Distribution-Free Statistical Methods

J. S. MARITZ

Distribution-Free Statistical Methods

J.S. MARITZ

Department of Mathematical Statistics
La Trobe University
Australia

CHAPMAN AND HALL

First published 1981 by
Chapman and Hall Ltd
11 New Fetter Lane, London EC4P 4EE

Published in the USA by
Chapman and Hall
in association with Methuen, Inc
733 Third Avenue New York NY 10017

Printed in Great Britain at the
University Press, Cambridge

British Library Cataloguing in Publication Data

Maritz, J S

Distribution-free statistical methods. (Monographs on
applied probability and statistics)
1 Probabilities 2 Mathematical statistics
I Title II Series
519 2 QA278 80-42136

ISBN 0 412-15940-6

Preface

The preparation of several short courses on distribution-free statistical methods for students at third and fourth year level in Australian universities led to the writing of this book My criteria for the courses were, firstly, that the subject should have a clearly recognizable underlying common thread rather than appear to be a collection of isolated techniques Secondly, some discussion of efficiency seemed essential, at a level where the students could appreciate the reasons for the types of calculations that are performed, and be able actually to do some of them. Thirdly, it seemed desirable to emphasize point and interval estimation rather more strongly than is the case in many of the fairly elementary books in this field

Randomization, or permutation, is the fundamental idea that connects almost all of the methods discussed in this book Application of randomization techniques to original observations, or simple transformations of the observations, leads generally to conditionally distribution-free inference Certain transformations, notably 'sign' and 'rank' transformations may lead to unconditionally distribution-free inference An attendant advantage is that useful tabulations of null distributions of test statistics can be produced

In my experience students find the notion of asymptotic relative efficiency of testing difficult Therefore it seemed worthwhile to give a rather informal introduction to the relevant ideas and to concentrate on the Pitman 'efficacy' as a measure of efficiency

Most of the impetus to use distribution-free methods was originally in hypothesis testing. It is now well recognized that adaptation of some of the ideas to point estimation can be advantageous from the points of view of efficiency and robustness Pedagogically there are also advantages in emphasizing estimation One of them is that one can adopt the straightforward approach of defining relative efficiency in terms of variances of estimates Another is that using the notion of an estimating equation makes it easy to relate the distribution-free techniques to methods which will have been encountered in the

standard statistics courses Examples include the method of moments and large sample approximations to standard errors of estimates

The aim of this book is to give an introduction to the distribution-free way of thinking and sufficient detail of some standard techniques to be useful as a practical guide It is not intended as a compendium of distribution-free techniques and some readers may find that a technique which they regard as important is not mentioned For the most part the book deals with problems of location and location shift They include one- and two-sample location problems, and some aspects of regression and of the 'analysis of variance'

Although some of the presentation is somewhat different from what appears to have become the standard in this field, very little, if any, of the material is original Much has been gleaned from various texts Direct acknowledgement of my indebtedness to the authors of these works is made by the listing of general references Through these, and other references, I also acknowledge indirectly the work of other authors whose names may not appear in the bibliography No serious attempt has been made to attribute ideas to their originators Specific references are given only where it is felt that readers may be particularly interested in more detail

While the origins of this book are in undergraduate teaching I do hope that some experienced statisticians will find parts of it interesting In particular, developments in point and interval estimation, and noting of their connections with 'robust' methods have taken place fairly recently Some interesting problems of estimating standard errors as yet not fully resolved are touched upon in several places

Many of my colleagues have helped me, by discussion and by reading sections of manuscript Dr D G Kildea read the first draft of Chapter 2 and his detailed comments led to many improvements Dr B M Brown was not only a patient listener on many occasions but also generously provided Appendix A

Melbourne November 1980 J S Maritz

Contents

CHAPTER 1

Basic concepts in distribution-free methods

1.1 Introduction

In the broadest sense a distribution-free statistical method is one that does not rely for its validity or its utility on any assumptions about the form of distribution that is taken to have generated the sample values on the basis of which inferences about the population distribution are to be made. Obviously a method cannot be useful unless it is valid, but the converse is not true. The terms validity and utility are used in a semi-technical sense and relate to the usual statistical notions of consistency and efficiency respectively. The great attractions of distribution-free methods are:

(i) that they are, by definition, valid under minimal assumptions about underlying distributional forms;
(ii) the aesthetic appeal of their being based for the most part on very simple permutation or randomization ideas;
(iii) the fact that they have very satisfactory efficiency and robustness properties.

Distribution-free methods, especially the simpler ones, have gained widespread acceptance, but they are by no means the first weaponry of most practising statisticians. Perhaps the main impediments to their even greater popularity are:

(a) the results of distribution-free tests are often not as readily interpretable in terms of physical quantities as are the results of parametric analysis;
(b) in some of the more complex situations severe computational difficulties can arise; although many distribution-free methods are 'quick' and 'easy' they do not all share these properties.

It should also be noted, of course, that many distribution-free methods are relatively new; this applies particularly to the estimation methods. Therefore they are not yet well known in the popular sense.

This book is written for undergraduate students, as an instruction manual in the use of some of the standard distribution-free methods, but its main objective is to serve as an introduction to the underlying ideas, and perhaps to stimulate further reading in the subject. Consequently the emphasis is on randomization (or permutation) as the underlying unifying notion in the development of testing methods, and associated methods of estimation. These ideas are certainly not new, and have been developed in great detail in various special contexts. Nevertheless, the use of 'signs' and 'ranks' is still commonly thought to characterize distribution-free methods, if not by statisticians, then by very many non-professional users of statistical methods.

The selection of topics that are treated in ensuing chapters is influenced strongly by consideration (a) above. In fact the emphasis is heavily on questions of location and location shift. They are not only among the most important from the practical viewpoint, but also represent a class of problems where it is clearly easy, and sensible, to visualize the quantities that are subject to inference, without the need to specify underlying distributions in close detail. This is the only excuse offered for not including many 'standard' procedures, such as runs tests, some of the tests of dispersion, general tests of distribution functions, such as the Kolmogorov–Smirnov test.

Very few of the so-called distribution-free methods are truly distribution-free. Many of the arguments are simplified if the underlying distribution can be taken as continuous, and this is commonly done. This assumption will be made throughout this book. Other assumptions are necessary, depending on the problem. For example, in one-sample location problems the assumption of symmetry plays a major role. Thus the term distribution-free must be interpreted with some qualification. The methods are developed without detailed parametric specification of distributions; we may assume that a density $f(x)$ is symmetric about θ, but need not say that it is, for example, $1/[\pi\{1+(x-\theta)^2\}]$. The term 'nonparametric' is preferred to 'distribution-free' by some, but since we are actually trying to make inferences about parameters the latter term seems more appropriate here.

1.2 Randomization and exact tests

Although the randomization basis of test and other methods will be restated for specific procedures in later chapters, we shall illustrate it

here in a simple example. This will enable us to define certain terms rather conveniently.

Consider the well-known simple 'paired comparison' experiment in which two treatments are allocated at random, one each to a pair of subjects. The two subjects of the same pairs are chosen to be as alike as possible. For example, in an experiment on sheep a natural pairing would be to use twins for each pair. Suppose that the results obtained for the ith pair are measurements y_{iA} and y_{iB} for the members receiving treatments A and B respectively. The differential effect of the treatments for the ith pair could now be measured by $d_i = y_{iA} - y_{iB}$, with $i = 1, 2, \ldots, n$.

Now suppose that we are to test the null hypothesis H_0, according to which the effects of treatments A and B are identical. Since our allocation of treatments within pairs is random, correctness of H_0 would mean that d_i could have had the value $-d_i = y_{iB} - y_{iA}$. Further, if we denote by D_i the random variable being the ith difference obtained in the experiment, then

$$\Pr(D_i = +|d_i|) = \Pr(D_i = -|d_i|) = \tfrac{1}{2} \tag{1.1}$$

The probabilities in (1.1) are conditional probabilities, the conditioning being on the ith pair whose difference has magnitude $|d_i|$.

Since the randomization is performed independently for each pair, it is now a simple matter to conceive the joint distribution of $D_1, D_2, \ldots, D_n$, for the random variables $D_1, D_2, \ldots, D_n$ are independent with individual distributions given by (1.1). Again we note that it is a conditional distribution. A natural test statistic for H_0 is $T = D_1 + D_2 + \ldots + D_n$, and from the preceding discussion it is clear that tabulation of the exact conditional distribution of T is a straightforward matter; the 2^n possible sign combinations to be attached to the magnitudes $|d_i|$, $i = 1, 2, \ldots, n$, have to be listed, and for each of these the value of T computed. This generates 2^n possible values of T occurring with equal probabilities 2^{-n}, and thus establishes the exact conditional distribution of T. Let τ be such that $\Pr(T \geqslant \tau) = r/2^n$. If we test H against a one-sided alternative and take as critical region all $T_i \geqslant \tau$, then the size of this critical region (the level of significance) is exactly $r/2^n$.

A test is said to be *exact* if the actual significance level is exactly that which is nominated. In our example, the test is an exact level $r/2^n$ test. Moreover, it is important to note that although the exact significance level derives from the exact conditional distribution, the unconditional significance level is also exactly $r/2^n$. This is true simply

because, for every possible set $|d_i|$, $i = 1, 2, \ldots n$, the probability of rejecting H_0 is $r/2^n$.

The distribution of T under H_0 will often be referred to as the null distribution of T. In our example it is a conditional null distribution. The test of H_0 is carried out by referring the observed value of T to its conditional null distribution. The derivation of the distribution in our example was achieved by using the randomization argument, and did not depend on an assumption of the form of distribution that individual y_i values might follow. So, the conditional null distribution of T does not depend on an underlying distributional assumption, and consequently the significance level is exact, free from such an assumption.

Although the exactness of the test is not affected by its being based on a conditional null distribution, it should be kept in mind that the conditional distribution of T will change from sample to sample. Therefore the unconditional distribution of T will be a mixture of the conditional distributions and its form will depend on the underlying distribution of $|d|$ values. When the exact conditional distribution of a statistic, obtained by a randomization procedure, is not invariant with respect to the realized sample values, the associated distribution-free methods are said to be *conditionally distribution-free*.

By transformations such as rank and sign transformations it is often possible to derive methods that are unconditionally distribution-free from those that are conditionally distribution-free. If every $|d_i|$ in the example that we have been discussing is replaced by 1 we obtain the well known 'sign test' and it is clear that the conditional distribution of $S = \sum_{i=1}^{n} \text{sgn}(D_i)$, remains exactly the same for every possible set of realized results.

From the point of view of exactness of significance levels, there is no obvious advantage in a test being unconditionally distribution-free. Since the distributions of test statistics can be tabulated once and for all if they are unconditionally distribution-free, there can be worthwhile computational advantages in such tests. We shall see, also, that there can be gains in efficiency by astute choice of transformation. However, our starting point is randomization and its natural consequence is to produce, in the first instance, conditionally distribution-free methods.

Enumeration of exact null distributions can be a totally impractical task for large sample sizes, hence it is quite common to approximate null distributions by some standard distribution, usually a normal distribution, and so to obtain approximate values of significance

levels Here the approximation is a mathematical convenience and does not affect the exactness of the method in principle Hence such procedures will be called 'exact' whether or not some mathematical approximation is used for convenience However, there are circumstances where approximations, of essentially unknown precision, have to be made They are usually occasioned by nuisance parameters, whose values, while not of direct concern do affect the null distributions of interest

1.3 Consistency of tests

If the null hypothesis in the example of Section 1 2 had specified a difference θ between the two treatments, the d_i values would have been replaced by $d_i - \theta$, $i = 1, 2, \quad, n$, the argument otherwise remaining the same In order to show the dependence of T on $D_1, D_2, \quad, D_n$ and θ we may write it $I(\mathbf{D}, \theta)$ Taking differences of the paired values can be regarded as reducing the problem to a one-sample problem, and in the remainder of this section we shall be discussing one-sample, one-parameter problems

Suppose that random sampling from a population with parameter of interest θ produces the results $X_1\ X_2, \quad, X_n$, where $X_1, X_2, \quad, X_n$ can be taken as independent and identically distributed Let the statistic to be used in testing a hypothesis about θ be $S(\mathbf{X}, t)$, defined such that its conditional and its unconditional null distributions have mean 0 if t is replaced by θ

Suppose that we propose to test $H_0\ \ \theta = \theta_0$ against $H_1\ \ \theta = \theta_1 > 0$ at level α and that the test procedure is to reject H_0 if observed $S(\mathbf{X}, \theta_0) > C_\alpha(\mathbf{X}, \theta_0)$ The value of $C_\alpha(\mathbf{X}, \theta_0)$ is determined from the randomization distribution of $S(\mathbf{X}, \theta_0)$, therefore it generally depends on $\mathbf{X}$ and on θ_0 We shall assume that $S(\mathbf{X}, \theta_0)$ is so scaled with respect to n that $C_\alpha(\mathbf{X}, \theta_0) \xrightarrow{P} 0$ i e , in probability, as $n \to \infty$

Questions of consistency have to be answered in terms of the behaviour of the unconditional distributions of the relevant statistics and we shall assume that in the unconditional distribution of $S(\mathbf{X}, \theta_0)$, when $\theta = \theta_1$,

$$E\{S(\mathbf{X}, \theta_0)|\theta_1)\} = \Delta(\theta_1, \theta_0) > 0$$
$$\text{var}\{S(\mathbf{X}, \theta_0)|\theta_1)\} = \sigma^2(\theta_1, \theta_0)\ n, \text{ with } \sigma^2(\theta_1\ \ \theta_0) \text{ bounded}$$

We shall say that the test of H_0 against H_1 is *consistent* if its power can be made arbitrarily close to 1 by increasing n

Lemma 1 1 Under the assumptions given above about the distribution of $S(\mathbf{X}, t)$, the test of H_0 against H_1 is consistent

The proof of this lemma is obtained simply from the assumptions by noting that $S(\mathbf{X}, \theta_0) \xrightarrow{P} \Delta(\theta_1, \theta_0) > 0$, while $C_\alpha(\mathbf{X}, \theta_0) \xrightarrow{P} 0$

We may be concerned with values of θ_1 close to θ_0 in which case it is useful if we assume

$$E\{S(\mathbf{X}, t)|\theta\} = \mu(t, \theta)$$

with $\mu(t, \theta)$ continuous in t and differentiable near θ
Then we can put

$$\Delta(\theta_1, \theta_0) \simeq (\theta_0 - \theta_1)\mu(\theta_1, \theta_1),$$

noting that $\mu(\theta_0, \theta_0) = 0$

For Lemma 1 1 to hold we now have to replace the earlier assumption about $E\{S(\mathbf{X}, \theta_0)|\theta_1\}$ by the assumption $\mu'(\theta, \theta) < 0$

Let us reconsider the example of Section 1 2 in two versions

(i) $S(\mathbf{X}, t) = T^*(\mathbf{D}, t) = T(\mathbf{D}, t)/n = (D_1 + \quad + D_n)/n - t$, suppose, for our illustration, that the distribution of every D_i is normal with mean θ and variance σ^2, and that $\theta_0 = 0$ In this case, as we shall see, the conditional randomization distribution of S can be taken as approximately normal for large n, with variance $\sum d_i^2/n^2$, if $d_1, d_2, \quad , d_n$ are the observed differences So, with u_α an appropriate normal quantile,

$$C_\alpha(\mathbf{X}, \theta_0) = u_\alpha\{\sum D_i^2/n\}^{1/2}/\sqrt{n},$$

and it is easy to see that $\sum D_i^2/n \xrightarrow{P} \sigma^2 + \theta_1^2$ under H_1 so that $C_\alpha(\mathbf{X}, \theta_0) \xrightarrow{P} 0$

Further,

$$E\{S(\mathbf{X}, \theta_0)|\theta_1\} = \theta_1$$
$$\text{var}\ \{S(\mathbf{X}_1, \theta_0)|\theta_1\} = \sigma^2/n$$

note also that $E\{S(\mathbf{X}, t)|\theta\} = \theta - t$, $\mu'(\theta, \theta) = -1$ The test is consistent, in fact, as we shall show in Chapter 2 the test statistic can be written as a function of the usual t-statistic whose consistency and other properties are well known

It should be noted, however, that with certain distributions for the D_i the test may not be consistent

(ii) Let $S(\mathbf{X}, t) = (1/n)\sum_{i=1}^{n} \text{sgn}(D_i - t)$ and suppose that the distribution function of every D_i is $F(d, \theta)$, with density $f(d, \theta)$, symmetric

about θ The null distribution of S has variance $1/n$, n even, and S, being a linear function of a binomial random variable, has an approximately normal distribution
So,

$$C_\alpha(\mathbf{X}, \theta_0) = u_\alpha/\sqrt{n}$$

Also

$$\begin{aligned} E\{S(\mathbf{X}, t)|\theta\} &= (1/n) \sum E\{\operatorname{sgn}(D_i - t)|\theta\} \\ &= (1/n) \sum \{(-1)\ P(D_i < t|\theta) + (1)\ P(D_i > t|\theta)\} \\ &= 1 - 2F(t, \theta) \end{aligned}$$

Therefore $\mu'(\theta, \theta) = -2f(\theta, \theta)$ Under H_1 the statistic S can again be expressed as a binomial random variable, hence $\sigma^2(\theta_1, \theta_0) \leqslant \frac{1}{4}$, so that the conditions of Lemma 1 1 are satisfied

Two points about the examples above are worth remarking upon First we see that the quantity $C_\alpha(\mathbf{X}, \theta_0)$ is simply a constant, that is non-random, in case (ii) Second, the only restriction on F is that $f(\theta, \theta) \neq 0$, so that the 'sign' test will be consistent in many cases where the usual t-test is not

1.4 Point estimation of a single parameter

Since the statistic $S(\mathbf{X}, t)$ is defined so that $E\{S(\mathbf{X}, \theta)\} = 0$, a natural procedure for finding a point estimate of θ is suggested by the method of moments, namely, to take as point estimate $\hat{\theta}$ of θ the solution $t = \hat{\theta}$ of the estimating equation

$$S(\mathbf{X}, t) = 0 \tag{1 2}$$

As we shall see in later chapters, some of the statistics $S(\mathbf{X}, t)$, regarded as functions of t for fixed $\mathbf{X}$ are not continuous in t so that a unique solution of (1 2) has to be decided upon by a suitable convention An appropriate one is usually obvious in context Weak consistency of $\hat{\theta}$ is easily checked, in fact we have

Lemma 1 2 If $E\{S(\mathbf{X}, t)|\theta\} = \mu(t, \theta)$ is continuous in t and differentiable near $t > \theta$, and if the other conditions for the applicability of Lemma 1 1 hold, then $\hat{\theta} \xrightarrow{P} \theta$ as $n \to \infty$.

1.5 Confidence limits

Provision of a measure of precision of an estimate is an essential part of statistical inference, and one way of doing this is to give confidence

limits Confidence limits, or confidence sets, can be determined by the well-known procedure of taking a point θ in the parameter space to belong to the confidence set if the null hypothesis that $\theta = \theta'$ is accepted Briefly, the argument is as follows

To test H_0 $\theta = \theta_0$ against some alternative H_1, a critical region of size α, $R(\theta_0)$, is determined such that

$$Pr\{S(\mathbf{X}, \theta_0) \in R(\theta_0) | H_0\} = 1 - \alpha \tag{1 3}$$

Now, for a given $\mathbf{X} = \mathbf{x}$ find the set $C_\theta(\mathbf{x})$ of all θ such that $S(\mathbf{x}, \theta) \in R(\theta)$ The true value θ_0 will belong to $C_\theta(\mathbf{x})$ if $S(\mathbf{x}, \theta_0) \in R(\theta_0)$ But the probability of this event is $1 - \alpha$, whatever the value of θ_0, according to our definition (1 3) The set $C_\theta(\mathbf{x})$ is a $100(1 - \alpha)\%$ confidence set for θ The shape of $C_\theta(\mathbf{x})$ is determined by the shape of $R(\theta)$ In the one-parameter case that we are considering, $R(\theta)$ is typically an interval, and so is $C_\theta(\mathbf{x})$, one- or two-sided confidence limits are obtained according to whether the test is one- or two-sided The probability $1 - \alpha$ is sometimes called the *confidence coefficient*

A notable feature of the procedure outlined above is that only the null distribution of S is needed In distribution-free methods this is particularly useful because the null distributions are usually exact, often very easy to obtain, and of course in many instances already tabulated Moreover, whether a conditional or unconditional null distribution is used the confidence coefficient is the value $1 - \alpha$ associated with the hypothesis test, and if the probability $1 - \alpha$ is exact, then so is the confidence coefficient We shall say that a confidence region is exact if the confidence coefficient $1 - \alpha$ is exact One of the great attractions of distribution-free methods is that they enable one to determine, often fairly easily, exact confidence limits for certain parameters with minimal assumptions about distributional forms

1.6 Efficiency considerations in the one-parameter case

1 6 1 *Estimation*

Efficiency of estimation will be measured in terms of $\mathrm{var}(\hat{\theta})$ The relative efficiency of two estimators will be measured by the ratio of their variances In some cases it will be possible to express $\hat{\theta}$ fairly simply in terms of $X_1, X_2, \ldots, X_n$ so that an exact expression for $\mathrm{var}(\hat{\theta})$ may be given However, for the most part we shall have to deal with cases where such a simple expression cannot be obtained, in fact we

may not even be able to express $\hat{\theta}$ explicitly in terms of $X_1, X_2, \ldots, X_n$. Then the best we can do is obtain a large-sample approximation formula for $\text{var}(\hat{\theta})$.

To simplify notation we shall assume in what follows that expectation and variance are derived at the true value of the parameter θ. Thus we put $E\{S(\mathbf{X}, t)|\theta\} = E\{S(\mathbf{X}, t)\}$, etc. The main assumptions that we shall make are

(i) $E\{S(\mathbf{X}, t)\} = \mu(t, \theta)$, is continuous in t and differentiable near $t = \theta$.

(ii) the statistic $S(\mathbf{X}, t)$, treated as a function of t for fixed $\mathbf{X}$, either is continuous and differentiable for t near θ, or it can be replaced by an approximating function which has these properties,

(iii) $\text{var}\{S(\mathbf{X}, t)\} = \sigma^2(t, \theta)/n$, $\sigma^2(t, \theta)$ continuous in t and bounded.

Assumption (ii) is needed because statistics $S(\mathbf{X}, t)$ obtained after rank or sign transformations are typically discontinuous step functions of t. However, it is also typically true of them that if they are scaled such that $E\{S(\mathbf{X}, t)\} = \mu(t, \theta)$, that is, not dependent on n, then the number of steps increase with n, and their heights decrease. The sign statistic of example (ii) in Section 1.3 is a case in point, a simple transformation of $S(\mathbf{X}, t)$ is the sample distribution function which is known to have the desired property.

Now consider a small but finite neighbourhood of θ, the interval $(\theta - h/2, \theta + h/2)$ with h held constant, so that we can put

$$\left\{\frac{\partial S(\mathbf{X}, t)}{\partial t}\right\}_{t=\theta} \simeq \frac{S(\mathbf{X}, \theta + h/2) - S(\mathbf{X}, \theta - h/2)}{h}$$

$$= \frac{S(\theta + h/2, \theta) - S(\theta - h/2, \theta) + O(1/\sqrt{n})}{h}$$

$$\simeq \mu(\theta, \theta) + O(1/\sqrt{n})$$

in view of assumption (iii).

Write

$$S(\mathbf{X}, t) \simeq S(\mathbf{X}, \theta) + (t - \theta)\{\mu'(\theta, \theta) + O(1/\sqrt{n})\} \tag{1.4}$$

and note that $S(\mathbf{X}, \hat{\theta}) = 0$. Then we have approximately

$$\text{var}(\hat{\theta}) = \text{var}\{S(\mathbf{X}, \theta)\}\,[\partial E\{S(\mathbf{X}, t)\}/\partial t]^{-2}_{t=\theta} \tag{1.5}$$

Formula (1.5), or approximate methods similar to those used in its derivation, occur in sundry standard situations. For example if it is applied to the estimating equation in the case of 'regular' maximum-

likelihood estimation, the usual large-sample variance formula for maximum-likelihood estimators is obtained

A simple example is the following suppose that $F_n(x)$ is the usual sample distribution function based on n independent observations from a population with continuous distribution function $F(x)$, density $f(x)$ and median θ The estimating equation for θ is

$$S(\mathbf{X}, t) = F_n(t) - 1/2 = 0$$
$$E\{S(X, t)\} = F(t) - 1/2$$
$$[\partial E\{S(\mathbf{X}, t)\}/\partial t]_{t=\theta} = r(\theta)$$
$$\text{var}\{S(\mathbf{X}, \theta)\} = 1/4n$$

giving the approximate formula for the variance of the sample median, $\hat{\theta}$,

$$\text{var}(\hat{\theta}) \simeq 1/(4nf^2(\theta))$$

Another consequence of the 'linearization' represented by (1 3) is that the distribution of $\hat{\theta}$ will be approximately normal if the distribution of $S(\mathbf{X}, \theta)$ is approximately normal Linearization ideas have become important in distribution-free methods, and a more rigorous approach is outlined in Appendix A

1 6 2 *Hypothesis testing*

Consider two statistics S_1 and S_2 satisfying the conditions (i), (ii), (iii) given in Section 1 6 1 Suppose that we scale these two statistics so that their null distributions have the same dispersion at a certain value θ_0 of θ, that is, we replace S_1 by $S_1(\mathbf{X}, t)/\sigma_1(\theta_0, \theta_0) = S_1^*(X, t)$ and $S_2(X, t)$ by $S_2(X, t)/\sigma_2(\theta_0, \theta_0) = S_2^*(\mathbf{X}, t)$ Inspection of formula (1 4) shows that the ratio of variances of estimates of θ_0 based on S_1^* and S_2^* is determined by the slopes $[\partial E\{S_r^*(\mathbf{X}, t)\}/\partial t]_{t=\theta_0}$, $r = 1, 2$

Without doing formal power calculations, it is clear that the power of a test of H_0 $\theta = \theta_0$ against H_1 $\theta = \theta_0 + \Delta$ Δ small, will be determined largely by the slope given above associated with the statistic S_r^* used for the test For any statistic S, the slope at θ_0

$$e_S(\theta_0) = |[\partial E\{S^*(\mathbf{X}, t)\}/\partial t]_{t=\theta} = \mu(\theta_0, \theta_0)\ \sigma(\theta_0, \theta_0) \qquad (1\ 6)$$

is, therefore a natural measure of its efficiency for testing H_0 against a close alternative H_1 If we put $\Delta = 1/\sqrt{n}$ then $e_S(\theta_0)$ is the displacement of the distribution of S from its null location under H_1, standardized with respect to its standard deviation under H_0

Making the approximation $\sigma(\theta_0 + \Delta, \theta_0) \simeq \sigma(\theta_0, \theta_0)$ and assuming