nu}\, g^{(rs)}(\xi_{0\nu}, \eta_{0\nu}) + R_\nu, \tag{5.2.2}$$

where, for any positive integer $\varkappa$,

$$|R_\nu| < M(\psi_{k\nu} + \varepsilon^{-\varkappa} \psi_{\varkappa\nu}), \tag{5.2.3}$$

and M *does not depend upon* ν *and* ε. *If, in particular,* $\varkappa = k$, *then*

$$|R_\nu| < M_\varepsilon \psi_{k\nu}, \tag{5.2.4}$$

where M_ε *may depend upon* ε *but not upon* ν.

In order to prove this lemma, consider the integral I_ν for some fixed ν. For convenience, the index ν is dropped in F_ν, etc., in the course of the proof. We observe that, as Ω is finite, moments μ_{rs} and γ_{rs} of any order exist and are finite.

Divide Ω into two parts Ω_1 and Ω_2, where Ω_1 consists of the points in the square $|\xi - \xi_0| \leqslant \varepsilon$, $|\eta - \eta_0| \leqslant \varepsilon$, and Ω_2 of all other points in Ω. Then

$$I_\nu = \int_{\Omega_1} + \int_{\Omega_2} g\, dF = J_1 + J_2. \tag{5.2.5}$$

First, we shall estimate J_2. For any $\varepsilon > 0$ and any positive integer $\varkappa$, we have by Tchebycheff's inequality

$$P(|\xi - \xi_0| > \varepsilon) \leqslant \varepsilon^{-\varkappa} \gamma_{\varkappa 0}$$

and a corresponding relation for η. Hence

$$P(\Omega_2) \leqslant \varepsilon^{-\varkappa} (\gamma_{\varkappa 0} + \gamma_{0 \varkappa}) \leqslant \varepsilon^{-\varkappa} \psi_\varkappa. \tag{5.2.6}$$

As g is bounded,

$$|J_2| \leqslant M \varepsilon^{-\varkappa} \psi_\varkappa, \tag{5.2.7}$$

where M does not depend upon ν and ε.

Secondly, we turn our attention to J_1. According to condition (b) we have the following Taylor expansion for any ξ, η in Ω_1

$$g(\xi, \eta) = \sum_{r+s \leqslant k-1} \frac{1}{r!\, s!} g^{(r\, s)}(\xi_0, \eta_0) (\xi - \xi_0)^r (\eta - \eta_0)^s +$$

$$+ \sum_{r+s=k} \frac{1}{r!\, s!} g^{(r\, s)}(\xi', \eta') (\xi - \xi_0)^r (\eta - \eta_0)^s,$$

where the point ξ', η' also belongs to Ω_1. Integrating over Ω_1 and using the identity

$$\mu_{r s} = \int_{\Omega_1} + \int_{\Omega_2} (\xi - \xi_0)^r (\eta - \eta_0)^s \, dF,$$

we get

$$J_1 = \sum_{r+s \leqslant k-1} \frac{1}{r!\, s!} \mu_{r s}\, g^{(r\, s)}(\xi_0, \eta_0) - H_1 + H_2, \tag{5.2.8}$$

where

$$H_1 = \sum_{r+s \leqslant k-1} \frac{1}{r!\, s!} g^{(r\, s)}(\xi_0, \eta_0) \int_{\Omega_2} (\xi - \xi_0)^r (\eta - \eta_0)^s \, dF,$$

$$H_2 = \sum_{r+s=k} \frac{1}{r!\, s!} \int_{\Omega_1} g^{(r\, s)}(\xi', \eta') (\xi - \xi_0)^r (\eta - \eta_0)^s \, dF.$$

We now observe that, as ξ and η are bounded, the integrals appearing in H_1 are also bounded. Hence by condition (b) and (5.2.6)

$$|H_1| < M \varepsilon^{-\varkappa} \psi_\varkappa. \tag{5.2.9}$$

Furthermore, by condition (b)

$$|H_2| < M \psi_k, \tag{5.2.10}$$

where M does not depend upon ν.

Combining (5.2.7)–(5.2.10), we obtain (5.2.2) and (5.2.3), and the lemma is proved.

Note that it is essential for the proof that Ω is finite. If the components have an infinite range of variation, the lemma may, however, still be true. It holds good, for example, if condition (b) is satisfied for any point in Ω; in that case the term containing ε in (5.2.3) disappears.

In the second lemma we remove the condition that g shall be bounded everywhere in Ω. We then need the concept of a majorant of a sequence of random variables with respect to a function $g(\xi, \eta)$.

The sequence $\{X'_\nu\} = \{\xi'_\nu, \eta'_\nu\}$, $(\nu = 1, 2, \ldots)$, *of random variables will be called a majorant of the sequence* $\{X_\nu\}$ *with respect to* $g(\xi, \eta)$, *if for any Borel-set* S *contained in* Ω

$$\int_S |g(\xi, \eta)| \, dF_\nu < M' \int_S dF'_\nu, \quad (\nu = 1, 2, \ldots), \tag{5.2.11}$$

where M' *does not depend upon* ν *and* S. (Here and subsequently primed letters are used in order to denote the cdf. and other concepts connected with X'_ν.) In the particular case where X_ν and X'_ν have distributions of the continuous type, the condition is equivalent to requiring that

$$|g(\xi, \eta)| \, f_\nu(\xi, \eta) < M' f'_\nu(\xi, \eta), \tag{5.2.12}$$

where f_ν and f'_ν are the corresponding fr.f:s.

Lemma 2. *If the sequence of random variables* $\{X_\nu\}$, $(\nu = 1, 2, \ldots)$, *has a majorant* $\{X'_\nu\}$ *with respect to* $g(\xi, \eta)$, *and if, in addition, conditions* (b) *and* (c) *of Lemma 1 are fulfilled, then* (5.2.2) *is true with*

$$|R_\nu| < M(\psi_{k\nu} + \varepsilon^{-\varkappa} \psi_{\varkappa\nu}) + M' \varepsilon^{-\varkappa} \psi'_{\varkappa\nu}, \tag{5.2.13}$$

where $\varkappa$ *is an arbitrary positive integer, and* M *and* M' *do not depend upon* ν *and* ε. *If, in particular,* $\varkappa = k$, *then*

$$|R_\nu| < M_\varepsilon(\psi_{k\nu} + \psi'_{k\nu}), \tag{5.2.14}$$

where M_ε *may depend upon* ε *but not upon* ν.

The proof follows the same pattern as that of Lemma 1. We use (5.2.5) and begin by estimating J_2. From the definition of the majorant we conclude that, for any $F = F_\nu$,

$$|J_2| < M' \int_{\Omega_2} dF' = M' P'(\Omega_2).$$

By analogy to (5.2.6),

$$P'(\Omega_2) \leqslant \varepsilon^{-\varkappa} \psi'_{\varkappa},$$

whence

$$|J_2| \leqslant M' \varepsilon^{-\varkappa} \psi'_{\varkappa}. \tag{5.2.15}$$

J_1 is treated exactly as before. Combining (5.2.15) with the previously obtained relations (5.2.8), (5.2.9), and (5.2.10), we get (5.2.2) and (5.2.13), and the lemma is proved.

In the special case where g is a product of two functions depending only upon ξ and η respectively, each factor can be expanded separately in a Taylor series. Proceeding in other respects as before, we get

LEMMA 3. *If*

(a) *the sequence of random variables* $\{X_\nu\}$, $(\nu = 1, 2, \ldots)$, *has a majorant* $\{X'_\nu\}$ *with respect to the function*

$$[g_1(\xi) - g_1(\xi_{0\nu})] [g_2(\eta) - g_2(\eta_{0\nu})],$$

(b) *a quantity* $\varepsilon > 0$ *can be found such that, in the* ε*-neighbourhood of the sequence* $\{\xi_{0\nu}\}$, $(\nu = 1, 2, \ldots)$, *the function* $g_1(\xi)$ *and its* $k-1$ *first derivatives are bounded and continuous and, further, its kth derivative is bounded (and correspondingly for* $g_2(\eta)$*)*,

(c) *the integral*

$$I_\nu = \int_\Omega [g_1(\xi) - g_1(\xi_{0\nu})] [g_2(\eta) - g_2(\eta_{0\nu})] \, dF_\nu(\xi, \eta) \tag{5.2.16}$$

exists for any $\nu = 1, 2, \ldots$,

then

$$I_\nu = \sum_{1 \leqslant r, s \leqslant k-1} \frac{1}{r!\, s!} \mu_{rs\nu} \, g_1^{(r)}(\xi_{0\nu}) \, g_2^{(s)}(\eta_{0\nu}) + R_\nu, \tag{5.2.17}$$

where for any positive integer $\varkappa$

$$|R_\nu| < M \left(\sum \gamma_{rs\nu} + \varepsilon^{-\varkappa} \psi_{\varkappa\nu}\right) + M' \varepsilon^{-\varkappa} \psi'_{\varkappa\nu}. \tag{5.2.18}$$

Here $\sum$ *denotes summation for* $1 \leqslant r \leqslant k$, $s = k$ *and* $1 \leqslant s \leqslant k$, $r = k$.

5.3. An approximation formula for the mean of a transformed beta-variable

Consider the relation

$$E\, x_i = G\left(\frac{i}{n+1}\right) + R_i. \tag{5.3.1}$$

We shall later often use a special notation for the leading term in this relation, viz.

$$\lambda_i = G\left(\frac{i}{n+1}\right). \tag{5.3.2}$$

THEOREM. *If*

(a) *$G(u)$ is H-bounded,*[1]

(b) *the functions $G(u)$ and $G'(u)$ are bounded and continuous, and $G''(u)$ is bounded in $\mathfrak{C}$, where $\mathfrak{C}$ is equal to the interval $0 \leqslant u \leqslant 1$ with the possible exception of a finite number of points $\mathfrak{C}^*$,*

(c) *x_i belongs to any c-sequence when $c \in \mathfrak{C}$, and to a less than $n^{-\frac{1}{2}}$-convergent c-sequence when $c \in \mathfrak{C}^*$,*

then the error term in (5.3.1) *satisfies the inequality*

$$|R_i| < \frac{M}{n}, \tag{5.3.3}$$

where M does not depend upon i and n.

For the proof we use the one-dimensional version of Lemma 2. We take $\xi_\nu = u_i$ and $\xi'_\nu = u'_i$, where u'_i is distributed according to the fr.f. $\beta_{i-a}(u; n-a-b)$ (cf. 4.2). Further, take $g = G(u)$ and $k = 2$.

First, we shall prove that the sequence $\{u'_i\}$, $(i = 1, \ldots, n)$, is a majorant of the sequence $\{u_i\}$ with respect to $G(u)$. To this end, we observe that, by (4.2.2) and (3.5.1), the fr.f:s of u_i and u'_i satisfy the inequality

$$|G(u)|\, \beta_i(u; n) < Q_{in}\, \beta_{i-a}(u; n-a-b),$$

where
$$Q_{in} = M \frac{n!}{(i-1)!\,(n-i)!} \cdot \frac{(i-a-1)!\,(n-i-b)!}{(n-a-b)!}.$$

We can always assume that $a, b \geqslant 0$. Hence

$$Q_{in} \leqslant M \frac{n!}{(n-a-b)!} < M_1 n^{a+b}.$$

It follows that (5.2.12) holds good in this case if we take

$$M' = M_1 n^{a+b}. \tag{5.3.4}$$

Thus the sequence $\{u'_i\}$ is a majorant as stated.

[1] For the definition of a H-bounded function, see 3.5.

We now have two cases. First, suppose that $c \in \mathfrak{C}$. Let ϑ be an arbitrarily small positive quantity, and choose n so large that the two inequalities

$$\left|\frac{i}{n+1} - c\right| < \frac{\vartheta}{2}, \tag{5.3.5}$$

$$a < i < n - b + 1 \tag{5.3.6}$$

are satisfied simultaneously. We shall prove that this is possible. As to the first inequality, we observe that it follows from condition (c) of the theorem that it is always fulfilled if n is large enough. Turning to (5.3.6), let us assume that it were not true, so that, for example, $i \leqslant a$ even for large n. We would then have $c = 0$, which, as seen from condition (b), implies that $G(u)$ would be finite for $u = 0$. This in its turn would mean that, in (3.5.1), a could be put equal to zero. This would entail $i = 0$, which is impossible. Similarly, it is shown that $i \geqslant n - b + 1$ leads to a contradiction. Thus (5.3.6) is satisfied for large n.

Since, by the definition of continuity, condition (b) holds good in the ϑ-neighbourhood of c if ϑ is small enough, it is by (5.3.5) satisfied in the $\vartheta/2$-neighbourhood of $\xi_{0\nu} = i/(n+1)$. Thus condition (b) of Lemma 1 is fulfilled, if we take $\varepsilon = \vartheta/2$ and $k = 2$. Finally, it follows from (5.3.6) and condition (a) of the theorem that $E x_i$ is finite (cf. 4.3), and thus condition (c) of the lemma is satisfied.

Applying Lemma 2, we obtain (5.3.1), where, as seen from (5.2.13),

$$|R_i| < M\left[\gamma_2 + \left(\frac{\vartheta}{2}\right)^{-\varkappa} \gamma_\varkappa\right] + M'\left(\frac{\vartheta}{2}\right)^{-\varkappa} \gamma'_\varkappa.$$

Here γ_r and γ'_r are the absolute central moments of the variables u_i and u'_i respectively. Replacing $\varkappa$ by $2\varkappa$ and using (4.2.8) and (5.3.4), we get

$$|R_i| < M\left(\frac{1}{n} + \frac{1}{\vartheta^{2\varkappa}\, n^{\varkappa - a - b}}\right). \tag{5.3.7}$$

Since $\varkappa$ is arbitrary, this is equivalent to (5.3.3).

Secondly, suppose that $c \in \mathfrak{C}^*$. By the definition of a less than $n^{-\frac{1}{2}}$-convergent c-sequence (see 3.4), formula (5.3.5) is replaced by

$$\left|\frac{i}{n+1} - c\right| > \frac{\vartheta}{2},$$

where

$$\vartheta = \text{const.}\; n^{-\frac{1}{2}+\rho}, \tag{5.3.8}$$

and ρ is an arbitrarily small positive quantity. Observing that $\mathfrak{C}^*$ consists of a finite number of points, we conclude that, when n is large enough, the distance from the point $i/(n+1)$ to any member of $\mathfrak{C}^*$ is at least $\vartheta/2$. Arguing in other respects as before, we obtain (5.3.1) and (5.3.7) with ϑ given by (5.3.8). Now take

$$\varkappa > \frac{1+a+b}{2\rho}$$

in (5.3.7). It is realized that the second term is then at most of the order $n^{-1+\rho}$. As ρ is arbitrary, we obtain (5.3.3), and the theorem is proved.

Finally, we shall make a remark, which will be found useful later (cf. 8.7). If condition (b) is replaced by the weaker condition that $G(u)$ is bounded and continuous and $G'(u)$ is bounded in $\mathfrak{C}$, then (5.3.3) should be replaced by

$$|R_i| < \frac{M}{\sqrt{n}}.$$

To prove this, use Lemma 2 with $k=1$. In other respects the proof is the same as that given above.

5.4. An approximation formula for variances and covariances of transformed beta-variables

An exact expression for the covariance of a pair of *TRB*-variables was given in 4.3. This expression contains one double and two single integrals and is for the most part difficult to use in practical applications. In this section we shall examine the formula

$$\operatorname{cov}(x_i, x_j) = \frac{1}{n+2} p_i (1-p_j) G'(p_i) G'(p_j) + R_{ij}, \quad (i \leq j). \qquad (5.4.1)$$

Note that this formula includes the variance of x_i.

THEOREM. *If*

(a) *$G(u)$ is H-bounded,*

(b) *Condition A in 3.5 is satisfied with $\mathfrak{C}$ equal to the interval $0 \leq u \leq 1$ with the possible exception of a finite number of points $\mathfrak{C}^*$,*

(c) *x_i belongs to any c_1-sequence when $c_1 \in \mathfrak{C}$, and to a less than $n^{-\frac{1}{2}}$-convergent c_1-sequence when $c_1 \in \mathfrak{C}^*$; x_j belongs to any c_2-sequence $(c_1 \leq c_2)$ when $c_2 \in \mathfrak{C}$, and to a less than $n^{-\frac{1}{2}}$-convergent c_2-sequence when $c_2 \in \mathfrak{C}^*$,* *then the error term in* (5.4.1) *satisfies the inequality*

$$|R_{ij}| < \frac{M}{n^2}, \qquad (5.4.2)$$

where M does not depend upon i, j, and n.

To prove this theorem, we copy the proof of the theorem in 5.3 as closely as possible. Write

$$\operatorname{cov}(x_i, x_j) = E[(G(u_i) - G(p_i))(G(u_j) - G(p_j))] - \\ - [E\,G(u_i) - G(p_i)]\,[E\,G(u_j) - G(p_j)] = J_1 - J_2.$$

We apply Lemma 3 to J_1. For this purpose, take $X_\nu = (u_i, u_j)$ and $X'_\nu = (u'_i, u'_j)$, where u'_i and u'_j have the two-dimensional fr.f.

$$\beta_{i-2a\ j-2a}(u, v;\ n - 2a - 2b).$$

Further, take $g_1 = G(u_i)$ and $g_2 = G(u_j)$. It is found after some calculation that (5.2.12) is fulfilled with M' given by

$$M' = M\,n^{2(a+b)}. \qquad (5.4.3)$$

Hence X'_ν is a majorant of X_ν as required by condition (a) of the lemma.

Let us first suppose that $c_1, c_2 \in \mathfrak{C}$. Then n can be chosen so large that (5.3.5) is valid for both i and j, and also (5.3.6) with a and b replaced by $2a$ and $2b$ respectively. By means of these inequalities it can be verified that conditions (b) and (c) of the lemma are satisfied for $k = 3$.

Applying Lemma 3 to J_1, we get

$$J_1 = \mu_{11} G'(p_i)\,G'(p_j) + R_1 + R_2,$$

where

$$R_1 = \tfrac{1}{2}\mu_{21} G''(p_i)\,G'(p_j) + \tfrac{1}{2}\mu_{12} G'(p_i)\,G''(p_j),$$

and

$$|R_2| < M\left[\Sigma\gamma_{rs} + \left(\frac{\vartheta}{2}\right)^{-\varkappa}\psi_\varkappa\right] + M'\left(\frac{\vartheta}{2}\right)^{-\varkappa}\psi'_\varkappa.$$

Here the quantities μ_{rs} denote the central moments of the variables u_i and u_j. Further, $\psi_\varkappa$ is the sum of their absolute central moments γ_{rs} of order $\varkappa$, and correspondingly for $\psi'_\varkappa$. The range of the summation in $\Sigma\gamma_{rs}$ is clear from the lemma.

We shall study the behaviour of J_1 when n is large. By (4.2.7)

$$|R_1| < \frac{M}{n^2}.$$

Further, in the first term in the upper limit of R_2 we have $r+s \geqslant 4$. Using this observation together with (4.2.8) and (5.4.3), we find, replacing $\varkappa$ by $2\varkappa$,

$$|R_2| < M\left(\frac{1}{n^2} + \frac{1}{\vartheta^{2\varkappa}\, n^{\varkappa-2a-2b}}\right).$$

Finally, as seen from formula (5.3.7), which is evidently applicable in this case,

$$|J_2| < M\left(\frac{1}{n} + \frac{1}{\vartheta^{2\varkappa}\, n^{\varkappa-a-b}}\right)^2.$$

Combining all these results and substituting μ_{11} by its expression as given by (4.2.6), we get (5.4.1) with

$$|R_{ij}| < M\left(\frac{1}{n^2} + \frac{1}{\vartheta^{2\varkappa}\, n^{\varkappa-2a-2b}} + \frac{1}{\vartheta^{4\varkappa}\, n^{2\varkappa-2a-2b}}\right). \tag{5.4.4}$$

Since $\varkappa$ is arbitrary, this is equivalent to (5.4.2).

By an argument similar to that employed in 5.3 we use (5.4.4) to prove that the theorem holds good also when c_1 or c_2 (or both) $\in \mathfrak{C}^*$.

5.5. Weighted differences between consecutive transformed beta-variables

Differences between consecutive *TRB*-variables were briefly discussed in 4.4. Here we shall consider another type of difference, which does not seem to have been studied before in the literature. Write

$$y_i = \theta_{i+1}\, x_{i+1} - \theta_i\, x_i, \tag{5.5.1}$$

$$(i = 0, 1, \ldots, n;\ x_0 = A;\ x_{n+1} = B),$$

where the weights θ_i are defined by

$$\theta_i = [G'(p_i)]^{-1},\ (i = 1, \ldots, n),\ \ \theta_0 = \theta_{n+1} = 0. \tag{5.5.2a}$$

Alternatively, we may write

$$\theta_i = f(\lambda_i),\ \ (i = 1, \ldots, n),\ \ \theta_0 = \theta_{n+1} = 0, \tag{5.5.2b}$$

where λ_i is given by (5.3.2).

It would be more adequate to write y_{in} instead of y_i, but we prefer the simpler notation. The interest of the variables y_i lies chiefly in the fact that, provided that a sufficient number of derivatives of $G(u)$ are continuous, their variances and covariances are independent of $G(u)$ in large samples. More precisely, *suppose that the conditions of the theorem in* 5.4 *are fulfilled, with the exception that Condition A is replaced by Condition B in* 3.5. Note that the latter condition is more restrictive than the former.

Expanding $x_i = G(u_i)$ and $x_{i+1} = G(u_{i+1})$ in Taylor series around p_i and p_{i+1} respectively, we see that y_i can be divided into three parts

$$y_i = m_i + \left(\delta_i - \frac{1}{n+1}\right) + \eta_i, \tag{5.5.3}$$

where
$$m_i = \theta_{i+1}\, G(p_{i+1}) - \theta_i\, G(p_i),$$
$$\delta_i = u_{i+1} - u_i,$$

and η_i is an error variable.

The first two moments of the variables η_i satisfy the inequalities

$$|E\,\eta_i| < \frac{M}{n^2}, \quad \text{var}\ \eta_i < \frac{M}{n^3}, \quad |\text{cov}\ (\eta_i, \eta_j)| < \frac{M}{n^4}, \quad (i \neq j). \tag{5.5.4}$$

Further,

$$|\text{cov}\ (\delta_i, \eta_i)| < \frac{M}{n^{5/2}}, \quad |\text{cov}\ (\delta_i, \eta_j)| < \frac{M}{n^{7/2}}, \quad (i \neq j).$$

These results are proved essentially by the same method as that used in the proofs of the theorems in 5.3 and 5.4. The calculations are, however, more lengthy, five terms of the Taylor expansions being required in this case. Since no new principles are involved, we shall not reproduce any details of the proof.

By means of the partition (5.5.3) and the above inequalities it is easy to derive expressions for the second moments of the variables y_i. In fact, we obtain, using (4.2.12) and (4.2.13),

$$\begin{aligned} \text{var}\ y_i &= \frac{n}{(n+1)^2\,(n+2)} + \Delta_{ii}, \\ \text{cov}\ (y_i, y_j) &= -\frac{1}{(n+1)^2\,(n+2)} + \Delta_{ij}, \quad (i \neq j), \end{aligned} \tag{5.5.5}$$

where
$$|\Delta_{ii}| < \frac{M}{n^{5/2}}, \quad |\Delta_{ij}| < \frac{M}{n^{7/2}}. \tag{5.5.6}$$

Note that, in the special case of a rectangular distribution, η_i, Δ_{ii}, and

Δ_{ij} are zero. The expressions (5.5.5) will be of much use in Chapter 8 in connection with our study of linear combinations.

Finally, we shall derive the limiting distribution of y_i when $n \to \infty$. For this purpose, note that, according to (4.2.11), the variable

$$(n\,\delta_1, n\,\delta_2, \ldots, n\,\delta_k)$$

has the k-dimensional fr.f.

$$\text{const.} \left(1 - \frac{x_1}{n} - \frac{x_2}{n} - \ldots - \frac{x_k}{n}\right)^{n-k}.$$

Consequently, the variables $n\,\delta_i$ are independent in the limit, each having the fr.f. e^{-x}. Moreover, by (5.5.4) and Tchebycheff's inequality, the variable

$$(n\,\eta_1, n\,\eta_2, \ldots, n\,\eta_k)$$

converges in probability to $(0, \ldots, 0)$. Using (5.5.3), we conclude (cf. Cramér, 1946*a*, p. 254) that the variables $n(y_i - m_i) + 1$ are also independent in the limit and exponentially distributed according to the common fr.f. e^{-x}.

CHAPTER 6

THE α, β-CORRECTION

6.1. Introduction

In this chapter we shall consider a generalization of the results deduced in 5.3–5.5. This extension is obtained by introducing what will be called the *α, β-correction* into the asymptotic mean and covariance formulae obtained in these sections. The generalized formulae for the means and covariances of *TRB*-variables will be deduced in 6.2 and 6.3. In 6.4 we shall generalize the weighted differences introduced in 5.5 in a similar way. In 6.5 we shall return to the generalized mean value formula and shall make a detailed investigation of its properties in the rest of the chapter. The formula provides a simple and useful approximation to the mean, which will be discussed in 6.9. In the remaining part of the chapter, the theory will be applied to several particular distributions.

When this chapter was planned, it was deemed desirable to make the proofs as rigorous as possible. Inevitably, this approach has made the exposition rather long. For this reason, anyone whose interest lies primarily in the practical field is advised to study the examples given in sections 6.10–6.15 and consult the preceding sections when necessary.

6.2. A generalized mean value formula

We shall consider a generalization of (5.3.1), viz.

$$E\, x_i = G(\pi_i) + R_i', \tag{6.2.1}$$

where

$$\pi_i = \frac{i - \alpha_i}{n - \alpha_i - \beta_i + 1}. \tag{6.2.2}$$

α_i and β_i are two quantities which may or may not depend upon i and n. We shall sometimes write α_{in}, β_{in} instead of α_i, β_i. When the same values are used for $i = 1, \ldots, n$, we must have $\alpha_i \leqslant 1$, $\beta_i \leqslant 1$. We shall sometimes use the notation[1]

$$\lambda_i' = G(\pi_i).$$

---

[1] In the sequel we shall generally denote "α, β-corrected" quantities by primed letters. An exception to this rule is afforded by the symbol π_i, which is used instead of p_i'.

Further, we shall sometimes write

$$n' = n - \alpha_i - \beta_i. \tag{6.2.3}$$

We shall see in 6.9 that formula (6.2.1) is of practical interest, because the first term generally furnishes a better approximation to the mean than λ_i defined by (5.3.2), if *a single* suitable pair of values $\alpha_i = \alpha$ and $\beta_i = \beta$ is chosen in an appropriate way. Before entering upon a discussion of this approximation, we shall prepare the ground by deriving some properties of the formula in the general case where α_i and β_i depend upon both i and n.

THEOREM. *If*

(a) *the conditions of the theorem in* 5.3 *are satisfied,*

(b)
$$\begin{aligned} a \leqslant \alpha_i &< i \\ b \leqslant \beta_i &< n - i + 1, \end{aligned} \tag{6.2.4}$$

where a, b are the constants appearing in formula (3.5.1),

(c)
$$\begin{aligned} &\alpha_i \to 0 \;\; \textit{when} \;\; i/n \to 0, \\ &\beta_i \to 0 \;\; \textit{when} \;\; i/n \to 1, \end{aligned} \tag{6.2.5}$$

then

$$R_i' = O\left(\frac{1}{n}\right).$$

Note, first, that conditions (b) and (c) are compatible. If, for example, $c = 0$ belongs to $\mathfrak{C}$, the transform must be bounded for $u = 0$. Hence by (3.5.1) we can take $a = 0$. The case $c = 1$ is analogous.

To prove the theorem, rewrite the mean value formula (4.3.1) as follows.

$$E x_i = \frac{n!}{(i-1)!\,(n-i)!} \int_0^1 G(u)\, H(u)\, u^{i-\alpha-1} (1-u)^{n-i-\beta}\, du =$$

$$= \frac{\int_0^1 G(u)\, H(u)\, \beta_{i-\alpha}(u;\, n-\alpha-\beta)\, du}{\int_0^1 H(u)\, \beta_{i-\alpha}(u;\, n-\alpha-\beta)\, du} = \frac{E'\, G H}{E' H}, \tag{6.2.6}$$

where $H(u) = H(u;\, \alpha, \beta)$ is defined by (3.5.1), and $\alpha = \alpha_i$, $\beta = \beta_i$. Further, E' denotes expectation with respect to the beta-variable u_i' with the fr.f. $\beta_{i-\alpha}(u;\, n-\alpha-\beta)$. The mean of u_i' is, clearly, equal to π_i in (6.2.2).

Note that $GH = G(u)\, H(u;\, \alpha_i, \beta_i)$ is bounded for $0 \leqslant u \leqslant 1$, for, by condition (a) in the theorem in 5.3, $G(u)\, H(u;\, a, b)$ is bounded, and $\alpha_i \geqslant a$, $\beta_i \geqslant b$ by (6.2.4). Hence the two integrals in (6.2.6) are both finite.

The theorem in 5.3 is now applied to each of these integrals. Hence

$$E\,x_i = \frac{G(\pi_i)\,H(\pi_i) + O(n^{-1})}{H(\pi_i) + O(n^{-1})}. \qquad (6.2.7)$$

If x_i belongs to a c-sequence with $0<c<1$, the function H remains greater than zero when $n\to\infty$. Thus, clearly, (6.2.7) is equivalent to (6.2.1) with $R_i' = O(n^{-1})$ as asserted in the theorem. If, on the other hand, $c=0$ or $c=1$, we use condition (c), and the same conclusion follows.

We shall return to the subject considered here in 6.5.

6.3. A generalized covariance formula

The α, β-correction can also be introduced into the covariance formula (5.4.1), which then takes the form

$$\operatorname{cov}(x_i, x_j) = \frac{\pi_i(1-\pi_j)}{n'+2}\,G'(\pi_i)\,G'(\pi_j) + R_{ij}', \quad (i \leqslant j), \qquad (6.3.1)$$

where

$$\pi_\nu = \frac{\nu - \alpha_{ij}}{n'+1}, \quad (\nu = i, j),$$

and

$$n' = n - \alpha_{ij} - \beta_{ij}.$$

α_{ij} and β_{ij} may depend on one or more of the parameters i, j, and n or be independent of them. Corresponding to the theorem in 6.2, we have the following result.

THEOREM. *If*

(a) *the conditions of the theorem in* 5.4 *are satisfied,*

(b)
$$2a \leqslant \alpha_{ij} < i, \qquad (6.3.2)$$
$$2b \leqslant \beta_{ij} < n - j + 1,$$

(c)
$$\alpha_{ij} \to 0 \ \textit{when}\ i/n \to 0,$$
$$\beta_{ij} \to 0 \ \textit{when}\ j/n \to 1,$$

then

$$R_{ij}' = O\left(\frac{1}{n^2}\right).$$

The proof follows the same pattern as that of the theorem in 6.2 and will not be given here.

6.4. Generalized weighted differences

In 5.5 we considered weighted differences between consecutive *TRB*-variables with weights θ_i equal to $[G'(p_i)]^{-1}$. A more general class of differences $y_i = \theta_{i+1} x_{i+1} - \theta_i x_i$ is obtained if, in the definition of the weights, p_i is replaced by π_i. Then

$$\theta_i = [G'(\pi_i)]^{-1}, \ (i = 1, \ldots, n), \quad \theta_0 = \theta_{n+1} = 0,$$

where π_i is given by (6.2.2). We shall suppose that the sum $\alpha_i + \beta_i$ is independent of i (but not necessarily of n).

Arguing as in 5.5, we get after some easy calculation

$$\operatorname{var} \ y_i = \frac{\rho_i (n' + 1 - \rho_i)}{(n'+1)^2 (n'+2)} + \Delta'_{ii},$$

$$\operatorname{cov} \ (y_i, y_j) = -\frac{\rho_i \rho_j}{(n'+1)^2 (n'+2)} + \Delta'_{ij}, \ (i < j), \tag{6.4.1}$$

where

$$\rho_i = 1 - \alpha_{i+1} + \alpha_i = 1 - \beta_i + \beta_{i+1}, \tag{6.4.2}$$

$$(i = 0, 1, \ldots, n; \ \alpha_0 = 0; \ \beta_{n+1} = 0).$$

Note that the sum of the ρ_i's is equal to $n' + 1$. In the particular case where α_i is independent of i we have

$$\rho_0 = 1 - \alpha, \ \rho_1 = \cdots = \rho_{n-1} = 1, \ \rho_n = 1 - \beta. \tag{6.4.3}$$

The conclusions made in 5.5 concerning the error terms are valid even in this more general situation. An application of (6.4.1) will be made in 8.5.

6.5. On the error term in the generalized mean value formula

We shall return to the mean value formula (6.2.1). For a purpose which will become clear in the following sections, the term of order n^{-1} in the remainder term of this formula will be separated from the rest of this term. To this end, define a function $Q(u; \alpha, \beta)$ by

$$Q(u; \alpha, \beta) = \tfrac{1}{2} u (1 - u) G''(u) + [\alpha (1 - u) - \beta u] G'(u). \tag{6.5.1}$$

For simplicity, we shall often write $Q(u)$ instead of $Q(u; \alpha, \beta)$. Later, we shall also find it convenient to use the notation

$$L(u) = -\tfrac{1}{2} u (1 - u) \frac{G''(u)}{G'(u)}. \tag{6.5.2}$$

Introducing $f(x)$ and its first derivative in this expression, we have the alternative form

$$L(u) = \tfrac{1}{2} u (1 - u) \frac{f'(x)}{[f(x)]^2}, \tag{6.5.3}$$

where

$$x = G(u).$$

Using (6.5.2), we can write (6.5.1) in the form

$$Q(u; \alpha, \beta) = [\alpha(1-u) - \beta u - L(u)]\, G'(u). \tag{6.5.4}$$

THEOREM. *If the conditions of the theorem in* 6.2 *are satisfied, and if, in addition,* $G''(u)$ *is continuous and* $G'''(u)$ *is bounded in* $\mathfrak{C}$, *then*

$$E\,x_i = G(\pi_i) + \frac{1}{n} Q(\pi_i; \alpha_i, \beta_i) + O(n^{-3/2}). \tag{6.5.5}$$

To prove this, apply Lemma 1 in 5.2 with $k=3$ to each of the integrals in (6.2.6). We then obtain, e.g.,

$$E'\,H = H(\pi_i) + \tfrac{1}{2}\mu_2' H''(\pi_i) + R, \tag{6.5.6}$$

where

$$|R| < M\gamma_3'.$$

Here μ_2' and γ_3' denote moments of the beta-variable u_i' defined in 6.2. According to (4.2.4) and (4.2.8),

$$\mu_2' = \frac{1}{n'+2}\pi_i(1-\pi_i), \quad \gamma_3' = O(n^{-3/2}).$$

Inserting these expressions in (6.5.6), and treating the integral in the numerator in (6.2.6) analogously, we find after some calculation that the term of order n^{-1} in the expansion of $E\,x_i$ is

$$\frac{1}{2(n'+2)}\pi_i(1-\pi_i)\, G\left[\frac{(GH)''}{GH} - \frac{H''}{H}\right],$$

where $G = G(\pi_i)$ and $H = \pi_i^{\alpha_i}(1-\pi_i)^{\beta_i}$. According to (6.5.1), this is equal to $Q(\pi_i)/(n'+2)$. The order of the error term in (6.5.5) is determined by the same reasoning as that used in the last paragraphs of 6.2. This completes the proof.

6.6. A limiting property of the generalized mean value formula

Assume that, corresponding to any *TRB*-variable in a *c*-sequence, the quantities α_{in} and β_{in} are chosen such that the error term in (6.2.1) is *exactly zero*. Generally, this choice can be made in an infinity of ways, since two parameters are used (cf., however, the discussion of the symmetrical case in 6.7, in which case the choice is unique). For any x_i in the *c*-sequence, we then have

$$E\,x_i = G(\pi_i), \tag{6.6.1}$$

where π_i is given by (6.2.2). The following limiting property of equation (6.6.1) will prove useful in the sequel.

THEOREM. *Suppose that* α_{in} *and* β_{in} *are chosen such that* (6.6.1) *is satisfied and that, in addition, quantities* $\alpha(c)$ *and* $\beta(c)$ *exist such that*

$$\alpha_{in} \to \alpha(c), \quad \beta_{in} \to \beta(c) \tag{6.6.2}$$

when $i/n \to c$. *If, further, the conditions of the theorem in* 6.5 *are fulfilled, then*

$$Q(c; \alpha, \beta) = 0 \tag{6.6.3}$$

for $\alpha = \alpha(c)$, $\beta = \beta(c)$.

To prove this theorem, multiply (6.5.5) by n and let $n \to \infty$.

6.7. Application of the limiting property

In the first half of this section, the relation (6.6.1) will be analysed in the special but important *symmetrical* case. A considerable simplification of the discussion is then possible, which is seen as follows. When in (6.2.1) i is replaced by $n-i+1$, we evidently wish to obtain the same expression with a reversed sign. It is immediately realized that, to achieve this, we must have $\alpha = \beta$. Thus (6.6.1) is simplified to

$$E x_i = G\left(\frac{i-\alpha}{n-2\alpha+1}\right). \tag{6.7.1}$$

For any fixed pair of values i and n, this equation in α has generally a single root $\alpha = \alpha_{in}$. (Such is the case, for example, when $G(u)$ is non-decreasing.) When i varies from 1 to n, a sequence of roots $\alpha_{1n}, \alpha_{2n}, \ldots, \alpha_{nn}$ is obtained, where $\alpha_{in} = \alpha_{n-i+1\,n}$. The smallest and the greatest of the roots form end-points of an interval, which we denote by I_n. To different sample-sizes $n = 2, 3, \ldots$ correspond intervals $I_2, I_3, \ldots$

We now use the theorem in 6.6 for studying the behaviour of I_n when $n \to \infty$. *Suppose that the conditions of the theorem in* 6.5 *are satisfied and that the set* $\mathfrak{C}$ *consists of all points in the open interval* $0 < u < 1$. It then follows from the theorem in 6.6 that I_n tends to a limit interval I, which contains all points obtained by solving the relation

$$Q(c; \alpha, \alpha) = 0 \tag{6.7.2}$$

with respect to $\alpha = \alpha(c)$, and letting c vary from 0 to 1. The explicit solution is, as seen from (6.5.4), provided that $G'(c) \neq 0$,

$$\alpha(c) = \frac{L(c)}{1-2c}, \tag{6.7.3}$$

where $L(c)$ is defined by (6.5.2).

Consider the greatest lower bound $\alpha_{\inf}$ and the least upper bound $\alpha_{\sup}$

of $\alpha(c)$ when c varies from 0 to 1. The limit interval I evidently consists of all points α in the interval

$$\alpha_{\inf} \leqslant \alpha \leqslant \alpha_{\sup}. \tag{6.7.4}$$

We observe that, when $f(x)$ has a single maximum, I consists only of non-negative values of α.

We now turn to the general, *non-symmetrical* case. Instead of considering the mean of a single *TRB*-variable, we shall then study the means of two "symmetrically" situated variables x_i and x_{n-i+1} simultaneously; i.e., we replace equation (6.7.1) by a system of equations in α and β

$$E\,x_\nu = G\left(\frac{\nu-\alpha}{n-\alpha-\beta+1}\right), \quad (\nu = i \text{ and } n-i+1). \tag{6.7.5}$$

For any fixed values of i and n, this system has generally a unique solution $\alpha = \alpha_{in}$, $\beta = \beta_{in}$, which has similar properties as the solution of (6.7.1). When n increases, we obtain in the limit not a single limit interval I but two intervals I_α and I_β, which consist of all points α and β, respectively, obtained by solving the system of equations

$$\begin{aligned} Q(c;\alpha,\beta) &= 0, \\ Q(1-c;\alpha,\beta) &= 0 \end{aligned} \tag{6.7.6}$$

with respect to $\alpha = \alpha(c)$ and $\beta = \beta(c)$. As seen from (6.5.4), the solution of the system is

$$\begin{aligned} \alpha(c) &= \frac{(1-c)\,L(c) - c\,L(1-c)}{1-2c}, \\ \beta(c) &= \frac{c\,L(c) - (1-c)\,L(1-c)}{1-2c}, \end{aligned} \tag{6.7.7}$$

where $L(c)$ is given by (6.5.2). In the symmetrical case, the solution is specialized to (6.7.3).

By assigning different values to c, it is possible to determine the intervals I_α and I_β from (6.7.7). The limits $\alpha(0)$ and $\beta(0)$ are often one of the end-points of I_α and I_β respectively. As these limits will be used later, they will be determined below for *AL*- and *AP*-transforms.

If $G(u)$ is an *AL*-transform at both end-points (cf. 3.5), the limits are equal to

$$\alpha(0) = \beta(0) = \tfrac{1}{2}. \tag{6.7.8}$$

If $G(u)$ is an *AP*-transform at both end-points,

$$\alpha(0) = \beta(0) = \frac{1-k}{2}. \tag{6.7.9}$$

Note that, if $G(u)$ is an AL- or AP-transform only at one end-point, $u=1$ say, and G''/G' is finite for $u=0$, then only $\beta(0)$ is given by (6.7.8) or (6.7.9), while $\alpha(0)$ is equal to 0. The exponential distribution is a case in point.

Transforms with short intervals I_α and I_β are of interest both theoretically and practically. We shall make some comments upon such transforms in 6.11 and 6.15.

6.8. A large-sample inequality

We have hitherto assumed that one pair of quantities α_i, β_i is used for each TRB-variable, and certain theoretical results concerning such quantities have been deduced. In this section we shall assume that α and β in the expression $G(\pi_i)$ are independent of i and n.

Consider the inequality

$$G(\pi_{i1}) \leqslant E\,x_i \leqslant G(\pi_{i2}), \tag{6.8.1}$$

where

$$\pi_{i\nu} = \frac{i-\alpha_\nu}{n-\alpha_\nu-\beta_\nu+1}, \quad (\nu = 1, 2), \tag{6.8.2}$$

and α_1, β_1 and α_2, β_2 are two pairs of constants. We wish to determine these constants so that (6.8.1) holds good for all i (or, at least, for a succession of i-values) when $n=\infty$. When (6.8.1) has this property, we shall call it a large-sample inequality for $E\,x_i$.

Suppose that the conditions of the theorem in 6.5 *are satisfied.* It then follows from the expansion (6.5.5) that the determination of the constants consists essentially in investigating the sign of Q for various combinations of α and β. The problem is solved if we can find constants such that the inequalities

$$Q_1 = Q(c;\, \alpha_1, \beta_1) \geqslant 0,$$
$$Q_2 = Q(c;\, \alpha_2, \beta_2) \leqslant 0$$

are valid for any c in the interval $0<c<1$ (or in some part hereof). Since Q is a linear function of α and β, this procedure presents no great difficulties in practice. It is, of course, desirable that the choice is made so that the resulting limits are as close together as possible.

Note that, when $G(u)$ is non-decreasing and $G'(u)$ is symmetrical, we already know the answer to the question of the best choice of the constants. Combining (6.5.4) and (6.7.3) and taking $\alpha=\beta$, we get

$$Q(c) = [\alpha - \alpha(c)]\,(1-2c)\,G'(c).$$

Suppose, for instance, that $c \geqslant \frac{1}{2}$. Insert the end-points $\alpha_{\inf}$ and $\alpha_{\sup}$ of the limit interval I in this relation. It then follows that $Q(c) \geqslant 0$ for $\alpha = \alpha_{\inf}$ and $Q(c) \leqslant 0$ for $\alpha = \alpha_{\sup}$. Accordingly, we take

$$\alpha_1 = \beta_1 = \alpha_{\inf}, \quad \alpha_2 = \beta_2 = \alpha_{\sup}$$

in (6.8.2). Clearly, the resulting limits in (6.8.1) cannot be improved. The modifications required when $c \leqslant \frac{1}{2}$ are obvious.

Finally, we make a general remark concerning the inequality (6.8.1). The question of the validity of an inequality of this form for *any* value of n leads to an interesting problem, which will be mentioned only in passing. If constants α_ν, β_ν have been found such that (6.8.1) is true for large n, it is, of course, possible to change the constants when necessary, so that the inequality holds for all values of n (and the limits may still be not very far apart so that the formula has practical interest). The determination of the best values of the constants can, however, be an intricate problem.

The following simple remark is of some interest in this connection: If the function $G(u)$ is continuous and convex for any u in the interval $0 \leqslant u \leqslant 1$, then for all i and n

$$E\,x_i \geqslant G\left(\frac{i}{n+1}\right).$$

Thus, we can take $\alpha_1 = \beta_1 = 0$ in (6.8.2). The truth of the statement follows immediately from (4.3.1) and the properties of convex functions (see, e.g., Hardy & Littlewood & Polya, 1934, Ch. III).

6.9. An approximate mean value formula

We now carry the simplification a final step forward and use only one pair of constants α, β in $G(\pi_i)$ for all values of i and n.

We then have the simple approximation

$$E\,x_i \sim G(\pi_i), \tag{6.9.1}$$

where

$$\pi_i = \frac{i-\alpha}{n-\alpha-\beta+1}.$$

Following the same line of thought as in the preceding part of the chapter, we try to find constants which make Q small. Generally speaking, this requirement is evidently satisfied, if α and β are chosen somewhere between the values α_ν, β_ν, $(\nu = 1, 2)$, used in the asymptotic inequality discussed in 6.8. The final choice obviously depends upon the desirability of making Q small for certain c-sequences. For instance, if we are interested in TRB-variables with ranks near 1 or n, we make the choice so that $Q(c; \alpha, \beta)$ is zero when c is close to 0 or close to 1. This argument suggests the following simple practical rule: *Determine α and β in* (6.9.1) *from the solution* (6.7.7) *of the system* (6.7.6) *with a suitable fixed value of c.* Geometrically speaking, this is equivalent to saying that two corresponding points are chosen in the limit intervals I_α and I_β. In the symmetrical case, we take as usual $\alpha=\beta$. The determination of α then consists in selecting a point in the limit interval I.

Finally, we make the practically important remark that the procedure described above presupposes that n is fairly large, as we have altogether neglected the term of order $n^{-3/2}$ in (6.5.5). It is, however, a notable feature of the formula (6.9.1) that, in many practical cases, it is useful even when n is quite small. In order to attain the highest possible acccuracy for small n, it may, however, be necessary to choose values of α and β which are situated slightly outside the limit intervals I_α and I_β. The examples discussed in the following sections support these assertions.

If the accuracy of the formula is found to be unsatisfactory when α and β are constant, it is, at least theoretically, possible to choose a special combination of constants for each value of n. We shall, however, not consider this possibility.

In the remaining part of the chapter, the theory will be applied to several examples. In all these examples, $G(u)$ is an increasing function of u.

6.10. Normal distribution

In the normal case we have $G(u)=\Phi^{-1}(u)$,

$$\Phi(x)=\int_{-\infty}^{x}\varphi(t)\,dt,$$

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}.$$

As seen from (6.5.3),

$$L(u)=-\tfrac{1}{2}u(1-u)\cdot\frac{x}{\varphi(x)}.$$

Hence by (6.7.3)

$$\alpha(c)=\tfrac{1}{2}\frac{c(1-c)}{2c-1}\cdot\frac{x}{\varphi(x)},$$

where $x=\Phi^{-1}(c)$.

TABLE 1.

Normal distribution. Root α_{in} of equation (6.10.2).

i	Sample-size n						
	2	4	6	8	10	15	20
1	0.330	0.347	0.355	0.360	0.364	0.370	0.374
2		0.359	0.368	0.374	0.378	0.385	0.390
3			0.370	0.375	0.379	0.386	0.391
4				0.375	0.379	0.386	0.390
5					0.379	0.384	0.389
6						0.383	0.388
7						0.382	0.387
8							0.386
9							0.386
10							0.386

When c increases from 0 to 1, the function $\alpha(c)$ decreases from $\frac{1}{2}$ to a minimum equal to $\frac{\pi}{8}$ for $c=\frac{1}{2}$, and then increases to the same value $\frac{1}{2}$. Consequently, the limit interval I (cf. 6.7) is equal to approximately (0.39, 0.50), and we obtain from (6.8.1) for $i \geqslant (n+1)/2$ the large-sample inequality

$$\Phi^{-1}\left(\frac{i-0.39}{n+0.22}\right) \leqslant E\,x_i \leqslant \Phi^{-1}\left(\frac{i-0.5}{n}\right). \qquad (6.10.1)$$

When $i \leqslant (n+1)/2$, the inequality signs should be reversed.

Next, we make a numerical study of the equation (6.7.1), which becomes in this case

$$E\,x_i = \Phi^{-1}\left(\frac{i-\alpha}{n-2\alpha+1}\right). \qquad (6.10.2)$$

The root $\alpha=\alpha_{in}$ of this equation has been determined for some values of i and n by means of a table of means of normal order statistics (Godwin, 1949 *b*). The result is given in Table 1. It is interesting to observe that α_{in} does not change very much with i and n when $n \leqslant 20$, the smallest and largest values being 0.33 and 0.39 respectively. When n tends to infinity, the root tends to some point within the interval I considered above. Thus we are led to suppose that α_{in} is always situated in the interval (0.33, 0.50). This would be equivalent to saying that the inequality, which is obtained by replacing $(i-0.39)/(n+0.22)$ by

TABLE 2.

Normal distribution. Calculation of Ex_i by means of the right member in (6.10.2).

n	i	Exact value	α		
			0	3/8	1/2
5	1	1.163	0.967	1.180	1.282
	2	0.495	0.431	0.497	0.524
10	1	1.539	1.335	1.547	1.645
	2	1.001	0.908	1.000	1.036
	3	0.656	0.605	0.655	0.674
	4	0.376	0.349	0.375	0.385
	5	0.123	0.114	0.123	0.126
15	1	1.736	1.534	1.739	1.834
	2	1.248	1.150	1.241	1.282
	3	0.948	0.887	0.946	0.967
	4	0.715	0.674	0.714	0.728
	5	0.516	0.489	0.515	0.524
	6	0.335	0.319	0.335	0.341
	7	0.165	0.157	0.165	0.168

All entries should be given negative signs.

$(i-0.33)/(n+0.33)$ in (6.10.1), is true for *any* i and n. The truth of this statement has, however, not been analytically verified.

Because of the small variation of α_{in} it may be expected that (6.9.1) will be fairly accurate when used with a constant value of α. As a compromise between different possible values we suggest $\alpha=\frac{3}{8}$, which leads to the formula

$$Ex_i \sim \Phi^{-1}\left(\frac{i-\frac{3}{8}}{n+\frac{1}{4}}\right). \qquad (6.10.3)$$

This simple expression is quite useful as an overall approximation to the mean. Some information concerning its accuracy is conveyed by Table 2, which also contains accurate values taken from existing tables and values computed from the expressions

$$Ex_i \sim \Phi^{-1}\left(\frac{i}{n+1}\right), \qquad (6.10.4)$$

$$E x_i \sim \Phi^{-1}\left(\frac{i-\frac{1}{2}}{n}\right). \tag{6.10.5}$$

The first expression has been included in the calculations, because it is sometimes asserted that it furnishes a good approximation to the mean (cf. the discussion in 12.4 and 12.6).

It is seen from the table that (6.10.3) is quite accurate in view of the small sample-sizes. (6.10.4) is unsatisfactory when n is small For instance, when $n=10$, it underestimates the true value by 7–13%, depending on i. For $n=10$, the expression (6.10.5) overestimates the mean by 3–7%.

6.11. An optimal property of the normal distribution

We found in 6.10 that the limit interval I is comparatively short when the distribution is normal. It will now be proved that the normal distribution has the shortest interval of the family of distributions with the symmetrical fr.f.

$$f(x)=C(m)\exp\left(\frac{-|x|^m}{m}\right), \quad (-\infty<x<\infty;\ m>0).$$

The transform $G(u)$ corresponding to this fr.f. is evidently an *AL*-transform. Thus $\alpha(c)$ satisfies (6.7.8) when c tends to 0 or 1.

Further, let us determine the value of $\alpha(c)$ when $c=\frac{1}{2}$. From the relation

$$2F(x)-1=2\int_0^x f(t)\,dt \sim 2C(m)\,x,$$

which is valid for values near $x=0$, we obtain for c near $\frac{1}{2}$

$$\alpha(c) \sim \frac{|x|^{m-2}}{16\,[C(m)]^2}.$$

Observing that $C(2)=(2\pi)^{-\frac{1}{2}}$, we find

$$\alpha(\tfrac{1}{2})=\begin{cases}\infty & \text{for } 0<m<2\\ \dfrac{\pi}{8} & \text{for } m=2\\ 0 & \text{for } m>2.\end{cases} \tag{6.11.1}$$

Combining this result with (6.7.8), we conclude that the length of I is at least $\frac{1}{2}$, except for $m=2$ when (as already stated in 6.10) it is equal to $\frac{1}{2}-\frac{\pi}{8}=0.11$. This proves the assertion that I is shortest when the distribution is normal. It might be added that, when $m=1$, we obtain Laplace's distribution. It follows from (6.11.1) that I has then infinite length.

6.12. Cauchy's distribution

When ξ follows Cauchy's distribution, we have

$$f(x)=\frac{1}{\pi}\cdot\frac{1}{1+x^2}, \quad (-\infty<x<\infty).$$

Hence

$$G(u)=\operatorname{tg}\pi(u-\tfrac{1}{2}).$$

Formula (6.7.3) is specialized to

$$\alpha(c)=\frac{\pi c(1-c)}{2c-1}\operatorname{tg}\pi(c-\tfrac{1}{2}).$$

Hence $\alpha(0)=\alpha(1)=1$ and $\alpha(\frac{1}{2})=\pi^2/8$, these values being end-points of I. The former limit can also be obtained by taking $k=-1$ in (6.7.9). The large-sample inequality (6.8.1) thus becomes

$$\operatorname{tg}\pi\left(\frac{i-1.23}{n-1.46}-\tfrac{1}{2}\right)\leqslant E\,x_i\leqslant\operatorname{tg}\pi\left(\frac{i-1}{n-1}-\tfrac{1}{2}\right),\ \left(\frac{n+1}{2}\leqslant i\leqslant n-1\right).$$

For $2\leqslant i\leqslant(n+1)/2$, the inequality signs should be reversed.

6.13. Extreme-value distribution

The extreme-value distribution has the cdf.

$$F(x)=\exp(-e^{-x}), \quad (-\infty<x<\infty). \tag{6.13.1}$$

Hence

$$G(u)=-\log\log\frac{1}{u}. \tag{6.13.2}$$

The properties of this distribution, which was first derived by Fisher & Tippett (1928) in connection with a study of the extreme values from a normal distribution (cf. Chapter 7), have been studied in later years by Gumbel (1954) and Lieblein (1954), among others. A table of $E\,x_i$ is available (Table of the first moment of ranked extremes, 1951).

First, we shall determine the intervals I_α and I_β defined in 6.7. For this purpose, consider the general formula

$$L(u)=(1-u)\left[a-b\left(\log\frac{1}{u}\right)^{-1}\right]. \tag{6.13.3}$$

It is easily seen from (6.5.2) and (6.13.2) that, by taking $a=b=\frac{1}{2}$ in this expression, we obtain $L(u)$ for the extreme-value distribution. Inserting (6.13.3) into (6.7.7) and letting c vary from 0 to 1, it is found after some calculation that the end-points of I_α and I_β are

$$\begin{aligned} &\alpha(0)=\alpha(1)=a, \quad \alpha(\tfrac{1}{2})=\frac{a}{2\log^2 2}=1.04\,a, \\ &\beta(0)=\beta(1)=b, \quad \beta(\tfrac{1}{2})=b-a\left(\frac{1}{\log 2}-\frac{1}{2\log^2 2}\right)=b-0.40\,a. \end{aligned} \tag{6.13.4}$$

It follows that, in the case of the extreme-value distribution, $I_\alpha=(0.50, 0.52)$, $I_\beta=(0.30, 0.50)$.

Secondly, we shall determine the limits in the asymptotic inequality. (6.5.4) is specialized to

$$Q(u;\alpha,\beta)=\frac{1-u}{\log 1/u}\left(\frac{\alpha-\frac{1}{2}}{u}-\frac{\beta}{1-u}+\frac{1}{2u\log 1/u}\right).$$

We now seek constants α, β which make $Q\geqslant 0$ or $Q\leqslant 0$ for any u, and then use the double inequality

$$\frac{t}{1-t/2}<\log\frac{1}{1-t}<\frac{t}{1-t}, \quad (0<t<1).$$

Replacing t by $1-u$ and making a rearrangement, we obtain the new inequality

$$\frac{1}{1-u}<\frac{1}{u\log 1/u}<\frac{1}{1-u}+\frac{1}{2u}, \quad (0<u<1).$$

Hence

$$Q(u;\tfrac{1}{2},\tfrac{1}{2})>0, \quad Q(u;\tfrac{1}{4},\tfrac{1}{2})<0.$$

Accordingly, we can take $\alpha_1=\frac{1}{2}$, $\beta_1=\frac{1}{2}$ and $\alpha_2=\frac{1}{4}$, $\beta_2=\frac{1}{2}$ in (6.8.1) and obtain the large-sample inequality

$$G\left(\frac{i-\frac{1}{2}}{n}\right)\leqslant E\,x_i\leqslant G\left(\frac{i-\frac{1}{4}}{n+\frac{1}{4}}\right), \tag{6.13.5}$$

where $G(u)$ is given by (6.13.2). We observe that α_1 is identical with the left end-point $\alpha(0)$ of I_α and β_1 with the right end-point $\beta(0)$ of I_β.

Some numerical illustrations are given in Table 3, where the limits in (6.13.5) have been calculated for $n=5$ and $n=10$. Either of the limits can be used as an approximation to the mean, the upper limit being, however, more accurate for these sample-sizes.

TABLE 3.

Extreme-value distribution. Calculation of $E x_i$ by means of the limits in (6.13.5).

n	i	Exact value	Lower limit	Upper limit
5	1	−0.69	−0.83	−0.67
	2	−0.11	−0.19	−0.10
	3	0.43	0.37	0.44
	4	1.07	1.03	1.09
	5	2.19	2.25	2.30
10	1	−0.99	−1.10	−0.96
	2	−0.58	−0.64	−0.57
	3	−0.28	−0.33	−0.28
	4	−0.01	−0.05	−0.01
	5	0.26	0.23	0.26
	6	0.54	0.51	0.55
	7	0.87	0.84	0.87
	8	1.27	1.25	1.27
	9	1.83	1.82	1.85
	10	2.88	2.97	2.99

6.14. Weibull's distribution

The cdf. of Weibull's distribution (see Weibull, 1951) can be written

$$F(x) = 1 - \exp(-x^m), \quad (x \geqslant 0; m \geqslant 1). \tag{6.14.1}$$

Hence

$$G(u) = \left(\log \frac{1}{1-u}\right)^{1/m}. \tag{6.14.2}$$

We shall determine I_α and I_β. We find from (6.5.2) that $L(u)$ is obtained by taking

$$a = \tfrac{1}{2}\left(1 - \frac{1}{m}\right), \quad b = \tfrac{1}{2} \tag{6.14.3}$$

in the formula

$$L(u) = u\left[a\left(\log \frac{1}{1-u}\right)^{-1} - b\right]. \tag{6.14.4}$$

Introducing this expression into (6.7.7) and letting c vary from 0 to 1, we easily find that I_α and I_β have the same end-points (6.13.4) as when $L(u)$ is given by (6.13.3). Inserting the values (6.14.3) into (6.13.4), we conclude that I_α has the end-points

$$\alpha(0) = 0.5\left(1 - \frac{1}{m}\right), \quad \alpha(\tfrac{1}{2}) = 0.52\left(1 - \frac{1}{m}\right),$$

and I_β the end-points

$$\beta(\tfrac{1}{2}) = 0.5 - 0.2\left(1 - \frac{1}{m}\right), \quad \beta(0) = 0.5.$$

6.15. *B*-transforms

In the preceding examples the choice of constants in (6.9.1) necessitated a compromise between conflicting interests, the best values being functions of i and n. In one special case the determination is easily made, namely when $G(u)$ is a B-transform (see 3.5).

Taking logarithms of both members of (3.5.4) and differentiating with respect to u, we infer that a B-transform satisfies the differential equation

$$\frac{G''(u)}{G'(u)} + \frac{\tau_1}{u} - \frac{\tau_2}{1-u} = 0.$$

By this relation and (6.5.1) we conclude that, identically for any u,

$$Q(u; \tau_1/2, \ \tau_2/2) = 0.$$

Thus the intervals I_α and I_β degenerate to the points $\alpha = \tau_1/2$ and $\beta = \tau_2/2$.

We use these points as constants in the mean value expansion (6.5.5), which becomes

$$E\,x_i = G(\pi_i) + O(n^{-3/2}), \tag{6.15.1}$$

where $G(u)$ is obtained by integrating (3.5.4), and

$$\pi_i = \frac{i - \tau_1/2}{n - \tau_1/2 - \tau_2/2 + 1}. \tag{6.15.2}$$

Note that an exact expression for $E\,x_i$ is obtained by summing (4.4.5) from 0 to $i-1$. By identifying the exact expression with the leading term in (6.15.1), we obtain a relation, which has a certain interest of its own.

In order to illustrate the nature of the approximation (6.15.1), we take the special case $\tau_1 = 0$, $\tau_1 = 1$, which, as mentioned in 3.5, leads to the exponential distribution. $G(\pi_i)$ is then specialized to $\log[(n+\frac{1}{2})/(n-i+\frac{1}{2})]$, which expression can also be obtained by approximating the exact value, given by the first formula (4.5.3), by means of the Euler–Maclaurin sum formula.

CHAPTER 7

SINGULAR TRANSFORMED BETA-VARIABLES

7.1 Introduction

The study of singular *TRB*-variables undertaken in this chapter will concern singular i-sequences (see 3.4). The foundation of this subject was laid by Fisher & Tippett (1928), who derived limiting distributions for the extreme variables x_1 and x_n when $G(u)$ is specialized in various ways. Their work was extended by Gumbel in a series of papers (1936, 1944, 1946, 1947). Frechét (1927) and Smirnoff (1935) have also contributed to the development of the subject. Several other papers might be mentioned, e.g. those by Elfving (1947) and Cox (1948) dealing with the asymptotic properties of the range in normal samples.

We shall determine the first two moments of *TRB*-variables belonging to singular i-sequences in the practically important cases when $G(u)$ is an *AL*- or *AP*-transform. These cases are closely related to cases I and II, respectively, studied by Fisher & Tippett. The α, β-correction introduced in Chapter 6 will prove useful even in the present connection. In the last section of the chapter, the results will be generalized to more general sequences than i-sequences.

To make the exposition perfectly rigorous, we would need similar tools as in Chapter 5 and, in particular, certain lemmas of the kind derived in 5.2. This would have made the chapter rather long, and we have preferred to omit certain details of the derivations.

7.2. An expansion valid for *AL*-transforms

The *AL*-transform can be asymptotically logarithmic at $u=0$ or at $u=1$ (or at both these points). We shall confine the discussion to the first-mentioned case, the other alternatives being similar.

Suppose, then, that the first formula (3.5.2) is valid. Consider a *TRB*-variable x_i, where i is a fixed number. An asymptotic expansion for x_i can be derived as follows. We have from (3.5.2)

$$x_i = -c_0 \left(\log \frac{1}{u_i}\right)^k \left[1 + O\left(\frac{1}{\log 1/u_i}\right)\right] \tag{7.2.1}$$

and, using the notation (5.3.2),

$$\lambda_i = -c_0 \left(\log \frac{1}{p_i}\right)^k \left[1 + O\left(\frac{1}{\log 1/p_i}\right)\right]. \tag{7.2.2}$$

Define random variables ω_i, $(i = 1, 2, \ldots)$, by

$$\omega_i = \log u_i - \log p_i. \tag{7.2.3}$$

Inserting ω_i into (7.2.1), we find that the leading term can be written

$$-c_0 \left(\log \frac{1}{p_i}\right)^k \left(1 - \frac{\omega_i}{\log 1/p_i}\right)^k.$$

Expanding the last factor in a Taylor series, we obtain after some calculation

$$x_i = \lambda_i + k c_0^{1/k} (-\lambda_i)^{1-1/k} \omega_i + R(\omega_i), \tag{7.2.4}$$

where we can make the following statement with respect to the error term: Corresponding to any positive quantity ε, there is a quantity M_ε, which does not depend upon i, such that, if

$$\left|\frac{\omega_i}{\lambda_i^{1/k}}\right| < \varepsilon,$$

then

$$|R(\omega_i)| < M_\varepsilon \omega_i^2 |\lambda_i|^{1-2/k}.$$

Note that by (7.2.2)

$$\lambda_i = O\left((\log n)^k\right). \tag{7.2.5}$$

Hence, when ω_i is bounded,

$$R(\omega_i) = O\left((\log n)^{k-2}\right). \tag{7.2.6}$$

It follows from (7.2.4) that, apart from the error term, x_i is a linear function of ω_i. Using the fact (cf. Cramér, 1946 *a*, p. 370 ff.) that the variable $-\log n u_i$ has, in the limit, the fr.f.

$$\frac{1}{\Gamma(i)} \exp(-ix - e^{-x}), \tag{7.2.7}$$

we obtain the well-known result that, apart from scale and location factors, x_i also has this limiting distribution. Our principal use of (7.2.4) will, however, be to derive moments of x_i, and then we need not invoke (7.2.7).

It should be noted that, if we take $i = 1$ in (7.2.7), we obtain the fr.f. for the extreme-value distribution (cf. 6.13).

7.3. *AL*-transforms. First two moments of transformed beta-variables

We obtain from (7.2.4) an approximation to the mean of x_i, viz.

$$E\,x_i = \lambda_i + k\,c_0^{1/k}\,(-\lambda_i)^{1-1/k}\,E\,\omega_i + R_i, \tag{7.3.1}$$

where

$$|R_i| < M\,|\lambda_i|^{1-2/k}\,E\,\omega_i^2. \tag{7.3.2}$$

By (7.2.3) and the first relation (4.5.4)

$$E\,\omega_i = -\sum_{\nu=i}^{n} \frac{1}{\nu} - \log p_i. \tag{7.3.3}$$

It can be shown that

$$|E\,\omega_i| < \frac{C}{i}, \qquad E\,\omega_i^2 < \frac{M}{i}, \tag{7.3.4}$$

where C is Euler's constant, and M is a constant which does not depend upon i and n. By the latter of these relations and (7.2.5)

$$R_i = O\left((\log n)^{k-2}\right). \tag{7.3.5}$$

The expression (7.3.3) is inconvenient when n is large, and we shall derive an approximation. Take $N = n$ in the formula (cf. Cramér, 1946 *a*, p. 125)

$$\sum_{\nu=1}^{N} \frac{1}{\nu} = \log\,(N + \tfrac{1}{2}) + C + O\,(N^{-2}), \tag{7.3.6}$$

where, as before, C is Euler's constant. Hence

$$E\,\omega_i = \sum_{\nu=1}^{i-1} \frac{1}{\nu} - \log i - C + O\,(n^{-1}). \tag{7.3.7}$$

For a later purpose, we shall also give another version of the expansion (7.3.1). Combining the last two terms in the expansion and using the second inequality (7.3.4), we obtain

$$E\,x_i = \lambda_i + R_i, \tag{7.3.8}$$

where

$$|R_i| < M\,|\lambda_i|^{1-1/k}\left(|E\,\omega_i| + \frac{1}{i}\,|\lambda_i|^{-1/k}\right). \tag{7.3.9}$$

We have used the same notation for the error term as in (7.3.1), which, it is hoped, will not cause confusion. By the first result (7.3.4) we have, more simply,

$$|R_i| < \frac{M}{i} |\lambda_i|^{1-1/k}. \qquad (7.3.9\,\text{a})$$

The quantities M in these relations do not depend upon i and n.

Approximations to the covariance of x_i and x_j can be derived from (7.2.4) in a similar manner. Omitting certain details of calculation, we find

$$\text{cov}(x_i, x_j) = k^2 c_0^{2/k} (\lambda_i \lambda_j)^{1-1/k} \text{cov}(\omega_i, \omega_j) + R_{ij}, \quad (i \leqslant j), \qquad (7.3.10)$$

where, by the second formula (4.5.4),

$$\text{cov}(\omega_i, \omega_j) = \sum_{\nu=j}^{n} \frac{1}{\nu^2} = \frac{\pi^2}{6} - \sum_{\nu=1}^{j-1} \frac{1}{\nu} + O(n^{-1}). \qquad (7.3.11)$$

The latter of these expressions is convenient to use when n is large. Further,

$$R_{ij} = O\left((\log n)^{2k-3}\right). \qquad (7.3.12)$$

It is easily seen that

$$\text{cov}(\omega_i, \omega_j) < \frac{\pi^2}{6j}. \qquad (7.3.13)$$

It follows from this relation and from an analysis of R_{ij} that, corresponding to (7.3.8) and (7.3.9 a), we have the simple result

$$|\text{cov}(x_i, x_j)| < \frac{M}{j} |\lambda_i \lambda_j|^{1-1/k}, \quad (i \leqslant j). \qquad (7.3.14)$$

All the formulae derived above can easily be adapted to the case where $G(u)$ is an AL-transform at the point $u=1$.

Normal distribution.

We shall finish the section by applying the results to the normal distribution. As already mentioned in 3.5, we then have $c_0 = \sqrt{2}$, $k = \frac{1}{2}$. Thus (7.3.1) and (7.3.10) are specialized to

$$\begin{aligned} E\,x_i &= \lambda_i - \lambda_i^{-1} E\,\omega_i + O\left((\log n)^{-3/2}\right), \\ \text{cov}(x_i, x_j) &= (\lambda_i \lambda_j)^{-1} \text{cov}(\omega_i, \omega_j) + O\left((\log n)^{-2}\right), \end{aligned} \qquad (7.3.15)$$

where

$$\lambda_i = \Phi^{-1}\left(\frac{i}{n+1}\right).$$

The principal use of the formulae deduced in this section will be made in 8.9.

7.4. *AL*-transforms. Use of the α, β-correction

We saw in Chapter 6 that a simple approximation to the mean of a non-singular *TRB*-variable can be obtained by means of the α, β-correction. This correction is of interest also in the case of singular *TRB*-variables, as will be demonstrated below for *AL*-transforms.

An inspection of 7.2 and 7.3 shows that we may everywhere replace p_i by π_i defined by (6.2.2). In particular, (7.3.1) and (7.3.10) are valid if we replace λ_i and ω_i by

$$\lambda_i' = G(\pi_i) = G\left(\frac{i-\alpha_i}{n-\alpha_i-\beta_i+1}\right)$$

and

$$\omega_i' = \log u_i - \log \pi_i,$$

respectively. We observe that (7.3.7) is then replaced by

$$E\,\omega_i' = \sum_{\nu=1}^{i-1}\frac{1}{\nu} - \log(i-\alpha_i) - C + O(n^{-1}). \tag{7.4.1}$$

We shall make some comments with respect to the choice of α_i. Applying the same principle as in Chapter 6, we determine α_i so that, apart from the error term in (7.4.1), the mean $E\,\omega_i'$ is zero. Using (7.3.5), we then conclude that the α, β-corrected form of (7.3.1) becomes

$$E\,x_i = \lambda_i' + O\left((\log n)^{k-2}\right), \tag{7.4.2}$$

independently of the choice of β_i. We infer from (7.4.1) that α_i in this formula is given by

$$\alpha_i = i - \exp\left(\sum_{\nu=1}^{i-1}\frac{1}{\nu} - C\right). \tag{7.4.3}$$

For $i=1, 2,$ and 3 we obtain $\alpha_1 = 0.438$, $\alpha_2 = 0.474$, and $\alpha_3 = 0.483$. For large i we have by an application of (7.3.6)

$$E\,\omega_i' = \log(i-\tfrac{1}{2}) - \log(i-\alpha_i) + O(i^{-2}) + O(n^{-1}). \tag{7.4.4}$$

Thus α_i tends to $\frac{1}{2}$ when i tends to infinity. As seen from (6.7.8) we obtained the same result in our study of non-singular *TRB*-variables. These two identical results are interesting, because they form a connecting link between singular and non-singular *TRB*-variables, which in other respects behave very differently.

It might also be noted that the values α_i given above are identical with the limits of the roots α_{in} given in Table 1 when $n \to \infty$ and i is constant. As seen from the table, p. 70, and from the values of

α_1, etc., given above, the roots converge fairly slowly to the limiting values when $G(u)$ is the inverse of the normal cdf.

If $G(u)$ is an AL-transform at $u=1$, the mean Ex_{n-i+1} can be submitted to a similar analysis as that performed above. We then find that β_{n-i+1} should be determined from the same expression (7.4.3) as α_i. In the symmetrical case we take as usual $\alpha_i=\beta_i$. If, on the other hand, $G(u)$ is finite for $u=1$, we take $\beta_i=0$.

We shall illustrate these ideas by an example.

Normal distribution.

By (7.4.2) and (7.4.3), the mean of the range of a sample from a normal distribution is approximately equal to $-2\Phi^{-1}[(1-\alpha_1)/(n-2\alpha_1+1)]$, where $\alpha_1=0.438$. For $n=20$, this expression furnishes the value 3.82, the correct value being 3.73. The approximations indicated by Cox (1948; § 3) and Elfving (1947) lead to the values 3.83 and 3.76, respectively. (It might be mentioned that if, instead of α_1, we choose the value $\alpha=\frac{3}{8}$ proposed in 6.10, the value 3.82 is changed to 3.74.)

7.5. *AP*-transforms. First two moments of transformed beta-variables

We shall suppose that $G(u)$ is an AP-transform at $u=0$. Then the first formula (3.5.3) can be used for deriving approximations to the first two moments of the TRB-variables. We have

$$x_i=-c_0u_i^k[1+O(u_i)], \tag{7.5.1}$$

$$\lambda_i=-c_0p_i^k[1+O(p_i)]. \tag{7.5.2}$$

Introducing the random variable

$$v_i=\left(\frac{u_i}{p_i}\right)^k-1,$$

we obtain

$$Ex_i=\lambda_i+\lambda_i Ev_i+R_i, \tag{7.5.3}$$

where
$$|R_i|<M|\lambda_i|^{1+1/k},$$

and
$$\operatorname{cov}(x_i,x_j)=\lambda_i\lambda_j\operatorname{cov}(v_i,v_j)+R_{ij},\quad (i\leqslant j), \tag{7.5.4}$$

where
$$|R_{ij}|<M|\lambda_i\lambda_j|(|\lambda_i|^{1/k}+|\lambda_j|^{1/k}).$$

Note that
$$R_i=O(n^{-k-1}),\qquad R_{ij}=O(n^{-2k-1}). \tag{7.5.5}$$

The mean and the variance of v_i can be exactly determined by means of the fr.f. (4.2.2) of u_i. We find, for example,

$$E\,v_i = \frac{(i+k-1)!}{i^k\,(i-1)!}\cdot\frac{(n+1)^k\,n!}{(n+k)!} - 1. \tag{7.5.6}$$

It can be shown that

$$\begin{aligned} |E\,v_i| &< \frac{M}{i} \quad \text{for } i > -k, \\ 0 \leqslant \operatorname{cov}(v_i, v_j) &< \frac{M}{j} \quad \text{for } j \geqslant i > -2k, \end{aligned} \tag{7.5.7}$$

where the constants M do not depend upon i, j, and n. For other values of i and j, the moments do not exist.

An application of the results obtained in this section will be given in 8.11.

7.6. *AP*-transforms. Use of the α, β-correction

As in the case of *AL*-transforms, we can generalize the results obtained in 7.5 by replacing p_i by π_i. In (7.5.3) and (7.5.4) we then write λ_i' instead of λ_i and adjust the definition of v_i accordingly. (7.5.6) is then replaced by

$$E\,v_i' = \frac{(i+k-1)!}{(i-\alpha_i)^k\,(i-1)!}\cdot\frac{(n-\alpha_i-\beta_i+1)^k\,n!}{(n+k)!} - 1. \tag{7.6.1}$$

Arguing in the same way as in 7.4, we choose α_i such that the mean of v_i' is small for large n. It is seen from (7.6.1) that we should then determine α_i so that the first factor in the first term of this expression is equal to unity. This leads to

$$\alpha_i = i - \left(\frac{(i+k-1)!}{(i-1)!}\right)^{1/k}.$$

If, for example, $k=-1$, we get $\alpha_i = 1$, independently of i. When i tends to infinity, α_i tends to the limit $(1-k)/2$ in accordance with (6.7.9). If $G(u)$ is an *AP*-transform at $u=1$, the constant β_{n-i+1} can be determined by similar considerations.

7.7. A generalization

The results obtained in 7.2–7.6 were deduced on the assumption that the *TRB*-variable belongs to an i-sequence, i.e. the rank i was supposed to be a fixed number. It is an important feature of these results that, with a few exceptions, they hold good for more general sequences.

Suppose, in fact, that x_i belongs to *any* c-sequence such that $c=0$ or $c=1$. An examination of the proofs shows that all the results stated in 7.2–7.6 are still valid, except those statements concerning error terms which contain the sample-size n explicitly. The latter remark concerns formulae (7.2.5), (7.2.6), (7.3.5), (7.3.7), etc. It is easily seen how these relations should be modified. Suppose, e.g., that i/n tends to zero as $1/\log n$. It is then recognized that (7.2.6), for example, should be replaced by $R(\omega_i)=O((\log\log n)^{k-2})$. It should also be remarked that the discussion concerning the limit distribution (7.2.7) applies only to the case where i is fixed.

The generalization is of particular interest in the special case where the sequence is at least $n^{-\frac{1}{2}}$-convergent (cf. 3.4). (In the case of less than $n^{-\frac{1}{2}}$-convergent sequences, the extension is of no great value, since more powerful results were obtained in Chapters 5 and 6.)

In this connection, we shall make a special remark concerning the α, β-corrected version of (7.3.8) (cf. the dicussion in 7.4). Introducing the α, β-correction into (7.3.8), we find

$$E x_i = \lambda_i' + R_i', \tag{7.7.1}$$

where, by analogy to (7.3.9),

$$|R_i'| < M\,|\lambda_i'|^{1-1/k}\left(|E\,\omega_i'| + \frac{1}{i}\,|\lambda_i'|^{-1/k}\right). \tag{7.7.2}$$

Let us now perform the generalization referred to above. More precisely, suppose that x_i belongs to a sequence of *TRB*-variables which is at least $n^{-\frac{1}{2}}$-convergent when $i/n\to 0$. We then have from (7.4.4)

$$|E\,\omega_i'| < \begin{cases} \dfrac{M}{i} & \text{for } \alpha_i \neq \frac{1}{2} \\[2ex] \dfrac{M}{i^2} & \text{for } \alpha_i = \frac{1}{2}. \end{cases}$$

Applying the latter result to (7.7.2), we infer that, when the sequence is at least $n^{-\frac{1}{2}}$-convergent, and we choose the special value $\alpha_i=\frac{1}{2}$, the mean of the *TRB*-variable x_i satisfies relation (7.7.1) with

$$|R_i'| < \frac{M}{i^2}\,|\lambda_i'|^{1-1/k} + \frac{M}{i}\,|\lambda_i'|^{1-2/k}, \tag{7.7.3}$$

where the constants M do not depend upon i and n.

This result, which holds true with an obvious modification also when $i/n\to 1$, will be used in 8.9.

CHAPTER 8

LINEAR COMBINATIONS OF TRANSFORMED BETA-VARIABLES

8.1. Introduction

Let $x_{1n}, x_{2n}, \ldots, x_{nn}$ be a set of *TRB*-variables (cf. 3.1). Consider a linear combination

$$T = \sum_{i=1}^{n} g_{in} x_{in} \tag{8.1.1}$$

of these variables, where the coefficients g_{in} are real quantities. To condense the notation, we shall generally omit the index n, always bearing in mind, however, that the coefficients and variables depend both upon i and n.

Linear expressions of the form (8.1.1) are of great importance in the theory of linear estimation. In spite of their practical importance, no general investigation of the properties of such combinations seems to have been undertaken. In the present chapter we shall perform such an analysis, the results of which will be applied to linear estimation problems in Part III.

In 8.2–8.6 we shall derive exact and approximate expressions for the first two moments of T. Three alternatives will be discussed, all of which occur in practice.

(a) The means, variances, and covariances of the *TRB*-variables are accurately known (tabulated).

(b) Only the means are accurately known.

(c) Neither the means nor the variances and covariances are accurately known.

Sections 8.7–8.12 contain a detailed account of the asymptotic behaviour of the error terms in the approximate expressions. The question of the asymptotic distribution of a linear combination is also discussed. In 8.13 some remarks are made concerning linear combinations defined by means of continuous weight functions.

8.2. An alternative form of the linear combination

We shall sometimes find it convenient to replace the variables x_i in (8.1.1) by the weighted differences y_i introduced in 5.5. As will be shown in 8.4, this substitution enables us to derive a useful expression for the variance of T. We write, then, T as a linear combination

$$T = -\sum_{i=0}^{n} h_i y_i \tag{8.2.1}$$

of the $n+1$ variables

$$y_i = \theta_{i+1} x_{i+1} - \theta_i x_i, \quad (i = 0, 1, \dots, n).$$

Here and subsequently we use h_i instead of the complete notation h_{in}.

The coefficients g_i and h_i are evidently connected by the relations

$$g_i = \theta_i (h_i - h_{i-1}), \quad (i = 1, \dots, n), \tag{8.2.2}$$

where the weights θ_i are defined by (5.5.2a) or by the equivalent expressions (5.5.2b). When the coefficients g_i are given and $\theta_i \neq 0$ $(i = 1, \dots, n)$, the quantities h_i are determined, apart from an arbitrary additive constant. In the sequel, we shall use the coefficients g_i and h_i alternately. The fundamental relations (8.2.2) should then always be remembered. Note, however, that the representation (8.2.1) of T is somewhat less general than (8.1.1), since, in the former case, we must suppose that $G(u)$ has a non-zero first derivative.

8.3. Mean and variance of the linear combination

Denoting the mean and the variance of T by $E\,T$ and var T, we have from (8.1.1) the general expressions

$$E\,T = \sum_{i=1}^{n} g_i\, E\, x_i, \tag{8.3.1}$$

$$\operatorname{var} T = \sum_{i,j=1}^{n} g_i g_j \operatorname{cov}(x_i, x_j). \tag{8.3.2}$$

The second expression can, alternatively, be written

$$\operatorname{var} T = \sum_{i,j=0}^{n} h_i h_j \operatorname{cov}(y_i, y_j). \tag{8.3.3}$$

The mean of T can, of course, also be expressed in terms of the quantities h_i, but we shall not need this expression. The expressions given above lose their meaning if any of the moments of the *TRB*-variables do not exist. When this is the case, the linear combination can be replaced by a censored combination which does not contain these variables (cf. 8.6).

8.4. Approximations

When the means or the variances and covariances of the variables x_i are not accurately known, we make use of the approximative procedures developed in Chapter 5, or of the α, β-corrected forms of the approximations discussed in Chapter 6. In this section we shall treat the former of these alternatives.

First, consider the mean of T. Inserting (5.3.1) into (8.3.1), we obtain

$$E\,T = \sum_{i=1}^{n} g_i\, G(p_i) + R_{\text{mean}}, \tag{8.4.1}$$

where by R_{mean} we denote the remainder term

$$R_{\text{mean}} = \sum_{i=1}^{n} g_i\, R_i. \tag{8.4.2}$$

If this term is small enough, the first term in (8.4.1) can be used as an approximation to the mean of T. However, in this section we make no assumption with respect to the magnitude of the remainder term and regard the expression (8.4.1) and the relations to be derived below as purely formal relations.

Secondly, consider the variance of T given by the first alternative formula, (8.3.2). Replacing the variances and covariances of x_i and x_j by the expressions (5.4.1), we obtain

$$\operatorname{var} T = \frac{1}{n+2} \sum_{i,j=1}^{n} g_i\, g_j\, p_i\,(1-p_j)\, G'(p_i)\, G'(p_j) + R_{\text{var}}, \tag{8.4.3}$$

where

$$R_{\text{var}} = \sum_{i,j=1}^{n} g_i\, g_j\, R_{ij}. \tag{8.4.4}$$

Note that the double summation in these expressions should be interpreted as

$$\sum_{i,j=1}^{n} = \sum_{i=j=1}^{n} + 2 \sum_{i<j}.$$

The first term in (8.4.3) is inconvenient for both numerical and analytical investigations, and we shall seldom use it in the sequel. Starting from the second alternative, (8.3.3), and using the expressions (5.5.5), we obtain the alternative formula

$$\operatorname{var} T = \frac{\operatorname{var} h}{n+2} + R_{\text{var}}, \tag{8.4.5}$$

where

$$\operatorname{var} h = \frac{1}{n+1} \sum_{i=0}^{n} (h_i - \bar{h})^2, \tag{8.4.6}$$

$$\bar{h} = \frac{1}{n+1} \sum_{i=0}^{n} h_i,$$

and

$$R_{\text{var}} = \sum_{i,j=0}^{n} h_i h_j \Delta_{ij}. \tag{8.4.7}$$

The letter h in var h may be regarded, e.g., as symbolizing the row vector $(h_0, h_1, \ldots, h_n)$. The remark concerning the summation made after formula (8.4.4) is relevant also in this case. The quantities Δ_{ij} in (8.4.7) are the error terms appearing in (5.5.5). Thus we have proved that, if the error term is discarded, the variance of T is proportional to the variance of the coefficients $h_0, \ldots, h_n$. The formula (8.4.5) is remarkably simple and will be used extensively in Part III.

We shall also consider the covariance of two linear combinations

$$T_1 = \sum_{i=1}^{n} g_{1i} x_i, \quad T_2 = \sum_{i=1}^{n} g_{2i} x_i. \tag{8.4.8}$$

Defining h_{1i} and h_{2i} by relations of the same kind as (8.2.2), we obtain an expression of the same simplicity as (8.4.5), viz.

$$\operatorname{cov}(T_1, T_2) = \frac{\operatorname{cov}(h_1, h_2)}{n+2} + R_{\text{cov}}, \tag{8.4.9}$$

where

$$\operatorname{cov}(h_1, h_2) = \frac{1}{n+1} \sum_{i=0}^{n} (h_{1i} - \bar{h}_1)(h_{2i} - \bar{h}_2), \tag{8.4.10}$$

$$\bar{h}_r = \frac{1}{n+1} \sum_{i=0}^{n} h_{ri}, \quad (r = 1, 2),$$

and R_{cov} is a remainder term given by

$$R_{\text{cov}} = \sum_{i,j=0}^{n} h_{1i} h_{2j} \Delta_{ij}. \tag{8.4.11}$$

Finally, it should be noted that, when $G(u)$ is specialized to the inverse of the rectangular distribution, the remainder terms in all the expressions given above are zero (cf. 4.2 and 5.5).

8.5. Use of the α, β-correction

The approximations given in the preceding section can often be improved by introducing the α, β-correction discussed in Chapter 6. It is easily seen how the formulae should then be modified.

We first consider the mean and variance expressions (8.4.1) and (8.4.3). As seen from (6.2.1) and (6.3.1), we have only to replace p_i by π_i, given by (6.2.2), and n by n', given by (6.2.3). The mean value formula, for example, becomes

$$E\,T = \sum_{i=1}^{n} g_i\, G(\pi_i) + R'_{\text{mean}}, \tag{8.5.1}$$

where

$$R'_{\text{mean}} = \sum_{i=1}^{n} g_i\, R'_i\,. \tag{8.5.2}$$

Secondly, consider the alternative variance formula (8.4.5). Note that the expressions derived in 8.2 and 8.3 hold good without any modification, if the quantities y_i are defined as the generalized weighted differences introduced in 6.4. Thus we may apply the results obtained in that section. We find after some calculation that, *if the sum* $\alpha_i + \beta_i$ *does not depend on* i, then

$$\text{var}\, T = \frac{V(h, \rho)}{n'+2} + R_V, \tag{8.5.3}$$

where

$$V(h, \rho) = \frac{1}{n'+1}\left[\sum_{i=0}^{n} \rho_i h_i^2 - \frac{1}{n'+1}\left(\sum_{i=0}^{n} \rho_i h_i\right)^2\right] \tag{8.5.4}$$

is a 'weighted' variance of the h_i's, and

$$R_V = \sum_{i,j=0}^{n} h_i\, h_j\, \Delta'_{ij}\,. \tag{8.5.5}$$

The quantities ρ_i are defined by (6.4.2), and the components Δ'_{ij} of R_V are the error terms in (6.4.1). The letter ρ in $V(h, \rho)$ symbolizes the row vector $(\rho_0, \rho_1, \ldots, \rho_n)$. We see that, when $\rho_i = 1$ for all i, formula (8.5.3) reduces to (8.4.5).

It is realized that the expression (8.4.9) for the covariance of two linear combinations can be generalized in a similar way by means of the α, β-correction.

The expressions (8.5.1) and (8.5.3) are very general, since they are valid even when the quantities α_i depend upon both i and n. In the remaining part of the section, we shall make some comments upon the simple case where $\alpha_i = \alpha$, $\beta_i = \beta$, independently of i and n. We notice that the quantities ρ_i appearing in (8.5.4) are then given by the simple expressions (6.4.3). Then the only difference between (8.4.5) and (8.5.3) is that, in the latter expression, n' is used instead of n, and, furthermore, that weights are used for $i = 0$ and $i = n$.

The constants α and β ought to be chosen, so that the leading terms in the expressions for the mean and the variance of T are as good approximations to these quantities as possible.

Consider, e.g., (8.5.1). We saw in the examples in Chapter 6 that it is sometimes possible to determine α and β such that $G(\pi_i)$ is a good overall approximation to $E\,x_i$. When such values of α and β are known, they should, of course, be used also in (8.5.1). Theoretically, we might go a step further and adjust these values of α and β, so that the term of order n^{-1} in R_{mean} is zero. As seen from (6.5.5), this leads to the condition

$$\sum_{i=1}^{n} g_i\, Q(\pi_i;\, \alpha, \beta) = 0.$$

In the symmetrical case, this relation could even be used for an exact determination of $\alpha = \beta$. From a computational point of view, the relation is, however, generally too complicated to be used for this purpose.

Similarly, α and β in (8.5.3) might, theoretically, be determined by analyzing the error term in this formula. The computational difficulties are, however, still greater than in the previous case, and no attempt has been made to apply the idea to a numerical example. In the applications of the theory made in Part III, we shall either use the variance formula without any α, β-correction or, if this correction is used in the mean value formula, with the same correction in both expressions.

Summing up the contents of 8.3–8.5, we now have the following tools with which to handle the three situations mentioned in 8.1. When tables of the means and covariances of the *TRB*-variables are available, the first two moments of T can be determined from (8.3.1) and (8.3.2). If only the means are accurately known, (8.3.2) is replaced by the approximation afforded by the leading term of (8.4.3) or by the equivalent expression (8.4.5). When neither the means nor the covariances are accurately known, we also replace (8.3.1) by an approximation, namely by the leading term of (8.4.1). The accuracy is often improved by introducing a suitable α, β-correction into the leading terms.

8.6. A note concerning censored linear combinations

If some of the coefficients in the linear combination (8.1.1) are zero, we have a *censored linear combination.*

The expressions for the mean and the variance of a linear combination given in 8.3–8.5 are, of course, valid also for censored combinations. It should only be noted that, in the expressions containing the coefficients h_i explicitly, we must, for any zero coefficients $g_i, g_j, g_k, \dots$, impose the conditions

$$h_i = h_{i-1}, \quad h_j = h_{j-1}, \quad h_k = h_{k-1}, \dots \tag{8.6.1}$$

8.7. Bounded transforms. Order of remainder terms

Up to the present, the discussion has been given in terms of a general transform. In the remaining part of the chapter, we shall investigate the approach to zero of the remainder terms appearing in the formulae deduced in the preceding sections. The question of the asymptotic distribution of a linear combination will also be discussed. In order to arrive at useful results, certain conditions must be imposed upon $G(u)$. In this and the following section, $G(u)$ will be assumed to be bounded; later, we shall remove this condition.

THEOREM. *If $G(u)$ satisfies Condition A in 3.5 with $\mathfrak{C}$ equal to the closed interval $0 \leqslant u \leqslant 1$, then the remainder terms in the mean value formula (8.4.1) and the equivalent variance formulae (8.4.3) and (8.4.5) satisfy the relations*

$$\begin{aligned} R_{\text{mean}} &= O\left(\frac{1}{n}\sum_{i=1}^{n} |g_i|\right), \\ R_{\text{var}} &= O\left(\frac{1}{n}\sum_{i=1}^{n} |g_i|\right)^2. \end{aligned} \tag{8.7.1}$$

If, in addition, constants M and d can be found such that

$$|g_i| < \frac{M}{n^{d-1}} \tag{8.7.2}$$

for any $i = 1, \dots, n$ and

$$|g_i| < \frac{M}{n^{d}} \tag{8.7.2 a}$$

for all but a finite number of ranks i, then

$$R_{\text{mean}} = O\left(\frac{1}{n^d}\right), \tag{8.7.3}$$

$$R_{\text{var}} = O\left(\frac{1}{n^{2d}}\right). \tag{8.7.4}$$

The proof is simple. Consider, for example, the mean value formula (8.4.1). We apply the theorem in 5.3, p. 53, with $\mathfrak{C}$ equal to the unit interval. Since by (8.4.2) the remainder term R_{mean} is a linear combination of the errors R_i, we immediately obtain the first relation (8.7.1). Similarly, the second relation (8.7.1) is proved by means of the theorem in 5.4, p. 55. Finally, we obtain (8.7.3) and (8.7.4) by applying (8.7.2) and (8.7.2 a) to the relations (8.7.1).

We shall make some comments upon the theorem. First, it should be remarked that there is a similar result as (8.7.4) which holds good for the remainder term R_{cov} in the formula (8.4.9) for the covariance of two linear combinations T_1 and T_2. If the coefficients of both these combinations satisfy the relations (8.7.2) and (8.7.2 a), we find by a slight extension of the proof that

$$R_{\text{cov}} = O\left(\frac{1}{n^{2d}}\right).$$

Secondly, it should be mentioned that (8.7.3) holds good even if $G'(u)$ has a finite number of discontinuity points in the unit interval, provided, however, that (8.7.2 a) is satisfied for *any* i. To prove this statement, surround each discontinuity point w of $G'(u)$ by an interval with the breadth 2δ. Apply the result in the final paragraph of 5.3, p. 55, to those *TRB*-variables in the set which belong to c-sequences such that $|c-w|<\delta$, and treat the remaining variables as before. We then obtain

$$|R_{\text{mean}}| < \frac{M}{\sqrt{n}}\Sigma|g_i| + \frac{M}{n}\Sigma|g_i|,$$

where the first sum contains approximately $2\delta n$ terms. Now use (8.7.2 a) and take $\delta = n^{-\frac{1}{2}}$. (This choice of δ is allowed, for the variables corresponding to the second term then belong to c-sequences which are less than $n^{-\frac{1}{2}}$-convergent in the singular point w.) Hence we conclude that R_{mean} has at most the order n^{-d} as stated.

It would be of great value if the scope of (8.7.4) could be enlarged in a similar manner. Unfortunately, this does not seem possible, the continuity of $G'(u)$ being essential for the validity of the theorem in 5.4.

Thirdly, we shall comment upon the use of (8.7.3) and (8.7.4) in practice. It should be remembered that, in all situations of practical interest, the leading terms in the expressions (8.4.3) and (8.4.5) tend to zero when n tends to infinity. We must therefore require that R_{var} *should tend to zero more rapidly than the leading term.* Thus we find that (8.7.4) is useful if the leading term var $h/(n+2)$ in (8.4.5) converges to zero less rapidly than n^{-2d}. This is, indeed, a very common case, but there are also other possibilities.

Suppose, for instance, that all but a finite number of the quantities h_i are equal. The example $h_0=1,\ h_n=1,\ h_1=\cdots=h_{n-1}=0$ is a case in point (cf. 9.2, final paragraphs). Then $d=1$, and var $h/(n+2)$ has the order n^{-2}, and thus (8.7.4) is not powerful enough. By similar examples it may be shown that (8.7.3) is not always applicable. To overcome these difficulties, we shall determine the order of magnitude of the error terms in a different way. The assumptions concerning the transform made in the above theorem will then be replaced by slightly different conditions.

THEOREM. *If* $G(u)$ *satisfies Condition* B *in* 3.5 *with* $\mathfrak{C}$ *equal to the closed interval* $0\leqslant u\leqslant 1$, *then, for any choice of the coefficients* g_i, *the remainder terms in the mean value formula* (8.4.1) *and the equivalent variance formulae* (8.4.3) *and* (8.4.5) *satisfy the relations*

$$R_{\text{mean}}=O\left(\frac{(\text{var }h)^{\frac{1}{2}}}{n}\right), \tag{8.7.5}$$

$$R_{\text{var}}=O\left(\frac{\text{var }h}{n^{\frac{3}{2}}}\right). \tag{8.7.6}$$

The theorem is proved by applying the results obtained in 5.5 concerning the weighted differences y_i. First, we observe that, by (5.5.3) and (8.4.1) (cf. also the proof of the theorem in 8.8),

$$R_{\text{mean}}=-\sum_{i=0}^{n} h_i\, E\,\eta_i. \tag{8.7.7}$$

Hence by the first inequality (5.5.4)

$$|R_{\text{mean}}|<\frac{M}{n^2}\sum_{i=0}^{n}|h_i|. \tag{8.7.8}$$

We may replace h_i by $h_i - \bar{h}$, since these coefficients depend upon an arbitrary additive factor. By an application of Cauchy's inequality we then obtain (8.7.5).

Secondly, applying (5.5.6) to the representation (8.4.7) of R_{var}, we find

$$|R_{\text{var}}| < \frac{M}{n^{5/2}} \left(\sum_{i=0}^{n} h_i^2 + \frac{1}{n} \sum_{i \neq j} |h_i h_j| \right).$$

By Cauchy's inequality

$$|R_{\text{var}}| < \frac{2M}{n^{5/2}} \sum_{i=0}^{n} h_i^2.$$

Replacing h_i by $h_i - \bar{h}$, we obtain (8.7.6), and the theorem is proved.

As a consequence of the theorem, (8.4.5) can be given the convenient form

$$\operatorname{var} T = \frac{\operatorname{var} h}{n+2} \left[1 + O\left(\frac{1}{\sqrt{n}} \right) \right]. \tag{8.7.9}$$

This formula evidently solves the difficulty referred to earlier. It shows that R_{var} tends to zero faster than the leading term, regardless of the order of this term. It should, however, be remembered that, to arrive at this result, we have used somewhat more restrictive assumptions concerning $G(u)$ than before.

The considerations which led to (8.7.6) can, of course, be extended to the error term in (8.4.9), which then takes the form

$$R_{\text{cov}} = O\left(\frac{(\operatorname{var} h_1)^{\frac{1}{2}} (\operatorname{var} h_2)^{\frac{1}{2}}}{n^{\frac{3}{2}}} \right). \tag{8.7.10}$$

Finally, we point out that the results obtained in this section may easily be extended so as to apply to the α, β-corrected expressions (8.5.1) and (8.5.3) for the mean and the variance of T.

8.8. Bounded transforms. Asymptotic distribution of a linear combination

We shall now show by the method used in 8.7 that, given certain conditions, a linear combination is asymptotically normally distributed. Set

$$\varrho_r = \sum_{i=0}^{n} |h_i - \bar{h}|^r. \tag{8.8.1}$$

Theorem. *If $G(u)$ satisfies Condition B in 3.5 with $\mathfrak{C}$ equal to the closed interval $0 \leqslant u \leqslant 1$, and if, in addition,*

$$\lim_{n\to\infty} \frac{\varrho_3}{\varrho_2^{\frac{3}{2}}} = 0, \qquad (8.8.2)$$

then

$$T = \sum_{i=1}^{n} g_i x_i$$

is asymptotically normally distributed with the mean $\sum_{i=1}^{n} g_i G(p_i)$ *and the variance* var h/n.

The relation between the coefficients g_i and h_i is as usual given by (8.2.2). Before proving the theorem, we point out that we may of course, if we please, introduce the quantities h_i also in the expression for the mean of T.

For the proof, we write as in 8.2

$$T = -\sum_{i=0}^{n} h_i y_i$$

and apply the partition (5.5.3) to each component y_i. Hence

$$T = -\sum_{i=0}^{n} h_i m_i - \sum_{i=0}^{n} h_i \left(\delta_i - \frac{1}{n+1}\right) - \sum_{i=0}^{n} h_i \eta_i .$$

Replacing in the first term h_i by g_i and making a rearrangement, we obtain

$$-\left(\frac{\text{var } h}{n}\right)^{-\frac{1}{2}} \left[T - \sum_{i=1}^{n} g_i G(p_i)\right] = Z_1 + Z_2,$$

where

$$Z_1 = \left(\frac{\text{var } h}{n}\right)^{-\frac{1}{2}} \sum_{i=0}^{n} h_i \left(\delta_i - \frac{1}{n+1}\right),$$

$$Z_2 = \left(\frac{\text{var } h}{n}\right)^{-\frac{1}{2}} \sum_{i=0}^{n} h_i \eta_i .$$

We first prove that Z_2 converges in probability to zero. Applying the last two results (5.5.4) and Cauchy's inequality, we find after a simple calculation (cf. the estimation of R_{var} in 8.7) that

$$\text{var } Z_2 = O\left(\frac{1}{n}\right).$$

Further, by (8.7.7) and (8.7.5)

$$E\, Z_2 = O\left(\frac{1}{\sqrt{n}}\right).$$

The truth of the assertion then follows from Tchebycheff's theorem (Cramér, 1946 *a*, p. 253). Hence the standardized variable $-T$ behaves asymptotically in the same way as Z_1.

To determine the asymptotic distribution of Z_1, we consider the linear combination

$$\zeta = \left(\sum_{i=0}^{n} t_i^2\right)^{-\frac{1}{2}} \sum_{i=0}^{n} t_i z_i,$$

where the variables z_i are independent, each having the fr.f. $e^{-(x+1)}$, $(x \geqslant -1)$. Applying the Central Limit Theorem in the form given by Liapounoff (see Cramér, 1946 *a*, p. 215 ff.), we see that, if

$$\lim_{n\to\infty} \frac{\sum_{i=0}^{n} |t_i|^3}{\left(\sum_{i=0}^{n} t_i^2\right)^{\frac{3}{2}}} = 0, \tag{8.8.3}$$

then ζ is asymptotically normally distributed with mean 0 and variance 1.

Let us now compare the distributions of ζ and Z_1 in the special case $t_i = h_i - \bar{h}$. Condition (8.8.2) then ensures that (8.8.3) holds good; hence ζ is asymptotically normal. With respect to Z_1, we use the result, proved in 5.5, that any finite number of variables $n\,\delta_i$ are, in the limit, independent and exponentially distributed with the fr.f. e^{-x}. Consequently, any moment of Z_1 coincides, in the limit, with the corresponding moment for ζ. As the normal distribution is uniquely determined by its moments, it follows that, when ζ is asymptotically normal, this must also be true of Z_1. This completes the proof.

A consequence of the proof is that a linear combination of the type studied in this section behaves asymptotically in the same way as a weighted sum of independent χ^2-variables, each with two degrees of freedom.

It should be remarked that the conditions of the theorem can undoubtedly be weakened. This observation raises an interesting problem, which, however, will not be discussed here, the result established above being sufficiently general for our purposes.

Note that the asymptotic distribution of a linear combination is not always normal. As a simple example, take $G(u) = u$ and $h_0 = h_n = 1$, $h_1 = \cdots = h_{n-1} = 0$. It is seen that (8.8.2) then breaks down. We have in this case $T = u_n - u_1$; hence we see that $n(1-T)$ is, in the limit, gamma-distributed with the fr.f. $x\,e^{-x}$, $(x \geqslant 0)$.

We shall also consider the asymptotic joint distribution of two linear combinations of the form (8.4.8). Set

$$\varrho_{rs} = \sum_{i=0}^{n} |h_{1i} - \bar{h}_1|^r \, |h_{2i} - \bar{h}_2|^s. \tag{8.8.4}$$

THEOREM. *If $G(u)$ satisfies Condition B in 3.5 with $\mathfrak{C}$ equal to the closed interval $0 \leqslant u \leqslant 1$, and if, in addition,*

$$\lim_{n\to\infty} \frac{\varrho_{rs}}{\varrho_{20}^{r/2}\, \varrho_{02}^{s/2}} = 0 \tag{8.8.5}$$

for any non-negative integers r, s with the sum $r+s=3$, then

$$T_1 = \sum_{i=1}^{n} g_{1i}\, x_i, \quad T_2 = \sum_{i=1}^{n} g_{2i}\, x_i$$

are asymptotically jointly normally distributed with means $\sum_{i=1}^{n} g_{ri}\, G(p_i)$, $(r=1, 2)$, variances var h_r/n, $(r=1, 2)$, *and covariance* cov $(h_1, h_2)/n$.

This theorem is proved by a slight extension of the method used by Liapounoff in his proof of the Central Limit Theorem.

8.9. *AL*-transforms. Order of remainder terms

An essential condition for the validity of the results established in 8.7 and 8.8 is that the transform should be bounded in the unit interval and possess a sufficient number of bounded derivatives. In this and the following three sections, we shall study the behaviour of the linear combination when these conditions are replaced by weaker assumptions.

Beginning with *AL*-transforms, we shall prove the following theorem. The symbol δ used in the theorem denotes a small quantity > 0.

THEOREM. *If*

a) *$G(u)$ satisfies Condition A in 3.5 with $\mathfrak{C}$ equal to the half-open interval $0 < u \leqslant 1$,*

b) *$G(u)$ is an AL-transform at $u=0$,*

c) *constants M, d, and m can be found such that $mk \geqslant 1$ and, for any $i = 1, \ldots, n$,*

$$|g_i| < \frac{M}{n^d} |\lambda_i^\circ|^{m-1}, \tag{8.9.1}$$

where $\lambda_i^\circ = \lambda_i = G(p_i)$ *when* $i/n < \delta$, *and* $\lambda_i^\circ = 1$ *otherwise,*

then the remainder terms in the mean value formula (8.4.1) *and the equivalent variance formulae* (8.4.3) *and* (8.4.5) *satisfy the relations*

$$R_{\text{mean}} = O\left(\frac{1}{n^d} (\log n)^{mk}\right), \tag{8.9.2}$$

$$R_{\text{var}} = O\left(\frac{1}{n^{2d-\frac{3}{4}}} (\log n)^{mk-1}\right). \tag{8.9.3}$$

For other values of mk, *the relations still hold, with the modification that the exponents of* $\log n$ *in* (8.9.2) *and* (8.9.3) *should be put equal to zero when* $mk < 0$ *or* $mk < 1$, *respectively.*

If $G(u)$ is an AL-transform at $u = 1$, the conditions should be slightly reformulated. (In $i/n < \delta$, the letter i should then be replaced by $n - i$.)

To prove the theorem, we divide the linear combination (8.1.1) into two parts

$$T = T_0 + T_1, \tag{8.9.4}$$

where

$$T_0 = \sum_{i=1}^{s} g_i x_i, \qquad T_1 = \sum_{i=s+1}^{n} g_i x_i \tag{8.9.5}$$

and $s = [\delta n^{\frac{1}{2}+\varepsilon}]$. Here ε is an arbitrarily small positive quantity.

Let us determine the mean of T by making separate calculations of the means of T_0 and T_1. The latter combination contains non-singular TRB-variables, for, by the choice of s, all variables belonging to c-sequences with $c = 0$ are less than $n^{-\frac{1}{2}}$-convergent (cf. 3.4). The conditions of the theorem in 5.3, p. 53, are evidently fulfilled, and thus we find by applying this theorem to each component of T_1

$$E\,T_1 = \sum_{i=s+1}^{n} g_i\, G(p_i) + R_1, \tag{8.9.6}$$

where, by analogy to the first relation (8.7.1),

$$R_1 = O\left(\frac{1}{n} \sum_{i=s+1}^{n} |g_i|\right). \tag{8.9.7}$$

By condition (c)

$$R_1 = O\left(\frac{1}{n^{d+1}} \sum_{i=s+1}^{n} |\lambda_i^\circ|^{m-1}\right).$$

Since the sum
$$\frac{1}{n}\sum_{i=s+1}^{n}|\lambda_i^{\circ}|^{m-1}$$

remains finite when n approaches infinity, we conclude that

$$R_1 = O\left(\frac{1}{n^d}\right). \qquad (8.9.8)$$

Next, consider T_0, which is a sum of singular *TRB*-variables. In view of the remark made in 7.7, the results stated in 7.3 apply to these variables. Hence by (7.3.8)

$$E\,T_0 = \sum_{i=1}^{s} g_i\,G(p_i) + R_0, \qquad (8.9.9)$$

where, as seen from condition (8.9.1) and formula (7.3.9 a),

$$|R_0| < \frac{M}{n^d}\sum_{i=1}^{s}\frac{1}{i}\,|\lambda_i|^{m-1/k}.$$

It follows from (7.2.2) that the sum in the right member has the same order of magnitude as the sum

$$\sum_{i=1}^{s}\frac{1}{i}\left(\log\frac{1}{p_i}\right)^{mk-1}.$$

Hence

$$R_0 = O\left(\frac{1}{n^d}(\log n)^{mk}\right). \qquad (8.9.10)$$

Adding (8.9.6) and (8.9.9), we obtain the mean value formula (8.4.1) with an error term, which, as seen from (8.9.8) and (8.9.10), satisfies formula (8.9.2). Thus the first statement in the theorem is proved.

In the remaining part of the proof, we shall make a similar examination of the variance formula (8.4.5). T is divided into two parts T_0 and T_1 exactly as before. We have

$$\operatorname{var} T = \operatorname{var} T_0 + \operatorname{var} T_1 + 2\,\vartheta\,(\operatorname{var} T_0 \operatorname{var} T_1)^{\frac{1}{2}}, \qquad (8.9.11)$$

where
$$|\vartheta| \leqslant 1.$$

It follows from (8.4.5) and from the note in 8.6 concerning censored combinations that

$$\operatorname{var} T_1 = \frac{\operatorname{var}_c h}{n+2} + R_{\operatorname{var}_c}. \qquad (8.9.12)$$

Here $\text{var}_c h$ is the "censored" variance of the coefficients $h_0, \ldots, h_n$ obtained by setting

$$h_0 = h_1 = \cdots = h_s$$

in (8.4.6).

We shall determine the order of magnitude of the leading term in (8.9.12), which can be done in different ways. We may, for example, write this term in the form of a double sum similar to the leading term in (8.4.3), but with an obvious restriction on the summation. Applying the first formula (3.5.2) to this sum and using condition (c), we obtain after some calculation

$$\frac{\text{var}_c h}{n+2} = O\left(\frac{1}{n^{2d-1}}\right). \tag{8.9.13}$$

Further, by analogy to the second result (8.7.1),

$$R_{\text{var}_c} = O\left(\left(\frac{1}{n} \sum_{i=s+1}^{n} |g_i|\right)^2\right).$$

Proceeding as in the discussion concerning the mean of T_1, we find

$$R_{\text{var}_c} = O\left(\frac{1}{n^{2d}}\right). \tag{8.9.14}$$

Next, consider the variance of T_0. Applying (7.3.14) to the singular TRB-variables entering into T_0, we find, using condition (c),

$$\text{var } T_0 < \frac{M}{n^{2d}} \sum_{1 \leqslant i \leqslant j \leqslant s} \frac{1}{j} |\lambda_i \lambda_j|^{m-1/k}.$$

By (7.2.2) the right member has the same order as

$$\frac{1}{n^{2d}} \sum_{1 \leqslant i \leqslant j \leqslant s} \frac{1}{j} \left(\log \frac{1}{p_i} \log \frac{1}{p_j}\right)^{mk-1},$$

which is less than

$$\frac{s}{n^{2d}} [\log (n+1)]^{2mk-2}.$$

Since $s = [\delta n^{\frac{1}{2}+\varepsilon}]$ and ε is arbitrarily small, we obtain

$$\text{var } T_0 = O\left(\frac{1}{n^{2d-\frac{1}{2}}} (\log n)^{2mk-2}\right). \tag{8.9.15}$$

Inserting (8.9.12) and (8.9.15) into (8.9.11), and using (8.9.13), (8.9.14), (8.9.15), and the condition $mk \geqslant 1$, we find after some rearrangement

$$\operatorname{var} T = \frac{\operatorname{var}_c h}{n+2} + O\left(\frac{1}{n^{2d-\frac{1}{2}}}(\log n)^{mk-1}\right). \tag{8.9.16}$$

If $mk<1$, the exponent of $\log n$ should be put equal to zero. It is intuitively clear and can, in fact, easily be rigorously proved that the statement concerning the error term is not affected if $\operatorname{var} h$ is substituted for $\operatorname{var}_c h$ in the leading term of this relation. This proves the theorem.

It should be pointed out that the theorem applies also to the α, β-corrected expressions for the mean and the variance of T given in 8.5. Let us, for simplicity, suppose that $\alpha_i = \alpha$, $\beta_i = \beta$, independently of i and n, and, furthermore, that $\alpha<1$, $\beta<1$. If, in addition, the conditions of the theorem are fulfilled with the obvious modification that, in (8.9.1), λ_i is replaced by $\lambda_i' = G(\pi_i)$, then R'_{mean} *in* (8.5.1) *and* R_V *in* (8.5.3) *satisfy the relations* (8.9.2) *and* (8.9.3). These two statements are proved entirely as above.

The first statement, however, deserves special comment. It is, indeed, a remarkable fact that, if $G(u)$ is an AL-transform, and *we choose the special value* $\alpha = \frac{1}{2}$, the remainder term R'_{mean} in (8.5.1) satisfies a somewhat stronger relation than (8.9.2), viz.

$$R'_{\text{mean}} = O\left(\frac{1}{n^d}(\log n)^{mk-1}\right), \qquad (mk \geqslant 1). \tag{8.9.17}$$

To prove this, only a slight modification of the proof given above is required. Applying (7.7.1) to each TRB-variable in T_0, we obtain, instead of (8.9.9), the following expression

$$E\,T_0 = \sum_{i=1}^{s} g_i\, G(\pi_i) + R_0',$$

where by (8.9.1) and (7.7.3)

$$|R_0'| < \frac{M}{n^d}\left[\sum_{i=1}^{s} \frac{1}{i^2}|\lambda_i'|^{m-1/k} + \sum_{i=1}^{s} \frac{1}{i}|\lambda_i'|^{m-2/k}\right].$$

Hence

$$R_0' = O\left(\frac{1}{n^d}(\log n)^{mk-1}\right).$$

Proceeding in other respects as before, we obtain (8.5.1) with the error term satisfying (8.9.17).

As, consequently, the α, β-correction affects the upper limit of the order of the remainder term in the mean value formula, this correction has a deeper significance than might possibly have been anticipated when it was first introduced in Chapter 6.

Though of wide application, the theorem proved above is, like the corresponding theorem in 8.7, p. 93, not always sufficiently general. Particularly, the formulation of condition (c) puts a limit on the use of the theorem (cf. 13.6). For this reason, we shall prove a result which is more general in this respect. On the other hand, we must then impose somewhat more restrictive conditions upon $G(u)$ than before.

THEOREM. *If*

(a) $G(u)$ *satisfies Condition B in* 3.5 *with* $\mathfrak{C}$ *equal to the half-open interval* $0 < u \leqslant 1$,

(b) $G(u)$ *is an AL-transform at* $u=0$,

(c) *constants* M, d, *and* m *can be found such that*

$$\operatorname{var} h = O\left(\frac{1}{n^{2d-2}}\right), \tag{8.9.18}$$

and, further, for any $i \leqslant [\sqrt{n}]$

$$|g_i| < \frac{M}{n^d} |\lambda_i|^{m-1}, \tag{8.9.19}$$

then formulae (8.9.2) *and* (8.9.3) *are still valid.*

An important feature of this theorem is that only coefficients g_i with low indices are assumed to have an upper bound. The theorem is proved by a combination of the methods used in the proofs of the second theorem in 8.7, p. 93, and the theorem stated earlier in this section.

Consider, for example, the mean of T. We start as before from the partition (8.9.4). The second part, T_1, satisfies (8.9.6), where by the remark in 8.6 concerning censored combinations and by (8.7.5)

$$R_1 = O\left(\frac{1}{n} (\operatorname{var}_c h)^{\frac{1}{2}}\right).$$

Here $\operatorname{var}_c h$ has the same meaning as on p. 100. As this quantity is asymptotically equivalent to $\operatorname{var} h$, we obtain by condition (c)

$$R_1 = O\left(\frac{1}{n^d}\right).$$

The rest of the proof concerning the mean of T is the same as before.

The statement regarding the variance of T is proved in an analogous way by means of (8.7.6).

In the two theorems proved in this section we have considered the mean and the variance of a single linear combination. It should be pointed out that the result (8.9.3) holds good also for R_{cov} in the formula (8.4.9) for the covariance of two linear combinations, provided that the coefficients in each combination satisfy (8.9.1) or, in the second theorem, (8.9.18) and (8.9.19). This is proved by a straightforward extension of the methods used in the section.

8.10. *AL*-transforms. Asymptotic distribution of a linear combination

Let as in the preceding section $G(u)$ be an AL-transform. The question of the asymptotic distribution of a linear combination is answered by the following theorem, which is sufficiently general for our purposes.

THEOREM. *If*

(a) $G(u)$ *satisfies Condition B in 3.5 with* $\mathfrak{C}$ *equal to the half-open interval* $0 < u \leqslant 1$,

(b) $G(u)$ *is an AL-transform at* $u = 0$,

(c) *constants* M *and* m *can be found such that, for any* $i = 1, \ldots, n$,

$$|g_i| < \frac{M}{n} |\lambda_i^\circ|^{m-1}, \tag{8.10.1}$$

(d) *condition* (8.8.2) *is satisfied, and, in addition,* $\operatorname{var} h$ *remains greater than zero when* n *approaches infinity,*

then T *is asymptotically normally distributed with the mean* $\sum_{i=1}^{n} g_i G(p_i)$ *and the variance* $\operatorname{var} h/n$.

We shall give a sketch of the proof. Write

$$\left(\frac{\operatorname{var} h}{n}\right)^{-\frac{1}{2}} [T - \sum_{i=1}^{n} G(p_i)] = \left(\frac{\operatorname{var} h}{n}\right)^{-\frac{1}{2}} (T_0 - ET_0) +$$

$$+ \left(\frac{\operatorname{var} h}{n}\right)^{-\frac{1}{2}} R_0 + \left(\frac{\operatorname{var} h}{n}\right)^{-\frac{1}{2}} [T_1 - \sum_{i=s+1}^{n} g_i G(p_i)],$$

where T_0 and T_1 are defined by (8.9.5), and R_0 is the remainder term in (8.9.9).

Using (8.9.15) with $d=1$, we see that the variance of the first term in the right member has at most the same order as the expression

$$\frac{1}{\operatorname{var} h} \frac{(\log n)^{2mk-2}}{n^{\frac{1}{2}}}.$$

Thus, by condition (d) the variance tends to zero when $n \to \infty$. Hence by Tchebycheff's theorem the first term tends to zero in probability. It is also inferred from (8.9.10) with $d=1$ that the second term has the same property.

By a method quite similar to that developed in 8.8, it is demonstrated that T_1 is asymptotically normal. The variance of T_1 is given by (8.9.12), and is thus asymptotically equivalent to var h/n.

As in 8.8 there is a corresponding theorem for the asymptotic joint distribution of two linear combinations. Since the formulation of this theorem presents no great difficulties, it will not be given here.

8.11. *AP*-transforms. Order of remainder terms

We shall state a theorem valid for *AP*-transforms, which is very similar to the first theorem in 8.9. It should, however, be noted in this case that if, in the definition (3.5.3) of the transform, the constant k satisfies $k \leqslant -1$, the linear combination must be censored, since otherwise $E x_1$ or $E x_n$ and hence $E T$ do not exist. Similarly, var T does not exist for $k \leqslant -\frac{1}{2}$ unless the combination is censored. λ_i° is defined as in the first theorem in 8.9, p. 98.

THEOREM. *If*

(a) $G(u)$ *satisfies Condition A in 3.5 with* $\mathfrak{C}$ *equal to the half-open interval* $0 < u \leqslant 1$,

(b) $G(u)$ *is an AP-transform at* $u=0$,

(c) *the coefficients* g_i *are zero for* $i \leqslant -2k$ *and, further, constants* M, d, *and* m *can be found such that* $mk > -\frac{1}{2}$ *and, for any* $i > -2k$,

$$|g_i| < \frac{M}{n^d} |\lambda_i^\circ|^{m-1}, \tag{8.11.1}$$

then the remainder terms in the mean value formula (8.4.1) *and the equivalent variance formulae* (8.4.3) *and* (8.4.5) *satisfy the relations*

$$R_{\text{mean}} = \begin{cases} O(n^{-d}) & \text{for } mk > 0 \\ O(n^{-d}\log n) & \text{for } mk = 0 \\ O(n^{-mk-d}) & \text{for } -\tfrac{1}{2} < mk < 0, \end{cases} \tag{8.11.2}$$

$$R_{\text{var}} = \begin{cases} O\left(n^{-2d+\frac{3}{4}}\right) & \text{for } mk \geqslant 0 \\ O\left(n^{-\frac{mk}{2}-2d+\frac{3}{4}}\right) & \text{for } -\tfrac{1}{2} < mk < 0. \end{cases} \tag{8.11.3}$$

When $G(u)$ is an AP-transform at $u=1$, or at both $u=0$ and $u=1$, there is an obvious change in the formulation of the theorem.

The method of proof is entirely the same as that used in the proof of the first theorem in 8.9, and we shall therefore omit most details. We define T_0 and T_1 as before. Note that, as stated in condition (c), certain of the coefficients g_i may be zero. We also observe that, in view of the remark in 7.7, the results obtained in 7.5 can be applied to the present problem.

Formulae (8.9.6) and (8.9.7) are valid even in this case. By (7.5.2), the remainder term R_1 has at most the same order as S_1, where

$$S_1 = \frac{1}{n^{d+1}} \sum_{i=s+1}^{n} \left(\frac{i}{n}\right)^{k(m-1)}.$$

Furthermore, we obtain (8.9.9), where, as seen from (7.5.3),

$$|R_0| < \frac{M}{n^d} \sum_{i=s_1}^{s} |\lambda_i^m E v_i| + \frac{M}{n^d} \sum_{i=s_1}^{s} |\lambda_i|^{m+1/k}$$

with

$$s_1 = [-2k] + 1, \qquad s = [\delta n^{+\frac{1}{2}+\varepsilon}].$$

It follows from (7.5.2) and from the first relation (7.5.7) that the first and second terms converge at least as rapidly as S_2 and S_2/n, respectively, where

$$S_2 = \frac{1}{n^d} \sum_{i=s_1}^{n} \frac{1}{i} \left(\frac{i}{n}\right)^{mk}.$$

Adding (8.9.6) and (8.9.9), we obtain the mean value formula (8.4.1). Since S_1 has a smaller order than S_2, we conclude that R_{mean} has at most the same order as S_2. It is easy to verify that S_2 satisfies (8.11.2), and hence the first part of the theorem is proved.

We shall make a few remarks as to the second part of the theorem. Formulae (8.9.11)–(8.9.14) are true even in this case. By (7.5.4) and condition (c) we conclude that

$$\operatorname{var} T_0 < \frac{M}{n^{2d}} \Big[\sum_{s_1 \leqslant i \leqslant j \leqslant s} |\lambda_i \lambda_j|^m (|\operatorname{cov}(v_i, v_j)| + |\lambda_i|^{1/k} + |\lambda_j|^{1/k}) \Big],$$

where s_1 and s are defined as before. Using (7.5.2) and the second relation (7.5.7), we see that the variance of T_0 has at most the same order as

$$\frac{1}{n^{2d}} \sum_{s_1 \leqslant i \leqslant j \leqslant s} \frac{1}{j} \left(\frac{i}{n}\right)^{mk} \left(\frac{j}{n}\right)^{mk}.$$

The order of magnitude of this expression is determined by a straightforward analysis. Using (8.9.11) in the same way as in the final stage of the proof in 8.9, we finally obtain (8.11.3), and the theorem is proved.

There is a further theorem, analogous to the second theorem in 8.9, p. 102, which presupposes that (8.11.1) holds only for certain ranks i. On the other hand, we must then impose some restriction upon var h, for example that it satisfies condition (8.9.18). The formulation of the theorem will be clear from these suggestions.

8.12. *AP*-transforms. Asymptotic distribution of a linear combination

Corresponding to the theorem in 8.10, we have the following result.

THEOREM. *If*

(a) *$G(u)$ satisfies Condition B in 3.5 with $\mathfrak{C}$ equal to the half-open interval $0 < u \leqslant 1$,*

(b) *$G(u)$ is an AP-transform at $u = 0$,*

(c) *the coefficients g_i are zero for $i \leqslant -2k$, and, further, constants M and m can be found such that $mk > -\frac{1}{2}$ and, for any $i > -2k$,*

$$|g_i| < \frac{M}{n} |\lambda_i^\circ|^{m-1}, \tag{8.12.1}$$

(d) *condition (8.8.2) is satisfied, and, in addition,* var h *remains greater than zero when n approaches infinity,*

then T is asymptotically normally distributed with the mean $\sum_{i=1}^{n} g_i\, G(p_i)$ and the variance var h/n.

It is obvious how the formulation should be modified when $G(u)$ is an *AP*-transform at $u = 1$, or at both $u = 0$ and $u = 1$.

The proof of the theorem is quite similar to that of the theorem in 8.10 and will not be given here.

8.13. Linear combinations defined by means of continuous weight functions

When the asymptotic properties of the linear combination T in (8.1.1) are subject to investigation, it is sometimes convenient, when possible, to represent the mean and the variance of T by integrals instead of by sums. Such approximations have been used by Jung (1955), whose results are, however, less general and were obtained by a different method than that indicated below. As approximations by means of integrals will not be much used in the sequel, only a few remarks will be made on this question.

Assume that the coefficients g_i in T are obtained as particular values of a continuous weight function. More specifically, define g_i from

$$g_i = \frac{1}{n} h'(\lambda_i), \tag{8.13.1}$$

where $h'(x)$ is the first derivative of a function $h(x)$, defined over the range of variation of ξ, and λ_i is as usual given by (5.3.2). We shall suppose that $h(x)$ and $h'(x)$ are continuous for any x. Formula (8.13.1) is analogous to the expression (8.2.2), which connects the quantities g_i and h_i in the general case.

Replacing the sums in (8.4.1) and (8.4.5) by integrals and making a transformation of variables, we find

$$\begin{aligned} E\,T &= \int x h'(x) f(x)\,dx + R^{(1)}_{\text{mean}} + R^{(2)}_{\text{mean}}, \\ \operatorname{var} T &= \frac{1}{n}\left[\int [h(x)]^2 f(x)\,dx - \left(\int h(x) f(x)\,dx\right)^2\right] + R^{(1)}_{\text{var}} + R^{(2)}_{\text{var}}, \end{aligned} \tag{8.13.2}$$

vhere, under general conditions (cf. Cramér, 1946 a, p. 123 ff.),

$$R^{(2)}_{\text{mean}} = o(1), \qquad R^{(2)}_{\text{var}} = o\left(\frac{1}{n}\right).$$

he behaviour of $R^{(1)}_{\text{mean}}$ and $R^{(1)}_{\text{var}}$ is clear from the discussion in 8.7, 8.9, nd 8.11.

We shall use these results in 12.3 and 12.6.

PART III

TRANSFORMED BETA-VARIABLES APPLICATIONS TO THE THEORY OF LINEAR ESTIMATION

CHAPTER 9

LINEAR ESTIMATES OF LOCATION AND SCALE PARAMETERS

9.1. Introduction

Let z be a random variable with the cdf. $F[(z-\mu)/\sigma]$, where μ and σ are unknown parameters. In applications, μ and σ are often the mean and the standard deviation of z, but there are also other possibilities. μ may, for example, be defined as one of the end-points of the range of variation of z or, more generally, as any percentage point of the listribution, and σ may be defined, e.g., as the range of variation of z.

Let us represent by

$$\xi = \frac{z-\mu}{\sigma} \tag{9.1.1}$$

ie reduced variable with the cdf. $F(x)$. Throughout Part III, we shall ippose that $F(x)$ is differentiable for any x. The corresponding fr.f. ill be denoted by $f(x)$.

Suppose that we have at our disposal an ordered random sample

$$z_{(1)} \leqslant z_{(2)} \leqslant \cdots \leqslant z_{(n)}$$

z-values. We want to estimate the location parameter μ and the le parameter σ by means of linear expressions of the form

$$U = \sum_{i=1}^{n} g_i z_{(i)}, \tag{9.1.2}$$

re the coefficients g_i do not depend on μ and σ. Any such expres- will be called a *linear estimate.*

Putting

$$z_{(i)} = \mu + \sigma x_i, \tag{9.1.3}$$

we have

$$U = \mu \sum_{i=1}^{n} g_i + \sigma \sum_{i=1}^{n} g_i x_i. \tag{9.1.4}$$

It is immediately realized (cf. Chapter 3) that the variables $x_1, x_2, \ldots, x_n$, defined by (9.1.3), constitute a set of transformed beta-variables with the transform $x = G(u)$, where $G(u)$ is the inverse of $F(x)$. Hence U is linearly dependent upon a linear combination

$$T = \sum_{i=1}^{n} g_i x_i$$

of the type studied in Chapter 8. Consequently, we can apply the results deduced in the chapter also to U. Note that, in this application of the general theory of TRB-variables, $G(u)$ is increasing and differentiable.

In the present chapter, which has a preliminary character, we shall derive expressions for the moments of U. Depending upon our knowledge about the means and covariances of the TRB-variables, these expressions will be exact or approximate. We shall also use these results to formulate conditions with respect to the g_i's which ensure that unbiased or, at least, nearly unbiased estimates of the parameters are obtained.

For the benefit of the reader who is mainly interested in the theory of linear estimation, we shall reproduce certain results obtained in Part II, which are of basic importance for the theory expounded in Part III.

9.2. Unbiased estimates

We shall suppose that the parameter to be estimated is $k_1 \mu + k_2 \sigma$, where k_1 and k_2 are two given quantities. For $k_1 = 1$, $k_2 = 0$ and $k_1 = 0$, $k_2 = 1$, the parameter is specialized to μ and σ respectively. Other values of k_1 and k_2 may also be of interest; such is the case, for example, when percentage points $\neq \mu$ are estimated (cf. 10.9).

It follows from (9.1.4) that U in (9.1.2) is an unbiased estimate of the parameter, if the coefficients are chosen such that

$$\sum_{i=1}^{n} g_i = k_1, \qquad \sum_{i=1}^{n} g_i E\, x_i = k_2. \tag{9.2.1}$$

Further, it follows from (9.1.4) and (8.3.2) that

$$\text{var}\ U = \sigma^2 \sum_{i,j=1}^{n} g_i g_j \ \text{cov}\ (x_i, x_j). \tag{9.2.2}$$

If the covariances of x_i and x_j are accurately known, the variance of U can be determined from this expression (apart from the factor σ^2). If the covariances are unknown, or if we find the exact formula inconvenient, we make use of either the approximation formulae described in 8.4 or of the α, β-corrected expressions given in 8.5. Both alternatives will be treated below.

Note that the expressions given above are meaningful only when the means and the covariances of the *TRB*-variables exist for all i and j. If this is not the case, the sample may have to be censored. Formulae (9.2.1) and (9.2.2) are then still true, provided that we set $g_i = 0$ corresponding to any discarded observation. The same remark applies to the case where certain observations are missing.

(a) *α, β-correction is not used.*

As described in detail in Chapter 8, we replace the coefficients g_i in (9.1.2) by new quantities

$$h_0, h_1, \ldots, h_n$$

lefined, apart from an additive constant, by the relations

$$g_i = \theta_i (h_i - h_{i-1}). \tag{9.2.3}$$

s already mentioned in 5.5, the weights θ_i are given by

$$\theta_i = [G'(p_i)]^{-1}, \quad (i = 1, \ldots, n), \quad \theta_0 = \theta_{n+1} = 0, \tag{9.2.4}$$

ere $p_i = i/(n+1)$. Using the notation

$$\lambda_i = G(p_i), \tag{9.2.5}$$

may write, alternatively,

$$\theta_i = f(\lambda_i), \quad (i = 1, \ldots, n), \quad \theta_0 = \theta_{n+1} = 0. \tag{9.2.6}$$

e that, when the sample is incomplete, we have $h_i = h_{i-1}$, correding to any censored or missing observation $z_{(i)}$ (cf. 8.6). the coefficients h_i are used instead of the g_i's, the variance of T be given a convenient approximate form, as seen from (8.4.5). It vs immediately from this formula and from (9.2.2) that the vari- of U can be written

$$\operatorname{var} U = \frac{\sigma^2 \operatorname{var} h}{n+2} + \sigma^2 R_{\mathrm{var}}, \tag{9.2.7}$$

where

$$\operatorname{var} h = \frac{1}{n+1} \sum_{i=0}^{n} (h_i - \bar{h})^2, \tag{9.2.8}$$

$$\bar{h} = \frac{1}{n+1} \sum_{i=0}^{n} h_i,$$

and R_{var} is an error term given by the equivalent expressions (8.4.4) and (8.4.7).

It is sometimes convenient to use the quantities h_i not only for the computation of the variance of U but also in the conditions for unbiasedness (9.2.1). For this purpose, we introduce the following two sets of $n+1$ quantities

$$C_{1i} = \theta_i - \theta_{i+1}, \qquad (i = 0, 1, \ldots, n), \tag{9.2.9}$$

$$C_{2i} = \theta_i E x_i - \theta_{i+1} E x_{i+1}, \qquad (i = 0, 1, \ldots, n). \tag{9.2.10}$$

We observe that

$$\sum_{i=0}^{n} C_{ri} = 0, \qquad (r = 1, 2). \tag{9.2.11}$$

Further, when ξ has a symmetrical distribution,

$$\sum_{i=0}^{n} C_{1i} C_{2i} = 0. \tag{9.2.12}$$

Hence the sets are then orthogonal. Replacing the g_i's in (9.2.1) by the h_i's given by (9.2.3), we find

$$\sum_{i=0}^{n} C_{ri} h_i = k_r, \quad (r = 1, 2). \tag{9.2.13}$$

If these conditions are satisfied, U is an unbiased estimate of

$$k_1 \mu + k_2 \sigma.$$

(b) *α, β-correction is used.*

As seen from 8.5, the use of the α, β-correction leads to the following modifications of the procedure described in (a).

The weights θ_i are defined by

$$\theta_i = [G'(\pi_i)]^{-1}, \quad (i = 1, \ldots, n), \qquad \theta_0 = \theta_{n+1} = 0, \tag{9.2.14}$$

where
$$\pi_i = \frac{i - \alpha_i}{n - \alpha_i - \beta_i + 1}. \tag{9.2.15}$$

The sum $\alpha_i + \beta_i$ will be assumed to be independent of i. Formulae (9.2.7) and (9.2.8) are replaced by

$$\operatorname{var} U = \frac{\sigma^2 V(h, \rho)}{n' + 2} + \sigma^2 R_V \tag{9.2.16}$$

and
$$V(h, \rho) = \frac{1}{n' + 1} \left[\sum_{i=0}^{n} \rho_i h_i^2 - \frac{1}{n' + 1} \left(\sum_{i=0}^{n} \rho_i h_i \right)^2 \right], \tag{9.2.17}$$

where
$$n' = n - \alpha_i - \beta_i,$$

$$\rho_i = 1 - \alpha_{i+1} + \alpha_i = 1 - \beta_i + \beta_{i+1},$$

$$(\alpha_0 = 0;\ \beta_{n+1} = 0),$$

and R_V is given by (8.5.5). As we saw in 8.5, the case $\alpha_i = \alpha$, $\beta_i = \beta$ is of special interest because of the simple character of the resulting formulae. We then have $\rho_0 = 1 - \alpha$ and $\rho_n = 1 - \beta$, all other ρ_i's being equal to unity.

The formulae presented in this section are very general, since we have made no assumptions with respect to the transform $G(u)$ besides the general conditions stated in 9.1. The approximate expressions (9.2.7) and (9.2.16) are, however, of practical interest only when the error terms are small, at least for large n. For this purpose, a separate investigation is needed for each class of transforms. The reader is referred to 8.7, 8.9, and 8.11, where the asymptotic behaviour of the error terms has been studied in detail for some important families of distributions.

To bring out the essential features of the formulae reproduced in (a), we shall give a simple example.

Rectangular distribution.

Suppose that z is distributed according to the rectangular distribution, and let μ be the mean and σ the range of variation of z. It follows from (9.1.1) that ξ has mean 0 and range of variation 1. Hence

$$E\, x_i = \frac{i}{n+1} - \tfrac{1}{2}.$$

As seen from (9.2.4), we have $\theta_i = 1, (i = 1, \ldots, n)$. Hence (9.2.9) and (9.2.10) are specialized to

$$C_{1i} = \begin{cases} -1 \\ 0, \\ 1 \end{cases} \qquad C_{2i} = \begin{cases} \dfrac{n-1}{2(n+1)} & \text{for } i = 0 \\ -\dfrac{1}{n+1} & \text{for } 1 \leqslant i \leqslant n-1 \\ \dfrac{n-1}{2(n+1)} & \text{for } i = n. \end{cases}$$

Two pairs of estimates μ^* and σ^* of μ and σ will be considered.

(a) *Arithmetic mean and Gini's mean difference.*

We take
$$\mu^* = \bar{z} = \frac{1}{n} \sum_{i=1}^{n} z_{(i)},$$

$$\sigma^* = \frac{\sum_{i=1}^{n} \left(\dfrac{i}{n+1} - \dfrac{1}{2} \right) z_{(i)}}{\sum_{i=1}^{n} \left(\dfrac{i}{n+1} - \dfrac{1}{2} \right)^2} = \frac{12}{n(n-1)} \sum_{i=1}^{n} \left(i - \frac{n+1}{2} \right) z_{(i)}.$$

The second estimate is Gini's mean difference in the rectangular case. These estimates are obtained, if we take, respectively,

$$h_i = \frac{i}{n}, \quad h_i = \frac{6(i - n/2)^2}{n(n-1)}.$$

That the side conditions (9.2.13) are fulfilled is easily checked; the estimates are therefore unbiased (this is, of course, also seen directly from the definition of the estimates). Inserting the expressions for the coefficients h_i in (9.2.7), we find after some calculation the following variances of the estimates.

$$\text{var } \mu^* = \frac{\sigma^2}{12n}, \quad \text{var } \sigma^* = \frac{(n+3)\sigma^2}{5n(n-1)}. \qquad (9.2.18)$$

Equality signs may be used, the error terms being zero. The latter expression seems to have been first given by Nair (1936).

(b) *Estimates based upon extreme sample-values.*

We take
$$\mu^* = \tfrac{1}{2}(z_{(1)} + z_{(n)}),$$
$$\sigma^* = \frac{n+1}{n-1}(z_{(n)} - z_{(1)}),$$

which estimates are obtained for

$$h_i = \begin{cases} -\frac{1}{2} \\ 0, \\ \frac{1}{2} \end{cases} \qquad h_i = \begin{cases} \frac{n+1}{n-1} & \text{for } i=0 \\ 0 & \text{for } 1 \leqslant i \leqslant n-1 \\ \frac{n+1}{n-1} & \text{for } i=n. \end{cases}$$

The variance formula (9.2.7) gives the known expressions

$$\operatorname{var} \mu^* = \frac{\sigma^2}{2(n+1)(n+2)}, \qquad \operatorname{var} \sigma^* = \frac{2\sigma^2}{(n-1)(n+2)}. \tag{9.2.19}$$

The main application of the results reviewed in the section will be made in Chapter 10.

9.3. Nearly unbiased estimates

Throughout the preceding section it was assumed that accurate values of the means of the *TRB*-variables can be obtained from tables or other sources. When this is not the case, the approximations discussed in Chapter 6 and in 8.5 are used.

The conditions (9.2.1) then become

$$\sum_{i=1}^{n} g_i = k_1, \qquad \sum_{i=1}^{n} g_i\, G(\pi_i) = k_2, \tag{9.3.1}$$

where π_i is given by (9.2.15). In the sequel we shall sometimes find it convenient to use the notation

$$\lambda_i' = G(\pi_i).$$

The same modification is applied to the definition (9.2.10) of C_{2i}, which is replaced by

$$C_{2i}' = \theta_i\, G(\pi_i) - \theta_{i+1}\, G(\pi_{i+1}), \tag{9.3.2}$$

where θ_i should now be defined by (9.2.14).

When the relations (9.3.1) are valid, U is a biased estimate of

$$k_1 \mu + k_2 \sigma$$

with the mean (cf. 6.2 and 8.5)

$$E\,U = k_1 \mu + k_2 \sigma + \sigma R'_{\text{mean}},$$

where R'_{mean} is given by (8.5.2).

If α and β are chosen such that $G(\pi_i)$ is a good overall approximation to $E\,x_i$ (cf. the discussion in 6.9 and the following sections), the bias may be expected to be quite small.

CHAPTER 10

NEARLY BEST LINEAR ESTIMATES

10.1. Introduction

Let U be a linear estimate of the parameter $k_1 \mu + k_2 \sigma$. In Chapter 9 we have stated some results concerning the class of functions U which are unbiased or nearly unbiased estimates of the parameter. In this chapter we consider the following problem. We wish to select the function U which has the smallest possible variance of all functions in the class. The unbiased linear estimate with this property will be termed the *best linear estimate.*

Much attention has been paid in the literature to this minimum variance problem. For exactly unbiased estimates the general solution has been given by Lloyd (1952). When applied to special distributions, the solution involves for the most part considerable numerical work. The covariances of the *TRB*-variables must be computed for all combinations of i and j, and the coefficients g_i in U are then obtained as solutions of n linear equations with the covariances as coefficients. A separate calculation is needed for each sample-size.

Several studies of particular distributions have been published. Exact analytical solutions have been given for the rectangular distribution (Lloyd, 1952), the exponential distribution (Sarhan, 1954), and the generalized geometric distribution (Downton, 1954). Numerical solutions for certain small sample-sizes have been derived for, among others, the normal distribution (Godwin, 1949 *a*), the extreme-value distribution (Lieblein, 1954; see also Lieblein & Zelen, 1956), the triangular distribution (Sarhan, 1954), and Laplace's distribution (Sarhan, 1954). The case of censored samples has been studied numerically by Sarhan (1955) and Sarhan & Greenberg (1956).

The present chapter contains one of the main results of the thesis. A procedure will be developed which furnishes approximations to the exact solution of the minimum variance problem. The method will be used for both complete and censored samples and illustrated by several numerical examples. The resulting linear estimates will be called *nearly best linear estimates* or, shorter, nearly best estimates.

The main contents of the chapter have been summarized in an earlier publication (Blom, 1956).

10.2. A preliminary remark

When we seek a convenient method of solving the minimum variance problem, our first thought is, perhaps, to write U, as given by (9.1.2), as a linear combination with unknown coefficients of the unweighted differences $\delta_i = x_{i+1} - x_i$ (cf. 4.4). If the covariances of the variables δ_i are known, an expression for the variance of U is then obtained, which can be minimized in the usual way. This method works, in fact, excellently when $G(u)$ is a B_0-transform (cf. 3.5) because of the explicit expressions (4.4.6) which are then obtained for the expected values $E\,\delta_i\,\delta_j$, and hence for the covariances of the differences δ_i and δ_j. In this special case it is even possible to obtain by this method an exact analytical solution of the minimum variance problem. We shall not reproduce the solution, which, as mentioned in 10.1, has already been derived by other authors for the three members of the B_0-family, though the relationship between these distributions seems to have passed unnoticed.

In the general case, the expected values referred to above are difficult to calculate, and it is better to abandon this approach to the problem. Instead, we use the formula (9.2.7), which is based upon the properties of the weighted differences y_i introduced in 5.5.

10.3. Determination of unbiased nearly best estimates

Let U be a linear estimate of the general form (9.1.2). It is an unbiased estimate of $k_1\mu + k_2\sigma$, if the conditions (9.2.13) are satisfied. We shall assume that the means $E\,x_i$ appearing in these expressions are accurately known from tables, etc. (later we shall remove this condition). As seen from (9.2.7) and (9.2.8), the variance of U is, apart from the error term in (9.2.7), proportional to

$$Z = \sum_{i=0}^{n} h_i^2 - \frac{1}{n+1}\left(\sum_{i=0}^{n} h_i\right)^2. \tag{10.3.1}$$

Let now Z be minimized with respect to the h_i's subject to the side conditions (9.2.13). Generally, the true minimum of var U will not be attained in this way, since the remainder term is discarded. Later, we shall examine the consequences of this approximation both empirically (cf. 10.8) and analytically (cf. Chapter 13). For the present, we merely point out that, as will emerge from the empirical study, good results may be obtained even when the remainder term is considerable.

The solution of the minimum value problem can, in terms of two Lagrange multipliers, be written

$$h_i - \bar{h} = a_1 C_{1i} + a_2 C_{2i}, \quad (i = 0, 1, \ldots, n), \tag{10.3.2}$$

where

$$\bar{h} = \frac{1}{n+1} \sum_{i=0}^{n} h_i .$$

The determination of a_1 and a_2 is made in the traditional way. Multiplying (10.3.2) first by C_{1i} and adding from 0 to n, secondly by C_{2i} and again adding from 0 to n, we get, using (9.2.11) and (9.2.13),

$$\begin{aligned} a_1 d_{11} + a_2 d_{12} &= k_1, \\ a_1 d_{21} + a_2 d_{22} &= k_2, \end{aligned} \tag{10.3.3}$$

where

$$d_{rs} = \sum_{i=0}^{n} C_{ri} C_{si}, \quad (r, s = 1, 2). \tag{10.3.4}$$

Set

$$D = \{d_{rs}\}, \quad D^{-1} = \{d^{rs}\} = \frac{1}{|D|} \begin{Bmatrix} d_{22} & -d_{12} \\ -d_{21} & d_{11} \end{Bmatrix}, \tag{10.3.5}$$

where $$|D| = d_{11} d_{22} - d_{12}^2 .$$

If $|D| \neq 0$, the solution of the system (10.3.3) can be written

$$\begin{aligned} a_1 &= d^{11} k_1 + d^{12} k_2, \\ a_2 &= d^{21} k_1 + d^{22} k_2 . \end{aligned} \tag{10.3.6}$$

Finally, (10.3.2) is multiplied by h_i, and the $n+1$ relations are added. Hence we obtain the minimum of Z

$$Z_{\min} = a_1 k_1 + a_2 k_2 . \tag{10.3.7}$$

It is in most cases convenient to translate the solution from h_i to g_i. By (9.2.3) and (10.3.2)

$$g_i = \theta_i \left[a_1 (C_{1i} - C_{1\,i-1}) + a_2 (C_{2i} - C_{2\,i-1})\right] \tag{10.3.8}$$

or, alternatively, in central difference notation

$$g_i = -\theta_i \left[a_1 \delta^2 \theta_i + a_2 \delta^2 (\theta_i E x_i)\right]. \tag{10.3.9}$$

The estimate

$$U^* = \sum_{i=1}^{n} g_i z_{(i)} \tag{10.3.10}$$

with these coefficients will be termed the *unbiased nearly best linear estimate* of $k_1 \mu + k_2 \sigma$. Its variance is given by

$$\text{var } U^* = \frac{\sigma^2}{(n+1)(n+2)} \cdot Z_{\min} + \sigma^2 R_{\text{var}}, \qquad (10.3.11)$$

where $Z_{\min}$ is given by (10.3.7), and R_{var} is an error term (cf. 8.4).

By taking $k_1 = 1$, $k_2 = 0$ or $k_1 = 0$, $k_2 = 1$, we obtain the nearly best estimates

$$\mu^* = \sum_{i=1}^{n} g_{1i} z_{(i)}, \qquad \sigma^* = \sum_{i=1}^{n} g_{2i} z_{(i)} \qquad (10.3.12)$$

of μ and σ. The coefficients are given by

$$g_{ri} = \theta_i [d^{r1} (C_{1i} - C_{1\,i-1}) + d^{r2} (C_{2i} - C_{2\,i-1})], \quad (r = 1, 2). \qquad (10.3.13)$$

The variances of μ^* and σ^* are, to the accuracy provided by the leading term in (10.3.11),

$$\text{var } \mu^* \sim \frac{\sigma^2}{(n+1)(n+2)} \frac{d_{22}}{|D|}, \qquad \text{var } \sigma^* \sim \frac{\sigma^2}{(n+1)(n+2)} \frac{d_{11}}{|D|}. \qquad (10.3.14)$$

Using the expressions (8.4.9) and (8.4.10) or by a direct calculation, we further obtain

$$\text{cov } (\mu^*, \sigma^*) \sim -\frac{\sigma^2}{(n+1)(n+2)} \frac{d_{12}}{|D|}. \qquad (10.3.15)$$

The procedure developed above can be summarized in the following numerical rules.

Step 1. Compute the weights θ_i according to (9.2.4).

Step 2. Compute $\theta_i E x_i$, where $E x_i$ is the mean of the ith order statistic in a sample of n values from the reduced cdf. $F(x)$.

Step 3. Compute the quantities C_{1i} and C_{2i} according to (9.2.9) and (9.2.10).

Step 4. Compute the elements d_{rs} of the matrix D and the elements d^{rs} of the reciprocal D^{-1} according to (10.3.4) and (10.3.5). Note that, in the symmetrical case, $d_{12} = 0$. Hence $d_{22}/|D| = 1/d_{11}$, etc.

Step 5. Compute the coefficients of the nearly best estimates μ^* and σ^* by inserting the values of d^{rs}, θ_i, and C_{ri} in (10.3.13).

Step 6. If the variances of μ^* and σ^* are required, use (9.2.2) or, if the variances and covariances of the variables x_i are unknown, the approximate expressions (10.3.14). (As mentioned in the beginning of the section, these approximations may be crude, even though the method works excellently in other respects.)

Numerical illustrations of these rules will be given in 10.8.

10.4. A single unknown parameter

When only one of the parameters μ and σ is unknown, the method described in the foregoing section should be modified.

(a) *μ unknown, σ known.*

Z in (10.3.1) is minimized subject to the single side condition

$$\sum_{i=0}^{n} C_{1i}\, h_i = k_1 .$$

The solution is found to be

$$h_i - \bar{h} = a_1\, C_1 \quad ,$$

where

$$a_1 = \frac{k_1}{d_{11}} .$$

(10.3.11) is valid even in this case with

$$Z_{\min} = a_1\, k_1 .$$

Further, the coefficients of the resulting nearly best estimate U_1^* are given by

$$g_i = \theta_i\, a_1\, (C_{1i} - C_{1\,i-1}). \tag{10.4.1}$$

It is easily found that the mean of U_1^* is given by

$$E\, U_1^* = k_1\, \mu + \sigma \sum_{i=0}^{n} C_{2i}\, h_i = k_1 \left(\mu + \sigma \frac{d_{12}}{d_{11}}\right) .$$

Thus we obtain an unbiased nearly best estimate μ^* of μ by taking $k_1 = 1$ and subtracting the known term $\sigma\, d_{12}/d_{11}$ from U_1^*. The variance is given approximately by

$$\text{var } \mu^* \sim \frac{\sigma^2}{(n+1)(n+2)} \frac{1}{d_{11}}. \quad (10.4.2)$$

(b) μ *known*, σ *unknown*.

The coefficients of the nearly best estimate U_2^* are given by

$$g_i = \theta_i a_2 (C_{2i} - C_{2\,i-1}), \quad (10.4.3)$$

where
$$a_2 = \frac{k_2}{d_{22}}.$$

The mean of U_2^* is

$$E\,U_2^* = k_2 \left(\frac{\mu\, d_{12}}{d_{22}} + \sigma\right).$$

An unbiased nearly best estimate σ^* of σ is obtained by taking $k_2 = 1$ and subtracting the known term $\mu\, d_{12}/d_{22}$ from U_2^*. The variance is approximately

$$\text{var } \sigma^* \sim \frac{\sigma^2}{(n+1)(n+2)} \frac{1}{d_{22}}. \quad (10.4.4)$$

10.5. Censored samples

The procedure developed in 10.3 can easily be extended to the case of censored samples. Suppose, e.g., that the r_1 smallest and r_2 largest observations $z_{(i)}$ are missing; the estimation of μ and σ should then be based upon the observations $z_{(r_1+1)}, \ldots, z_{(n-r_2)}$. Thus we must require that, in (9.1.2), the g_i's corresponding to missing observations are zero.

Consequently, in (10.3.1) we take

$$h_0 = h_1 = \cdots = h_{r_1} \text{ and } h_{n-r_2} = \cdots = h_n.$$

Hence

$$Z = (r_1+1)\,h_{r_1}^2 + \sum_{i=r_1+1}^{n-r_2-1} h_i^2 + (r_2+1)\,h_{n-r_2}^2 -$$
$$- \frac{1}{n+1}\left[(r_1+1)\,h_{r_1} + \sum_{i=r_1+1}^{n-r_2-1} h_i + (r_2+1)\,h_{n-r_2}\right]^2.$$

The side conditions (9.2.13) are changed in a similar way.

Determining the minimum of Z in the same way as before, we see after some calculation that (10.3.2) is still valid, provided that we replace C_{1i} and C_{2i} by C_{1i}^* and C_{2i}^* according to the following definition.

$$C_{1i}^* = \begin{cases} -\dfrac{1}{r_1+1}\theta_{r_1+1} & \text{for } 0 \leqslant i \leqslant r_1 \\ C_{1i} & \text{for } r_1+1 \leqslant i \leqslant n-r_2-1 \\ \dfrac{1}{r_2+1}\theta_{n-r_2} & \text{for } n-r_2 \leqslant i \leqslant n, \end{cases}$$

$$C_{2i}^* = \begin{cases} -\dfrac{1}{r_1+1}\theta_{r_1+1}\, E\, x_{r_1+1} & \text{for } 0 \leqslant i \leqslant r_1 \\ C_{2i} & \text{for } r_1+1 \leqslant i \leqslant n-r_2-1 \\ \dfrac{1}{r_2+1}\theta_{n+r_2}\, E\, x_{n-r_2} & \text{for } n-r_2 \leqslant i \leqslant n. \end{cases} \tag{10.5.1}$$

With this modification the other formulae in 10.3 also hold good.

Other types of censoring are similarly handled. We realize that the extreme cases of censoring occurring when the estimation is based upon a few selected observations can also be treated by this method.

10.6. α, β-corrected nearly best estimates

The nearly best estimate, which has been discussed in the preceding sections of this chapter, was derived on the assumption that the variance of the linear estimate (9.1.2) is approximated by the leading term of (9.2.7). We saw, however, in 9.2 that the variance can also be written in the α, β-corrected form (9.2.16). By using the leading term of this formula as the starting-point for derivations of the same kind as in 10.3, we are able to construct a more general estimate, which will be called the *α, β-corrected nearly best linear estimate.* Since the theory of this estimate has as yet not been fully explored, the following discussion has a preliminary character.

We shall assume that the sample is non-censored. Further, we shall suppose that α_i and β_i do not depend upon i, so that $\alpha_i = \alpha$ and $\beta_i = \beta$. As seen from (9.2.16), we have then to minimize

$$Z' = \sum_{i=0}^{n} \rho_i h_i^2 - \frac{1}{n'+1}\left(\sum_{i=0}^{n} \rho_i h_i\right)^2,$$

where $\rho_0 = 1-\alpha, \quad \rho_1 = \cdots = \rho_{n-1} = 1, \quad \rho_n = 1-\beta,$

with the usual side conditions (9.2.13). It should be observed that the quantities θ_i, which appear in these conditions, should now be defined by (9.2.14).

The following modifications of the formulae given in 10.3 then arise. The h_i-coefficients which minimize Z' are given by

$$h_i - \bar{h} = a_1 \cdot \frac{C_{1i}}{\rho_i} + a_2 \cdot \frac{C_{2i}}{\rho_i},$$

where
$$\bar{h} = \frac{1}{n'+1} \sum_{i=0}^{n} \rho_i h_i .$$

The multipliers a_1 and a_2 are given by (10.3.6), provided that we replace (10.3.4) by

$$d_{rs} = \sum_{i=0}^{n} \frac{C_{ri}\, C_{si}}{\rho_i}, \quad (r, s = 1, 2).$$

The coefficients of the α, β-corrected nearly best estimate $U^{*\prime}$ are

$$g_i = \frac{\theta_i}{\rho_i} [a_1 (C_{1i} - C_{1\; i-1}) + a_2 (C_{2i} - C_{2\; i-1})],$$

where θ_i is given by (9.2.14).

Formula (10.3.11) should be replaced by

$$\operatorname{var} U^{*\prime} = \frac{\sigma^2}{(n'+1)(n'+2)} \cdot Z_{\min} + \sigma^2 R_V,$$

where $n' = n - \alpha - \beta$, and R_V is an error term. The remaining expressions in 10.3 are valid with the modifications already mentioned.

The estimate $U^{*\prime}$ contains two quantities α and β, which have to be determined. Theoretically, they might be chosen in such a manner that the approximation to the best estimate becomes as good as possible. This principle has, however, no practical interest at the present state of our knowledge, as no method of calculating the optimal values is known. Moreover, the numerical study described in 10.8 lends support to the assertion that the efficiency of the nearly best estimate derived in 10.3 is often so high that any substantial improvement of the efficiency cannot possibly be effected by using an α, β-corrected estimate. (However, there may, of course, exist distributions for which the correction is of practical importance from this point of view.)

Quite apart from these remarks, there is often good reason for using α, β-corrected nearly best estimates, as we shall see in the next section.

10.7. Nearly unbiased nearly best estimates

We have hitherto assumed that the means $E x_i$ are accurately known. When, as is generally the case, such detailed information is not available, it is not possible to calculate exactly unbiased estimates, and a certain bias must be tolerated. To reduce the bias, we use the α, β-correction in the way described in 9.3 (cf. also Chapter 6). We then proceed as follows.

First, the quantities α and β are determined such that $G(\pi_i)$ is a good overall approximation to $E x_i$.

Secondly, the nearly best estimate is derived by either (a) the method described in 10.3 or (b) the method described in 10.6. Regardless of which alternative is chosen, C_{2i} should be replaced by C'_{2i} defined by (9.3.2). In all other respects the rules are unaltered. When alternative (b) is used, the question of the choice of α and β, which was left open in 10.6, is thus solved.

It should be observed that, from the numerical point of view, alternative (a) is less attractive to use than alternative (b). In the former case we must, as an examination of the rules immediately shows, compute the values of $G(u)$ for $u = \pi_i$, $(i = 1, \ldots, n)$, and the values of $[G'(u)]^{-1}$ for $u = p_i$. If, on the other hand, alternative (b) is chosen, the calculations have to be made for $u = \pi_i$ in both cases.

In either way we obtain estimates which may be denoted *nearly unbiased nearly best linear estimates.* As already mentioned, they have often a very small bias and then practically coincide with the unbiased nearly best estimates treated in the preceding sections of the chapter.

10.8. Examples of unbiased nearly best estimates

The method for finding nearly best estimates can be expected to yield satisfactory results, if the error term in (9.2.7) or (9.2.16) is small compared to the leading term. In Chapter 13 we shall show that, given certain general conditions, this is true asymptotically in the sense that the quotient of the error term and the leading term tends to zero when $n \to \infty$.

It is, however, a notable feature of the method that the approximation may often be very good even though the sample-size is small. We shall show this empirically for seven different distributions. These distributions have been chosen, because best linear estimates are known

TABLE 4.

Right triangular distribution. Calculation of unbiased nearly best estimates μ^* and σ^*. Sample size $n=5$.

i	θ_i	$\theta_i\, E x_i$	C_{1i}	C_{2i}	$\theta_i\,(C_{1i} - C_{1\,i-1})$	$\theta_i\,(C_{2i} - C_{2\,i-1})$	g_{1i}	g_{2i}
0	0	0	−0.1925	0.2427				
1	0.1925	−0.2427	−0.0797	−0.1127	0.02170	−0.06841	0.3679	−0.3588
2	0.2722	−0.1300	−0.0611	−0.1667	0.00505	−0.01469	0.0816	−0.0783
3	0.3333	0.0367	−0.0516	−0.1942	0.00320	−0.00916	0.0513	−0.0490
4	0.3849	0.2309	−0.0454	−0.2117	0.00236	−0.00673	0.0377	−0.0361
5	0.4303	0.4426	0.4303	0.4426	0.20474	0.28156	0.4615	0.5222
6	0	0						

$$D=\begin{Bmatrix} \Sigma C_{1i}^2 & \Sigma C_{1i}\, C_{2i} \\ \Sigma C_{1i}\, C_{2i} & \Sigma C_{2i}^2 \end{Bmatrix}=\begin{Bmatrix} 0.2370 & 0.1826 \\ 0.1826 & 0.3778 \end{Bmatrix}, \qquad D^{-1}=\begin{Bmatrix} 6.720 & -3.247 \\ -3.247 & 4.216 \end{Bmatrix}.$$

Calculation of two last columns: $g_{11}=6.720 \cdot 0.02170+(-3.247)\cdot(-0.06841)$, etc.
$g_{21}=-3.247 \cdot 0.02170+4.216 \cdot (-0.06841)$, etc.

The expression for θ_i is given in (10.8.1). The values of $E x_i$ have been taken from Table I in Downton (1954). The entries in the above table have been obtained by rounding off entries taken from the original work sheet.

at least for certain sample-sizes, and thus the efficiency of the proposed method can be evaluated in each case.

We shall assume that the means $E x_i$ are accurately known and shall apply the procedures described in 10.3, 10.4, and 10.5.

(1) *Normal, triangular, extreme-value, right triangular, and exponential distributions.*

The method developed in 10.3 has been applied to the five distributions mentioned above in the special case $n=5$. In all these cases, μ and σ are the mean and the standard deviation of z.

As an illustration, we shall reproduce the numerical calculations in detail for the right triangular distribution, which is a special case of the generalized geometric distribution (cf. 3.5). The cdf. of the reduced variable can be written

$$F(x)=\frac{1}{18}(x+2\sqrt{2})^2, \quad (-2\sqrt{2}\leqslant x\leqslant \sqrt{2}).$$

TABLE 5.

Variances and efficiencies of unbiased nearly best linear estimates μ^* and σ^*. Sample-size $n=5$.

Distribution	Estimate	Variances of μ^* and σ^*		Variance of best linear estimate	Percentage efficiency of nearly best linear estimate
		Exact value	Approximative value		
Rectangular	μ^*	0.1428	0.1428	0.1428	100
	σ^*	0.0714	0.0714	0.0714	100
Normal	μ^*	0.2003	0.1554	0.2000	99.8
	σ^*	0.1337	0.0922	0.1333	99.7
Triangular	μ^*	0.2049	0.1684	0.1934	94.4
	σ^*	0.1083	0.1023	0.1080	99.6
Extreme-value	μ^*	0.2007	0.1374	0.1983	98.8
	σ^*	0.1725	0.1014	0.1667	96.6
Right triangular	μ^*	0.1692	0.1600	0.1691	99.9
	σ^*	0.1053	0.1004	0.1051	99.8
Exponential	μ^*	0.2058	0.1187	0.2000	97.2
	σ^*	0.2592	0.1409	0.2500	96.5

μ and σ are the mean and the standard deviation of the distribution. Each variance should be multiplied by σ^2. The last decimal may be incorrect in some cases.

Hence

$$G(u)=\sqrt{18u}-\sqrt{8}$$

and

$$\theta_i=\frac{1}{3}\sqrt{\frac{2i}{n+1}},\quad (i=0,\ 1,\ \ldots,\ n). \qquad (10.8.1)$$

The calculations required for the determination of μ^* and σ^* are given in Table 4. Further, the variances of the estimates have been determined both approximately according to (10.3.14) and accurately according to the general formula (9.2.2). In the latter case, use has been made of the table of variances and covariances of the *TRB*-variables constructed by Downton (1954, p. 310). As a measure of the efficiency we take the quotient of the variance of the best linear estimate and the variance of the nearly best estimate. We shall call this quotient the *efficiency* (*in the linear sense*) of the nearly best estimate.

Similar calculations have been performed for the other four distributions mentioned above. In the case of the normal and the triangular distributions, simplifications are possible, since these distributions are symmetrical. The resulting variances and efficiencies are reproduced in Table 5. The efficiencies must be considered very good.

Note that the approximations (10.3.14) are crude in most cases, which, as already mentioned in 10.3, shows that the method may yield satisfactory results even when the error term in (9.2.7) is considerable.

The two following examples illustrate the case of one unknown parameter.

(2) *Exponential distribution.*

Let z have the fr.f. $\frac{1}{\sigma}e^{-(z-\mu)/\sigma}$, $(z \geqslant \mu)$, where σ is known. Note that the definition of μ is not the same as that used above.

The best linear estimate of μ is $z_{(1)} - \sigma/n$. The nearly best estimate can be constructed by the method described in 10.4. We have

$$\theta_i = \begin{cases} 0 & \text{for } i = 0 \\ 1 - \dfrac{i}{n+1} & \text{for } 1 \leqslant i \leqslant n+1, \end{cases}$$

$$C_{1i} = \begin{cases} -\dfrac{n}{n+1} & \text{for } i = 0 \\ \dfrac{1}{n+1} & \text{for } 1 \leqslant i \leqslant n, \end{cases}$$

and $d_{11} = n/(n+1)$. Setting $k_1 = 1$ in (10.4.1), we get $U_1^* = z_{(1)}$. The mean of U_1^* is $\mu + \sigma/n$. Thus the nearly best unbiased estimate μ^* is $z_{(1)} - \sigma/n$, and hence it coincides with the best estimate.

The variance of μ^* is, as seen from (10.4.2), approximately $\sigma^2/[n(n+2)]$, the exact value being σ^2/n^2.

When σ is unknown, the nearly best estimate of μ is no longer 100 % efficient. If, e.g., $n = 5$, the efficiency (in the linear sense) is 99.3 %.

(3) *Laplace's distribution.*

Suppose that z has the fr.f. $\frac{1}{2}e^{-|z-\mu|}$, where μ is unknown. We find

$$\theta_i = \begin{cases} \dfrac{i}{n+1} & \text{for } i \leqslant \dfrac{n+1}{2} \\ 1 - \dfrac{i}{n+1} & \text{for } i \geqslant \dfrac{n+1}{2}. \end{cases}$$

Let us assume that n is odd, so that $n=2m+1$. Then

$$C_{1i}=\begin{cases}-\dfrac{1}{n+1} & \text{for } 0\leqslant i\leqslant m\\[2ex] \dfrac{1}{n+1} & \text{for } m+1\leqslant i\leqslant n\end{cases}$$

and $d_{11}=1/(n+1)$. Applying (10.4.1), we see that the nearly best estimate is the median

$$\mu^*=z_{(m)}.$$

Similarly, when $n=2m$, $\quad \mu^*=\tfrac{1}{2}(z_{(m)}+z_{(m+1)})$.

In this example the nearly best estimate is identical with the maximum likelihood estimate (cf. 14.5).

As seen from Sarhan (1954, p. 321), the efficiency (in the linear sense) of the median is 98.9 % when $n=4$ and 90.2 % when $n=5$.

(4) *Rectangular distribution.*

We take μ as the mean and σ as the range of variation of z. Suppose that we have a censored sample of the type considered in 10.5 with $r_1=r_2=r$.

The values of θ_i, C_{1i}, and C_{2i} were given in 9.2, p. 113. By (10.5.1)

$$C_{1i}^*=\begin{cases}-\dfrac{1}{r+1}\\[2ex] 0,\\[1ex] \dfrac{1}{r+1}\end{cases}\qquad C_{2i}^*=\begin{cases}\dfrac{n-2r-1}{2(n+1)(r+1)} & \text{for } 0\leqslant i\leqslant r\\[2ex] -\dfrac{1}{n+1} & \text{for } r+1\leqslant i\leqslant n-r-1\\[2ex] \dfrac{n-2r-1}{2(n+1)(r+1)} & \text{for } n-r\leqslant i\leqslant n.\end{cases}$$

Inserting these values in (10.3.4), we find

$$d_{11}=\frac{2}{r+1},\ d_{12}=0,\ d_{22}=\frac{n-2r-1}{2(n+1)(r+1)}.$$

Hence by (10.3.5)

$$d^{11}=\frac{2(n+1)(r+1)}{n-2r-1},\ d^{12}=0,\ d^{22}=\frac{r+1}{2}.$$

Thus we obtain from (10.3.13) the nearly best estimates

$$\mu^*=\tfrac{1}{2}(z_{(n-r)}+z_{(r+1)}),$$

$$\sigma^*=\frac{n+1}{n-2r-1}(z_{(n-r)}-z_{(r+1)}),$$

which, in this special case, coincide with the best linear estimates. Further, by (10.3.14)

$$\text{var}\ \mu^* = \frac{(r+1)\,\sigma^2}{2\,(n+1)\,(n+2)}, \quad \text{var}\ \sigma^* = \frac{2\,(r+1)\,\sigma^2}{(n+2)\,(n-2\,r-1)},$$

where we may use equality signs, since the error terms are zero.

If no sample-values are missing, we take $r=0$, and μ^* and σ^* are specialized to the well-known estimates already considered in 9.2. For comparison, the variances of μ^* and σ^* obtained for $r=0$ and $n=5$ have been included in Table 5.

The examples given in this section concern distributions of very different shapes. In all cases the nearly best estimate is remarkably efficient (in the linear sense) in view of the small sample-size used in the numerical calculations; in several cases it is extraordinarily good, the variance being practically the same as that of the best linear estimate. The results seem to warrant the conclusion that *the exact method for finding best estimates can, in many cases encountered in practical applications, be replaced by the approximate method without any appreciable loss of efficiency.*

It should, however, be pointed out that it is possible that other distributions behave less favourably than those considered here. Further research is needed to clarify this question.

Finally, it should be mentioned that we may compare the nearly best estimate not only with the unbiased best linear estimate but also with the unbiased best general estimate. By the latter term we denote the estimate with the smallest variance in the class of all unbiased estimates, linear or non-linear. By dividing the variance of this estimate with the variance of the nearly best estimate, we obtain the *efficiency* (*in the general sense*) of the latter estimate.

As an example, we consider once more the estimation of σ in a normal cdf. We saw in Table 5 that the efficiency of σ^* (in the linear sense) is 99.7 %. The corresponding efficiency (in the general sense) is 98.6 % (cf. Chernoff & Lieberman, 1954, Table II). Thus σ^* compares very favourably to any other unbiased estimate.

We shall return to these questions in Chapters 13 and 14.

10.9. Estimation of percentage points

In the above applications of the general method for constructing nearly best estimates we have everywhere assumed that only μ and σ should be estimated. In this section we shall draw attention to a fact, which is obvious, though it has not been mentioned hitherto, viz. that any percentage point of the distribution can be estimated by the same method. For brevity, we shall only discuss the situation considered in 10.3, i.e. we assume that both μ and σ are unknown and, further, that the means $E x_i$ are accurately known.

Let z_P be an arbitrary percentage point defined by

$$F\left(\frac{z_P-\mu}{\sigma}\right)=P, \quad (0 \leqslant P \leqslant 1).$$

It follows from 10.3 that, in order to estimate z_P, we have only to identify this quantity with the general parameter $k_1\mu+k_2\sigma$ studied in this section. Evidently we have in this case

$$k_1=1, \quad k_2=G(P). \tag{10.9.1}$$

We insert these values in (10.3.6) and the resulting values of a_1 and a_2 in (10.3.8). The estimate z_P^* with the coefficients g_i, which are then obtained, is an unbiased nearly best linear estimate of z_P. Obviously,

$$z_P^*=\mu^*+\sigma^* G(P). \tag{10.9.2}$$

It is interesting to consider the variance of z_P^* when P is given different values between 0 and 1. Inserting (10.3.6) in (10.3.7) with the values of k_1 and k_2 given above, we infer from the latter formula and from (10.3.11) that, apart from the remainder term, the variance of z_P^* is proportional to the quadratic

$$d^{22}[G(P)]^2+2d^{12}G(P)+d^{11}.$$

Letting P vary from 0 to 1, we see that this expression attains a minimum when P is equal to a value P_0 defined by

$$G(P_0)=-\frac{d^{12}}{d^{22}}=\frac{d_{12}}{d_{11}}.$$

It should be observed that P_0 is a function of n, which commonly converges to a limit when $n \to \infty$.

The percentage point corresponding to P_0 is of some special interest, because, of all such points, it is "easiest" to estimate (for any given sample-size). The more P differs from P_0, the greater becomes the variance of the corresponding percentage point estimate. Note, however, that these statements are true only in an approximate sense, as we have neglected the influence of the remainder term in (10.3.11).

Moreover, it is interesting to observe that, as is readily seen, the variance of the 'optimal' percentage point estimate $z^*_{P_0}$ is given approximately by the same expression (10.4.2) as was obtained for the variance of the nearly best estimate of μ when σ is known *a priori*. Thus, apart from the error term, var z_P is always greater than or equal to this expression.

CHAPTER 11

LINEAR ESTIMATES WITH NEARLY MINIMUM MEAN SQUARE DEVIATION

11.1. Introduction

In Chapters 9 and 10 we have tacitly assumed that only unbiased or nearly unbiased estimates are worth considering. In this chapter we shall (a) drop the condition of unbiasedness and (b) use the *mean square deviation* (*MSD*) *about the true value of the parameter* as our measure of the precision of an estimate. An estimate will be considered good, if its *MSD* is small.

The problem of constructing estimates with the least possible *MSD* has been studied by Goodman (1953) and Chernoff & Lieberman (1954), among others.

The discussion in this chapter will concern the estimation of σ in the cdf. $F[(z-\mu)/\sigma]$ by means of a linear estimate of the form (9.1.1). More specifically, we shall consider the class of all linear estimates σ^* of σ the means of which are proportional to σ. From among this class, we select the estimate which has the property that its *MSD* about σ

$$E(\sigma^* - \sigma)^2$$

is as small as possible.

Just as in the corresponding problem for unbiased linear estimates, the exact solution is laborious to apply in practice, and we shall therefore construct a nearly best estimate in the sense that the *MSD* is nearly minimum. It will be shown that this may be accomplished by a slight variation of the procedure developed in Chapter 10.

We shall only consider the situation which arises when the variances and covariances of the *TRB*-variables are unknown but their means are accurately known from tables, etc. The case in which no such accurate values are available is treated in the same manner as in 10.6 and will not be discussed here.

11.2. Determination of an estimate with nearly minimum mean square deviation

We shall follow the exposition in 10.3 as closely as possible, using the expressions (9.2.7) and (9.2.13) as our starting-point.

The relations (9.2.13) express the conditions which a linear function U of the ordered sample-values must satisfy in order to be an unbiased estimate of $k_1\mu + k_2\sigma$, where k_1 and k_2 are given quantities. In the present problem, the mean of U should be proportional to σ. Hence $k_1 = 0$, and the first relation (9.2.13) becomes

$$\sum_{i=0}^{n} C_{1i} h_i = 0. \tag{11.2.1}$$

We see, further, that k_2 is no longer a fixed quantity but a function of the coefficients h_i determined by

$$k_2 = \sum_{i=0}^{n} C_{2i} h_i. \tag{11.2.2}$$

The formula (9.2.7) can be used for determining the *MSD*. Replacing U by σ^*, we find after a simple calculation

$$E(\sigma^* - \sigma)^2 = \operatorname{var}\sigma^* + (E\sigma^* - \sigma)^2 = \frac{\sigma^2 Y}{(n+1)(n+2)} + \sigma^2 R_{\text{var}}, \tag{11.2.3}$$

where
$$Y = \sum_{i=0}^{n} h_i^2 - \frac{1}{n+1}\left(\sum_{i=0}^{n} h_i\right)^2 + (n+1)(n+2)(k_2 - 1)^2 \tag{11.2.4}$$

with k_2 given by (11.2.2).

We now determine the coefficients h_i, so that (11.2.1) is satisfied and Y is minimized. The same formal solution (10.3.2) is obtained as in 10.3. In the present case, however, only a_1 is a multiplier, a_2 being given by

$$a_2 = (n+1)(n+2)(1 - k_2). \tag{11.2.5}$$

The system of equations (10.3.3) is valid if we take $k_1 = 0$. The coefficients d_{rs} are given by the usual expressions (10.3.4). Using (11.2.5), we find

$$\begin{aligned} a_1 &= -\left[|D| + \frac{d_{11}}{(n+1)(n+2)}\right]^{-1} d_{12}, \\ a_2 &= \left[|D| + \frac{d_{11}}{(n+1)(n+2)}\right]^{-1} d_{11}. \end{aligned} \tag{11.2.6}$$

Further,
$$Y_{\min} = a_2 k_2 + (n+1)(n+2)(k_2 - 1)^2 = a_2. \tag{11.2.7}$$

The coefficients g_i of the estimate σ_1^* with nearly minimum *MSD* have the general form (10.3.8) with a_1 and a_2 determined by (11.2.6).

As seen from (11.2.3) and (11.2.7), the *MSD* about σ is given approximately by

$$E(\sigma_1^* - \sigma)^2 \sim \frac{\sigma^2 d_{11}}{d_{11} + (n+1)(n+2)|D|}. \tag{11.2.8}$$

Let us now compare σ_1^* with the unbiased nearly best estimate σ^* derived in 10.3. We see that the estimates are identical, apart from a multiplicative factor. We have, in fact,

$$\sigma_1^* = \left[1 + \frac{d_{11}}{(n+1)(n+2)|D|}\right]^{-1} \sigma^*. \tag{11.2.9}$$

Further, a comparison of (11.2.8) and the second expression (10.3.14) shows that the leading term of the *MSD* of σ_1^* is less than the corresponding term of var σ^*. This does not prove, however, that σ_1^* is more efficient (in the sense of having a smaller *MSD*), since the error terms have then to be taken into consideration. A gain of efficiency can, however, be obtained in this way, as shown by the following simple example.

Rectangular distribution.

The nearly best estimate σ^* of the range of variation σ is obtained by taking $r=0$ in 10.8, Ex. (4). It is seen that

$$1 + \frac{d_{11}}{(n+1)(n+2)|D|} = \frac{n(n+1)}{(n-1)(n+2)}.$$

Using (11.2.9), we obtain

$$\sigma_1^* = \frac{n+2}{n}(z_{(n)} - z_{(1)}).$$

The error term discarded in (11.2.8) is zero in this special case. Hence

$$E(\sigma_1^* - \sigma)^2 = \frac{2\sigma^2}{n(n+1)},$$

which, as seen from 10.8, is less than the variance of σ^*.

A second example will be given in 12.4.

CHAPTER 12

MODIFIED NEARLY BEST LINEAR ESTIMATES

12.1. Introduction

The method for constructing nearly best estimates is, as we have shown in Chapter 10, of wide application. In certain cases, however, it is of interest to change the general procedure. The modification consists in a substitution of derivatives for differences in the expressions for the nearly best estimate. Certain continuity conditions must then be imposed upon $G(u)$ and its derivatives.

The modified method can be applied to both complete and censored samples. As we shall see later, it bears an interesting relation to the work of Jung (1955) and, in the special case of censored samples from a normal distribution, to the work of Gupta (1952).

In 12.6 we shall study the relationship between modified nearly best estimates and graphical estimates obtained by means of probability papers.

12.2. Derivation of the modified estimates

Suppose that Condition A in 3.5 is satisfied with $\mathfrak{C}$ equal to the open interval $0<u<1$ (note that $G(u)$ may be unbounded for $u=0$ or 1). Moreover, let us assume that the functions $f(x)=1/G'(u)$ and $xf(x)=G(u)/G'(u)$ tend to 0 when u tends 0 or 1 (cf. 13.2). It is then possible to replace the differences in the definitions (9.2.9) and (9.2.10) of C_{1i} and C_{2i} by derivatives, which, for convenience, will be expressed in terms of $f(x)$.

By (9.2.6) and (9.2.9) we have

$$C_{1i}=f(\lambda_i)-f(\lambda_{i+1}).$$

It should be noted that this relation is satisfied not only for $i=1, \ldots, n-1$ but also for $i=0$ and $i=n$, since $f(x)$ is zero at the ends of the range of variation of ξ. Using the notation (9.2.5), we have approximately

$$C_{1i}\sim -\frac{1}{n+1}f'(\lambda_i)\,G'\left(\frac{i}{n+1}\right)=-\frac{1}{n+1}\,\frac{f'(\lambda_i)}{f(\lambda_i)}.$$

Similarly,

$$C_{2i} \sim \lambda_i f(\lambda_i) - \lambda_{i+1} f(\lambda_{i+1}) \sim -\frac{1}{n+1}\left[1 + \frac{\lambda_i f'(\lambda_i)}{f(\lambda_i)}\right].$$

In accordance with Jung (1955), we use the notations

$$\gamma_1(x) = -\frac{d \log f(x)}{dx}, \quad \gamma_2(x) = -1 - x\frac{d \log f(x)}{dx},$$
$$\gamma_1'(x) = \frac{d\gamma_1(x)}{dx}, \quad \gamma_2'(x) = \frac{d\gamma_2(x)}{dx}. \tag{12.2.1}$$

Hence $$C_{ri} \sim \frac{1}{n+1}\gamma_r(\lambda_i), \quad (r=1, 2;\ i=0, 1, \ldots, n). \tag{12.2.2}$$

Moreover, we find

$$\theta_i(C_{ri} - C_{r\,i-1}) \sim \frac{1}{(n+1)^2}\gamma_r'(\lambda_i), \quad (r=1, 2;\ i=0, 1, \ldots, n). \tag{12.2.3}$$

Thus the coefficients (10.3.8) of the nearly best estimate can be approximated by

$$\tilde{g}_i = a_1\gamma_1'(\lambda_i) + a_2\gamma_2'(\lambda_i), \tag{12.2.4}$$

a factor $(n+1)^{-2}$ having being included in the multipliers. More directly, this expression can also be obtained from (10.3.9). The estimate

$$\tilde{U} = \sum_{i=1}^{n} \tilde{g}_i z_{(i)} \tag{12.2.5}$$

with these coefficients will be called a *modified nearly best linear estimate* of $k_1\mu + k_2\sigma$. The question of the determination of the multipliers will be treated in 12.4.

When only one of the parameters μ and σ is unknown, the same reasoning holds good with the following modifications: Suppose, e.g., that μ is unknown. The function $f(x)$ (but not necessarily $x f(x)$) should then vanish in the end-points. Moreover, (12.2.4) is replaced by

$$\tilde{g}_i = a_1\gamma_1'(\lambda_i).$$

The case of an unknown σ is analogous.

12.3. Discussion

It is interesting to observe that, if the discussion is limited to linear estimates defined by means of a *continuous* weight function, the coefficients $\tilde{g}_i$ can also be derived by means of formulae (8.13.2). It follows, indeed, from these approximate expressions that the mean and the variance of a linear estimate

$$U = \sum_{i=1}^{n} g_i z_{(i)} = \sum_{i=1}^{n} g_i (\mu + \sigma x_i)$$

are, to the same degree of approximation, given by

$$E\,U \sim \int (\mu + \sigma x)\, h'(x)\, f(x)\, dx,$$

$$\operatorname{var} U \sim \frac{\sigma^2}{n} \left[\int [h(x)]^2 f(x)\, dx - \left(\int h(x)\, f(x)\, dx\right)^2\right].$$

Setting the first expression equal to $k_1 \mu + k_2 \sigma$, we obtain two side conditions, which correspond to the conditions (9.2.13) obtained in the general case. By the calculus of variation we determine $h(x)$ so that the second expression is minimized. Applying (8.13.1) to the resulting function, we obtain (12.2.4). Note that this method is less general than that used in Chapter 10 and in 12.2, where we have used no *a priori* assumption concerning the choice of the coefficients g_i.

The general form of the coefficients (12.2.4) is previously known (Jung, 1955), the derivation just outlined being a simplified version of his method.

12.4. Determination of the multipliers. Examples

There are two alternatives, depending on whether the means $E\,x_i$ have been tabulated or not.

(a) *Unbiased estimates.*

$\tilde{U}$ in (12.2.5) is an unbiased estimate of $k_1 \mu + k_2 \sigma$, if the multipliers a_1 and a_2 in (12.2.4) are determined from the system

$$\begin{aligned} a_1 \sum_{i=1}^{n} \gamma_1'(\lambda_i) + a_2 \sum_{i=1}^{n} \gamma_2'(\lambda_i) &= k_1, \\ a_1 \sum_{i=1}^{n} \gamma_1'(\lambda_i)\, E\,x_i + a_2 \sum_{i=1}^{n} \gamma_2'(\lambda_i)\, E\,x_i &= k_2. \end{aligned} \qquad (12.4.1)$$

In the symmetrical case we see that

$$a_1 = k_1 \left[\sum_{i=1}^{n} \gamma_1'(\lambda_i)\right]^{-1}, \qquad a_2 = k_2 \left[\sum_{i=1}^{n} \gamma_2'(\lambda_i)\, E\,x_i\right]^{-1}. \qquad (12.4.2)$$

(b) *Nearly unbiased estimates.*

In (12.4.1), replace λ_i and $E x_i$ by $\lambda_i' = G(\pi_i)$, where π_i is given by (9.2.15). The resulting estimate is only slightly biased if α and β are chosen appropriately (cf. Chapter 6).

We shall give two examples.

(1) *Normal distribution.*

The normal cdf. satisfies the conditions stated in the first paragraph of 12.2. By (12.2.1)

$$\gamma_1(x) = x, \qquad \gamma_2(x) = x^2 - 1.$$

(a) *Unbiased estimates.*

It follows from (12.2.4) and (12.4.2) that the modified unbiased nearly best estimates of μ and σ are

$$\tilde{\mu} = \bar{z}, \qquad \tilde{\sigma} = \frac{\sum_{i=1}^{n} \lambda_i z_{(i)}}{\sum_{i=1}^{n} \lambda_i E x_i},$$

where

$$\lambda_i = \Phi^{-1}\left(\frac{i}{n+1}\right).$$

The first estimate is 100 % efficient. The second estimate has practically the same efficiency for all sample-sizes as σ^* derived in 10.8, Ex. (1). For example, if $n=5$, the efficiency is 99.9 % (in the linear sense) and 98.7 % (in the general sense) (cf. pp. 127 and 130). The high efficiency of $\tilde{\sigma}$, together with its simplicity, makes this estimate very suited for practical application.

(b) *Nearly unbiased estimates.*

The estimate $\tilde{\sigma}$ is altered to

$$\tilde{\sigma}' = \frac{\sum_{i=1}^{n} \lambda_i' z_{(i)}}{\sum_{i=1}^{n} \lambda_i'^2},$$

where

$$\lambda_i' = \Phi^{-1}\left(\frac{i-\alpha}{n-2\alpha+1}\right).$$

A numerical investigation has been performed, which is intended to demonstrate that, in addition to its main purpose of providing a nearly unbiased estimate of high efficiency, $\tilde{\sigma}'$ may, by a change of α, also provide a biased estimate with nearly minimum *MSD* (cf. Chapter 11).

TABLE 6.

Normal distribution. Mean of nearly unbiased nearly best estimate $\tilde{\sigma}'$.

n	α			
	0	0.35	0.375	0.5
5	1.193	1.002	0.987	0.913
6	1.174	1.004	0.991	0.924
7	1.159	1.005	0.993	0.932
8	1.147	1.006	0.995	0.939
9	1.137	1.006	0.996	0.944
10	1.128	1.007	0.997	0.948
15	1.099	1.007	1.001	0.963

Each entry should be multiplied by σ.

In Table 6 the mean of $\tilde{\sigma}'$ has been tabulated for some sample-sizes. In accordance with the results obtained in 6.10, the bias is small for $n = 5 - 15$ when α lies in the interval 0.350–0.375. When $\alpha = 0$ or $\alpha = \frac{1}{2}$, there is a considerable bias for $n = 5$, which diminishes when n increases.

As a measure of the precision of $\tilde{\sigma}'$ we use the *MSD* about σ. Table 7 gives the *MSD* for the same values of α as in Table 6 and, in addition, the variance of the unbiased best linear estimate. The entries for $\alpha = 0$ and $\alpha = \frac{1}{2}$ have been taken from Chernoff & Lieberman (1954, Table II). The moments of the normal order statistics required for constructing Table 6 and Table 7 have been obtained from Godwin (1949 *b*) and Teichroew (1956).

TABLE 7.

Normal distribution. Mean square deviation about σ of nearly unbiased nearly best estimate $\tilde{\sigma}'$.

n	α				Variance of best linear estimate
	0	0.35	0.375	0.5	
5	0.228	0.134	0.130	0.119	0.133
6	0.176	0.107	0.104	0.096	0.106
7	0.143	0.088	0.086	0.081	0.088
8	0.120	0.076	0.074	0.070	0.075
9	0.103	0.066	0.065	0.061	0.065
10	0.090	0.058	0.057	0.054	0.058

Each *MSD* should be multiplied by σ^2.

It is observed from Table 7 that

(a) the precision of $\tilde{\sigma}'$ for $\alpha=0.35$ and $\alpha=\frac{3}{8}=0.375$ is about the same as that of the best linear estimate.

As a suitable single value of α we suggest the value $\alpha=\frac{3}{8}$ already considered in 6.10.

(b) the biased estimate obtained for $\alpha=\frac{1}{2}$ has the smallest *MSD* of the estimates entered in the table.

The latter statement illustrates the situation discussed in Chapter 11: If some bias is allowed, it is possible to construct estimates which have greater precision than the best unbiased estimate. It might be added that, by considering values in the neighbourhood of $\alpha=\frac{1}{2}$, it is possible to find estimates with a *MSD* which is somewhat smaller than the *MSD* for $\alpha=\frac{1}{2}$. We shall, however, not consider this possibility, because, as seen from column 3 in Chernoff & Lieberman's table referred to above, even the *MSD* for $\alpha=\frac{1}{2}$ is only very little in excess of the minimum *MSD* which can be attained by any estimate, linear or non-linear.

We conclude from these statements that it is advantageous to use the modified estimate $\tilde{\sigma}'$, both when nearly unbiased estimates are required and when biased estimates can be used.

(2) *Weibull's distribution.*

We shall discuss the estimation of σ in the cdf. (cf. 6.14)

$$F\left(\frac{z}{\sigma}\right)=1-\exp\left[-\left(\frac{z}{\sigma}\right)^m\right], \quad (z\geqslant 0,\ m\geqslant 1).$$

This is a one-parameter situation of the type discussed at the end of 12.2. Condition A is satisfied for $0<u<1$. Further, the function $x\,f(x)$ tends to zero in the end-points. Hence we are entitled to apply the results in 12.2. By (12.2.1)

$$\gamma_2(x)=m\,(x^m-1).$$

(a) *Unbiased estimate.*

The modified unbiased nearly best estimate is

$$\tilde{\sigma}=\frac{\sum\limits_{i=1}^{n}\lambda_i^{m-1}\,z_{(i)}}{\sum\limits_{i=1}^{n}\lambda_i^{m-1}\,E\,x_i},$$

where

$$\lambda_i=\left(\log\frac{n+1}{n-i+1}\right)^{1/m}.$$

When $m=1$ Weibull's distribution is specialized to the exponential distribution; $\tilde{\sigma}$ is then identical with the best linear estimate $\bar{z}$.

(b) *Nearly unbiased estimate.*

We have
$$\tilde{\sigma}' = \frac{\sum_{i=1}^{n} \lambda_i'^{m-1} z_{(i)}}{\sum_{i=1}^{n} \lambda_i'^{m}},$$

where
$$\lambda_i = \left(\log \frac{n-\alpha-\beta+1}{n-i-\beta+1}\right)^{1/m}.$$

The question of the choice of α and β was briefly discussed in 6.14.

As the best linear estimate has not been determined numerically for Weibull's distribution, no investigation of the efficiency of $\tilde{\sigma}$ and $\tilde{\sigma}'$ has been performed.

12.5. Censored samples

The method described above can easily be adapted to the situation where the r_1 smallest and r_2 highest sample-values are missing. In the system (12.4.1) we have, in fact, only to alter the summation and let the sums run from r_1+1 to $n-r_2$. In all other respects the procedure is the same as before. We shall give an example.

Normal distribution.

We shall assume that both μ and σ are unknown.

(a) *Unbiased estimates.*

We obtain the system
$$a_1(n-r_1-r_2) + a_2 \Sigma \lambda_i = k_1,$$
$$a_1 \Sigma E x_i + a_2 \Sigma \lambda_i E x_i = k_2.$$

Here and subsequently the sums run from r_1+1 to $n-r_2$. After some calculation we obtain the following estimates
$$\tilde{\mu} = \Sigma g_{1i} z_{(i)}, \quad \tilde{\sigma} = \Sigma g_{2i} z_{(i)}.$$

The coefficients are determined by
$$g_{1i} = \frac{1}{n-r_1-r_2} - g_{2i} \overline{E x_i},$$
$$g_{2i} = \frac{\lambda_i - \bar{\lambda}}{\Sigma (\lambda_i - \bar{\lambda})(E x_i - \overline{E x_i})},$$

where
$$\bar{\lambda} = \frac{1}{n - r_1 - r_2} \Sigma \lambda_i,$$
$$\overline{E x_i} = \frac{1}{n - r_1 - r_2} \Sigma E x_i.$$

For $r_1 = r_2 = 0$ the estimates are specialized to those obtained in 12.4, p. 139.

$\tilde{\mu}$ and $\tilde{\sigma}$ should be compared with the closely related estimates which have been discussed by Gupta (1952) and investigated numerically by Sarhan & Greenberg (1956).

(b) *Nearly unbiased estimates.*

By solving the modified system (12.4.1) or, more simply, by substituting λ_i' for $E x_i$ in $\tilde{\mu}$ and $\tilde{\sigma}$, we obtain

$$\tilde{\mu}' = \Sigma g_{1i}' z_{(i)}, \quad \tilde{\sigma}' = \Sigma g_{2i}' z_{(i)},$$

where
$$g_{1i}' = \frac{1}{n - r_1 - r_2} - \bar{\lambda}' g_{2i}',$$
$$g_{2i}' = \frac{\lambda_i' - \bar{\lambda}'}{\Sigma (\lambda_i' - \bar{\lambda}')^2}.$$

12.6. Application to probability papers

As already mentioned in 12.1, there is a connection between certain of the results obtained in the preceding sections and the use of probability papers.

Let $F(x)$ be an arbitrary continuous cdf. and $G(u)$ its inverse. The probability paper is so designed that one of the scales, say the vertical, is proportional to $G(u)$ and consequently non-linear. This paper can be used for graphical estimation of the parameters in a cdf. of the form $F[(z-\mu)/\sigma]$ when a random sample of size n is available. The ordered sample-values $z_{(i)}$, $(i = 1, \ldots, n)$, are then plotted against P_i on the vertical axis, where $P_1, \ldots, P_n$ is a set of quantities to be discussed presently, and a straight line is fitted visually to the points.

In accordance with Chernoff & Lieberman (1954), we shall assume that *the fitted line is identical with the line obtained by minimizing the sum of squares of the horizontal deviations from the line.*

Let us introduce the notations

$$v = G(P), \qquad v_i = G(P_i). \tag{12.6.1}$$

The position of any point on the paper is determined by the abscissa z and the ordinate v. By means of these symbols the true line can be written

$$z=\mu+\sigma v,$$

and the fitted line $$z=\mu_g+\sigma_g v,$$

where μ_g and σ_g are the graphical estimates of μ and σ. Applying the usual formulae of regression analysis, we find

$$\mu_g=\bar{z}-\sigma_g \bar{v},$$
$$\sigma_g=\frac{\sum_{i=1}^{n}(v_i-\bar{v})\,z_{(i)}}{\sum_{i=1}^{n}(v_i-\bar{v})^2}, \tag{12.6.2}$$

where $$\bar{z}=\frac{1}{n}\sum_{i=1}^{n} z_{(i)}, \qquad \bar{v}=\frac{1}{n}\sum_{i=1}^{n} v_i.$$

Thus μ_g and σ_g are linear estimates. They depend upon the quantities P_i, which are as yet undetermined.

A common plotting rule is $P_i=i/(n+1)$, which seems to have been first proposed by Weibull (1939, p. 28 ff.). It has been much recommended by Gumbel (1954, p. 13 ff.). Another well-known rule is $P_i=(i-\frac{1}{2})/n$, which has been discussed by Bliss (1937) and Ipsen & Jerne (1944), among others. We shall consider the general expression

$$P_i=\pi_i=\frac{i-\alpha}{n-\alpha-\beta+1}, \quad (\alpha,\,\beta\leqslant 1), \tag{12.6.3}$$

of which the rules mentioned above are special cases. We observe that v_i is then specialized to $\lambda_i'=G(\pi_i)$. The given expression has the following advantages if α and β are suitably chosen.

(a) The graphical estimates are nearly unbiased.
(b) The points $(z_{(i)};\ P_i)$, $(i=1,\ldots,n)$, lie on the average on a line, which deviates only very little from a straight line.
(c) The graphical estimates are highly efficient in the important case of the normal distribution.

The truth of statements (a) and (b) follows from the discussion in Chapter 6. Formula (6.9.1) gives, in fact,

$$E\,z_{(i)}=\mu+\sigma E\,x_i\sim\mu+\sigma G(\pi_i).$$

In order to verify the truth of statement (c), we shall consider the graphical estimates of the mean and the standard deviation of a normal population obtained by aid of a normal probability paper. We then take $\alpha = \beta$ in (12.6.3); the choice of α will be discussed below.

It is seen that $\bar{v} = 0$ for any choice of α; hence μ_g in (12.6.2) is specialized to $\bar{z}$. Moreover, σ_g is seen to coincide with the modified nearly best estimate $\tilde{\sigma}'$, discussed in 12.4, Ex. (1). Applying the conclusions made in this example to the present problem, we find that, *in the special case of a normal distribution,*

(1) *the plotting rule*

$$P_i = \frac{i - \frac{3}{8}}{n + \frac{1}{4}}$$

leads to a practically unbiased estimate of σ with a MSD about σ which is about the same as that of the unbiased best linear estimate,

(2) *the plotting rule*

$$P_i = \frac{i - \frac{1}{2}}{n}$$

leads to a biased estimate of σ with nearly minimum MSD about σ.

The latter statement is due to Chernoff & Lieberman (1954). Accordingly, the first rule should be applied when unbiasedness is essential, and the second rule otherwise. The rule $P_i = i/(n+1)$, on the other hand, cannot be recommended for use in the normal case if high efficiency is aimed at and the sample-size is small.

It should be mentioned that the expression obtained by taking $\alpha = \beta$ in (12.6.3) has been recommended as a plotting rule for *any* probability paper by Benard & Bos-Levenbach (1953), who, for other reasons than those given above, proposed the value $\alpha = 0.3$, irrespective of the shape of the distribution.

It is interesting to observe that, if a censored sample from a normal population is plotted according to the rule (12.6.3), and the graphical estimates are determined by an application of the least-squares principle described above, these estimates coincide with $\tilde{\mu}'$ and $\tilde{\sigma}'$ derived in 12.5. The proof of this statement is simple and will not be given here.

Finally, though it is outside the scope of the present chapter, it should be mentioned that, if other probability papers than the normal are considered, the rule (12.6.3) generally does *not* furnish graphical estimates with high efficiency. As a simple example we may take the rectangular distribution. Suppose that an ordered random sample from this distribution is plotted according to the rule $P_i = i/(n+1)$ on a paper with linear scales in both directions. It then follows from (12.6.2) that the graphical estimates of the mean μ and the range of variation σ are identical with the arithmetic mean and Gini's mean difference, respectively, which, as seen from the example in 9.2, p. 113 ff., are very inefficient as compared with the best linear estimates based upon the extreme values of the sample.

More generally, it may be shown by an application of the formulae in 8.13 that, for any cdf. satisfying certain general conditions, the graphical estimate σ_g has approximately the variance

$$\frac{\sigma^2}{4n} \frac{\mu_4 - \mu_2^2}{\mu_2^2},$$

where μ_2 and μ_4 denote the second and fourth moments, respectively, of the reduced variable ξ around its mean.

Thus, when the sample-size is large, σ_g has approximately the same variance as $s/\sqrt{\mu_2}$, where s is the standard deviation

$$s = \frac{1}{n-1} \sqrt{\sum_{i=1}^{n} (z_i - \bar{z})^2}$$

of the sample[1] (cf. Cramér, 1946 *a*, p. 353). This relation "explains" in a way, why the normal probability paper works so well; we know that s is a good estimate in this case, and hence this is true also of σ_g. We also understand that the normal distribution is exceptional in this respect; s (and thus also σ_g) is for the most part an inefficient estimate of a scale parameter.

Similar remarks apply to the location parameter μ.

[1] Since $z = \mu + \sigma\xi$, the standard deviation satisfies $E\,s \sim \sigma\sqrt{\mu_2}$; this explains why we have divided s by $\sqrt{\mu_2}$.

CHAPTER 13

ASYMPTOTIC PROPERTIES OF NEARLY BEST LINEAR ESTIMATES

13.1. Introduction

The purpose of this chapter is three-fold.

(a) It will be shown that nearly best linear estimates of location and scale parameters are generally asymptotically efficient (in the general sense considered in Part I).

(b) The order of magnitude of the variances and covariances of nearly best linear estimates will be analysed in various situations.

(c) The asymptotic distribution of nearly best linear estimates will be determined in an important special case.

Both the case of two unknown parameters and the case of a single unknown parameter will be considered. In order to solve the problems involved in (a) and (b), we shall use the general results concerning the asymptotic behaviour of estimates presented in Part I of this thesis. The classification of cdf:s introduced in this part will also be extensively used in the chapter.

Some preliminary remarks will be made in 13.2 and 13.3. In 13.4 and 13.5, we shall investigate the asymptotic properties of the nearly best estimates when the cdf. belongs uniformly to Type 1. It will be demonstrated that, under general conditions, these estimates are asymptotically jointly efficient, asymptotically jointly normally distributed, and, in accordance with the general theory, have variances and, generally, a covariance of order n^{-1}. Similar results apply to the case of a single unknown parameter.

In 13.6 and 13.7, we shall study nearly best estimates from Type 2 distributions. As follows from Part I, such distributions occur when the fr.f. possesses discontinuity points in the interior or at the ends of the range of variation of the distribution. The discussion will be limited to the latter alternative. Both uniform Type 2 distributions and other Type 2 distributions will be considered. It will be shown that nearly best estimates from Type 2 distributions show a different behaviour than estimates from Type 1 distributions; they are not asymptotically normal, and their variances and covariance generally (but not always) converge as n^{-2}. However, they have an important property in common with estimates from Type 1 distributions, namely, they are generally asymptotically jointly efficient.

In 13.8, some remarks will be made concerning nearly best linear estimates from Type 3 distributions.

Finally, in 13.9, a summary of certain results is given and also a table illustrating the behaviour of nearly best estimates corresponding to B-transforms.

13.2. Classification of distributions with unknown location and scale parameters

As in the foregoing chapters in Part III, we shall consider the estimation of μ and σ in a differentiable cdf. $F[(z-\mu)/\sigma]$. The corresponding fr.f. is

$$f_1(z; \mu, \sigma) = \frac{1}{\sigma} f[(z-\mu)/\sigma].$$

Let (A, B) be the range of variation[1] of the reduced variable

$$\xi = (z-\mu)/\sigma.$$

Suppose that nearly best linear estimates of μ and σ have been constructed according to the method described in Chapter 10. In order to facilitate the study of the asymptotic properties of these estimates, we shall use the classification introduced in Part I. There are two situations, depending on whether both parameters are unknown, or whether only one of these is unknown. We shall reverse the order of presentation as compared with Part I and begin with the former alternative.

(a) *μ and σ are unknown.*

The classification relevant to this situation was given in 2.3. Identifying the general fr.f. $f(z; \alpha_1, \alpha_2, \ldots)$ considered there with $f_1(z; \mu, \sigma)$, we infer from (2.3.1) after a simple calculation that e_{rs} in this formula becomes equal to

$$\frac{1}{\sigma^2} \int_A^B \gamma_r(x)\, \gamma_s(x)\, f(x)\, dx,$$

[1] It should be noted that, in Part I, the interval (A, B) denotes the range of variation of z; the change of notation will, it is hoped, not cause confusion.

where, in accordance with (12.2.1),

$$\gamma_1(x) = -\frac{d \log f(x)}{d x}, \qquad \gamma_2(x) = -1 - x\frac{d \log f(x)}{d x}.$$

For formal reasons, we shall in this chapter include the factor $1/\sigma^2$ in e_{rs} and therefore write

$$\begin{aligned} e_{rs} &= \int_A^B \gamma_r(x)\, \gamma_s(x)\, f(x)\, dx, \quad (r, s = 1, 2), \\ |E| &= e_{11} e_{22} - e_{12}^2. \end{aligned} \qquad (13.2.1)$$

Moreover, Conditions C 1 and C 2 in 2.3 assume the following form:

(1) *With respect to μ.*

CONDITION C 1. *The function $f(x)$ is a continuous function of x in the interval $A < x < B$, and tends to zero when $x \to A$ or $x \to B$.*

CONDITION C 2. *The quantities e_{rs} exist, and $|E|$ is greater than zero.*

(2) *With respect to σ.*

CONDITION C 1. *The function $x f(x)$ is a continuous function of x in the interval $A < x < B$, and tends to zero when $x \to A$ or $x \to B$.*

CONDITION C 2. (The same condition as in (1).)

(b) *μ or σ is unknown.*

In this case we apply the general rules of classification introduced in 1.3. Suppose, for instance, that μ is unknown. As seen from 1.3, the conditions then become

CONDITION C 1. *The function $f(x)$ is a continuous function of x in the interval $A < x < B$, and tends to zero when $x \to A$ or $x \to B$.*

CONDITION C 2. *The quantity e_{11} exists and is greater than zero.*

Consequently, the conditions are, with a minor modification, the same as those stated in (1) above. The case where σ is unknown is analogous.

The various conditions stated above in (a) and (b) will play an important rôle in the following sections. They enable us to classify any distribution with unknown location and/or scale parameters into one of three types according to the definitions introduced in 2.3 and 1.3, respectively. Especially, the reader's attention is called to the definition of a uniform Type 1 distribution, etc., given in 2.3. As seen from this section, a cdf. of the form $F[(z-\mu)/\sigma]$ is, for example, a uniform Type 1 distribution when both μ and σ are unknown, and, in addition, the conditions stated above in (a), i.e. three conditions in all, are fulfilled. Similarly, F is a uniform Type 2 distribution if, in the pairs of conditions given in (a), the second condition is satisfied, but not the first one.

13.3. A general remark

Before entering upon a discussion of the three types of distribution considered in Part I, we shall make a remark concerning problem (a) in 13.1, which applies to any of these types.

Suppose that the parameters μ and σ are both unknown. The nearly best linear estimates corresponding to this situation were derived in 10.3, and their variances and covariance are, apart from certain error terms, given by (10.3.14) and (10.3.15). In order to obtain information as to the efficiency of these estimates when the sample-size is large, we use the theorem in 2.2, p. 24.

Let us determine the matrix I mentioned in the theorem. Since

$$\frac{\partial F}{\partial \mu} = -\frac{1}{\sigma} f\left(\frac{z-\mu}{\sigma}\right), \qquad \frac{\partial F}{\partial \sigma} = -\frac{1}{\sigma}\frac{z-\mu}{\sigma} f\left(\frac{z-\mu}{\sigma}\right),$$

we realize, using the notations (9.2.5) and (9.2.6), that the quantities I_{rs} in (2.2.8) reduce to

$$I_{11} = \frac{n^2}{\sigma^2}\sum_{i=0}^{n}(\theta_i - \theta_{i+1})^2,$$

$$I_{22} = \frac{n^2}{\sigma^2}\sum_{i=0}^{n}(\lambda_i\theta_i - \lambda_{i+1}\theta_{i+1})^2,$$

and a similar expression for I_{12}. We now use the fact, proved in 5.3, that, under very general conditions, the quantities λ_i and Ex_i are asymptotically equivalent. Then, by (9.2.9), (9.2.10), (10.3.4), and (10.3.5) we infer that the matrices I and D satisfy the relation (cf. p. 36).

$$I \sim \frac{n^2}{\sigma^2} D. \tag{13.3.1}$$

Now consider the inequalities (2.2.12), which contain lower bounds for the variances of any asymptotically unbiased estimates of μ and σ.

From (13.3.1) we immediately conclude that these bounds are asymptotically equivalent to the large-sample expressions (10.3.14) for the variances of the nearly best linear estimates μ^* and σ^*. Furthermore, using also (10.3.15), we see that, apart from the error terms, the generalized variance of μ^* and σ^* attains the lower bound given in (2.2.14).

Consequently, if it can be proved that the error terms discarded in (10.3.14) and (10.3.15) tend to zero in an appropriate way as compared with the leading terms, it follows from the theorem in 2.2 that, whenever this theorem is applicable, the nearly best estimates are asymptotically jointly efficient (in the general sense considered in Chapter 2). For this reason, one of our main tasks in the following sections will be to gain information concerning the behaviour of the error terms.

A similar problem arises when only one of the parameters is unknown. The nearly best linear estimate of μ (or σ) is then constructed according to the procedure described in 10.4. The asymptotic properties of the estimates can be studied by means of the theorem in 1.2. In the same way as above it is realized that the large-sample expressions (10.4.2) and (10.4.4) for the variances attain the lower bound in the general inequality (1.2.11). Hence, if the error terms converge faster than these expressions, we can assert that the nearly best estimate of μ (or σ) is asymptotically efficient (in the general sense considered in Chapter 1).

13.4. A theorem on Type 1 distributions

In the main part of this section we shall assume that both μ and σ are unknown. We shall prove a theorem valid for unbiased nearly best estimates of μ and σ when F is a uniform Type 1 distribution. By $x = G(u)$ we designate as usual the inverse function of $u = F(x)$.

Theorem. *Suppose that one of the following three sets of conditions is satisfied:*

(1) *$G(u)$ satisfies Condition B in 3.5 with $\mathfrak{C}$ equal to the closed interval $0 \leqslant u \leqslant 1$.*

(2) *$G(u)$ is an AL-transform at $u=0$ or $u=1$ (or both) and satisfies Condition B with $\mathfrak{C}$ equal to the remaining part of the unit interval.*

(3) *$G(u)$ is an AP-transform at $u=0$ or $u=1$ (or both) and satisfies Condition B with $\mathfrak{C}$ equal to the remaining part of the unit interval.*

Suppose, further, that the cdf. $F[(z-\mu)/\sigma]$ satisfies Conditions C 1 and C 2 stated in 13.2, alternative (a), *with respect to μ and σ.*

Let μ^* and σ^* be the unbiased nearly best estimates given by (10.3.12) *(with the modification that, when $G(u)$ is an AP-transform at $u=0$ or $u=1$, then $g_i=0$ for $i \leqslant -2k$ or $i \geqslant n+2k+1$, respectively). The second moments of these estimates satisfy the relations*

$$\begin{aligned} \operatorname{var} \mu^* &= \frac{\sigma^2}{n} \frac{e_{22}}{|E|} + o\left(\frac{1}{n}\right), \\ \operatorname{var} \sigma^* &= \frac{\sigma^2}{n} \frac{e_{11}}{|E|} + o\left(\frac{1}{n}\right), \\ \operatorname{cov}(\mu^*, \sigma^*) &= -\frac{\sigma^2}{n} \frac{e_{12}}{|E|} + o\left(\frac{1}{n}\right). \end{aligned} \tag{13.4.1}$$

If the general conditions of the theorem in 2.2 *are satisfied, μ^* and σ^* are asymptotically jointly efficient. Further, they are asymptotically jointly normally distributed with means μ and σ and variances and a covariance given by the leading terms in* (13.4.1).

The quantities e_{rs} and $|E|$ appearing in (13.4.1) are defined by (13.2.1). It should be noted that, when $G(u)$ is an AP-transform, we must have $k<\frac{1}{2}$. For other values of k, Condition C 2 is violated (see further 13.8).

Before proving the theorem, we may remark that it can be given a slightly more general formulation. All results except, possibly, the statement concerning the asymptotic normality are valid if Condition B is replaced by the somewhat more general Condition A. For simplicity, we shall, however, disregard this possibility.

We shall, first, consider alternative (1) in the theorem. Let us examine the behaviour of the matrix D in (10.3.5) when n approaches infinity. We have, for example, by Condition C 1 and by the definition (10.3.4) of the elements in D (cf. also 13.3)

$$d_{11} = \sum_{i=0}^{n} [f(\lambda_i) - f(\lambda_{i+1})]^2, \tag{13.4.2}$$

where as usual $\lambda_i = G[i/(n+1)]$. Replacing the differences by derivatives in the same way as in 12.2, and passing to the limit, we obtain

$$n\, d_{11} \to e_{11} \tag{13.4.3}$$

when $n \to \infty$. More generally, we find

$$\begin{aligned} n\, d_{rs} &\to e_{rs}, \quad (r, s = 1, 2), \\ n^2 |D| &\to |E|. \end{aligned} \tag{13.4.4}$$

Applying these results to the expressions (10.3.14) and (10.3.15) for the variances and the covariance of μ^* and σ^*, we obtain the leading terms in (13.4.1). More directly, this result can be obtained from (13.3.1) and the general statement in the first paragraph of 2.4.

It remains to prove the statements in (13.4.1) concerning the convergence of the error terms. For this purpose, we apply the general results deduced in 8.4 and 8.7 concerning variances and covariances of linear combinations. For example, in (8.4.5), T is identified as μ^* or σ^*, and analogously in (8.4.9). We infer from (8.7.6) and (8.7.10), which evidently apply to the present situation, that the error terms converge to zero at least as rapidly as $n^{-\frac{3}{2}}$. Thus formulae (13.4.1) are proved, and we conclude that μ^* and σ^* are asymptotically jointly efficient (cf. 13.3).

The assertion concerning the asymptotic normality of the estimates is a consequence of the second theorem in 8.8, p. 97, the condition (8.8.5) being fulfilled in this case.

Secondly, we shall consider alternative (2) in the theorem. To save space, we shall omit certain details of the proof. Formulae (13.4.4) are valid even in this case, and thus we obtain the leading terms of (13.4.1) in the same way as before. The remainder terms need, however, closer study. For this purpose, an upper limit is required for the coefficients (10.3.13) of μ^* and σ^*.

First, it follows from (13.4.4) and Condition C 2 in 13.2, alternative (a), that

$$d^{rs} = O(n), \quad (r, s = 1, 2). \tag{13.4.5}$$

Secondly, applying (3.5.2) to θ_i and C_{ri}, and using the notation λ_i° introduced on p. 98, we find after some calculation (cf. 12.2)

$$\begin{aligned} \theta_i \,|C_{1i} - C_{1\ i-1}| &< \frac{M}{n^2}\,|\lambda_i^\circ|^{1/k\,-2}, \\ \theta_i \,|C_{2i} - C_{2\ i-1}| &< \frac{M}{n^2}\,|\lambda_i^\circ|^{1/k\,-1}. \end{aligned} \tag{13.4.6}$$

Combining (13.4.5) and (13.4.6), we find that the coefficients in (10.3.13) satisfy the inequality

$$|g_{ri}| < \frac{M}{n}\,|\lambda_i^\circ|^{1/k\,-1}. \tag{13.4.7}$$

Thus we realize that the conditions of the first theorem in 8.9, p. 97, are fulfilled if we take $d=1$ and $m=1/k$ in (8.9.1). By (8.9.3) we get

$$R_{\text{var}} = O(n^{-5/4}).$$

Hence the error terms in the two variance formulae in (13.4.1) converge more rapidly than n^{-1} as stated. The remainder term in the covariance formula is examined in a similar way (cf. the final paragraph in 8.9).

The statement concerning the asymptotic normality of μ^* and σ^* is a consequence of the two-dimensional version of the theorem in 8.10, p. 103.

It remains to prove the theorem for alternative (3), which is made in the same way as above. Relation (13.4.5) holds good even in this case. The inequalities (13.4.6) are replaced by

$$\begin{aligned} \theta_i \,|C_{1i} - C_{1\;i-1}| &< \frac{M}{n^2}\,|\lambda_i^\circ|^{-2}, \\ \theta_i \,|C_{2i} - C_{2\;i-1}| &< \frac{M}{n^2}\,|\lambda_i^\circ|^{1/k-1}. \end{aligned} \tag{13.4.8}$$

Thus we have from (10.3.13)

$$|g_{ri}| < \frac{M}{n}\,|\lambda_i^\circ|^{m-1}, \tag{13.4.9}$$

where $m = -1$ for $-1 \leqslant k < \frac{1}{2}$

$m = \frac{1}{k}$ for $k \leqslant -1$.

By an application of the theorem in 8.11, p. 104, we conclude that

$$R_{\text{var}} = o\left(\frac{1}{n}\right).$$

It may be similarly shown that the remainder term in the covariance formula in (13.4.1) also converges faster than n^{-1}. Finally, applying the two-dimensional counterpart to the theorem in 8.12, we infer that the estimates are asymptotically jointly normal. This completes the proof.

We shall add a few words regarding the case of a single unknown parameter. Formulae (13.4.1) are then replaced by

$$\operatorname{var} \mu^* = \frac{\sigma^2}{n\,e_{11}} + o\left(\frac{1}{n}\right) \tag{13.4.10}$$

when μ is unknown, and by

$$\operatorname{var} \sigma^* = \frac{\sigma^2}{n\,e_{22}} + o\left(\frac{1}{n}\right) \tag{13.4.11}$$

when σ is unknown. It is evident how the theorem should be modified in other respects. As is easily seen, the leading terms in these expressions are identical with the lower limit in Fisher's inequality (1.1.1) (cf. also 1.3).

13.5. Further comments upon Type 1 distributions

We shall extend the scope of the theorem in 13.4 by proving that it is also true for nearly unbiased nearly best estimates (cf. 10.7).

We may without loss of generality suppose that the estimates are constructed according to alternative (a) in 10.7. The coefficients in the estimates are then given by (10.3.13) with C_{2i} replaced by C'_{2i}. It is realized that this modification of (10.3.13) does not affect the validity of the relations (13.4.1). Also the statement in the theorem concerning the asymptotic normality of the estimates holds good, provided, however, that we are able to show that the bias of the nearly unbiased estimates tends to zero when n approaches infinity. We shall prove that this is, indeed, the case.

Consider, first, alternative (1) in the theorem. An analysis of the expressions (10.3.13) with C_{2i} replaced by C'_{2i} shows that, for any i,

$$|g_{ri}| < \frac{M}{n}.$$

Thus it follows from the first (or from the second) theorem in 8.7 that the bias is at most of order n^{-1} and, consequently, tends to zero when $n \to \infty$.

Secondly, we turn our attention to the alternatives (2) and (3). We then proceed as before and apply (13.4.7) to the first theorem in 8.9, p. 97, and (13.4.9) to the theorem in 8.11, p. 104. In the former case we infer from (8.9.2) that the error term converges to zero at least as rapidly as $n^{-1} \log n$. In the latter case it follows from (8.11.2) that the error term converges faster than $n^{-\frac{1}{2}}$. Consequently, the estimates are asymptotically unbiased even in these cases, and we have proved the assertion that the theorem holds good for nearly unbiased nearly best estimates.

Examining the proofs of the theorem in 13.4 and the above extension, we infer that all these results are true also for estimates with nearly minimum *MSD* and for modified nearly best estimates (cf. Chapters 11 and 12).

In order to give some illustrations of the theory developed in this and in the preceding section, we shall inspect the examples given in

10.8 and 12.4 and find out which of them contain estimates from Type 1 distributions.

First, suppose that both μ and σ are unknown. Examples of this situation were given in 10.8, Ex. (1) and Ex. (4), and in 12.4, Ex. (1). It follows from the reasoning developed in 13.2 that, of the six distributions considered in these examples, the normal and the extreme-value distributions satisfy Conditions C 1 (with respect to μ and σ) and C 2, and are thus uniform Type 1 distributions. Some comments upon the remaining four distributions will be given in 13.7 and 13.8.

Secondly, assume that μ or σ is unknown. Examples were given in 10.8, Ex. (2) and Ex. (3), and in 12.4, Ex. (2). It is realized that Laplace's distribution (with μ unknown) is a Type 1 case, and also Weibull's distribution (with σ unknown), provided that $m > 1$. The exponential distribution, treated in the first of the three examples, will be discussed in 13.7.

13.6. Type 2 distributions

In the main part of this section, we shall assume that both μ and σ in the cdf. $F[(z-\mu)/\sigma]$ are unknown, and, furthermore, that F belongs to Type 2. We shall study the asymptotic behaviour of the unbiased nearly best estimates μ^* and σ^* derived in 10.3. As we shall see presently, several alternatives arise in this case, corresponding to the different situations outlined in 2.5. Some remarks upon the one-parameter case will be given at the end of the section.

When F is a Type 2 distribution, Condition C 2 given under (a) in 13.2 should be fulfilled. On the other hand, Condition C 1 breaks down at least for one of the parameters. When it is violated with respect to both parameters, F is a uniform Type 2 distribution.

There are two principal situations when Condition C 1 breaks down, viz.

(a) the functions $f(x)$ and/or $xf(x)$ are not zero at the ends A and B of the range of variation of x,

(b) the functions $f(x)$ and/or $xf(x)$ have discontinuity points in the interior of the interval (A, B).

In practice, situation (a) is much more common than (b), and our investigation in this and the next section will be limited to the former alternative. *Throughout* 13.6 *and* 13.7, *we shall assume that the conditions stated in the theorem in* 13.4 *are satisfied, with the above exception.*

It is convenient to use the notations introduced in 2.5. There are in this case at most two discontinuity points

$$w_1 = A, \qquad w_2 = B.$$

The jumps $f(A)$ and $-f(B)$ of $f(x)$ in these points are denoted by

$$\Delta_{1\nu}, \quad (\nu = 1, 2),$$

and the corresponding jumps of $x f(x)$ by

$$\Delta_{2\nu} = w_\nu \Delta_{1\nu}, \quad (\nu = 1, 2).$$

Note that some of these four quantities $\Delta_{r\nu}$ (but not all) may be zero.

The matrix Δ has two rows and two columns, and the product matrix $S = \Delta \Delta'$ has the elements

$$S_{rs} = \Delta_{r1} \Delta_{s1} + \Delta_{r2} \Delta_{s2}, \ (r, s = 1, 2).$$

We see that

$$|S| = (\Delta_{11} \Delta_{22} - \Delta_{12} \Delta_{21})^2.$$

Now consider the behaviour of d_{rs} in (10.3.4) when n approaches infinity. Note that d_{11}, for example, cannot be written in the form (13.4.2). Instead

$$d_{11} = [f(\lambda_1)]^2 + \sum_{i=1}^{n-1} [f(\lambda_i) - f(\lambda_{i+1})]^2 + [f(\lambda_n)]^2. \qquad (13.6.1)$$

We find by analyzing this expression and the related expressions for d_{12} and d_{22} (cf. 2.5 and 13.3)

$$d_{rs} = S_{rs} + \frac{a_{rs}}{n} + \frac{e_{rs}}{n} + o\left(\frac{1}{n}\right), \quad (r, s = 1, 2), \qquad (13.6.2)$$

where

$$a_{rs} = \sum_{\nu = 1, 2} [\Delta_{r\nu} \gamma_s (w_\nu) + \Delta_{s\nu} \gamma_r (w_\nu)].$$

There are three subcases, which will be briefly discussed below.

(1) *Two discontinuity points.*

When $f(x)$ is discontinuous at both ends, F is a uniform Type 2 distribution, for Condition C 1 breaks down both with respect to μ and σ. (Note that one of the quantities $\Delta_{2\nu}$, *but not both*, may be zero.) Moreover, the matrix S is non-singular. Evidently, $G(u)$ is bounded in this case, and only alternative (1) in 13.4, p. 151, is relevant. We have

$$\operatorname{var}\mu^* = \frac{\sigma^2}{n^2}\frac{S_{22}}{|S|} + o\left(\frac{1}{n^2}\right),$$

$$\operatorname{var}\sigma^* = \frac{\sigma^2}{n^2}\frac{S_{11}}{|S|} + o\left(\frac{1}{n^2}\right), \tag{13.6.3}$$

$$\operatorname{cov}(\mu^*, \sigma^*) = -\frac{\sigma^2}{n^2}\frac{S_{12}}{|S|} + o\left(\frac{1}{n^2}\right).$$

The leading terms in these relations can be obtained in two ways, viz. either by using the relation (13.3.1) and the general theory expounded in 2.5, or by combining (13.6.2) with the expressions (10.3.14) and (10.3.15) for the variances and covariance of μ^* and σ^*. The assertions concerning the error terms follow from the general formulae (8.7.6) and (8.7.10), which evidently may be applied to the present situation.

(2) *A single discontinuity point* $\neq 0$.

Suppose, for example, that $f(x)$ is discontinuous at the left end w_1. Since $w_1 \neq 0$, also $x f(x)$ is discontinuous at this point, and hence F is a uniform Type 2 distribution even in this case. The matrix S, however, is now singular, and so the situation is of the type considered in 2.5, (b), p. 29.

It is inferred from (13.6.2) after some reduction that

$$|D| \sim \frac{1}{n}\Delta_{11}^2 J_1,$$

where

$$J_1 = \int_A^B [w_1\gamma_1(x) - \gamma_2(x)]^2 f(x)\,dx = w_1^2 e_{11} - 2w_1 e_{12} + e_{22}.$$

Hence, if $J_1 \neq 0$,

$$\operatorname{var}\mu^* = \frac{\sigma^2}{n}\frac{w_1^2}{J_1} + o\left(\frac{1}{n}\right),$$

$$\operatorname{var}\sigma^* = \frac{\sigma^2}{n}\frac{1}{J_1} + o\left(\frac{1}{n}\right), \tag{13.6.4}$$

$$\operatorname{cov}(\mu^*, \sigma^*) = -\frac{\sigma^2}{n}\frac{w_1}{J_1} + o\left(\frac{1}{n}\right).$$

We shall prove the assertions concerning the error terms in these relations, which may be done in about the same way as in the proof of the theorem in 13.4.

When alternative (1) in 13.4, p. 157, is relevant, we use the general formulae (8.7.6) and (8.7.10). The remaining two alternatives are slightly more complicated. Suppose, for example, that $G(u)$ is an AL-transform at $u=1$. By (13.6.2) we obtain after a simple consideration

$$d_{rs}=O(1), \qquad d^{rs}=O(1), \qquad (r, s=1, 2).$$

We see that the formulae (13.4.6) hold, except for $i=1$. Hence also (13.4.7) is true with the same exception.

Now apply the second theorem in 8.9, p. 102, with the modifications caused by the fact that $G(u)$ is an AL-transform at $u=1$. Conditions (8.9.18) and (8.9.19) hold in this case for $d=1$ and $m=1/k$. Hence we obtain exactly the same result as in the corresponding situation in 13.4, viz.

$$R_{\text{var}}=O(n^{-5/4}).$$

Thus the remainder terms in the two first relations (13.6.4) converge more rapidly than n^{-1}. The assertion concerning the third relation is proved in a similar way.

If the discontinuity point is situated in the right end w_2 of the range of $f(x)$, the formulae (13.6.4) are still true, if only w_1 is replaced by w_2.

(3) *A single discontinuity point equal to zero.*

If $f(x)$ is discontinuous e.g. at the left end w_1 and this point is equal to zero, the situation is of some special interest. Then the function $x f(x)$ is continuous everywhere, and hence F is no longer a uniform Type 2 distribution.

Denoting as before the jump of $f(x)$ in w_1 by Δ_{11}, we find after calculations which are very similar to those performed above

$$\begin{aligned} \text{var}\ \mu^* &= \frac{\sigma^2}{n^2}\frac{1}{\Delta_{11}^2}+o\left(\frac{1}{n^2}\right), \\ \text{var}\ \sigma^* &= \frac{\sigma^2}{n\, e_{22}}+o\left(\frac{1}{n}\right), \\ \text{cov}\ (\mu^*, \sigma^*) &= o\left(\frac{1}{n^{3/2}}\right). \end{aligned} \qquad (13.6.5)$$

Thus we find that var σ^* satisfies the same formula (13.4.11) as when μ is known *a priori* and F is a Type 1 distribution.

The statements concerning the error terms are proved by similar methods as before. A new situation arises, however, when $G(u)$ is an

AL- or an *AP*-transform and the first and third formulae (13.6.5) are considered; we shall briefly discuss this case.

By (13.6.2) we have

$$d_{11} = O(1), \quad d_{12} = O\left(\frac{1}{n}\right), \quad d_{22} = O\left(\frac{1}{n}\right).$$

Consequently,

$$d^{11} = O(1), \quad d^{12} = O(1), \quad d^{22} = O(n),$$

Now suppose, for example, that $G(u)$ is an *AL*-transform, and consider the first formula (13.6.5). We find by similar calculations as in 13.4, p. 153, that the coefficients g_{1i} of μ^* given by (10.3.13) satisfy the inequality

$$|g_{1i}| < \frac{M}{n^2} |\lambda_i^\circ|^{1/k-1},$$

except for $i = 1$. The second theorem in 8.9, p. 102, is satisfied for $d = 3/2$ and $m = 1/k$. Note that the formulation of the theorem should be slightly modified, since $G(u)$ is an *AL*-transform at $u = 1$ in this case. It should be observed that we cannot take $d = 2$, as suggested by the above inequality, for then only (8.9.19) but not (8.9.18) would be satisfied. By (8.9.3), which is a consequence of the theorem, the remainder term in the first relation (13.6.5) converges at least as rapidly as $n^{-9/4}$, and thus faster than n^{-2}, as stated. Other alternatives are treated analogously.

A similar analysis as that performed in (1), (2), and (3) above may also be made for nearly unbiased nearly best estimates (cf. 10.7), all the results obtained in the three subcases being valid even in this case. Moreover, it is possible to prove by exactly the same method as that used in the corresponding analysis of Type 1 distributions in 13.5 that the bias tends to zero when n approaches infinity. To save space, we shall omit the details.

We shall also briefly mention the case of a single unknown parameter. The situation is then very much simpler than before, and we shall only quote the result of the calculations.

(1) μ *unknown*, σ *known*.

When μ is unknown and $f(x)$ is discontinuous at either or both ends, we have

$$\text{var } \mu^* = \frac{\sigma^2}{n^2 S_{11}} + o\left(\frac{1}{n^2}\right). \tag{13.6.6}$$

(2) *μ known, σ unknown.*

When σ is unknown and $xf(x)$ is discontinuous, we have

$$\operatorname{var} \sigma^* = \frac{\sigma^2}{n^2 S_{22}} + o\left(\frac{1}{n^2}\right). \tag{13.6.7}$$

If, on the other hand, $f(x)$ has a single discontinuity at $x=0$, the Type 1 formula (13.4.11) holds true, since $xf(x)$ is then continuous everywhere.

Summing up all the results obtained in the various subcases presented in this section, we may, in view of the general remark in 13.3, make the following statement. *If the italicized conditions on p.* 156 *are satisfied, then nearly best linear estimates from Type 2 distributions are asymptotically efficient (in the general sense considered in Part I).*

We shall finish the section by making two additional remarks.

First, an interesting consequence of the formulae should be pointed out. It is, of course, always necessary to define μ and σ in an unambiguous way, so that the fr.f. of the reduced variable ξ is clearly defined. The importance of this somewhat trivial statement can be aptly illustrated in the case where the cdf. of z has a single discontinuity at one end. Suppose that both μ and σ are unknown, and let us define μ as either any percentage point z_P, $(0 < P < 1)$, of the cdf. or as the discontinuity point. Applying the results obtained above, we infer that, in the former case, the variance of the nearly best estimate μ^* of μ has the order of magnitude n^{-1}, but, in the latter case, the order of magnitude is n^{-2}. Examples illustrating this situation will be given in the following section.

Secondly, we note that nearly best estimates corresponding to Type 2 distributions are generally not asymptotically normally distributed. It is easy to explain why this is so. We saw, in fact, in 8.8 that a linear combination of *TRB*-variables is, essentially, equal to a weighted sum of $n+1$ χ^2-variables, each with two degrees of freedom. To ensure that the sum is asymptotically normal, we introduced the condition (8.8.2), which prevents any component of the sum from dominating the scene. In the present case, (8.8.2) generally breaks down because of the discontinuity points of the fr.f., and the behaviour of the sum is, to a large extent, determined by the components $h_0 y_0$ and $h_n y_n$ corresponding to these points.

13.7. Further comments upon Type 2 distributions

The most common examples of Type 2 distributions are furnished by the rectangular and the exponential distributions (cf. 2.5, 9.2, 10.8, and Sarhan, 1954, p. 320 ff.). Let us suppose that both μ and σ are unknown.

Define, in the case of the rectangular distribution, the parameters as the mean and the range of variation of z. The rectangular fr.f. has two discontinuities. Hence the nearly best estimates satisfy the expressions (13.6.3). It is easily seen that the leading terms in these formulae are asymptotically equivalent to (9.2.19), as they should be.

In the case of the exponential distribution, we define, first, μ and σ as the mean and the standard deviation of z. Then, clearly, the reduced variable has the fr.f. e^{-x-1}. The fr.f. has a single discontinuity for $x = -1$, and thus the expressions (13.6.4) are relevant in this case. If, on the other hand, we define σ as before but μ as the left end of the range of variation of z, the reduced variable has the fr.f. e^{-x}. The discontinuity is then moved to the point $x = 0$, and thus μ^* and σ^* satisfy the relations (13.6.5). It is easy to check that the variances are asymptotically equivalent to those of the best linear estimates (cf. Sarhan, 1954, formulae (5.7) and (5.8)), as they should be.

As a further illustration of the results obtained in 13.6, we consider a truncated normal distribution. The parameters can be defined in various ways in this case. Let us, to fix the ideas, define them as the mean and the standard deviation of the corresponding non-truncated distribution. We must distinguish between three different methods of truncation.

If (a) both tails are truncated, the nearly best estimates satisfy (13.6.3), and thus have variances of order n^{-2}. If (b) the truncation is made at any single point $\neq \mu$, the relations (13.6.4) show that the variances converge as n^{-1}. If (c) the truncation is made in such a way that there remains exactly one half of the normal curve, μ^* has, as seen from (13.6.5), a variance of order n^{-2} but σ^* a variance of order n^{-1}. Note, however, that, if one of the parameters is known *a priori*, both case (a) and case (b) lead to estimates with variances of order n^{-2}; the conclusions concerning case (c) hold good without modification.

We saw in 13.4 and 13.5 that, when no truncation is performed, the nearly best estimates of μ and σ in a normal cdf. have variances which converge as n^{-1}. It follows from the above that the situation becomes entirely different if the tails are cut off even at a long distance from the origin, and the sampling is performed from the truncated distribution. The variances of the nearly best estimates are then of order n^{-2}, and thus converge much more rapidly. It is interesting to study the behaviour of the variances in the last-mentioned case when n is increased, let us say, by factors of two.

When n is small, a doubling of the sample-size results, approximately, in an increase of the precision by a factor $\sqrt{2}$. Gradually, the increase becomes greater, and, in the limit, it becomes equal to 2; thus, when n is large, a doubling of the sample-size results almost in a doubling of the precision. The reasoning developed in this example applies, of course, also to other distributions than the normal.

It has sometimes been argued that the estimates obtained in the rectangular or exponential case behave abnormally, have an unnaturally small variance, and so on. We see now that a wide class of distributions show a similar behaviour. For this reason, estimates of parameters in Type 2 distributions deserve a reasonable part of the huge amount of attention which has hitherto been reserved for what is sometimes called »regular» estimates, i.e. estimates from Type 1 distributions.

The investigation of Type 2 distributions, performed in this and the preceding section, has concerned case (a) mentioned at the beginning of 13.6. Though of no great importance from a practical point of view, it would be of considerable theoretical interest to undertake a similar study of case (b), i.e. of the situation arising when the fr.f. has discontinuities in the interior of its range of variation. It may be believed that such an investigation would yield results which are quite similar to those reported here; it seems highly probable that linear estimates of location and scale parameters may be constructed which generally have variances of order n^{-2}, and which are asymptotically efficient in the general sense considered in Part I. It is even possible that the nearly best linear estimates have these desirable properties also in this case; this is, however, as yet an open question.

13.8. Examples of Type 3 distributions

The discussion in this section has a preliminary character. We shall suppose that $G(u)$ is an AP-transform (cf. 3.5) at one or both end-points with

$$\tfrac{1}{2} \leqslant k < 1.$$

Writing e_{11} in (13.2.1) in the form

$$e_{11} = \int_0^1 \left[\frac{d\,[G'(u)^{-1}]}{d\,u}\right]^2 d\,u,$$

we see by means of (3.5.3) that e_{11} is finite when $k < \frac{1}{2}$, but that it does not exist when k lies in the above interval. Thus we have encountered a case where Condition C 2 breaks down, and a Type 3 situation arises.

We shall examine the leading terms in the expressions (10.3.14) for the variances of the nearly best linear estimates when $G(u)$ fulfills the condition given above. Because of the general remark in 13.3, these terms are asymptotically equivalent to the lower limits which can be attained by any estimate satisfying the general requirements described in the theorems in Part I. Hence, if it can be proved that the errors in (10.3.14), (10.4.2), (10.4.4) converge more rapidly than the leading terms, we know that the nearly best estimates are asymptotically efficient. Unfortunately, the theorem concerning AP-transforms proved in 8.11 is not strong enough for this purpose, and it has as yet not been extended so as to cover the present problem. For this reason, the question of the behaviour of the error terms will be left open, and the truth of the assertions made in the sequel concerning the variances of the nearly best estimates from Type 3 distributions is definitely established, *only after this question has been answered in the positive.* However, the assertions are true even without any complementary proof if μ^* and σ^* are used as symbols for any asymptotically efficient estimates of μ and σ.

Two alternatives will be considered.

(1) *Condition C 1 is satisfied with respect to μ and σ.*

We shall only consider the case where μ is unknown and σ is known *a priori.* Then

$$\operatorname{var}\mu^* = \begin{cases} \dfrac{\sigma^2}{n^{2k}}\,\dfrac{1}{V_k} + o\left(\dfrac{1}{n^{2k}}\right) & \text{for } \frac{1}{2} < k < 1 \\[2ex] \dfrac{\sigma^2}{n\,\log\,n}\,\dfrac{1}{V_k} + o\left(\dfrac{1}{n\,\log\,n}\right) & \text{for } k = \frac{1}{2}, \end{cases} \tag{13.8.1}$$

where

$$V_k = \left(\frac{1}{k} - 1\right)^2 c_0^{-2} \sum_{\nu=1}^{\infty} \frac{1}{\nu^{2k}}$$

for $\frac{1}{2} < k < 1$, and $V_k = c_0^{-2}$ for $k = \frac{1}{2}$.

TABLE 8.

B-transforms. Orders of magnitude of variances of nearly best linear estimates μ^* and σ^* when $0 \leqslant \tau_2 \leqslant \tau_1 \leqslant 1$.

μ or σ unknown		μ and σ unknown	
τ_2	Order	τ_1	Order
0	n^{-2}	0	n^{-2}
$>0, <\frac{1}{2}$	$n^{-2(1-\tau_2)}$	$>0, <\frac{1}{2}$	$n^{-2(1-\tau_1)}$
$\frac{1}{2}$	$(n \log n)^{-1}$	$\frac{1}{2}$	$(n \log n)^{-1}$
$>\frac{1}{2}$	n^{-1}	$>\frac{1}{2}$	n^{-1}

The estimates μ^* and σ^* are based upon a random sample from the cdf. $F[(z-\mu)/\sigma]$, where μ and σ are the mean and the standard deviation of the distribution and $F(x)$ = inverse of B-transform (cf. 3.5). If $\tau_1 < \tau_2$, the exponents should be interchanged.

The leading terms in these expressions are obtained by analyzing the behaviour of d_{11} in (10.4.2) for large n by means of (3.5.3). The result is remarkable, since it is very different from all results found before in the chapter.

The triangular distribution affords an example of this alternative, the second formula (13.8.1) being relevant in this case.

(2) *Condition C 1 is not satisfied with respect to μ and σ.*

Let us assume that the fr.f. has one discontinuity at one of the ends. Two results corresponding to this situation will be briefly stated, and no proofs will be given.

When μ or σ is unknown, the formulae (13.6.6) and (13.6.7) are valid, and hence the variances of μ^* and σ^* converge as n^{-2}.

If both μ and σ are unknown and $\mu \neq$ the discontinuity point, then μ^* and σ^* satisfy relations of the same kind as (13.8.1), though with other values of the factor V_k.

As an illustration, we take the generalized geometric distribution (cf. 3.5) with the exponent satisfying $0 < \tau_1 \leqslant \frac{1}{2}$. (Note that, in the special case $\tau_1 = \frac{1}{2}$, we obtain the right triangular distribution studied empirically in 10.8, Ex. (1).) If μ and σ are defined as the mean and the standard deviation of z, and if both these parameters are unknown, we conclude from what has just been said that μ^* and σ^* have variances of order $n^{-2(1-\tau_1)}$ when $\tau_1 < \frac{1}{2}$, and order $(n \log n)^{-1}$ when $\tau_1 = \frac{1}{2}$. Though outside the scope of this chapter, it is instructive to check by means of a result due to Downton (1954, p. 308 ff.), already mentioned in 10.1, that the latter result holds for the best linear estimates of μ and σ in a right triangular distribution.

13.9. Summary and a general example

Let us survey the results which have been reached in this chapter concerning the order of magnitude of the variances of nearly best linear estimates.[1]

It was proved in 13.4–13.7 that estimates from Type 1 distributions have variances of order n^{-1} and estimates from Type 2 distributions variances of order n^{-2}, except in a special case where the order is n^{-1}. In 13.8 we saw that, in the case of Type 3 distributions, the order of magnitude may equal any power $n^{-\varkappa}$ with $1 \leqslant \varkappa \leqslant 2$. As seen from the example $(n \log n)^{-1}$, there are, however, also other possibilities. It should also be remembered that our discussion in 13.8 was limited to distributions corresponding to AP-transforms, and that other Type 3 distributions may behave differently.

Summing up our findings in a single sentence, we may, somewhat loosely, say that Type 1 estimates converge slowest and Type 2 estimates fastest, with Type 3 estimates (corresponding to AP-transforms) filling the gap between these extremes.

A good illustration of the various alternatives investigated in this chapter is furnished by the class of cdf:s which are inverses of B-transforms (cf. 3.5). The reader is referred to Table 8, where the order of magnitude of the variances of the nearly best estimates is given for various combinations of the exponents τ_1 and τ_2. It is interesting to note that the order of the variances is determined by the smallest of the exponents in the one-parameter case, and by the largest of the exponents in the two-parameter case. The table includes the rectangular and exponential distributions discussed in 13.7 and the generalized geometric distribution mentioned in 13.8.

It might be added that the results reproduced in Table 8 are valid also for many other definitions of μ and σ than those stated in the table, with the following important exceptions: If $\tau_2 = 0$, $0 < \tau_1 < \frac{1}{2}$ and μ is defined as the right end-point of the range of the fr.f. (where the

[1] With respect to AP-transforms, note the remark on p. 164 concerning the convergence of the error terms.

fr.f is discontinuous), then (a) the "two-parameter" column should be used for var σ^* also when μ is known *a priori*, and (b) the "one-parameter" column should be used for var μ^* also when both parameters are unknown. These statements are closely related to those given on pp. 159–161.

CHAPTER 14

CONCLUDING REMARKS

14.1. Introduction

In this final chapter, the information obtained in the preceding chapters concerning the efficiency of nearly best linear estimates is surveyed, and an attempt is made to assess the value of such estimates in comparison with other methods of estimation. In particular, it is claimed that, in their region of application, nearly best estimates compare favourably with maximum likelihood estimates. Some comments are also made concerning the rôle of nearly best estimates in the theory of the testing of hypotheses.

14.2. Criteria for comparison between methods of estimation

Two methods of estimation may be compared from several points of view. In particular, it seems important that the resulting estimates should be efficient. This general term may be given many different meanings; we have, in fact, used the word in three different connections in this thesis, viz. asymptotic efficiency (in the general sense introduced in Part I), efficiency (in the general sense), and efficiency (in the linear sense). These concepts have been defined on pp. 18 and 25, p. 130, and p. 127, respectively. The last term is of interest only when two linear methods of estimation are compared, and has therefore, in the present connection, not quite the same importance as the other two.

Apart from being efficient, the estimates should be easy to calculate. Thus, it seems appropriate to compare two methods of estimation from at least three points of view, namely

(a) Asymptotic efficiency.
(b) Efficiency (in the general sense).
(c) Ease of computation.

We shall briefly compare the nearly best linear estimates with estimates obtained by three other methods in the light of these criteria.

14.3. Comparison with best linear estimates

In Chapter 13, some results have been proved, which all lead up to the same conclusion, viz. that nearly best estimates are asymptotically efficient in the sense introduced in Part I. The same is evidently true also of best linear estimates (cf. 10.1). Thus, by criterion (a), the two methods of estimation are equivalent.

Moreover, we saw in 10.8 that nearly best estimates have often high efficiency (in the linear sense), even when the sample-size is small; i.e. nearly best linear estimates and best linear estimates seem to be practically equivalent from the point of view of criterion (b).

Finally, it was pointed out in Chapter 10 that nearly best estimates are for the most part very much easier to calculate than best linear estimates. Therefore, by criterion (c), the former estimates are generally preferable to the latter.

14.4. Comparison with 'zero-one' linear estimates

In practical problems, it may be desirable, when possible, to estimate location and scale parameters by means of linear combinations

$$\text{const.} \sum_{i=1}^{n} g_i \, z_{(i)}$$

of the ordered sample-values, where the g_i's are equal to 0 or ± 1. It is known (cf., for example, Dixon, 1957) that highly efficient estimates can be constructed in this way when the sample is taken from a normal population.

From a practical point of view, such estimates are very convenient to use and should of course, when available and efficient enough, be used instead of nearly best linear estimates. It may, however, be supposed that no general and simple method of derivation can be found; in any case, 'zero-one' linear estimates have as yet a very limited area of application and, furthermore, have not the same theoretical interest as nearly best estimates.

14.5. Comparison with maximum likelihood estimates

The method of maximum likelihood, introduced by R. A. Fisher, has long been regarded as the most important general method of estimation. It seems to have reached this position for three main reasons put forward by Fisher (1921), namely (1) it rests upon a simple principle, which, at least theoretically, is convenient to apply to any particular problem, (2) the maximum likelihood estimate of a parameter in a Type 1 distribution is under general conditions consistent, asymptotically normal, and asymptotically efficient, and (3) when a sufficient statistic exists, the maximum likelihood estimate is a function of this statistic. Statement (2) has been proved by Cramér (1946 *a*, p. 498 ff.). Statement (3), which was proved by Fisher, may be replaced by the following stronger result, which is a consequence of a remarkable theorem due to Rao (1945), viz. (4) when a sufficient statistic exists, the maximum likelihood estimate is an efficient estimate (in the general sense) of its expected value.

We shall compare nearly best linear estimates of location and scale parameters with maximum likelihood estimates of such parameters. First, when the distribution is of Type 1, it follows from the above and from the theorem in 13.4 that both types of estimate are generally asymptotically efficient. It seems not unlikely that this equivalence of the two methods holds good even in the case of Type 2 and Type 3 distributions; this is, however, as yet an open question, since no general investigation of the asymptotic properties of maximum likelihood estimates from such distributions has been undertaken. Until the contrary has been proved, it may be assumed that, from an asymptotical point of view, there is no essential difference between the two methods of estimation.

Secondly, let us turn to criterion (b) in 14.2 and consider the case of finite samples. Unfortunately, our knowledge concerning the efficiency (in the general sense) of the two types of estimate is very limited, apart from the theoretically interesting but specialized statement (4) above. Therefore, it does not seem possible to settle the question of whether one of the methods is generally superior to the other from this point of view. It might be added that in 10.8, p. 130, we encountered an example indicating that nearly best estimates can be highly efficient (in the general sense). It was mentioned there that the nearly best estimate σ^* of σ in a normal cdf. has an efficiency (in the general sense) of 98.6 % when the sample-size is 5. It is the author's belief that this should not be regarded as only a stroke of luck, but that the normal distribution is only a particular case of a wide class of distributions displaying a similar behaviour. Much research will, however, be required in order to confirm or to disprove this statement.

Finally, from the point of view of criterion (c) in 14.2, nearly best estimates are, with some important exceptions, superior to maximum likelihood estimates. To see this, we may, for example, consider the list of nine distributions given in 3.3, p. 35. Maximum likelihood estimates of location and scale parameters are convenient to determine for the exponential, normal, and rectangular distributions, and for Laplace's distribution. For the other five distributions, these estimates are very difficult to calculate numerically, since an application of the maximum likelihood principle results in quite complicated equations. Similar complexities occur in many other cases. Nearly best estimates, on the other hand, are always determined by the same, rather simple rules (cf. p. 120 ff.). It seems to the author that this is an important argument in favour of the method presented in this thesis.

14.6. Unsolved problems

It might be appropriate to mention that nearly best linear estimates deserve a place not only in the theory of estimation but also in the theory of the testing of hypotheses. Because of their high efficiency, it may be supposed that they can with advantage replace other quantities commonly used in standard tests involving location and scale parameters. Furthermore, it seems possible that the use of these estimates may facilitate the extension of such tests to non-normal situations. For instance, the standard method for testing a location parameter in the presence of an unknown scale parameter can in this way be given a more general formulation. These remarks raise many interesting problems, which are, however, outside the scope of the present investigation.

REFERENCES

BENARD, A. and BOS-LEVENBACH, E. C. (1953), The plotting of observations on probability paper. *Statistica,* **7**, 163.

BHATTACHARYA, A. (1946), On some analogues of the amount of information and their use in statistical estimation. *Sankhyā,* 8, 1, 201, 315.

BLISS, C. I. (1937), The calculation of the time-mortality curve. *Ann. Appl. Biol.*, **24**, 815.

BLOM, G. (1954), Transformations of the binomial, negative binomial, Poisson and χ^2 distributions. *Biometrika,* **41**, 302.

—— (1956), On linear estimates with nearly minimum variance. *Arkiv för Matematik*, Bd **3** nr 31, 365.

CHERNOFF, H. and LIEBERMAN, G. J. (1954), Use of normal probability paper. *Journ. Am. Stat. Ass.*, **49**, 778.

CHU, J. T. and HOTELLING, H. (1955), The moments of the sample median. *Ann. Math. Stat.*, **26**, 593.

CORNISH, E. A. and FISHER, R. A. (1937), Moments and cumulants in the specification of distributions. *Rev. Inst. Int. Statist.*, **4**, 307.

COX, D. R. (1948), A note on the asymptotic distribution of range. *Biometrika,* **35**, 310.

CRAMÉR, H. (1946*a*), *Mathematical Methods of Statistics.* Princeton.

—— (1946*b*), A contribution to the theory of statistical estimation. *Skand. Aktuarietidskrift*, **29**, 85.

DAVID, F. N. and JOHNSON, N. L. (1954), Statistical treatment of censored data I. *Biometrika,* **41**, 228.

—— (1956), Some tests of significance with ordered variables. *Journ. Roy. Stat. Soc. B*, **18**, 1.

DEMING, W. E. (1950), *Some Theory of Sampling.* New York.

DIXON, W. J. (1957), Estimates of the mean and standard deviation of a normal population. *Ann. Math. Stat.*, **28**, 806.

DOWNTON, F. (1954), Least-squares estimates using ordered observations. *Ann. Math. Stat.*, **25**, 303.

ELFVING, G. (1947), The asymptotical distribution of range in samples from a normal population. *Biometrika,* **34**, 111.

FISHER, R. A. (1921), On the mathematical foundations of theoretical statistics. *Phil. Trans. Roy. Soc. A*, **222**, 309.

—— (1925), Theory of statistical estimation. *Proc. Cambr. Phil. Soc.*, **22**, 700.

—— (1935), The logic of inductive inference. *Journ. Roy. Stat. Soc.*, **98**, 39.

FISHER, R. A. and TIPPETT, L. H. C. (1928), Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc. Cambr. Phil. Soc.*, **24**, 180.

FRÉCHET, M. (1927), Sur la loi de probabilité de l'écart maximum. *Ann. Soc. Polon. Math.*, **6**, 92.

GODWIN, H. J. (1949*a*), On the estimation of dispersion by linear systematic statistics. *Biometrika,* **36**, 92.

—— (1949*b*), Some low moments of order statistics. *Ann. Math. Stat.*, **20**, 279.

GOODMAN, L. A. (1953), A simple method for improving some estimators. *Ann. Math. Stat.*, **24**, 114.

GUMBEL, E. J. (1936), Les valeurs extrêmes des distributions statistiques. *Annales de l'Institut Henri Poincaré,* **4**, 115.

—— (1937), Les intervalles extrêmes entre les émissions radio-actives. *Journ. de Physique et le Radium,* **8**, sér. 7, 321.

—— (1944), Ranges and midranges. *Ann. Math. Stat.*, **15**, 415.

—— (1946), On the independence of the extremes in a sample. *Ann. Math. Stat.*, **17**, 78.

—— (1947), The distribution of the range. *Ann. Math. Stat.*, **18**, 384.

—— (1954), Statistical theory of extreme values and some practical applications. Nat. Bur. Stand. Appl. Math. Ser. 33. Washington.

GUPTA, A. K. (1952), Estimation of the mean and standard deviation of a normal population from a censored sample. *Biometrika,* **39**, 260.

HARDY, G. H., LITTLEWOOD, J. E. and PÓLYA, G. (1934), *Inequalities.* Cambridge.

HOEFFDING, W. (1953), On the distribution of the expected values of the order statistics. *Ann. Math. Stat.*, **24**, 93.

IPSEN, J. and JERNE, N. K. (1944), Graphical evaluation of the distribution of small experimental series. *Acta Pathologica,* **21**, 343.

IRWIN, J. O. (1925), The further theory of Francis Galton's individual difference problem. *Biometrika,* **17**, 100.

JUNG, J. (1955), On linear estimates defined by a continuous weight function. *Arkiv för Matematik,* Bd **3** nr 15, 199.

LIEBLEIN, J. (1954), A new method of analyzing extreme-value data. *Nat. Adv. Comm. Aeronat.*, Techn. Note 3053.

LIEBLEIN, J. and ZELEN, M. (1956), Statistical investigation of the fatigue life of deep-groove ball bearings. *Journ. Res. Nat. Bur. Stand.*, **57**, 273.

LLOYD, E. H. (1952), Least-squares estimation of location and scale parameters using order statistics. *Biometrika,* **39**, 88.

MALMQUIST, S. (1950), On a property of order statistics from a rectangular distribution. *Skand. Aktuarietidskrift,* **33**, 214.

MOSTELLER, F. (1946), On some useful 'inefficient' statistics. *Ann. Math. Stat.*, **17**, 377.

NAIR, U. S. (1936), The standard error of Gini's mean difference. *Biometrika,* **28**, 428.

PEARSON, K. (1920), On the probable errors of frequency constants. *Biometrika,* **13**, 113.

PEARSON, K. and PEARSON, M. V. (1931, 1932), On the mean character and variance of a ranked individual, and on the mean and variance of the intervals between ranked individuals. *Biometrika,* **23**, 364; **24**, 203.

RAO, C. R. (1945), Information and accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.*, **37**, 81.

—— (1947), Minimum variance and the estimation of several parameters. *Proc. Cambr. Phil. Soc.*, **43**, 280.

—— (1952), *Advanced Statistical Methods in Biometric Research.* New York.

RÉNYI, A. (1953), On the theory of ordered samples. *Acta Math. Acad. Sci. Hungar.*, **4**, 191.

SARHAN, A. E. (1954, 1955), Estimation of the mean and standard deviation by order statistics. *Ann. Math. Stat.*, **25**, 317; **26**, 505; **26**, 576.

SARHAN, A. E. and GREENBERG, B. G. (1956), Estimation of location and scale parameters by order statistics from singly and doubly censored samples. I. *Ann. Math. Stat.*, **27**, 427.

SCHEFFÉ, H. and TUKEY, J. W. (1945), Non-parametric estimation. I. Validation of order statistics. *Ann. Math. Stat.*, **16**, 187.

SMIRNOFF, N. (1935), Über die Verteilung des allgemeinen Gliedes in der Variationsreihe. *Metron*, **12**, 59.

Table of the first moment of ranked extremes (1951). Nat. Bur. Stand., Report no. 1167 (mimeographed).

TEICHROEW, D. (1956), Tables of expected values of order statistics and products of order statistics for samples of size twenty and less from the normal distribution. *Ann. Math. Stat.*, **27**, 410.

WEIBULL, W. (1939), A statistical theory of the strength of materials. *Ingeniörsvetenskapsakademiens handlingar*, nr 151. Stockholm.

—— (1951), A statistical distribution of wide applicability. *Journ. Appl. Mech.*, **18**, 293.

WILKS, S. S. (1948), Order statistics. *Bull. Am. Math. Soc.*, **54**, 6.

POSTSCRIPT

The original edition of this book was published in Sweden in April, 1958. Listed below are some corrections and complementary statements, which for technical reasons have not been introduced into the text of the present edition.

On p. 15 and p. 21, λ_0 and λ_{n+1} should be defined as $-\infty$ and $+\infty$, respectively.

The definition of an *AL*-transform given on p. 37 is too restrictive. In fact, the exponent -1 in the error term within brackets should be replaced by $-1+\varepsilon$, where ε is a small positive quantity. This change affects the order of several error terms given in Chapter 7 in a rather obvious way, but not any results in the following chapters.

On p. 55, 8th line, $n^{-1+\rho}$ should be replaced by n^{-1}.

On p. 63, the statement in the second sentence from the end of 6.4 is true, provided that the α_i's are chosen such that

$$|\alpha_{i+1}-\alpha_i|<\frac{M}{n}, \quad (i=0, 1, \dots, n).$$

On p. 146, the factor $1/(n-1)$ in the second formula should, of course, be placed under the square root sign.

On p. 159, 5th line, the order of d^{rs} should be $O(n)$.

Professor G. Elfving has kindly informed me that the proof of the theorem in 8.8 is incomplete. The statement on p. 96 that "any moment of Z_1 coincides, in the limit, with the corresponding moment for ζ" requires, in fact, a special proof, which will be published by me elsewhere.

Finally, some additional references will be given. Fil. Lic. G. Kulldorff has pointed out that the papers by Davis, Ann. Math. Stat. Vol. 22, 1951, 43, and Kiefer, Ann. Math. Stat. Vol. 23, 1952, 627, ought to have been mentioned in 1.1 and 13.7. I am indebted to Professor Herbert A. David for bringing to my attention the paper by Hastings, Mosteller, Tukey and Winsor, Ann. Math. Stat. Vol. 18, 1947, 412, which might have been included in the survey of the literature given in 5.1.

It is interesting to note that the results recently published by Plackett in Ann. Math. Stat. Vol. 29, 1958, 131, are closely related to certain of those presented in Part III of this book. The reader is recommended to study Plackett's paper and to compare the two approaches to the problems of linear estimation.

Malmö, August 1958.

Gunnar Blom.