

# 1 Confidence Sets

Let  $X_i, i = 1, \dots, n$  be an i.i.d. sample of observations with distribution  $P \in \mathbf{P}$ . The family  $\mathbf{P}$  may be a parametric, nonparametric, or semiparametric family of distributions. We are interested in making inferences about some parameter  $\theta(P) \in \Theta = \{\theta(P) : P \in \mathbf{P}\}$ . Typical examples of  $\theta(P)$  are the mean of  $P$  or median of  $P$ , but, more generally, it could be any function of  $P$ . Specifically, we are interested in constructing a confidence set for  $\theta(P)$ ; that is, a random set,  $C_n = C_n(X_1, \dots, X_n)$  such that

$$P\{\theta(P) \in C_n\} \approx 1 - \alpha ,$$

at least for  $n$  sufficiently large.

The typical way of constructing such sets is based off of approximating the distribution of a *root*,  $R_n = R_n(X_1, \dots, X_n, \theta(P))$ . A root is simply any real-valued function depending on both the data,  $X_i, i = 1, \dots, n$ , and the parameter of interest,  $\theta(P)$ . The idea is that if the distribution of the root were known, then one could straightforwardly construct a confidence set for  $\theta(P)$ . To illustrate this idea, let  $J_n(x, P)$  denote the distribution of  $R_n$ ; that is

$$J_n(x, P) = P\{R_n \leq x\} .$$

The notation is intended to emphasize the fact that the distribution of the root depends on both the sample size,  $n$ , and the distribution of the data,  $P$ . Using  $J_n(x, P)$ , we may choose a constant  $c$  such that

$$P\{R_n \leq c\} \approx 1 - \alpha .$$

Given such a  $c$ , the set

$$C_n = \{\theta \in \Theta : R_n(X_1, \dots, X_n, \theta) \leq c\}$$

is a confidence set in the sense described above. We may also choose  $c_1$  and  $c_2$  so that

$$P\{c_1 \leq R_n \leq c_2\} \approx 1 - \alpha .$$

Given such  $c_1$  and  $c_2$ , the set

$$C_n = \{\theta \in \Theta : c_1 \leq R_n(X_1, \dots, X_n, \theta) \leq c_2\}$$

is a confidence set in the sense described above.

### 1.1 Pivots

In some rare instances,  $J_n(x, P)$  does not depend on  $P$ . In these instances, the root is said to be *pivotal* or a *pivot*. For example, if  $\theta(P)$  is the mean of  $P$  and  $\mathbf{P} = \{N(\theta, 1) : \theta \in \mathbf{R}\}$ , then the root

$$R_n = \sqrt{n}(\bar{X}_n - \theta(P))$$

is a pivot because  $R_n \sim N(0, 1)$ . In this case, we may construct confidence sets  $C_n$  with finite-sample validity; that is,

$$P\{\theta(P) \in C_n\} = 1 - \alpha$$

for all  $n$  and  $P \in \mathbf{P}$ . If it is known that  $J_n(x, P)$  does not depend on  $P$ , but its exact form is not known or is untractable, then one may resort to simulation to approximate  $J_n(x, P)$  to any desired degree of accuracy (since the distribution does not depend on  $P$ , just pick any  $P \in \mathbf{P}$  and simulate  $J_n(x, P)$  using that  $P$ ). An example of this is given by the Kolmogorov-Smirnov statistic: Remarkably, the distribution of

$$\sqrt{n} \sup_{x \in \mathbf{R}} |\hat{F}_n(x) - F(x)|$$

does not depend on  $F$  as long as  $F$  is continuous! We may use this to construct uniform confidence bands on  $F$  provided that we assume that  $F$  is continuous.

### 1.2 Asymptotic Pivots

Sometimes, the root may not be pivotal in the sense described above, but it may be *asymptotically pivotal* or an *asymptotic pivot* in that  $J_n(x, P)$  converges in distribution to a limit distribution  $J(x, P)$  that does not depend on

$P$ . For example, if  $\theta(P)$  is the mean of  $P$  and  $\mathbf{P}$  is the set of all distributions on  $\mathbf{R}$  with a finite, nonzero variance, then

$$R_n = \frac{\sqrt{n}(\bar{X}_n - \theta(P))}{\hat{\sigma}_n}$$

is asymptotically pivotal because it converges in distribution to  $J(x, P) = \Phi(x)$ . In this case, we may construct confidence sets that are asymptotically valid in the sense that

$$\lim_{n \rightarrow \infty} P\{\theta(P) \in C_n\} = 1 - \alpha$$

for all  $P \in \mathbf{P}$ .

### 1.3 Asymptotic Approximations

Typically, the root will be neither a pivot nor an asymptotic pivot. The distribution of the root,  $J_n(x, P)$ , will typically depend on  $P$ , and, when it exists, the limit distribution of the root,  $J(x, P)$ , will, too. For example, if  $\theta(P)$  is the mean of  $P$  and  $\mathbf{P}$  is the set of all distributions on  $\mathbf{R}$  with a finite, nonzero variance, then

$$R_n = \sqrt{n}(\bar{X}_n - \theta(P))$$

converges in distribution to  $J(x, P) = \Phi(x/\sigma(P))$ . In this case, we can approximate this limit distribution with  $\Phi(x/\hat{\sigma}_n)$ , which will lead to confidence sets that are asymptotically valid in the sense described above.

Note that this third approach depends very heavily on the limit distribution  $J(x, P)$  being both known and tractable. Even if it is known, the limit distribution may be difficult to work with (e.g., it could be the supremum of some complicated stochastic process with many nuisance parameters). Moreover, even if it is known and manageable, the method may be poor in finite-samples because it essentially relies on a double approximation: first,  $J_n(x, P)$  is approximated by  $J(x, P)$ , then  $J(x, P)$  is approximated in some way by estimating the unknown parameters of the limit distribution.

## 2 The Bootstrap

The bootstrap is a fourth, more general approach to approximating  $J_n(x, P)$ . The idea is very simple: Replace the unknown  $P$  with an estimate  $\hat{P}_n$ . Given  $\hat{P}_n$ , it is possible to compute (either analytically or using simulation to any desired degree of accuracy)  $J_n(x, \hat{P}_n)$ . In the case of i.i.d. data, a typical choice is the empirical distribution (though if  $P = P(\psi)$  for some finite-dimensional parameter  $\psi$ , then one may also use  $\hat{P}_n = P(\hat{\psi}_n)$  for some estimate  $\hat{\psi}_n$  of  $\psi$ ). The hope is that whenever  $\hat{P}_n$  is “close” to  $P$  (which may be ensured, for example, by the Glivenko-Cantelli Theorem),  $J_n(x, \hat{P}_n)$  is “close” to  $J_n(x, P)$ . Essentially, this requires that  $J_n(x, P)$ , when viewed as a function of  $P$ , is continuous in an appropriate neighborhood of  $P$ . Often, this turns out to be true, but, unfortunately, it is not true in general.

### 2.1 The Nonparametric Mean

We will now consider the case where  $P$  is a distribution on  $\mathbf{R}$  and  $\theta(P)$  is the mean of  $P$ . We will consider first the root  $R_n = \sqrt{n}(\bar{X}_n - \theta(P))$ . Let  $\hat{P}_n$  denote the empirical distribution of the  $X_i, i = 1, \dots, n$ . Under what conditions is  $J_n(x, \hat{P}_n)$  “close” to  $J_n(x, P)$ ?

The sequence of distributions  $\hat{P}_n$  is a random sequence, so, as before, it is more convenient to answer the question first for a nonrandom sequence  $P_n$ . The following theorem does exactly that.

**Theorem 2.1** Let  $\theta(P)$  be the mean of  $P$  and let  $\mathbf{P}$  denote the set of all distributions on  $\mathbf{R}$  with a finite, nonzero variance. Consider the root  $R_n = \sqrt{n}(\bar{X}_n - \theta(P))$ . Let  $P_n, n \geq 1$  be a nonrandom sequence of distributions such that  $P_n$  converges in distribution to  $P$ ,  $\theta(P_n) \rightarrow \theta(P)$  and  $\sigma^2(P_n) \rightarrow \sigma^2(P)$ . Then,

- (i)  $J_n(x, P_n)$  converges in distribution to  $J(x, P) = \Phi(x/\sigma(P))$ .
- (ii)  $J_n^{-1}(1 - \alpha, P_n) = \inf\{x \in \mathbf{R} : J_n(x, P_n) \geq 1 - \alpha\}$  converges to  $J^{-1}(1 - \alpha, P) = z_{1-\alpha}\sigma(P)$ .

PROOF: (i) For each  $n$ , let  $X_{i,n}, i = 1, \dots, n$  be an i.i.d. sequence of random variables with distribution  $P_n$ . We must show that

$$\sqrt{n}(\bar{X}_{n,n} - \theta(P_n))$$

converges in distribution to  $N(0, \sigma^2(P))$ . To this end, let

$$Z_{n,i} = \frac{X_{n,i} - \theta(P_n)}{\sigma(P_n)}$$

and apply Theorem 4.1 from the notes on “Asymptotic Comparisons of Tests” (i.e., apply the Lindeberg-Feller central limit theorem). We must show that

$$\lim_{n \rightarrow \infty} E[Z_{n,i}^2 I\{|Z_{n,i}| > \epsilon\sqrt{n}\}] = 0 .$$

Let  $\epsilon > 0$  be given. By the assumption that  $P_n$  converges in distribution to  $P$  and Slutsky’s Theorem,  $Z_{n,i}$  converges in distribution to  $Z = (X - \theta(P))/\sigma(P)$ . It follows that for any  $\lambda > 0$  for which the distribution of  $Z$  is continuous at  $\lambda$ , we have that

$$E[Z_{n,i}^2 I\{|Z_{n,i}| > \lambda\}] \rightarrow E[Z^2 I\{|Z| > \lambda\}] .$$

To see this, note that

$$E[Z_{n,i}^2 I\{|Z_{n,i}| > \lambda\}] = E[Z_{n,i}^2] - E[Z_{n,i}^2 I\{|Z_{n,i}| \leq \lambda\}] .$$

The first term on the right-hand side is always equal to one and the second term converges to  $E[Z^2 I\{|Z| \leq \lambda\}]$  because of the fact that  $Z_{n,i}^2 I\{|Z_{n,i}| \leq \lambda\}$  is a bounded random variable that converges in distribution to  $Z^2 I\{|Z| \leq \lambda\}$ . (In general, convergence in distribution *does not* imply convergence of moments! You should try to find an example of this phenomenon to convince yourself.) As  $\lambda \rightarrow \infty$ ,  $E[Z^2 I\{|Z| > \lambda\}] \rightarrow 0$ . To complete the proof, note that for any fixed  $\lambda > 0$

$$E[Z_{n,i}^2 I\{|Z_{n,i}| > \epsilon\sqrt{n}\}] \leq E[Z_{n,i}^2 I\{|Z_{n,i}| > \lambda\}]$$

for  $n$  sufficiently large. Thus,

$$\sqrt{n}\bar{Z}_{n,n} \rightarrow N(0, 1)$$

under  $P_n$ . The desired result now follows from Slutsky's Theorem and the fact that  $\sigma(P_n) \rightarrow \sigma(P)$ .

(ii) This follows from part (i) and Lemma ?? below applied to  $F_n(x) = J_n(x, P)$  and  $F(x) = J(x, P)$ . ■

**Lemma 2.1** Let  $F_n, n \geq 1$  and  $F$  be nonrandom of distribution functions on  $\mathbf{R}$  such that  $F_n$  converges in distribution to  $F$ . Suppose  $F$  is continuous and strictly increasing at  $F^{-1}(1 - \alpha) = \inf\{x \in \mathbf{R} : F(x) \geq 1 - \alpha\}$ . Then,  $F_n^{-1}(1 - \alpha) = \inf\{x \in \mathbf{R} : F_n(x) \geq 1 - \alpha\} \rightarrow F^{-1}(1 - \alpha)$ .

PROOF: Let  $q = F^{-1}(1 - \alpha)$ . Fix  $\delta > 0$  and choose  $\epsilon$  so that  $0 < \epsilon < \delta$  and  $F$  is continuous at  $q - \epsilon$  and  $q + \epsilon$ . This is possible because  $F$  is continuous at  $q$  and therefore continuous in a neighborhood of  $q$ . Hence,  $F_n(q - \epsilon) \rightarrow F(q - \epsilon) < 1 - \alpha$  and  $F_n(q + \epsilon) \rightarrow F(q + \epsilon) > 1 - \alpha$ , where the inequalities follow from the assumption that  $F$  is strictly increasing at  $q$ . For  $n$  sufficiently large, we thus have that  $F_n(q - \epsilon) < 1 - \alpha$  and  $F_n(q + \epsilon) > 1 - \alpha$ . It follows that  $q - \epsilon \leq F_n^{-1}(1 - \alpha) \leq q + \epsilon$  for such  $n$ . ■

We are now ready to pass from the nonrandom sequence  $P_n$  to the random sequence  $\hat{P}_n$ .

**Theorem 2.2** Let  $\theta(P)$  be the mean of  $P$  and let  $\mathbf{P}$  denote the set of all distributions on  $\mathbf{R}$  with a finite, nonzero variance. Consider the root  $R_n = \sqrt{n}(\bar{X}_n - \theta(P))$ . Then,

- (i)  $J_n(x, \hat{P}_n)$  converges in distribution to  $J(x, P) = \Phi(x/\sigma(P))$  a.s.
- (ii)  $J_n^{-1}(1 - \alpha, \hat{P}_n)$  converges to  $J^{-1}(1 - \alpha, P) = z_{1-\alpha}\sigma(P)$  a.s.

PROOF: By the Glivenko-Cantelli Theorem,

$$\sup_{x \in \mathbf{R}} |\hat{P}_n((-\infty, x]) - P((-\infty, x])| \rightarrow 0$$

a.s. This implies that  $\hat{P}_n$  converges in distribution to  $P$  a.s. Since  $|x| \leq 1 + x^2$  and that  $\sigma^2(P) < \infty$ , we have that  $E[|X|] \leq 1 + E[X^2] < \infty$ . Thus, we

may apply the Strong Law of Large Numbers to conclude that  $\theta(\hat{P}_n) = \bar{X}_n$  converges to  $\theta(P)$  a.s. and  $\sigma(\hat{P}_n)$  converges to  $\sigma(P)$  a.s. Thus, w.p. 1,  $\hat{P}_n$  satisfies the assumptions of Theorem ???. The conclusions of the theorem now follow. ■

We will now consider the root  $R_n = \frac{\sqrt{n}(\bar{X}_n - \theta(P))}{\hat{\sigma}_n}$ . The following theorem, which parallels Theorem ??, provides conditions under which  $J_n(x, \hat{P}_n)$  is “close” to  $J_n(x, P)$ . A key step in the proof will be to show that  $\hat{\sigma}_n$  converges in probability to  $\sigma(P)$  under an appropriate sequence of distributions. For this reason, we will need the following weak law of large numbers for a triangular array:

**Lemma 2.2** For each  $n$ , let  $Y_{n,i}, i = 1, \dots, n$  be an i.i.d. sequence of random variables with distribution  $F_n$  on  $\mathbf{R}$ . Suppose  $F_n$  converges in distribution to  $F$  and that  $E[|Y_{n,i}|] \rightarrow E[|Y|] < \infty$ , where  $Y \sim F$ . Then,  $\bar{Y}_{n,n}$  converges in probability to  $E[Y]$  under  $F_n$ .

The proof of this is nontrivial, so we postpone it until later.

**Theorem 2.3** Let  $\theta(P)$  be the mean of  $P$  and let  $\mathbf{P}$  denote the set of all distributions on  $\mathbf{R}$  with a finite, nonzero variance. Consider the root  $R_n = \frac{\sqrt{n}(\bar{X}_n - \theta(P))}{\hat{\sigma}_n}$ . Let  $P_n, n \geq 1$  be a nonrandom sequence of distributions such that  $P_n$  converges in distribution to  $P$ ,  $\theta(P_n) \rightarrow \theta(P)$  and  $\sigma^2(P_n) \rightarrow \sigma^2(P)$ . Then,

- (i)  $J_n(x, P_n)$  converges in distribution to  $J(x, P) = \Phi(x)$ .
- (ii)  $J_n^{-1}(1 - \alpha, P_n) = \inf\{x \in \mathbf{R} : J_n(x, P_n) \geq 1 - \alpha\}$  converges to  $J^{-1}(1 - \alpha, P) = z_{1-\alpha}$ .

PROOF: (i) For each  $n$ , let  $X_{n,i}, i = 1, \dots, n$  be an i.i.d. sequence of random variables with distribution  $P_n$ . The proof of Theorem ??? shows that

$$\sqrt{n}(\bar{X}_{n,n} - \theta(P_n))$$

converges in distribution to a  $N(0, \sigma^2(P))$  under  $P_n$ . Therefore, by Slutsky's Theorem, it suffices to show that

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{1 \leq i \leq n} X_{n,i}^2 - \bar{X}_{n,n}^2$$

converges in probability to  $\sigma^2(P)$  under  $P_n$ . To do this, we will apply Lemma ?? twice, once with  $Y_{n,i} = X_{n,i}^2$  and then with  $Y_{n,i} = X_{n,i}$ .

Let  $X \sim P$ . Recall that by assumption, (i)  $X_{n,i}$  converges in distribution to  $X$  (and thus by the Continuous Mapping Theorem,  $X_{n,i}^2$  converges in distribution to  $X^2$ ), (ii)  $\theta(P_n) = E[X_{n,i}] \rightarrow E[X] = \theta(P)$ , and (iii)  $\sigma^2(P_n) = E[X_{n,i}^2] - E[X_{n,i}]^2 \rightarrow E[X^2] - E[X]^2 = \sigma^2(P)$ . It follows that  $E[X_{n,i}^2] \rightarrow E[X^2]$ . This last fact implies (with a tiny bit of work) that  $E[|X_{n,i}|] \rightarrow E[|X|]$  (use the inequality  $|x| \leq 1+x^2$  and Lebesgue Dominated Convergence Theorem). Thus, we may apply Lemma ?? with both  $Y_{n,i} = X_{n,i}^2$  and then with  $Y_{n,i} = X_{n,i}$  to conclude that

$$\frac{1}{n} \sum_{1 \leq i \leq n} X_{n,i}^2$$

converges in probability to  $E[X^2]$  under  $P_n$  and  $\bar{X}_{n,n}$  converges in probability to  $E[X]$  under  $P_n$ . The desired result now follows.

(ii) This follows as before from part (i) and Lemma ?? applied to  $F_n(x) = J_n(x, P)$  and  $F(x) = J(x, P)$ . ■

We can now pass, as before, from the nonrandom sequence  $P_n$  to the random sequence  $\hat{P}_n$ .

**Theorem 2.4** Let  $\theta(P)$  be the mean of  $P$  and let  $\mathbf{P}$  denote the set of all distributions on  $\mathbf{R}$  with a finite, nonzero variance. Consider the root  $R_n = \frac{\sqrt{n}(\bar{X}_n - \theta(P))}{\hat{\sigma}_n}$ . Then,

- (i)  $J_n(x, \hat{P}_n)$  converges in distribution to  $J(x, P) = \Phi(x)$  a.s.
- (ii)  $J_n^{-1}(1 - \alpha, \hat{P}_n)$  converges to  $J^{-1}(1 - \alpha, P) = z_{1-\alpha}$  a.s.



PROOF: The proof follows as before from the Glivenko-Cantelli Theorem and the Strong Law of Large Numbers. ■

It now follows from Slutsky's Theorem that confidence sets of the form  $C_n = \{\theta \in \mathbf{R} : R_n(X_1, \dots, X_n, \theta) \leq J_n^{-1}(1 - \alpha, \hat{P}_n)\}$  or  $C_n = \{\theta \in \mathbf{R} : J_n^{-1}(\frac{\alpha}{2}, \hat{P}_n) \leq R_n(X_1, \dots, X_n, \theta) \leq J_n^{-1}(1 - \frac{\alpha}{2}, \hat{P}_n)\}$ , with  $R_n = \sqrt{n}(\bar{X}_n - \theta(P))$  or  $R_n = \sqrt{n} \frac{(\bar{X}_n - \theta(P))}{\hat{\sigma}_n}$  satisfy

$$P\{\theta(P) \in C_n\} \rightarrow 1 - \alpha \quad (1)$$

for all  $P \in \mathbf{P}$ .

Of course, even a confidence set  $C_n$  based off of the asymptotic normality of either root would satisfy (??). It can be shown under certain conditions (that ensure the existence of so-called Edgeworth expansions of  $J_n(x, P)$ ) that one-sided confidence sets  $C_n$  based off of such an asymptotic approximation satisfy

$$P\{\theta(P) \in C_n\} - (1 - \alpha) = O(n^{-1/2}) . \quad (2)$$

One-sided confidence sets based off of the bootstrap and the root  $R_n = \sqrt{n}(\bar{X}_n - \theta(P))$  also satisfy (??), though there is some evidence to suggest that it does a bit better in the size of  $O(n^{-1/2})$  term. On the other hand, one-sided confidence sets based off of the bootstrap and the root  $R_n = \frac{\sqrt{n}(\bar{X}_n - \theta(P))}{\hat{\sigma}_n}$  satisfy

$$P\{\theta(P) \in C_n\} - (1 - \alpha) = O(n^{-1}) . \quad (3)$$

One-sided confidence sets that satisfy only (??) are said to be first-order accurate, where as one-sided confidence sets that satisfy (??) are said to be second-order accurate. See Section 15.5 of Lehmann and Romano (2005) for further details.

We now return to the proof of Lemma ?? that was used in the proof of Theorem ??.

PROOF OF LEMMA ??: Note that  $E[|Y_{n,i}|] \rightarrow E|Y|$ , implies that  $E[Y_{n,i}] \rightarrow E[Y]$ . Therefore, it is enough to prove that  $\bar{Y}_{n,n} - E[Y_{n,i}]$  converges in probability to zero under  $F_n$ .

We may assume w.l.o.g. that  $E[Y_{n,i}] = 0$ . Arguing as in the proof of Theorem ??, we have that

$$\lim_{\lambda \rightarrow \infty} \limsup_{n \rightarrow \infty} E[|Y_{n,i}| I\{|Y_{n,i}| > \lambda\}] = 0 . \quad (4)$$

Let  $\epsilon > 0$  be given and define  $Z_{n,i} = Y_{n,i} I\{|Y_{n,i}| \leq n\}$ . Note that

$$P\{|\bar{Y}_{n,n}| > \epsilon\} \leq P\{|Z_{n,n}| > \epsilon\} + P\{\bar{Y}_{n,n} \neq \bar{Z}_{n,n}\} .$$

Note that

$$\begin{aligned} P\{\bar{Y}_{n,n} \neq \bar{Z}_{n,n}\} &\leq P\left\{ \bigcup_{1 \leq i \leq n} P\{Y_{n,i} \neq Z_{n,i}\} \right\} \\ &= nP\{Y_{n,i} \neq Z_{n,i}\} = nP\{|Y_{n,i}| > n\} . \end{aligned}$$

Furthermore, by Chebychev's Inequality,

$$P\{|\bar{Z}_{n,n}| > \epsilon\} \leq \frac{E[Z_{n,i}^2]}{n\epsilon^2} .$$

Thus,

$$P\{|\bar{Y}_{n,n}| > \epsilon\} \leq \frac{E[Z_{n,i}^2]}{n\epsilon^2} + nP\{|Y_{n,i}| > n\} .$$

For  $t > 0$ , let

$$\begin{aligned} \tau_n(t) &= tP\{|Y_{n,i}| > t\} = t(1 - F_n(t) + F_n(-t)) \\ \kappa_n(t) &= \frac{1}{t}E[Z_{n,i}^2] = \frac{1}{t}E[|Y_{n,i}| I\{|Y_{n,i}| \leq t\}] \\ &= \frac{1}{t} \int_{-t}^t x^2 dF_n . \end{aligned}$$

Hence,

$$P\{|\bar{Y}_{n,n}| > \epsilon\} \leq \frac{\kappa_n(n)}{\epsilon^2} + \tau_n(n) .$$

It therefore suffices to show that  $\tau_n(n) \rightarrow 0$  and  $\kappa_n(n) \rightarrow 0$ . Since

$$tP\{|Y_{n,i}| > t\} \leq E[|Y_{n,i}| I\{|Y_{n,i}| > t\}] , \quad (5)$$

$\tau_n(n) = nP\{|Y_{n,i}| > n\} \rightarrow 0$ . It is possible to show (using integration by parts and some persistence) that

$$\kappa_n(t) = -\tau_n(t) + \frac{2}{t} \int_0^t \tau_n(x) dx .$$

To complete the proof, we thus only need to show that

$$\frac{1}{n} \int_0^n \tau_n(x) dx \rightarrow 0 .$$

To this end, let  $\delta > 0$  be given. Note that (??) implies that

$$\frac{1}{n} \int_0^n \tau_n(x) dx \leq \frac{1}{n} \int_0^n E[|Y_{n,i}| I\{|Y_{n,i}| > x\}] dx . \quad (6)$$

Using (??) and (??), choose  $n_0$  and  $\lambda_0$  so that for all  $n > n_0$ ,

$$E[|Y_{n,i}| I\{|Y_{n,i}| > \lambda_0\}] < \frac{\delta}{2} .$$

For all  $x \geq \lambda_0$  and  $n > n_0$ , we thus have that

$$E[|Y_{n,i}| I\{|Y_{n,i}| > x\}] \leq E[|Y_{n,i}| I\{|Y_{n,i}| > \lambda_0\}] < \frac{\delta}{2} .$$

For all  $x \leq \lambda_0$  and  $n > n_0$ , we have that

$$\begin{aligned} E[|Y_{n,i}| I\{|Y_{n,i}| > x\}] &\leq E[|Y_{n,i}|] \\ &= E[|Y_{n,i}| I\{|Y_{n,i}| \leq \lambda_0\}] + E[|Y_{n,i}| I\{|Y_{n,i}| > \lambda_0\}] \\ &\leq \lambda_0 + \frac{\delta}{2} . \end{aligned}$$

Finally, using (??) and these last two inequalities, we have for  $n > n_0$  and  $n > \lambda_0$  that

$$\frac{1}{n} \int_0^n \tau_n(x) dx \leq \frac{\lambda_0(\lambda_0 + \frac{\delta}{2})}{n} + \frac{\delta}{2} ,$$

which is less than  $\delta$  for all  $n$  sufficiently large. ■

### 3 The Multivariate Nonparametric Mean

The convergence in distribution of random vectors in  $\mathbf{R}^k$  can be deduced from convergence in distribution of appropriate random variables on  $\mathbf{R}$  by means of the following useful result, known as the Cramer-Wold device:

**Theorem 3.1** A sequence of random vectors  $X_n$  on  $\mathbf{R}^k$  satisfies  $X_n \xrightarrow{d} X$  if and only if  $t'X_n \xrightarrow{d} t'X$  for all  $t \in \mathbf{R}^k$ .

Using the Cramer-Wold device it is straightforward to generalize the arguments of the preceding section to the multivariate case.

**Theorem 3.2** Let  $\theta(P)$  be the mean of  $P$  and let  $\mathbf{P}$  denote the set of distributions on  $\mathbf{R}^k$  whose variance,  $\Sigma(P)$ , exists and has at least one nonzero component. Consider the root  $R_n = \|\sqrt{n}(\bar{X}_n - \theta(P))\|$ , where  $\|\cdot\|$  is the usual norm on  $\mathbf{R}^k$ . Let  $P_n, n \geq 1$  be a nonrandom sequence of distributions such that  $P_n$  converges in distribution to  $P$ ,  $\theta(P_n) \rightarrow \theta(P)$ , and  $\Sigma(P_n) \rightarrow \Sigma(P)$ . Then,

- (i)  $J_n(x, P_n)$  converges in distribution to  $J(x, P)$ , the distribution of  $\|Z\|$ , where  $Z \sim N(0, \Sigma(P))$ .
- (ii)  $J_n^{-1}(1 - \alpha, P_n)$  converges to  $J^{-1}(1 - \alpha, P)$ .

PROOF: Let  $X_{i,n}, i = 1, \dots, n$  be an i.i.d. sequence of random variables with distribution  $P_n$ . We first show that

$$\sqrt{n}(\bar{X}_{n,n} - \theta(P_n)) \tag{7}$$

converges in distribution under  $P_n$  to a  $N(0, \Sigma(P))$ . To this end, we apply the Cramer-Wold device. Let  $t \in \mathbf{R}^k$  be given. There are two cases to consider:  $t'\Sigma(P)t > 0$  and  $t'\Sigma(P)t = 0$ . In the former case, the desired result follows immediately from part (i) of Theorem ???. In the latter case, we have by Chebychev's inequality that (??) converges in distribution to zero. The conclusion (i) now follows from the continuous mapping theorem; (ii) follows from Lemma ?? upon realizing that the requirement that  $\Sigma(P)$  has at least one nonzero component implies that  $J(x, P)$  is continuous and strictly increasing at its  $1 - \alpha$  quantile. ■

We can now pass, as before, from the nonrandom sequence to the random sequence  $\hat{P}_n$ .

**Theorem 3.3** Let  $\theta(P)$  be the mean of  $P$  and let  $\mathbf{P}$  denote the set of all distributions on  $\mathbf{R}^k$  whose variance,  $\Sigma(P)$ , exists and has at least one

nonzero component. Consider the root  $R_n = \|\sqrt{n}(\bar{X}_n - \theta(P))\|$ , where  $\|\cdot\|$  is the usual norm on  $\mathbf{R}^k$ . Then,

(i)  $J_n(x, \hat{P}_n)$  converges in distribution to  $J(x, P)$ , the distribution of  $\|Z\|$ , where  $Z \sim N(0, \Sigma(P))$ , a.s.

(ii)  $J_n^{-1}(1 - \alpha, \hat{P}_n)$  converges to  $J^{-1}(1 - \alpha, P)$  a.s.

PROOF: The proof follows as before from the Glivenko-Cantelli Theorem (in this case on  $\mathbf{R}^k$  over an appropriate class of sets) and the Strong Law of Large Numbers. ■

As in the univariate case, it is straightforward to extend the argument to the root

$$R_n = \|\hat{\Sigma}_n^{-1/2}(\sqrt{n}(\bar{X}_n - \theta(P)))\| ,$$

where  $\hat{\Sigma}_n = \Sigma(\hat{P}_n)$ . In this case, one must assume further that  $\Sigma(P)$  is nonsingular.

Note that the argument above did not depend on  $\|\cdot\|$  in any essential way; we could just as easily replace it with any other norm on  $\mathbf{R}^k$ .

## 4 The Sample Median

Let  $P$  be a distribution on  $\mathbf{R}$  and denote by  $\theta(P)$  its median. Let  $\hat{\theta}_n = \theta(\hat{P}_n)$  the sample median of the  $X_i, i = 1, \dots, n$  and consider the root  $R_n = \sqrt{n}(\hat{\theta}_n - \theta(P))$ .

It is useful first to consider the limiting behavior of  $R_n$  for a fixed distribution  $P$ . This is done in the following theorem (where, for expositional purposes, we've switched from  $P$  to  $F$ ):

**Theorem 4.1** Let  $X_i, i = 1, \dots, n$  be an i.i.d. sequence of random variables with c.d.f.  $F$ . Suppose  $F$  is differentiable at  $\theta(F)$  with derivative  $f$  and  $f(\theta(F)) > 0$ . Then,

$$\sqrt{n}(\hat{\theta}_n - \theta(F)) \xrightarrow{d} N\left(0, \frac{1}{4f^2(\theta(F))}\right) .$$

PROOF: Suppose first that  $n$  tends to infinity through odd values only; the proof for even values is similar. For any  $x \in \mathbf{R}$ ,

$$P\{\sqrt{n}(\hat{\theta}_n - \theta(F)) \leq x\} = P\{S_n \leq \frac{n-1}{2}\},$$

where  $S_n \sim \text{Binomial}(n, p_n)$  and

$$p_n = 1 - F(\theta(F) + \frac{x}{\sqrt{n}}).$$

Moreover,

$$P\{S_n \leq \frac{n-1}{2}\} = P\{\frac{S_n - np_n}{[np_n(1-p_n)]^{1/2}} \leq \frac{\sqrt{n}(\frac{1}{2} - p_n) - \frac{1}{2\sqrt{n}}}{[p_n(1-p_n)]^{1/2}}\}.$$

Because  $p_n \rightarrow 1/2$ , it follows from the Lindeberg-Feller Central Limit Theorem that

$$\frac{S_n - np_n}{[np_n(1-p_n)]^{1/2}} \xrightarrow{d} N(0, 1).$$

Furthermore,

$$\sqrt{n}(\frac{1}{2} - p_n) = \sqrt{n}(F(\theta(F) + \frac{x}{\sqrt{n}}) - \frac{1}{2}) \rightarrow xf(\theta(F)).$$

Hence,

$$\frac{\sqrt{n}(\frac{1}{2} - p_n) - \frac{1}{2\sqrt{n}}}{[p_n(1-p_n)]^{1/2}} \rightarrow x2f(\theta(F)).$$

The desired result now follows from Slutsky's Theorem. ■

We now need to determine for which sequences  $F_n$ ,  $\sqrt{n}(\hat{\theta}_n - \theta(F_n))$  converges in distribution to the *same* limiting distribution. By mimicking the proof of the above result, we see that this will be the case when

$$\sqrt{n}(F_n(\theta(F_n) + \frac{x}{\sqrt{n}}) - \frac{1}{2}) \rightarrow xf(\theta(F)).$$

Note that this convergence implies in particular that

$$F_n(\theta(F_n) + \frac{x}{\sqrt{n}}) \rightarrow \frac{1}{2}.$$

The following theorem summarizes this conclusion.

**Theorem 4.2** Suppose  $F$  is a distribution satisfying the assumptions of the preceding theorem and let  $F_n, n \geq 1$  be a sequence of distributions satisfying

$$\sqrt{n}(F_n(\theta(F_n) + \frac{x}{\sqrt{n}}) - \frac{1}{2}) \rightarrow xf(\theta(F)) .$$

Let  $J_n(x, F_n)$  denote the distribution of  $R_n = \sqrt{n}(\hat{\theta}_n - \theta(F))$  under  $F_n$ . Then,

(i)  $J_n(x, F_n) \xrightarrow{d} J(x, F) = \Phi(2f(\theta(F))x);$

(ii)  $J_n^{-1}(1 - \alpha, F_n) \rightarrow J^{-1}(1 - \alpha, F) = 2f(\theta(F))z_{1-\alpha}.$

PROOF: Follows essentially from the proof of the preceding theorem and earlier results. ■

Using an almost sure representation theorem, it is possible to show that there exist *versions* of  $\hat{F}_n$  (i.e., having the same distributional properties) that satisfy the requirements on  $F_n$  in the preceding theorem almost surely. This implies that for the *actual*  $\hat{F}_n$ , the conclusions of the theorem hold in probability, which is good enough for our purposes. See Beran (1984) for details.