15

Efficiency of Tests

It is shown that, given converging experiments, every limiting power function is the power function of a test in the limit experiment. Thus, uniformly most powerful tests in the limit experiment give absolute upper bounds for the power of a sequence of tests. In normal experiments such uniformly most powerful tests exist for linear hypotheses of codimension one. The one-sample location problem and the two-sample problem are discussed in detail, and appropriately designed (signed) rank tests are shown to be asymptotically optimal.

15.1 Asymptotic Representation Theorem

A randomized *test* (or *test function*) ϕ in an experiment $(\mathcal{X}, \mathcal{A}, P_h : h \in H)$ is a measurable map $\phi : \mathcal{X} \mapsto [0, 1]$ on the sample space. The interpretation is that if x is observed, then a null hypothesis is rejected with probability $\phi(x)$. The *power function* of a test ϕ is the function

$$h \mapsto \pi(h) = \mathrm{E}_h \phi(X).$$

This gives the probabilities that the null hypothesis is rejected. A test is of level α for testing a null hypothesis H_0 if its size $\sup\{\pi(h) : h \in H_0\}$ does not exceed α. The quality of a test can be judged from its power function, and classical testing theory is aimed at finding, among the tests of level α, a test with high power at every alternative.

The asymptotic quality of a sequence of tests may be judged from the limit of the sequence of local power functions. If the tests are defined in experiments that converge to a limit experiment, then a pointwise limit of power functions is necessarily a power function in the limit experiment. This follows from the following theorem, which specializes the asymptotic representation theorem, Theorem 9.3, to the testing problem. Applied to the special case of the local experiments $\mathcal{E}_n = (P^n_{\theta+h/\sqrt{n}} : h \in \mathbb{R}^k)$ of a differentiable parametric model as considered in Chapter 7, which converge to the Gaussian experiment $(N(h, I_\theta^{-1}), h \in \mathbb{R}^k)$, the theorem is the parallel for testing of Theorem 7.10.

15.1 ***Theorem.*** *Let the sequence of experiments $\mathcal{E}_n = (P_{n,h} : h \in H)$ converge to a dominated experiment $\mathcal{E} = (P_h : h \in H)$. Suppose that a sequence of power functions π_n of tests in $\mathcal{E}_n$ converges pointwise: $\pi_n(h) \to \pi(h)$, for every h and some arbitrary function π. Then π is a power function in the limit experiment: There exists a test ϕ in $\mathcal{E}$ with $\pi(h) = \mathrm{E}_h\phi(X)$ for every h.*

Proof. We give the proof for the special case of experiments that satisfy the following assumption: Every sequence of statistics T_n that is tight under every given parameter h possesses a subsequence (not depending on h) that converges in distribution to a limit under every h. See problem 15.2 for a method to extend the proof to the general situation.

The additional condition is valid in the case of local asymptotic normality. With the notation of the proof of Theorem 7.10, we argue first that the sequence (T_n, Δ_n) is uniformly tight under $h=0$ and hence possesses a weakly convergent subsequence by Prohorov's theorem. Next, by the expansion of the likelihood and Slutsky's lemma, the sequence $(T_n, \log dP_{n,h}/dP_{n,0})$ converges under $h = 0$ along the same sequence, for every h. Finally, we conclude by Le Cam's third lemma that the sequence T_n converges under h, along the subsequence.

Let ϕ_n be tests with power functions π_n. Because each ϕ_n takes its values in the compact interval $[0, 1]$, the sequence of random variables ϕ_n is certainly uniformly tight. By assumption, there exists a subsequence of $\{n\}$ along which ϕ_n converges in distribution under every h. Thus, the assumption of the asymptotic representation theorem, Theorem 9.3 or Theorem 7.10, is satisfied along some subsequence of the statistics ϕ_n. By this theorem, there exists a randomized statistic $T = T(X, U)$ in the limit experiment such that $\phi_n \overset{h}{\rightsquigarrow} T$ along the subsequence, for every h. The randomized statistic may be assumed to take its values in $[0, 1]$. Because the ϕ_n are uniformly bounded, $\mathrm{E}_h\phi_n \to \mathrm{E}_h T$. Combination with the assumption yields $\pi(h) = \mathrm{E}_h T$ for every h. The randomized statistic T is not a test function (it is a "doubly randomized" test). However, the test $\phi(x) = \mathrm{E}\big(T(X, U) \mid X = x\big)$ satisfies the requirements. ∎

The theorem suggests that the best possible limiting power function is the power function of the best test in the limit experiment. In classical testing theory an "absolutely best" test is defined as a *uniformly most powerful test* of the required level. Depending on the experiment, such a test may or may not exist. If it does not exist, then the classical solution is to find a uniformly most powerful test in a restricted class, such as the class of all unbiased or invariant tests; to use the maximin criterion; or to use a conditional test. In combination with the preceding theorem, each of these approaches leads to a criterion for asymptotic quality. We do not pursue this in detail but note that, in general, we would avoid any sequence of tests that is matched in the limit experiment by a test that is considered suboptimal.

In the remainder of this chapter we consider the implications for locally asymptotically normal models in more detail. We start with reviewing testing in normal location models.

15.2 Testing Normal Means

Suppose that the observation X is $N_k(h, \Sigma)$-distributed, for a known covariance matrix Σ and unknown mean vector h. First consider testing the null hypothesis $H_0: c^T h = 0$ versus the alternative $H_1: c^T h > 0$, for a known vector c. The "natural" test, which rejects H_0 for large values of $c^T X$, is uniformly most powerful. In other words, if π is a power function such that $\pi(h) \leq \alpha$ for every h with $c^T h = 0$, then for every h with $c^T h > 0$,

$$\pi(h) \leq \mathrm{P}_h(c^T X > z_\alpha \sqrt{c^T \Sigma c}) = 1 - \Phi\left(z_\alpha - \frac{c^T h}{\sqrt{c^T \Sigma c}}\right).$$

15.2 ***Proposition.*** *Suppose that X be $N_k(h, \Sigma)$-distributed for a known nonnegative-definite matrix Σ, and let c be a fixed vector with $c^T \Sigma c > 0$. Then the test that rejects H_0 if $c^T X > z_\alpha \sqrt{c^T \Sigma c}$ is uniformly most powerful at level α for testing $H_0 : c^T h = 0$ versus $H_1 : c^T h > 0$, based on X.*

Proof. Fix h_1 with $c^T h_1 > 0$. Define $h_0 = h_1 - (c^T h_1 / c^T \Sigma c)\Sigma c$. Then $c^T h_0 = 0$. By the Neyman-Pearson lemma, the most powerful test for testing the simple hypotheses $H_0 : h = h_0$ and $H_1 : h = h_1$ rejects H_0 for large values of

$$\log \frac{dN(h_1, \Sigma)}{dN(h_0, \Sigma)}(X) = \frac{c^T h_1}{c^T \Sigma c} c^T X - \frac{1}{2} \frac{(c^T h_1)^2}{c^T \Sigma c}.$$

This is equivalent to the test that rejects for large values of $c^T X$. More precisely, the most powerful level α test for $H_0 : h = h_0$ versus $H_1 : h = h_1$ is the test given by the proposition. Because this test does not depend on h_0 or h_1, it is uniformly most powerful for testing $H_0 : c^T h = 0$ versus $H_1 : c^T h > 0$. ∎

The natural test for the two-sided problem $H_0 : c^T h = 0$ versus $H_1 : c^T h \neq 0$ rejects the null hypothesis for large values of $|c^T X|$. This test is not uniformly most powerful, because its power is dominated by the uniformly most powerful tests for the two one-sided alternatives whose union is H_1. However, the test with critical region $\{x : |c^T x| \geq z_{\alpha/2} \sqrt{c^T \Sigma c}\}$ is uniformly most powerful among the unbiased level α tests (see problem 15.1).

A second problem of interest is to test a simple null hypothesis $H_0 : h = 0$ versus the alternative $H_1 : h \neq 0$. If the parameter set is one-dimensional, then this reduces to the problem in the preceding paragraph. However, if θ is of dimension $k > 1$, then there exists no uniformly most powerful test, not even among the unbiased tests. A variety of tests are reasonable, and whether a test is "good" depends on the alternatives at which we desire high power. For instance, the test that is most sensitive to detect the alternatives such that $c^T h > 0$ (for a given c) is the test given in the preceding theorem. Probably in most situations no particular "direction" is of special importance, and we would use a test that distributes the power over all directions. It is known that any test with as critical region the complement of a closed, convex set C is admissible (see, e.g., [138, p. 137]). In particular, complements of closed, convex, and symmetric sets are admissible critical regions and cannot easily be ruled out a priori. The shape of C determines the power function, the directions in which C extends little receiving large power (although the power also depends on Σ).

The most popular test rejects the null hypothesis for large values of $X^T \Sigma^{-1} X$. This test arises as the limit version of the Wald test, the score test, and the likelihood ratio test. One advantage is a simple choice of critical values, because $X^T \Sigma^{-1} X$ is chi square–distributed with k degrees of freedom. The power function of this test is, with Z a standard normal vector,

$$\pi(h) = \mathrm{P}_h\left(X^T \Sigma^{-1} X > \chi^2_{k,\alpha}\right) = \mathrm{P}\left(\|Z + \Sigma^{-1/2} h\|^2 > \chi^2_{k,\alpha}\right).$$

By the rotational symmetry of the standard normal distribution, this depends only on the *noncentrality parameter* $\|\Sigma^{-1/2} h\|$. The power is relatively large in the directions h for which $\|\Sigma^{-1/2} h\|$ is large. In particular, it increases most steeply in the direction of the eigenvector corresponding to the smallest eigenvalue of Σ. Note that the test does not distribute the power evenly, but dependent on Σ. Two optimality properties of this test are given in problems 15.3 and 15.4, but these do not really seem convincing.

Due to the lack of an acceptable optimal test in the limit problem, a satisfactory asymptotic optimality theory of testing simple hypotheses on multidimensional parameters is impossible.

15.3 Local Asymptotic Normality

A normal limit experiment arises, among others, in the situation of repeated sampling from a differentiable parametric model. If the model $(P_\theta : \theta \in \Theta)$ is differentiable in quadratic mean, then the local experiments converge to a Gaussian limit:

$$\left(P^n_{\theta_0+h/\sqrt{n}} : h \in \mathbb{R}^k\right) \to \left(N\left(h, I_{\theta_0}^{-1}\right) : h \in \mathbb{R}^k\right).$$

A sequence of power functions $\theta \mapsto \pi_n(\theta)$ in the original experiments induces the sequence of power functions $h \mapsto \pi_n(\theta_0 + h/\sqrt{n})$ in the local experiments. Suppose that $\pi_n(\theta_0 + h/\sqrt{n}) \to \pi(h)$ for every h and some function π. Then, by the asymptotic representation theorem, the limit π is a power function in the Gaussian limit experiment.

Suppose for the moment that θ is real and that the sequence π_n is of asymptotic level α for testing $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$. Then $\pi(0) = \lim \pi_n(\theta_0) \leq \alpha$ and hence π corresponds to a level α test for $H_0 : h = 0$ versus $H_1 : h > 0$ in the limit experiment. It must be bounded above by the power function of the uniformly most powerful level α test in the limit experiment, which is given by Proposition 15.2. Conclude that

$$\lim_{n\to\infty} \pi_n\left(\theta_0 + \frac{h}{\sqrt{n}}\right) \leq 1 - \Phi(z_\alpha - h\sqrt{I_{\theta_0}}), \quad \text{every } h > 0.$$

(Apply the proposition with $c = 1$ and $\Sigma = I_{\theta_0}^{-1}$.) We have derived an absolute upper bound on the local asymptotic power of level α tests.

In Chapter 14 a sequence of power functions such that $\pi_n(\theta_0 + h/\sqrt{n}) \to 1 - \Phi(z_\alpha - hs)$ for every h is said to have slope s. It follows from the present upper bound that the square root $\sqrt{I_{\theta_0}}$ of the Fisher information is the largest possible slope. The quantity

$$\frac{I_{\theta_0}}{s^2}$$

is the relative efficiency of the best test and the test with slope s. It can be interpreted as the number of observations needed with the given sequence of tests with slope s divided by the number of observations needed with the best test to obtain the same power.

With a bit of work, the assumption that $\pi_n(\theta_0 + h/\sqrt{n})$ converges to a limit for every h can be removed. Also, the preceding derivation does not use the special structure of i.i.d. observations but only uses the convergence to a Gaussian experiment. We shall rederive the result within the context of local asymptotic normality and also indicate how to construct optimal tests.

Suppose that at "time" n the observation is distributed according to a distribution $P_{n,\theta}$ with parameter ranging over an open subset Θ of $\mathbb{R}^k$. The sequence of experiments $(P_{n,\theta} : \theta \in \Theta)$ is locally asymptotically normal at θ_0 if

$$\log \frac{dP_{n,\theta_0+r_n^{-1}h}}{dP_{n,\theta_0}} = h^T \Delta_{n,\theta_0} - \frac{1}{2} h^T I_{\theta_0} h + o_{P_{n,\theta_0}}(1), \tag{15.3}$$

for a sequence of statistics Δ_{n,θ_0} that converges in distribution under θ_0 to a normal $N_k(0, I_{\theta_0})$-distribution.

15.4 Theorem. *Let $\Theta \subset \mathbb{R}^k$ be open and let $\psi : \Theta \mapsto \mathbb{R}$ be differentiable at θ_0, with nonzero gradient $\dot\psi_{\theta_0}$ and such that $\psi(\theta_0) = 0$. Let the sequence of experiments $(P_{n,\theta}: \theta \in \Theta)$ be locally asymptotically normal at θ_0 with nonsingular Fisher information, for constants $r_n \to \infty$. Then the power functions $\theta \mapsto \pi_n(\theta)$ of any sequence of level α tests for testing $H_0 : \psi(\theta) \le 0$ versus $H_1 : \psi(\theta) > 0$ satisfy, for every h such that $\dot\psi_{\theta_0} h > 0$,*

$$\limsup_{n\to\infty} \pi_n\left(\theta_0 + \frac{h}{r_n}\right) \le 1 - \Phi\left(z_\alpha - \frac{\dot\psi_{\theta_0} h}{\sqrt{\dot\psi_{\theta_0} I_{\theta_0}^{-1} \dot\psi_{\theta_0}^T}}\right).$$

15.5 Addendum. *Let T_n be statistics such that*

$$T_n = \frac{\dot\psi_{\theta_0} I_{\theta_0}^{-1} \Delta_{n,\theta_0}}{\sqrt{\dot\psi_{\theta_0} I_{\theta_0}^{-1} \dot\psi_{\theta_0}^T}} + o_{P_{n,\theta_0}}(1).$$

Then the sequence of tests that reject for values of T_n exceeding z_α is asymptotically optimal in the sense that the sequence $\mathrm{P}_{\theta_0 + r_n^{-1}h}(T_n \ge z_\alpha)$ converges to the right side of the preceding display, for every h.

Proofs. The sequence of localized experiments $(P_{n,\theta_0 + r_n^{-1}h} : h \in \mathbb{R}^k)$ converges by Theorem 7.10, or Theorem 9.4, to the Gaussian location experiment $\big(N_k(h, I_{\theta_0}^{-1}) : h \in \mathbb{R}^k\big)$.

Fix some h_1 such that $\dot\psi_{\theta_0} h_1 > 0$, and a subsequence of $\{n\}$ along which the lim sup $\pi(\theta_0 + h_1/r_n)$ is taken. There exists a further subsequence along which $\pi_n(\theta_0 + r_n^{-1}h)$ converges to a limit $\pi(h)$ for every $h \in \mathbb{R}^k$ (see the proof of Theorem 15.1). The function $h \mapsto \pi(h)$ is a power function in the Gaussian limit experiment. For $\dot\psi_{\theta_0} h < 0$, we have $\psi(\theta_0 + r_n^{-1}h) = r_n^{-1}\big(\dot\psi_{\theta_0} h + o(1)\big) < 0$ eventually, whence $\pi(h) \le \limsup \pi_n(\theta_0 + r_n^{-1}h) \le \alpha$. By continuity, the inequality $\pi(h) \le \alpha$ extends to all h such that $\dot\psi_{\theta_0} h \le 0$. Thus, π is of level α for testing $H_0 : \dot\psi_{\theta_0} h \le 0$ versus $H_1 : \dot\psi_{\theta_0} h > 0$. Its power function is bounded above by the power function of the uniformly most powerful test, which is given by Proposition 15.2. This concludes the proof of the theorem.

The asymptotic optimality of the sequence T_n follows by contiguity arguments. We start by noting that the sequence $(\Delta_{n,\theta_0}, \Delta_{n,\theta_0})$ converges under θ_0 in distribution to a (degenerate) normal vector (Δ, Δ). By Slutsky's lemma and local asymptotic normality,

$$\left(\Delta_{n,\theta_0}, \log \frac{dP_{n,\theta_0 + r_n^{-1}h}}{dP_{n,\theta_0}}\right) \overset{\theta_0}{\rightsquigarrow} \left(\Delta, h^T\Delta - \tfrac{1}{2} h^T I_{\theta_0} h\right)$$
$$\sim N\left(\begin{pmatrix} 0 \\ -\frac{1}{2} h^T I_{\theta_0} h \end{pmatrix}, \begin{pmatrix} I_{\theta_0} & I_{\theta_0} h \\ h^T I_{\theta_0} & h^T I_{\theta_0} h \end{pmatrix}\right).$$

By Le Cam's third lemma, the sequence Δ_{n,θ_0} converges in distribution under $\theta_0 + r_n^{-1}h$ to a $N(I_{\theta_0} h, I_{\theta_0})$-distribution. Thus, the sequence T_n converges under $\theta_0 + r_n^{-1}h$ in distribution to a normal distribution with mean $\dot\psi_{\theta_0} h / (\dot\psi_{\theta_0} I_{\theta_0}^{-1} \dot\psi_{\theta_0}^T)^{1/2}$ and variance 1. ∎

The point θ_0 in the preceding theorem is on the boundary of the null and the alternative hypotheses. If the dimension k is larger than 1, then this boundary is typically $(k-1)$-dimensional, and there are many possible values for θ_0. The upper bound is valid at every possible choice.

If $k = 1$, the boundary point θ_0 is typically unique and hence known, and we could use $T_n = I_{\theta_0}^{-1/2}\Delta_{n,\theta_0}$ to construct an optimal sequence of tests for the problem $H_0:\theta = \theta_0$. These are known as *score tests*.

Another possibility is to base a test on an estimator sequence. Not surprisingly, efficient estimators yield efficient tests.

15.6 ***Example (Wald tests).*** Let $X_1, \dots, X_n$ be a random sample in an experiment $(P_\theta:\theta \in \Theta)$ that is differentiable in quadratic mean with nonsingular Fisher information. Then the sequence of local experiments $(P^n_{\theta+h/\sqrt{n}}:h \in \mathbb{R}^k)$ is locally asymptotically normal with $r_n = \sqrt{n}$, I_θ the Fisher information matrix, and

$$\Delta_{n,\theta} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \dot\ell_\theta(X_i).$$

A sequence of estimators $\hat\theta_n$ is asymptotically efficient for estimating θ if (see Chapter 8)

$$\sqrt{n}(\hat\theta_n - \theta) = I_\theta^{-1}\Delta_{n,\theta} + o_{P_\theta}(1).$$

Under regularity conditions, the maximum likelihood estimator qualifies. Suppose that $\theta \mapsto I_\theta$ is continuous, and that ψ is continuously differentiable with nonzero gradient. Then the sequence of tests that reject $H_0:\psi(\theta) \le 0$ if

$$\sqrt{n}\psi(\hat\theta_n) \ge z_\alpha\sqrt{\dot\psi_{\hat\theta_n} I_{\hat\theta_n}^{-1}\dot\psi_{\hat\theta_n}^T}$$

is asymptotically optimal at every point θ_0 on the boundary of H_0. Furthermore, this sequence of tests is consistent at every θ with $\psi(\theta) > 0$.

These assertions follow from the preceding theorem, upon using the delta method and Slutsky's lemma. The resulting tests are called *Wald tests* if $\hat\theta_n$ is the maximum likelihood estimator. □

15.4 One-Sample Location

Let $X_1, \dots, X_n$ be a sample from a density $f(x-\theta)$, where f is symmetric about zero and has finite Fisher information for location I_f. It is required to test $H_0:\theta = 0$ versus $H_1:\theta > 0$. The density f may be known or (partially) unknown. For instance, it may be known to belong to the normal scale family.

For fixed f, the sequence of experiments $\left(\prod_{i=1}^n f(x_i-\theta):\theta \in \mathbb{R}\right)$ is locally asymptotically normal at $\theta = 0$ with $\Delta_{n,0} = -n^{-1/2}\sum_{i=1}^n (f'/f)(X_i)$, norming rate $\sqrt{n}$, and Fisher information I_f. By the results of the preceding section, the best asymptotic level α power function (for known f) is

$$1 - \Phi(z_\alpha - h\sqrt{I_f}).$$

This function is an upper bound for $\limsup \pi_n(h/\sqrt{n})$, for every $h > 0$, for every sequence of level α power functions. Suppose that T_n are statistics with

$$T_n = -\frac{1}{\sqrt{n}}\frac{1}{\sqrt{I_f}}\sum_{i=1}^n \frac{f'}{f}(X_i) + o_{P_0}(1). \tag{15.7}$$

Then, according to the second assertion of Theorem 15.4, the sequence of tests that reject the null hypothesis if $T_n \geq z_\alpha$ attains the bound and hence is asymptotically optimal. We shall discuss several ways of constructing test statistics with this property.

If the shape of the distribution is completely known, then the test statistics T_n can simply be taken equal to the right side of (15.7), without the remainder term, and we obtain the score test. It is more realistic to assume that the underlying distribution is only known up to scale. If the underlying density takes the form $f(x) = f_0(x/\sigma)/\sigma$ for a known density f_0 that is symmetric about zero, but for an unknown scale parameter σ, then

$$\frac{f'}{f}(x) = \frac{1}{\sigma}\frac{f_0'}{f_0}\Big(\frac{x}{\sigma}\Big), \qquad I_f = \frac{1}{\sigma^2}I_{f_0}, \qquad \frac{1}{\sqrt{I_f}}\frac{f'}{f}(x) = \frac{1}{\sqrt{I_{f_0}}}\frac{f_0'}{f_0}\Big(\frac{x}{\sigma}\Big).$$

15.8 ***Example (t-test).*** The standard normal density f_0 possesses score function $f_0'/f_0(x) = -x$ and Fisher information $I_{f_0} = 1$. Consequently, if the underlying distribution is normal, then the optimal test statistics should satisfy $T_n = \sqrt{n}\overline{X}_n/\sigma + o_{P_0}(n^{-1/2})$. The t-statistics $\overline{X}_n/S_n$ fulfill this requirement. This is not surprising, because in the case of normally distributed observations the t-test is uniformly most powerful for every finite n and hence is certainly asymptotically optimal. $\square$

The t-statistic in the preceding example simply replaces the unknown standard deviation σ by an estimate. This approach can be followed for most scale families. Under some regularity conditions, the statistics

$$T_n = -\frac{1}{\sqrt{n}}\frac{1}{\sqrt{I_{f_0}}}\sum_{i=1}^n \frac{f_0'}{f_0}\Big(\frac{X_i}{\hat\sigma_n}\Big)$$

should yield asymptotically optimal tests, given a consistent sequence of scale estimators $\hat\sigma_n$.

Rather than using score-type tests, we could use a test based on an efficient estimator for the unknown symmetry point and efficient estimators for possible nuisance parameters, such as the scale – for instance, the maximum likelihood estimators. This method is indicated in general in Example 15.8 and leads to the Wald test.

Perhaps the most attractive approach is to use signed rank statistics. We summarize some definitions and conclusions from Chapter 13. Let $R_{n1}^+, \dots, R_{nn}^+$ be the ranks of the absolute values $|X_1|, \dots, |X_n|$ in the ordered sample of absolute values. A *linear signed rank statistic* takes the form

$$T_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n a_{n,R_{ni}^+}\,\mathrm{sign}(X_i),$$

for given numbers $a_{n1}, \dots, a_{nn}$, which are called the *scores* of the statistic. Particular examples are the *Wilcoxon signed rank statistic*, which has scores $a_{ni} = i$, and the *sign statistic*, which corresponds to scores $a_{ni} = 1$. In general, the scores can be chosen to weigh the influence of the different observations. A convenient method of generating scores is through a fixed function $\phi : [0, 1] \mapsto \mathbb{R}$, by

$$a_{ni} = \mathrm{E}\phi(U_{n(i)}).$$

(Here $U_{n(1)}, \dots, U_{n(n)}$ are the order statistics of a random sample of size n from the uniform distribution on $[0, 1]$.) Under the condition that $\int_0^1 \phi^2(u)\, du < \infty$, Theorem 13.18 shows that, under the null hypothesis, and with $F^+(x) = 2F(x) - 1$ denoting the distribution function of $|X_1|$,

$$T_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi\Big(F^+\big(|X_i|\big)\Big) \operatorname{sign}(X_i) + o_{P_0}(1).$$

Because the score-generating function ϕ can be chosen freely, this allows the construction of an asymptotically optimal rank statistic for any given shape f. The choice

$$\phi(u) = -\frac{1}{\sqrt{I_f}} \frac{f'}{f}\big((F^+)^{-1}(u)\big). \tag{15.9}$$

yields the *locally most powerful scores*, as discussed in Chapter 13. Because $f'/f\big(|x|\big)$ sign $(x) = f'/f(x)$ by the symmetry of f, it follows that the signed rank statistics T_n satisfy (15.7). Thus, the locally most powerful scores yield asymptotically optimal signed rank tests. This surprising result, that the class of signed rank statistics contains asymptotically efficient tests for every given (symmetric) shape of the underlying distribution, is sometimes expressed by saying that the signs and absolute ranks are "asymptotically sufficient" for testing the location of a symmetry point.

15.10 ***Corollary.*** *Let T_n be the simple linear signed rank statistic with scores $a_{ni} = \mathrm{E}\phi(U_{n(i)})$ generated by the function ϕ defined in* (15.9). *Then T_n satisfies* (15.7) *and hence the sequence of tests that reject $H_0 : \theta = 0$ if $T_n \geq z_\alpha$ is asymptotically optimal at $\theta = 0$.*

Signed rank statistics were originally constructed because of their attractive property of being distribution free under the null hypothesis. Apparently, this can be achieved without losing (asymptotic) power. Thus, rank tests are strong competitors of classical parametric tests. Note also that signed rank statistics automatically adapt to the unknown scale: Even though the definition of the optimal scores appears to depend on f, they are actually identical for every member of a scale family $f(x) = f_0(x/\sigma)/\sigma$ (since $(F^+)^{-1}(u) = \sigma(F_0^+)^{-1}(u)$). Thus, no auxiliary estimate for σ is necessary for their definition.

15.11 ***Example (Laplace).*** The sign statistic $T_n = n^{-1/2}\sum_{i=1}^n \operatorname{sign}(X_i)$ satisfies (15.7) for f equal to the Laplace density. Thus the sign test is asymptotically optimal for testing location in the Laplace scale family. □

15.12 ***Example (Normal).*** The standard normal density has score function for location $f_0'/f_0(x) = -x$ and Fisher information $I_{f_0} = 1$. The optimal signed rank statistic for the normal scale family has score-generating function

$$\phi(u) = \mathrm{E}(\Phi^+)^{-1}(U_{n(i)}) = \mathrm{E}\Phi^{-1}\left(\frac{U_{n(i)} + 1}{2}\right) \approx \Phi^{-1}\left(\frac{i}{2n+2} + \frac{1}{2}\right).$$

We conclude that the corresponding sequence of rank tests has the same asymptotic slope as the t-test if the underlying distribution is normal. (For other distributions the two sequences of tests have different asymptotic behavior.) □

Even the assumption that the underlying distribution of the observations is known up to scale is often unrealistic. Because rank statistics are distribution-free under the null hypothesis, the level of a rank test is independent of the underlying distribution, which is the best possible protection of the level against misspecification of the model. On the other hand, the power of a rank test is not necessarily robust against deviations from the postulated model. This might lead to the use of the best test for the wrong model. The dependence of the power on the underlying distribution may be relaxed as well, by a procedure known as *adaptation*. This entails estimating the underlying density from the data and next using an optimal test for the estimated density. A remarkable fact is that this approach can be completely successful: There exist test statistics that are asymptotically optimal for any shape f. In fact, without prior knowledge of f (other than that it is symmetric with finite and positive Fisher information for location), estimators $\hat{\theta}_n$ and I_n can be constructed such that, for every θ and f,

$$\sqrt{n}(\hat{\theta}_n - \theta) = -\frac{1}{\sqrt{n}}\frac{1}{I_f}\sum_{i=1}^{n}\frac{f'}{f}(X_i - \theta) + o_{P_\theta}(1); \qquad I_n \overset{P_\theta}{\rightsquigarrow} I_f.$$

We give such a construction in section 25.8.1. Then the test statistics $T_n = \sqrt{n}\hat{\theta}_n I_n^{1/2}$ satisfy (15.7) and hence are asymptotically (locally) optimal at $\theta = 0$ for every given shape f. Moreover, for every $\theta > 0$, and every f,

$$\mathrm{P}_\theta(T_n > z_\alpha) = \mathrm{P}_\theta\big(\sqrt{n}(\hat{\theta}_n - \theta) > z_\alpha I_n^{-1/2} - \sqrt{n}\theta\big) \to 1.$$

Hence, the sequence of tests based on T_n is also consistent at every (θ, f) in the alternative hypothesis $H_1 : \theta > 0$.

15.5 Two-Sample Problems

Suppose we observe two independent random samples $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ from densities p_μ and q_ν, respectively. The problem is to test the null hypothesis $H_0 : \nu \leq \mu$ versus the alternative $H_1 : \nu > \mu$. There may be other unknown parameters in the model besides μ and ν, but we shall initially ignore such "nuisance parameters" and parametrize the model by $(\mu, \nu) \in \mathbb{R}^2$. Null and alternative hypotheses are shown graphically in Figure 15.1. We let $N = m + n$ be the total number of observations and assume that $m/N \to \lambda$ as $m, n \to \infty$.

15.13 ***Example (Testing shift).*** If $p_\mu(x) = f(x - \mu)$ and $q_\nu(y) = g(y - \nu)$ for two densities f and g that have the same "location," then we obtain the two-sample location problem. The alternative hypothesis asserts that the second sample is "stochastically larger." □

The alternatives of greatest interest for the study of the asymptotic performance of tests are sequences (μ_N, ν_N) that converge to the boundary between null and alternative hypotheses. In the study of relative efficiency, in Chapter 14, we restricted ourselves to

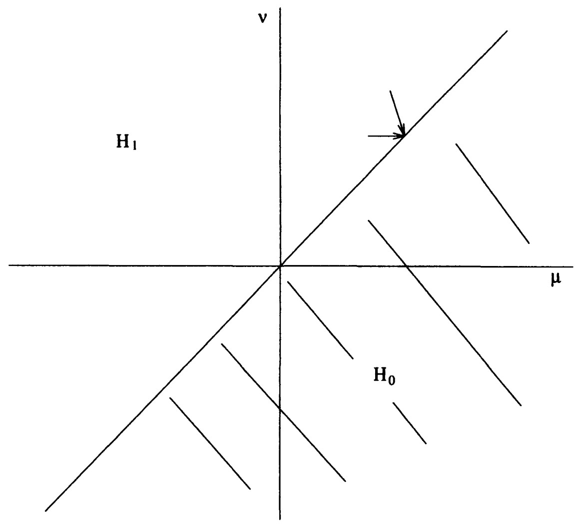

Figure 15.1. Null and alternative hypothesis.

vertical perturbations $(\theta, \theta + h/\sqrt{N})$. Here we shall use the sequences $(\theta + g/\sqrt{N}, \theta + h/\sqrt{N})$, which approach the boundary in the direction of a general vector (g, h).

If both p_μ and q_ν define differentiable models, then the sequence of experiments $\left(P_\mu^m \otimes P_\nu^n : (\mu, \nu) \in \mathbb{R}^2\right)$ is locally asymptotically normal with norming rate $\sqrt{N}$. If the score functions are denoted by $\dot{\kappa}_\mu$ and $\dot{\ell}_\nu$, and the Fisher informations by I_μ and J_ν, respectively, then the parameters of local asymptotic normality are

$$\Delta_{n,(\mu,\nu)} = \begin{pmatrix} \dfrac{\sqrt{\lambda}}{\sqrt{m}} \displaystyle\sum_{i=1}^{m} \dot{\kappa}_\mu(X_i) \\ \dfrac{\sqrt{1-\lambda}}{\sqrt{n}} \displaystyle\sum_{j=1}^{n} \dot{\ell}_\nu(Y_j) \end{pmatrix}, \qquad I_{(\mu,\nu)} = \begin{pmatrix} \lambda I_\mu & 0 \\ 0 & (1-\lambda) J_\nu \end{pmatrix}.$$

The corresponding limit experiment consists of observing two independent normally distributed variables with means g and h and variances $\lambda^{-1} I_\mu^{-1}$ and $(1-\lambda)^{-1} J_\nu^{-1}$, respectively.

15.14 ***Corollary.*** *Suppose that the models $(P_\mu : \mu \in \mathbb{R})$ and $(Q_\nu : \nu \in \mathbb{R})$ are differentiable in quadratic mean, and let $m, n \to \infty$ such that $m/N \to \lambda \in (0, 1)$. Then the power functions of any sequence of level α tests for $H_0 : \nu = \mu$ satisfies, for every μ and for every $h > g$,*

$$\limsup_{n,m\to\infty} \pi_{m,n}\left(\mu + \frac{g}{\sqrt{N}}, \mu + \frac{h}{\sqrt{N}}\right) \leq 1 - \Phi\left(z_\alpha - (h-g)\sqrt{\frac{\lambda(1-\lambda) I_\mu J_\mu}{\lambda I_\mu + (1-\lambda) J_\mu}}\right).$$

Proof. This is a special case of Theorem 15.4, with $\psi(\mu, \nu) = \nu - \mu$ and Fisher information matrix diag$\left(\lambda I_\mu, (1-\lambda) J_\mu\right)$. It is slightly different in that the null hypothesis $H_0 : \psi(\theta) = 0$ takes the form of an equality, which gives a weaker requirement on the sequence T_n. The proof goes through because of the linearity of ψ. ∎

It follows that the optimal slope of a sequence of tests is equal to

$$s_{\mathrm{opt}}(\mu) = \sqrt{\frac{\lambda(1-\lambda)I_\mu J_\mu}{\lambda I_\mu + (1-\lambda)J_\mu}}.$$

The square of the quotient of the actual slope of a sequence of tests and this number is a good absolute measure of the asymptotic quality of the sequence of tests.

According to the second assertion of Theorem 15.4, an optimal sequence of tests can be based on any sequence of statistics such that

$$T_N = s_{\mathrm{opt}}(\mu)\left(\frac{1}{\sqrt{1-\lambda}J_\mu}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\dot\ell_\mu(Y_j) - \frac{1}{\sqrt{\lambda}I_\mu}\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\dot\kappa_\mu(X_i)\right) + o_P(1).$$

(The multiplicative factor $s_{\mathrm{opt}}(\mu)$ ensures that the sequence T_N is asymptotically normally distributed with variance 1.) Test statistics with this property can be constructed using a variety of methods. For instance, in many cases we can use asymptotically efficient estimators for the parameters μ and ν, combined with estimators for possible nuisance parameters, along the lines of Example 15.6.

If $p_\mu = q_\mu = f_\mu$ are equal and are densities on the real line, then rank statistics are attractive. Let $R_{N1}, \dots, R_{NN}$ be the ranks of the pooled sample $X_1, \dots, X_m, Y_1, \dots, Y_n$. Consider the two-sample rank statistics

$$T_N = \frac{1}{\sqrt{N}}\sum_{i=m+1}^{N} a_{N,R_{Ni}}, \qquad a_{Ni} = \mathrm{E}\phi(U_{N(i)}),$$

for the score generating function

$$\phi(u) = \frac{1}{\sqrt{\lambda(1-\lambda)}\sqrt{I_\mu}}\dot\ell_\mu\big(F_\mu^{-1}(u)\big).$$

Up to a constant these are the locally most powerful scores introduced in Chapter 13. By Theorem 13.5, because $\bar a_N = \int_0^1 \phi(u)\,du = 0$,

$$T_N = -\frac{1}{\sqrt{I_\mu}}\left(\sqrt{1-\lambda}\frac{1}{\sqrt{m}}\sum_{i=1}^{m}\dot\ell_\mu(X_i) - \sqrt{\lambda}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\dot\ell_\mu(Y_j)\right) + o_{P_\mu}(1).$$

Thus, the locally most powerful rank statistics yield asymptotically optimal tests. In general, the optimal rank test depends on μ, and other parameters in the model, which must be estimated from the data, but in the most interesting cases this is not necessary.

15.15 ***Example (Wilcoxon statistic).*** For f_μ equal to the logistic density with mean μ, the scores $a_{N,i}$ are proportional to i. Thus, the Wilcoxon (or Mann-Whitney) two-sample statistic is asymptotically uniformly most powerful for testing a difference in location between two samples from logistic densities with different means. □

15.16 ***Example (Log rank test).*** The log rank test is asymptotically optimal for testing proportional hazard alternatives, given any baseline distribution. □

Notes

Absolute bounds on asymptotic power functions as developed in this chapter are less known than the absolute bounds on estimator sequences given in Chapter 8. Testing problems were nevertheless an important subject in Wald [149], who is credited by Le Cam for having first conceived of the method of approximating experiments by Gaussian experiments, albeit in a somewhat different way than later developed by Le Cam. From the point of view of statistical decision theory, there is no difference between testing and estimating, and hence the asymptotic bounds for tests in this chapter fit in the general theory developed in [99]. Wald appears to use the Gaussian approximation to transfer the optimality of the likelihood ratio and the Wald test (that is now named for him) in the Gaussian experiment to the sequence of experiments. In our discussion we use the Gaussian approximation to show that, in the multidimensional case, "asymptotic optimality" can only be defined in a somewhat arbitrary manner, because optimality in the Gaussian experiment is not easy to define. That is a difference of taste.

PROBLEMS

1. Consider the two-sided testing problem $H_0: c^T h = 0$ versus $H_1: c^T h \neq 0$ based on an $N_k(h, \Sigma)$-distributed observation X. A test for testing H_0 versus H_1 is called *unbiased* if $\sup_{h\in H_0} \pi(h) \leq \inf_{h\in H_1} \pi(h)$. The test that rejects H_0 for large values of $|c^T X|$ is uniformly most powerful among the unbiased tests. More precisely, for every power function π of a test based on X the conditions

$$\pi(h) \leq \alpha \quad \text{if } h^T c = 0 \qquad \text{and} \qquad \pi(h) \geq \alpha \quad \text{if } h^T c \neq 0,$$

imply that, for every $c^T h \neq 0$,

$$\pi(h) \leq \mathrm{P}\big(|c^T X| > z_{\alpha/2}\sqrt{c^T \Sigma c}\big) = 1 - \Phi\Big(z_{\alpha/2} - \frac{c^T h}{\sqrt{c^T \Sigma c}}\Big) + 1 - \Phi\Big(z_{\alpha/2} + \frac{c^T h}{\sqrt{c^T \Sigma c}}\Big).$$

Formulate an asymptotic upper bound theorem for two-sided testing problems in the spirit of Theorem 15.4.

2. (i) Show that the set of power functions $h \mapsto \pi_n(h)$ in a dominated experiment $(P_h : h \in H)$ is compact for the topology of pointwise convergence (on H).
   (ii) Give a full proof of Theorem 15.1 along the following lines. First apply the proof as given for every finite subset $I \subset H$. This yields power functions π_I in the limit experiment that coincide with π on I.
3. Consider testing $H_0: h = 0$ versus $H_1: h \neq 0$ based on an observation X with an $N(h, \Sigma)$-distribution. Show that the testing problem is invariant under the transformations $x \mapsto \Sigma^{1/2} O \Sigma^{-1/2} x$ for O ranging over the orthonormal group. Find the best invariant test.
4. Consider testing $H_0: h = 0$ versus $H_1: h \neq 0$ based on an observation X with an $N(h, \Sigma)$-distribution. Find the test that maximizes the minimum power over $\{h : \|\Sigma^{-1/2} h\| = c\}$. (By the Hunt-Stein theorem the best invariant test is maximin, so one can apply the preceding problem. Alternatively, one can give a direct derivation along the following lines. Let π be the distribution of $\Sigma^{1/2} U$ if U is uniformly distributed on the set $\{h : \|h\| = c\}$. Derive the Neyman-Pearson test for testing $H_0: N(0, \Sigma)$ versus $H_1: \int N(h, \Sigma)\, d\pi(h)$. Show that its power is constant on $\{h : \|\Sigma^{-1/2} h\| = c\}$. The minimum power of any test on this set is always smaller than the average power over this set, which is the power at $\int N(h, \Sigma)\, d\pi(h)$.)