

Chapter 17

Permutation and Randomization Tests



17.1 Introduction

In this chapter, and Chapter 18, we shall deal with situations where both the null hypothesis and the class of alternatives may be nonparametric and so, as a result, it may be difficult even to construct tests (or confidence regions) that satisfactorily control the level (exactly or asymptotically). For such situations, we shall develop methods which achieve this modest goal under fairly general assumptions. A secondary aim will then be to obtain some idea of the power of the resulting tests.

In Section 17.2, we consider the class of randomization tests as a generalization of permutation tests. Under the randomization hypothesis (see Definition 17.2.1), the empirical distribution of the values of a given statistic recomputed over transformations of the data serves as a null distribution; this leads to exact control of the level in such models. When the randomization hypothesis holds, the construction applies broadly to any statistic. The appeal of these methods in semiparametric or nonparametric problems stems from their finite-sample validity, but their validity may not hold when the randomization hypothesis fails. Asymptotic analysis allows one to study the robustness of the methods when assumptions may not hold, and it also allows one to study power properties of the methods. Asymptotic efficiency properties ensue if the statistic is chosen appropriately. Section 17.3 discusses two-sample permutation tests in some depth. Further examples are provided in Section 17.4. Section 17.5 extends the use of randomization tests to problems in multiple testing.

17.2 Permutation and Randomization Tests

Permutation tests were introduced in Chapter 5 as a robust means of controlling the level of a test if the underlying parametric model only holds approximately. For example, the two-sample permutation t -test for testing equality of means studied in

Section 5.11 of Chapter 5 has level α whenever the two populations have the same distribution under the null hypothesis (without the assumption of normality). In this section, we consider the large-sample behavior of permutation tests and, more generally, randomization tests. The use of the term randomization here is distinct from its meaning in Section 5.10. There, randomization was used as a device prior to collecting data, for example, by randomly assigning experimental units to treatment or control. Such a device allows for a meaningful comparison after the data has been observed, by considering the behavior of a statistic recomputed over permutations in the data. Thus, the term randomization referred to both the experimental design and the analysis of data by recomputing a statistic over permutations or randomizations (sometimes called rerandomizations) of the data. It is this latter use of randomization that we now generalize. Thus, the term randomization test will refer to tests obtained by recomputing a test statistic over transformations (not necessarily permutations) of the data.

A general test construction will be presented that yields an exact level α test for a fixed sample size, under a certain group invariance hypothesis. Then, two main questions will be addressed. First, we shall consider the robustness of the level. For example, in the two-sample problem just mentioned, the underlying populations may have the same mean under the null hypothesis, but differ in other ways, as in the classical Behrens–Fisher problem, where the underlying populations are normal but may not have the same variance. Then, the rejection probability under such populations is no longer α , and it becomes necessary to investigate the behavior of the rejection probability. In addition, we also consider the large-sample power of permutation and randomization tests. In the two-sample problem when the underlying populations are normal with common variance, for example, we should like to know whether there is a significant loss in power when using a permutation test as compared to the UMPU t -test.

17.2.1 The Basic Construction

Based on data X taking values in a sample space \mathcal{X} , it is desired to test the null hypothesis H that the underlying probability law P generating X belongs to a certain family Ω_0 of distributions. Let \mathbf{G} be a finite group of transformations g of \mathcal{X} onto itself. The following assumption, which we will call the *randomization hypothesis*, allows for a general test construction.

Definition 17.2.1 (Randomization Hypothesis) Under the null hypothesis, the distribution of X is invariant under the transformations in \mathbf{G} , that is, for every g in \mathbf{G} , gX and X have the same distribution whenever X has distribution P in Ω_0 .

The randomization hypothesis asserts that the null hypothesis parameter space Ω_0 remains invariant under g in \mathbf{G} . However, here we specifically do not require the alternative hypothesis parameter space to remain invariant (unlike what was assumed in Chapter 6).

As an example, consider testing the equality of distributions based on two independent samples (Y_1, \dots, Y_m) and (Z_1, \dots, Z_n) , which was previously considered in Sections 5.8–5.11. Under the null hypothesis that the samples are generated from the same probability law, the observations can be permuted or assigned at random to either of the two groups, and the distribution of the permuted samples is the same as the distribution of the original samples. (Note that a test that is invariant with respect to all permutations of the data would be useless here.)

To describe the general construction of a randomization test, let $T(X)$ be any real-valued test statistic for testing H . Suppose the group \mathbf{G} has M elements. Given $X = x$, let

$$T^{(1)}(x) \leq T^{(2)}(x) \leq \dots \leq T^{(M)}(x)$$

be the ordered values of $T(gx)$ as g varies in \mathbf{G} . Fix a nominal level α , $0 < \alpha < 1$, and let k be defined by

$$k = M - [M\alpha], \quad (17.1)$$

where $[M\alpha]$ denotes the largest integer less than or equal to $M\alpha$. Let $M^+(x)$ and $M^0(x)$ be the number of values $T^{(j)}(x)$ ($j = 1, \dots, M$) which are greater than $T^{(k)}(x)$ and equal to $T^{(k)}(x)$, respectively. Set

$$a(x) = \frac{M\alpha - M^+(x)}{M^0(x)}.$$

Since \mathbf{G} is a group, $T^{(k)}(gx) = T^{(k)}(x)$, and similarly for the functions $M^+(\cdot)$, $M^0(\cdot)$, and $a(\cdot)$.

Generalizing the construction presented in Section 5.8, define the randomization test function $\phi(x)$ to be equal to 1, $a(x)$, or 0 according to whether $T(x) > T^{(k)}(x)$, $T(x) = T^{(k)}(x)$, or $T(x) < T^{(k)}(x)$, respectively. By construction, for every x in \mathcal{X} ,

$$\sum_{g \in \mathbf{G}} \phi(gx) = M^+(x) + a(x)M^0(x) = M\alpha. \quad (17.2)$$

The following theorem shows that the resulting test is level α , under the hypothesis that X and gX have the same distribution whenever the distribution of X is in Ω_0 . Note that this result is true for *any* choice of test statistic T .

Theorem 17.2.1 *Suppose X has distribution P on \mathcal{X} and the problem is to test the null hypothesis $P \in \Omega_0$. Let \mathbf{G} be a finite group of transformations of \mathcal{X} onto itself. Suppose the randomization hypothesis holds, so that, for every $g \in \mathbf{G}$, X and gX have the same distribution whenever X has a distribution P in Ω_0 . Given a test statistic $T = T(X)$, let ϕ be the randomization test as described above. Then,*

$$E_P[\phi(X)] = \alpha \quad \text{for all } P \in \Omega_0. \quad (17.3)$$

PROOF. To prove (17.3), by (17.2),

$$M\alpha = E_P\left[\sum_g \phi(gX)\right] = \sum_g E_P[\phi(gX)].$$

By hypothesis $E_P[\phi(gX)] = E_P[\phi(X)]$, so that

$$M\alpha = \sum_g E_P[\phi(X)] = M E_P[\phi(X)],$$

and the result follows. ■

To gain further insight as to why the construction works, for any $x \in \mathcal{X}$, let \mathbf{G}^x denote the \mathbf{G} -orbit of x ; that is,

$$\mathbf{G}^x = \{gx : g \in \mathbf{G}\}.$$

Recall from Section 6.2 that these orbits partition the sample space. The hypothesis in Theorem 17.2.1 implies that the conditional distribution of X given $X \in \mathbf{G}^x$ is uniform on \mathbf{G}^x , as will be seen in the next theorem. Since this conditional distribution is the same for all $P \in \Omega_0$, a test can be constructed to be level α conditionally, which is then level α unconditionally as well. Because the event $\{X \in \mathbf{G}^x\}$ typically has probability zero for all x , we need to be careful about how we state a result. As x varies, the sets \mathbf{G}^x form a partition of the sample space. Let \mathcal{G} be the σ -field generated by this partition.

Theorem 17.2.2 *Under the null hypothesis of Theorem 17.2.1, for any real-valued statistic $T = T(X)$, any $P \in \Omega_0$, and any Borel subset B of the real line,*

$$P\{T(X) \in B | X \in \mathcal{G}\} = M^{-1} \sum_g I\{T(gX) \in B\} \quad (17.4)$$

with probability one under P . In particular, if the M values of $T(gx)$ as g varies in \mathbf{G} are all distinct, then the uniform distribution on these M values serves as a conditional distribution of $T(X)$ given that $X \in \mathbf{G}^x$.

PROOF. First, we claim that, for any $g \in \mathbf{G}$ and $E \in \mathcal{G}$, $gE = E$. To see why, assume $y \in E$. Then, $g^{-1}y \in E$, because $g^{-1}y$ is on the same orbit as y . Then, $gg^{-1}y \in gE$ or $y \in gE$. A similar argument shows that, if $y \in gE$, then $y \in E$, so that $gE = E$. Now, the right-hand side of (17.4) is clearly \mathcal{G} -measurable, since the right-hand side is constant on any orbit. We need to prove, for any $E \in \mathcal{G}$,

$$\int_E M^{-1} \sum_g I\{T(gx) \in B\} dP(x) = P\{T(X) \in B, X \in E\}.$$

But, the left-hand side is

$$\begin{aligned}
 M^{-1} \sum_g \int_E I\{T(gx) \in B\} dP(x) &= M^{-1} \sum_g P\{T(gX) \in B, X \in E\} \\
 &= M^{-1} \sum_g P\{T(gX) \in B, gX \in gE\} = M^{-1} \sum_g P\{T(gX) \in B, gX \in E\},
 \end{aligned}$$

since $gE = E$. Hence, this last expression becomes (by the randomization hypothesis)

$$M^{-1} \sum_g P\{T(X) \in B, X \in E\} = P\{T(X) \in B, X \in E\},$$

as was to be shown. ■

Example 17.2.1 (One-Sample Tests) Let $X = (X_1, \dots, X_n)$, where the X_i are i.i.d. real-valued random variables. Suppose that, under the null hypothesis, the distribution of the X_i is symmetric about 0. This applies, for example, to the parametric normal location model when the null hypothesis specifies the mean is 0, but it also applies to the nonparametric model that consists of all distributions with the null hypothesis specifying the underlying distribution is symmetric about 0. For $i = 1, \dots, n$, let ϵ_i take on either the value 1 or -1 . Consider a transformation $g = (\epsilon_1, \dots, \epsilon_n)$ of \mathbf{R}^n that takes $x = (x_1, \dots, x_n)$ to $(\epsilon_1 x_1, \dots, \epsilon_n x_n)$. Finally, let \mathbf{G} be the $M = 2^n$ collection of such transformations. Then, the randomization hypothesis holds, i.e., X and gX have the same distribution under the null hypothesis. ■

Example 17.2.2 (Two-Sample Tests) Suppose Y_1, \dots, Y_m are i.i.d. observations from a distribution P_Y and, independently, Z_1, \dots, Z_n are i.i.d. observations from a distribution P_Z . Here, $X = (Y_1, \dots, Y_m, Z_1, \dots, Z_n)$. Suppose that, under the null hypothesis, $P_Y = P_Z$. This applies, for example, to the parametric normal two-sample problem for testing equality of means when the populations have a common (possibly unknown) variance. Alternatively, it also applies to the parametric normal two-sample problem where the null hypothesis is that the means and variances are the same, but under the alternative either the means or the variances may differ; this model was advocated by Fisher (1935a, pp. 122–124). Lastly, this setup also applies to the nonparametric model where P_Y and P_Z may vary freely, but the null hypothesis is that $P_Y = P_Z$. To describe an appropriate \mathbf{G} , let $N = m + n$. For $x = (x_1, \dots, x_N) \in \mathbf{R}^N$, let $gx \in \mathbf{R}^N$ be defined by $(x_{\pi(1)}, \dots, x_{\pi(N)})$, where $(\pi(1), \dots, \pi(N))$ is a permutation of $\{1, \dots, N\}$. Let \mathbf{G} be the collection of all such g , so that $M = N!$. Whenever $P_Y = P_Z$, X and gX have the same distribution. In essence, each transformation g produces a new data set gx , of which the first m elements are used as the Y sample and the remaining n as the Z sample to recompute the test statistic. Note that, if a test statistic is chosen that is invariant under permutations within each of the Y and Z samples (which makes sense by sufficiency), it is enough to consider the $\binom{N}{m}$ transformed data sets obtained by taking m observations from all N as the Y observations

and the remaining n as the Z observations (which, of course, is equivalent to using a subgroup \mathbf{G}' of \mathbf{G}).

As a special case, suppose the observations are real-valued and the underlying distribution is assumed continuous. Suppose T is any statistic that is a function of the ranks of the combined observations, so that T is a *rank statistic* (previously studied in Sections 6.8 and 6.9). The randomization (or permutation) distribution can be obtained by recomputing T over all permutations of the ranks. In this sense, rank tests are special cases of permutation tests. ■

Example 17.2.3 (Tests of Independence) Suppose that X consists of i.i.d. random vectors $X = ((Y_1, Z_1), \dots, (Y_n, Z_n))$ having common joint distribution P and marginal distributions P_Y and P_Z . Assume, under the null hypothesis, Y_i and Z_i are independent, so that P is the product of P_Y and P_Z . This applies to the parametric bivariate normal model when testing that the correlation is zero, but it also applies to the nonparametric model when the null hypothesis specifies Y_i and Z_i are independent with arbitrary marginal distributions. To describe an appropriate \mathbf{G} , let $(\pi(1), \dots, \pi(n))$ be a permutation of $\{1, \dots, n\}$. Let g be the transformation that takes $((y_1, z_1), \dots, (y_n, z_n))$ to the value $((y_1, z_{\pi(1)}), \dots, (y_n, z_{\pi(n)}))$. Let \mathbf{G} be the collection of such transformations, so that $M = n!$. Whenever Y_i and Z_i are independent, X and gX have the same distribution. ■

In general, one can define a p -value \hat{p} of a randomization test by

$$\hat{p} = \frac{1}{M} \sum_g I\{T(gX) \geq T(X)\}. \tag{17.5}$$

It can be shown (Problem 17.2) that \hat{p} satisfies, under the null hypothesis,

$$P\{\hat{p} \leq u\} \leq u \quad \text{for all } 0 \leq u \leq 1. \tag{17.6}$$

Therefore, the nonrandomized test that rejects when $\hat{p} \leq \alpha$ is level α .

Because \mathbf{G} may be large, one may resort to an approximation to construct the randomization test, for example, by randomly sampling transformations g from \mathbf{G} with or without replacement. In the former case, for example, suppose g_1, \dots, g_{B-1} are i.i.d. and uniformly distributed on \mathbf{G} . Let

$$\tilde{p} = \frac{1}{B} \left[1 + \sum_{i=1}^{B-1} I\{T(g_i X) \geq T(X)\} \right]. \tag{17.7}$$

Then, it can be shown (Problem 17.3) that, under the null hypothesis,

$$P\{\tilde{p} \leq u\} \leq u \quad \text{for all } 0 \leq u \leq 1, \tag{17.8}$$

where this probability reflects variation in both X and the sampling of the g_i . Note that (17.8) holds for any B , and so the test that rejects when $\tilde{p} \leq \alpha$ is level α even

when a stochastic approximation is employed. Of course, the larger the value of B , the closer \hat{p} and \tilde{p} are to each other; in fact, $\hat{p} - \tilde{p} \rightarrow 0$ in probability as $B \rightarrow \infty$ (Problem 17.4). Approximations based on auxiliary randomization (such as the sampling of g_i) are known as stochastic approximations.

17.2.2 Asymptotic Results

We next study the limiting behavior of the randomization test in order to derive its large-sample power properties. For example, for testing the mean of a normal distribution is zero with unspecified variance, one would use the optimal t -test. But if we use the randomization test based on the transformations in Example 17.2.1, we will find that the randomization test has the same limiting power as the t -test against contiguous alternatives, and so is LAUMP. Of course, for testing the mean, the randomization test can be used without the assumption of normality, and we will study its asymptotic properties both when the underlying distribution is symmetric so that the randomization hypothesis holds, and also when the randomization hypothesis fails.

Consider a sequence of situations with $X = X^n$, $P = P_n$, $\mathcal{X} = \mathcal{X}_n$, $\mathbf{G} = \mathbf{G}_n$, $T = T_n$, etc. defined for $n = 1, 2, \dots$; notice we use a superscript for the data $X = X^n$. Typically, $X = X^n = (X_1, \dots, X_n)$ consists of n i.i.d. observations and the goal is to consider the behavior of the randomization test sequence as $n \rightarrow \infty$.

Let \hat{R}_n denote the *randomization distribution* of T_n defined by

$$\hat{R}_n(t) = M_n^{-1} \sum_{g \in \mathbf{G}_n} I\{T_n(gX^n) \leq t\}. \tag{17.9}$$

We seek the limiting behavior of $\hat{R}_n(\cdot)$ and its $1 - \alpha$ quantile, which we now denote $\hat{r}_n(1 - \alpha)$ (but in the previous subsection was denoted by $T^{(k)}(X)$); thus,

$$\hat{r}_n(1 - \alpha) = \hat{R}_n^{-1}(1 - \alpha) = \inf\{t : \hat{R}_n(t) \geq 1 - \alpha\}.$$

We will study the behavior of \hat{R}_n under the null hypothesis and under a sequence of alternatives. First, observe that

$$E[\hat{R}_n(t)] = P\{T_n(G_n X^n) \leq t\},$$

where G_n is a random variable that is uniform on \mathbf{G}_n . So, in the case the randomization hypothesis holds, $G_n X^n$ and X^n have the same distribution and so

$$E[\hat{R}_n(t)] = P\{T_n(X^n) \leq t\}.$$

Then, if T_n converges in distribution to a c.d.f. $R(\cdot)$ which is continuous at t , it follows that

$$E[\hat{R}_n(t)] \rightarrow R(t) .$$

In order to deduce $\hat{R}_n(t) \xrightarrow{P} R(t)$ (i.e., the randomization distribution asymptotically approximates the unconditional distribution of T_n), it is then enough to show $Var[\hat{R}_n(t)] \rightarrow 0$. This approach for proving consistency of $\hat{R}_n(t)$ and $\hat{r}_n(1 - \alpha)$ is used in the following result. The sufficiency part is due to Hoeffding (1952), and the necessity part is from Chung and Romano (2013). Note that the randomization hypothesis is not assumed.

Theorem 17.2.3 *Suppose X^n has distribution P_n in \mathcal{X}_n , and \mathbf{G}_n is a finite group of transformations from \mathcal{X}_n to \mathcal{X}_n . Let G_n be a random variable that is uniform on \mathbf{G}_n . Also, let G'_n have the same distribution as G_n , with X^n , G_n , and G'_n mutually independent.*

(i) *Suppose, under P_n ,*

$$(T_n(G_n X^n), T_n(G'_n X^n)) \xrightarrow{d} (T, T') , \tag{17.10}$$

where T and T' are independent, each with common c.d.f. $R(\cdot)$. Then, under P_n ,

$$\hat{R}_n(t) \xrightarrow{P} R(t) \tag{17.11}$$

for every t which is a continuity point of $R(\cdot)$. Let

$$r(1 - \alpha) = \inf\{t : R(t) \geq 1 - \alpha\} .$$

Suppose $R(\cdot)$ is continuous and strictly increasing at $r(1 - \alpha)$. Then, under P_n ,

$$\hat{r}_n(1 - \alpha) \xrightarrow{P} r(1 - \alpha) .$$

(ii) *Conversely, if (17.11) holds for some limiting c.d.f. $R^T(\cdot)$ whenever t is a continuity point, then (17.10) holds.*

PROOF. To prove (i), let t be a continuity point of $R(\cdot)$. Then,

$$E_{P_n}[\hat{R}_n(t)] = P_n\{T_n(G_n X^n) \leq t\} \rightarrow R(t) ,$$

by the convergence hypothesis (17.10). It therefore suffices to show that $Var_{P_n}[\hat{R}_n(t)] \rightarrow 0$ or, equivalently, that

$$E_{P_n}[\hat{R}_n^2(t)] \rightarrow R^2(t) .$$

But,

$$\begin{aligned} E_{P_n}[\hat{R}_n^2(t)] &= M_n^{-2} \sum_g \sum_{g'} P_n\{T_n(gX^n) \leq t, T_n(g'X^n) \leq t\} \\ &= P_n\{T_n(G_nX^n) \leq t, T_n(G'_nX^n) \leq t\} \rightarrow R^2(t), \end{aligned}$$

again by the convergence hypothesis (17.10). Hence, $\hat{R}_n(t) \rightarrow R(t)$ in P_n -probability. The convergence of $\hat{f}_n(1 - \alpha)$ now follows from Problem 11.30, which is a generalization of Lemma 11.2.1.

To prove (ii), let s and t be continuity points of $R^T(\cdot)$. Then,

$$\begin{aligned} P\{T_n(G_nX^n) \leq s, T_n(G'_nX^n) \leq t\} &= E[P\{T_n(G_nX^n) \leq s, T_n(G'_nX^n) \leq t | X^n\}] \\ &= E[\hat{R}_n^T(s)\hat{R}_n^T(t)] \rightarrow R^T(s)R^T(t), \end{aligned}$$

since convergence in probability of a bounded sequence of random variables implies convergence of moments. But, convergence in the plane for a dense set of rectangles entails weak convergence. ■

Note that, if the randomization hypothesis holds, then $T_n(X^n)$ and $T_n(G_nX^n)$ have the same distribution. Assumption (17.10) then implies the unconditional distribution of $T_n(X^n)$ under P_n converges to R in distribution. The conclusion is that the randomization distribution approximates this (unconditional) limit distribution in the sense that (17.11) holds.

Example 17.2.4 (One-Sample Test, continuation of Example 17.2.1) In Example 17.2.1, first consider $T_n = n^{1/2}\bar{X}_n$. If P denotes the common distribution of the X_i , then $P_n = P^n$ is the joint distribution of the sample. Let P be any distribution with mean 0 and finite nonzero variance $\sigma^2(P)$ (not necessarily symmetric). We will verify (17.10) with $R(t) = \Phi(t/\sigma(P))$. Let $\epsilon_1, \dots, \epsilon_n, \epsilon'_1, \dots, \epsilon'_n$ be mutually independent random variables, each 1 or -1 with probability $\frac{1}{2}$ each. We must find the limiting distribution of

$$n^{-1/2} \sum_i (\epsilon_i X_i, \epsilon'_i X_i).$$

But, the vectors $(\epsilon_i X_i, \epsilon'_i X_i)$, $1 \leq i \leq n$, are i.i.d. with

$$E_P(\epsilon_i X_i) = E_P(\epsilon'_i X_i) = E(\epsilon_i)E_P(X_i) = 0,$$

$$E_P[(\epsilon_i X_i)^2] = E(\epsilon_i^2)E_P(X_i^2) = \sigma^2(P) = E_P[(\epsilon'_i X_i)^2],$$

and

$$Cov_P(\epsilon_i X_i, \epsilon'_i X_i) = E_P(\epsilon_i \epsilon'_i X_i^2) = E(\epsilon_i)E(\epsilon'_i)E_P(X_i^2) = 0.$$

By the bivariate Central Limit Theorem,

$$n^{-1/2} \sum_i (\epsilon_i X_i, \epsilon'_i X_i) \xrightarrow{d} (T, T'),$$

where T and T' are independent, each distributed as $N(0, \sigma^2(P))$. Hence, by Theorem 17.2.3, we conclude

$$\hat{R}_n(t) \xrightarrow{P} \Phi(t/\sigma(P))$$

and

$$\hat{r}_n(1 - \alpha) \xrightarrow{P} \sigma(P)z_{1-\alpha}.$$

Let ϕ_n be the randomization test which rejects when $T_n > \hat{r}_n(1 - \alpha)$, accepts when $T_n < \hat{r}_n(1 - \alpha)$ and possibly randomizes when $T_n = \hat{r}_n(1 - \alpha)$. Since T_n is asymptotically normal, it follows by Slutsky's Theorem that

$$E_P(\phi_n) = P\{T_n > \hat{r}_n(1 - \alpha)\} + o(1) \rightarrow P\{\sigma(P)Z > \sigma(P)z_{1-\alpha}\} = \alpha,$$

where Z denotes a standard normal variable. In other words, we have deduced the following for the problem of testing the mean of P is zero versus the mean exceeds zero. By Theorem 17.2.1, ϕ_n is exact level α if the underlying distribution is symmetric about 0; otherwise, it is at least asymptotically pointwise level α as long as the variance is finite.

We now investigate the asymptotic power of ϕ_n against the sequence of alternatives that the observations are $N(hn^{-1/2}, \sigma^2)$. By the above, under $N(0, \sigma^2)$, $\hat{r}_n(1 - \alpha) \rightarrow \sigma z_{1-\alpha}$ in probability. By contiguity, it follows that, under $N(hn^{-1/2}, \sigma^2)$, $\hat{r}_n(1 - \alpha) \rightarrow \sigma z_{1-\alpha}$ in probability as well. Under $N(hn^{-1/2}, \sigma^2)$, T_n is $N(h, \sigma^2)$. Therefore, by Slutsky's Theorem, the limiting power of ϕ_n against $N(hn^{-1/2}, \sigma^2)$ is then

$$E_{P_n}(\phi_n) \rightarrow P\{\sigma Z + h > \sigma z_{1-\alpha}\} = 1 - \Phi\left(z_{1-\alpha} - \frac{h}{\sigma}\right).$$

In fact, this is also the limiting power of the optimal t -test for this problem. Thus, there is asymptotically no loss in efficiency when using the randomization test as opposed to the optimal t -test, but the randomization test has the advantage that its size is α over all symmetric distributions. In the terminology of Section 15.2, the efficacy of the randomization test is $1/\sigma$ and its ARE with respect to the t -test is 1. In fact, the ARE is 1 whenever the underlying family is a q.m.d. location family with finite variance (Problem 17.6).

Note that the randomization test that is based on T_n is identical to the randomization test that is based on the usual t -statistic t_n . To see why, first observe that the randomization test based on T_n is identical to the randomization test based on $S_n = T_n/(\sum_i X_i^2)^{1/2}$, simply because all "randomizations" of the data have the same value for the sum of squares. But, as was seen in Section 5.2, t_n is an increasing func-

tion of S_n for positive S_n . Hence, the one-sample t -test which rejects when t_n exceeds $t_{n-1, 1-\alpha}$, the $1 - \alpha$ quantile of the t -distribution with $n - 1$ degrees of freedom, is equivalent to a randomization test based on the statistic t_n , except that $t_{n-1, 1-\alpha}$ is replaced by the data-dependent value. Such an analogy was previously made for the two-sample test in Section 5.8.

One benefit of the randomization test is that one does not have to assume normality. In addition, the asymptotic results allow one to avoid the exact computation of the randomization distribution by approximating the critical value by the normal quantile $z_{1-\alpha}$ or even $t_{n-1, 1-\alpha}$. The problem of whether to use $z_{1-\alpha}$ or $t_{n-1, 1-\alpha}$ is discussed in Diaconis and Holmes (1994), who also give algorithms for the exact evaluation of the randomization distribution. In practice, critical values should be obtained from the exact randomization distribution, or its Monte Carlo approximation by randomly sampling elements of \mathbf{G} . In summary, two additional benefits are revealed by the asymptotics. First, the randomization test may be used in large samples even when the randomization hypothesis fails; in the one-sample case, this means the assumption of symmetry is not required. Second, asymptotics allow us to perform local power calculations and show that, even under normality, very little power is lost when using a randomization test as compared to the t -test; in fact, the randomization test and the t -test have the same limiting local power function against normal contiguous alternatives. ■

In the previous example, it was seen that the randomization distribution approximates the (unconditional) null distribution of T_n in the sense that

$$\hat{R}_n(t) - P\{T_n \leq t\} \xrightarrow{P} 0$$

if P has mean 0 and finite variance, since $P\{T_n \leq t\} \rightarrow \Phi(t/\sigma(P))$. The following is a more general version of this result.

Theorem 17.2.4 (i) *Suppose X_1, \dots, X_n are i.i.d. real-valued random variables with distribution P , assumed symmetric about 0. Assume T_n is asymptotically linear in the sense that, for some function ψ_P ,*

$$T_n = n^{-1/2} \sum_{i=1}^n \psi_P(X_i) + o_P(1), \tag{17.12}$$

where $E_P[\psi_P(X_i)] = 0$ and $\tau_P^2 = \text{Var}_P[\psi_P(X_i)] < \infty$. Also, assume ψ_P is an odd function. Let \hat{R}_n denote the randomization distribution based on T_n and the group of sign changes in Example 17.2.1. Then, the hypotheses of Theorem 17.2.3 hold with $P_n = P^n$ and $R(t) = \Phi(t/\tau(P))$, and so

$$\hat{R}_n(t) \xrightarrow{P} \Phi(t/\tau(P)).$$

(ii) If P is not symmetric about 0, let F denote its c.d.f. and define a symmetrized version \tilde{P} of P as the probability with c.d.f.

$$\tilde{F}(t) = \frac{1}{2}[F(t) + 1 - F(-t)].$$

Assume T_n satisfies (17.12) under \tilde{P} . Then, under P ,

$$\hat{R}_n(t) \xrightarrow{P} \Phi(t/\tau(\tilde{P})) \quad \text{and} \quad \hat{r}_n(1 - \alpha) \xrightarrow{P} \tau(\tilde{P})z_{1-\alpha}.$$

PROOF. Independent of $X^n = (X_1, \dots, X_n)$ let $\epsilon_1, \dots, \epsilon_n$ and $\epsilon'_1, \dots, \epsilon'_n$ be mutually independent, each ± 1 with probability $\frac{1}{2}$. Then, in the notation of Theorem 17.2.3, $G_n X^n = (\epsilon_1 X_1, \dots, \epsilon_n X_n)$. Set

$$\delta_n(X_1, \dots, X_n) = T_n - n^{-1/2} \sum \psi_P(X_i)$$

so that $\delta_n(X_1, \dots, X_n) \xrightarrow{P} 0$. Since $\epsilon_i X_i$ has the same distribution as X_i , it follows that $\delta_n(\epsilon_1 X_1, \dots, \epsilon_n X_n) \xrightarrow{P} 0$, and the same is true with ϵ_i replaced by ϵ'_i . Then,

$$(T_n(G_n X^n), T_n(G'_n X^n)) = n^{-1/2} \sum_{i=1}^n (\psi_P(\epsilon_i X_i), \psi_P(\epsilon'_i X_i)) + o_P(1).$$

But since ψ_P is odd, $\psi_P(\epsilon_i X_i) = \epsilon_i \psi_P(X_i)$. By the bivariate CLT,

$$n^{-1/2} \sum_{i=1}^n (\epsilon_i \psi_P(X_i), \epsilon'_i \psi_P(X_i)) \xrightarrow{d} (T, T'),$$

where (T, T') is bivariate normal, each with mean 0 and variance τ_P^2 , and

$$Cov(T, T') = Cov(\epsilon_i \psi_P(X_i), \epsilon'_i \psi_P(X_i)) = E(\epsilon_i)E(\epsilon'_i)E_P[\psi_P^2(X_i)] = 0,$$

and so (i) follows.

To prove (ii), observe that, if X has distribution P and \tilde{X} has distribution \tilde{P} , then $|X|$ and $|\tilde{X}|$ have the same distribution. But, the construction of the randomization distribution only depends on the values $|X_1|, \dots, |X_n|$. Hence, the behavior of \hat{R}_n under P and \tilde{P} must be the same. But, the behavior of \hat{R}_n under \tilde{P} is given in (i). ■

Example 17.2.5 (One-Sample Location Models) Suppose X_1, \dots, X_n are i.i.d. $f(x - \theta)$, where f is assumed symmetric about $\theta_0 = 0$. Assume the family is q.m.d. at θ_0 with score statistic Z_n . Thus, under θ_0 , $Z_n \xrightarrow{d} N(0, I(\theta_0))$. Consider the randomization test based on $T_n = Z_n$ (and the group of sign changes). It is exact level α for all symmetric distributions. Moreover, $Z_n = n^{-1/2} \sum_i \tilde{\eta}(X_i, \theta_0)$, where $\tilde{\eta}$ can

always be taken to be an odd function if f is even. So, the assumptions of Theorem 17.2.4 (i) hold. Hence, when $\theta_0 = 0$,

$$\hat{r}_n(1 - \alpha) \rightarrow I^{1/2}(\theta_0)z_{1-\alpha} .$$

By contiguity, the same is true under $\theta_{n,h} = hn^{-1/2}$. By Theorem 15.2.1, the efficacy of the randomization test is $I^{1/2}(\theta_0)$. By Corollary 15.2.1, the ARE of the randomization test with respect to the Rao test that uses the critical value $z_{1-\alpha}I^{1/2}(\theta_0)$ (or even an exact critical value based on the true unconditional distribution of Z_n under θ_0) is 1. Indeed, the randomization test is LAUMP. Therefore, there is no loss of efficiency in using the randomization test, and it has the advantage of being level α across symmetric distributions. ■

17.3 Two-Sample Permutation Tests

In this section, we derive the asymptotic behavior of some two-sample permutation tests introduced in Example 17.2.2. Recall the setup of Example 17.2.2 where Y_1, \dots, Y_m are i.i.d. P_Y and, independently, Z_1, \dots, Z_n are i.i.d. P_Z , and P_Y and P_Z are now assumed to be distributions on the real line. Let $\mu(P)$ and $\sigma^2(P)$ denote the mean and variance, respectively, of a distribution P . Consider the test statistic

$$T_{m,n} = m^{1/2}(\bar{Y}_m - \bar{Z}_n) = m^{-1/2}[\sum_{i=1}^m Y_i - \frac{m}{n} \sum_{j=1}^n Z_j] . \tag{17.13}$$

Assume $m/n \rightarrow \lambda \in (0, \infty)$ as $m, n \rightarrow \infty$. If the variances of P_Y and P_Z are finite and nonzero and $\mu(P_Y) = \mu(P_Z)$, then

$$T_{m,n} \xrightarrow{d} N(0, \sigma^2(P_Y) + \lambda\sigma^2(P_Z)) . \tag{17.14}$$

We wish to study the limiting behavior of the randomization test based on the test statistic $T_{m,n}$. If the null hypothesis implies that $P_Y = P_Z$, then the randomization test is exact level α , though we may still require an approximation to its power. On the other hand, we may consider using the randomization test for testing the null hypothesis $\mu(P_Y) = \mu(P_Z)$, and the randomization test is no longer exact if the distributions differ.

Let $N = m + n$ and write

$$(X_1, \dots, X_N) = (Y_1, \dots, Y_m, Z_1, \dots, Z_n) .$$

Independent of the X s, let $(\pi(1), \dots, \pi(N))$ and $(\pi'(1), \dots, \pi'(N))$ be independent random permutations of $1, \dots, N$. In order to verify the conditions for Theorem 17.2.3, we need to determine the joint limiting behavior of

$$(T_{m,n}, T'_{m,n}) = m^{-1/2} \left(\sum_{i=1}^N X_i W_i, \sum_{i=1}^N X_i W'_i \right), \quad (17.15)$$

where $W_i = 1$ if $\pi(i) \leq m$ and $W_i = -m/n$ otherwise; W'_i is defined with π replaced by π' . Note that $E(W_i) = E(X_i W_i) = 0$. Moreover, an easy calculation (Problem 17.8) gives

$$\text{Var}(T_{m,n}) = \frac{m}{n} \sigma^2(P_Y) + \sigma^2(P_Z) \quad (17.16)$$

and

$$\text{Cov}(T_{m,n}, T'_{m,n}) = m^{-1} \sum_{i=1}^N \sum_{j=1}^N E(X_i X_j W_i W'_j) = 0, \quad (17.17)$$

by the independence of the W_i and the W'_i . These calculations suggest the following result.

Theorem 17.3.1 *Assume the above setup with $m/n \rightarrow \lambda \in (0, \infty)$. If $\sigma^2(P_Y)$ and $\sigma^2(P_Z)$ are finite and nonzero and $\mu(P_Y) = \mu(P_Z)$, then (17.15) converges in law to a bivariate normal distribution with independent, identically distributed marginals having mean 0 and variance*

$$\tau^2 = \lambda \sigma^2(P_Y) + \sigma^2(P_Z). \quad (17.18)$$

Hence, the randomization distribution $R_{m,n}(\cdot)$ based on the statistic $T_{m,n}$ satisfies

$$\hat{R}_{m,n}(t) \xrightarrow{P} \Phi(t/\tau).$$

PROOF. Assume without loss of generality that $\mu(P_Y) = 0$. By the Cramér–Wold device (Theorem 11.2.3), it suffices to show, for any a and b (not both 0)

$$m^{-1/2} \sum_{i=1}^N X_i (a W_i + b W'_i) \xrightarrow{d} N(0, (a^2 + b^2)\tau^2).$$

Write the left side as

$$m^{-1/2} \sum_{i=1}^m Y_i (a W_i + b W'_i) + m^{-1/2} \sum_{j=1}^n Z_j (a W_{m+j} + b W'_{m+j}), \quad (17.19)$$

which conditional on the W_i and W'_i is a sum of two independent terms, with each term a linear combination of independent variables. We can handle each of the two terms in (17.19) by appealing to Lemma 13.2.3. So, we must verify

$$\frac{\max_i |aW_i + bW'_i|^2}{\sum_{i=1}^m (aW_i + bW'_i)^2} \xrightarrow{P} 0. \tag{17.20}$$

But, the numerator is bounded above by $2a^2 + 2b^2(m/n)^2$, which is uniformly bounded since $m/n \rightarrow \lambda$. Thus, it suffices to show that

$$\frac{1}{m} \sum_{i=1}^m (aW_i + bW'_i)^2 = \frac{1}{m} a^2 \sum_{i=1}^m W_i^2 + \frac{1}{m} 2ab \sum_{i=1}^m W_i W'_i + \frac{1}{m} b^2 \sum_{i=1}^m W_i'^2 \tag{17.21}$$

converges in probability to a positive constant. On the right side of (17.21) we now consider the first of the three terms and show

$$\frac{1}{m} \sum_{i=1}^m W_i^2 \xrightarrow{P} \lambda. \tag{17.22}$$

But,

$$E(W_i^2) = 1 \cdot \frac{m}{N} + \frac{m^2}{n^2} \cdot \frac{n}{N} = \frac{m}{n} \rightarrow \lambda. \tag{17.23}$$

But, $\frac{1}{m} \sum_{i=1}^m W_i$ is an average of m uniformly bounded random variables taken without replacement from a population with m items labeled one and n items labeled $(m/n)^2$ s, so that its variance tends to zero by (12.2). By Chebyshev's inequality, (17.22) follows. Since the W_i s and W'_i s have the same distribution, it also follows that

$$\frac{1}{m} \sum_{i=1}^m W_i'^2 \xrightarrow{P} \lambda. \tag{17.24}$$

We now claim that

$$\frac{1}{m} \sum_{i=1}^m W_i W'_i \xrightarrow{P} 0. \tag{17.25}$$

Since the left side of (17.25) has mean 0, it suffices to show its variance tends to 0. First, note that

$$Var(W_i W'_i) = E(W_i^2 W_i'^2) = [E(W_i^2)]^2 = \frac{m^2}{n^2},$$

by (17.23). Also,

$$Cov(W_1 W'_1, W_2 W'_2) = [E(W_1 W_2)]^2$$

and

$$E(W_1 W_2) = \frac{m}{N} \frac{m-1}{N-1} + \frac{n}{N} \frac{n-1}{N-1} \left(\frac{m}{n}\right)^2 - 2 \frac{m}{n} \frac{m}{N} \frac{n}{N-1} = -\frac{m}{n} \frac{1}{N-1}.$$

Therefore,

$$\text{Var} \left(\frac{1}{m} \sum_{i=1}^m W_i W_i' \right) = \frac{1}{m} \left[\left(\frac{m}{n} \right)^2 + m(m-1) \left(\frac{m^2}{n^2(N-1)^2} \right) \right] = O\left(\frac{1}{m}\right) \rightarrow 0.$$

It now follows that (17.21) converges in probability to $(a^2 + b^2)\lambda > 0$, as was required. Thus, (17.20) holds and the left side of (17.19) converges in distribution to $N(0, \lambda(a^2 + b^2)\sigma^2(P_Y))$. The right side of (17.19) is similar. Thus, Lemma 13.2.3 can be applied (conditionally) to each term in (17.19) and the result follows by Problem 11.73. ■

Note that the proof also shows that the result holds with $\lambda = 0$ as long as $m \rightarrow \infty$. An alternative proof of the limiting behavior of the permutation distribution can be based on Theorem 12.2.3 (Problem 17.11).

The problem of testing equality of means in the two-sample problem without imposing parametric assumptions on the underlying distributions can be viewed as a nonparametric version of the Behrens–Fisher problem. Theorem 17.2.3 and Theorem 17.3.1 show that, under the null hypothesis that $\mu(P_Y) = \mu(P_Z)$, the randomization distribution is, in large samples, approximately a normal distribution with mean 0 and variance τ^2 . Hence, the critical value of the randomization test that rejects for large values of $T_{m,n}$ converges in probability to $z_{1-\alpha}\tau$. On the other hand, the true sampling distribution of $T_{m,n}$ is approximately normal with mean 0 and variance

$$\sigma^2(P_Y) + \lambda\sigma^2(P_Z),$$

if $\mu(P_Y) = \mu(P_Z)$. These two distributions are identical if and only if $\lambda = 1$ or $\sigma^2(P_Y) = \sigma^2(P_Z)$. Therefore, for testing equality of means (and not distributions), the randomization test will be pointwise consistent in level even if P_Y and P_Z differ, as long as the variances of the populations are the same, or the sample sizes are roughly the same. In particular, when the underlying distributions have the same variance (as in the normal-theory model assumed in Section 5.3 for which the two-sample t -test is UMPU), the two-sample t -test is asymptotically equivalent to the corresponding randomization test. This equivalence is not limited to the behavior under the null hypothesis; see Problem 17.10.

In order to gain some insight into Theorem 17.3.1, the permutation distribution is invariant under permutations, and therefore its behavior under m observations from P_Y and n from P_Z should not be too different from the permutation distribution based on $N = m + n$ observations from the mixture distribution, where each observation is taken from P_Y with probability m/N and from P_Z with probability n/N . Consider the mixture distribution

$$\bar{P} = \frac{\lambda}{1+\lambda} P_Y + \frac{1}{1+\lambda} P_Z. \quad (17.26)$$

Note that when $\mu(P_Y) = \mu(P_Z)$,

$$\sigma^2(\bar{P}) = \frac{\lambda}{1 + \lambda} \sigma^2(P_Y) + \frac{1}{1 + \lambda} \sigma^2(P_Z). \tag{17.27}$$

Since the permutation test is exact in the i.i.d. case when all N observations are from \bar{P} , one might expect that the permutation distribution for $T_{m,n}$ in this case behaves like its unconditional distribution. Its limiting distribution is (from 17.14) given by

$$N(0, \sigma^2(\bar{P}) + \lambda \sigma^2(\bar{P})) = N(0, (1 + \lambda) \sigma^2(\bar{P})) = N(0, \tau^2),$$

which agrees with (17.18) in Theorem 17.3.1.

If the underlying variances differ and $\lambda \neq 1$, the permutation test based on $T_{m,n}$ given in (17.13) will have rejection probability that does not tend to α . Further results are given in Romano (1990). For example, two-sample permutations tests based on sample medians lead to tests that are not even pointwise consistent in level, unless the strict randomization hypothesis of equality of distributions holds. Thus, if testing equality of population medians based on the difference between sample medians, the asymptotic rejection probability of the randomization test need not be α even with the underlying populations have the same median.

However, if one replaces $T_{m,n}$ by the studentized version

$$\tilde{T}_{m,n} = T_{m,n} / D_N, \tag{17.28}$$

where

$$D_{m,n}^2 = D_{m,n}^2(X_1, \dots, X_N) = S_Y^2 + \frac{m}{n} S_Z^2, \tag{17.29}$$

$$S_Y^2 = \frac{1}{m} \sum_{i=1}^m (Y_i - \bar{Y}_m)^2 \quad \text{and} \quad S_Z^2 = \frac{1}{n} \sum_{j=1}^n (Z_j - \bar{Z}_n)^2,$$

then the permutation test is pointwise consistent in level for testing equality of means, even when the underlying distributions have possibly different variances and the sample sizes differ. This important result is due to Janssen (1997) (with a different proof than the one below).

In order to prove this result, note that the unconditional distribution of (17.28) is asymptotically standard normal under the assumption of finite variances. Indeed, this is a simple exercise in applying Slutsky’s Theorem. When considering the randomization distribution of (17.28), the following result can be viewed as Slutsky’s Theorem for randomization distributions.

Given sequences of statistics T_n , A_n , and B_n , let $\hat{R}_n^{AT+B}(\cdot)$ denote the randomization distribution corresponding to the statistic sequence $A_n T_n + B_n$, i.e., replace T_n in (17.9) by $A_n T_n + B_n$, so

$$\hat{R}_n^{AT+B}(t) \equiv \frac{1}{|G_n|} \sum_{g \in G_n} I\{A_n(gX^n)T_n(gX^n) + B_n(gX^n) \leq t\}. \tag{17.30}$$

Theorem 17.3.2 (Slutsky's Theorem for Randomization Distributions) Let G_n and G'_n be independent and uniformly distributed over \mathbf{G}_n (and independent of X^n). Assume T_n satisfies (17.10). Further assume that, for constants a and b ,

$$A_n(G_n X^n) \xrightarrow{P} a \quad (17.31)$$

and

$$B_n(G_n X^n) \xrightarrow{P} b. \quad (17.32)$$

Let $R^{aT+b}(\cdot)$ denote the distribution of $aT + b$, where T is the limiting random variable in (17.10). Then,

$$\hat{R}_n^{AT+B}(t) \xrightarrow{P} R^{aT+b}(t),$$

if the distribution $R^{aT+b}(\cdot)$ of $aT + b$ is continuous at t . (Note $R^{aT+b}(t) = R^T(\frac{t-b}{a})$ if $a \neq 0$.)

PROOF. The assumptions imply (see Problem 11.34) that

$$(T_n(G_n X^n), A_n(G_n X^n), B_n(G_n X^n), T_n(G'_n X^n), A_n(G'_n X^n), B_n(G'_n X^n))$$

converges in distribution to $(T, a, b, T', a, b,)$, where (T, T') are independent and given in Assumption (17.10). By the Continuous Mapping Theorem (Theorem 11.3.2), it follows that

$$(A_n(G_n X^n)T_n(G_n X^n) + B_n(G_n X^n), (A_n(G'_n X^n)T_n(G'_n X^n) + B_n(G'_n X^n))$$

converges in distribution to the distribution of $(aT + b, aT' + b)$, so that the asymptotic independence condition in Theorem 17.2.3 holds. Therefore, the result follows by Theorem 17.2.3. ■

Returning to the two-sample problem studied in Theorem 17.3.1, we are now in a position to provide the limiting behavior of the randomization based on the studentized statistic $\tilde{T}_{m,n}$ in (17.28). By Theorem 17.3.2, the problem is reduced to studying the statistic $D_{m,n}$ given in (17.29).

Theorem 17.3.3 Under the setup of Theorem 17.3.1, let $\hat{R}_{m,n}^{\tilde{T}}$ denote the randomization distribution for $\tilde{T}_{m,n}$. Assume $H_0 : \mu(P_Y) = \mu(P_Z)$. Then,

$$\hat{R}_{m,n}^{\tilde{T}}(t) \xrightarrow{P} \Phi(t). \quad (17.33)$$

Since also

$$\tilde{T}_{m,n} \xrightarrow{d} N(0, 1),$$

the randomization test based on $\tilde{T}_{m,n}$ is pointwise consistent in level.

PROOF. Combining Theorems 17.3.1 and 17.3.2, it suffices to show that, when $(\pi(1), \dots, \pi(N))$ is a random permutation, then $D_{m,n}$ given by (17.29) satisfies

$$D_{m,n}^2(X_{\pi(1)}, \dots, X_{\pi(N)}) \xrightarrow{P} \tau^2,$$

where τ^2 is given in (17.18). Equivalently, it suffices to show that each of S_Y^2 and S_Z^2 under random permutation converges in probability to $\sigma^2(\bar{P})$ given in (17.27). By symmetry it suffices to look at a randomly permuted value of

$$S_Y^2 = \frac{1}{m} \sum_{i=1}^m X_i^2 - \left(\frac{1}{m} \sum_{i=1}^m X_i \right)^2.$$

Thus, it suffices to show that, for $j = 1, 2$,

$$\frac{1}{m} \sum_{i=1}^m X_{\pi(i)}^j \xrightarrow{P} E(M^j), \tag{17.34}$$

where $M \sim \bar{P}$. Looking at $j = 1$, it is easy to calculate (Problem 17.14) that

$$E\left(\frac{1}{m} \sum_{i=1}^m X_{\pi(i)}\right) = \frac{m}{N} \mu(P_Y) + \frac{n}{N} \mu(P_Z) \rightarrow \frac{\lambda}{1+\lambda} \mu(P_Y) + \frac{1}{\lambda} \mu(P_Z) = \mu(\bar{P})$$

and that $Var\left(\frac{1}{m} \sum_{i=1}^m X_{\pi(i)}\right) \rightarrow 0$, so that

$$\frac{1}{m} \sum_{i=1}^m X_{\pi(i)} \xrightarrow{P} \mu(\bar{P}).$$

The argument for $j = 2$ is left for Problem 17.14. ■

To summarize, Theorem 17.3.3 shows that for testing the equality of population means, the studentized permutation controls the probability of a Type 1 error asymptotically, but also retains exact Type 1 error control when the underlying distributions are equal (because in this case the randomization hypothesis holds). A test based on an asymptotic normal approximation does not have such a property.

In order to generalize Theorem 17.3.3 to other test statistics, it is important to understand the intuition behind the analysis. For a given test statistic $T = T_{m,n}$, let $J_{m,n}(P_Y, P_Z)$ denote the distribution of $T_{m,n}$ based on m observations from P_Y and n from P_Z . Just as in the case of the unstudentized test statistic, the asymptotic behavior of the randomization distribution $\hat{R}_{m,n}^T$ based on m observations from P_Y and n observations from P_Z should be the same as when all observations are i.i.d. from the mixture distribution \bar{P} defined in (17.26). But the latter should be approximately equal to the true sampling distribution $J_{m,n}(\bar{P}, \bar{P})$ because the permutation test is

exact under the randomization hypothesis that both distributions are the same. If we assume that $J_{m,n}(P_Y, P_Z)$ converges to a limit law J which does not depend on (P_Y, P_Z) , then we have the randomization distribution $\hat{R}_{m,n}^T$ is approximately equal to J , which is also approximately equal to $J_{m,n}(P_Y, P_Z)$. Therefore, the conclusion is that one should construct test statistics that have a limiting distribution free of any unknown parameters, at least under the null hypothesis. In the two-sample problem for testing differences of means, this is easily achieved by studentization. In fact, these ideas hold quite generally and apply to broad classes of test statistics; see Chung and Romano (2013, 2016a, 2016b).

17.4 Further Examples

In this section, randomization and permutation tests are applied to some other situations.

Example 17.4.1 (Testing Means From Independent Observations) Assume X_1, \dots, X_n are independent, but not necessarily i.i.d. Let $E(X_i) = \mu_i$, assumed to exist. Also, let

$$\bar{\mu}_n = \frac{1}{n} \sum_{i=1}^n \mu_i,$$

$\text{Var}(X_i) = \sigma_i^2$ and

$$\bar{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n \sigma_i^2.$$

First, examine the case where $\mu_i = \mu$ for all i . For testing $H_0 : \mu = 0$, consider the group of sign changes as in Example 17.2.4. Even if the X_i have distinct distributions but these distributions are symmetric about μ , then the randomization hypothesis holds. Therefore, one may construct exact level α randomization tests under symmetry.

The question we would now like to investigate is Type 1 error control when symmetry does not hold. Let $T_n = \sqrt{n} \bar{X}_n$. Assume $\bar{\sigma}_n^2 \rightarrow \sigma_\infty^2 > 0$ and that, for some $\delta > 0$, $\sup_i E(|X_i - \mu_i|^{2+\delta}) < \infty$. By an argument similar to Example 17.2.4, one can show (Problem 17.18) that the conditions of Theorem 17.2.3 hold with the limit distribution R equal to $N(0, \sigma_\infty^2)$, where $\sigma_\infty^2 = \lim_n \bar{\sigma}_n^2$. On the other hand, by Example 11.2.2, under $\mu = 0$,

$$\sqrt{n} \bar{X}_n \xrightarrow{d} N(0, \sigma_\infty^2). \quad (17.35)$$

Thus, the true sampling distribution and the randomization distribution are asymptotically equal, and therefore, under the above moment assumptions, the probability of a Type 1 error tends to the nominal level even under asymmetry.

Under the same setup and assumptions, we now consider the problem of testing the null hypothesis that $\bar{\mu}_n = 0$ (so that the μ_i may differ even under the null hypothesis). In this case, the randomization test has exact Type 1 error control when all the underlying distributions are symmetric about 0. So, we wish to also consider the asymptotic behavior under asymmetry as well as heterogeneity in means (and underlying distributions). Assume that

$$\frac{1}{n} \sum_{i=1}^n \mu_i^2 \rightarrow v_\infty < \infty .$$

Then, the conditions of Theorem 17.2.3 hold with R now equal to $N(0, \sigma_\infty^2 + v_\infty)$. It then follows that the critical value, $\hat{r}_n(1 - \alpha)$, based on the randomization distribution satisfies

$$\hat{r}_n(1 - \alpha) \xrightarrow{P} z_{1-\alpha} \sqrt{\sigma_\infty^2 + v_\infty} .$$

On the other hand, under the null hypothesis $\bar{\mu}_n = 0$, (17.35) still holds. Hence, if $v_\infty > 0$, we see that these limiting distributions do not match. Therefore, under the null hypothesis, by Slutsky's Theorem,

$$P \left\{ \sqrt{n} \bar{X}_n > \hat{r}_n(1 - \alpha) \right\} \rightarrow 1 - \Phi \left(z_{1-\alpha} \sqrt{1 + \frac{v_\infty}{\sigma_\infty^2}} \right) .$$

The limiting probability is $\leq \alpha$ and only equals α when $v_\infty = 0$, so that the resulting randomization test is in general conservative. ■

Example 17.4.2 (Matched Pairs) In paired comparisons, data (Y_i, Z_i) , $i = 1, \dots, n$, are observed. For example, pairing may arise in studies of twins, where one receives treatment and the other does not. Alternatively, paired data may represent before and after treatment of the same unit. In general, units may be matched according to other observed covariates, such as age, sex, blood pressure, etc. If the Z_i s represents the treated observations and Y_i s the untreated, then the differences $D_i = Z_i - Y_i$ may be used to make inferences about the treatment effect $E(D_i)$. (More generally, matched pairs may be viewed as a special case of a randomized block design, where units are divided into subgroups or blocks according to covariates, so that smaller variability within blocks leads to more efficient estimate of treatment effects.)

Upon reduction to the D_i s, the one-sample tests studied in Examples 17.2.4 and 17.4.1 may apply. However, if observations are paired according to covariates, then the D_i may no longer be independent, and so the analysis becomes more involved. If observations are assigned to treatment at random within pairs once pairs are formed, then the D_i are conditionally independent given the covariates, in which case the analysis of Example 17.4.1 may be used as a starting point. The details are beyond the scope here; see Bugni et al. (2019) and Bai et al. (2021). ■

Example 17.4.3 (Hotelling Test for Multivariate Mean) Let X_1, \dots, X_n be i.i.d. random vectors with distribution P on \mathbf{R}^p . Assume $E_p|X_i|^2 < \infty$. Let $\mu = \mu(P)$ be the mean vector and let $\Sigma = \Sigma(P)$ be the covariance matrix, assumed positive definite. The problem is to test the null hypothesis $H_0 : \mu(P) = 0$ versus the alternative hypothesis $H_1 : \mu(P) \neq 0$ (or possibly a subset thereof).

Under the assumption of multivariate normality, one can perform an exact test using Hotelling's T-squared statistic. Specifically, let \bar{X}_n denote the sample mean vector, and let $\hat{\Sigma}_n$ denote the sample covariance matrix, defined as

$$\hat{\Sigma}_n = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)(X_i - \bar{X}_n)^\top.$$

Note that $\hat{\Sigma}_n$ is invertible with probability one. Then, Hotelling's T-squared statistic is defined as

$$T_n = T_n(X_1, \dots, X_n) = n\bar{X}_n^\top \hat{\Sigma}_n^{-1} \bar{X}_n.$$

If H_0 is true, then under the assumption of multivariate normality, T_n has Hotelling's T-squared distribution with parameters p and $n-1$, from which exact critical values may be obtained. As $n \rightarrow \infty$, this distribution tends to the Chi-squared distribution with p degrees of freedom.

Instead, we may construct exact tests without the multivariate normality assumption by constructing a randomization test. Under normality and if H_0 is true, it follows that X_i and $-X_i$ have the same distribution. But, this holds for many more distributions. In fact, exact Type 1 error control holds for distributions with mean 0 satisfying X_i and $-X_i$ have the same distribution; that is, the randomization hypothesis holds for this family of distributions with respect to the group of sign changes. Of course, this holds for any p , so that it applies even in the high-dimensional setting.

The question we now consider is the asymptotic behavior of the randomization test if the randomization hypothesis fails (for large n , and fixed p). We wish to show that the probability of rejecting H_0 tends to the nominal level α as $n \rightarrow \infty$, without assuming X_i and $-X_i$ have the same distribution under H_0 .

To do this, let $\epsilon_1, \dots, \epsilon_n$ and $\epsilon'_1, \dots, \epsilon'_n$ be mutually independent and identically distributed, and independent of the X_i s, each ϵ_i either 1 or -1 with probability $1/2$ each. We claim that

$$(T_n(\epsilon_1 X_1, \dots, \epsilon_n X_n), T_n(\epsilon'_1 X_1, \dots, \epsilon'_n X_n)) \xrightarrow{d} (T, T'), \quad (17.36)$$

where T and T' are i.i.d., each with the Chi-squared distribution with p degrees of freedom. The result then follows immediately from Theorem 17.2.3. To prove (17.36), we need the following lemma.

Lemma 17.4.1 *Assume X_1, \dots, X_n are i.i.d. P with positive definite covariance matrix $\Sigma = \Sigma(P)$. Under the above assumptions, and if $\mu(P) = 0$, then*

$$n^{-1/2} \left(\sum_{i=1}^n \epsilon_i X_i, \sum_{i=1}^n \epsilon'_i X_i \right) \xrightarrow{d} (Z_1, Z_2), \tag{17.37}$$

where Z_1 and Z_2 are i.i.d., each with the multivariate normal distribution with mean 0 and covariance matrix Σ .

Remark 17.4.1 Even if $\mu(P) \neq 0$, then the same argument applies with the same result as long as Σ is replaced by the matrix with (j, k) component given by $E(X_{i,j} X_{i,k})$.

PROOF OF LEMMA 17.4.1. Apply the Cramér–Wold Device. Fix vectors a and b . It suffices to show the unconditional distribution of

$$n^{-1/2} \sum_{i=1}^n (\epsilon_i a^\top X_i + \epsilon'_i b^\top X_i) \tag{17.38}$$

tends in distribution to that of $a^\top Z_1 + b^\top Z_2$, which of course is $N(0, a^\top \Sigma a + b^\top \Sigma b)$. But (17.38) is a normalized sum of i.i.d. real-valued random variables, with mean 0 and variance

$$\begin{aligned} E[(\epsilon_i a^\top X_i + \epsilon'_i b^\top X_i)^2] &= E\epsilon_i^2 E[(a^\top X_i)^2] + E((\epsilon'_i)^2) E[(b^\top X_i)^2] + E\epsilon_i E\epsilon'_i E(a^\top X_i b^\top X_i) \\ &= E[(a^\top X_i)^2] + E[(b^\top X_i)^2] = E(a^\top X_i X'_i a) + E(b^\top X_i X'_i b) = a^\top \Sigma a + b^\top \Sigma b. \end{aligned}$$

The result follows from the ordinary Central Limit Theorem. ■

Next, we consider the randomization distribution based on a modified T-squared statistic. Instead of $\hat{\Sigma}_n$, define $\bar{\Sigma}_n$ to be

$$\bar{\Sigma}_n = \frac{1}{n} \sum_{i=1}^n X_i X_i^\top.$$

Let

$$\bar{T}_n = \bar{T}_n(X_1, \dots, X_n) = n \bar{X}_n^\top \bar{\Sigma}_n^{-1} \bar{X}_n,$$

so that $\hat{\Sigma}_n$ replaces $\bar{\Sigma}_n$ in T_n , and the denominator $n - 1$ is changed to n . There are two reasons for considering this modification. First, $\bar{\Sigma}_n$ is, under H_0 , an unbiased and consistent estimator for Σ . More importantly for our purposes, $\bar{\Sigma}_n$ is invariant with respect to sign changes of the observations; that is, replacing X_i by $-X_i$ for any of the i results in the same estimator.

Lemma 17.4.2 Under the assumptions of Lemma 17.4.1 and if $\mu(P) = 0$, we have

$$(\bar{T}_n(\epsilon_1 X_1, \dots, \epsilon_n X_n), \bar{T}_n(\epsilon'_1 X_1, \dots, \epsilon'_n X_n)) \xrightarrow{d} (Z_1^\top \Sigma^{-1} Z_1, Z_2^\top \Sigma^{-1} Z_2),$$

and hence the limiting joint distribution is a product of independent Chi-squared distributions each with p degrees of freedom.

PROOF OF LEMMA 17.4.2. As mentioned, for any ϵ_i , $\bar{\Sigma}_n(\epsilon_1 X_1, \dots, \epsilon_n X_n) = \bar{\Sigma}_n(X_1, \dots, X_n)$, with each term converging in probability to that of Σ . The result follows by the Continuous Mapping Theorem. ■

Next, we consider Hotelling’s T-squared statistic.

Lemma 17.4.3 *Under the assumptions of Lemma 17.4.1 and if $\mu(P) = 0$, we have*

$$(T_n(\epsilon_1 X_1, \dots, \epsilon_n X_n), T_n(\epsilon'_1 X_1, \dots, \epsilon'_n X_n)) \xrightarrow{d} (Z_1^\top \Sigma^{-1} Z_1, Z_2^\top \Sigma^{-1} Z_2),$$

and hence the limiting joint distribution is a product of independent Chi-squared distributions each with p degrees of freedom.

PROOF OF LEMMA 17.4.3. Trivially,

$$\frac{n-1}{n} \hat{\Sigma}_n(\epsilon_1 X_1, \dots, \epsilon_n X_n) = \bar{\Sigma}_n - \left(\frac{1}{n} \sum_{i=1}^n \epsilon_i X_i\right) \left(\frac{1}{n} \sum_{i=1}^n \epsilon_i X_i\right)^\top.$$

But, $n^{-1} \sum_i \epsilon_i X_i$ is an average of mean 0 random vectors with finite second moments, and hence converges in probability to 0. Since we already established $\bar{\Sigma}_n$ is consistent, it follows that $\hat{\Sigma}_n$ is a consistent estimator of Σ . The result then follows again by the Continuous Mapping Theorem. ■

Thus, the conditions for Theorem 17.2.3 have been verified and so the permutation distribution is asymptotically Chi-squared with p degrees of freedom. Since the true unconditional limiting distribution is also Chi-squared with p degrees of freedom, the conclusion is that the randomization test based on the T-squared statistic (or modified T-squared statistic) has rejection probability tending to the nominal level under any P with $\mu(P) = 0$ (and having second moments). Moreover, it retains exact control of the Type 1 error as long as the randomization hypothesis holds, i.e., X_i and $-X_i$ have the same distribution (actually even if second moments don’t exist). ■

Example 17.4.4 (Maximum Test for Multivariate Mean) Assume the same setup as Example 17.4.3. But rather than Hotelling’s T-squared statistic, consider the maximum studentized statistic M_n , given by

$$M_n = M_n(X_1, \dots, X_n) = \max_{1 \leq j \leq p} T_{n,j}, \tag{17.39}$$

where

$$T_{n,j} = \frac{\sqrt{n} |\bar{X}_{n,j}|}{s_{n,j}}, \tag{17.40}$$

$\bar{X}_{n,j}$ is the j th component of the sample mean vector \bar{X}_n , and $s_{n,j}^2$ is the (j, j) component of $\hat{\Sigma}_n$. As before, when X_i and $-X_i$ have the same distribution under

the null hypothesis, the rejection probability is exactly the nominal level (and this is true for any choice of test statistic). We would like to examine the asymptotic Type 1 error rate of the randomization test based on the test statistic M_n under the null hypothesis H_0 that specifies $\mu(P) = 0$. For this, we assume as in Example 17.4.3 that Σ exists, and also that all diagonal elements of Σ are positive.

Under H_0 , the true sampling distribution of M_n satisfies (Problem 17.19)

$$M_n \xrightarrow{d} \max_{1 \leq j \leq p} (|Y_1|, \dots, |Y_p|), \quad (17.41)$$

where (Y_1, \dots, Y_p) is multivariate normal with mean 0 and covariance matrix Σ' , where Σ' is the correlation matrix corresponding to Σ . In other words, the (i, j) component of Σ' is the (i, j) component of Σ divided by $\sigma_i \sigma_j$, and σ_i^2 is the i th diagonal element of Σ .

We claim that the randomization distribution asymptotically approximates the distribution of $\max_j |Y_j|$ under H_0 . To see this, we again apply Theorem 17.2.3. First, note that, under H_0 ,

$$\frac{n-1}{n} s_{n,j}^2(\epsilon_1 X_1, \dots, \epsilon_n X_n) = \frac{1}{n} \sum_{i=1}^n X_{i,j}^2 - \bar{X}_{n,j}^2 \xrightarrow{P} \sigma_j^2. \quad (17.42)$$

Therefore, by Lemma 17.4.1 and (17.42), we can apply the argument used in the proof of Theorem 17.3.2 to deduce that

$$(M_n(\epsilon_1 X_1, \dots, \epsilon_n X_n), M_n(\epsilon'_1 X_1, \dots, \epsilon'_n X_n)) \xrightarrow{d} (M, M'), \quad (17.43)$$

where M and M' are i.i.d. with distribution that of $\max_j (|Y_j|)$ given in (17.41). Thus, the conditions of Theorem 17.2.3 hold. Therefore, similar conclusions apply to the randomization test based on M_n as for Hotelling's T-squared statistic. That is, the probability of a Type 1 error control is exactly the nominal level when X_i and $-X_i$ have the same distribution, but otherwise, the probability of a Type 1 error tends to the nominal level.

17.5 Randomization Tests and Multiple Testing

So far, randomization and permutation tests have been developed for tests of a single null hypothesis. Extensions to multiple testing are possible and desirable. Some obvious ways to do this are as follows:

- Since randomization tests can be used to generate p -values of individual tests, such as by (17.5) or (17.7), one can apply any of a number of multiple testing procedures based on marginal p -values. Many such tests were discussed in Chapter 9. For example, one may apply:

- The Holm method to control the FWER (Theorem 9.1.2).
- The Benjamini–Yekutieli method to control the FDR (Theorem 9.3.2).
- Apply the closure method discussed in Section 9.2, where tests of intersection hypotheses are constructed via randomization tests.

Example 17.5.1 (Example 17.4.3, continued) In Example 17.4.3, a randomization test based on Hotelling’s T-squared statistic was discussed for testing the null hypothesis that a mean vector $\mu(P) = 0$. If one wishes to know which components of the mean vector might be nonzero, then the problem should be regarded as a multiple testing problem. Let H_i specify the i th component, $\mu_i(P)$, of $\mu(P)$ is 0. The closure method may be applied here as a means of constructing a procedure that controls the FWER. All that is needed are tests of the intersection hypotheses H_K , where

$$H_K : \mu_i(P) = 0 \quad \text{for all } i \in K .$$

Example 17.4.3 shows how to test H_K when $K = \{1, \dots, p\}$. For general K , one can simply apply the same test, but just to those components specified by K . If I denotes the set of indices i of hypotheses corresponding to $\mu_i(P) = 0$, and $X_I = (X_i, i \in I)$, then the FWER is controlled exactly if X_I and $-X_I$ have the same distribution. Without the symmetry assumption, the FWER tends to α (under the second moment assumption considered in Example 17.4.3). ■

In the previous example, the number of subsets K required to test may be of order 2^p . By using the maximum statistic, we now instead develop a stepdown method (as in Example 9.1.7) that is feasible for large p . By using randomization tests, the assumption of multivariate normality is not needed. For power considerations when comparing the two test statistics, see Section 13.5.4.

Example 17.5.2 (Example 17.4.4, continued) We now develop a stepdown method for testing means based on the maximum statistic M_n defined in (17.39). The method is a special case of Procedure 9.1.1, though we provide the details here. For $i = 1, \dots, p$, H_i specifies $\mu_i(P) = 0$. With $T_{n,j}$ defined in (17.40), order the observed test statistics as

$$T_{n,r_1} \geq T_{n,r_2} \geq \dots \geq T_{n,r_p}$$

and let $H_{(1)}, H_{(2)}, \dots, H_{(p)}$ be the corresponding hypotheses.

The stepdown procedure begins with the most significant test statistic T_{n,r_1} , which is also M_n . First, test the joint null hypothesis $H_{\{1,\dots,p\}}$ that all null hypotheses are true, using the randomization test based on the maximum statistic M_n described in Example 17.4.4. This hypothesis is rejected if T_{n,r_1} is large. If it is not large, accept all hypotheses; otherwise, reject the hypothesis $H_{(1)}$ corresponding to the largest test statistic. Once a hypothesis is rejected, remove it and test the remaining hypotheses by rejecting for large values of the maximum of the remaining test statistics, and so on.

We just need to specify the construction of critical values in each step. When testing H_K , let $\hat{c}_{n,K}(1 - \alpha)$ denote the $1 - \alpha$ quantile of the randomization distribution corresponding to the statistic

$$M_{n,K} = \max_{j \in K} T_{n,j} .$$

(Note that one could use the exact randomization test, which allows randomization in order to achieve exact level α , but for simplicity we opt for the slightly conservative procedure that does not randomize. The method can be adapted to maintain exact error control if desired.) Then, the stepdown algorithm can be described as follows.

Procedure 1 (Stepdown Method Based on Randomization Tests)

1. Let $K_1 = \{1, \dots, p\}$. If $T_{n,r_1} \leq \hat{c}_{n,K_1}(1 - \alpha)$, then accept all hypotheses and stop; otherwise, reject $H_{(1)}$ and continue.
2. Let K_2 be the indices of the hypotheses not previously rejected. If $T_{n,r_2} \leq \hat{c}_{n,K_2}(1 - \alpha)$, then accept all remaining hypotheses and stop; otherwise, reject $H_{(2)}$ and continue.
- ⋮
- j. Let K_j be the indices of the hypotheses not previously rejected. If $T_{n,r_j} \leq \hat{c}_{n,K_j}(1 - \alpha)$, then accept all remaining hypotheses and stop; otherwise, reject $H_{(j)}$ and continue.
- ⋮
- s. If $T_{n,r_p} \leq \hat{c}_{n,K_p}(1 - \alpha)$, then accept $H_{(p)}$; otherwise, reject $H_{(p)}$.

Theorem 9.1.3 shows that the FWER is controlled at level α under the symmetry assumption that X_I and $-X_I$ have the same distribution, just as in Example 17.5.1. Moreover, even if the symmetry assumption does not hold, the FWER tends to α under the assumption of finite nonzero variances. ■

17.6 Problems

Section 17.2

Problem 17.1 Generalize Theorem 17.2.1 to the case where \mathbf{G} is an infinite group.

Problem 17.2 With \hat{p} defined in (17.5), show that (17.6) holds.

Problem 17.3 (i) Suppose Y_1, \dots, Y_B are exchangeable real-valued random variables; that is, their joint distribution is invariant under permutations. Let \tilde{q} be defined by

$$\tilde{q} = \frac{1}{B} \left[1 + \sum_{i=1}^{B-1} I\{Y_i \geq Y_B\} \right].$$

Show, $P\{\tilde{q} \leq u\} \leq u$ for all $0 \leq u \leq 1$. *Hint:* Condition on the order statistics.

(ii) With \tilde{p} defined in (17.7), show that (17.8) holds.

(iii) How would you construct a p -value based on sampling without replacement from \mathbf{G} ?

Problem 17.4 With \hat{p} and \tilde{p} defined in (17.5) and (17.7), respectively, show that $\hat{p} - \tilde{p} \rightarrow 0$ in probability.

Problem 17.5 As an approximation to (17.9), let g_1, \dots, g_{B-1} be i.i.d. and uniform on \mathbf{G} . Also, set g_B to be the identity. Define

$$\tilde{R}_{n,B}(t) = \frac{1}{B} \sum_{i=1}^B I\{T_n(g_i X) \leq t\}.$$

Show, conditional on X ,

$$\sup_t |\tilde{R}_{n,B}(t) - \hat{R}_n(t)| \rightarrow 0$$

in probability as $B \rightarrow \infty$, and so

$$\sup_t |\tilde{R}_{n,B}(t) - \hat{R}_n(t)| \rightarrow 0$$

in probability (unconditionally) as well. Do these results hold only under the null hypothesis? *Hint:* Apply Theorem 11.4.3. For a similar result based on sampling without replacement, see Problem 12.15.

Problem 17.6 Suppose X_1, \dots, X_n are i.i.d. according to a q.m.d. location model with finite variance. Show the ARE of the one-sample t -test with respect to the randomization t -test (based on sign changes) is 1 (even if the underlying density is not normal).

Problem 17.7 In Theorem 17.2.4, show the conclusion may fail if ψ_P is not an odd function.

Problem 17.8 Verify (17.16) and (17.17). *Hint:* Let S be the number of positive integers $i \leq m$ with $W_i = 1$, and condition on S .

Problem 17.9 (i) Assume X_1, X_2, \dots are independent, with $X_i \sim N(\mu_i, 1)$, with $\mu_i \geq 0$. For testing the null hypothesis that all $\mu_i = 0$, compute the limiting power of the one-sided t -test against alternatives μ_i such that $\sum_i \mu_i^2 < \infty$. (Even though the variance is known, you are asked to consider the t -test.)

(ii) Rather than being normally distributed, suppose X_i has density f_i , where f_i is symmetric about some μ_i , but is otherwise unknown. Assume $\mu_i \geq 0$. For testing $\mu_i = 0$ for all i , how can you construct a randomization test that is level α in finite samples? (Note the X_i can have different distributions even under the null hypothesis.) Show your test is reasonable by calculating its limiting power against the same alternatives as in (i) when the X_i are normally distributed but heterogeneous with different means.

Section 17.3

Problem 17.10 In the two-sample problem of Section 17.3, suppose the underlying distributions are normal with common variance. For testing $\mu(P_Y) = \mu(P_Z)$ against $\mu(P_Y) > \mu(P_Z)$ compute the limiting power of the randomization test based on the test statistic $T_{m,n}$ given in (17.13) against contiguous alternatives of the form $\mu(P_Y) = \mu(P_Z) + hn^{-1/2}$. Show this is the same as the optimal two-sample t -test. Argue that the two tests are asymptotically equivalent in the sense of Problem 15.25.

Problem 17.11 In the two-sample problem of Section 17.3, reprove the limiting behavior of the permutation distribution by using Theorem 12.2.3.

Problem 17.12 Show (17.27). Does it hold if $\mu(P_Y) \neq \mu(P_Z)$?

Problem 17.13 Use Theorem 17.3.2 to deduce the limiting behavior of the randomization distribution using the classical t -statistic in Example 17.2.4.

Problem 17.14 Complete the argument to show (17.34) for $j = 1$ and $j = 2$.

Problem 17.15 Under the setting of Problem 11.59 for testing equality of Poisson means λ_i based on the test statistic T , show how to construct a randomization test based on T . Examine the limiting behavior of the randomization distribution under the null hypothesis and contiguous alternatives.

Problem 17.16 Suppose $(X_1, Y_1), \dots, (X_n, Y_n)$ are i.i.d. bivariate observations in the plane, and let ρ denote the correlation between X_1 and Y_1 . Let $\hat{\rho}_n$ be the sample correlation

$$\hat{\rho}_n = \frac{\sum (X_i - \bar{X}_n)(Y_i - \bar{Y}_n)}{[\sum_i (X_i - \bar{X}_n)^2 \sum_j (Y_j - \bar{Y}_n)^2]^{1/2}}$$

(i) For testing independence of X_i and Y_i , construct a randomization test based on the test statistic $T_n = n^{1/2}|\hat{\rho}_n|$.

(ii) For testing $\rho = 0$ versus $\rho > 0$ based on the test statistic $\hat{\rho}_n$, determine the limit behavior of the randomization distribution when the underlying population is bivariate Gaussian with correlation $\rho = 0$. Determine the limiting power of the randomization test under local alternatives $\rho = hn^{-1/2}$. Argue that the randomization

test and the optimal UMPU test (5.75) are asymptotically equivalent in the sense of Problem 15.25.

(iii) Investigate what happens if the underlying distribution has correlation 0, but X_i and Y_i are dependent.

Problem 17.17 Prove a version of the Continuous Mapping Theorem (Theorem 11.2.10) for randomization distributions. That is, assume the randomization distribution $\hat{R}_n(\cdot)$ of some test statistic T_n satisfies $\hat{R}_n(t)$ converges in probability to $R(t)$ for all t for which $R(\cdot)$ is continuous. Let g be a continuous function, at least on a set of points where R has probability one. Prove a limit result for the randomization distribution based on the statistic $g(T_n)$.

Section 17.4

Problem 17.18 In Example 17.4.1, for testing the null hypothesis that all $\mu_i = 0$, verify the asymptotic behavior of the randomization distribution under the null hypothesis. *Hint: Problem 11.12.*

Problem 17.19 Show (17.41).

Problem 17.20 Prove (17.43).

17.7 Notes

Early references to permutations tests were provided at the end of Chapter 5. An elementary account is provided by Good (1994), who provides an extensive bibliography, and Edgington (1995). Multivariate permutation tests are developed in Pesarin (2001). The present large-sample approach originated in Hoeffding (1952). Applications to block experiments are discussed in Robinson (1973). Expansions for the power of rank and permutation tests in the one- and two-sample problems are obtained in Albers et al. (1976) and Bickel and Van Zwet (1978), respectively. A full account of the large-sample theory of rank statistics is given in Hájek et al. (1999). Robust two-sample permutation tests are obtained in Lambert (1985). The role of studentization in providing Type 1 error control for permutation tests was first recognized in Neuhaus (1993) and Janssen (1997, 1999), Janssen and Pauls (2005). A growing literature allows for application of randomization tests even when the randomization hypothesis fails. Some general results were provided in Chung and Romano (2013, 2016a, 2016b). Also, see Neubert and Brunner (2007) and Janssen and Pauls (2003a, 2005). Omelka and Pauly (2012) use permutation tests to compare correlations. Jentsch and Pauly (2015) apply randomization tests to testing equality of spectral densities. DiCiccio and Romano (2017) apply permutation tests for

inference about correlation and regression coefficients. Bugni et al. (2019) apply randomization schemes to randomized controlled studies where units are stratified into a finite number of strata according to covariates. Bai et al. (2021) apply randomization methods to paired comparisons where the number of strata grows with sample size. Ritzwoller and Romano (2021) develop permutation tests for tests of streakiness in Bernoulli sequences with applications to the “hot hand” hypothesis. Romano and Tirlea (2020) consider permutation tests for dependence in time series models. Randomization methods have been extended to situations where the randomization hypothesis only holds in an asymptotic sense; see Canay et al. (2017).