1 More on Local Asymptotic Power

Earlier we introduced the idea of local asymptotic power, i.e., the power of a test under a sequence of distributions in the alternative hypothesis, as a way of approximating the finite-sample power function of a test. In the examples we considered, namely, the t-test and the sign test in a symmetric location model, we were able to compute the local asymptotic power using direct arguments. Unfortunately, it is easy to write down even simple situations in which direct arguments are too cumbersome to be fruitful. We now consider one such example and provide an alternative way of carrying out the required computation. This alternative way is based on the theory of contiguous probability measures developed by Lucien Le Cam.

2 Wilcoxon Signed Rank Statistic

Consider again the symmetric location model from earlier, i.e., one observes data $X_i, i = 1, \ldots, n$ i.i.d. with distribution $P \in \mathbf{P} = \{P_\theta : \theta \in \Theta\}$, where P_θ is the distribution with density $f(x - \theta)$ on the real line and f is symmetric about zero. One wishes to test the null hypothesis $H_0 : \theta = 0$ versus the alternative $H_1 : \theta > 0$. We considered two tests of this null hypothesis, namely the t-test and sign test. We now consider a third test based on the following statistic:

$$W_n = \frac{1}{\sqrt{n}} \sum_{1 \le i \le n} \frac{R^+_{i,n}}{n} \text{sign}(X_i) \ ,$$

where

$$\text{sign}(X_i) = \begin{cases} 1 & \text{if } X_i \ge 0 \\ 0 & \text{otherwise} \end{cases}$$

and $R^+_{i,n}$ is the rank of $|X_i|$ among $|X_1|, \ldots, |X_n|$.

In order to determine the appropriate critical value with which to compare this statistic, we must analyze its behavior under the null hypothesis, i.e., when $\theta = 0$. To this end, note that

$$\frac{R^+_{i,n}}{n} = \frac{1}{n} \sum_{1 \le j \le n} I\{|X_j| \le |X_i|\} \ .$$

Hence, it is reasonable to suspect that

$$W_n - \frac{1}{\sqrt{n}} \sum_{1 \le i \le n} U_i \text{sign}(X_i) = o_P(1) \ ,$$

where $U_i = G(|X_i|)$ and G is the c.d.f. of $|X_i|$. Therefore, the limit distribution of W_n is the same as the limit distribution of

$$\frac{1}{\sqrt{n}} \sum_{1 \le i \le n} U_i \text{sign}(X_i) \ ,$$

which is relatively easy to determine. Observe that under the null hypothesis U_i and $\text{sign}(X_i)$ are independent and $E_0[\text{sign}(X_i)] = 0$, which implies that

$$E_0[U_i \text{sign}(X_i)] = 0 \ .$$

Moreover,

$$V_0[U_i \text{sign}(X_i)] = E_0[U_i^2] = E_0[G(|X_i|)^2] \ .$$

It is possible to show that $E_0[G(|X_i|)^2] = \frac{1}{3}$. Therefore,

$$W_n \xrightarrow{d} N(0, \frac{1}{3})$$

under P_0. Hence, the test

$$\phi_{3,n} = I\{W_n > \frac{z_{1-\alpha}}{\sqrt{3}}\}$$

is asymptotically of level α, i.e.

$$E_0[\phi_{3,n}] \to \alpha \ .$$

In order to compute the local asymptotic power of this test, we must analyze its behavior under a sequence θ_n tending to zero at an appropriate rate. Unfortunately, the above argument, which relies heavily on the symmetry about zero implied by $\theta = 0$, does not generalize to such sequences. Luckily, the theory of contiguity, which we will now discuss, will provide a rather elegant means of tackling this problem.

3 Contiguity

Let's first consider a non-asymptotic version of the problem at hand. Let P and Q be two probability distributions on some measurable space. Suppose Q is absolutely continuous w.r.t. P, i.e., $P\{E\} = 0$ implies that $Q\{E\} = 0$ for all measurable sets E. If we denote by p and q densities of P and Q w.r.t. some common dominating measure μ (e.g., we could take $\mu = P+Q$), then it is easy to see that Q is absolutely continuous w.r.t. P if and only if $Q\{p = 0\} = 0$.

Let $T = T(X)$ be some function of X of interest. Our goal is to compute, say, $E_Q[f(T(X))]$ for some function f, but we only have the ability to compute expectations under P. Note that

$$\begin{aligned} E_Q[f(T(X))] &= \int f(T(x))q(x)d\mu(x) \\ &= \int_{p(x)>0} f(T(x))q(x)d\mu(x) \\ &= \int_{p(x)>0} f(T(x))\frac{q(x)}{p(x)}p(x)d\mu(x) \\ &= E_P[f(T(X))L(X)] \ , \end{aligned}$$

where

$$L(x) = \begin{cases} \frac{q(x)}{p(x)} & \text{if } p(x) > 0 \\ \infty & \text{if } p(x) = 0 < q(x) \\ 1 & \text{if } p(x) = q(x) = 0 \ . \end{cases}$$

The quantity $L(x)$ is actually independent of the choice of dominating measure and typically denoted by $\frac{dQ}{dP}(x)$. In mathematics, it is more commonly known as the Radon-Nikodym derivative; in statistics, people refer to it as likelihood ratio. Note that in particular we could choose f to be the indicator of some measurable set. Hence, provided that Q is absolutely continuous w.r.t. P, we can deduce the distribution of $T(X)$ under Q from knowledge of the joint distribution of $T(X)$ and $L(X)$ under P.

More generally, if we denote by F the joint distribution of T and L under

P, the above analysis implies that

$$E_Q[f(T,L)] = E_P[f(T,L)L] = \int f(t,r) r dF(t,r) \tag{1}$$

for any measurable function f.

We are interested in an asymptotic version of the above problem. For each $n \geq 1$, let P_n and Q_n be probability distributions on some measurable space. Our goal is to be able to deduce the limiting distribution of T_n under Q_n from the joint limiting distribution of T_n and the likelihood ratio, $L_n = \frac{dQ_n}{dP_n}$, under P_n. In light of the preceding discussion, it is perhaps not surprising that the correct requirement on the sequences of probability distributions is an asymptotic version of absolute continuity known as contiguity. We say that Q_n is contiguous w.r.t. P_n if $P_n\{E_n\} \to 0$ implies that $Q_n\{E_n\} \to 0$. If Q_n is contiguous w.r.t. P_n and P_n is contiguous w.r.t. Q_n, then we say that Q_n and P_n are mutually contiguous.

Before proceeding with the main result, it is useful to state some alternative characterizations of contiguity. These are useful in particular for checking whether Q_n is contiguous w.r.t. P_n. Recall that a sequence of random variables T_n with distribution P_n is tight if $\lim_{B\to\infty} \inf_n P_n\{|T_n| \leq B\} = 1$.

Theorem 3.1 *The following statements are equivalent:*

(i) Q_n *is contiguous w.r.t.* P_n*;*

(ii) $T_n \overset{P_n}{\to} 0$ *implies that* $T_n \overset{Q_n}{\to} 0$*;*

(iii) if T_n *is tight under* P_n*, then it is also tight under* Q_n*;*

(iv) if L_n *converges in distribution to* G *under* P_n *along a subsequence, then* G *has mean 1.*

For a proof of this theorem, see Theorem 12.3.2 of Lehmann and Romano (2005). Similar results are also provided in Chapter 6 of van der Vaart (1998). To shed some light on part (iv) of the result, note that

$$E_{P_n}[L_n] = \int_{p_n>0} L_n p_n d\mu_n = \int_{p_n>0} q_n d\mu_n = 1 - Q_n\{p_n = 0\} \leq 1 \ ,$$

with equality if and only if Q_n is absolutely continuous w.r.t. P_n. The equivalence between (i) and (iv) of the theorem is essentially an asymptotic version of this statement.

An important implication of this theorem is the following corollary:

Corollary 3.1 *Consider sequences of probability distributions P_n and Q_n with likelihood ratio $L_n = \frac{dQ_n}{dP_n}$. Suppose $\log L_n \stackrel{d}{\to} N(\mu, \sigma^2)$ under P_n. Then, Q_n and P_n are mutually contiguous if and only if $\mu = -\frac{1}{2}\sigma^2$.*

PROOF: With Theorem 3.1 in hand, the only hard part of the proof is showing that $\mu = -\frac{1}{2}\sigma^2$ implies that P_n is contiguous w.r.t. Q_n. We will prove the following more general assertion: If $L_n \stackrel{d}{\to} W$ under P_n and $\Pr\{W = 0\} = 0$, then P_n is continguous w.r.t. Q_n.

To this end, suppose $Q_n\{E_n\} \to 0$. Note that

$$P_n\{E_n\} = \int_{E_n \cap q_n > 0} dP_n + \int_{E_n \cap q_n = 0} dP_n \ .$$

Note further that

$$\int_{E_n \cap q_n = 0} dP_n \le P_n\{L_n = 0\} \to \Pr\{W = 0\} = 0$$

since the distribution of W is continuous at zero and $L_n \stackrel{d}{\to} W$ under P_n. Now consider $\int_{E_n \cap q_n > 0} dP_n$. For any $\epsilon > 0$,

$$\int_{E_n \cap q_n > 0} dP_n = \int_{E_n \cap q_n > 0 \cap p_n/q_n < \epsilon} dP_n + \int_{E_n \cap q_n > 0 \cap p_n/q_n \ge \epsilon} dP_n \ .$$

Note that

$$\int_{E_n \cap q_n > 0 \cap p_n/q_n < \epsilon} dP_n = \int_{E_n \cap q_n > 0 \cap p_n/q_n < \epsilon} \frac{p_n}{q_n} dQ_n \le \epsilon Q_n\{E_n\} \to 0 \ .$$

On the other hand,

$$\int_{E_n \cap q_n > 0 \cap p_n/q_n \ge \epsilon_n} dP_n \le P_n\{L_n \le 1/\epsilon\} \ .$$

To complete the argument, replace ϵ with a sequence ϵ_n tending to infinity sufficiently slowly so that $\epsilon_n Q_n\{E_n\} \to 0$. For such a sequence ϵ_n, $P_n\{L_n \leq 1/\epsilon_n\} \to \Pr\{W = 0\} = 0$. Thus, $P_n\{E_n\} \to 0$. ■

We now return to the issue at hand, i.e., computing the limiting distribution of T_n under Q_n from the joint limiting distribution of T_n and L_n under P_n. The following theorem provides us with the desired answer.

Theorem 3.2 *Suppose Q_n is contiguous w.r.t. P_n. Suppose (T_n, L_n) converges in distribution under P_n to a distribution F. Then, the limiting distribution of (T_n, L_n) under Q_n has density $rdF(t,r)$, i.e., for any continuous, bounded, real-valued f,*

$$E_{Q_n}[f(T_n, L_n)] \to \int f(t,r) r dF(t,r) \ .$$

Note that the conclusion of the theorem is essentially an asymptotic version of (1). The following useful characterization of convergence in distribution known as the Portmanteau Theorem will be useful in the proof of the preceding theorem.

Theorem 3.3 *Suppose X_n and X are random vectors. The following statements are equivalent:*

(i) $X_n \xrightarrow{d} X$;

(ii) $E[f(X_n)] \to E[f(X)]$ *for all continuous, bounded, real-valued* f.

(iii) for any open set O, $\liminf_{n\to\infty} P\{X_n \in O\} \geq P\{X \in O\}$;

(iv) for any closed set C, $\limsup_{n\to\infty} P\{X_n \in C\} \leq P\{X \in C\}$;

(v) for any set E *satisfying* $P\{X \in \partial E\} = 0$, $P\{X_n \in E\} \to P\{X \in E\}$;

(vi) $\liminf_{n\to\infty} E[f(X_n)] \geq E[f(X)]$ *for any nonnegative, continuous* f.

Proof of Theorem 3.2: We must first establish that the distribution with density $rdF(t,r)$ is a proper probability distribution on $\mathbf{R} \times \mathbf{R}_+$. First note that $rdF(t,r)$ is nonnegative on $\mathbf{R} \times \mathbf{R}_+$. Moreover, it follows from Theorem 3.1 (iv) that

$$\int_{\mathbf{R} \times \mathbf{R}_+} rdF(t,r) = 1 \ .$$

Hence, it is a proper probability distribution on $\mathbf{R} \times \mathbf{R}_+$.

Let G_n and F_n be the distributions of (T_n, L_n) under Q_n and P_n, respectively. By the Portmanteau Theorem, it suffices to show that

$$\liminf_{n \to \infty} \int f(t,r) dG_n(t,r) \geq \int f(t,r) rdF(t,r)$$

for all nonnegative, continuous f on $\mathbf{R} \times \mathbf{R}_+$.

Let p_n and q_n be densities of Q_n and P_n w.r.t. some measure μ. We have that

$$\begin{aligned}
\int f(t,r) dG_n(t,r) &= \int f(T_n, L_n) dQ_n \\
&\geq \int_{\{p_n > 0\}} f(T_n, L_n) q_n d\mu \\
&= \int f(T_n, L_n) L_n p_n d\mu \\
&= \int f(T_n, L_n) L_n dP_n \\
&= \int f(t,r) rdF_n(t,r) \ ,
\end{aligned}$$

so it suffices to show that

$$\liminf_{n \to \infty} \int f(t,r) rdF_n(t,r) \geq \int f(t,r) rdF(t,r) \ .$$

But $rf(t,r)$ is also a nonnegative, continuous function on $\mathbf{R} \times \mathbf{R}_+$, so the desired result follows again from the Portmanteau Theorem. ■

While the preceding theorem gives us the answer to our question, we will apply it most often in the special case when $(T_n, \log L_n)$ tends in distribution to a normal distribution under P_n. The following corollary specializes to this case:

Corollary 3.2 *Suppose $(T_n, \log L_n)$ tends in distribution to (T, Z) under P_n where (T, Z) is bivariate normal with mean $(\mu_1, \mu_2)'$ and variance*

$$\begin{pmatrix} \sigma_1^2 & \sigma_{1,2} \\ \sigma_{1,2} & \sigma_2^2 \end{pmatrix} .$$

Suppose further that $\mu_2 = -\frac{\sigma_2^2}{2}$, so Q_n is contiguous w.r.t. P_n. Then,

$$T_n \xrightarrow{d} N(\mu_1 + \sigma_{1,2}, \sigma_1^2)$$

under Q_n.

Proof: Let $\bar{F}$ be the limiting distribution of (T, Z). By Theorem 3.2, the limiting distribution of T_n under Q_n has density $\int_{\mathbf{R}_+} \exp(r) d\bar{F}(t, r)$. Let $\tilde{T}$ be a random variable with this distribution.

Recall that the distribution of a random variable is uniquely determined by its characteristic function. We can therefore deduce that $\tilde{T}$ has the desired distribution by showing that its characteristic function agrees with the characteristic function of the desired distribution. For an arbitrary random variable V, its characteristic function is given by $E[\exp(it'V)]$. A useful fact is that the characteristic function of a multivariate normal random variable is given by

$$\exp(i\mu' t - \frac{1}{2} t' \Sigma t)] . \tag{2}$$

Note that the characteristic function of

$$E[\exp(i\lambda \tilde{T})] = \int_{\mathbf{R} \times \mathbf{R}_+} \exp(i\lambda x) \exp(r) d\bar{F}(t, r) = E[\exp(i\lambda T + Z)] ,$$

which is simply the characteristic function of (T, Z) evaluated at $t = (\lambda, -i)$. Plugging this into (2) and using the fact that $\mu_2 = -\frac{\sigma_2^2}{2}$, we find that

$$E[\exp(i\lambda \tilde{T})] = \exp(i(\mu_1 + \sigma_{1,2}) - \frac{1}{2} \lambda^2 \sigma_1^2) ,$$

which is simply the characteristic function of a $N(\mu_1 + \sigma_{1,2}, \sigma_1^2)$ random variable, as desired. ■

4 The Wilcoxon Signed Rank Statistic Revisited

Recall the setup of the symmetric location model. Earlier we argued that under the null hypothesis that $\theta = 0$,

$$W_n = \frac{1}{\sqrt{n}} \sum_{1 \le i \le n} U_i \text{sign}(X_i) + o_{P_0}(1) \ ,$$

where $U_i = G(|X_i|)$ and G is the c.d.f. of $|X_i|$, and used this to deduce its asymptotic normality under the null hypothesis. We now wish to analyze the behavior of W_n under a sequence of alternatives of the form $P_{h/\sqrt{n}}$. For simplicity, suppose further that P_θ is the $N(\theta, 1)$ distribution.

In order to apply the above results, we must first derive the limiting disdribution of $(W_n, \log L_n)$ under P_0. To this end, let $\ell_n(\theta)$ denote the likelihood of $X_1, \ldots, X_n$ under θ. It is easy to see that

$$\ell_n(\theta) = \prod_{1 \le i \le n} \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}(X_i - \theta)^2) \ .$$

Hence,

$$\log L_n = \log(\frac{\ell_n(h/\sqrt{n})}{\ell_n(0)}) = \frac{h}{\sqrt{n}} \sum_{1 \le i \le n} X_i - \frac{h^2}{2} \ .$$

Under P_0, this tends in distribution to a $N(-h^2/2, h^2)$, so P_0 and $P_{h/\sqrt{n}}$ are mutually contiguous. From our earlier calculations, we have that

$$(W_n, \log L_n) = (\frac{1}{\sqrt{n}} \sum_{1 \le i \le n} U_i \text{sign}(X_i), \frac{h}{\sqrt{n}} \sum_{1 \le i \le n} X_i - \frac{h^2}{2}) + o_{P_0}(1) \ .$$

By the usual central limit theorem, the righthand side of the last expression tends in distribution to a bivariate normal distribution with covariance

$$\sigma_{1,2} = hE_0[G(|X_i|)|X_i|] \ .$$

It is possible to show that $\sigma_{1,2} = h/\sqrt{\pi}$. From Corollary 3.2, we therefore have that

$$W_n \xrightarrow{d} N(\frac{h}{\sqrt{\pi}}, \frac{1}{3})$$

under $h/\sqrt{n}$.

More generally, suppose f is differentiable a.e. w.r.t. Lebesgue measure and that

$$0 < I_0 = \int \frac{f'(x)^2}{f(x)} dx < \infty \ .$$

Then, it is possible to show that

$$\log L_n = \frac{1}{\sqrt{n}} \sum_{1 \le i \le n} -\frac{h f'(X_i)}{f(X_i)} - \frac{h^2}{2} I_0 + o_{P_0}(1) \ .$$

Under P_0, this tends in distribution to a $N(-\frac{h^2}{2} I_0, h^2 I_0)$, so, as before, P_0 and $P_{h/\sqrt{n}}$ are mutually contiguous. In this case, $(W_n, \log L_n)$ tends in distribution to a bivariate normal distribution with covariance

$$\sigma_{1,2} = E_0[-hG(|X_i|)\text{sign}(X_i) f'(X_i)/f(X_i)] \ .$$

Using integration by parts and the identify $G(x) = 2F(x) - 1$, one can deduce that

$$\sigma_{1,2} = 2h \int f^2(x) dx \ .$$

Thus, by Corollary 3.2,

$$W_n \xrightarrow{d} N(2h \int f^2(x) dx, \frac{1}{3})$$

under $h/\sqrt{n}$. It follows that the local asymptotic power function of W_n is given by

$$1 - \Phi(z_{1-\alpha} - 2\sqrt{3} h \int f^2(x) dx) \ .$$