

# Chapter 13

## Applications to Inference



### 13.1 Introduction

In this chapter, we apply some basic asymptotic concepts and results to gain insight into the large-sample behavior of some fundamental tests. Section 13.2 considers the robustness of some classical tests, like the  $t$ -test, when the underlying assumptions may not hold. For example, for testing the null hypothesis that the underlying mean is zero, it is seen that the  $t$ -test is pointwise asymptotically level  $\alpha$  for any distribution  $F$  with finite nonzero variance. However, such a robustness of validity result does not always extend to other parametric procedures, as will be seen. Although the probability of a Type 1 error tends to the nominal level as  $n \rightarrow \infty$ , it is also important to investigate the speed of convergence. For this reason, Edgeworth expansions are discussed in Section 13.3. Further results are developed for testing a univariate mean in a nonparametric setting in Section 13.4. The question of whether or not the  $t$ -test is uniformly asymptotically level  $\alpha$  is investigated, where uniformity refers to some broad nonparametric family. In order to obtain uniformity, some restrictions are needed for any method, as demonstrated by a result of Bahadur and Savage. Section 13.5 serves as an introduction to testing many means in a high-dimensional setting.

### 13.2 Robustness of Some Classical Tests

Optimality theory postulates a statistical model and then attempts to determine a best procedure for that model. Since model assumptions tend to be unreliable, it is necessary to go a step further and ask how sensitive the procedure and its optimality are to the assumptions. In the normal models of Chapters 4–7, three assumptions are made: independence, identity of distribution, and normality. In the two-sample  $t$ -test, there is the additional assumption of equality of variance. We shall consider

the effects of nonnormality and inequality of variance in the first subsection, and that of dependence in the next subsection.

The natural first question to ask about the robustness of a test concerns the behavior of the significance level. If an assumption is violated, is the significance level still approximately valid? Such questions concerning robustness of validity are typically answered by combining two methods of attack. The actual significance level under some alternative distribution is either calculated exactly or, more usually, estimated by simulation. In addition, asymptotic results are obtained which provide approximations to the true significance level for a wide variety of models. We here restrict ourselves to a brief sketch of the latter approach.

### 13.2.1 Effect of Distribution

Consider the one-sample problem where  $X_1, \dots, X_n$  are independently distributed as  $N(\xi, \sigma^2)$ . Tests of  $H : \xi = \xi_0$  are based on the test statistic

$$t_n = t_n(X_1, \dots, X_n) = \frac{\sqrt{n}(\bar{X}_n - \xi_0)}{S_n} = \frac{\sqrt{n}(\bar{X}_n - \xi_0)}{\sigma} \bigg/ \frac{S_n}{\sigma}, \quad (13.1)$$

where

$$S_n^2 = \sum (X_i - \bar{X}_n)^2 / (n - 1);$$

see Section 5.2. When  $\xi = \xi_0$  and the  $X$ 's are normal,  $t_n$  has the  $t$ -distribution with  $n - 1$  degrees of freedom. Suppose, however, that the normality assumption fails and the  $X$ 's instead are distributed according to some other distribution  $F$  with mean  $\xi_0$  and finite variance. Then by the Central Limit Theorem,  $\sqrt{n}(\bar{X}_n - \xi_0)/\sigma$  has the limit distribution  $N(0, 1)$ ; furthermore  $S_n/\sigma$  tends to 1 in probability by Example 11.3.3. Therefore, by Slutsky's Theorem,  $t_n$  has the limit distribution  $N(0, 1)$  regardless of  $F$ . This shows in particular that the  $t$ -distribution with  $n - 1$  degrees of freedom tends to  $N(0, 1)$  as  $n \rightarrow \infty$ .

To be specific, consider the one-sided  $t$ -test which rejects when  $t_n \geq t_{n-1, 1-\alpha}$ , where  $t_{n-1, 1-\alpha}$  is the  $1 - \alpha$  quantile of the  $t$ -distribution with  $n - 1$  degrees of freedom. It follows from Corollary 11.3.1 and the asymptotic normality of the  $t$ -distribution that (see Problem 11.47 (ii))

$$t_{n-1, 1-\alpha} \rightarrow z_{1-\alpha} = \Phi^{-1}(1 - \alpha).$$

In fact, the difference  $t_{n-1, 1-\alpha} - z_{1-\alpha}$  is  $O(n^{-1})$ , as will be seen in Section 13.3.

Let  $\alpha_n(F)$  be the true probability of the rejection region  $t_n \geq t_{n-1, 1-\alpha}$  when the distribution of the  $X$ 's is  $F$ . Then

$$\alpha_n(F) = P_F\{t_n \geq t_{n-1, 1-\alpha}\}$$

has the same limit as  $P_\Phi\{t_n \geq z_{1-\alpha}\}$ , which is  $\alpha$ . Thus, the  $t$ -test is pointwise asymptotically level  $\alpha$ , assuming the underlying distribution has a finite nonzero variance. However, the  $t$ -test is not uniformly asymptotically level  $\alpha$ . This issue will be studied more closely in Section 13.4. For sufficiently large  $n$ , the actual rejection probability  $\alpha_n(F)$  will be close to the nominal level  $\alpha$ ; how close depends on  $F$  and  $n$ . For entries to the literature dealing with this dependence, see Cressie (1980), Tan (1982), Benjamini (1983), and Edelman (1990). Other robust approaches for testing the mean are discussed in Sutton (1993) and Chen (1995). The use of permutation and resampling methods will be deferred to Chapters 17 and 18.

To study the corresponding test of variance, suppose first that the mean  $\xi$  is 0. When  $F$  is normal, the UMP test of  $H : \sigma = \sigma_0$  against  $\sigma > \sigma_0$  rejects when  $\sum X_i^2/\sigma_0^2$  is too large, where the null distribution of  $\sum X_i^2/\sigma_0^2$  is  $\chi_n^2$ . By the Central Limit Theorem,

$$\frac{1}{\sqrt{n}}(\sum X_i^2 - n\sigma_0^2) \xrightarrow{d} N(0, 2\sigma_0^4)$$

as  $n \rightarrow \infty$ , since  $\text{Var}(X_i^2) = 2\sigma_0^4$ . If the rejection region is written as

$$\frac{\sum X_i^2 - n\sigma_0^2}{\sqrt{2n\sigma_0^2}} \geq C_n,$$

it follows that  $C_n \rightarrow z_{1-\alpha}$ .

Suppose now instead that the  $X$ 's are distributed according to a distribution  $F$  with  $E(X_i) = 0$ ,  $E(X_i^2) = \text{Var}(X_i) = \sigma^2$ , and  $\text{Var}(X_i^2) = \gamma^2$ . Then,

$$\sum (X_i^2 - \sigma^2)/\sqrt{n} \xrightarrow{d} N(0, \gamma^2)$$

when  $\sigma = \sigma_0$ , and the rejection probability  $\alpha_n(F)$  of the test tends to

$$\lim P \left\{ \frac{\sum X_i^2 - n\sigma_0^2}{\sqrt{2n\sigma_0^2}} \geq z_{1-\alpha} \right\} = 1 - \Phi \left( \frac{z_{1-\alpha}\sqrt{2\sigma_0^2}}{\gamma} \right).$$

Depending on  $\gamma$ , which can take on any positive value, the sequence  $\alpha_n(F)$  can thus tend to any limit  $< \frac{1}{2}$ . Even asymptotically and under rather small departures from normality (if they lead to big changes in  $\gamma$ ), the size of the  $\chi^2$ -test is thus completely uncontrolled.

For sufficiently large  $n$ , the difficulty can be overcome by Studentization,<sup>1</sup> where one divides the test statistic by a consistent estimate of the asymptotic standard deviation. Letting  $Y_i = X_i^2$  and  $E(Y_i) = \eta = \sigma^2$ , the test statistic then reduces to  $\sqrt{n}(\bar{Y} - \eta_0)$ . To obtain an asymptotically valid test, it is only necessary to divide by a suitable estimator of  $\sqrt{\text{Var}Y_i}$  such as  $\sqrt{\sum (Y_i - \bar{Y})^2/n}$ . (However, since  $Y_i^2 = X_i^4$ ,

---

<sup>1</sup> Studentization is defined in a more general context at the end of Section 7.3.

small changes in the tail of  $X_i$  may have large effects on  $Y_i^2$ , and  $n$  may have to be rather large for the asymptotic result to give a good approximation.)

When  $\xi$  is unknown, the normal theory test for  $\sigma^2$  is based on  $\sum(X_i - \bar{X}_n)^2$ , and the sequence

$$\frac{1}{\sqrt{n}} \left[ \sum (X_i - \bar{X}_n)^2 - n\sigma_0^2 \right] = \frac{1}{\sqrt{n}} \left( \sum X_i^2 - n\sigma_0^2 \right) - \frac{1}{\sqrt{n}} n\bar{X}^2$$

again has the limit distribution  $N(0, \gamma^2)$ . To see this, note that the distribution of  $\sum(X_i - \bar{X}_n)^2$  is independent of  $\xi$  and put  $\xi = 0$ . Since  $\sqrt{n}\bar{X}$  has a (normal) limit distribution,  $n\bar{X}^2$  is bounded in probability and so  $n\bar{X}^2/\sqrt{n}$  tends to zero in probability. The result now follows from that for  $\xi = 0$  and Slutsky's Theorem.

The above results carry over to the corresponding two-sample problems that were considered in Section 5.3. Consider the two-sample  $t$ -statistic given by (5.28). An extension of the one-sample argument shows that as  $m, n \rightarrow \infty$ ,

$$\frac{\bar{Y}_n - \bar{X}_m}{\sigma\sqrt{1/m + 1/n}} \xrightarrow{d} N(0, 1)$$

while

$$\frac{\sum(X_i - \bar{X}_m)^2 + \sum(Y_j - \bar{Y}_n)^2}{(m + n - 2)\sigma^2} \xrightarrow{P} 1$$

for samples  $X_1, \dots, X_m; Y_1, \dots, Y_n$  from any common distribution  $F$  with finite variance. Thus, the rejection probability  $\alpha_{m,n}(F)$  tends to  $\alpha$  for any such  $F$ . As will be seen in Section 13.2.3, the same robustness property for the UMP invariant test of equality of  $s$  means also holds.

On the other hand, the  $F$ -test for variances, just like the one-sample  $\chi^2$ -test, is extremely sensitive to the assumption of normality. To see this, express the rejection region in terms of  $\log S_Y^2 - \log S_X^2$ , where

$$S_X^2 = \frac{\sum(X_i - \bar{X}_m)^2}{m - 1}$$

and

$$S_Y^2 = \frac{\sum(Y_j - \bar{Y}_n)^2}{n - 1}.$$

Also, suppose that as  $m$  and  $n \rightarrow \infty$ ,  $m/(m+n)$  remains fixed at  $\rho$ . By the result for the one-sample problem and the delta method with  $g(u) = \log u$  (Theorem 11.3.4), it is seen that  $\sqrt{m}[\log S_X^2 - \log \sigma^2]$  and  $\sqrt{n}[\log S_Y^2 - \log \sigma^2]$  both tend in law to  $N(0, \gamma^2/\sigma^4)$  when the  $X$ 's and  $Y$ 's are distributed as  $F$ , and hence that  $\sqrt{m+n}[\log S_Y^2 - \log S_X^2]$  tends in law to the normal distribution with mean 0 and variance

$$\frac{\gamma^2}{\sigma^4} \left( \frac{1}{\rho} + \frac{1}{1-\rho} \right) = \frac{\gamma^2}{\rho(1-\rho)\sigma^4} .$$

In the particular case that  $F$  is normal,  $\gamma^2 = 2\sigma^4$  and the variance of the limit distribution is  $2/\rho(1-\rho)$ . For other distributions  $\gamma^2/\sigma^4$  can take on any positive value and, as in the one-sample case,  $\alpha_n(F)$  can tend to any limit less than  $\frac{1}{2}$ . [For an entry into the extensive literature on more robust alternatives, see for example Conover et al. (1981), Tiku and Balakrishnan (1984), Boos and Brownie (1989), Baker (1995), Hall and Padmanabhan (1997), and Section 2.10 of Hettmansperger and McKean (1998)].

Having found that the rejection probability of the one- and two-sample  $t$ -tests is relatively insensitive to nonnormality (at least for large samples), let us turn to the corresponding question concerning the power of these tests. By similar asymptotic calculations, it can be shown that the same conclusion holds. Power values of the  $t$ -tests obtained under normality are asymptotically valid also for all other distributions with finite variance. This is a useful result if it has been decided to employ a  $t$ -test and one wishes to know what power it will have against a given alternative  $\xi/\sigma$  or  $(\eta - \xi)/\sigma$ , or what sample sizes are required to obtain a given power.

Recall that there exists a modification of the  $t$ -test, the *permutation* version of the  $t$ -test discussed in Section 5.9, whose size is independent of  $F$  not only asymptotically but exactly. Moreover, we will see in Section 17.2 that its asymptotic power is equal to that of the  $t$ -test. It may seem that the permutation  $t$ -test has all the properties one could hope for. However, this overlooks the basic question of whether the  $t$ -test itself, which is optimal under normality, will retain a high standing with respect to its competitors under other distributions. The  $t$ -tests are in fact not robust in this sense. Some tests which are preferable when a broad spectrum of distributions  $F$  is considered possible were discussed in Section 6.9. A permutation test with this property has been proposed by Lambert (1985).

As a last problem, consider the level of the two-sample  $t$ -test when the variances  $\text{Var}(X_i) = \sigma^2$  and  $\text{Var}(Y_j) = \tau^2$  may differ (as in the Behrens–Fisher problem), and the assumption of normality may fail as well. As before, one finds that  $(\bar{Y}_m - \bar{X}_n)/\sqrt{\sigma^2/m + \tau^2/n}$  tends in law to  $N(0, 1)$  as  $m, n \rightarrow \infty$ , while  $S_X^2 = \sum(X_i - \bar{X}_m)^2/(m - 1)$  and  $S_Y^2 = \sum(Y_i - \bar{Y}_n)^2/(n - 1)$ , respectively, tend to  $\sigma^2$  and  $\tau^2$  in probability. If  $m$  and  $n$  tend to  $\infty$  through a sequence with fixed proportion  $m/(m + n) = \rho$ , the squared denominator of the  $t$ -statistic,

$$D^2 = \frac{m - 1}{m + n - 2} S_X^2 + \frac{n - 1}{m + n - 2} S_Y^2 ,$$

tends in probability to  $\rho\sigma^2 + (1 - \rho)\tau^2$ , and the limit of

$$t = \frac{1}{\sqrt{\frac{1}{m} + \frac{1}{n}}} \left( \frac{\bar{Y}_n - \bar{X}_m}{\sqrt{\frac{\sigma^2}{m} + \frac{\tau^2}{n}}} \cdot \frac{\sqrt{\frac{\sigma^2}{m} + \frac{\tau^2}{n}}}{D} \right)$$

is normal with mean zero and variance

$$\frac{(1 - \rho)\sigma^2 + \rho\tau^2}{\rho\sigma^2 + (1 - \rho)\tau^2}. \quad (13.2)$$

The ratio (13.2) is exactly one if and only if  $\rho = \frac{1}{2}$  or  $\sigma = \tau$ . When  $m = n$ , so that  $\rho = \frac{1}{2}$ , the  $t$ -test thus has approximately the right level even if  $\sigma$  and  $\tau$  are far apart. The accuracy of this approximation for different values of  $m = n$  and  $\tau/\sigma$  is discussed by Ramsey (1980) and Posten et al. (1982). However, when  $\rho \neq \frac{1}{2}$ , the actual size of the test can differ greatly from the nominal level  $\alpha$  even for large  $m$  and  $n$ . An approximate test of the hypothesis  $H : \eta = \xi$  when  $\sigma, \tau$  are not assumed equal, which asymptotically is free of this difficulty, can be obtained through Studentization, i.e., by replacing  $D^2$  with  $(1/m)S_X^2 + (1/n)S_Y^2$  and referring the resulting statistic to the standard normal distribution. This approximation is very crude, and not reliable unless  $m$  and  $n$  are fairly large. A refinement, the *Welch approximate  $t$ -test*, refers the resulting statistic not to the standard normal but to the  $t$ -distribution with a random number of degrees of freedom  $f$  given by

$$\frac{1}{f} = \left( \frac{R}{1 + R} \right)^2 \frac{1}{m - 1} + \frac{1}{(1 + R)^2} \cdot \frac{1}{n - 1},$$

where  $R = (nS_X^2)/(mS_Y^2)$ .<sup>2</sup> When the  $X$ 's and  $Y$ 's are normal, the actual level of this test has been shown to be quite close to the nominal level for sample sizes as small as  $m = 4, n = 8$  and  $m = n = 6$  [see Wang (1971)]. A further refinement will be mentioned in Section 18.5. A simple but crude approach that controls the level is to use as degrees of freedom the smaller of  $n - 1$  and  $m - 1$ , as remarked by Scheffé (1970). Two-sample permutation tests will be studied in Section 17.3.

The robustness of the level of Welch's test against nonnormality is studied by Yuen (1974), who shows that for heavy-tailed distributions the actual level tends to be considerably smaller than the nominal level (which leads to an undesirable loss of power), and who proposes an alternative. Some additional results are discussed in Scheffé (1970) and in Tiku and Singh (1981). The robustness of some quite different competitors of the  $t$ -test is investigated in Pratt (1964).

For testing the equality of  $s$  normal means with  $s > 2$ , the classical test based on the  $F$ -statistic (7.19) is not robust, even if all the observations are normally distributed, regardless of the sample sizes (Scheffé (1959), Problem 13.25); again, the problem is due to the assumption of a common variance. More appropriate tests for this generalized Behrens–Fisher problem have been proposed by Welch (1951), James (1951), and Brown and Forsythe (1974a), and is further discussed by Clinch and Kesselman (1982), Hettmansperger and McKean (1998) and Chapter 10 of Pesarin (2001). The corresponding robustness problem for more general linear hypotheses is treated by James (1954a, 1954b) and Johansen (1980); see also Rothenberg (1984).

<sup>2</sup> For a variant see Fenstad (1983).

### 13.2.2 Effect of Dependence

The one-sample  $t$ -test arises when a sequence of measurements  $X_1, \dots, X_n$ , is taken of a quantity  $\xi$ , and the  $X$ 's are assumed to be independently distributed as  $N(\xi, \sigma^2)$ . The effect of nonnormality on the level of the test was discussed in the preceding subsection. Independence may seem like a more innocuous assumption. However, it has been found that observations occurring close in time or space are often positively correlated [Student (1927), Hotelling (1961), Cochran (1968)]. The present section will therefore be concerned with the effect of this type of dependence.

**Lemma 13.2.1** *Let  $X_1, \dots, X_n$  be jointly normally distributed with common marginal distribution  $N(0, \sigma^2)$  and with correlation coefficients  $\rho_{i,j} = \text{corr}(X_i, X_j)$ . Assume that*

$$\frac{1}{n} \sum_{i \neq j} \rho_{i,j} \rightarrow \gamma \tag{13.3}$$

and

$$\frac{1}{n^2} \sum_{i \neq j} \rho_{i,j}^2 \rightarrow 0 \tag{13.4}$$

as  $n \rightarrow \infty$ . Then,

(i) the distribution of the  $t$ -statistic  $t_n$  defined in Equation (13.1) (with  $\xi_0 = 0$ ) tends to the normal distribution  $N(0, 1 + \gamma)$ ;

(ii) if  $\gamma \neq 0$ , the level of the  $t$ -test is not robust even asymptotically as  $n \rightarrow \infty$ . Specifically, if  $\gamma > 0$ , the asymptotic level of the  $t$ -test carried out at nominal level  $\alpha$  is

$$1 - \Phi\left(\frac{z_{1-\alpha}}{\sqrt{1+\gamma}}\right) > 1 - \Phi(z_{1-\alpha}) = \alpha.$$

PROOF. (i): Since the  $X_i$  are jointly normal, the numerator  $\sqrt{n}\bar{X}_n$  of  $t_n$  is also normal, with mean zero and variance

$$\text{Var}(\sqrt{n}\bar{X}) = \sigma^2 \left[ 1 + \frac{1}{n} \sum_{i \neq j} \rho_{i,j} \right] \rightarrow \sigma^2(1 + \gamma), \tag{13.5}$$

and hence tends in law to  $N(0, \sigma^2(1 + \gamma))$ . The denominator of  $t_n$  is the square root of

$$S_n^2 = \frac{1}{n-1} \sum X_i^2 - \frac{n}{n-1} \bar{X}_n^2.$$

By (13.5),  $\text{Var}(\bar{X}_n) \rightarrow 0$  and so  $\bar{X}_n \xrightarrow{P} 0$ . A calculation similar to (13.5) shows that  $\text{Var}(n^{-1} \sum_{i=1}^n X_i^2) \rightarrow 0$  (Problem 13.4). Thus,  $n^{-1} \sum_{i=1}^n X_i^2 \xrightarrow{P} \sigma^2$  and so  $S_n \xrightarrow{P} \sigma$ . By Slutsky's Theorem, the distribution of  $t_n$  therefore tends to  $N(0, 1 + \gamma)$ .

The implications (ii) are obvious. ■

Under the assumptions of Lemma 13.2.1, the joint distribution of the  $X$ 's is determined by  $\sigma^2$  and the correlation coefficients  $\rho_{i,j}$ , with the asymptotic level of the  $t$ -test depending only on  $\gamma$ . The following examples illustrating different correlation structures show that even under rather weak dependence of the observations, the assumptions of Lemma 13.2.1 are satisfied with  $\gamma \neq 0$ , and hence that the level of the  $t$ -test is quite sensitive to the assumption of independence.

MODEL A. (CLUSTER SAMPLING) Suppose the observations occur in  $s$  groups (or clusters) of size  $m$ , and that any two observations within a group have a common correlation coefficient  $\rho$ , while those in different groups are independent. (This may be the case, for instance, when the observations within a group are those taken on the same day or by the same observer, or involve some other common factor.) Then (Problem 13.6),

$$\text{Var}(\bar{X}) = \frac{\sigma^2}{ms} [1 + (m - 1)\rho],$$

which tends to zero as  $s \rightarrow \infty$ . The conditions of the lemma hold with  $\gamma = (m - 1)\rho$ , and the level of the  $t$ -test is not asymptotically robust as  $s \rightarrow \infty$ . In particular, the test overstates the significance of the results when  $\rho > 0$ .

To provide a specific structure leading to this model, denote the observations in the  $i$ th group by  $X_{i,j}$  ( $j = 1, \dots, m$ ), and suppose that  $X_{i,j} = A_i + U_{i,j}$ , where  $A_i$  is a factor common to the observations in the  $i$ th group. If the  $A$ 's and  $U$ 's (none of which are observable) are all independent with normal distributions  $N(\xi, \sigma_A^2)$  and  $N(0, \sigma_0^2)$ , respectively, then the joint distribution of the  $X$ 's is that prescribed by Model A with  $\sigma^2 = \sigma_A^2 + \sigma_0^2$  and  $\rho = \sigma_A^2/\sigma^2$ .

MODEL B. (MOVING-AVERAGE PROCESS) When the dependence of nearby observations is not due to grouping as in Model A, it is often reasonable to assume that  $\rho_{i,j}$  depends only on  $|j - i|$  and is nonincreasing in  $|j - i|$ . Let  $\rho_{i,i+k}$  then be denoted by  $\rho_k$ , and suppose that the correlation between  $X_i$  and  $X_{i+k}$  is negligible for  $k > m$  ( $m$  an integer  $< n$ ), so that one can put  $\rho_k = 0$  for  $k > m$ . Then the conditions for Lemma 13.2.1 are satisfied with

$$\gamma = 2 \sum_{k=1}^m \rho_k.$$

In particular, if  $\rho_1, \dots, \rho_m$  are all positive, the  $t$ -test is again too liberal.

A specific structure leading to Model B is given by the moving-average process

$$X_i = \xi + \sum_{j=0}^m \beta_j U_{i+j},$$

where the  $U$ 's are independent  $N(0, \sigma_0^2)$ . Such a process was discussed in Example 12.4.1 and Theorem 12.4.1. The variance  $\sigma^2$  of the  $X$ 's is then  $\sigma^2 = \sigma_0^2 \sum_{j=0}^m \beta_j^2$  and



$$\rho_k = \begin{cases} \frac{\sum_{i=0}^{m-k} \beta_i \beta_{i+k}}{\sum_{j=0}^m \beta_j^2} & \text{for } k \leq m, \\ 0 & \text{for } k > m. \end{cases}$$

MODEL C. (FIRST-ORDER AUTOREGRESSIVE PROCESS) A simple model for dependence in which the  $|\rho_k|$  are decreasing in  $k$  but  $\neq 0$  for all  $k$  is the *first-order autoregressive process* previously introduced in Example 12.4.2. Here, we assume the underlying distribution of the observations is normal. Define

$$X_{i+1} = \xi + \beta(X_i - \xi) + U_{i+1}, \quad |\beta| < 1, \quad i = 1, \dots, n,$$

with the  $U_i$  independent  $N(0, \sigma_0^2)$ . If  $X_1$  is  $N(\xi, \tau^2)$ , the marginal distribution of  $X_i$  for  $i > 1$  is normal with mean  $\xi$  and variance  $\sigma_i^2 = \beta^2 \sigma_{i-1}^2 + \sigma_0^2$ . The variance of  $X_i$  will thus be independent of  $i$  provided  $\tau^2 = \sigma_0^2 / (1 - \beta^2)$ . For the sake of simplicity we shall assume this to be the case, and take  $\xi$  to be zero. From

$$X_{i+k} = \beta^k X_i + \beta^{k-1} U_{i+1} + \beta^{k-2} U_{i+2} + \dots + \beta U_{i+k-1} + U_{i+k}$$

it then follows that  $\rho_k = \beta^k$ , so that the correlation between  $X_i$  and  $X_j$  decreases exponentially with increasing  $|j - i|$ . The assumptions of Lemma 13.2.1 are again satisfied, and  $\gamma = 2\beta / (1 - \beta)$ . Thus, in this case too, the level of the  $t$ -test is not asymptotically robust. [Some values of the actual asymptotic level when the nominal level is 0.05 or 0.01 are given by Gastwirth and Rubin (1971).]

In Models A, B, and C, we have seen that the null rejection probability may be far from the nominal level, even in large samples and when the underlying distributions are normal. One can consider alternatives to both normality and independence simultaneously, using the results in Section 12.4, but the general conclusions remain the same. In summary, the effect of dependence on the level of the  $t$ -test is more serious than that of nonnormality.

In order to robustify the test against general dependence through studentization (as was done in the two-sample case with unequal variances), it is necessary to consistently estimate  $\gamma$ , which implicitly depends on estimation of all the  $\rho_{i,j}$ . Unfortunately, the number of parameters  $\rho_{i,j}$  exceeds the number of observations. However, robustification is possible against some types of dependence. For example, it may be reasonable to assume a model such as A–C so that it is only required to estimate a reduced number of correlations.<sup>3</sup> Some specific procedures of this type are discussed by Albers (1978), [and for an associated sign test by Falk and Kohne (1984)]. Such robust procedures will in fact often also be insensitive to the assumption of normality, as can be shown by appealing to an appropriate Central Limit Theorem for dependent variables, such as those in Section 12.4. The validity of these procedures is of

---

<sup>3</sup> Models of a sequence of dependent observations with various covariance structures are discussed in books on time series such as Brockwell and Davis (1991), Hamilton (1994), Fuller (1996) or McElroy and Politis (2020).

course limited to the particular model assumed, including the value of a parameter such as  $m$  in Models A and B. In fact, robustification is achievable for fairly general classes of models with dependence by using an appropriate bootstrap method; see Problem 18.19 and Lahiri (2003). Alternatively, one can use subsampling, as in Romano and Thoms (1996); see Section 18.7.

The results of the present section easily extend to the case of the two-sample  $t$ -test, when each of the two series of observations shows dependence of the kind considered here.

### 13.2.3 Robustness in Linear Models

In this section, we consider the large-sample robustness properties of some of the linear model tests discussed in Chapter 7. As in Section 13.2.1, we focus on the effect of distribution.

A large class of these testing situations is covered by the following general model, which was discussed in Problem 7.8. Let  $X_1, \dots, X_n$  be independent with  $E(X_i) = \xi_i$  and  $Var(X_i) = \sigma^2 < \infty$ , where we assume the vector  $\xi$  to lie in an  $s$ -dimensional subspace  $\Pi_\Omega$  of  $\mathbb{R}^n$ , defined by the following parametric set of equations

$$\xi_i = \sum_{j=1}^s a_{i,j} \beta_j, \quad i = 1, \dots, n. \quad (13.6)$$

Here the  $a_{i,j}$  are known coefficients and the  $\beta_j$  are unknown parameters. In matrix form, the  $n \times 1$  vector  $\xi$  with  $i$ th component  $\xi_i$  satisfies

$$\xi = A\beta, \quad (13.7)$$

where  $A$  is an  $n \times s$  matrix having  $(i, j)$  entry  $a_{i,j}$  and  $\beta$  is an  $s \times 1$  vector with  $j$ th component  $\beta_j$ . It is assumed  $A$  is known and of rank  $s$ . In the asymptotics below, the  $a_{i,j}$  may depend on  $n$ , as will  $\xi$ , but  $s$  remains fixed. Throughout, the notation will suppress this dependence on  $n$ . The parameter vector  $\beta$  does not change with  $n$ .

The least squares estimators  $\hat{\xi}_1, \dots, \hat{\xi}_n$  of  $\xi_1, \dots, \xi_n$  are defined as the values of  $\xi_i$  minimizing

$$\sum_{i=1}^n (X_i - \xi_i)^2$$

subject to  $\xi \in \Pi_\Omega$ , where  $\Pi_\Omega$  is the space spanned by the  $s$  columns of  $A$ . Correspondingly, the least squares estimators  $\hat{\beta}_1, \dots, \hat{\beta}_s$  of  $\beta_1, \dots, \beta_s$  are the values of  $\beta_j$  minimizing

$$\sum_{i=1}^n \left( X_i - \sum_{j=1}^s a_{i,j} \beta_j \right)^2.$$

By taking partial derivatives of this last expression with respect to the  $\beta_j$ , it is seen that the  $\hat{\beta}_j$  are solutions of the equations

$$A^T A \beta = A^T X$$

and so

$$\hat{\beta} = (A^T A)^{-1} A^T X .$$

(The fact that  $A^T A$  is nonsingular follows from Problem 6.3.) Thus,

$$\hat{\xi} = P X ,$$

where

$$P = A(A^T A)^{-1} A^T . \quad (13.8)$$

In fact,  $\hat{\xi}$  is the projection of  $X$  into the space  $\Pi_\Omega$ . (These estimators formed the basis of optimal invariant tests studied in Chapter 7.) Some basic properties of  $P$  and  $\hat{\xi}$  are recorded in the following lemma.

**Lemma 13.2.2** (i) *The matrix  $P$  defined by (13.8) is symmetric ( $P = P^T$ ) and idempotent ( $P^2 = P$ ).*

(ii)  *$X - \hat{\xi}$  is orthogonal to  $\hat{\xi}$ ; that is,*

$$\hat{\xi}^T (X - \hat{\xi}) = 0 .$$

PROOF. The proof of (i) follows by matrix algebra (Problem 13.10). To prove (ii), note that

$$\begin{aligned} \hat{\xi}^T (X - \hat{\xi}) &= (P X)^T (X - P X) = X^T P^T (X - P X) \\ &= X^T P^T X - X^T P^T P X = 0 , \end{aligned}$$

since by (i)  $P^T P = P^T$ . ■

Note that  $\hat{\beta}_j$  is a linear combination of the  $X_i$ . Thus, if the  $X_i$  are normally distributed, so are the  $\hat{\beta}_j$ . However, we would like to understand their properties when then the  $X_i$  are not normally distributed. We shall now suppose that the model (13.6) is embedded in a sequence of such models defined by matrices  $A_{i,j}^{(n)}$ , with  $s$  fixed and  $n \rightarrow \infty$ . Suppose that the  $X$ s are not normal but given by

$$X_i = U_i + \xi_i ,$$

where the  $U$ s are i.i.d. according to a distribution  $F$  with mean 0 and variance  $\sigma^2 < \infty$ . Since  $E(X) = \xi = A\beta$ , the least squares estimator  $\hat{\beta}$  is unbiased; that is,

$$E(\hat{\beta}) = (A^T A)^{-1} A^T E(X) = \beta . \quad (13.9)$$

Without the assumption of normality, the asymptotic normality of  $\hat{\beta}_j$  can be established by the following lemma, which can be obtained as a consequence of the Lindeberg Central Limit Theorem (Problem 13.11).

**Lemma 13.2.3** *Let  $Y_1, Y_2, \dots$  be independent and identically distributed with mean zero and finite variance  $\sigma^2$ . (i) Let  $c_1, c_2, \dots$  be a sequence of constants. Then a sufficient condition for*

$$\frac{\sum_{i=1}^n c_i Y_i}{\sqrt{\sum c_i^2}} \xrightarrow{d} N(0, \sigma^2)$$

is that

$$\frac{\max_{i=1, \dots, n} c_i^2}{\sum_{j=1}^n c_j^2} \rightarrow 0 \quad \text{as } n \rightarrow \infty. \quad (13.10)$$

(ii) More generally, suppose  $C_{n,1}, \dots, C_{n,n}$  is a sequence of random variables, independent of  $Y_1, \dots, Y_n$ . Then, a sufficient condition for

$$\frac{\sum_{i=1}^n C_{n,i} Y_i}{\sqrt{\sum C_{n,i}^2}} \xrightarrow{d} N(0, \sigma^2)$$

is

$$\frac{\max_{i=1, \dots, n} C_{n,i}^2}{\sum_{j=1}^n C_{n,j}^2} \xrightarrow{P} 0 \quad \text{as } n \rightarrow \infty. \quad (13.11)$$

Moreover, (13.11) implies that

$$P \left\{ \frac{C_{n,i} Y_i}{\sqrt{\sum C_{n,i}^2}} \leq z | C_{n,1}, \dots, C_{n,n} \right\} \xrightarrow{P} \Phi\left(\frac{z}{\sigma}\right).$$

Condition (13.10) prevents the  $c$ 's from increasing so fast that the last term essentially dominates the sum, in which case there is no reason to expect asymptotic normality.

**Example 13.2.1** Suppose  $U_1, U_2, \dots$  are i.i.d. with mean 0 and finite nonzero variance  $\sigma^2$ . Consider the simple regression model

$$X_i = \alpha + \beta t_i + U_i,$$

where the  $t_i$  are known and not all equal. The least squares estimator  $\hat{\beta}$  of  $\beta$  satisfies

$$\hat{\beta} - \beta = \frac{\sum (X_i - \alpha - \beta t_i)(t_i - \bar{t})}{\sum (t_i - \bar{t})^2} .$$

By Lemma 13.2.3,

$$\frac{(\hat{\beta} - \beta)\sqrt{\sum (t_i - \bar{t})^2}}{\sigma} \xrightarrow{d} N(0, 1)$$

provided

$$\frac{\max(t_i - \bar{t})^2}{\sum (t_j - \bar{t})^2} \rightarrow 0 . \quad (13.12)$$

Condition (13.12) holds in the case of equal spacing  $t_i = a + i\Delta$ , but not when the  $t$ 's grow exponentially, for example, when  $t_i = 2^i$  (Problem 13.12). ■

Consider the hypothesis

$$H : \theta = \sum_{j=1}^s b_j \beta_j = 0 , \quad (13.13)$$

where the  $b$ 's are known constants with  $\sum b_j^2 = 1$ . Assume without loss of generality that  $A^T A = I$ , the identity matrix, so that the columns of  $A$  are mutually orthogonal and of length one. The least squares estimator of  $\theta$  is given by

$$\hat{\theta} = \sum_{j=1}^s b_j \hat{\beta}_j = \sum_{i=1}^n d_i X_i , \quad (13.14)$$

where by (13.7)

$$d_i = \sum_{j=1}^s a_{i,j} b_j \quad (13.15)$$

(Problem 13.13). By the assumption that the columns of  $A$  are orthogonal,  $\sum d_i^2 = \sum b_j^2 = 1$ . So, under  $H$ ,

$$E(\hat{\theta}) = \sum_{j=1}^s E(b_j \hat{\beta}_j) = \sum_{j=1}^s b_j \beta_j = 0$$

and

$$\text{Var}(\hat{\theta}) = \text{Var}\left(\sum_{i=1}^n d_i X_i\right) = \sigma^2 \sum_{i=1}^n d_i^2 = \sigma^2 .$$

The uniformly most powerful invariant test rejects  $H$  when the  $t$ -statistic satisfies

$$\frac{|\hat{\theta}|}{\sqrt{\sum (X_i - \hat{\xi}_i)^2 / (n - s)}} \geq C . \quad (13.16)$$

We would now like to examine the level when then the  $X_i$  are not normally distributed. Assume

$$X_i = U_i + \xi_i ,$$

where the  $U$ s are i.i.d. according to a distribution  $F$  with mean 0 and variance  $\sigma^2 < \infty$ .

The denominator of (13.16) tends in probability to  $\sigma$ . To see why, with  $s$  fixed, it suffices to show

$$\frac{1}{n} \sum (X_i - \hat{\xi}_i)^2 \xrightarrow{P} \sigma^2 .$$

But, the left side is

$$\frac{\sum (X_i - \xi_i)^2}{n} + \frac{2 \sum (X_i - \xi_i)(\xi_i - \hat{\xi}_i)}{n} + \frac{\sum (\xi_i - \hat{\xi}_i)^2}{n} .$$

The first term tends in probability to  $\sigma^2$ , by the Weak Law of Large Numbers. By the Cauchy–Schwarz inequality, half the middle term is bounded by the square root of the product of the first and third terms. Therefore, it suffices to show the third term tends to 0 in probability. Since this term is nonnegative, it suffices to show its expectation tends to 0, by Markov’s Inequality (Problem 11.27). But its expectation is the trace of the covariance matrix of  $\hat{\xi}$  divided by  $n$ . Letting  $I_n$  denote the  $n \times n$  identity matrix, the covariance matrix of  $\hat{\xi} = PX$  is

$$\sigma^2 P I_n P^\top = \sigma^2 P P^\top = \sigma^2 P .$$

But, the trace of  $P$  is

$$\text{tr}(P) = \text{tr}(A(A^\top A)^{-1}A^\top) = \text{tr}(A^\top A(A^\top A)^{-1}) = \text{tr}(I_s) = s ,$$

since  $\text{tr}(BC) = \text{tr}(CB)$  for any  $n \times s$  matrix  $B$  and  $s \times n$  matrix  $C$ . Hence, the denominator of (13.16) converges in probability to  $\sigma$ . By Lemma 13.2.3, the numerator of (13.16) converges in distribution to  $N(0, \sigma^2)$  provided

$$\max d_i^2 \rightarrow 0 \quad \text{as } n \rightarrow \infty . \quad (13.17)$$

Under this condition, the level of the  $t$ -test is therefore robust against nonnormality.

So far,  $b = (b_1, \dots, b_s)^\top$  has been fixed. To determine when the level of (13.16) is robust for all  $b$  with  $\sum b_j^2 = 1$ , it is only necessary to find the maximum value of  $d_i^2$  as  $b$  varies. By the Cauchy–Schwarz inequality

$$d_i^2 = \left( \sum_j a_{i,j} b_j \right)^2 \leq \sum_{j=1}^s a_{i,j}^2,$$

with equality holding when  $b_j = a_{i,j} / \sqrt{\sum_k a_{i,k}^2}$ . The desired maximum of  $d_i^2$  is therefore  $\sum_j a_{i,j}^2$ , and

$$\max_i \sum_{j=1}^s a_{i,j}^2 \rightarrow 0 \quad \text{as } n \rightarrow \infty \quad (13.18)$$

is a sufficient condition for the asymptotic normality of every  $\hat{\theta}$  of the form (13.14).

Condition (13.18) depends on the particular parametrization (13.6) chosen for  $\Pi_{\Omega}$ . Note however that

$$\sum_{j=1}^s a_{i,j}^2 = \Pi_{i,i}, \quad (13.19)$$

where  $\Pi_{i,j}$  is the  $(i, j)$  element of the projection matrix  $P$ .

This shows that the value of  $\Pi_{i,i}$  is coordinate free, i.e., it is unchanged by an arbitrary change of coordinates  $\beta^* = B^{-1}\beta$ , where  $B$  is a nonsingular matrix, since

$$\xi = A\beta = AB\beta^* = A^*\beta^*$$

with  $A^* = AB$ , and

$$P^* = AB(B^T A^T AB)^{-1} B^T A^T = ABB^{-1}(A^T A)^{-1}(B^T)^{-1} BA = P.$$

Hence, (13.18) is equivalent to the coordinate-free Huber condition

$$\max_i \Pi_{i,i} \rightarrow 0 \quad \text{as } n \rightarrow \infty. \quad (13.20)$$

For evaluating  $\Pi_{i,i}$ , it is helpful to note that

$$\hat{\xi}_i = \sum_{j=1}^n \Pi_{i,j} X_j \quad (i = 1, \dots, n),$$

so that  $\Pi_{i,i}$  is simply the coefficient of  $X_i$  in  $\hat{\xi}_i$ , which must be calculated in any case to carry out the test.

If  $\Pi_{i,i} \leq M_n$  for all  $i = 1, \dots, n$ , then also  $\Pi_{i,j} \leq M_n$  for all  $i$  and  $j$ . This follows from the fact that there exists a nonsingular  $E$  with  $P = EE^T$ , on applying the Cauchy–Schwarz inequality to the  $(i, j)$  element of  $EE^T$ . Condition (13.20) is therefore equivalent to

$$\max_{i,j} \Pi_{i,j} \rightarrow 0 \quad \text{as } n \rightarrow \infty. \quad (13.21)$$

**Example 13.2.2 (Example 13.2.1, continued)** In Example 13.2.1, the coefficient of  $X_i$  in  $\hat{\xi}_i = \hat{\alpha} + \hat{\beta}t_i$  is

$$\Pi_{i,i} = \frac{1}{n} + \frac{(t_i - \bar{t})^2}{\sum (t_j - \bar{t})^2}$$

and the Huber condition reduces to Condition (13.12) found earlier. ■

**Example 13.2.3 (Two-way Layout)** Consider the two-way layout with  $m$  observations per cell and the additive model

$$\xi_{i,j,k} = E(X_{i,j,k}) = \mu + \alpha_i + \beta_j$$

with

$$\sum_i \alpha_i = \sum_j \beta_j = 0,$$

$i = 1, \dots, a; j = 1, \dots, b; k = 1, \dots, m$ . It is easily seen (Problem 13.14) that, for fixed  $a$  and  $b$ , the Huber condition is satisfied as  $m \rightarrow \infty$ . ■

Let us next generalize the hypothesis (13.13) to hypotheses which impose several linear constraints. Without loss of generality, choose the parametrization in (13.6) in such a way that the  $s$  columns of  $A$  are orthogonal and of length one and make the transformation

$$Y = CX$$

(as used in (7.1)), where  $C$  is orthogonal and the first  $s$  rows of  $C$  are equal to those of  $A^\top$ , say

$$C = \begin{pmatrix} A^\top \\ D \end{pmatrix} \quad (13.22)$$

for some  $(n-s) \times n$  matrix  $D$ . If  $\eta_i = E(Y_i)$ , we then have that

$$\eta = \begin{pmatrix} A^\top \\ D \end{pmatrix} A\beta = (\beta_1, \dots, \beta_s, 0, \dots, 0)^\top. \quad (13.23)$$

By the orthogonality of  $C$ , the  $Y_i$  are independent with  $Y_i$  distributed as  $N(\eta_i, \sigma^2)$ , where  $\eta_i = \beta_i$  for  $i = 1, \dots, s$  and  $\eta_i = 0$  for  $i = s+1, \dots, n$ . We want to test

$$H: \sum_{j=1}^s \alpha_{i,j} \eta_j = 0; \quad i = 1, \dots, r$$

where we shall assume that the  $r$  vectors  $(\alpha_{i,1}, \dots, \alpha_{i,s})^\top$  are orthogonal and of length one. Then the variables



$$Z_i = \begin{cases} \sum_{j=1}^n \alpha_{i,j} Y_j & i = 1, \dots, r \\ Y_i & i = s + 1, \dots, n \end{cases} \tag{13.24}$$

are independent  $N(\zeta_i, \sigma^2)$  with

$$\zeta_i = \begin{cases} \sum_{j=1}^s \alpha_{i,j} \eta_j & i = 1, \dots, r \\ \eta_i & i = r + 1, \dots, s \\ 0 & i = s + 1, \dots, n \end{cases} \tag{13.25}$$

The standard UMPI test of  $H : \zeta_1 = \dots = \zeta_r = 0$  rejects when

$$\frac{\sum_{i=1}^r Z_i^2 / r}{\sum_{j=s+1}^n Z_j^2 / (n - s)} > k, \tag{13.26}$$

where  $k$  is determined so that the probability of (13.26) is  $\alpha$  when the  $Z$ s are normal and  $H$  holds.

As before, suppose that the model (13.6) is embedded in a sequence of models defined by matrices  $A_{i,j}^{(n)}$ , with  $s$  fixed and  $n \rightarrow \infty$ . Suppose that the  $X$ s satisfy

$$X_i = U_i + \xi_i,$$

where the  $U$ s are i.i.d. according to a distribution  $F$  with mean 0 and variance  $\sigma^2 < \infty$ . We then have the following robustness result.

**Theorem 13.2.1** *Let  $\alpha_n(F)$  denote the rejection probability of the test (13.26) when the  $U$ s have distribution  $F$  and the null hypothesis constraints are satisfied. Then,  $\alpha_n(F) \rightarrow \alpha$  provided*

$$\max_i \sum_{j=1}^s (a_{i,j}^{(n)})^2 \rightarrow 0 \tag{13.27}$$

or equivalently

$$\max \Pi_{i,i}^{(n)} \rightarrow 0,$$

where  $\Pi_{i,i}^{(n)}$  is the  $i$ th diagonal element of  $P = A(A^\top A)^{-1} A^\top$ .

PROOF. We must show that the limiting distribution of (13.26) is the same as when  $F$  is normal. First, we shall show that the denominator of (13.26) satisfies

$$\frac{1}{n - s} \sum_{j=s+1}^n Z_j^2 \xrightarrow{P} \sigma^2. \tag{13.28}$$

Note that  $X = C^\top Y$  and  $Y = QZ$  where  $C^\top$  and  $Q$  are both orthogonal. Therefore,

$$\begin{aligned} \frac{1}{n-s} \sum_{j=s+1}^n Z_j^2 &= \frac{n}{n-s} \left[ \frac{1}{n} \sum_{i=1}^n Z_i^2 \right] - \frac{1}{n-s} \sum_{i=1}^s Z_i^2 \\ &= \frac{n}{n-s} \cdot \frac{1}{n} \sum_{i=1}^n X_i^2 - \frac{1}{n-s} \sum_{i=1}^s Z_i^2. \end{aligned}$$

To see that this tends to  $\sigma^2$  in probability, we first show that

$$\frac{1}{n} \sum_{i=1}^n X_i^2 \xrightarrow{P} \sigma^2.$$

But,

$$\frac{\sum_{i=1}^n X_i^2}{n} = \frac{\sum_{i=1}^n (X_i - \xi_i)^2}{n} + \frac{2 \sum_{i=1}^n \xi_i X_i}{n} - \frac{\sum_{i=1}^n \xi_i^2}{n}.$$

The first term on the right tends to  $\sigma^2$  in probability, by the Weak Law of Large Numbers. By the orthogonality of  $C$ , the last term is equal to  $\sum_{i=1}^s \beta_i^2/n$ , which tends to 0 since  $s$  is fixed. It is easily checked that the middle term has a mean and variance which tend to 0. Hence,  $\sum X_i^2/n$  tends in probability to  $\sigma^2$ . Next, we show that

$$\frac{\sum_{i=1}^s Z_i^2}{n} \xrightarrow{P} 0.$$

It suffices to show

$$\frac{\sum_{i=1}^s E(Z_i^2)}{n} = \frac{\sum_{i=1}^s \text{Var}(Z_i)}{n} + \frac{\sum_{i=1}^s [E(Z_i)]^2}{n} \rightarrow 0.$$

Since  $s$  is fixed and  $\text{Var}(Z_i) = \sigma^2$ , we only need to show

$$\frac{\sum_{i=1}^s [E(Z_i)]^2}{n} \rightarrow 0.$$

For  $i \leq r$ ,

$$E(Z_i) = \sum_{j=1}^s \alpha_{i,j} \eta_j = \sum_{j=1}^s \alpha_{i,j} \beta_j$$

and

$$[E(Z_i)]^2 \leq \sum_{j=1}^s \alpha_{i,j}^2 \sum_{j=1}^s \beta_j^2 = \sum_{j=1}^s \beta_j^2.$$

For  $r+1 \leq i \leq s$ ,  $E(Z_i) = \beta_i$ , in which case the same bound holds. Therefore,

$$\frac{\sum_{i=1}^s [E(Z_i)]^2}{n} \leq \frac{s \sum_{j=1}^s \beta_j^2}{n} \rightarrow 0,$$

and the result (13.28) follows.

Next, we consider the numerator of (13.26). We show the joint asymptotic normality of  $(Z_1, \dots, Z_r)$ . By the Cramér–Wold device, it suffices to show that, for any constants  $\gamma_1, \dots, \gamma_r$  with  $\sum_i \gamma_i^2 = 1$ ,

$$\sum_{i=1}^r \gamma_i Z_i \xrightarrow{d} N(0, \sigma^2).$$

Indeed, since the columns of  $A$  are orthogonal,  $\hat{\beta}_i = Y_i$  for  $1 \leq i \leq s$  and so  $Z_i$  is a linear combination of  $\hat{\beta}_1, \dots, \hat{\beta}_s$ . But then so is  $\sum_i \gamma_i Z_i$  and asymptotic normality follows from the argument for  $\hat{\theta}$  of the form (13.14). ■

**Example 13.2.4 (Test of Homogeneity)** Let  $X_{i,j}$  ( $j = 1, \dots, n_i; i = 1, \dots, s$ ) be independently distributed as  $N(\mu_i, \sigma^2)$ . The problem is to test the null hypothesis

$$H : \mu_1 = \dots = \mu_s .$$

In this case, the test (13.26) is UMP invariant and reduces to

$$W^* = \frac{\sum n_i (X_{i.} - X_{..})^2 / (s - 1)}{\sum \sum (X_{i,j} - X_{i.})^2 / (n - s)}, \tag{13.29}$$

where

$$X_{i.} = \sum_j X_{i,j} / n_i, \quad X_{..} = \sum_i \sum_j X_{i,j} / n$$

and  $n = \sum_i n_i$ . If instead of  $X_{i,j}$  being  $N(\mu_i, \sigma^2)$ , assume that  $X_{i,j}$  has a distribution  $F(x - \mu_i)$ , where  $F$  is an arbitrary distribution with finite variance. Then, the theorem implies that, if  $\min_i n_i \rightarrow \infty$ , then the rejection probability tends to  $\alpha$ . In fact, the distributions may even vary within each sample, but it is important that the different samples have a common variance or the result fails; see Problems 13.24 and 13.25. ■

### 13.3 Edgeworth Expansions

Suppose  $X_1, \dots, X_n$  are i.i.d. with c.d.f.  $F$ . Let  $\mu(F)$  denote the mean of  $F$ , and consider the problem of testing  $\mu(F) = 0$ . As in Section 13.2.1, let  $\alpha_n(F)$  denote the actual rejection probability of the one-sided  $t$ -test under  $F$ . It was seen that the  $t$ -test is pointwise consistent in level in the sense that  $\alpha_n(F) \rightarrow \alpha$  whenever  $F$  has a

finite nonzero variance  $\sigma^2(F)$ . We shall now examine the rate at which the difference  $\alpha_n(F) - \alpha$  tends to 0.

In order to study this problem, we will consider expansions of the distribution function of the sample mean, as well as its studentized version. Such expansions are known as *Edgeworth expansions*. Let  $\Phi(\cdot)$  denote the standard normal c.d.f. and  $\varphi(\cdot)$  the standard normal density. Also let

$$\gamma = \gamma(F) = \frac{E_F[(X_i - \mu(F))^3]}{\sigma^3(F)}$$

and

$$\kappa = \kappa(F) = \frac{E_F[X_i - \mu(F)]^4]}{\sigma^4(F)} - 3.$$

The values  $\gamma$  and  $\kappa$  are known as the *skewness* and *kurtosis* of  $F$ , respectively.

**Theorem 13.3.1** *Assume  $E_F(|X_i|^{k+2}) < \infty$ . Let  $\psi_F$  denote the characteristic function of  $F$ , and assume*

$$\limsup_{|s| \rightarrow \infty} |\psi_F(s)| < 1. \tag{13.30}$$

Then,

$$P_F\left\{\frac{n^{1/2}[\bar{X}_n - \mu(F)]}{\sigma(F)} \leq x\right\} = \Phi(x) + \sum_{j=1}^k n^{-j/2} \varphi(x) p_j(x, F) + r_n(x, F), \tag{13.31}$$

where  $r_n(x, F) = o(n^{-k/2})$  and  $p_j(x, F)$  is a polynomial in  $x$  of degree  $3j - 1$  which depends on  $F$  through its first  $j + 2$  moments. In particular,

$$p_1(x, F) = -\frac{1}{6} \gamma(x^2 - 1), \tag{13.32}$$

and

$$p_2(x, F) = -x \left[ \frac{1}{24} \kappa(x^2 - 3) + \frac{1}{72} \gamma^2(x^4 - 10x^2 + 15) \right]. \tag{13.33}$$

Moreover, the expansion holds uniformly in  $x$  in the sense that, for fixed  $F$ ,

$$n^{k/2} \sup_x |r_n(x, F)| \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Assumption (13.30) is known as *Cramér's condition* and can be viewed as a smoothness assumption on  $F$ . It holds, for example, if  $F$  is absolutely continuous (or more generally is nonsingular) but fails if  $F$  is a lattice distribution, i.e.,  $X_1$  can only take on values of the form  $a + jb$  for some fixed  $a$  and  $b$  as  $j$  varies through the integers. A proof of Theorem 13.3.1 can be found in Feller (1971, Section XVI.4)

or Bhattacharya and Rao (1976), who also provide formulae for the  $p_j(x, F)$  when  $j > 2$ . The proofs hinge on expansions of characteristic functions.

Note that the term of order  $n^{-1/2}$  is zero if and only if the underlying skewness  $\gamma(F)$  is zero. This shows that the dominant error in using a standard normal approximation to the distribution of the standardized sample mean is due to skewness of the underlying distribution. Expansions such as these hold for many classes of statistics and provide more information than a weak convergence result, such as that provided by the Central Limit Theorem. As an example, the following result provides an Edgeworth expansion for the studentized sample mean. Let  $S_n^2 = \sum_i (X_i - \bar{X}_n)^2 / (n - 1)$ .

**Theorem 13.3.2** *Assume  $E_F(|X_i|^{k+2}) < \infty$  and that  $F$  is absolutely continuous.<sup>4</sup> Then, uniformly in  $t$ ,*

$$P_F\left\{\frac{n^{1/2}[\bar{X}_n - \mu(F)]}{S_n} \leq t\right\} = \Phi(t) + \sum_{j=1}^k n^{-j/2} \varphi(t) q_j(t, F) + \bar{r}_n(t, F), \quad (13.34)$$

where  $n^{k/2} \sup_t |\bar{r}_n(t, F)| \rightarrow 0$  and  $q_j(t, F)$  is a polynomial which depends on  $F$  through its first  $j + 2$  moments. In particular,

$$q_1(t, F) = \frac{1}{6} \gamma(2t^2 + 1), \quad (13.35)$$

and

$$q_2(t, F) = t \left[ \frac{1}{12} \kappa(t^2 - 3) - \frac{1}{18} \gamma^2(t^4 + 2t^2 - 3) - \frac{1}{4}(t^2 + 1) \right]. \quad (13.36)$$

Note that some authors prefer to provide an Edgeworth expansion as in (13.34), except that  $S_n^2$  replaces its denominator  $n - 1$  with  $n$ ; then,  $q_2(t, F)$  would have to be modified as well.

**Example 13.3.1 (Expansion for the  $t$ -distribution)** Suppose  $F$  is normal  $N(\mu, \sigma^2)$ . Let  $t_n = n^{1/2}(\bar{X}_n - \mu)/S_n$ . Then,  $\gamma(F) = \kappa(F) = 0$ . By Theorem 13.3.2,

$$P_F\{t_n \leq t\} = \Phi(t) - \frac{1}{4n}(t + t^3)\varphi(t) + o(n^{-1}). \quad (13.37)$$

This result implies a corresponding expansion for the quantiles of the  $t$ -distribution, known as a *Cornish–Fisher expansion*. Specifically, let  $t = t_{n-1, 1-\alpha}$  be the  $1 - \alpha$  quantile of the  $t$ -distribution with  $n - 1$  degrees of freedom. We would like to determine  $c = c_{1-\alpha}$  such that

$$t_{n-1, 1-\alpha} = z_{1-\alpha} + \frac{c_{1-\alpha}}{n} + o(n^{-1}).$$

---

<sup>4</sup> Alternatively, one can assume  $E_F(|X_i|^{2j+2}) < \infty$  and the distribution of  $(X_i, X_i^2)$  satisfies the multivariate analogue of Cramér’s condition; see Hall (1992), Chapter 2.

When  $t = t_{n-1, 1-\alpha}$ , the left side of (13.37) is  $1 - \alpha$  and the right side is by a Taylor expansion,

$$\Phi(z) + \frac{c}{n}\varphi(z) - \frac{1}{4n}(z + z^3)\varphi(z) + o(n^{-1}),$$

where  $z = z_{1-\alpha}$ . Since  $\Phi(z) = 1 - \alpha$ , we must have

$$\frac{c}{n}\varphi(z) - \frac{1}{4n}(z + z^3)\varphi(z) = o(n^{-1})$$

so that

$$c = c_{1-\alpha} = \frac{1}{4}z_{1-\alpha}(1 + z_{1-\alpha}^2).$$

Therefore,

$$n(t_{n-1, 1-\alpha} - z_{1-\alpha}) \rightarrow \frac{1}{4}z_{1-\alpha}(1 + z_{1-\alpha}^2). \blacksquare \quad (13.38)$$

In Section 13.2.1, we showed that the  $t$ -test has error in rejection probability tending to 0 as long as the underlying distribution has a finite nonzero variance. We will now make use of Edgeworth expansions in order to determine the orders of error in rejection probability for tests of the mean. All tests considered are based on the  $t$ -statistic  $t_n$ . In order to study this problem, we consider three factors: the one-sided case which rejects for large  $t_n$  versus the two-sided case which rejects for large  $|t_n|$ ; the use of a normal critical value versus a  $t$  critical value; and the dependence on  $F$ , especially whether  $\gamma(F)$  is 0 or not. For  $j = 1, 2$ , let  $\alpha_{n,j}^z(F)$  denote the error in rejection probability under  $F$  of the  $j$ -sided test using the normal quantile, and let  $\alpha_{n,j}^t(F)$  denote the analogous quantity using the appropriate  $t$ -quantile. For example,

$$\alpha_{n,2}^t(F) = P_F\{|t_n| \geq t_{n-1, 1-\frac{\alpha}{2}}\}.$$

We assume  $E_F(X_i^4) < \infty$  and that  $F$  is absolutely continuous so that we can apply the Edgeworth expansions in Theorems 13.3.1 and 13.3.2 with  $k = 2$ .

*The One-sided Case.* First, consider the test using the normal quantile. By (13.34),

$$\alpha_{n,1}^z(F) - \alpha = n^{-1/2}\varphi(z_{1-\alpha})q_1(z_{1-\alpha}, F) + n^{-1}\varphi(z_{1-\alpha})q_2(z_{1-\alpha}, F) + o(n^{-1}).$$

It follows that

$$\alpha_{n,1}^z(F) - \alpha = O(n^{-1/2}).$$

However, if  $\gamma(F) = 0$ , then  $q_1(z_{1-\alpha}, F) = 0$  and so

$$\alpha_{n,1}^z(F) - \alpha = O(n^{-1})$$

in this case. Using the  $t$ -quantiles instead of the normal quantiles yields

$$\alpha_{n,1}^t(F) - \alpha = \Phi(t_{n-1,\alpha}) - \alpha + n^{-1/2}\varphi(t_{n-1,1-\alpha})q_1(t_{n-1,1-\alpha}, F) + O(n^{-1}).$$

Then, applying (13.38),  $t_{n-1,1-\alpha} - z_{1-\alpha} = O(n^{-1})$ , so that a Taylor's expansion yields

$$\alpha_{n,1}^t(F) - \alpha = n^{-1/2}\varphi(z_{1-\alpha})q_1(z_{1-\alpha}, F) + O(n^{-1}).$$

Therefore,

$$\alpha_{n,1}^t(F) - \alpha = O(n^{-1/2}),$$

but the error in rejection probability is  $O(n^{-1})$  if  $\gamma(F) = 0$ .

*The Two-sided Case.* Let  $z = z_{1-\frac{\alpha}{2}}$ . Then, using the fact that  $\varphi(z) = \varphi(-z)$ ,

$$\begin{aligned}\alpha_{n,2}^z(F) &= P_F\{|t_n| \geq z\} = 1 - [P_F\{t_n \leq z\} - P_F\{t_n \leq -z\}] \\ &= \alpha + n^{-1/2}\varphi(z)[q_1(z, F) - q_1(-z, F)] + O(n^{-1}).\end{aligned}$$

But,  $q_1(\cdot, F)$  is an even function, which implies

$$\alpha_{n,2}^z(F) - \alpha = O(n^{-1}),$$

even if  $\gamma(F)$  is not zero. Similarly, it can be shown that (Problem 13.30)

$$\alpha_{n,2}^t(F) - \alpha = O(n^{-1}). \quad (13.39)$$

## 13.4 Nonparametric Inference for the Mean

### 13.4.1 Uniform Behavior of $t$ -test

It was seen in Section 13.2.1 that the classical  $t$ -test of the mean is asymptotically pointwise consistent in level for the class  $\mathbf{F}$  of all distributions with finite nonzero variance. In Section 13.3, the orders of error in rejection probability were obtained for a given  $F$ . However, these results are not reassuring unless the convergence is uniform in  $F$ . If it is not, then for any  $n$ , no matter how large, there will exist  $F$  in  $\mathbf{F}$  for which the rejection probability under  $F$ ,  $\alpha_n(F)$ , is not even close to  $\alpha$ . We shall show below that the convergence is not uniform and that the situation is even worse than what this negative result suggests. Namely, we shall show that for any  $n$ , there exist distributions  $F$  for which  $\alpha_n(F)$  is arbitrarily close to 1; that is, the size of the  $t$ -test is 1.

Suppose  $X_1, \dots, X_n$  are i.i.d. real-valued random variables with unknown c.d.f.  $F \in \mathbf{F}$ , where  $\mathbf{F}$  is a large nonparametric class of distributions. Let  $\mu(F)$  denote

the mean of  $F$  and  $\sigma^2(F)$  the variance of  $F$ . The goal is to test the null hypothesis  $\mu(F) = 0$  versus  $\mu(F) > 0$ , or perhaps the two-sided alternative  $\mu(F) \neq 0$ .

**Theorem 13.4.1** *For every  $n$ , the size of the  $t$ -test is 1 for the family  $\mathbf{F}_0$  of all distributions with finite variance.*

PROOF. Let  $c$  be an arbitrary positive constant less than one and let  $p_n = 1 - c^{1/n}$  so that  $(1 - p_n)^n = c$ . Let  $F = F_{n,c}$  be the distribution that places mass  $1 - p_n$  at  $p_n$  and mass  $p_n$  at  $p_n - 1$ , so that  $\mu(F) = 0$ . With probability  $c$ , we have all observations equal to  $p_n$ . For such a sample, the numerator  $n^{1/2}\bar{X}_n$  of the  $t$ -statistic is  $n^{1/2}p_n > 0$  while the denominator is 0. Thus, the  $t$ -statistic blows up and the hypothesis will be rejected. The probability of rejection is therefore  $\geq c$ , and by taking  $c$  arbitrarily close to 1 the theorem is proved. (Note that one can modify the distributions  $F_{n,c}$  used in the proof to be continuous rather than discrete.) ■

It follows that the  $t$ -test is not even uniformly asymptotically level  $\alpha$  for the family  $\mathbf{F}_0$ .

Instead of  $\mathbf{F}_0$ , one may wish to consider the behavior of the  $t$ -test against other nonparametric families. If  $\mathbf{F}_2$  is the family of all symmetric distributions with finite variance, it turns out that the  $t$ -test is still not uniformly level  $\alpha$ , and this is true even if the symmetric distributions have their support on  $(-1, 1)$  or any other fixed compact set; see Romano (2004). In fact, the size of the  $t$ -test under symmetry is one for moderate values of  $\alpha$ ; see Basu and DasGupta (1995). However, it can be shown that the size of the  $t$ -test is bounded away from 1 for small values of  $\alpha$ , by a result of Edelman (1990). Basu and DasGupta (1995) also show that if  $\mathbf{F}_3$  is the family of all symmetric unimodal distributions (with no moment restrictions), then the largest rejection probability under  $F$  of the  $t$ -test occurs when  $F$  is uniform on  $[-1, 1]$ , at least in the case of very small  $\alpha$ .

On the other hand, we will now show that the  $t$ -test is uniformly consistent over certain large subfamilies of distributions with two finite moments. For this purpose, consider a family of distributions  $\tilde{\mathbf{F}}$  on the real line satisfying

$$\lim_{\lambda \rightarrow \infty} \sup_{F \in \tilde{\mathbf{F}}} E_F \left[ \frac{|X - \mu(F)|^2}{\sigma^2(F)} I \left\{ \frac{|X - \mu(F)|}{\sigma(F)} > \lambda \right\} \right] = 0. \quad (13.40)$$

For example, for any  $\epsilon > 0$  and  $b > 0$ , let  $\mathbf{F}_b^{2+\epsilon}$  be the set of distributions satisfying

$$E_F \left[ \frac{|X - \mu(F)|^{2+\epsilon}}{\sigma^{2+\epsilon}(F)} \right] \leq b.$$

Then,  $\tilde{\mathbf{F}} = \mathbf{F}_b^{2+\epsilon}$  satisfies (13.40). To see why, take expectations of both sides of the inequality

$$\lambda^\epsilon Y^2 I\{|Y| > \lambda\} \leq |Y|^{2+\epsilon}.$$



**Lemma 13.4.1** Suppose  $X_{n,1}, \dots, X_{n,n}$  are i.i.d.  $F_n$  with  $F_n \in \tilde{\mathbf{F}}$ , where  $\tilde{\mathbf{F}}$  satisfies (13.40). Let  $\bar{X}_n = \sum_{i=1}^n X_{n,i}/n$ . Then, under  $F_n$ ,

$$\frac{n^{1/2}[\bar{X}_n - \mu(F_n)]}{\sigma(F_n)} \xrightarrow{d} N(0, 1).$$

PROOF. Let  $Y_{n,i} = [X_{n,i} - \mu(F_n)]/\sigma(F_n)$ . We verify the Lindeberg Condition (11.11), which in the case of  $n$  i.i.d. variables reduces to showing

$$\limsup_n E[Y_{n,i}^2 I\{|Y_{n,i}| > \epsilon n^{1/2}\}] = 0$$

for every  $\epsilon > 0$ . But, for every  $\lambda > 0$ ,

$$\limsup_n E[Y_{n,i}^2 I\{|Y_{n,i}| > \epsilon n^{1/2}\}] \leq \limsup_n E[Y_{n,i}^2 I\{|Y_{n,i}| > \lambda\}].$$

Let  $\lambda \rightarrow \infty$  and the right side tends to zero. ■

**Lemma 13.4.2** Let  $Y_{n,1}, \dots, Y_{n,n}$  be i.i.d. with c.d.f.  $G_n$  and finite mean  $\mu(G_n)$  satisfying

$$\lim_{\beta \rightarrow \infty} \limsup_{n \rightarrow \infty} E_{G_n} [ |Y_{n,i} - \mu(G_n)| I\{|Y_{n,i} - \mu(G_n)| \geq \beta\} ] = 0. \quad (13.41)$$

Let  $\bar{Y}_n = \sum_{i=1}^n Y_{n,i}/n$ . Then, under  $G_n$ ,  $\bar{Y}_n - \mu(G_n) \rightarrow 0$  in probability.

PROOF. Without loss of generality, assume  $\mu(G_n) = 0$ . Define

$$Z_{n,i} = Y_{n,i} I\{|Y_{n,i}| \leq n\}.$$

Let  $m_n = E(Z_{n,i})$  and  $\bar{Z}_n = \sum_{i=1}^n Z_{n,i}/n$ . Then, the event  $\{|\bar{Y}_n - m_n| > \epsilon\}$  implies either  $\{|\bar{Z}_n - m_n| > \epsilon\}$  occurs or  $\{\bar{Y}_n \neq \bar{Z}_n\}$  occurs. Hence, for any  $\epsilon > 0$ ,

$$P\{|\bar{Y}_n - m_n| > \epsilon\} \leq P\{|\bar{Z}_n - m_n| > \epsilon\} + P\{\bar{Y}_n \neq \bar{Z}_n\}. \quad (13.42)$$

The last term is bounded above by

$$P\left\{\bigcup_{i=1}^n \{Y_{n,i} \neq Z_{n,i}\}\right\} \leq \sum_{i=1}^n P\{Y_{n,i} \neq Z_{n,i}\} = nP\{|Y_{n,1}| > n\}.$$

The first term on the right side of (13.42) can be bounded by Chebyshev's inequality, so that

$$P\{|\bar{Z}_n - m_n| > \epsilon\} \leq (n\epsilon^2)^{-1} E(Z_{n,1}^2) + nP\{|Y_{n,1}| > n\}. \quad (13.43)$$

For  $t > 0$ , let

$$\tau_n(t) = t[1 - G_n(t) + G_n(-t)]$$

and

$$\kappa_n(t) = \frac{1}{t} \int_{-t}^t x^2 dG_n(t) = -\tau_n(t) + \frac{2}{t} \int_0^t \tau_n(x) dx ; \quad (13.44)$$

the last equality follows by integration by parts (Problem 13.37) and corrects (7.7), p. 235 of Feller (1971). Hence,

$$P\{|\bar{Y}_n - m_n| > \epsilon\} \leq \epsilon^{-2} \kappa_n(n) + \tau_n(n) . \quad (13.45)$$

But, for any  $t > 0$ ,

$$\tau_n(t) \leq E[|Y_{n,1}| I\{|Y_{n,1}| \geq t\}] ,$$

so  $\tau_n(n) \rightarrow 0$  by (13.41). Fix any  $\delta > 0$  and let  $\beta_0$  be such that

$$\limsup_n E[|Y_{n,1}| I\{|Y_{n,1}| > \beta_0\}] < \frac{\delta}{4} .$$

Then, there is an  $n_0$  such that, for all  $n \geq n_0$ ,

$$E[|Y_{n,1}| I\{|Y_{n,1}| > \beta_0\}] < \frac{\delta}{2} ,$$

and so

$$E|Y_{n,1}| \leq \beta_0 + \frac{\delta}{2}$$

for all  $n \geq n_0$  as well. Then, if  $n \geq n_0 > \beta_0$ ,

$$\begin{aligned} \frac{1}{n} \int_0^n \tau_n(x) dx &\leq \frac{1}{n} \int_0^n E[|Y_{n,1}| I\{|Y_{n,1}| \geq x\}] dx \\ &\leq \frac{1}{n} \int_0^{\beta_0} E|Y_{n,1}| dx + \frac{1}{n} \int_{\beta_0}^n \frac{\delta}{2} dx \leq \frac{\beta_0(\beta_0 + \frac{\delta}{2})}{n} + \frac{\delta}{2} , \end{aligned}$$

which is less than  $\delta$  for all sufficiently large  $n$ . Thus,  $\kappa_n(n) \rightarrow 0$  as  $n \rightarrow \infty$  and so (13.45) tends to 0 as well. Therefore,  $\bar{Y}_n - m_n \rightarrow 0$  in probability. Finally,  $m_n \rightarrow 0$ ; to see why, observe

$$0 = E(Y_{n,i}) = m_n + E[Y_{n,1} I\{|Y_{n,1}| > n\}] ,$$

so that

$$|m_n| \leq E[|Y_{n,1}| I\{|Y_{n,1}| > n\}] \rightarrow 0 ,$$

by assumption (13.41). ■

**Lemma 13.4.3** Let  $\tilde{\mathbf{F}}$  be a family of distributions satisfying (13.40). Suppose  $X_{n,1}, \dots, X_{n,n}$  are i.i.d.  $F_n \in \tilde{\mathbf{F}}$  and  $\mu(F_n) = 0$ . Then, under  $F_n$ ,

$$\frac{\frac{1}{n} \sum_{i=1}^n X_{n,i}^2}{\sigma^2(F_n)} \rightarrow 1 \quad \text{in probability.}$$

PROOF. Apply Lemma 13.4.2 to  $Y_{n,i} = [X_{n,i}^2/\sigma^2(F_n)] - 1$ . To see that Lemma 13.4.2 applies, note that if  $\beta > 1$ , then the event  $\{|Y_{n,i}| > \beta\}$  implies  $X_{n,i}^2/\sigma^2(F_n) > \beta + 1$  (since  $X_{n,i}^2/\sigma^2(F_n) > 0$ ) and also  $|Y_{n,i}| < X_{n,i}^2/\sigma^2(F_n)$ . Hence, for  $\beta > 1$ ,

$$E[|Y_{n,i}| I\{|Y_{n,i}| \geq \beta\}] \leq E\left[\frac{X_{n,i}^2}{\sigma^2(F_n)} I\left\{\frac{|X_{n,i}|}{\sigma(F_n)} > \sqrt{\beta + 1}\right\}\right].$$

The sup over  $n$  then tends to 0 as  $\beta \rightarrow \infty$  by the assumption  $F_n \in \tilde{\mathbf{F}}$ . ■

We are now in a position to study the behavior of the  $t$ -test uniformly across a fairly large class of distributions.

**Theorem 13.4.2** Let  $F_n \in \tilde{\mathbf{F}}$ , where  $\tilde{\mathbf{F}}$  satisfies (13.40). Assume

$$n^{1/2}\mu(F_n)/\sigma(F_n) \rightarrow \delta \quad \text{as } n \rightarrow \infty$$

(where  $|\delta|$  is allowed to be  $\infty$ ). Let  $X_1, \dots, X_n$  be i.i.d. with c.d.f.  $F_n$ , and consider the  $t$ -statistic

$$t_n = n^{1/2}\bar{X}_n/S_n,$$

where  $\bar{X}_n$  is the sample mean and  $S_n^2$  is the sample variance. If  $|\delta| < \infty$ , then under  $F_n$ ,

$$t_n \xrightarrow{d} N(\delta, 1).$$

If  $\delta \rightarrow \infty$  (respectively,  $-\infty$ ), then  $t_n \rightarrow \infty$  (respectively,  $-\infty$ ) in probability under  $F_n$ .

PROOF. Write

$$t_n = \frac{n^{1/2}[\bar{X}_n - \mu(F_n)]}{S_n} + \frac{n^{1/2}\mu(F_n)/\sigma(F_n)}{S_n/\sigma(F_n)}.$$

The proof will follow if we show  $S_n/\sigma(F_n) \rightarrow 1$  in probability under  $F_n$  and if

$$\frac{n^{1/2}[\bar{X}_n - \mu(F_n)]}{\sigma(F_n)} \xrightarrow{d} N(0, 1). \quad (13.46)$$

But the latter follows by Lemma 13.4.1. To show  $S_n^2/\sigma^2(F_n) \rightarrow 1$  in probability, use Lemma 13.4.3 (Problem 13.34). ■

Theorem 13.4.2 now allows us to deduce that the  $t$ -test is uniformly consistent in level, and it also yields a limiting power calculation.

**Theorem 13.4.3** *Let  $\tilde{\mathbf{F}}$  satisfy (13.40) and let  $\tilde{\mathbf{F}}_0$  be the set of  $F$  in  $\tilde{\mathbf{F}}$  with  $\mu(F) = 0$ . For testing  $\mu(F) = 0$  versus  $\mu(F) > 0$ , the  $t$ -test that rejects when  $t_n > z_{1-\alpha}$  (or  $t_{n-1, 1-\alpha}$ ) is uniformly asymptotically level  $\alpha$  over  $\tilde{\mathbf{F}}_0$ ; that is,*

$$|\sup_{F \in \tilde{\mathbf{F}}_0} P_F\{t_n > z_{1-\alpha}\} - \alpha| \rightarrow 0 \quad (13.47)$$

as  $n \rightarrow \infty$ . Also, the limiting power against  $F_n \in \tilde{\mathbf{F}}$  with  $n^{1/2}\mu(F_n)/\sigma(F_n) \rightarrow \delta$  is given by

$$\lim_n P_{F_n}\{t_n > z_{1-\alpha}\} = 1 - \Phi(z_{1-\alpha} - \delta). \quad (13.48)$$

Furthermore,

$$\inf_{\{F \in \tilde{\mathbf{F}}: n^{1/2}\mu(F)/\sigma(F) \geq \delta\}} P_F\{t_n > z_{1-\alpha}\} \rightarrow 1 - \Phi(z_{1-\alpha} - \delta). \quad (13.49)$$

PROOF. To prove (13.47), if the result failed, one could extract a subsequence  $\{F_n\}$  with  $F_n \in \tilde{\mathbf{F}}_0$  such that

$$P_{F_n}\{t_n > z_{1-\alpha}\} \rightarrow \beta \neq \alpha.$$

But this contradicts Theorem 13.4.2 since  $t_n$  is asymptotically standard normal under  $F_n$ . The proof of (13.48) follows from Theorem 13.4.2 as well. To prove (13.49), again argue by contradiction and assume there exists a subsequence  $\{F_n\}$  with  $n^{1/2}\mu(F_n)/\sigma(F_n) \geq \delta$  such that

$$P_{F_n}\{t_n > z_{1-\alpha}\} \rightarrow \gamma < 1 - \Phi(z_{1-\alpha} - \delta).$$

The result follows from (13.48) if  $n^{1/2}\mu(F_n)/\sigma(F_n)$  has a limit; otherwise, pass to any convergent subsequence and apply the same argument. ■

Note that (13.49) does not hold if  $\tilde{\mathbf{F}}$  is replaced by all distributions with finite second moments or finite fourth moments, or even the more restricted family of distributions supported on a compact set. In fact, there exists a sequence of distributions  $\{F_n\}$  supported on a fixed compact set and satisfying  $n^{1/2}\mu(F_n)/\sigma(F_n) \geq \delta$  such that the limiting power of the  $t$ -test against this sequence of alternatives is  $\alpha$ ; see Problem 13.38 for a construction. Nevertheless, the  $t$ -test behaves well for typical distributions, as demonstrated in Theorem 13.4.3. However, it is important to realize the  $t$ -test does not behave uniformly well across distributions with large skewness, as the limiting normal theory fails.

### 13.4.2 A Result of Bahadur and Savage

The negative results for the  $t$ -test under the families of all distributions with finite variance, or even the family of symmetric distributions with infinitely many moments are perhaps unexpected in view of the fact that the  $t$ -test is pointwise consistent in level for any distribution with finite (nonzero) variance, but they should not really be surprising. After all, the  $t$ -test was designed for the family of normal distributions and not for nonparametric families. This raises the question whether there do exist more satisfactory tests of the mean for nonparametric families.

For the family of distributions with finite variance and for some related families, this question was answered by Bahadur and Savage (1956). The desired results follow from the following basic lemma.

**Lemma 13.4.4** *Let  $\mathbf{F}$  be a family of distributions on  $\mathbb{R}$  satisfying:*

- (i) *For every  $F \in \mathbf{F}$ ,  $\mu(F)$  exists and is finite.*
- (ii) *For every real  $m$ , there is an  $F \in \mathbf{F}$  with  $\mu(F) = m$ .*
- (iii) *The family  $\mathbf{F}$  is convex in the sense that, if  $F_i \in \mathbf{F}$  and  $\gamma \in [0, 1]$ , then  $\gamma F_1 + (1 - \gamma)F_2 \in \mathbf{F}$ .*

*Let  $X_1, \dots, X_n$  be i.i.d.  $F \in \mathbf{F}$  and let  $\phi_n = \phi_n(X_1, \dots, X_n)$  be any test function. Let  $\mathbf{G}_m$  denote the set of distributions  $F \in \mathbf{F}$  with  $\mu(F) = m$ . Then,*

$$\inf_{F \in \mathbf{G}_m} E_F(\phi_n) \quad \text{and} \quad \sup_{F \in \mathbf{G}_m} E_F(\phi_n)$$

*are independent of  $m$ .*

PROOF. To show the result for the sup, fix  $m_0$  and let  $F_j \in \mathbf{G}_{m_0}$  be such that

$$\lim_j E_{F_j}(\phi_n) = \sup_{F \in \mathbf{G}_{m_0}} E_F(\phi_n) \equiv s.$$

Fix  $m_1$ . The goal is to show

$$\sup_{F \in \mathbf{G}_{m_1}} E_F(\phi_n) = s.$$

Let  $H_j$  be a distribution in  $\mathbf{F}$  with mean  $h_j$  satisfying

$$m_1 = \left(1 - \frac{1}{j}\right)m_0 + \frac{1}{j}h_j$$

and define

$$G_j = \left(1 - \frac{1}{j}\right)F_j + \frac{1}{j}H_j.$$

Thus,  $G_j \in \mathbf{G}_{m_1}$ . An observation from  $G_j$  can be obtained through a two-stage procedure. First, a coin is flipped with probability of heads  $1/j$ . If the outcome is a

head, then the observation has the distribution  $H_j$ ; otherwise, the observation is from  $F_j$ . So, with probability  $[1 - (1/j)]^n$ , a sample of size  $n$  from  $G_j$  is just a sample from  $F_j$ . Then,

$$\sup_{G \in \mathbf{G}_{m_1}} E_G(\phi_n) \geq E_{G_j}(\phi_n) \geq (1 - \frac{1}{j})^n E_{F_j}(\phi_n) \rightarrow s$$

as  $j \rightarrow \infty$ . Thus,

$$\sup_{G \in \mathbf{G}_{m_1}} E_G(\phi_n) \geq \sup_{G \in \mathbf{G}_{m_0}} E_G(\phi_n).$$

Interchanging the roles of  $m_0$  and  $m_1$  and applying the same argument makes the last inequality an equality. The result for the inf can be obtained by applying the argument to  $1 - \phi_n$ . ■

**Theorem 13.4.4** *Let  $\mathbf{F}$  satisfy (i)–(iii) of Lemma 13.4.4.*

(i) *Any test of  $H : \mu(F) = 0$  which has size  $\alpha$  for the family  $\mathbf{F}$  has power  $\leq \alpha$  for any alternative  $F$  in  $\mathbf{F}$ .*

(ii) *Any test of  $H : \mu(F) = 0$  which has power  $\beta$  against some alternative  $F$  in  $\mathbf{F}$  has size  $\geq \beta$ .*

Among the families satisfying (i)–(iii) of Lemma 13.4.4 is the family  $\mathbf{F}_0$  of distributions with finite second moment and that with infinitely many moments. Part (ii) of the above theorem provides an alternative proof of Theorem 13.4.1 since the power of the  $t$ -test against the normal alternatives  $N(\mu, 1)$  tends to 1 as  $\mu \rightarrow \infty$ . Theorem 13.4.4 now shows that the failure of the  $t$ -test for the family of all distributions with finite variance is not the fault of the  $t$ -test; in this setting, there exists no reasonable test of the mean. The reason is that slight changes in the tails of the distribution can result in enormous changes in the mean.

### 13.4.3 Alternative Tests

Another family satisfying conditions (i)–(iii) of Theorem 13.4.4 is the family of all distributions with compact support. However, the family of all distributions on a fixed compact set is excluded because it does not satisfy Condition (ii). In fact, the following construction due to Anderson (1967) shows that reasonable tests of the mean do exist if we assume the family of distributions is supported on a specified compact set. Specifically, let  $\mathbf{G}$  be the family of distributions supported on  $[-1, 1]$ , and let  $\mathbf{G}_0$  be the set of distributions on  $[-1, 1]$  having mean 0. We will exhibit a test that has size  $\alpha$  for any fixed sample size  $n$  and all  $F \in \mathbf{G}_0$ , and is pointwise consistent in power. First, recall the Kolmogorov–Smirnov confidence band  $R_{n,1-\alpha}$  given by (11.36). This leads to a conservative confidence interval  $I_{n,1-\alpha}$  for  $\mu(F)$  as follows. Include the value  $\mu$  in  $I_{n,1-\alpha}$  if and only if there exists some  $G$  in  $R_{n,1-\alpha}$  with  $\mu(G) = \mu$ . Then,

$$\{F \in R_{n,1-\alpha}\} \subseteq \{\mu(F) \in I_{n,1-\alpha}\}$$

and so

$$P_F\{\mu(F) \in I_{n,1-\alpha}\} \geq P_F\{F \in R_{n,1-\alpha}\} \geq 1 - \alpha ,$$

where the last inequality follows by construction of the Kolmogorov–Smirnov confidence bands. Finally, for testing  $\mu(F) = 0$  versus  $\mu(F) \neq 0$ , let  $\phi_n$  be the test that accepts the null hypothesis if and only if the value 0 falls in  $I_{n,1-\alpha}$ . By construction,

$$\sup_{F \in \mathbf{G}_0} E_F(\phi_n) \leq \alpha .$$

We claim that

$$I_{n,1-\alpha} \subseteq \bar{X}_n \pm 2n^{-1/2}s_{n,1-\alpha} , \tag{13.50}$$

where  $s_{n,1-\alpha}$  is the  $1 - \alpha$  quantile of the null distribution of the Kolmogorov–Smirnov test statistic. The result (13.50) follows from the following lemma.

**Lemma 13.4.5** *Suppose  $F$  and  $G$  are distributions on  $[-1, 1]$  with*

$$\sup_t |F(t) - G(t)| \leq \epsilon .$$

*Then,  $|\mu(F) - \mu(G)| \leq 2\epsilon$ .*

For a proof, see Problem 13.35. The result (13.50) now follows by applying the lemma to  $F$  and the empirical cdf  $\hat{F}_n$ .

Let  $F$  be a distribution with mean  $\mu(F) \neq 0$ . Suppose without loss of generality that  $\mu(F) > 0$ . Also, let  $L_{n,1-\alpha}$  be the lower endpoint of the interval  $I_{n,1-\alpha}$ . Then,

$$E_F(\phi_n) \geq P_F\{L_{n,1-\alpha} > 0\} \geq P_F\{\bar{X}_n > 2n^{-1/2}s_{n,1-\alpha}\} \rightarrow 1 , \tag{13.51}$$

by Slutsky’s Theorem, since  $\bar{X}_n \rightarrow \mu(F) > 0$  and  $n^{-1/2}s_{n,1-\alpha} \rightarrow 0$ . Thus, the test is pointwise consistent in power against any distribution in  $\mathbf{G}$  having nonzero mean. In fact, if  $\{F_n\}$  is such that  $|n^{1/2}\mu(F_n)| \rightarrow \infty$ , then the limiting power against such a sequence is one (Problem 13.36).

While Anderson’s method controls the level and is pointwise consistent in power, it is not efficient; an efficient test construction which is of exact level  $\alpha$  can be based on the confidence interval construction of Romano and Wolf (2000).

Let us next consider the family of symmetric distributions. Here the mean coincides with the center of symmetry, and reasonable level  $\alpha$  tests for this center exist. They can, for example, be based on the signed ranks. The one-sample Wilcoxon test is an example, studied in Examples 12.3.6 and 14.3.11. A large family of randomization tests that control the level is discussed in Section 17.2.

Finally, we mention a quite different approach to the problem considered in this section concerning the validity of the  $t$ -test in a nonparametric setting. Originally,

the  $t$ -test was derived for testing the mean,  $\mu$ , on the basis of a sample  $X_1, \dots, X_n$  from  $N(\mu, \sigma^2)$ . But,  $\mu$  is not only the mean of the normal distribution but it is also, for example, its median. Instead of embedding the normal family in the family of all distributions with finite mean (and perhaps finite variance), we could obtain a different viewpoint by embedding it in the family of all continuous distributions  $F$ , and then test the hypothesis that the median of  $F$  is 0. A suitable test is then the sign test.

## 13.5 Testing Many Means: The Gaussian Sequence Model

In this section, the problem of testing many normal means is considered. Assume  $X_1, X_2, \dots, X_n$  are independent with  $X_i \sim N(\mu_i, 1)$ . The problem is to test the global null hypothesis  $H_0 : \mu_i = 0$  for  $i = 1, \dots, n$  against some class of alternatives. There is no UMP test, nor is there a UMPU test if  $n > 1$ . However, there do exist UMPI and maximin tests, which depend on the choice of group and the class of alternatives, respectively. Since the procedures depend on the choice of optimality criteria, we consider the high-dimensional situation where the number of parameters  $n$  tends to infinity. Such an approach clarifies the type of alternatives where the procedures offer good power. We first review the Chi-squared test, and then consider some alternatives.

### 13.5.1 Chi-Squared Test

Let  $T_n = \sum_{i=1}^n X_i^2$ . The problem is invariant with respect to the group of orthogonal transformations, resulting in the UMPI test that rejects when  $T_n > c_{n,1-\alpha}$ , where  $c_{n,1-\alpha}$  is the  $1 - \alpha$  quantile of the Chi-squared distribution with  $n$  degrees of freedom. In addition, for any fixed  $\delta > 0$ , this test is maximin against alternatives defined by

$$\omega_1 = \{(\mu_1, \dots, \mu_n) : \sum_{i=1}^n \mu_i^2 = \delta^2\}, \quad (13.52)$$

as well as

$$\omega_2 = \{(\mu_1, \dots, \mu_n) : \sum_{i=1}^n \mu_i^2 \geq \delta^2\}. \quad (13.53)$$

Moreover, the test maximizes average power with respect to the uniform distribution on  $\omega_1$ .

By Problem 11.13, we can calculate its limiting power against alternatives for which  $\delta_n^2/\sqrt{2n} \rightarrow h$ , where  $\delta_n^2 = \sum_{i=1}^n \mu_i^2$ . In particular, under such a sequence of alternatives,

$$P\{T_n > c_{n,1-\alpha}\} \rightarrow 1 - \Phi(z_{1-\alpha} - h).$$



Therefore,  $\delta_n^2/\sqrt{2n}$  must not tend to 0 in order to get the limiting power to exceed  $\alpha$ , i.e.,  $\delta_n^2$  must be of strict order  $\sqrt{n}$ . Note that if  $\delta_n^2/\sqrt{2n} \rightarrow \infty$ , then the limiting power tends to one; see Problem 13.39. Therefore, in the special case that  $\mu_i = \mu$  is constant and nonzero, then the power of the test tends to one. Or, if  $p$  of the  $n$  means has constant value  $\mu$ , then  $p$  must be of strict order  $\sqrt{n}$  in order to achieve nontrivial power. More generally, the Chi-squared test performs reasonably well when, roughly, there are many contributions from many of the  $\mu_i$ , resulting in a larger value of  $\delta_n^2$ .

On the other hand, the Chi-squared test’s ability to detect sparse alternatives where a great majority of the  $\mu_i$  are zero is poor. In the extreme case where only one  $\mu_i$  is nonzero and is equal to  $\mu$ , then  $\mu$  must be of strict order at least  $n^{1/4}$ . As we will soon see, there are tests that can detect a much smaller value of  $\mu$ .

### 13.5.2 Maximin Test for Sparse Alternatives

Fix  $\delta > 0$ . Consider the maximin test, not for alternatives (13.52) and (13.53), but for alternatives

$$\omega_3 = \{(\mu_1, \dots, \mu_n) : \text{exactly one } \mu_i = \delta \text{ and remaining } \mu_j = 0\} \tag{13.54}$$

or

$$\omega_4 = \{(\mu_1, \dots, \mu_n) : \max_i \mu_i \geq \delta\}. \tag{13.55}$$

In such a sparse setting where only one mean is nonzero, the problem is sometimes referred to as the problem of detecting the “needle” in a “haystack.” The least favorable distribution places equal mass on the  $n$  points in  $\omega_3$  and the maximin test rejects for large values of the (average) likelihood ratio  $L_n$  given by

$$L_n = \frac{1}{n} \sum_{i=1}^n \exp(\delta X_i - \frac{\delta^2}{2}); \tag{13.56}$$

see Problem 8.24.

The question we now address is the following. If one of the  $\mu_i = \delta$  and the remaining are zero (but it is not known for which  $i$  that  $\mu_i$  is the nonzero value), what is the order of the smallest value of  $\delta$  for which the test rejects  $H_0$  with probability tending to one. First, the following lemma is needed.

**Lemma 13.5.1** *Assume the above Gaussian setup. Fix  $r > 0$  and let*

$$\delta = \delta_n = \sqrt{2r \log n}. \tag{13.57}$$

Under  $H_0$  and  $r < 1$ ,

$$L_n \xrightarrow{P} 1. \quad (13.58)$$

PROOF. Note that  $L_n$  is an average of i.i.d. random variables with mean 1, so the result is somewhat expected. However,  $\text{Var}(L_n)$  need not tend to 0 (depending on the value of  $r$ ), and so a careful argument is required. A proof based on truncation is given in Problem 13.43, or one can apply the triangular array law of large numbers stated in Lemma 13.4.2; see Problem 13.44. ■

We now claim that if  $0 < r < 1$  in (13.57), then the limiting power of the optimal maximin test based on  $L_n$  against an alternative where exactly one  $X_i$  has mean  $\delta_n$  and the remaining have mean 0 tends to  $\alpha$ . In other words, the test is essentially no better than the randomized test which rejects with probability  $\alpha$ . Let  $d_{n,1-\alpha}$  be the  $1 - \alpha$  quantile of the distribution of  $L_n$  under  $H_0$ .

**Theorem 13.5.1** *Assume the above Gaussian setup with  $\mu_1 = \delta_n$  specified by (13.57) and the remaining  $\mu_i = 0$ . If  $0 < r < 1$ , then*

$$P\{\text{reject } H_0\} = P\{L_n \geq d_{n,1-\alpha}\} \rightarrow \alpha.$$

PROOF. Let  $P_{n,1}$  denote the joint distribution of  $(X_1, \dots, X_n)$  specified by the mixture distribution where, with probability  $1/n$ , the mean vector has  $\delta_n$  in the  $i$ th component and 0 in all others. Also, let  $P_{n,0}$  denote the joint distribution when all  $X_i$  are i.i.d.  $N(0, 1)$ . Then, the likelihood ratio is  $dP_{n,1}/dP_{n,0} = L_n$ . Moreover, the power under  $P_{n,1}$  is the same as the power when  $\mu_1$  is the nonzero mean. But, the power under  $P_{n,1}$  is given by

$$\begin{aligned} 1 - P_{n,1}\{L_n < d_{n,1-\alpha}\} &= 1 - \int I\{L_n < d_{n,1-\alpha}\} \frac{dP_{n,1}}{dP_{n,0}} dP_{n,0} \\ &= 1 - \int I\{L_n < d_{n,1-\alpha}\} L_n dP_{n,0} \\ &= \alpha - \int I\{L_n < d_{n,1-\alpha}\} (L_n - 1) dP_{n,0}. \end{aligned} \quad (13.59)$$

By Lemma 13.5.1, under  $P_{n,0}$ ,  $L_n \xrightarrow{P} 1$ , so that

$$I\{L_n < d_{n,1-\alpha}\} (L_n - 1) \xrightarrow{P} 0. \quad (13.60)$$

Moreover,  $d_{n,1-\alpha}$  is bounded (by Problem 11.67), so that the left side of (13.60) is bounded. By bounded convergence, the last integral in (13.59) tends to 0. ■

Therefore, for testing against the alternatives  $\omega_3$  or  $\omega_4$ , no test can have better limiting maximin power than  $\alpha$  if  $\delta_n = \sqrt{2r \log n}$  and  $r < 1$ . We will soon see that

a test based on  $\max_i X_i$  does have limiting power one when  $r > 1$ , and therefore so must the maximin test based on  $L_n$ . But, to be clear, a test based on  $L_n$  must specify  $\delta_n$ , whereas the test based on  $\max_i X_i$  does not.

### 13.5.3 Test Based on Maximum and Bonferroni

As in the sparse setting of the previous section where one of the means is nonzero but positive, an intuitive test is one that rejects for large values of

$$M_n = \max(X_1, \dots, X_n).$$

(Or course, if the nonzero mean could also be negative, then one could based a test on  $|M_n|$ .) It is worth noting that, under  $H_0$ ,  $M_n$  has a limiting distribution; see Galambos (1977).

**Theorem 13.5.2** *Assume  $X_1, X_2, \dots$  are i.i.d.  $N(0, 1)$  and  $M_n = \max(X_1, \dots, X_n)$ . Then, for  $-\infty < t < \infty$ ,*

$$P \left\{ \sqrt{2 \log n} \left( M_n - \sqrt{2 \log n} + \frac{\log \log n + \log 4\pi}{2\sqrt{2 \log n}} \right) \leq t \right\} \rightarrow G(t) = e^{-e^{-t}},$$

where the c.d.f.  $G$  is the Gumbel distribution.

Let  $m_{n,1-\alpha}$  be the  $1 - \alpha$  quantile of the distribution of  $M_n$  under  $H_0$ . Then, it is easy to check (Problem 13.45) that

$$m_{n,1-\alpha} = z_{(1-\alpha)^{1/n}}. \tag{13.61}$$

Alternatively, an approximate conservative critical value may be used in place of  $m_{n,1-\alpha}$ . Under  $H_0$ , by Bonferroni,

$$P\{M_n \geq c\} = P\left\{\bigcup_{i=1}^n \{X_i \geq c\}\right\} \leq \sum_{i=1}^n P\{X_i \geq c\} = n[1 - \Phi(c)].$$

In order for the right-hand side to be no bigger than the nominal level  $\alpha$ , we should take  $c = z_{1-\frac{\alpha}{n}}$ . This method which reject when  $M_n \geq z_{1-\frac{\alpha}{n}}$  is then called the Bonferroni method. Note, by Problem 13.40,

$$z_{1-\frac{\alpha}{n}} \sim \sqrt{2 \log n},$$

where  $a_n \sim b_n$  means  $a_n/b_n \rightarrow 1$ . In fact,  $\sqrt{2 \log n}$  is, for any fixed  $\alpha \in (0, 1)$  an upper bound for all large  $n$  (even though this approximation does not depend on  $\alpha$ ).

An alternative description of the Bonferroni method is based on  $p$ -values computed from each  $X_i$ . To that end, when testing  $\mu_i = 0$  against a positive alternative based on  $X_i$ , the resulting  $p$ -value is  $\hat{p}_i = 1 - \Phi(X_i)$ . Then, the Bonferroni method described above is equivalent to the test that rejects  $H_i$  if  $\min \hat{p}_i \leq \alpha/n$  (Problem 13.46).

We now show that the Bonferroni test has power tending to one when one of the  $\mu_i$  is as large as  $\delta_n = \sqrt{2r \log n}$ , if  $r > 1$ . Hence, the same is true of the test that rejects when  $M_n \geq m_{n,1-\alpha}$  (as well as the minimax test in Section 13.5.2).

**Theorem 13.5.3** *Assume the Gaussian setup where one of the  $\mu_i = \delta_n$  specified by (13.57) and the remaining  $\mu_i = 0$ . If  $r > 1$ , then the Bonferroni test has power satisfying*

$$P\{\text{reject } H_0\} = P\{M_n \geq z_{1-\frac{\alpha}{n}}\} \rightarrow 1 .$$

PROOF. Without loss of generality, assume  $\mu_1 = \delta_n$  and  $\mu_i = 0$  for  $i > 1$ . By Problem 13.40,  $z_{1-\frac{\alpha}{n}} \leq \sqrt{2 \log n}$  for sufficiently large  $n$ . Therefore, for sufficiently large  $n$ ,

$$P\{M_n \geq z_{1-\frac{\alpha}{n}}\} \geq P\{M_n \geq \sqrt{2 \log n}\} .$$

But,

$$P\{M_n \geq \sqrt{2 \log n}\} \geq P\{X_1 \geq \sqrt{2 \log n}\} =$$

$$1 - \Phi(\sqrt{2 \log n} - \delta_n) = 1 - \Phi((1 - \sqrt{r})\sqrt{2 \log n}) \rightarrow 1 . \blacksquare$$

To summarize, by Theorem 13.5.1, no level  $\alpha$  test can have minimum power tending to one against all alternatives in  $\omega_3$  (or  $\omega_4$ ), where at least one of the means is  $\delta_n = \sqrt{2r \log n}$  and  $r < 1$ . On the other hand, the Bonferroni test, or the test based on  $M_n$ , has power tending to one when  $r > 1$ . Therefore,  $\sqrt{2 \log n}$  can be viewed as a sharp threshold for detecting the nonzero mean.

### 13.5.4 Some Comparisons and the Higher Criticism

The following comparisons between the Chi-squared test and the Bonferroni test can be made. In order for the Chi-squared test to be powerful, the quantity  $\sum_i \mu_i^2 / \sqrt{2n}$  needs to be large. As an example, if all of the  $\mu_i$  are equal to  $\mu = Cn^{-1/4}$  with  $C$  large, then the Chi-squared test has large power. However, in this setting, the power of the Bonferroni test is poor. On the other hand, in the sparse setting where  $o(n^{1/2})$  of the means are as large as  $\sqrt{2r \log n}$  with  $r > 1$ , Bonferroni is powerful (as long as at least one of the means is  $\sqrt{2r \log n}$ ). But, the Chi-squared test has poor limiting power in this setting. In summary, we can say that roughly the Chi-squared test performs well for many (possibly) small effects while Bonferroni is better for a smaller number of large effects.

Fortunately, there exists a method that performs well in both settings, dating back to Tukey (1953). The test is based on Tukey’s Higher Criticism statistic, which we now motivate. First, recall  $p$ -values  $\hat{p}_1, \dots, \hat{p}_n$ , where in the context of many means,  $\hat{p}_i = 1 - \Phi(X_i)$ . (Note that the Higher Criticism approach applies more generally whenever one has  $p$ -values that are i.i.d.  $U(0, 1)$  under a global null hypothesis  $H_0$ .) Let  $\hat{F}_n(\cdot)$  be the empirical distribution of the  $p$ -values, so that

$$\hat{F}_n(t) = \frac{1}{n} \sum_{i=1}^n I\{\hat{p}_i \leq t\}.$$

For a given level of significance  $\beta$  and under  $H_0$ ,  $n\hat{F}_n(\beta)$  is distributed as binomial based on  $n$  trials and success probability  $\beta$ . So,  $H_0$  can be rejected for large values of  $n\hat{F}_n(\beta)$ , or equivalently, large values of

$$\frac{\sqrt{n}[\hat{F}_n(\beta) - \beta]}{\sqrt{\beta(1 - \beta)}}.$$

Such a binomial test can be traced back to Clopper and Pearson (1934). But, rather than using a fixed pre-specified level of significance  $\beta$ , the Higher Criticism statistic rejects for large values of  $HC_n$  defined by

$$HC_n = \sup_{0 < \beta < \beta_0} \left[ \frac{\sqrt{n}[\hat{F}_n(\beta) - \beta]}{\sqrt{\beta(1 - \beta)}} \right],$$

where  $\beta_0$  is a tuning parameter. A value of  $\beta_0 = 0.5$  is suggested in Donoho and Jin (2015).

We now briefly describe the optimality of  $HC_n$ . As before,  $H_0$  specifies that  $X_1, \dots, X_n$  are i.i.d.  $N(0, 1)$ . The alternative  $H_1$  specifies a mixture model where the  $X_1, \dots, X_n$  are i.i.d. according to the mixture distribution  $(1 - \epsilon_n)N(0, 1) + \epsilon_n N(\delta, 1)$ . Let

$$\epsilon_n = n^{-\gamma} \quad \frac{1}{2} < \gamma < 1$$

and

$$\delta_n = \sqrt{2r \log n} \quad 0 < r < 1.$$

Note the needle in haystack problem essentially corresponds to  $\gamma = 1$  and  $r = 1$  while the many small effects case corresponds to  $\gamma = 1/2$ . Let

$$\rho^*(\gamma) = \begin{cases} \gamma - \frac{1}{2} & \text{for } \frac{1}{2} < \gamma \leq \frac{3}{4} \\ (1 - \sqrt{1 - \gamma})^2 & \text{for } \frac{3}{4} \leq \gamma \leq 1 \end{cases} \tag{13.62}$$

Ingster (1999) and Jin (2003) showed that for  $r > \rho^*(\gamma)$ , there exists a test sequence such that the probabilities of both Type 1 and Type 2 errors tend to 0.

On the other hand, they also showed that, for  $r < \rho^*(\gamma)$ , the limiting sum of the probabilities of Type 1 and Type 2 errors is bounded below by 1. Thus, the function  $\rho^*(\gamma)$  gives the threshold values of  $r$  for detecting  $H_1$ . Moreover, Donoho and Jin (2004) proved that, for  $r > \rho^*(\gamma)$ , the test based on  $HC_n$  is optimal in that a critical value can be chosen (such as  $\sqrt{2 \log \log n}$ ), so that the sum of error probabilities tends to 0. Importantly, the test does not require knowledge of  $\epsilon_n$  or  $\delta_n$ , or equivalently  $\gamma$  and  $r$ , and thus is optimal across a broad range of sparse alternatives.

## 13.6 Problems

### Section 13.2

**Problem 13.1** (i) Let  $X_1, \dots, X_n$  be a sample from  $N(\xi, \sigma^2)$ . For testing  $\xi = 0$  against  $\xi > 0$ , show that the power of the one-sided one-sample  $t$ -test against a sequence of alternatives  $N(\xi_n, \sigma^2)$  for which  $n^{1/2}\xi_n/\sigma \rightarrow \delta$  tends to  $1 - \Phi(z_{1-\alpha} - \delta)$ .

(ii) The result of (i) remains valid if  $X_1, \dots, X_n$  are a sample from any distribution with mean  $\xi$  and finite variance  $\sigma^2$ .

**Problem 13.2** Generalize the previous problem to the two-sample  $t$ -test.

**Problem 13.3** Let  $(Y_i, Z_i)$  be i.i.d. bivariate random vectors in the plane, with both  $Y_i$  and  $Z_i$  assumed to have finite nonzero variances. Let  $\mu_Y = E(Y_1)$  and  $\mu_Z = E(Z_1)$ , let  $\rho$  denote the correlation between  $Y_1$  and  $Z_1$ , and let  $\hat{\rho}_n$  denote the sample correlation, as defined in (11.29).

(i). Under the assumption  $\rho = 0$ , show directly (without appealing to Example 11.3.6) that  $n^{1/2}\hat{\rho}_n$  is asymptotically normal with mean 0 and variance

$$\tau^2 = \text{Var}[(Y_1 - \mu_Y)(Z_1 - \mu_Z)] / \text{Var}(Y_1)\text{Var}(Z_1).$$

(ii). For testing that  $Y_1$  and  $Z_1$  are independent, consider the test that rejects when  $n^{1/2}|\hat{\rho}_n| > z_{1-\alpha/2}$ . Show that the asymptotic rejection probability is  $\alpha$ , without assuming normality, but under the sole assumption that  $Y_1$  and  $Z_1$  have arbitrary distributions with finite nonzero variances.

(iii). However, for testing  $\rho = 0$ , the above test is not asymptotically robust. Show that there exist bivariate distributions for  $(Y_1, Z_1)$  for which  $\rho = 0$  but the limiting variance  $\tau^2$  can take on any given positive value.

(iv). For testing  $\rho = 0$  against  $\rho > 0$ , define a denominator  $D_n$  and a critical value  $c_n$  such that the rejection region  $n^{1/2}\hat{\rho}_n/D_n \geq c_n$  has probability tending to  $\alpha$ , under any bivariate distribution with  $\rho = 0$  and finite, nonzero marginal variances.

**Problem 13.4** Under the assumptions of Lemma 13.2.1, compute  $\text{Cov}(X_i^2, X_j^2)$  in terms of  $\rho_{i,j}$  and  $\sigma^2$ . Show that  $\text{Var}(n^{-1} \sum_{i=1}^n X_i^2) \rightarrow 0$  and hence  $n^{-1} \sum_{i=1}^n X_i^2 \xrightarrow{P} \sigma^2$ .

**Problem 13.5** (i) Given  $\rho$ , find the smallest and largest value of (13.2) as  $\sigma^2/\tau^2$  varies from 0 to  $\infty$ .

(ii) For nominal level  $\alpha = 0.05$  and  $\rho = 0.1, 0.2, 0.3, 0.4$ , determine the smallest and the largest asymptotic level of the  $t$ -test as  $\sigma^2/\tau^2$  varies from 0 to  $\infty$ .

**Problem 13.6** Verify the formula for  $Var(\bar{X})$  in Model A.

**Problem 13.7** In Model A, suppose that the number of observations in group  $i$  is  $n_i$ . if  $n_i \leq M$  and  $s \rightarrow \infty$  show that the assumptions of Lemma 13.2.1 are satisfied and determine  $\gamma$ .

**Problem 13.8** Show that the conditions of Lemma 13.2.1 are satisfied and  $\gamma$  has the stated value: (i) in Model B; (ii) in Model C.

**Problem 13.9** Determine the maximum asymptotic level of the one-sided  $t$ -test when  $\alpha = .05$  and  $m = 2, 4, 6$ : (i) in Model A; (ii) in Model B.

**Problem 13.10** Prove (i) of Lemma 13.2.2.

**Problem 13.11** Prove Lemma 13.2.3. *Hint:* For part (ii), use Problem 11.72.

**Problem 13.12** Verify the claims made in Example 13.2.1.

**Problem 13.13** Verify (13.15).

**Problem 13.14** In Example 13.2.3, verify the Huber Condition holds.

**Problem 13.15** Let  $X_{ijk}$  ( $k = 1, \dots, n_{ij}$ ;  $i = 1, \dots, a$ ;  $j = 1, \dots, b$ ) be independently normally distributed with mean  $E(X_{ijk}) = \xi_{ij}$  and variance  $\sigma^2$ . Then the test of any linear hypothesis concerning the  $\xi_{ij}$  has a robust level provided  $n_{ij} \rightarrow \infty$  for all  $i$  and  $j$ .

**Problem 13.16** In the two-way layout of the preceding problem give examples of submodels  $\Pi_{\Omega}^{(1)}$  and  $\Pi_{\Omega}^{(2)}$  of dimensions  $s_1$  and  $s_2$ , both less than  $ab$ , such that in one case Condition (13.20) continues to require  $n_{ij} \rightarrow \infty$  for all  $i$  and  $j$  but becomes a weaker requirement in the other case.

**Problem 13.17** Suppose (13.20) holds for some particular sequence  $\Pi_{\Omega}^{(n)}$  with fixed  $s$ . Then it holds for any sequence  $\Pi'_{\Omega}{}^{(n)} \subseteq \Pi_{\Omega}^{(n)}$  of dimension  $s' < s$ .

*Hint:* If  $\Pi_{\Omega}$  is spanned by the  $s$  columns of  $A$ , let  $\Pi'_{\Omega}$  be spanned by the first  $s'$  columns of  $A$ .

**Problem 13.18** Show that (13.10) holds whenever  $c_n$  tends to a finite nonzero limit, but the condition need not hold if  $c_n \rightarrow 0$ .

**Problem 13.19** Let  $\{c_n\}$  and  $\{c'_n\}$  be two increasing sequences of constants such that  $c'_n/c_n \rightarrow 1$  as  $n \rightarrow \infty$ . Then  $\{c_n\}$  satisfies (13.10) if and only if  $\{c'_n\}$  does.

**Problem 13.20** Let  $c_n = u_0 + u_1n + \dots + u_kn^k$ ,  $u_i \geq 0$  for all  $i$ . Then  $c_n$  satisfies (13.10). What if  $c_n = 2^n$ ? *Hint:* Apply Problem 13.19 with  $c'_n = n^k$ .

**Problem 13.21** If  $\xi_i = \alpha + \beta t_i + \gamma u_i$ , express Condition (13.20) in terms of the  $t$ 's and  $u$ 's.

**Problem 13.22** If  $\Pi_{i,i}$  are defined as in (13.19), show that  $\sum_{i=1}^n \Pi_{i,i}^2 = s$ .  
*Hint:* Since the  $\Pi_{i,i}$  are independent of  $A$ , take  $A$  to be orthogonal.

**Problem 13.23** The size of each of the following tests is robust against nonnormality:

1. the test (7.24) as  $b \rightarrow \infty$ ,
2. the test (7.26) as  $mb \rightarrow \infty$ ,
3. the test (7.28) as  $m \rightarrow \infty$ .

**Problem 13.24** For  $i = 1, \dots, s$  and  $j = 1, \dots, n_i$ , let  $X_{i,j}$  be independent, with  $X_{i,j}$  having distribution  $F_i$ , where  $F_i$  is an arbitrary distribution with mean  $\mu_i$  and finite common variance  $\sigma^2$ . Consider testing  $\mu_1 = \dots = \mu_s$  based on the test statistic (13.29), which is UMPI under normality. Show the test remains robust with respect to the rejection probability under  $H_0$  even if the  $F_i$  differ and are not normal.

**Problem 13.25** In the preceding problem, investigate the rejection probability when the  $F_i$  have different variances. Assume  $\min n_i \rightarrow \infty$  and  $n_i/n \rightarrow \rho_i$ .

**Problem 13.26** Show that the test derived in Problem 11.56 is not robust against nonnormality.

**Problem 13.27** Let  $X_1, \dots, X_n$  be a sample from  $N(\xi, \sigma^2)$ , and consider the UMP invariant level- $\alpha$  test of  $H : \xi/\sigma \leq \theta_0$  (Section 6.4). Let  $\alpha_n(F)$  be the actual significance level of this test when  $X_1, \dots, X_n$  is a sample from a distribution  $F$  with  $E(X_i) = \xi$ ,  $Var(X_i) = \sigma^2 < \infty$ . Then the relation  $\alpha_n(F) \rightarrow \alpha$  will not in general hold unless  $\theta_0 = 0$ . *Hint:* First find the limiting joint distribution of  $\sqrt{n}(\bar{X} - \xi)$  and  $\sqrt{n}(S^2 - \sigma^2)$ .

### Section 13.3

**Problem 13.28** When sampling from a normal distribution, one can derive an Edgeworth expansion for the  $t$ -statistic as follows. Suppose  $X_1, \dots, X_n$  are i.i.d.  $N(\mu, \sigma^2)$  and let  $t_n = n^{1/2}(\bar{X}_n - \mu)/S_n$ , where  $S_n^2$  is the usual unbiased estimate of  $\sigma^2$ . Let  $\Phi$  be the standard normal c.d.f. and let  $\Phi' = \varphi$ . Show

$$P\{t_n \leq t\} = \Phi(t) - \frac{1}{4n}(t + t^3)\varphi(t) + O(n^{-2}) \quad (13.63)$$

as follows. It suffices to let  $\mu = 0$  and  $\sigma = 1$ . By conditioning on  $S_n$ , we can write

$$P\{t_n \leq t\} = E\{\Phi[t(1 + S_n^2 - 1)^{1/2}]\}.$$



By Taylor expansion inside the expectation, along with moments of  $S_n^2$ , one can deduce (13.63).

**Problem 13.29** In Theorem 13.3.2, suppose  $S_n^2$  is defined with its denominator  $n - 1$  replaced by  $n$ . Derive the explicit form for  $q_2(t, F)$  in the corresponding Edgeworth expansion.

**Problem 13.30** Assuming  $F$  is absolutely continuous with 4 moments, verify (13.39).

**Problem 13.31** Let  $\phi_n$  be the classical  $t$ -test for testing the mean is zero versus the mean is positive, based on  $n$  i.i.d. observations from  $F$ . Consider the power of this test against the distribution  $N(\mu, 1)$ . Show the power tends to one as  $\mu \rightarrow \infty$ .

### Section 13.4

**Problem 13.32** In Lemma 13.4.2, show that Condition (13.41) can be replaced by the assumption that, for some  $\beta_n = o(n^{1/2})$ ,

$$\limsup_{n \rightarrow \infty} E_{G_n}[|Y_{n,i} - \mu(G_n)| I\{|Y_{n,i} - \mu(G_n)| \geq \beta_n\}] = 0.$$

Moreover, this condition only needs to hold if  $\beta_n = o(n)$  if it is also known that  $\sup_n E_{G_n}|Y_{n,i} - \mu(G_n)| < \infty$ .

**Problem 13.33** Suppose  $\mathbf{F}$  satisfies the conditions of Theorem 13.4.4. Assume there exists  $\phi_n$  such that

$$\sup_{F \in \mathbf{F}: \mu(F)=0} E_F(\phi_n) \rightarrow \alpha .$$

Show that

$$\limsup_n E_F(\phi_n) \leq \alpha$$

for every  $F \in \mathbf{F}$ .

**Problem 13.34** In the proof of Theorem 13.4.2, prove  $S_n/\sigma(F_n) \rightarrow 1$  in probability.

**Problem 13.35** Prove Lemma 13.4.5.

**Problem 13.36** Consider the problem of testing  $\mu(F) = 0$  versus  $\mu(F) \neq 0$ , for  $F \in \mathbf{F}_0$ , the class of distributions supported on  $[0, 1]$ . Let  $\phi_n$  be Anderson's test.

(i) If

$$|n^{1/2}\mu(F_n)| \geq \delta > 2s_{n,1-\alpha} ,$$

then show that

$$E_{F_n}(\phi_n) \geq 1 - \frac{1}{2(2s_{n,1-\alpha} - \delta)^2},$$

where  $s_{n,1-\alpha}$  is the  $1 - \alpha$  quantile of the null distribution of the Kolmogorov–Smirnov statistic. *Hint:* Use (13.51) and Chebyshev’s inequality.

(ii) Deduce that the minimum power of  $\phi_n$  over  $\{F : n^{1/2}|\mu(F)| \geq \delta\}$  is at least  $1 - [2(2s_{n,1-\alpha} - \delta)^{-2}]$  if  $\delta > 2s_{n,1-\alpha}$ .

(iii) Use (ii) to show that, if  $F_n \in \mathbf{F}_0$  is any sequence of distributions satisfying  $n^{1/2}|\mu(F_n)| \rightarrow \infty$ , then  $E_{F_n}(\phi_n) \rightarrow 1$ .

**Problem 13.37** Prove the second equality in (13.44). In the proof of Lemma 13.4.2, show that  $\kappa_n(n) \rightarrow 0$ .

**Problem 13.38** Let  $Y_{n,1}, \dots, Y_{n,n}$  be i.i.d. bernoulli variables with success probability  $p_n$ , where  $np_n = \lambda$  and  $\lambda^{1/2} = \delta$ . Let  $U_{n,1}, \dots, U_{n,n}$  be i.i.d. uniform variables on  $(-\tau_n, \tau_n)$ , where  $\tau_n^2 = 3p_n^2$ . Then, let  $X_{n,i} = Y_{n,i} + U_i$ , so that  $F_n$  is the distribution of  $X_{n,i}$ . (Note that  $n^{1/2}\mu(F_n)/\sigma(F_n) = \delta$ .)

(i) If  $t_n$  is the  $t$ -statistic, show that, under  $F_n$ ,  $t_n \xrightarrow{d} V^{1/2}$ , where  $V$  is Poisson with mean  $\delta^2$ , and so if  $z_{1-\alpha}$  is not an integer,

$$P_{F_n}\{t_n > t_{n-1,1-\alpha}\} \rightarrow P\{V^{1/2} > z_{1-\alpha}\}.$$

(ii) Show, for  $\alpha < 1/2$ , the limiting power of the  $t$ -test against  $F_n$  satisfies

$$P\{V^{1/2} > z_{1-\alpha}\} \leq 1 - P\{V = 0\} = \exp(-\delta^2).$$

This is strictly smaller than  $1 - \Phi(z_{1-\alpha} - \delta)$  if and only if

$$\Phi(z_{1-\alpha} - \delta) < \exp(-\delta^2).$$

Certainly, for small  $\delta$ , this inequality holds, since the left-hand side tends to  $1 - \alpha$  as  $\delta \rightarrow 0$  while the right-hand side tends to 1.

### Section 13.5

**Problem 13.39** For the Chi-squared test discussed in Section 13.5.1, assume that  $\delta_n^2/\sqrt{2n} \rightarrow \infty$ . Show that the limiting power of the Chi-squared test against such an alternative sequence tends to one.

**Problem 13.40** (i) If  $\phi(\cdot)$  denotes the standard normal density and  $Z \sim N(0, 1)$ , then for any  $t > 0$ ,

$$\left(\frac{1}{t} - \frac{1}{t^3}\right)\phi(t) < P\{Z \geq t\} \leq \frac{\phi(t)}{t}. \tag{13.64}$$

Prove the right-hand inequality.

(ii) Prove the left inequality in (13.64). *Hint:* Feller (1968) p.179 notes the negative of the derivative of the left side  $(\frac{1}{t} - \frac{1}{t^3})\phi(t)$  is equal to  $(1 - 3t^{-4})\phi(t)$ , which is certainly less than  $\phi(t)$ .

(iii) Use (13.64) to show that, for any fixed  $\alpha$ , any  $\delta > 0$ , and all large enough  $n$ :

$$\sqrt{(1 - \delta)2 \log n} \leq z_{1-\frac{\alpha}{n}} \leq \sqrt{2 \log n} . \tag{13.65}$$

**Problem 13.41** Let  $X_1, \dots, X_n$  be i.i.d.  $N(0, 1)$ . Let  $M_n = \max(X_1, \dots, X_n)$ .

(i) Show that  $P\{M_n \geq \sqrt{2 \log n}\} \rightarrow 0$ .

(ii) Compute the limit of  $P\{M_n \geq z_{1-\frac{\alpha}{n}}\}$ .

**Problem 13.42** Under the setting of Lemma 13.5.1 calculate  $Var(L_n)$  and determine which values of  $r$  it tends to 0.

**Problem 13.43** Prove Lemma 13.5.1 as follows. Let  $\eta = 1 - \sqrt{r}$ . Let

$$\tilde{L}_n = \frac{1}{n} \sum_{i=1}^n \exp(\delta_n X_i - \frac{\delta_n^2}{2}) I\{X_i \leq \sqrt{2 \log n}\} .$$

First, show  $L_n - \tilde{L}_n \xrightarrow{P} 0$  (using Problem 13.41). Then, show

$$E(\tilde{L}_n) = \Phi(\eta\sqrt{2 \log(n)}) \rightarrow 1 .$$

The proof then follows by showing  $Var(\tilde{L}_n) \rightarrow 0$ . To this end, show

$$\begin{aligned} Var(\tilde{L}_n) &\leq \frac{1}{n} E[X_i^2 I\{X_i \leq \sqrt{2 \log n}\}] = \frac{1}{n} \exp(\delta_n^2) \Phi((2\eta - 1)\sqrt{2 \log n}) \\ &\leq \frac{1}{n} \exp(\delta_n^2) \phi((1 - 2\eta)\sqrt{2 \log n}) = \frac{1}{\sqrt{2\pi}} \exp[-\eta^2 \log n] \rightarrow 0 . \end{aligned}$$

**Problem 13.44** Prove Lemma 13.5.1 by using Problem 13.32. That is, if  $1 < \beta_n = o(n)$  and

$$Y_{n,i} = \exp(\delta_n X_i - \frac{\delta_n^2}{2}) ,$$

show that

$$E[|Y_{n,i} - 1| I\{|Y_{n,i} - 1| > \beta_n\}] \rightarrow 0 . \tag{13.66}$$

Since  $Y_{n,i} > 0$  and  $\beta_n > 1$ , this is equivalent to showing

$$E[(Y_{n,i} - 1) I\{Y_{n,i} > \beta_n + 1\}] \rightarrow 0 . \tag{13.67}$$

The event  $\{Y_{n,i} > \lambda + 1\}$  is equivalent to  $\{X_i > b_n(\beta_n)\}$ , where

$$b_n(\beta_n) = \frac{\log(\beta_n + 1)}{\delta_n} + \frac{\delta_n}{2}.$$

Show the left side of (13.67) is equal to

$$\int_{b_n(\beta_n)}^{\infty} [\exp(\delta_n x - \frac{\delta_n^2}{2}) - 1] \phi(x) dx = \Phi(b_n(\beta_n)) - \Phi(b_n(\beta_n) - \delta_n),$$

and show this last expression tends to zero by appropriate choice of  $\beta_n$ .

**Problem 13.45** Prove (13.61).

**Problem 13.46** In the setting of Section 13.5.3, show that the Bonferroni test that rejects  $H_0$  when  $M_n \geq z_{1-\frac{\alpha}{n}}$  is equivalent to the test that rejects  $H_i$  if  $\min \hat{p}_i \leq \alpha/n$ , where  $\hat{p}_i = 1 - \Phi(X_i)$ .

## 13.7 Notes

Concern about the robustness of classical normal theory tests began to be voiced in the 1920s (Neyman and Pearson, 1928; Shewhart and Winters, 1928; Sophister, 1928; Pearson, 1929) and has been an important topic ever since. Particularly influential were Box (1953), where the term *robustness* was introduced; also see Scheffé (1959, 10), Tukey (1960) and Hotelling (1961). The robustness of regression tests studied in Section 13.2.3 is based on Huber (1973).

As remarked in Example 13.2.4, the  $F$ -test for testing equality of means is not robust if the underlying variances differ, even if the sample sizes are equal and  $s > 2$ ; see Scheffé (1959). More appropriate tests for this generalized Behrens–Fisher problem have been proposed by Welch (1951), James (1951), and Brown and Forsythe (1974b), and are further discussed by Clinch and Kesselman (1982). The corresponding robustness problem for more general linear hypotheses is treated by James (1954a, 1954b) and Johansen (1980); see also Rothenberg (1984).

The linear model  $F$ -test—as was seen to be the case for the  $t$ -test—is highly nonrobust against dependence of the observations. Tests of the hypothesis that the covariance matrix is proportional to the identity against various specified forms of dependence are considered in King and Hillier (1985). For recent work on robust testing in linear models, see Müller (1998) and the references cited there.

The usual test for equality of variances is Bartlett’s test, which is discussed in Cyr and Monoukian (1982) and Glaser (1982). Bartlett’s test is highly sensitive to the assumption of normality, and therefore is rarely appropriate. More robust tests for this latter hypothesis are reviewed in Conover et al. (1981). For testing homogeneity of covariance matrices, see Beran and Srivastava (1985) and Zhang and Boos (1992).

Robustness properties of the  $t$ -test are studied in Efron (1969), Lehmann and Loh (1990), Basu and DasGupta (1995), Basu (1999) and Romano (2004). The nonexistence results of Bahadur and Savage (1956), and also Hoeffding (1956),

have been generalized to other problems; see Donoho (1988) and Romano (2004) and the references there.

The idea of expanding the distribution of the sample mean in order to study the error in normal approximation can be traced to Chebyshev (1890) and Edgeworth (1905). But it was not until later that Cramér (1928, 1937) provided some rigorous results. The fundamental theory of Edgeworth expansions is developed in Bhattacharya and Rao (1978); also see Bickel (1974), Bhattacharya and Ghosh (1978), Hall (1992) and Hall and Jing (1995).

Section 13.5 was inspired by class notes of Emmanuel Candès. Much more general results are available in Arias-Castro et al. (2011). The “needles” in “haystack” problem is attributed to Johnstone and Silverman (2004). Much further discussion of the Higher Criticism can be found in Donoho and Jin (2004). Extensions to the sparse regression setting appear in Ingster and Tsybakov (2010).