CHAPTER 10

Some Further Results of Hypothesis Testing

10.1 INTRODUCTION

In this chapter we study some commonly used procedures in the theory of testing of hypotheses. In Section 10.2 we describe the classical procedure for constructing tests based on likelihood ratios. This method is sufficiently general to apply to multiparameter problems and is especially useful in the presence of *nuisance parameters*. These are unknown parameters in the model which are of no inferential interest. Most of the normal theory tests described in Sections 10.3 to 10.5 and those in Chapter 12 can be derived by using methods of Section 10.2. In Sections 10.3 to 10.5 we list some commonly used normal theory-based tests. In Section 10.3 we also deal with goodness-of-fit tests. In Section 10.6 we look at the hypothesis testing problem from a decision-theoretic viewpoint and describe Bayes and minimax tests.

10.2 GENERALIZED LIKELIHOOD RATIO TESTS

In Chapter 9 we saw that UMP tests do not exist for some problems of hypothesis testing. In was suggested that we restrict attention to smaller classes of tests and seek UMP tests in these subclasses or, alternatively, seek tests that are optimal against local alternatives. Unfortunately, some of the reductions suggested in Chapter 9, such as invariance, do not apply to all families of distributions.

In this section we consider a classical procedure for constructing tests that has some intuitive appeal and that frequently, though not necessarily, leads to optimal tests. Also, the procedure leads to tests that have some desirable large-sample properties.

Recall that for testing $H_0: \mathbf{X} \sim f_0$ against $H_1: \mathbf{X} \sim f_1$, the Neyman–Pearson MP test is based on the ratio $f_1(\mathbf{x})/f_0(\mathbf{x})$. If we interpret the numerator as the best possible explanation of $\mathbf{x}$ under H_1, and the denominator as the best possible explanation of $\mathbf{X}$ under H_0, it is reasonable to consider the ratio

$$r(\mathbf{x}) = \frac{\sup_{\boldsymbol{\theta}\in\Theta_1} L(\theta;\mathbf{x})}{\sup_{\boldsymbol{\theta}\in\Theta_0} L(\theta;\mathbf{x})} = \frac{\sup_{\boldsymbol{\theta}\in\Theta_1} f_\theta(\mathbf{x})}{\sup_{\boldsymbol{\theta}\in\Theta_0} f_\theta(\mathbf{x})}$$

as a test statistic for testing $H_0\colon \boldsymbol{\theta} \in \Theta_0$ against $H_1\colon \boldsymbol{\theta} \in \Theta_1$. Here $L(\boldsymbol{\theta};\mathbf{x})$ is the likelihood function of $\mathbf{X}$. Note that for each $\mathbf{x}$ for which the MLEs of $\boldsymbol{\theta}$ under Θ_1 and Θ_0, exist, the ratio is well defined and free of $\boldsymbol{\theta}$ and can be used as a test statistic. Clearly, we should reject H_0 if $r(\mathbf{x}) > c$.

The statistic r is hard to compute; only one of the two suprema in the ratio may be attained. Let $\boldsymbol{\theta} \in \Theta \subseteq \mathcal{R}_k$ be a vector of parameters, and let $\mathbf{X}$ be a random vector with PDF (PMF) $f_{\boldsymbol{\theta}}$. Consider the problem of testing the null hypothesis $H_0\colon \mathbf{X} \sim f_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta_0$ against the alternative $H_1\colon \mathbf{X} \sim f_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \Theta_1$.

Definition 1. For testing H_0 against H_1, a test of the form: reject H_0 if and only if $\lambda(\mathbf{x}) < c$, where c is a constant, and

$$\lambda(\mathbf{x}) = \frac{\sup_{\boldsymbol{\theta}\in\Theta_0} f_{\boldsymbol{\theta}}(x_1, x_2, \ldots, x_n)}{\sup_{\boldsymbol{\theta}\in\Theta} f_{\boldsymbol{\theta}}(x_1, x_2, \ldots, x_n)}$$

is called a *generalized likelihood ratio* (GLR) *test.*

We leave the reader to show that the statistics $\lambda(\mathbf{X})$ and $r(\mathbf{X})$ lead to the same criterion for rejecting H_0.

The numerator of the likelihood ratio λ is the best *explanation* of $\mathbf{X}$ (in the sense of maximum likelihood) that the null hypothesis H_0 can provide, and the denominator is the best possible explanation of $\mathbf{X}$. H_0 is rejected if there is a much better explanation of $\mathbf{X}$ than the best one provided by H_0.

It is clear that $0 \le \lambda \le 1$. The constant c is determined from the size restriction

$$\sup_{\boldsymbol{\theta}\in\Theta_0} P_{\boldsymbol{\theta}}\{\lambda(\mathbf{X}) < c\} = \alpha.$$

If the distribution of λ is continuous (that is, the DF is absolutely continuous), any size α is attainable. If, however, $\lambda(\mathbf{X})$ is a discrete RV, it may not be possible to find a likelihood ratio test whose size exactly equals α. This problem arises because of the nonrandomized nature of the likelihood ratio test and can be handled by randomization. The following result holds.

Theorem 1. If for given α, $0 \le \alpha \le 1$, nonrandomized Neyman–Pearson and likelihood ratio tests of a simple hypothesis against a simple alternative exist, they are equivalent.

The proof is left as an exercise.

Theorem 2. For testing $\boldsymbol{\theta} \in \Theta_0$ against $\boldsymbol{\theta} \in \Theta_1$, the likelihood ratio test is a function of every sufficient statistic for $\boldsymbol{\theta}$.

Theorem 2 follows from the factorization theorem for sufficient statistics.

Example 1. Let $X \sim b(n, p)$, and we seek a level α likelihood ratio test of $H_0: p \le p_0$ against $H_1: p > p_0$:

$$\lambda(x) = \frac{\sup\limits_{p \le p_0} \binom{n}{x} p^x (1-p)^{n-x}}{\sup\limits_{0 \le p \le 1} \binom{n}{x} p^x (1-p)^{n-x}}.$$

Now

$$\sup_{0 \le p \le 1} p^x (1-p)^{n-x} = \left(\frac{x}{n}\right)^x \left(1 - \frac{x}{n}\right)^{n-x}.$$

The function $p^x(1-p)^{n-x}$ first increases, then achieves its maximum at $p = x/n$, and finally decreases, so that

$$\sup_{p \le p_0} p^x (1-p)^{n-x} = \begin{cases} p_0^x (1-p_0)^{n-x} & \text{if } p_0 < \dfrac{x}{n}, \\ \left(\dfrac{x}{n}\right)^x \left(1 - \dfrac{x}{n}\right)^{n-x} & \text{if } \dfrac{x}{n} \le p_0. \end{cases}$$

It follows that

$$\lambda(x) = \begin{cases} \dfrac{p_0^x (1-p_0)^{n-x}}{(x/n)^x [1 - (x/n)]^{n-x}} & \text{if } p_0 < \dfrac{x}{n}, \\ 1 & \text{if } \dfrac{x}{n} \le p_0. \end{cases}$$

Note that $\lambda(x) \le 1$ for $np_0 < x$ and $\lambda(x) = 1$ if $x \le np_0$, and it follows that $\lambda(x)$ is a decreasing function of x. Thus $\lambda(x) < c$ if and only if $x > c'$, and the GLR test rejects H_0 if $x > c'$.

The GLR test is of the type obtained in Section 9.4 for families with an MLR except for the boundary $\lambda(x) = c$. In other words, if the size of the test happens to be exactly α, the likelihood ratio test is a UMP level α test. Since X is a discrete RV, however, to obtain size α may not be possible. We have

$$\alpha = \sup_{p \le p_0} P_p\{X > c'\} = P_{p_0}\{X > c'\}.$$

If such a c' does not exist, we choose an integer c' such that

$$P_{p_0}\{X > c'\} \le \alpha \quad \text{and} \quad P_{p_0}\{X > c' - 1\} > \alpha.$$

The situation in Example 1 is not unique. For a one-parameter exponential family it can be shown (Birkes [6]) that a GLR test of $H_0: \theta \le \theta_0$ against $H_1: \theta > \theta_0$ is UMP of its size. The result holds also for the dual $H_0': \theta \ge \theta_0$ and, in fact, for a much wider class of one-parameter family of distributions.

The GLR test is specially useful when $\boldsymbol{\theta}$ is a multiparameter and we wish to test hypothesis concerning one of the parameters. The remaining parameters act as nuisance parameters.

Example 2. Consider the problem of testing $\mu = \mu_0$ against $\mu \neq \mu_0$ in sampling from $\mathcal{N}(\mu, \sigma^2)$, where both μ and σ^2 are unknown. In this case $\Theta_0 = \{(\mu_0, \sigma^2): \sigma^2 > 0\}$ and $\Theta = \{(\mu, \sigma^2): -\infty < \mu < \infty, \sigma^2 > 0\}$. We write $\boldsymbol{\theta} = (\mu, \sigma^2)$:

$$\begin{aligned}\sup_{\boldsymbol{\theta}\in\Theta_0} f_{\boldsymbol{\theta}}(\mathbf{x}) &= \sup_{\sigma^2>0}\left\{\frac{1}{(\sigma\sqrt{2\pi})^n}\exp\left[-\frac{\sum_1^n (x_i-\mu_0)^2}{2\sigma^2}\right]\right\}\\ &= f_{\hat{\sigma}_0^2}(\mathbf{x}),\end{aligned}$$

where $\hat{\sigma}_0^2$ is the MLE, $\hat{\sigma}_0^2 = (1/n)\sum_{i=1}^n (x_i - \mu_0)^2$. Thus

$$\sup_{\boldsymbol{\theta}\Theta_0} f_{\boldsymbol{\theta}}(\mathbf{x}) = \frac{1}{(2\pi/n)^{n/2}\left[\sum_1^n (x_i-\mu_0)^2\right]^{n/2}} e^{-n/2}.$$

The MLE of $\boldsymbol{\theta} = (\mu, \sigma^2)$ when both μ and σ^2 are unknown is $(\sum_1^n x_i/n, \sum_1^n (x_i - \overline{x})^2/n)$. If follows that

$$\begin{aligned}\sup_{\boldsymbol{\theta}\in\Theta} f_{\boldsymbol{\theta}}(\mathbf{x}) &= \sup_{\mu,\sigma^2}\left\{\frac{1}{(\sigma\sqrt{2\pi})^n}\exp\left[-\frac{\sum_1^n (x_i-\mu)^2}{2\sigma^2}\right]\right\}\\ &= \frac{1}{(2\pi/n)^{n/2}\left[\sum_1^n (x_i-\overline{x})^2\right]^{n/2}} e^{-n/2}.\end{aligned}$$

Thus

$$\begin{aligned}\lambda(\mathbf{x}) &= \left[\frac{\sum_1^n (x_i-\overline{x})^2}{\sum_1^n (x_i-\mu_0)^2}\right]^{n/2}\\ &= \left\{\frac{1}{1 + [n(\overline{x}-\mu_0)^2/\sum_1^n (x_i-\overline{x})^2]}\right\}^{n/2}.\end{aligned}$$

The GLR test rejects H_0 if

$$\lambda(\mathbf{x}) < c,$$

and since $\lambda(\mathbf{x})$ is a decreasing function of $n(\overline{x}-\mu_0)^2/\sum_1^n n(x_i-\overline{x})^2$, we reject H_0 if

$$\left|\frac{\overline{x}-\mu_0}{\sqrt{\sum_1^n (x_i-\overline{x})^2}}\right| > c',$$

that is, if

$$\left|\frac{\sqrt{n}(\overline{x}-\mu_0)}{s}\right| > c'',$$

where $s^2 = (n-1)^{-1}\sum_1^n (x_i-\overline{x})^2$. The statistic

$$t(\mathbf{X}) = \frac{\sqrt{n}(\overline{X}-\mu_0)}{S}$$

has a t-distribution with $n-1$ d.f. Under $H_0\colon \mu=\mu_0$, $t(\mathbf{X})$ has a central $t(n-1)$ distribution, but under $H_1\colon \mu\neq\mu_0$, $t(\mathbf{X})$ has a noncentral t-distribution with $n-1$ d.f. and noncentrality parameter $\delta = (\mu-\mu_0)/\sigma$. We choose $c'' = t_{n-1,\alpha/2}$ in accordance with the distribution of $t(\mathbf{X})$ under H_0. Note that the two-sided t-test obtained here is UMP unbiased. Similarly, one can obtain one-sided t-tests also as likelihood ratio tests.

The computations in Example 2 could be slightly simplified by using Theorem 2. Indeed, $T(X) = (\overline{X}, S^2)$ is a minimal sufficient statistic for θ, and since $\overline{X}$ and S^2 are independent, the likelihood is the product of the PDFs of $\overline{X}$ and S^2. We note that $\overline{X}\sim\mathcal{N}(\mu,\sigma^2/n)$ and $S^2\sim[\sigma^2/(n-1)]\chi^2_{n-1}$. We leave it to the reader to carry out the details.

Example 3. Let $X_1, X_2, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$ be independent random samples from $\mathcal{N}(\mu_1,\sigma_1^2)$ and $\mathcal{N}(\mu_2,\sigma_2^2)$, respectively. We wish to test the null hypothesis $H_0\colon \sigma_1^2=\sigma_2^2$ against $H_1\colon \sigma_1^2\neq\sigma_2^2$. Here

$$\Theta = \{(\mu_1,\sigma_1^2,\mu_2,\sigma_2^2)\colon -\infty<\mu_i<\infty, \sigma_i^2>0, i=1,2\}$$

and

$$\Theta_0 = \{(\mu_1,\sigma_1^2,\mu_2,\sigma_2^2)\colon -\infty<\mu_i<\infty, i=1,2, \sigma_1^2=\sigma_2^2>0\}.$$

Let $\boldsymbol{\theta} = (\mu_1,\sigma_1^2,\mu_2,\sigma_2^2)$. Then the joint PDF is

$$f_{\boldsymbol{\theta}}(\mathbf{x},\mathbf{y}) = \frac{1}{(2\pi)^{(m+n)/2}\sigma_1^m\sigma_2^n}\exp\left[-\frac{1}{2\sigma_1^2}\sum_1^m (x_i-\mu_1)^2 - \frac{1}{2\sigma_2^2}\sum_1^n (y_i-\mu_2)^2\right].$$

Also,

$$\log f_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{y}) = -\frac{m+n}{2}\log 2\pi - \frac{m}{2}\log\sigma_1^2 - \frac{n}{2}\log\sigma_2^2 - \frac{\sum_1^m (x_i - \mu_1)^2}{2\sigma_1^2} - \frac{1}{2\sigma_2^2}\sum_1^n (y_i - \mu_2)^2.$$

Differentiating with respect to μ_1 and μ_2, we obtain the MLEs

$$\hat{\mu}_1 = \overline{x}, \qquad \hat{\mu}_2 = \overline{y}.$$

Differentiating with respect to σ_1^2 and σ_2^2, we obtain the MLEs

$$\hat{\sigma}_1^2 = \frac{1}{m}\sum_1^m (x_i - \overline{x})^2, \qquad \hat{\sigma}_2^2 = \frac{1}{n}\sum_1^n (y_i - \overline{y})^2.$$

If, however, $\sigma_1^2 = \sigma_2^2 = \sigma^2$, the MLE of σ^2 is

$$\hat{\sigma}^2 = \frac{\sum_1^m (x_i - \overline{x})^2 + \sum_1^n (y_i - \overline{y})^2}{m+n}.$$

Thus

$$\sup_{\boldsymbol{\theta}\in\Theta_0} f_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{y}) = \frac{e^{-(m+n)/2}}{[2\pi/(m+n)]^{(m+n)/2}\left[\sum_1^m (x_i - \overline{x})^2 + \sum_1^n (y_i - \overline{y})^2\right]^{(m+n)/2}}$$

and

$$\sup_{\boldsymbol{\theta}\in\Theta} f_{\boldsymbol{\theta}}(\mathbf{x}, \mathbf{y}) = \frac{e^{-(m+n)/2}}{(2\pi/m)^{m/2}(2\pi/n)^{n/2}\left[\sum_1^m (x_i - \overline{x})^2\right]^{m/2}\left[\sum_1^n (y_i - \overline{y})^2\right]^{n/2}},$$

so that

$$\lambda(\mathbf{x}, \mathbf{y}) = \left(\frac{m}{m+n}\right)^{m/2}\left(\frac{n}{m+n}\right)^{n/2}\frac{\left[\sum_1^m (x_i - \overline{x})^2\right]^{m/2}\left[\sum_1^n (y_i - \overline{y})^2\right]^{n/2}}{\left[\sum_1^m (x_i - \overline{x})^2 + \sum_1^n (y_i - \overline{y})^2\right]^{(m+n)/2}}.$$

Now

$$\frac{\left[\sum_1^m (x_i - \overline{x})^2\right]^{m/2}\left[\sum_1^n (y_i - \overline{y})^2\right]^{n/2}}{\left[\sum_1^m (x_i - \overline{x})^2 + \sum_1^n (y_i - \overline{y})^2\right]^{(m+n)/2}}$$
$$= \frac{1}{\left[1 + \sum_1^m (x_i - \overline{x})^2 / \sum_1^n (y_i - \overline{y})^2\right]^{n/2}\left[1 + \sum_1^n (y_i - \overline{y})^2 / \sum_1^m (x_i - \overline{x})^2\right]^{m/2}}.$$

Writing

$$f = \frac{\sum_1^m (x_i - \overline{x})^2/(m-1)}{\sum_1^n (y_i - \overline{y})^2/(n-1)},$$

we have

$$\lambda(\mathbf{x}, \mathbf{y}) = \left(\frac{m}{m+n}\right)^{m/2} \left(\frac{n}{m+n}\right)^{n/2} \times \frac{1}{\{1 + [(m-1)/(n-1)]f\}^{n/2}\{1 + [(n-1)/(m-1)](1/f)\}^{m/2}}.$$

We leave the reader to check that $\lambda(\mathbf{x}, \mathbf{y}) < c$ is equivalent to $f < c_1$ or $f > c_2$. (Take logarithms, and use properties of convex functions. Alternatively, differentiate $\log \lambda$.)

Under H_0, the statistic

$$F = \frac{\sum_1^m (X_i - \overline{X})^2/(m-1)}{\sum_1^n (Y_i - \overline{Y})^2/(n-1)}$$

has an $F(m-1, n-1)$ distribution, so that c_1, c_2 can be selected. It is usual to take

$$P\{F \le c_1\} = P\{F \ge c_2\} = \frac{\alpha}{2}.$$

Under H_1, $(\sigma_2^2/\sigma_1^2)F$ has an $F(m-1, n-1)$ distribution.

In Example 3 we can obtain the same GLR test by focusing attention on the joint sufficient statistic $(\overline{X}, \overline{Y}, S_X^2, S_Y^2)$, where S_X^2 and S_Y^2 are sample variances of the X's and the Y's, respectively. In order to write down the likelihood function, we note that $\overline{X}, \overline{Y}, S_X^2, S_Y^2$ are independent RVs. The distributions $\overline{X}$ and S_X^2 are the same as in Example 2 except that m is the sample size. Distributions of $\overline{Y}$ and S_Y^2 require appropriate modifications. We leave the reader to carry out the details. It turns out that the GLR test coincides with the UMP unbiased test in this case.

In certain situations the GLR test does not perform well. We reproduce here an example due to Stein and Rubin.

Example 4. Let X be a discrete RV with PMF

$$P_{p=0}\{X = x\} = \begin{cases} \dfrac{\alpha}{2} & \text{if } x = \pm 2, \\ \dfrac{1-2\alpha}{2} & \text{if } x = \pm 1, \\ \alpha & \text{if } x = 0, \end{cases}$$

under the null hypothesis $H_0\colon p = 0$, and

$$P_p\{X = x\} = \begin{cases} pc & \text{if } x = -2, \\ \dfrac{1-c}{1-\alpha}\left(\dfrac{1}{2} - \alpha\right) & \text{if } x = \pm 1, \\ \alpha\left(\dfrac{1-c}{1-\alpha}\right) & \text{if } x = 0, \\ (1-p)c & \text{if } x = 2, \end{cases}$$

under the alternative $H_1\colon p \in (0, 1)$, where α and c are constants with

$$0 < \alpha < \frac{1}{2} \quad \text{and} \quad \frac{\alpha}{2-\alpha} < c < \alpha.$$

To test the simple null hypothesis against the composite alternative at the level of significance α, let us compute the likelihood ratio λ. We have

$$\lambda(2) = \frac{P_0\{X = 2\}}{\sup_{0 \le p < 1} P_p\{X = 2\}} = \frac{\alpha/2}{c} = \frac{\alpha}{2c}$$

since $\alpha/2 < c$. Similarly, $\lambda(-2) = \alpha/(2c)$. Also,

$$\lambda(1) = \lambda(-1) = \frac{\frac{1}{2} - \alpha}{[(1-c)/(1-\alpha)](\frac{1}{2} - \alpha)} = \frac{1-\alpha}{1-c}, \qquad \alpha < \frac{1}{2},$$

and

$$\lambda(0) = \frac{1-\alpha}{1-c}.$$

The GLR test rejects H_0 if $\lambda(x) < k$, where k is to be determined so that the level is α. We see that

$$P_0\left\{\lambda(X) < \frac{1-\alpha}{1-c}\right\} = P_0\{X = \pm 2\} = \alpha,$$

provided that $\alpha/2c < [(1-\alpha)/(1-c)]$. But $\alpha/(2-\alpha) < c < \alpha$ implies that $\alpha < 2c - c\alpha$, so that $\alpha - c\alpha < 2c - 2c\alpha$, or $\alpha(1-c) < 2c(1-\alpha)$, as required. Thus the GLR size α test is to reject H_0 if $X = \pm 2$. The power of the GLR test is

$$P_p\left\{\lambda(X) < \frac{1-\alpha}{1-c}\right\} = P_p\{X = \pm 2\} = pc + (1-p)c = c < \alpha$$

for all $p \in (0, 1)$. The test is not unbiased and is even worse than the trivial test $\varphi(x) \equiv \alpha$.

Another test that is better than the trivial test is to reject H_0 whenever $x = 0$ (this is opposite to what the likelihood ratio test says). Then

$$P_0\{X = 0\} = \alpha, \quad \text{and} \quad P_p\{X = 0\} = \alpha\frac{1-c}{1-\alpha} > \alpha \qquad (\text{since } c < \alpha),$$

for all $p \in (0, 1)$, and the test is unbiased.

We will use the generalized likelihood ratio procedure quite frequently hereafter because of its simplicity and wide applicability. The exact distribution of the test statistic under H_0 is generally difficult to obtain (despite what we saw in Examples 1 to 3 above), and evaluation of power function is also not possible in many problems. Recall, however, that under certain conditions the asymptotic distribution of the MLE is normal. This result can be used to prove the following large-sample property of the GLR under H_0, which solves the problem of computation of the cutoff point c at least when the sample size is large.

Theorem 3. Under some regularity conditions on $f_{\boldsymbol{\theta}}(\mathbf{x})$, the random variable $-2\log\lambda(\mathbf{X})$ under H_0 is asymptotically distributed as a chi-square RV with degrees of freedom equal to the difference between the number of independent parameters in Θ and the number in Θ_0.

We will not prove this result here; the reader is referred to Wilks [117, p. 419]. The regularity conditions are essentially those associated with Theorem 8.7.4. In Example 2 the number of parameters unspecified under H_0 is 1 (namely, σ^2), and under H_1 two parameters are unspecified (μ and σ^2), so that the asymptotic chi-square distribution will have 1 d.f. Similarly, in Example 3, the d.f. $= 4 - 3 = 1$.

Example 5. In Example 2 we showed that in sampling from a normal population with unknown mean μ and unknown variance σ^2, the likelihood ratio for testing $H_0: \mu = \mu_0$ against $H_1: \mu \neq \mu_0$ is

$$\lambda(\mathbf{x}) = \left[1 + \frac{n(\overline{x} - \mu_0)^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}\right]^{-n/2}.$$

Thus

$$-2\log\lambda(\mathbf{X}) = n\log\left[1 + n\frac{(\overline{X} - \mu_0)^2}{\sum_1^n(X_i - \overline{X})^2}\right].$$

Under H_0, $\sqrt{n}(\overline{X} - \mu_0)/\sigma \sim \mathcal{N}(0, 1)$ and $\sum_1^n(X_i - \overline{X})^2/\sigma^2 \sim \chi^2(n-1)$. Also, $\sum_{i=1}^{n}(X_i - \overline{X})^2/[(n-1)\sigma^2] \xrightarrow{P} 1$. It follows that if $Z \sim \mathcal{N}(0, 1)$, then $-2\log\lambda(\mathbf{X})$ has the same limiting distribution as $n\log[1 + Z^2/(n-1)]$. Moreover,

$$\left(1 + \frac{Z^2}{n-1}\right)^n \xrightarrow{L} \exp(Z^2)$$

and since logarithm is a continuous function, we see that

$$n \log\left(1 + \frac{Z^2}{n-1}\right) \xrightarrow{L} Z^2.$$

Thus $-2\log\lambda(X) \xrightarrow{L} Y$, where $Y \sim \chi^2(1)$. This result is consistent with Theorem 3.

PROBLEMS 10.2

1. Prove Theorems 1 and 2.
2. A random sample of size n is taken from the PMF $P(X_j = x_j) = p_j$, $j = 1, 2, 3, 4$, $0 < p_j < 1$, $\sum_{j=1}^{4} p_j = 1$. Find the form of the GLR test of $H_0\colon p_1 = p_2 = p_3 = p_4 = \frac{1}{4}$ against $H_1\colon p_1 = p_2 = p/2$, $p_3 = p_4 = (1-p)/2$, $0 < p < 1$.
3. Find the GLR test of $H_0\colon p = p_0$ against $H_1\colon p \neq p_0$, based on a sample of size 1 from $b(n, p)$.
4. Let $X_1, X_2, \dots, X_n$ be a sample from $\mathcal{N}(\mu, \sigma^2)$, where both μ and σ^2 are unknown. Find the GLR test of $H_0\colon \sigma = \sigma_0$ against $H_1\colon \sigma \neq \sigma_0$.
5. Let $X_1, X_2, \dots, X_k$ be a sample from the PMF

$$P_N\{X = j\} = \frac{1}{N}, \qquad j = 1, 2, \dots, N, \ N \geq 1 \text{ is an integer.}$$

   (a) Find the GLR test of $H_0\colon N \leq N_0$ against $H_1\colon N > N_0$.
   (b) Find the GLR test of $H_0\colon N = N_0$ against $H_1\colon N \neq N_0$.
6. For a sample of size 1 from the PDF

$$f_\theta(x) = \frac{2}{\theta^2}(\theta - x), \qquad 0 < x < \theta,$$

   find the GLR test of $\theta = \theta_0$ against $\theta \neq \theta_0$.
7. Let $X_1, X_2, \dots, X_n$ be a sample from $G(1, \beta)$.
   (a) Find the GLR test of $\beta = \beta_0$ against $\beta \neq \beta_0$.
   (b) Find the GLR test of $\beta \leq \beta_0$ against $\beta > \beta_0$.
8. Let $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ be a random sample from a bivariate normal population with $EX_i = \mu_1$, $EY_i = \mu_2$, $\text{var}(X_i) = \sigma^2$, $\text{var}(Y_i) = \sigma^2$, and $\text{cov}(X_i, Y_i) = \rho\sigma^2$. Show that the likelihood ratio test of the null hypothesis $H_0\colon \rho = 0$ against $H_1\colon \rho \neq 0$ reduces to rejecting H_0 if $|R| > c$, where $R = 2S_{11}/(S_1^2 + S_2^2)$, S_{11}, S_1^2, and S_2^2 being the sample covariance and the sample variances, respectively. (For the PDF of the test statistic R, see Problem 7.7.1.)

9. Let $X_1, X_2, \ldots, X_m$ be iid $G(1, \theta)$ RVs and let $Y_1, Y_2, \ldots, Y_n$ be iid $G(1, \mu)$ RVs, where θ and μ are unknown positive real numbers. Assume that the X's and the Y's are independent. Develop an α-level GLR test for testing $H_0: \theta = \mu$ against $H_1: \theta \neq \mu$.

10. A die is tossed 60 times in order to test H_0: $P\{j\} = 1/6, j = 1, 2, \ldots, 6$ (die is fair) against H_1: $P\{2\} = P\{4\} = P\{6\} = \frac{2}{9}$, $P\{1\} = P\{3\} = P\{5\} = \frac{1}{9}$. Find the GLR test.

11. Let $X_1, X_2, \ldots, X_n$ be iid with the common PDF $f_\theta(x) = \exp[-(x - \theta)]$, $x > \theta$ and $= 0$ otherwise. Find the level α GLR test for testing $H_0: \theta \leq \theta_0$ against $H_1: \theta > \theta_0$.

12. Let $X_1, X_2, \ldots, X_n$ be iid RVs with the common Pareto PDF $f_\theta(x) = \theta/x^2$ for $x > \theta$, and $= 0$ elsewhere. Show that the family of joint PDFs has MLR in $X_{(1)}$ and find a size α test of $H_0: \theta = \theta_0$ against $H_1: \theta > \theta_0$. Show that the GLR test coincides with the UMP test.

10.3 CHI-SQUARE TESTS

In this section we consider a variety of tests where the test statistic has an exact or a limiting chi-square distribution. Chi-square tests are also used for testing some nonparametric hypotheses and are taken up again in Chapter 13.

We begin with tests concerning variances in sampling from a normal population. Let $X_1, X_2, \ldots, X_n$ be iid $\mathcal{N}(\mu, \sigma^2)$ RVs where σ^2 is unknown. We wish to test a hypothesis of the type $\sigma^2 \geq \sigma_0^2$, $\sigma^2 \leq \sigma_0^2$, or $\sigma^2 = \sigma_0^2$, where σ_0 is some given positive number. We summarize the tests in the following table:

			Reject H_0 at Level α if:	
	H_0	H_1	μ Known	μ Unknown
I.	$\sigma \geq \sigma_0$	$\sigma < \sigma_0$	$\sum_1^n (x_i - \mu)^2 \leq \chi^2_{n,1-\alpha}\sigma_0^2$	$s^2 \leq \dfrac{\sigma_0^2}{n-1}\chi^2_{n-1,1-\alpha}$
II.	$\sigma \leq \sigma_0$	$\sigma > \sigma_0$	$\sum_1^n (x_i - \mu)^2 \geq \chi^2_{n,\alpha}\sigma_0^2$	$s^2 \geq \dfrac{\sigma_0^2}{n-1}\chi^2_{n-1,\alpha}$
III.	$\sigma = \sigma_0$	$\sigma \neq \sigma_0$	$\sum_1^n (x_i - \mu)^2 \leq \chi^2_{n,1-\alpha/2}\sigma_0^2$ or $\sum_1^n (x_i - \mu)^2 \geq \chi^2_{n,\alpha/2}\sigma_0^2$	$s^2 \leq \dfrac{\sigma_0^2}{n-1}\chi^2_{n-1,1-\alpha/2}$ or $s^2 \geq \dfrac{\sigma_0^2}{n-1}\chi^2_{n-1,\alpha/2}$

Remark 1. All these tests can be derived by the standard likelihood ratio procedure. If μ is unknown, tests I and II are UMP unbiased (and UMP invariant). If μ is known, tests I and II are UMP (see Example 9.4.5). For tests III we have chosen

constants c_1, c_2 so that each tail has probability $\alpha/2$. This is the customary procedure, even though it destroys the unbiasedness property of the tests, at least for small samples.

Example 1. A manufacturer claims that the lifetime of a certain brand of batteries produced by his factory has a variance of 5000 (hours)^2. A sample of size 26 has a variance of 7200 (hours)^2. Assuming that it is reasonable to treat these data as a random sample from a normal population, let us test the manufacturer's claim at the $\alpha = 0.02$ level. Here $H_0\colon \sigma^2 = 5000$ is to be tested against $H_1 : \sigma^2 \neq 5000$. We reject H_0 if either

$$s^2 = 7200 \le \frac{\sigma_0^2}{n-1}\chi^2_{n-1,1-\alpha/2} \quad \text{or} \quad s^2 > \frac{\sigma_0^2}{n-1}\chi^2_{n-1,\alpha/2}.$$

We have

$$\frac{\sigma_0^2}{n-1}\chi^2_{n-1,1-\alpha/2} = \frac{5000}{25} \times 110.524 = 2304.8$$

and

$$\frac{\sigma_0^2}{n-1}\chi^2_{n-1,\alpha/2} = \frac{5000}{25} \times 44.314 = 8862.8$$

Since s^2 is neither $\le$ 2304.8 nor $\ge$ 8862.8, we cannot reject the manufacturer's claim at the 0.02 level.

A test based on a chi-square statistic is also used for testing the equality of several proportions. Let $X_1, X_2, \ldots, X_k$ be independent RVs with $X_i \sim b(n_i, p_i)$, $i = 1, 2, \ldots, k$, $k \ge 2$.

Theorem 1. The RV $\sum_{i=1}^{k}[(X_i - n_i p_i)/\sqrt{n_i p_i(1-p_i)}]^2$ converges in distribution to the $\chi^2(k)$ RV as $n_1, n_2, \ldots, n_k \to \infty$.

The proof is left as an exercise.

If $n_1, n_2, \ldots, n_k$ are large, we can use Theorem 1 to test $H_0\colon p_1 = p_2 = \cdots = p_k = p$ against all alternatives. If p is known, we compute

$$y = \sum_1^k \left[\frac{x_i - n_i p}{\sqrt{n_i p(1-p)}}\right]^2,$$

and if $y \ge \chi^2_{k,\alpha}$, we reject H_0. In practice, p will be unknown. Let $\mathbf{p} = (p_1, p_2, \ldots, p_k)$. Then the likelihood function is

$$L(\mathbf{p}; x_1, \ldots, x_k) = \prod_1^k \left[\binom{n_i}{x_i} p_i^{x_i}(1-p_i)^{n_i - x_i}\right]$$

so that

$$\log L(\mathbf{p}; x) = \sum_{i=1}^{k} \log \binom{n_i}{x_i} + \sum_{i=1}^{k} x_i \log p_i + \sum_{i=1}^{k} (n_i - x_i) \log(1 - p_i).$$

The MLE $\hat{p}$ of p under H_0 is therefore given by

$$\frac{\sum_1^k x_i}{p} - \frac{\sum_1^k (n_i - x_i)}{1 - p} = 0,$$

that is,

$$\hat{p} = \frac{x_1 + x_2 + \cdots + x_k}{n_1 + n_2 + \cdots + n_k}.$$

Under certain regularity assumptions (see Cramér [16, pp. 426–427]) it can be shown that the statistic

$$(1) \qquad Y_1 = \sum_1^k \frac{(X_i - n_i \hat{p})^2}{n_i \hat{p}(1 - \hat{p})}$$

is asymptotically $\chi^2(k - 1)$. Thus the test rejects $H_0\colon p_1 = p_2 = \cdots = p_k = p$, p unknown, at level α if $y_1 \geq \chi^2_{k-1,\alpha}$.

It should be remembered that the tests based on Theorem 1 are all large-sample tests and hence not exact, in contrast to the tests concerning the variance discussed above, which are all exact tests. In the case $k = 1$, UMP tests of $p \geq p_0$ and $p \leq p_0$ exist and can be obtained by the MLR method described in Section 9.4. For testing $p = p_0$, the usual test is UMP unbiased.

In the case $k = 2$, if n_1 and n_2 are large, a test based on the normal distribution can be used instead of Theorem 1. In this case the statistic

$$(2) \qquad Z = \frac{X_1/n_1 - X_2/n_2}{\sqrt{\hat{p}(1 - \hat{p})(1/n_1 + 1/n_2)}},$$

where $\hat{p} = (X_1 + X_2)/(n_1 + n_2)$ is asymptotically $\mathcal{N}(0, 1)$ under $H_0\colon p_1 = p_2 = p$. If p is known, one uses p instead of $\hat{p}$. It is not too difficult to show that Z^2 is equal to Y_1, so that the two tests are equivalent.

For small sampies the *Fisher–Irwin test* is commonly used and is based on the conditional distribution of X_1 given $T = X_1 + X_2$. Let $\rho = [p_1(1 - p_2)]/[p_2(1 - p_1)]$. Then

$$\begin{aligned} P(X_1 + X_2 = t) &= \sum_{j=0}^{t} \binom{n_1}{j} p_1^j (1 - p_1)^{n_1 - j} \binom{n_2}{t - j} p_2^{t-j} (1 - p_2)^{n_1 - t + j} \\ &= \sum_{j=0}^{t} \binom{n_1}{j} \binom{n_2}{t - j} \rho^j a(n_1, n_2) \end{aligned}$$

where

$$a(n_1, n_2) = (1 - p_1)^{n_1}(1 - p_2)^{n_2}\left(\frac{p_2}{1 - p_2}\right)^t.$$

It follows that

$$P\{X_1 = x | X_1 + X_2 = t\} = \frac{\binom{n_1}{x} p_1^x (1 - p_1)^{n_1 - x} \binom{n_2}{t - x} p_2^{t-x} (1 - p_2)^{n_2 - t + x}}{a(n_1, n_2) \sum_{j=0}^{t} \binom{n_1}{j}\binom{n_2}{t - j}\rho^j}$$

$$= \frac{\binom{n_1}{j}\binom{n_2}{t - j}\rho^x}{\sum_{j=0}^{t} \binom{n_1}{j}\binom{n_2}{t - j}\rho^j}.$$

On the boundary of any of the hypotheses $p_1 = p_2$, $p_1 \le p_2$ or $p_1 \ge p_2$, we note that $\rho = 1$, so that

$$P\{X_1 = x | X_1 + X_2 = t\} = \frac{\binom{n_1}{x}\binom{n_2}{t - x}}{\binom{n_1 + n_2}{t}},$$

which is a hypergeometric distribution. For testing $H_0\colon p_1 \le p_2$ this conditional test rejects if $X_1 \le k(t)$ where $k(t)$ is the largest integer for which $P\{X_1 \le k(T) | T = t\} \le \alpha$. Obvious modifications yield critical regions for testing $p_1 = p_2$, and $p_1 \ge p_2$ against corresponding alternatives.

In applications a wide variety of problems can be reduced to the multinomial distribution model. We therefore consider the problem of testing the parameters of a multinomial distribution. Let $(X_1, X_2, \ldots, X_{k-1})$ be a sample from a multinomial distribution with parameters n, $p_1, p_2, \ldots, p_{k-1}$, and let us write $X_k = n - X_1 - \cdots - X_{k-1}$, and $p_k = 1 - p_1 - \cdots - p_{k-1}$. The difference between the model of Theorem 1 and the multinomial model is the independence of the X_i's.

Theorem 2. Let $(X_1, X_2, \ldots, X_{k-1})$ be a multinomial RV with parameters n, $p_1, p_2, \ldots, p_{k-1}$. Then the RV

$$U_k = \sum_{i=1}^{k} \frac{(X_i - np_i)^2}{np_i} \tag{3}$$

is asymptotically distributed as a $\chi^2(k - 1)$ RV (as $n \to \infty$).

Proof. For the general proof we refer the reader to Cramér [16, pp. 417–419] or Ferguson [26, p. 61]. We will consider here the $k = 2$ case to make the result a little more plausible. We have

$$U_2 = \frac{(X_1 - np_1)^2}{np_1} + \frac{(X_2 - np_2)^2}{np_2} = \frac{(X_1 - np_1)^2}{np_1} + \frac{[n - X_1 - n(1 - p_1)]^2}{n(1 - p_1)}$$

$$= (X_1 - np_1)^2 \left[\frac{1}{np_1} + \frac{1}{n(1 - p_1)}\right]$$

$$= \frac{(X_1 - np_1)^2}{np_1(1 - p_1)}.$$

It follows from Theorem 1 that $U_2 \xrightarrow{L} Y$ as $n \to \infty$, where $Y \sim \chi^2(1)$.

To use Theorem 2 to test $H_0\colon p_1 = p_1', \dots, p_k = p_k'$, we need only to compute the quantity

$$u = \sum_1^k \frac{(x_i - np_i')^2}{np_i'}$$

from the sample; if n is large, we reject H_0 if $u > \chi^2_{k-1,\alpha}$.

Example 2. A die is rolled 120 times with the following results:

Result	1	2	3	4	5	6
Frequency:	20	30	20	25	15	10

Let us test the hypothesis that the die is fair at level $\alpha = 0.05$. The null hypothesis is $H_0\colon p_i = \frac{1}{6}$, $i = 1, 2, \dots, 6$, where p_i is the probability that the face value is i, $1 \le i \le 6$. By Theorem 2, we reject H_0 if

$$u = \sum_1^6 \frac{[x_i - 120(\frac{1}{6})]^2}{120(\frac{1}{6})} > \chi^2_{5,0.05}.$$

We have

$$u = 0 + \frac{10^2}{20} + 0 + \frac{5^2}{20} + \frac{5^2}{20} + \frac{10^2}{20} = 12.5.$$

Since $\chi_{5,0.05} = 11.07$, we reject H_0. Note that if we choose $\alpha = 0.025$, then $\chi_{5,0.025} = 12.8$, and we cannot reject at this level.

Theorem 2 has much wider applicability, and we will later study its application to contingency tables. Here we consider the application of Theorem 2 to testing the null hypothesis that the DF of an RV X has a specified form.

Theorem 3. Let $X_1, X_2, \dots, X_n$ be a random sample on X. Also, let $H_0\colon X \sim F$, where the functional form of the DF F is known completely. Consider a collection of disjoint Borel sets $A_1, A_2, \ldots, A_k$ that form a partition of the real line. Let $P\{X \in A_i\} = p_i$, $i = 1, 2, \ldots, k$, and assume that $p_i > 0$ for each i. Let $Y_j =$ number of X_i's in A_j, $j = 1, 2, \ldots, k$, $i = 1, 2, \ldots, n$. Then the joint distribution of $(Y_1, Y_2, \ldots, Y_{k-1})$ is multinomial with parameters $n, p_1, p_2, \ldots, p_{k-1}$. Clearly, $Y_k = n - Y_1 - \cdots - Y_{k-1}$ and $p_k = 1 - p_1 - \cdots - p_{k-1}$.

The proof of Theorem 3 is obvious. One frequently selects $A_1, A_2, \ldots, A_k$ as disjoint intervals. Theorem 3 is especially useful when one or more of the parameters associated with the DF F are unknown. In that case the following result is useful.

Theorem 4. Let $H_0\colon X \sim F_{\boldsymbol{\theta}}$, where $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_r)$ is unknown. Let $X_1, X_2, \ldots, X_n$ be independent observations on X, and suppose that the MLEs of $\theta_1, \theta_2, \ldots, \theta_r$ exist and are, respectively, $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_r$. Let $A_1, A_2, \ldots, A_k$ be a collection of disjoint Borel sets that cover the real line, and let

$$\hat{p}_i = P_{\hat{\boldsymbol{\theta}}}\{X \in A_i\} > 0 \qquad i = 1, 2, \ldots, k,$$

where $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_r)$, and $P_{\boldsymbol{\theta}}$ is the probability distribution associated with $F_{\boldsymbol{\theta}}$. Let $Y_1, Y_2, \ldots, Y_k$ be the RVs, defined as follows: $Y_i =$ number of $X_1, X_2, \ldots, X_n$ in A_i, $i = 1, 2, \ldots, k$.

Then the RV

$$V_k = \sum_{n=1}^{k} \frac{(Y_i - n\hat{p}_i)^2}{n\hat{p}_i}$$

is asymptotically distributed as a $\chi^2(k - r - 1)$ RV (as $n \to \infty$).

The proof of Theorem 4 and some regularity conditions required on $F_{\boldsymbol{\theta}}$ are given in Rao [86, pp. 391–392].

To test $H_0\colon X \sim F$, where F is completely specified, we reject H_0 if

$$u = \sum_{1}^{k} \frac{(y_i - np_i)^2}{np_i} > \chi^2_{k-1,\alpha},$$

provided that n is sufficiently large. If the null hypothesis is $H_0\colon X \sim F_{\boldsymbol{\theta}}$, where $F_{\boldsymbol{\theta}}$ is known except for the parameter $\boldsymbol{\theta}$, we use Theorem 4 and reject H_0 if

$$v = \sum_{i=1}^{k} \frac{(y_i - n\hat{p}_i)^2}{n\hat{p}_i} > \chi^2_{k-r-1,\alpha},$$

where r is the number of parameters estimated.

Example 3. The following data were obtained from a table of random numbers of normal distribution with mean 0 and variance 1.

0.464	0.137	2.455	−0.323	−0.068
0.906	−0.513	−0.525	0.595	0.881
−0.482	1.678	−0.057	−1.229	−0.486
−1.787	−0.261	1.237	1.046	−0.508

We want to test the null hypothesis that the DF F from which the data came is normal with mean 0 and variance 1. Here F is completely specified. Let us choose three intervals $(-\infty, -0.5]$, $(-0.5, 0.5]$, and $(0.5, \infty)$. We see that $Y_1 = 5$, $Y_2 = 8$, and $Y_3 = 7$.

Also, if Z is $\mathcal{N}(0, 1)$, then $p_1 = 0.3085$, $p_2 = 0.3830$, and $p_3 = 0.3085$. Thus

$$\begin{aligned} u &= \sum_{i=1}^{3} \frac{(y_i - np_i)^2}{np_i} \\ &= \frac{(5 - 20 \times 0.3085)^2}{6.17} + \frac{(8 - 20 \times 0.383)^2}{7.66} + \frac{(7 - 20 \times 0.3085)^2}{6.17} \\ &< 1. \end{aligned}$$

Also, $\chi^2_{2,0.05} = 5.99$, so we cannot reject H_0 at level 0.05.

Example 4. In a 72-hour period on a long holiday weekend, there was a total of 306 fatal automobile accidents. The data are as follows:

Number of Fatal Accidents per Hour	Number of Hours
0 or 1	4
2	10
3	15
4	12
5	12
6	6
7	6
8 or more	7

Let us test the hypothesis that the number of accidents per hour is a Poisson RV. Since the mean of the Poisson RV is not given, we estimate it by

$$\hat{\lambda} = \overline{x} = \frac{306}{72} = 4.25.$$

Let us now estimate $\hat{p}_i = P_{\hat{\lambda}}\{X = i\}$, $i = 0, 1, 2, \ldots$, $\hat{p}_0 = e^{-\hat{\lambda}} = 0.0143$. Note that

$$\frac{P_{\hat{\lambda}}\{X = x + 1\}}{P_{\hat{\lambda}}\{X = x\}} = \frac{\hat{\lambda}}{x + 1},$$

so that $\hat{p}_{i+1} = [\hat{\lambda}/(i+1)]\hat{p}_i$. Thus

$$\hat{p}_1 = 0.0606, \hat{p}_2 = 0.1288, \hat{p}_3 = 0.1825, \hat{p}_4 = 0.1939,$$
$$\hat{p}_5 = 0.1648, \hat{p}_6 = 0.1167, \hat{p}_7 = 0.0709, \hat{p}_8 = 1 - 0.9325 = 0.0675.$$

The observed and expected frequencies are as follows:

	i							
	0 or 1	2	3	4	5	6	7	8 or More
Observed frequency, 0_i	4	10	15	12	12	6	6	7
Expected frequency $= 72\hat{p}_i = e_i$	5.38	9.28	13.14	13.96	11.87	8.41	5.10	4.86

$$\begin{aligned} u &= \sum_{i=1}^{8} \frac{(0_i - e_i)^2}{e_i} \\ &= 2.74. \end{aligned}$$

Since we estimated one parameter, the number of degrees of freedom is $k - r - 1 = 8 - 1 - 1 = 6$. From Table ST3, $\chi^2_{6,0.05} = 12.6$, and since $2.74 < 12.6$, we cannot reject the null hypothesis.

Remark 2. Any application of Theorem 3 or 4 requires that we choose sets $A_1, A_2, \ldots, A_k$, and frequently these are chosen to be disjoint intervals. As a rule of thumb, we choose the length of each interval in such a way that the probability $P\{X \in A_i\}$ under H_0 is approximately $1/k$. Moreover, it is desirable to have $n/k \geq 5$ or, rather, $e_i \geq 5$ for each i. If any of the e_i's is < 5, the corresponding interval is pooled with one or more adjoining intervals to make the cell frequency at least 5. If any pooling is done, the number of degrees of freedom is the number of classes after pooling, minus 1, minus the number of parameters estimated.

Finally, we consider a test of *homogeneity* of several multinomial distributions. Suppose that we have c samples of sizes $n_1, n_2, \ldots, n_c$ from c multinomial distributions. Let the associated probabilities with the jth population be $(p_{1j}, p_{2j}, \ldots, p_{rj})$ where $\sum_{i=1}^{r} p_{ij} = 1$, $j = 1, 2, \ldots, c$. Given observations N_{ij}, $i = 1, 2, \ldots, r$, $j = 1, 2, \ldots, c$ with $\sum_{i=1}^{r} N_{ij} = n_j$, $j = 1, 2, \ldots, c$ we wish to test $H_0\colon p_{ij} = p_i$, for $j = 1, 2, \ldots, c$, $i = 1, 2, \ldots, r - 1$. The case $c = 1$ is covered by Theorem 2. By Theorem 2 for each j,

$$U_r = \sum_{i=1}^{r} \frac{(N_{ij} - n_j p_i)^2}{n_j p_i}$$

has a limiting χ^2_{r-1} distribution. Since samples are independent, the statistic

$$U_{rc} = \sum_{j=1}^{c} \sum_{i=1}^{r} \frac{(N_{ij} - n_j p_i)^2}{n_j p_i}$$

has a limiting $\chi^2_{c(r-1)}$ distribution. If p_i's are unknown, we use the MLEs

$$\hat{p}_i = \frac{\sum_{j=1}^{c} N_{ij}}{\sum_{j=1}^{c} n_j}, \qquad i = 1, 2, \ldots, r-1$$

for p_i, and we see that the statistic

$$V_{rc} = \sum_{j=1}^{c} \sum_{i=1}^{r} \frac{(N_{ij} - n_j \hat{p}_i)^2}{n_j \hat{p}_i}$$

has a chi-square distribution with $c(r-1) - (r-1) = (c-1)(r-1)$ d.f. We reject H_0 at (approximate) level α is $V_{rc} > \chi^2_{(r-1)(c-1),\alpha}$.

Example 5. A market analyst believes that there is no difference in preferences of television viewers among the four Ohio cities of Toledo, Columbus, Cleveland, and Cincinnati. To test this belief, independent random samples of 150, 200, 250, and 200 persons were selected from the four cities and asked, "What type of program do you prefer most: mystery, soap, comedy, or news documentary?" The following responses were recorded:

	City			
Program Type	Toledo	Columbus	Cleveland	Cincinnati
Mystery	50	70	85	60
Soap	45	50	58	40
Comedy	35	50	72	67
News	20	30	35	33
Sample size	150	200	250	200

Under the null hypothesis that the proportions of viewers who prefer the four types of programs are the same in each city, the maximum likelihood estimates of $p_i, i = 1, 2, 3, 4$ are given by

$$\hat{p}_1 = \frac{50 + 70 + 85 + 60}{150 + 200 + 250 + 200} = \frac{265}{800} = 0.33,$$

$$\hat{p}_2 = \frac{45 + 50 + 58 + 40}{800} = \frac{193}{800} = 0.24,$$

$$\hat{p}_3 = \frac{35 + 50 + 72 + 67}{800} = \frac{224}{800} = 0.28,$$

$$\hat{p}_4 = \frac{20 + 30 + 35 + 33}{800} = \frac{118}{800} = 0.15.$$

Here p_1 =proportion of people who prefer mystery, and so on. The following table gives the expected frequencies under H_0:

Program Type	Expected Number of Responses Under H_0			
	Toledo	Columbus	Cleveland	Cincinnati
Mystery	$150 \times 0.33 = 49.5$	$200 \times 0.33 = 66$	$250 \times 0.33 = 82.5$	$200 \times 0.33 = 66$
Soap	$150 \times 0.24 = 36$	$200 \times 0.24 = 48$	$250 \times 0.24 = 60$	$200 \times 0.24 = 48$
Comedy	$150 \times 0.28 = 42$	$200 \times 0.28 = 56$	$250 \times 0.28 = 70$	$200 \times 0.28 = 56$
News	$150 \times 0.15 = 22.5$	$200 \times 0.15 = 30$	$250 \times 0.15 = 37.5$	$200 \times 0.15 = 30$
Sample size	150	200	250	200

It follows that

$$\begin{aligned} u_{44} = {} & \frac{(50-49.5)^2}{49.5} + \frac{(45-36)^2}{36} + \frac{(35-42)^2}{42} + \frac{(20-22.5)^2}{22.5} \\ & + \frac{(70-66)^2}{66} + \frac{(50-48)^2}{48} + \frac{(50-56)^2}{56} + \frac{(30-30)^2}{30} \\ & + \frac{(85-82.5)^2}{82.5} + \frac{(58-60)^2}{60} + \frac{(72-70)^2}{70} + \frac{(35-37.5)^2}{37.5} \\ & + \frac{(60-66)^2}{66} + \frac{(40-48)^2}{48} + \frac{(67-56)^2}{56} + \frac{(33-30)^2}{30} \\ = {} & 9.37. \end{aligned}$$

Since $c = 4$ and $r = 4$, the number of degrees of freedom is $(4-1)(4-1) = 9$ and we note that under H_0

$$0.30 < P\{U_{44} \geq 9.37\} < 0.50.$$

With such a large P-value we can hardly reject H_0. The data do not offer any evidence to conclude that the proportions in the four cities are different.

PROBLEMS 10.3

1. The standard deviation of capacity for batteries of a standard type is known to be 1.66 ampere-hours. The following capacities (ampere-hours) were recorded

for 10 batteries of a new type: 146, 141, 135, 142, 140, 143, 138, 137, 142, 136. Does the new battery differ from the standard type with respect to variability of capacity? (Natrella [73, p. 4-1])

2. A manufacturer recorded the cutoff bias (volts) of a sample of 10 tubes as follows: 12.1, 12.3, 11.8, 12.0, 12.4, 12.0, 12.1, 11.9, 12.2, 12.2. The variability of cutoff bias for tubes of a standard type as measured by the standard deviation is 0.208 volt. Is the variability of the new tube with respect to cutoff bias less than that of the standard type? (Natrella [73, p. 4-5])

3. Approximately equal numbers of four different types of meters are in service and all types are believed to be equally likely to break down. The actual numbers of breakdowns reported are as follows:

Type of Meter	1	2	3	4
Number of Breakdowns Reported	30	40	33	47

Is there evidence to conclude that the chances of failure of the four types are not equal? (Natrella [73, p. 9-4])

4. Every clinical thermometer is classified into one of four categories, A, B, C, D, on the basis of inspection and test. From past experience it is known that thermometers produced by a certain manufacturer are distributed among the four categories in the following proportions:

Category	A	B	C	D
Proportion	0.87	0.09	0.03	0.01

A new lot of 1336 thermometers is submitted by the manufacturer for inspection and test and the following distribution into the four categories results:

Category	A	B	C	D
Number of Thermometers Reported	1188	91	47	10

Does this new lot of thermometers differ from the previous experience with regard to proportion of thermometers in each category? (Natrella [73, p. 9-2])

5. A computer program is written to generate random numbers, X, uniformly in the interval $0 \le X < 10$. From 250 consecutive values the following data are obtained:

X-Value	0–1.99	2–3.99	4–5.99	6–7.99	8–9.99
Frequency	38	55	54	41	62

Do these data offer any evidence that the program is not written properly?

6. A machine working correctly cuts pieces of wire to a mean length of 10.5 cm with a standard deviation of 0.15 cm. Sixteen samples of wire were drawn at random from a production batch and measured with the following results (cen-

timeters): 10.4, 10.6, 10.1, 10.3, 10.2, 10.9, 10.5, 10.8, 10.6, 10.5, 10.7, 10.2, 10.7, 10.3, 10.4, 10.5. Test the hypothesis that the machine is working correctly.

7. An experiment consists in tossing a coin until the first head shows up. One hundred repetitions of this experiment are performed. The frequency distribution of the number of trials required for the first head is as follows:

Number of Trials	1	2	3	4	5 or more
Frequency	40	32	15	7	6

Can we conclude that the coin is fair?

8. Fit a binomial distribution to the following data:

x	0	1	2	3	4
Frequency	8	46	55	40	11

9. Prove Theorem 1.

10. Three dice are rolled independently 360 times each with the following results.

Face Value	Die 1	Die 2	Die 3
1	50	62	38
2	48	55	60
3	69	61	64
4	45	54	58
5	71	78	73
6	77	50	67
Sample size	360	360	360

Are all the dice equally loaded? That is, test the hypothesis H_0: $p_{i1} = p_{i2} = p_{i3}$, $i = 1, 2, \ldots, 6$, where p_{i1} is the probability of getting an i with die 1, and so on.

11. Independent random samples of 250 Democrats, 150 Republicans, and 100 Independent voters were selected one week before a nonpartisan election for mayor of a large city. Their preference for candidates Albert, Basu, and Chatfield were recorded as follows.

	Party Affiliation		
Preference	Democrat	Republican	Independent
Albert	160	70	90
Basu	32	45	25
Chatfield	30	23	15
Undecided	28	12	20
Sample size	250	150	150

Are the proportions of voters in favor of Albert, Basu, and Chatfield the same within each political affiliation?

12. Of 25 income tax returns audited in a small town, 10 were from low- and middle-income families and 15 from high-income families. Two of the low-income families and four fo the high-income families were found to have underpaid their taxes. Are the two proportions of families who underpaid taxes the same?

13. A candidate for a congressional seat checks her progress by taking a random sample of 20 voters each week. Last week, six reported to be in her favor. This week nine reported to be in her favor. Is there evidence to suggest that her campaign is working?

14. Let $\{X_{11}, X_{21}, \dots, X_{r1}\}, \dots, \{X_{1c}, X_{2c}, \dots, X_{rc}\}$ be independent multinomial RVs with parameters $(n_1, p_{11}, p_{21}, \dots, p_{r1}), \dots, (n_c, p_{1c}, p_{2c}, \dots, p_{rc})$, respectively. Let $X_{i\cdot} = \sum_{j=1}^{c} X_{ij}$ and $\sum_{j=1}^{c} n_j = n$. Show that the GLR test for testing H_0: $p_{ij} = p_j$, for $j = 1, 2, \dots, c$, $i = 1, 2, \dots, r-1$, where p_j's are unknown against all alternatives can be based on the statistic

$$\lambda(\mathbf{X}) = \prod_{i=1}^{r} \left(\frac{X_{i\cdot}}{n}\right)^{X_{i\cdot}} \Bigg/ \prod_{i=1}^{r} \prod_{j=1}^{c} \left(\frac{X_{ij}}{n_j}\right)^{X_{ij}}.$$

10.4 t-TESTS

In this section we investigate one of the most frequently used types of tests in statistics, the tests based on a t-statistic. Let $X_1, X_2, \dots, X_n$ be a random sample from $\mathcal{N}(\mu, \sigma^2)$, and, as usual, let us write

$$\overline{X} = n^{-1} \sum_{1}^{n} X_i, \qquad S^2 = (n-1)^{-1} \sum_{1}^{n} (X_i - \overline{X})^2.$$

The tests for usual null hypotheses about the mean can be derived using the GLR method. In the following table we summarize the results.

			Reject H_0 at Level α if:	
	H_0	H_1	σ^2 Known	σ^2 Unknown
I.	$\mu \le \mu_0$	$\mu > \mu_0$	$\overline{X} \ge \mu_0 + \frac{\sigma}{\sqrt{n}} z_\alpha$	$\overline{x} \ge \mu_0 + \frac{s}{\sqrt{n}} t_{n-1,\alpha}$
II.	$\mu \ge \mu_0$	$\mu < \mu_0$	$\overline{X} \le \mu_0 + \frac{\sigma}{\sqrt{n}} z_{1-\alpha}$	$\overline{x} \le \mu_0 + \frac{s}{\sqrt{n}} t_{n-1,1-\alpha}$
III.	$\mu = \mu_0$	$\mu \ne \mu_0$	$\lvert\overline{x} - \mu_0\rvert \ge \frac{\sigma}{\sqrt{n}} z_{\alpha/2}$	$\lvert\overline{x} - \mu_0\rvert \ge \frac{s}{\sqrt{n}} t_{n-1,\alpha/2}$

Remark 1. A test based on a t-statistic is called a *t-test*. The t-tests in I and II are called *one-tailed tests*; the t-test in III, a *two-tailed test*.

Remark 2. If σ^2 is known, tests I and II are UMP and test III is UMP unbiased. If σ^2 is unknown, the t-tests are UMP unbiased and UMP invariant.

Remark 3. If n is large, we may use normal tables instead of t-tables. The assumption of normality may also be dropped because of the central limit theorem. For small samples care is required in applying the proper test, since the tail probabilities under normal distribution and t-distribution differ significantly for small n (see Remark 7.4.2).

Example 1. Nine determinations of copper in a certain solution yielded a sample mean of 8.3 percent with a standard deviation of 0.025 percent. Let μ be the mean of the population of such determinations. Let us test $H_0: \mu = 8.42$ against $H_1: \mu < 8.42$ at level $\alpha = 0.05$.

Here $n = 9, \bar{x} = 8.3, s = 0.025, \mu_0 = 8.42$, and $t_{n-1,1-\alpha} = -t_{8,0.05} = -1.860$. Thus

$$\mu_0 + \frac{s}{\sqrt{n}} t_{n-1,1-\alpha} = 8.42 - \frac{0.025}{3} 1.86 = 8.4045.$$

We reject H_0 since $8.3 < 8.4045$.

We next consider the two-sample case. Let $X_1, X_2, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$ be independent random samples from $\mathcal{N}(\mu_1, \sigma_1^2)$ and $\mathcal{N}(\mu_2, \sigma_2^2)$, respectively. Let us write

$$\overline{X} = m^{-1} \textstyle\sum_1^m X_i, \qquad \overline{Y} = n^{-1} \sum_1^n Y_i,$$
$$S_1^2 = (m-1)^{-1} \textstyle\sum_1^m (X_i - \overline{X})^2, \qquad S_2^2 = (n-1)^{-1} \sum_1^n (Y_i - \overline{Y})^2,$$

and

$$S_p^2 = \frac{(m-1)S_1^2 + (n-1)S_2^2}{m+n-2}.$$

S_p^2 is sometimes called the *pooled sample variance*. The following table summarizes the two sample tests comparing μ_1 and μ_2:

	H_0	H_1	Reject H_0 at Level α if:	
	(δ = known constant)		σ_1^2, σ_2^2 Known	σ_1^2, σ_2^2 Unknown, $\sigma_1 = \sigma_2$
I.	$\mu_1 - \mu_2 \le \delta$	$\mu_1 - \mu_2 > \delta$	$\bar{x} - \bar{y} \ge \delta + z_\alpha\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}$	$\bar{x} - \bar{y} \ge \delta + t_{m+n-2,\alpha}\, s_p\sqrt{\frac{1}{m} + \frac{1}{n}}$
II.	$\mu_1 - \mu_2 \ge \delta$	$\mu_1 - \mu_2 < \delta$	$\bar{x} - \bar{y} \le \delta - z_\alpha\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}$	$\bar{x} - \bar{y} \le \delta - t_{m+n-2,\alpha} \cdot s_p\sqrt{\frac{1}{m} + \frac{1}{n}}$
III.	$\mu_1 - \mu_2 = \delta$	$\mu_1 - \mu_2 \ne \delta$	$\lvert\bar{x} - \bar{y} - \delta\rvert \ge z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}$	$\lvert\bar{x} - \bar{y} - \delta\rvert \ge t_{m+n-2,\alpha/2} \cdot s_p\sqrt{\frac{1}{m} + \frac{1}{n}}$

Remark 4. The case of most interest is that in which $\delta = 0$. If σ_1^2, σ_2^2 are unknown and $\sigma_1^2 = \sigma_2^2 = \sigma^2$, σ^2 unknown, then S_p^2 is an unbiased estimate of σ^2. In this case all the two-sample t-tests are UMP unbiased and UMP invariant. Before applying the t-test, one should first make sure that $\sigma_1^2 = \sigma_2^2 = \sigma^2, \sigma^2$ unknown. This means applying another test on the data. We consider this test in the next section.

Remark 5. If $m + n$ is large, we use normal tables; if both m and n are large, we can drop the assumption of normality, using the CLT.

Remark 6. The problem of equality of means in sampling from several populations will be considered in Chapter 12.

Remark 7. The two sample problem when $\sigma_1 \ne \sigma_2$, both unknown, is commonly referred to as Behrens–Fisher problem. The *Welch approximate t-test* of $H_0: \mu_1 = \mu_2$ is based on a random number of d.f. f given by

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

where

$$R = \frac{S_1^2/m}{S_2^2/n},$$

and the t-statistic

$$T = \frac{(\overline{X} - \overline{Y}) - (\mu_1 - \mu_2)}{\sqrt{S_1^2/m + S_2^2/n}}$$

with f d.f. This approximation has been found to be quite good even for small samples. The formula for f generally leads to noninteger d.f. Linear interpolation in t-tables can be used to obtain the required percentiles for f d.f.

Example 2. The mean life of a sample of 9 light bulbs was observed to be 1309 hours with a standard deviation of 420 hours. A second sample of 16 bulbs chosen from a different batch showed a mean life of 1205 hours with a standard deviation of 390 hours. Let us test to see whether there is a significant difference between the means of the two batches, assuming that the population variances are the same (see also Example 10.5.1).

Here $H_0: \mu_1 = \mu_2$, $H_1: \mu_1 \neq \mu_2$, $m = 9$, $n = 16$, $\overline{x} = 1309$, $s_1 = 420$, $\overline{y} = 1205$, $s_2 = 390$, and let us take $\alpha = 0.05$. We have

$$s_p = \sqrt{\frac{8(420)^2 + 15(390)^2}{23}}$$

so that

$$t_{m+n-2,\alpha/2} s_p \sqrt{\frac{1}{m} + \frac{1}{n}} = t_{23,0.025} \sqrt{\frac{8(420)^2 + 15(390)^2}{23}} \sqrt{\frac{1}{9} + \frac{1}{16}} = 345.44.$$

Since $|\overline{x} - \overline{y}| = |1309 - 1205| = 104 \not> 345.44$, we cannot reject H_0 at level $\alpha = 0.05$.

Quite frequently, one samples from a bivariate normal population with means μ_1, μ_2, variances σ_1^2, σ_2^2, and correlation coefficient ρ, the hypothesis of interest being $\mu_1 = \mu_2$. Let $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ be a sample from a bivariate normal distribution with parameters $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$, and ρ. Then $X_j - Y_j$ is $\mathcal{N}(\mu_1 - \mu_2, \sigma^2)$, where $\sigma^2 = \sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2$. We can therefore treat $D_j = (X_j - Y_j)$, $j = 1, 2, \ldots, n$, as a sample from a normal population. Let us write

$$\overline{d} = \frac{\sum_1^n d_i}{n} \quad \text{and} \quad s_d^2 = \frac{\sum_1^n (d_i - \overline{d})^2}{n-1}.$$

The following table summarizes the resulting tests:

	H_0	H_1	Reject H_0 at Level α if:
	(d_0 = known constant)		
I.	$\mu_1 - \mu_2 \geq d_0$	$\mu_1 - \mu_2 < d_0$	$\overline{d} \leq d_0 + \frac{s_d}{\sqrt{n}} t_{n-1,1-\alpha}$
II.	$\mu_1 - \mu_2 \leq d_0$	$\mu_1 - \mu_2 > d_0$	$\overline{d} \geq d_0 + \frac{s_d}{\sqrt{n}} t_{n-1,\alpha}$
III.	$\mu_1 - \mu_2 = d_0$	$\mu_1 - \mu_2 \neq d_0$	$\lvert\overline{d} - d_0\rvert \geq \frac{s_d}{\sqrt{n}} t_{n-1,\alpha/2}$

Remark 8. The case of most importance is that in which $d_0 = 0$. All the t-tests, based on D_j's, are UMP unbiased and UMP invariant. If σ is known, one can base the test on a standardized normal RV, but in practice such an assumption is quite unrealistic. If n is large, one can replace t-values by the corresponding critical values under the normal distribution.

Remark 9. Clearly, it is not necessary to assume that $(X_1, Y_1), \ldots, (X_n, Y_n)$ is a sample from a bivariate normal population. It suffices to assume that the differences D_i form a sample from a normal population.

Example 3. Nine adults agreed to test the efficacy of a new diet program. Their weights (pounds) were measured before and after the program and found to be as follows:

	Participant								
	1	2	3	4	5	6	7	8	9
Before	132	139	126	114	122	132	142	119	126
After	124	141	118	116	114	132	145	123	121

Let us test the null hypothesis that the diet is not effective, $H_0: \mu_1 - \mu_2 = 0$, against the alternative, $H_1: \mu_1 - \mu_2 > 0$, that it is effective at level $\alpha = 0.01$. We compute

$$\bar{d} = \frac{8 - 2 + 8 - 2 + 8 + 0 - 3 - 4 + 5}{9} = \frac{18}{9} = 2,$$

$$s_d^2 = 26.75, \quad \text{and} \quad s_d = 5.17.$$

Thus

$$d_0 + \frac{s_d}{\sqrt{n}} t_{n-1,\alpha} = 0 + \frac{5.17}{\sqrt{9}} t_{8,0.01} = \frac{5.17}{3} \times 2.896 = 4.99$$

Since $\bar{d} \not> 4.99$, we cannot reject hypothesis H_0 that the diet is not very effective.

PROBLEMS 10.4

1. The manufacturer of a certain subcompact car claims that the average mileage of this model is 30 miles per gallon of regular gasoline. For nine cars of this model driven in an identical manner, using 1 gallon of regular gasoline, the mean distance traveled was 26 miles with a standard deviation of 2.8 miles. Test the manufacturer's claim if you are willing to reject a true claim no more than twice in 100.

2. The nicotine contents of five cigarettes of a certain brand showed a mean of 21.2 milligrams with a standard deviation of 20.05 milligrams. Test the hypothesis

that the average nicotine content of this brand of cigarettes does not exceed 19.7 milligrams. Use $\alpha = 0.05$.

3. The additional hours of sleep gained by eight patients in an experiment with a certain drug were recorded as follows:

Patient	1	2	3	4	5	6	7	8
Hours Gained	0.7	−1.1	3.4	0.8	2.0	0.1	−0.2	3.0

Assuming that these patients form a random sample from a population of such patients and that the number of additional hours gained from the drug is a normal random variable, test the hypothesis that the drug has no effect at level $\alpha = 0.10$.

4. The mean life of a sample of 8 light bulbs was found to be 1432 hours with a standard deviation of 436 hours. A second sample of 19 bulbs chosen from a different batch produced a mean life of 1310 hours with a standard deviation of 382 hours. Making appropriate assumptions, test the hypothesis that the two samples came from the same population of light bulbs at level $\alpha = 0.05$.

5. A sample of 25 observations has a mean of 57.6 and a variance of 1.8. A further sample of 20 values has a mean of 55.4 and a variance of 20.5. Test the hypothesis that the two samples came from the same normal population.

6. Two methods were used in a study of the latent heat of fusion of ice. Both method A and method B were conducted with the specimens cooled to −0.72°C. The following data represent the change in total heat from −0.72°C to water, 0°C, in calories per gram of mass:

Method A: 79.98, 80.04, 80.02, 80.04, 80.03, 80.03, 80.04, 79.97, 80.05, 80.03, 80.02, 80.00, 80.02
Method B: 80.02, 79.74, 79.98, 79.97, 79.97, 80.03, 79.95, 79.97

Perform a test at level 0.05 to see whether the two methods differ with regard to their average performance. (Natrella [73, p. 3-23])

7. In Problem 6, if it is known from past experience that the standard deviations of the two methods are $\sigma_A = 0.024$ and $\sigma_B = 0.033$, test the hypothesis that the methods are same with regard to their average performance at level $\alpha = 0.05$.

8. During World War II bacterial polysaccharides were investigated as blood plasma extenders. Sixteen samples of hydrolyzed polysaccharides supplied by various manufacturers in order to assess two chemical methods for determining the average molecular weight yielded the following results:

Method A: 62,700; 29,100; 44,400; 47,800; 36,300; 40,000; 43,400; 35,800; 33,900; 44,200; 34,300; 31,300; 38,400; 47,100; 42,100; 42,200
Method B: 56,400; 27,500; 42,200; 46,800; 33,300; 37,100; 37,300; 36,200; 35,200; 38,000; 32,200; 27,300; 36,100; 43,100; 38,400; 39,900

Perform an appropriate test of the hypothesis that the two averages are the same against a one-sided alternative that the average of method A exceeds that of method B. Use $\alpha = 0.05$. (Natrella [73, p. 3-38])

9. The following grade-point averages were collected over a period of 7 years to determine whether membership in a fraternity is beneficial or detrimental to grades:

	Year						
	1	2	3	4	5	6	7
Fraternity	2.4	2.0	2.3	2.1	2.1	2.0	2.0
Nonfraternity	2.4	2.2	2.5	2.4	2.3	1.8	1.9

Assuming that the populations were normal, test at the 0.025 level of significance whether membership in a fraternity is detrimental to grades.

10. Consider the two-sample t-statistic $T = (\overline{X} - \overline{Y})/[S_p\sqrt{1/m + 1/n}]$, where $S_p^2 = [(m-1)S_1^2 + (n-1)S_2^2]/(m+n-2)$. Suppose that $\sigma_1 \neq \sigma_2$. Let $m, n \to \infty$ such that $m/(m+n) \to \rho$. Show that under $\mu_1 = \mu_2$, $T \xrightarrow{L} U$, where $U \sim \mathcal{N}(0, \tau^2)$ with $\tau^2 = [(1-\rho)\sigma_1^2 + \rho\sigma_2^2]/[\rho\sigma_1^2 + (1-\rho)\sigma_2^2]$. Thus when $m \approx n$, $\rho \approx \frac{1}{2}$ and $\tau^2 \approx 1$, and T is approximately $\mathcal{N}(0,1)$ as $m(\approx n) \to \infty$. In this case, a t-test based on T will have approximately the right level.

10.5 F-TESTS

The term *F-tests* refers to tests based on an F-statistic. Let $X_1, X_2, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$ be independent samples from $\mathcal{N}(\mu_1, \sigma_1^2)$ and $\mathcal{N}(\mu_2, \sigma_2^2)$, respectively. We recall that $\sum_1^m (X_i - \overline{X})/\sigma_1^2 \sim \chi^2(m-1)$ and $\sum_1^n (Y_i - \overline{Y})^2/\sigma_2^2 \sim \chi^2(n-1)$ are independent RVs, so that the RV

$$F(\mathbf{X}, \mathbf{Y}) = \frac{\sum_1^m (X_i - \overline{X})^2}{\sum_1^n (Y_i - \overline{Y})^2} \frac{\sigma_2^2(n-1)}{\sigma_1^2(m-1)} = \frac{\sigma_2^2}{\sigma_1^2} \frac{S_1^2}{S_2^2}$$

is distributed as $F(m-1, n-1)$.

The following table summarizes the F-tests:

			Reject H_0 at Level α if:	
	H_0	H_1	μ_1, μ_2 Known	μ_1, μ_2 Unknown
I.	$\sigma_1^2 \leq \sigma_2^2$	$\sigma_1^2 > \sigma_2^2$	$\dfrac{\sum_1^m (x_i - \mu_1)^2}{\sum_1^n (y_i - \mu_2)^2} \geq \dfrac{m}{n} F_{m,n,\alpha}$	$\dfrac{s_1^2}{s_2^2} \geq F_{m-1,n-1,\alpha}$
II.	$\sigma_1^2 \geq \sigma_2^2$	$\sigma_1^2 < \sigma_2^2$	$\dfrac{\sum_1^n (y_i - \mu_2)^2}{\sum_1^m (x_i - \mu_1)^2} \geq \dfrac{n}{m} F_{n,m,\alpha}$	$\dfrac{s_2^2}{s_1^2} \geq F_{n-1,m-1,\alpha}$

III.	$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 \neq \sigma_2^2$	$\begin{cases} \dfrac{\sum_1^m (x_i - \mu_1)^2}{\sum_1^n (y_i - \mu_2)^2} & \geq \dfrac{m}{n} F_{m,n,\alpha/2} \\ \text{or} & \leq \dfrac{m}{n} F_{m,n,1-\alpha/2} \end{cases}$	$\begin{cases} \dfrac{s_1^2}{s_2^2} & \geq F_{m-1,n-1,\alpha/2} \\ \text{or} & \leq F_{m-1,n-1,1-\alpha/2} \end{cases}$

Remark 1. Recall (Remark 7.4.5) that

$$F_{m,n,1-\alpha} = \{F_{n,m,\alpha}\}^{-1}.$$

Remark 2. The tests described above can easily be obtained from the likelihood ratio procedure. Moreover, in the important case where μ_1, μ_2 are unknown, tests I and II are UMP unbiased and UMP invariant. For test III we have chosen equal tails, as is customarily done for convenience even though the unbiasedness property of the test is thereby destroyed.

***Example 1* (Example 10.4.2 continued).** In Example 10.4.2 let us test the validity of the assumption on which the t-test was based, namely, that the two populations have the same variance at level 0.05. We compute $s_1^2/s_2^2 = (420/390)^2 = 196/169 = 1.16$. Since $F_{m-1,n-1,\alpha/2} = F_{8,15,0.025} = 3.20$, we cannot reject $H_0\colon \sigma_1 = \sigma_2$.

An important application of the F-test involves the case where one is testing the equality of means of two normal populations under the assumption that the variances are the same, that is, testing whether the two samples come from the same population. Let $X_1, X_2, \ldots, X_m$ and $Y_1, Y_2, \ldots, Y_n$ be independent samples from $\mathcal{N}(\mu_1, \sigma_1^2)$ and $\mathcal{N}(\mu_2, \sigma_2^2)$, respectively. If $\sigma_1^2 = \sigma_2^2$ but is unknown, the t-test rejects $H_0\colon \mu_1 = \mu_2$ if $|T| > c$, where c is selected so that $\alpha_2 = P\{|T| > c \mid \mu_1 = \mu_2, \sigma_1 = \sigma_2\}$, that is, $c = t_{m+n-2,\alpha_2/2} s_p \sqrt{(1/m + 1/n)}$, where

$$s_p^2 = \frac{(m-1)s_1^2 + (n-1)s_2^2}{m+n-2},$$

s_1, s_2 being the sample variances. If first an F-test is performed to test $\sigma_1 = \sigma_2$, and then a t-test to test $\mu_1 = \mu_2$ at levels α_1 and α_2, respectively, the probability of accepting both hypotheses when they are true is

$$P\{|T| \leq c, c_1 < F < c_2 | \mu_1 = \mu_2, \sigma_1 = \sigma_2\};$$

and if F is independent of T, this probability is $(1-\alpha_1)(1-\alpha_2)$. It follows that the combined test has a significance level $\alpha = 1 - (1-\alpha_1)(1-\alpha_2)$. We see that

$$\alpha = \alpha_1 + \alpha_2 - \alpha_1\alpha_2 \leq \alpha_1 + \alpha_2$$

and $\alpha \geq \max(\alpha_1, \alpha_2)$. In fact, α will be closer to $\alpha_1 + \alpha_2$, since for small α_1 and α_2, $\alpha_1\alpha_2$ will be closer to 0.

We show that F is independent of T whenever $\sigma_1 = \sigma_2$. The statistic $V = (\overline{X}, \overline{Y}, \sum_1^m (X_i - \overline{X})^2 + \sum_1^n (Y_i - \overline{Y})^2)$ is a complete sufficient statistic for the parameter $(\mu_1, \mu_2, \sigma_1 = \sigma_2)$ (see Theorem 8.3.2). Since the distribution of F does not depend on μ_1, μ_2, and $\sigma_1 = \sigma_2$, it follows (Problem 5) that F is independent of V whenever $\sigma_1 = \sigma_2$. But T is a function of V alone, so that F must be independent of T also.

In Example 1, the combined test has a significance level of

$$\alpha = 1 - (0.95)(0.95) = 1 - 0.9025 = 0.0975.$$

PROBLEMS 10.5

1. For the data of Problem 10.4.4, is the assumption of equality of variances on which the t-test is based, valid?

2. Answer the same question for Problems 10.4.5 and 10.4.6.

3. The performance of each of two different dive-bombing methods is measured a dozen times. The sample variances for the two methods are computed to be 5545 and 4073, respectively. Do the two methods differ in variability?

4. In Problem 3, does the variability of the first method exceed that of the second method?

5. Let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be a random sample from a distribution with PDF (PMF) $f(x, \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$ where Θ is an interval in $\mathcal{R}_k$. Let $T(\mathbf{X})$ be a complete sufficient statistic for the family $\{f(\mathbf{x}; \boldsymbol{\theta}) : \boldsymbol{\theta} \in \Theta\}$. If $U(\mathbf{X})$ is a statistic (not a function of T alone) whose distribution does not depend on $\boldsymbol{\theta}$, show that U is independent of T.

10.6 BAYES AND MINIMAX PROCEDURES

Let $X_1, X_2, \ldots, X_n$ be a sample from a probability distribution with PDF (PMF) f_θ, $\theta \in \Theta$. In Section 8.8 we described the general decision problem, namely, once the statistician observes $\mathbf{x}$, she has a set $\mathcal{A}$ of options available. The problem is to find a decision function d that minimizes the risk $R(\theta, \delta) = E_\theta L(\theta, \delta)$ in some sense. Thus a minimax solution requires the minimization of max $R(\theta, \delta)$, while a Bayes solution requires the minimization of $R(\pi, \delta) = ER(\theta, \delta)$, where π is the a priori distribution on Θ. In Remark 9.2.1 we considered the problem of hypothesis-testing as a special case of the general decision problem. The set $\mathcal{A}$ contains two points, a_0 and a_1; a_0 corresponds to the acceptance of $H_0: \theta \in \Theta_0$, and a_1 corresponds to the rejection of H_0. Suppose that the loss function is defined by

$$(1)\qquad \begin{cases} L(\theta, a_0) = a(\theta) & \text{if } \theta \in \Theta_1, \quad a(\theta) > 0, \\ L(\theta, a_1) = b(\theta) & \text{if } \theta \in \Theta_0, \quad b(\theta) > 0, \\ L(\theta, a_0) = 0 & \text{if } \theta \in \Theta_0, \\ L(\theta, a_1) = 0 & \text{if } \theta \in \Theta_1. \end{cases}$$

Then

$$(2)\qquad R(\theta, \delta(\mathbf{X})) = L(\theta, a_0)P_\theta\{\delta(\mathbf{X}) = a_0\} + L(\theta, a_1)P_\theta\{\delta(\mathbf{X}) = a_1\}$$

$$(3)\qquad = \begin{cases} a(\theta)P_\theta\{\delta(\mathbf{X}) = a_0\} & \text{if } \theta \in \Theta_1, \\ b(\theta)P_\theta\{\delta(\mathbf{X}) = a_1\} & \text{if } \theta \in \Theta_0. \end{cases}$$

A *minimax solution* to the problem of testing $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1$, where $\Theta = \Theta_0 + \Theta_1$, is to find a rule δ that minimizes

$$\max_\theta [a(\theta)P_\theta\{\delta(\mathbf{X}) = a_0\}, \qquad b(\theta)P_\theta\{\delta(\mathbf{X}) = a_1\}].$$

We will consider here only the special case of testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$. In that case we want to find a rule δ that minimizes

$$(4)\qquad \max[aP_{\theta_1}\{\delta(\mathbf{X}) = a_0\}, \qquad bP_{\theta_0}\{\delta(\mathbf{X}) = a_1\}].$$

We will show that the solution is to reject H_0 if

$$(5)\qquad \frac{f_{\theta_1}(\mathbf{x})}{f_{\theta_0}(\mathbf{x})} \geq k,$$

provided that the constant k is chosen so that

$$(6)\qquad R(\theta_0, \delta(\mathbf{X})) = R(\theta_1, \delta(\mathbf{X})),$$

where δ is the rule defined in (5); that is, the minimax rule δ is obtained if we choose k in (5) so that

$$(7)\qquad aP_{\theta_1}\{\delta(\mathbf{X}) = a_0\} = bP_{\theta_0}\{\delta(\mathbf{X}) = a_1\},$$

or, equivalently, we choose k so that

$$(8)\qquad aP_{\theta_1}\left\{\frac{f_{\theta_1}(\mathbf{X})}{f_{\theta_0}(\mathbf{X})} < k\right\} = bP_{\theta_0}\left\{\frac{f_{\theta_1}(\mathbf{X})}{f_{\theta_0}(\mathbf{X})} \geq k\right\}.$$

Let δ^* be any other rule. If $R(\theta_0, \delta) < R(\theta_0, \delta^*)$, then $R(\theta_0, \delta) = R(\theta_1, \delta) < \max[R(\theta_0, \delta^*), R(\theta_1, \delta^*)]$ and δ^* cannot be minimax. Thus $R(\theta_0, \delta) \geq R(\theta_0, \delta^*)$, which means that

$$(9)\qquad P_{\theta_0}\{\delta^*(\mathbf{X}) = a_1\} \leq P_{\theta_0}\{\delta(\mathbf{X}) = a_1\} = P\{\text{reject } H_0 \mid H_0 \text{ true}\}.$$

By the Neyman–Pearson lemma, rule δ is the most powerful of its size, so that its power must be at least that of δ^*, that is,

$$P_{\theta_1}\{\delta(\mathbf{X}) = a_1\} \geq P_{\theta_1}\{\delta^*(\mathbf{X}) = a_1\}$$

so that

$$P_{\theta_1}\{\delta(\mathbf{X}) = a_0\} \leq P_{\theta_1}\{\delta^*(\mathbf{X}) = a_0\}.$$

It follows that

$$aP_{\theta_1}\{\delta(\mathbf{X}) = a_0\} \leq aP_{\theta_1}\{\delta^*(\mathbf{X}) = a_0\}$$

and hence that

$$R(\theta_1, d) \leq R(\theta_1, \delta^*). \tag{10}$$

This means that

$$\max[R(\theta_0, \delta),\ R(\theta_1, \delta)] = R(\theta_1, \delta) \leq R(\theta_1, \delta^*)$$

and thus

$$\max[R(\theta_0, \delta),\ R(\theta_1, \delta)] \leq \max[R(\theta_0, \delta^*),\ R(\theta_1, \delta^*)].$$

Note that in the discrete case one may need some randomization procedure in order to achieve equality in (8).

Example 1. Let $X_1, X_2, \ldots, X_n$ be iid $\mathcal{N}(\mu, 1)$ RVs. To test $H_0\colon \mu = \mu_0$ against $H_1\colon \mu = \mu_1$ ($> \mu_0$), we should choose k so that (8) is satisfied. This is the same as choosing c, and thus k, so that

$$aP_{\mu_1}\{\overline{X} < c\} = bP_{\mu_0}\{\overline{X} \geq c\}$$

or

$$aP_{\mu_1}\left\{\frac{\overline{X} - \mu_1}{1/\sqrt{n}} < \frac{c - \mu_1}{1/\sqrt{n}}\right\} = bP_{\mu_0}\left\{\frac{\overline{X} - \mu_0}{1/\sqrt{n}} \geq \frac{c - \mu_0}{1/\sqrt{n}}\right\}.$$

Thus

$$a\Phi[\sqrt{n}(c - \mu_1)] = b\{1 - \Phi[\sqrt{n}(c - \mu_0)]\},$$

where Φ is the DF of an $\mathcal{N}(0, 1)$ RV. This can easily be accomplished with the help of normal tables once we know a, b, μ_0, μ_1, and n.

We next consider the problem of testing $H_0\colon \theta \in \Theta_0$ against $\mu_1\colon \theta \in \Theta_1$ from a Bayesian point of view. Let $\pi(\theta)$ be the a priori probability distribution on Θ. Then

$$(11)\qquad R(\pi,\delta)=E_\theta R(\theta,\delta(\mathbf{X}))$$
$$=\begin{cases}\int_\Theta R(\theta,\delta)\pi(\theta)d\theta & \text{if } \pi \text{ is a PDF,}\\ \sum_\Theta R(\theta,\delta)\pi(\theta) & \text{if } \pi \text{ is a PMF,}\end{cases}$$
$$=\begin{cases}\int_{\Theta_0} b(\theta)\pi(\theta)P_\theta\{\delta(\mathbf{X})=a_1\}d\theta+ & \\ \qquad \int_{\Theta_1} a(\theta)\pi(\theta)P_\theta\{\delta(\mathbf{X})=a_0\}d\theta & \text{if } \pi \text{ is a PDF,}\\ \sum_{\Theta_0} b(\theta)\pi(\theta)P_\theta\{\delta(\mathbf{X})=a_1\}+ & \\ \qquad \sum_{\Theta_1} a(\theta)\pi(\theta)P_\theta\{\delta(\mathbf{X})=a_0\} & \text{if } \pi \text{ is a PMF.}\end{cases}$$

The Bayes solution is a decision rule that minimizes $R(\pi,\delta)$. In what follows we restrict our attention to the case where both H_0 and H_1 have exactly one point each, that is, $\Theta_0=\{\theta_0\}$, $\Theta_1=\{\theta_1\}$. Let $\pi(\theta_0)=\pi_0$ and $\pi(\theta_1)=1-\pi_0=\pi_1$. Then

$$(12)\qquad R(\pi,\delta)=b\pi_0 P_{\theta_0}\{\delta(\mathbf{X})=a_1\}+a\pi_1 P_{\theta_1}\{\delta(\mathbf{X})=a_0\},$$

where $b(\theta_0)=b$, $a(\theta_1)=a$; $(a,b>0)$.

Theorem 1. Let $\mathbf{X}=(X_1,X_2,\ldots,X_n)$ be an RV of the discrete (continuous) type with PMF (PDF) f_θ, $\theta\in\Theta=\{\theta_0,\theta_1\}$. Let $\pi(\theta_0)=\pi_0$, $\pi(\theta_1)=1-\pi_0=\pi_1$ be the a priori probability mass function on Θ. A Bayes solution for testing $H_0\colon \mathbf{X}\sim f_{\theta_0}$ against $H_1\colon \mathbf{X}\sim f_{\theta_1}$, using the loss function (1), is to reject H_0 if

$$(13)\qquad \frac{f_{\theta_1}(\mathbf{x})}{f_{\theta_0}(\mathbf{x})}\geq\frac{b\pi_0}{a\pi_1}.$$

Proof. We wish to find δ that minimizes

$$R(\pi,\delta)=b\pi_0 P_{\theta_0}\{\delta(\mathbf{X})=a_1\}+a\pi_1 P_{\theta_1}\{\delta(\mathbf{X})=a_0\}.$$

Now

$$R(\pi,\delta)=E_\theta R(\theta,\delta)$$
$$=E\{E_\theta\{L(\theta,\delta)|\mathbf{X}\}\},$$

so it suffices to minimize $E_\theta\{L(\theta,\delta)|\mathbf{X}\}$.

The a posteriori distribution of θ is given by

$$(14)\qquad h(\theta|\mathbf{x})=\frac{\pi(\theta)f_\theta(\mathbf{x})}{\sum_\theta f_\theta(\mathbf{x})\pi(\theta)}$$
$$=\frac{\pi(\theta)f_\theta(\mathbf{x})}{\pi_0 f_{\theta_0}(\mathbf{x})+\pi_1 f_{\theta_1}(\mathbf{x})}$$

$$= \begin{cases} \dfrac{\pi_0 f_{\theta_0}(\mathbf{x})}{\pi_0 f_{\theta_0}(\mathbf{x}) + \pi_1 f_{\theta_1}(\mathbf{x})} & \text{if } \theta = \theta_0, \\ \dfrac{\pi_1 f_{\theta_1}(\mathbf{x})}{\pi_0 f_{\theta_0}(\mathbf{x}) + \pi_1 f_{\theta_1}(\mathbf{x})} & \text{if } \theta = \theta_1. \end{cases}$$

Thus

$$E_\theta\{L(\theta, \delta(\mathbf{X}))|\mathbf{X} = \mathbf{x}\} = \begin{cases} bh(\theta_0|\mathbf{x}), & \theta = \theta_0, \delta(\mathbf{X}) = a_1, \\ ah(\theta_1|\mathbf{x}), & \theta = \theta_1, \delta(\mathbf{X}) = a_0, \end{cases}$$

It follows that we reject H_0, that is, $\delta(\mathbf{X}) = a_1$ if

$$bh(\theta_0|\mathbf{x}) \le ah(\theta_1|\mathbf{x}),$$

which is the case if and only if

$$b\pi_0 f_{\theta_0}(\mathbf{x}) \le a\pi_1 f_{\theta_1}(\mathbf{x}),$$

as asserted.

Remark 1. In the Neyman–Pearson lemma we fixed $P_{\theta_0}\{\delta(\mathbf{X}) = a_1\}$, the probability of rejecting H_0 when it is true, and minimized $P_{\theta_1}\{\delta(\mathbf{X}) = a_0\}$, the probability of accepting H_0 when it is false. Here we no longer have a fixed level α for $P_{\theta_0}\{\delta(\mathbf{X}) = a_1\}$. Instead, we allow it to assume any value as long as $R(\pi, \delta)$, defined in (12), is minimum.

Remark 2. It is easy to generalize Theorem 1 to the case of multiple decisions. Let $\mathbf{X}$ be an RV with PDF (PMF) f_θ, where θ can take any of the k values $\theta_1, \theta_2, \ldots, \theta_k$. The problem is to observe $\mathbf{x}$ and decide which of the θ_i's is the correct value of θ. Let us write $H_i: \theta = \theta_i$, $i = 1, 2, \ldots, k$, and assume that $\pi(\theta_i) = \pi_i$, $i = 1, 2, \ldots k$, $\sum_1^k \pi_i = 1$, is the prior probability distribution on $\Theta = \{\theta_1, \theta_2, \ldots, \theta_k\}$. Let

$$L(\theta_i, \delta) = \begin{cases} 1 & \text{if } \delta \text{ chooses } \theta_j,\ j \ne i. \\ 0 & \text{if } \delta \text{ chooses } \theta_i. \end{cases}$$

The problem is to find a rule δ that minimizes $R(\pi, \delta)$. We leave the reader to show that a Bayes solution is to accept $H_i: \theta = \theta_i$ $(i = 1, 2, \ldots, k)$ if

(15) $$\pi_i f_{\theta_i}(\mathbf{x}) \ge \pi_j f_{\theta_j}(\mathbf{x}) \qquad \text{for all } j \ne i,\ j = 1, 2, \ldots, k,$$

where any point lying in more than one such region is assigned to any one of them.

Example 2. Let $X_1, X_2, \ldots, X_n$ be iid $\mathcal{N}(\mu, 1)$ RVs. To test $H_0: \mu = \mu_0$ against $H_1: \mu = \mu_1$ $(> \mu_0)$, let us take $a = b$ in the loss function (1). Then Theorem 1 says that the Bayes rule is one that rejects H_0 if

$$\frac{f_{\theta_1}(\mathbf{x})}{f_{\theta_0}(\mathbf{x})} \geq \frac{\pi_0}{1-\pi_0},$$

that is,

$$\exp\left[-\frac{\sum_1^n (x_i-\mu_1)^2}{2} + \frac{\sum_1^n (x_i-\mu_0)^2}{2}\right] \geq \frac{\pi_0}{1-\pi_0}$$

and

$$\exp\left[(\mu_1-\mu_0)\sum_1^n x_i + \frac{n(\mu_0^2-\mu_1^2)}{2}\right] \geq \frac{\pi_0}{1-\pi_0}.$$

This happens if and only if

$$\frac{1}{n}\sum_1^n x_i \geq \frac{1}{n}\frac{\log[\pi_0/(1-\pi_0)]}{\mu_1-\mu_0} + \frac{\mu_0+\mu_1}{2},$$

where the logarithm is to the base e. It follows that, if $\pi_0 = \frac{1}{2}$, the rejection region consists of

$$\overline{x} \geq \frac{\mu_0+\mu_1}{2}.$$

Example 3. This example illustrates the result described in Remark 2. Let $X_1, X_2, \ldots, X_n$ be a sample from $\mathcal{N}(\mu, 1)$, and suppose that μ can take any one of the three values μ_1, μ_2, or μ_3. Let $\mu_1 < \mu_2 < \mu_3$. Assume, for simplicity, that $\pi_1 = \pi_2 = \pi_3$. Then we accept $H_i: \mu = \mu_i$, $i = 1, 2, 3$, if

$$\pi_i \exp\left[-\sum_{k=1}^n \frac{(x_k-\mu_i)^2}{2}\right] \geq \pi_j \exp\left[-\sum_{k=1}^n \frac{(x_k-\mu_j)^2}{2}\right]$$
$$\text{for each } j \neq i,\ j = 1, 2, 3.$$

It follows that we accept H_i if

$$(\mu_i - \mu_j)\overline{x} + \frac{\mu_j^2-\mu_i^2}{2} \geq 0, \qquad j = 1, 2, 3 \quad (j \neq i),$$

that is,

$$\overline{x}(\mu_i-\mu_j) \geq \frac{(\mu_i-\mu_j)(\mu_i+\mu_j)}{2}, \qquad j = 1, 2, 3 \quad (j \neq i).$$

Thus the acceptance region of H_1 is given by

$$\bar{x} \le \frac{\mu_1 + \mu_2}{2} \quad \text{and} \quad \bar{x} \le \frac{\mu_1 + \mu_3}{2}.$$

Also, the acceptance region of H_2 is given by

$$\bar{x} \ge \frac{\mu_1 + \mu_2}{2} \quad \text{and} \quad \bar{x} \le \frac{\mu_2 + \mu_3}{2}$$

and that of H_3 by

$$\bar{x} \ge \frac{\mu_1 + \mu_3}{2} \quad \text{and} \quad \bar{x} \ge \frac{\mu_2 + \mu_3}{2}.$$

In particular, if $\mu_1 = 0, \mu_2 = 2, \mu_3 = 4$, we accept H_1 if $\bar{x} \le 1$, H_2 if $1 \le \bar{x} \le 3$, and H_3 if $\bar{x} \ge 3$. In this case, boundary points 1 and 3 have zero probability, and it does not matter where we include them.

PROBLEMS 10.6

1. In Example 1, let $n = 15$, $\mu_0 = 4.7$, and $\mu_1 = 5.2$, and choose $a = b > 0$. Find the minimax test and compute its power at $\mu = 4.7$ and $\mu = 5.2$.
2. A sample of five observations is taken on a $b(1, \theta)$ RV to test $H_0: \theta = \frac{1}{2}$ against $H_1: \theta = \frac{3}{4}$.
   (a) Find the most powerful test of size $\alpha = 0.05$.
   (b) If $L(\frac{1}{2}, \frac{1}{2}) = L(\frac{3}{4}, \frac{3}{4}) = 0$, $L(\frac{1}{2}, \frac{3}{4}) = 1$, and $L(\frac{3}{4}, \frac{1}{2}) = 2$, find the minimax rule.
   (c) If the prior probabilities of $\theta = \frac{1}{2}$ and $\theta = \frac{3}{4}$ are $\pi_0 = \frac{1}{3}$ and $\pi_1 = \frac{2}{3}$, respectively, find the Bayes rule.
3. A sample of size n is to be used from the PDF

$$f_\theta(x) = \theta e^{-\theta x}, \qquad x > 0,$$

   to test $H_0: \theta = 1$ against $H_1: \theta = 2$. If the a priori distribution on θ is $\pi_0 = \frac{2}{3}$, $\pi_1 = \frac{1}{3}$, and $a = b$, find the Bayes solution. Find the power of the test at $\theta = 1$ and $\theta = 2$.
4. Given two normal densities with variances 1 and with means -1 and 1, respectively, find the Bayes solution based on a single observation when $a = b$ and (a) $\pi_0 = \pi_1 = \frac{1}{2}$, and (b) $\pi_0 = \frac{1}{4}, \pi_1 = \frac{3}{4}$.
5. Given three normal densities with variances 1 and with means $-1, 0, 1$, respectively, find the Bayes solution to the multiple decision problem based on a single observation when $\pi_1 = \frac{2}{5}, \pi_2 = \frac{2}{5}, \pi_3 = \frac{1}{5}$.
6. For the multiple decision problem described in Remark 2, show that a Bayes solution is to accept $H_i: \theta = \theta_i$ $(i = 1, 2, \ldots, k)$ if (15) holds.