5.3 Asymptotic Distribution of the Sample Variance

We seek a WLLN and a CLT for S_n^2, the sample variance.

THEOREM 5.3.1 *Assume $X_1, \ldots, X_n \ldots \overset{iid}{\sim} (\mu, \sigma^2, \mu_3, \mu_4)$. Then $S_n^2 \xrightarrow{P} \sigma^2$, $S_n \xrightarrow{P} \sigma$, and $\overline{X}_n/S_n \xrightarrow{P} \mu/\sigma$.*

PROOF $S_n^2 = [n/(n-1)]\left(\frac{\Sigma X_i^2}{n} - \overline{X}_n^2\right)$. Have $\overline{X}_n \xrightarrow{P} \mu$ which implies (Corollary 5.1.3.b with $g(z) = z^2$) $\overline{X}_n^2 \xrightarrow{P} \mu^2$. Also, WLLN on $X_1^2, \ldots, X_n^2, \ldots$ give $\frac{1}{n}\Sigma_1^n X_i^2 \xrightarrow{P} \mu_2'$. Corollary 5.1.3.c with $g(u,v) = u - v$ gives $\frac{\Sigma X_i^2}{n} - \overline{X}_n^2 \xrightarrow{P} \mu_2' - \mu^2 = \sigma^2$ and since $n/(n-1) \longrightarrow 1$, $S_n^2 = [n/(n-1)]\left(\frac{\Sigma X_i^2}{n} - \overline{X}_n^2\right) \xrightarrow{P} \sigma^2$. $S_n^2 \xrightarrow{P} \sigma^2$ implies $S_n \xrightarrow{P} \sigma$ by Corollary 5.1.3.b with $g(z) = z^{1/2}$. Finally, $\overline{X}_n \xrightarrow{P} \mu$ and $S_n \xrightarrow{P} \sigma$ implies $\overline{X}_n/S_n \xrightarrow{P} \mu/\sigma$ again via Corollary 5.1.3.c with $g(u,v) = u/v$.
////

EXAMPLE 17 Assume $X_1, \ldots, X_n, \ldots \overset{iid}{\sim} (\mu, \sigma^2)$. Define $T_n = \sqrt{n}(\overline{X}_n - \mu)/S_n$, which is a t-type statistic. By the CLT, $\sqrt{n}(\overline{X}_n - \mu)/\sigma \xrightarrow{d} N_{0,1}$. By Theorem 5.3.1, $S_n \xrightarrow{P} \sigma$ which implies $\sigma/S_n \xrightarrow{P} 1$ by Corollary 5.1.3.b. Hence, by Slutzky's Theorem,

$$(\sigma/S_n) \cdot \sqrt{n}(\overline{X}_n - \mu)/\sigma = T_n \xrightarrow{d} 1 \cdot N_{0,1} = N_{0,1}$$

This result will be used in Chapter VIII to construct large sample confidence interval estimator for μ of the form $\overline{X}_n \pm z\, S_n/\sqrt{n}$, where z is a quantile of the standard normal determined by the desired confidence level.
////

THEOREM 5.3.2 *Assume $X_1, \ldots, X_n, \ldots \overset{iid}{\sim} (\mu, \sigma^2, \mu_3, \mu_4)$. Let S_n^2 be the sample variance of $X_1, \ldots, X_n$. Then $\sqrt{n}(S_n^2 - \sigma^2)/\sigma^2\sqrt{\kappa - 1} \xrightarrow{d} N_{0,1}$, where $\kappa = \mu_4/\sigma^4$. Or, $S_n^2 \overset{asymp}{\sim} N(\sigma^2, \sigma^4(\kappa - 1)/n)$.*
5.3.1

PROOF Write

$$\begin{aligned}
S_n^2 &= \tfrac{1}{n-1}\Sigma(X_i - \overline{X}_n)^2 \\
&= \tfrac{1}{n}\Sigma(X_i - \overline{X}_n)^2 + \tfrac{1}{n}S_n^2 \\
&= \tfrac{1}{n}\Sigma[(X_i - \mu) - (\overline{X}_n - \mu)]^2 + \tfrac{1}{n}S_n^2 \\
&= \tfrac{1}{n}\left\{\Sigma(X_i - \mu)^2 - 2(\overline{X}_n - \mu)\Sigma(X_i - \mu) + n(\overline{X}_n - \mu)^2\right\} + \tfrac{1}{n}S_n^2 \\
&= \tfrac{1}{n}\Sigma(X_i - \mu)^2 - (\overline{X}_n - \mu)^2 + \tfrac{1}{n}S_n^2, \qquad \text{which implies} \\
\sqrt{n}(S_n^2 - \sigma^2) &= \tfrac{1}{\sqrt{n}}\Sigma(X_i - \mu)^2 - \sqrt{n}\,\sigma^2 - \sqrt{n}(\overline{X}_n - \mu)^2 + \tfrac{1}{\sqrt{n}}\,S_n^2 \\
&= \sqrt{n}\left[\tfrac{\Sigma(X_i - \mu)^2}{n} - \sigma^2\right] - \sqrt{n}(\overline{X}_n - \mu)(\overline{X}_n - \mu) + \tfrac{1}{\sqrt{n}}\,S_n^2
\end{aligned}$$
5.3.2

Lets consider the three summands on the RHS of Eq. 5.3.2 in reverse order. $(1/\sqrt{n})S_n^2 \xrightarrow{P} 0$ since $(1/\sqrt{n}) \longrightarrow 0$ and $S_n^2 \xrightarrow{P} \sigma^2$. $\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{d} N_{0,\sigma^2}$ by CLT and $\overline{X}_n - \mu \xrightarrow{P} 0$ by WLLN, so by Slutzky's Theorem $(\overline{X}_n - \mu)\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{d} 0 \cdot N_{0,\sigma^2} = 0$. That is, both the middle summand and the last summand converge in probability to zero, hence (Slutzky again) the limiting distribution of $\sqrt{n}(S_n^2 - \sigma^2)$ is the same as the limiting distribution of $\sqrt{n}\left[\frac{\Sigma(X_i-\mu)^2}{n} - \sigma^2\right]$. But, employing the CLT on $(X_1 - \mu)^2, \ldots, (X_n - \mu)^2, \ldots$ we get $\sqrt{n}\left[\frac{\Sigma(X_i-\mu)^2}{n} - \sigma^2\right] \xrightarrow{d} N_{0,\,\text{var}[(X_1-\mu)^2]}$ noting that $\mathcal{E}[(X_1 - \mu)^2] = \sigma^2$. But $\text{var}[(X_1 - \mu)^2] = \mathcal{E}[(X_1 - \mu)^4] - \mathcal{E}^2[(X_1 - \mu)^2] = \mu_4 - \sigma^4 = \sigma^4(\kappa - 1)$. Summarizing, we have

$$\sqrt{n}(S_n^2 - \sigma^2) \xrightarrow{d} N_{0,\sigma^4(\kappa-1)}, \quad \text{or}$$

$$\sqrt{n}(S_n^2 - \sigma^2)/\sigma^2\sqrt{\kappa - 1} \xrightarrow{d} N_{0,1}, \quad \text{or}$$

$$S_n^2 \overset{\text{asymp}}{\sim} N(\sigma^2, \sigma^4(\kappa - 1)/n) \qquad ////$$

Remark The variance of the asymptotic distribution of S_n^2, which is $(1/n)[\mu_4 - \sigma^4]$, is close to the exact variance of S_n^2, which is $(1/n)[\mu_4 - (n-3)\sigma^4/(n-1)]$, as given in Theorem 3.3. ////

Remark Theorem 5.3 in conjunction with Slutzky's Theorem will be used to construct a large sample confidence interval estimator of σ^2 of the form $S_n^2 \pm z\, S_n^2\sqrt{K_n - 1}/\sqrt{n}$, where K_n is the sample kurtosis.

PROOF We know $S_n^2 \xrightarrow{P} \sigma^2$; it can be shown that $K_n \xrightarrow{P} \kappa$ (see the Problems). Hence, $\frac{\sigma^2\sqrt{\kappa - 1}}{S_n^2\sqrt{K_n - 1}} \xrightarrow{P} 1$, so that $\frac{\sigma^2\sqrt{\kappa - 1}}{S_n^2\sqrt{K_n - 1}} \cdot \frac{\sqrt{n}(S_n^2 - \sigma^2)}{\sigma^2\sqrt{\kappa - 1}} = \frac{\sqrt{n}(S_n^2 - \sigma^2)}{S_n^2\sqrt{K_n - 1}} \xrightarrow{d} N_{0,1}$ which will be used to construct a large sample confidence interval estimator of σ^2. ////

COROLLARY 5.3 Under the same conditions as Theorem 5.3, the *sample standard deviation*, S_n satisfies

$$\sqrt{n}(S_n - \sigma) \xrightarrow{d} N_{0,\sigma^2(\kappa-1)/4}, \quad \textit{or}$$

$$S_n \overset{\text{asymp}}{\sim} N(\sigma, \sigma^2(\kappa - 1)/4n).$$

PROOF We have $\sqrt{n}(S_n^2 - \sigma^2) \xrightarrow{d} N_{0,\sigma^4(\kappa-1)}$. Take $g(z) = \sqrt{z}$, so $g'(z) = 1/2\sqrt{z}$, then Theorem 5.1.4 gives $\sqrt{n}(S_n - \sigma) \xrightarrow{d} (1/2\sqrt{\sigma^2})N_{0,\sigma^4(\kappa-1)} = N_{0,\sigma^2(\kappa-1)/4}$. ////

5.4 Higher Dimensions

Heretofore we have considered convergence of sequences of random variables; in this section we generalize some of our results from random variables to random vectors. Our goal is to gain some acquaintence with asymptotic *joint* distributions. For example, can one find an asymptotic joint distribution of, say, the sample coefficient of variation and the sample skewness?

5.4.1 Convergences in Distribution and Probability

This section generalizes most of the results in Section 5.1. There we addressed sequences of random variables, here we consider sequences of random vectors. Let $\{\underline{X}_n\}_{n=1}^{a}$ be a sequence of k-dimensional random vectors and let $\underline{X}$ be a k-dimensional random vector.

Definition 16 $\underline{X}_n \xrightarrow{d} \underline{X}$ if and only if $F_{\underline{X}_n}(\underline{z}) \longrightarrow F_{\underline{X}}(\underline{z})$ as $n \longrightarrow \infty$ for every $\underline{z}$ a continuity point of $F_{\underline{X}}(\cdot)$.

$\underline{X}_n \xrightarrow{P} \underline{X}$ if and only if $P\left[\sqrt{\sum_{j=1}^{k}(X_{j,n} - X_j)^2} > \varepsilon\right] \longrightarrow 0$ as $n \longrightarrow \infty$ for every $\varepsilon > 0$, where

$$\underline{X}_n = \begin{pmatrix} X_{1,n} \\ \vdots \\ X_{k,n} \end{pmatrix} \text{ and } \underline{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_k \end{pmatrix}. \qquad 5.4.1$$

$\underline{X}$ is called the *limiting random vector* and $F_{\underline{X}}(\cdot)$ is called the *limiting joint distribution.*

////

Convergence in probability of random vectors is closely related to convergence in probability of random variables. In fact, the following Remark follows from the respective definitions.

Remark $\underline{X}_n \xrightarrow{P} \underline{X}$ if and only if $X_{j,n} \xrightarrow{P} X_j$ for $j = 1, \ldots, k$, where $\underline{X}_n$ and $\underline{X}$ are as in Eq. 5.4.1. That is, vector convergence in probability is equivalent to component convergence in probability for each component.

////

We now state the vector versions of Theorems 5.1.2, 5.1.3, and 5.1.4, without proofs.

THEOREM 5.4.1 *(Continuity Theorem for moment generating functions in higher dimensions.) Under existence of all pertinent moment generating functions in a neighborhood of* $\underline{0}$,

$$\underline{X}_n \xrightarrow{d} \underline{X} \text{ if } m_{\underline{X}_n}(\underline{t}) \longrightarrow m_{\underline{X}}(\underline{t})$$

as $n \longrightarrow \infty$ *for all* $\underline{t}$ *in some neighborhood of* $\underline{0}$.

////

THEOREM 5.4.2 *(Higher dimensional Slutzky's Theorem.) For sequence* $\{\underline{X}_n\}$ *of* k_1*-dimensional vectors and sequence* $\{\underline{Y}_n\}$ *of* k_2*-dimensional vectors and function* $g(\cdot,\cdot)$ *defined as:*

$$k_1\text{-space} \times k_2\text{-space} \xrightarrow{g(\cdot,\cdot)} r\text{-space},$$

if $g(\cdot,\underline{a})$ *is continuous and if* $\underline{X}_n \xrightarrow{d} \underline{X}$ *and if* $\underline{Y}_n \xrightarrow{P} \underline{a}$ *then* $g(\underline{X}_n,\underline{Y}_n) \xrightarrow{d} g(\underline{X},\underline{a})$. ////

Comment Theorem 5.4.2 is relatively easy to state, but to make it useful one has to recognize the need to derive the r-variate joint distribution of $g(\underline{X},\underline{a})$.

COROLLARY 5.4.2.1 Take $\underline{X}$ and $\underline{X}_n$ k-dimensional and $g(\cdot)$ continuous and k-space $\xrightarrow{g(\cdot)}$ r-space. If $\underline{X}_n \xrightarrow{d} \underline{X}$ then $g(\underline{X}_n) \xrightarrow{d} g(\underline{X})$. ////

COROLLARY 5.4.2.2 Again take $\underline{X}_n$ and $\underline{a}$ k-dimensional $g(\cdot)$ continuous at $\underline{a}$ and k-space $\xrightarrow{g(\cdot)}$ r-space. If $\underline{X}_n \xrightarrow{P} \underline{a}$ then $g(\underline{X}_n) \xrightarrow{P} g(\underline{a})$. ////

COROLLARY 5.4.2.3 $\underline{X}_n \xrightarrow{d} \underline{X}$ and $\underline{X}_n - \underline{Y}_n \xrightarrow{P} \underline{0}$ implies $\underline{Y}_n \xrightarrow{d} \underline{X}$. Or, $\underline{X}_n \xrightarrow{d} \underline{X}$ and $\underline{Y}_n \xrightarrow{P} \underline{a}$ implies $\underline{X}_n + \underline{Y}_n \xrightarrow{d} \underline{X} + \underline{a}$. (We are assuming that the dimensions are matched as needed for meaning.) ////

THEOREM 5.4.3 *(Taylor's Theorem Implication) Let* $\{\underline{X}_n\}$ *be a sequence of k-dimensional random vectors. Let* $\{b_n\}$ *be a sequence of positive constants satisfying* $b_n \longrightarrow \infty$ *as* $n \longrightarrow \infty$. *Finally, let* $g(\cdot)$ *be such that k-space* $\xrightarrow{g(\cdot)}$ *r-space. Assume that* $g(\cdot)$ *is differential at* $\underline{a}$, *meaning that all first order partial-derivatives of* $g(\cdot)$ *exist at* $\underline{a}$. *Let* $J_g(\underline{a})$ *denote the* $r \times k$ *matrix of partial derivatives of* $g(\cdot)$ *evaluated at* $\underline{a}$. *Then* $b_n(\underline{X}_n - \underline{a}) \xrightarrow{d} \underline{X}$ *implies* $b_n(g(\underline{X}_n) - g(\underline{a})) \xrightarrow{d} J_g(\underline{a})\underline{X}$. ////

Note that if $g(\cdot)$ is written as

$$g(\underline{x}) = \begin{pmatrix} g_1(x_1,\ldots,x_k) \\ g_2(x_1,\ldots,x_k) \\ \vdots \\ g_r(x_1,\ldots,x_k) \end{pmatrix} \qquad \text{then}$$

$$J_g(\underline{x}) = \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \cdots & \frac{\partial g_1}{\partial x_k} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \cdots & \frac{\partial g_2}{\partial x_k} \\ \vdots & \vdots & & \vdots \\ \frac{\partial g_r}{\partial x_1} & \frac{\partial g_r}{\partial x_2} & \cdots & \frac{\partial g_r}{\partial x_k} \end{pmatrix}, \qquad 5.4.2$$

which is sometimes called the *Jacobian* or *Jacobian matrix*, of the transformation $g(\cdot)$.

COROLLARY 5.4.3 If $\{\underline{X}_n\}$ is a sequence of k-dimensional random vectors and g has Jacobian matrix $J_g(\cdot)$ and if $\sqrt{n}(\underline{X}_n - \underline{a}) \xrightarrow{d} N_{\underline{0},\Sigma}$, then $\sqrt{n}\,(g(\underline{X}_n) - g(\underline{a})) \xrightarrow{d} J_g(\underline{a})N_{\underline{0},\Sigma} = N_{\underline{0},J_g(\underline{a})\Sigma J_g'(\underline{a})}$, where $N_{\underline{0},\underline{\Sigma}}$ is a k-variate normal random vector with mean vector $\underline{0}$ and variance-covariance matrix $\underline{\Sigma}$.

PROOF The first part is just a special case of Theorem 5.4.3. The last equality follows from (e) of Theorem 4.1. ////

We remarked earlier than convergence in probability of random vectors is tantamount to convergence in probability of random variables. This result makes it easy to state a WLLN for higher dimensions.

THEOREM 5.4.4 *(WLLN) Assume* $\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_n, \ldots \overset{\text{indep}}{\sim} (\underline{\mu}, \underline{\Sigma})$ *and define* $\overline{\underline{X}}_n = (1/n)\Sigma_{i=1}^n \underline{X}_i$*; that is, if the* $\underline{X}_n$ *are as in Eq. 5.4.1, then* $\overline{\underline{X}}_n = ((1/n)\Sigma X_{1,i}, (1/n)\Sigma X_{2,i}, \ldots, (1/n)\Sigma X_{k,i})'$*. Then* $\overline{\underline{X}}_n \xrightarrow{P} \underline{\mu}$.

PROOF Have the WLLNs componentwise. ////

Remark As was the case for random variables, this WLLN is valid for models much less stringent than the iid model hypothesized in Theorem 5.4.4. For our purposes, we can get by with this version which is the k-variate analogue of Corollary 5.2.1.2. ////

The following theorem shows that convergence in distribution of random vectors is related to convergence in distribution of random variables.

THEOREM 5.4.5 *(Cramer-Wold device)* $\underline{X}_n \xrightarrow{d} \underline{X}$ *if and only if* $\underline{a}'\underline{X}_n \xrightarrow{d} \underline{a}'\underline{X}$ *for every* $\underline{a} \neq \underline{0}$*. Here* $\underline{X}_n$*,* $\underline{X}$*, and* $\underline{a}$ *are all k-dimensional.*

PROOF Note that $\underline{a}'\underline{X}_n$ and $\underline{a}'\underline{X}$ are random variables. Assume existence of moment generating functions and use Theorem 5.4.1. Assume $\underline{X}_n \xrightarrow{d} \underline{X}$. So, $m_{\underline{a}'\underline{X}_n}(t) = \mathcal{E}[e^{t\Sigma_1^k a_j X_{j,n}}] = m_{\underline{X}_n}(ta_1, \ldots, ta_k) \longrightarrow$ (by assumption) $m_{\underline{X}}(ta_1, \ldots, ta_k) = \mathcal{E}[e^{\Sigma_1^k ta_j X_j}] = \mathcal{E}[e^{t\Sigma_1^k a_j X_j}] = m_{\sum a_j X_j}(t) = m_{\underline{a}'\underline{X}}(t)$ for any $\underline{a} \neq \underline{0}$.

Now assume $\underline{a}'\underline{X}_n \xrightarrow{d} \underline{a}'\underline{X}$ for any $\underline{a} \neq \underline{0}$. Show $m_{\underline{X}_n}(\underline{t}) \longrightarrow m_{\underline{X}}(\underline{t})$ for $\underline{t} = (t_1, \ldots, t_k)'$. We have by assumption $m_{\underline{a}'\underline{X}_n}(t) \longrightarrow m_{\underline{a}'\underline{X}}(t)$ for every $\underline{a} \neq \underline{0}$, so pick $\underline{a}$ so that $a_1 t = t_1, \ldots, a_k t = t_k$ and then $m_{\underline{X}_n}(\underline{t}) = m_{\underline{X}_n}(t_1, \ldots, t_k) = \mathcal{E}[e^{\Sigma(a_j t)X_{j,n}}] = \mathcal{E}[e^{(\Sigma a_j X_{j,n})t}] = m_{\underline{a}'\underline{X}_n}(t) \longrightarrow$ (by assumption) $m_{\underline{a}'\underline{X}}(t) = \mathcal{E}[e^{(\Sigma a_j X_j)t}] = \mathcal{E}[e^{\Sigma(a_j t)X_j}] = m_{\underline{X}}(a_1 t, \ldots, a_k t) = m_{\underline{X}}(\underline{t})$. ////

Comment This reduction of a k-dimensional convergence problem to a uni-dimensional convergence problem is called the *Cramer-Wold device.*

COROLLARY 5.4.5 $\underline{X}_n \xrightarrow{d} \underline{X} \sim \text{MVN}(\underline{\mu}, \underline{\Sigma})$ if and only if $\underline{a}'\underline{X}_n \xrightarrow{d} \underline{a}'\underline{X} \sim N(\underline{a}'\underline{\mu}, \underline{a}'\underline{\Sigma}\,\underline{a})$ for all $\underline{a} \neq \underline{0}$. ////

5.4.2 Central Limit Theorem

THEOREM 5.4.6 *(Central Limit Theorem)* *If* $\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_n, \ldots \overset{iid}{\sim} (\underline{\mu}, \underline{\Sigma})$ *then* $\sqrt{n}(\overline{\underline{X}}_n - \underline{\mu}) \xrightarrow{d} N_{(\underline{0},\underline{\Sigma})}$ *where* $\overline{\underline{X}}_n$ *is as in Theorem 5.4.4.*

PROOF Apply the Cramer-Wold device as in Corollary 5.4.5 in conjunction with the univariate CLT. The details are omitted. ////

EXAMPLE 18 Our repertoire of multivariate random variables is rather limited. The point multinomial was our k-variate extension of the Bernoulli. Let $\underline{X}_1, \ldots, \underline{X}_n, \ldots \overset{iid}{\sim}$ point multinomial $(p_1, \ldots, p_k)$. That is, each $\underline{X}_n$ has the $k+1$ values $(1,0,0,\ldots,0)$, $(0,1,0,\ldots,0)$, ..., $(0,\ldots,0,0,1)$, $(0,0,\ldots,0,0)$ with respective probabilities $p_1, p_2, \ldots, p_k, p_{k+1}$, where $p_i > 0$ for $i = 1, \ldots, k+1$ and $\sum_1^k p_j + p_{k+1} = 1$.

Note
$$\underline{\mu} = \begin{pmatrix} p_1 \\ \vdots \\ p_k \end{pmatrix} = \underline{p} \text{ and } \underline{\Sigma} = \begin{pmatrix} p_1(1-p_1) & -p_1p_2 & \cdots & -p_1p_k \\ -p_1p_2 & p_2(1-p_2) & & \vdots \\ \vdots & & & \vdots \\ -p_1p_k & \cdots & \cdots & p_k(1-p_k) \end{pmatrix}.$$

The WLLN says $\overline{\underline{X}}_n \xrightarrow{P} \underline{p}$ and the CLT says

$$\sqrt{n}\left(\begin{pmatrix} \overline{X}_{1,n} \\ \vdots \\ \overline{X}_{k,n} \end{pmatrix} - \begin{pmatrix} p_1 \\ \vdots \\ p_k \end{pmatrix}\right) \xrightarrow{d} N_{\underline{0},\underline{\Sigma}}$$

where $\overline{X}_{j,n} = (1/n)\sum_{i=1}^n X_{j,i}$, $j = 1, \ldots, k$. We say that $\overline{\underline{X}}_n \overset{\text{asymp}}{\sim} \text{MVN}(\underline{p}, (1/n)\underline{\Sigma})$. What can be said about the asymptotic distribution of $\overline{\underline{X}}_n(\underline{1} - \overline{\underline{X}}_n)$? Define $g(\cdot)$ by

$$g(\underline{z}) = \begin{pmatrix} z_1(1-z_1) \\ \vdots \\ z_k(1-z_k) \end{pmatrix} \quad \text{so that} \quad J_g(\underline{z}) = \begin{pmatrix} 1-2z_1 & 0 & \cdots & 0 \\ 0 & 1-2z_2 & & \vdots \\ \vdots & & & 0 \\ 0 & \cdots & \cdots & 1-2z_k \end{pmatrix}$$

Now, via Corollary 5.4.3,

$$\sqrt{n}\left(\begin{pmatrix}\overline{X}_{1,n}(1-\overline{X}_{1,n})\\ \vdots \\ \overline{X}_{k,n}(1-\overline{X}_{k,n})\end{pmatrix}-\begin{pmatrix}p_1(1-p_1)\\ \vdots \\ p_k(1-p_k)\end{pmatrix}\right)\xrightarrow{d} N_{\underline{0},J_g(\underline{a})\underline{\Sigma}J_g'(\underline{a})}$$

that is

$$\begin{pmatrix}\overline{X}_{1,n}(1-\overline{X}_{1,n})\\ \vdots \\ \overline{X}_{k,n}(1-\overline{X}_{k,n})\end{pmatrix}\overset{\text{asymp}}{\sim}$$

$$\text{MVN}\left(\begin{pmatrix}p_1(1-p_1)\\ \vdots \\ p_k(1-p_k)\end{pmatrix};\frac{1}{n}\begin{pmatrix}(1-2p_1)^2p_1(1-p_1) & \cdots & -(1-2p_1)(1-2p_k)p_1p_k\\ \vdots & & \vdots \\ -(1-2p_1)(1-2p_k)p_1p_k & \cdots & (1-2p_k)^2p_k(1-p_k)\end{pmatrix}\right) \qquad ////$$

EXAMPLE 19 Assume $X_1,\ldots,X_n,\ldots,\overset{iid}{\sim}F(\cdot)$, and let $F_n(\cdot)$ be the sample cumulative distribution function of $X_1,\ldots,X_n$. Recalling that $F_n(x)=(1/n)\Sigma_1^n I_{(-\infty,x]}(X_i)$, the CLT gives

$$\sqrt{n}\left(\begin{pmatrix}F_n(x_1)\\ \vdots \\ F_n(x_k)\end{pmatrix}-\begin{pmatrix}F(x_1)\\ \vdots \\ F(x_k)\end{pmatrix}\right)\xrightarrow{d} N_{\underline{0},\underline{\Sigma}}$$

where

$$\underline{\Sigma}=\begin{pmatrix}\text{var}[I_{(-\infty,x_1]}(X_1)] & \cdots\cdots & \text{cov}[I_{(-\infty,x_1]}(X_1),I_{(-\infty,x_k]}(X_1)]\\ \vdots & & \vdots \\ \vdots & & \vdots \\ \text{cov}[I_{(-\infty,x_1]}(X_1),I_{(-\infty,x_k]}(X_1)] & \cdots\cdots & \text{var}[I_{(-\infty,x_k]}(X_1)]\end{pmatrix}$$

and $\quad \text{var}[I_{(-\infty,x_j]}(X_1)]=F(x_j)(1-F(x_j))$

and $\quad \text{cov}[I_{(-\infty,x_i]}(X_1),I_{(-\infty,x_j]}(X_1)]=F(\min[x_i,x_j])-F(x_i)F(x_j).$ $\qquad ////$

EXAMPLE 20 Assume $\begin{pmatrix} X_1 \\ Y_1 \end{pmatrix}, \ldots, \begin{pmatrix} X_n \\ Y_n \end{pmatrix}, \ldots \overset{iid}{\sim} \left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} ; \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \right)$. Find the asymptotic distribution of $\overline{X}_n/\overline{Y}_n$ for $\mu_2 \neq 0$. By the bivariate CLT we have:

$$\sqrt{n}\left(\begin{pmatrix} \overline{X}_n \\ \overline{Y}_n \end{pmatrix} - \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \right) \xrightarrow{d} N_{\underline{0},\underline{\Sigma}},$$

where $\underline{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$. Take $g(x,y) = x/y$ in Corollary 5.4.3. Note that $k = 2$ and $r = 1$. $J_g = (1/y, -x/y^2)$ and $\underline{a} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$ so $J_g(\underline{a}) = (1/\mu_2, -\mu_1/\mu_2^2)$. Corollary 5.4.3 says $\sqrt{n}((\overline{X}_n/\overline{Y}_n) - (\mu_1/\mu_2)) \xrightarrow{d} N_{0,J_g(\underline{a})\underline{\Sigma}J_g'(\underline{a})}$ but

$$\begin{aligned} J_g(\underline{a})\underline{\Sigma}J_g'(\underline{a}) &= (1/\mu_2, -\mu_1/\mu_2^2)\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\begin{pmatrix} 1/\mu_2 \\ -\mu_1/\mu_2^2 \end{pmatrix} \\ &= (\sigma_1^2/\mu_2^2) + 2\rho\sigma_1\sigma_2(1/\mu_2)(-\mu_1/\mu_2^2) + \sigma_2^2(-\mu_1/\mu_2^2)^2 \\ &= (1/\mu_2^2)[\sigma_1^2 - 2(\mu_1/\mu_2)\rho\sigma_1\sigma_2 + (\mu_1/\mu_2)^2\sigma_2^2]. \end{aligned}$$

So

$$\begin{aligned} \overline{X}_n/\overline{Y}_n &\overset{asymp}{\sim} N(\mu_1/\mu_2, (1/n\mu_2^2)[\sigma_1^2 - 2(\mu_1/\mu_2)\rho\sigma_1\sigma_2 + (\mu_1/\mu_2)^2\sigma_2^2]) \\ &= N\left(\frac{\mu_1}{\mu_2}; \frac{1}{n}\left(\frac{\mu_1}{\mu_2}\right)^2 \left[\frac{\sigma_1^2}{\mu_1^2} - \frac{2\rho\sigma_1\sigma_2}{\mu_1\mu_2} + \frac{\sigma_2^2}{\mu_2^2}\right]\right). \end{aligned}$$

////

5.4.3 Joint Asymptotic Distribution of Sample Mean and Variance

Assume $X_1, X_2, \ldots, X_n, \ldots \overset{iid}{\sim} (\mu, \sigma^2, \mu_3, \mu_4)$. We seek the joint asymptotic distribution of $\overline{X}_n$ and S_n^2. We might anticipate the answer to be a bivariate normal inasmuch as we already know that the two marginal distributions are normals and we know the covariance of $\overline{X}_n$ and S_n^2. To confirm, consider

$$\begin{pmatrix} X_1 - \mu \\ (X_1 - \mu)^2 \end{pmatrix}, \begin{pmatrix} X_2 - \mu \\ (X_2 - \mu)^2 \end{pmatrix}, \ldots, \begin{pmatrix} X_n - \mu \\ (X_n - \mu)^2 \end{pmatrix}, \ldots \overset{iid}{\sim} \left(\begin{matrix} 0 \\ \sigma^2 \end{matrix} ; \underline{\Sigma} \right)$$

where

$$\underline{\Sigma} = \begin{pmatrix} \mathrm{var}[X_1 - \mu] & \mathrm{cov}[(X_1 - \mu), (X_1 - \mu)^2] \\ \mathrm{cov}[(X_1 - \mu), (X_1 - \mu)^2] & \mathrm{var}[(X_1 - \mu)^2] \end{pmatrix}.$$

The bivariate CLT gives

$$\sqrt{n}\left(\begin{pmatrix}(1/n)\Sigma_1^n(X_i-\mu)\\(1/n)\Sigma_1^n(X_i-\mu)^2\end{pmatrix}-\begin{pmatrix}0\\\sigma^2\end{pmatrix}\right)\xrightarrow{d}N_{\underline{0},\underline{\Sigma}}$$

We saw in the proof of Theorem 5.3.2 that $\sqrt{n}[S_n^2-(1/n)\Sigma_1^n(X_i-\mu)^2]\xrightarrow{P}0$. Hence

$$\sqrt{n}\left(\begin{pmatrix}\overline{X}_n\\S_n^2\end{pmatrix}-\begin{pmatrix}\mu\\\sigma^2\end{pmatrix}\right)=$$

$$\sqrt{n}\left(\begin{pmatrix}(1/n)\Sigma_1^n(X_i-\mu)\\(1/n)\Sigma_1^n(X_i-\mu)^2\end{pmatrix}-\begin{pmatrix}0\\\sigma^2\end{pmatrix}\right)+\sqrt{n}\begin{pmatrix}0\\S_n^2-(1/n)\Sigma_1^n(X_i-\mu)^2\end{pmatrix}.$$

But $\sqrt{n}\begin{pmatrix}0\\S_n^2-(1/n)\Sigma_1^n(X_i-\mu)^2\end{pmatrix}\xrightarrow{P}\begin{pmatrix}0\\0\end{pmatrix}$

and $\sqrt{n}\left(\begin{pmatrix}(1/n)\Sigma_1^n(X_i-\mu)\\(1/n)\Sigma_1^n(X_i-\mu)^2\end{pmatrix}-\begin{pmatrix}0\\\sigma^2\end{pmatrix}\right)\xrightarrow{d}N_{\underline{0},\underline{\Sigma}}$,

and hence by Slutzky's Theorem

$$\sqrt{n}\left(\begin{pmatrix}\overline{X}_n\\S_n^2\end{pmatrix}-\begin{pmatrix}\mu\\\sigma^2\end{pmatrix}\right)\xrightarrow{d}N_{\underline{0},\underline{\Sigma}};\quad\text{that is}$$

$$\begin{pmatrix}\overline{X}_n\\S_n^2\end{pmatrix}\overset{\text{asymp}}{\sim}\text{BVN}\left(\begin{matrix}\mu\\\sigma^2\end{matrix};\frac{1}{n}\begin{pmatrix}\sigma^2&\mu_3\\\mu_3&\mu_4-\sigma^4\end{pmatrix}\right)$$

since $\text{var}[X_1-\mu]=\sigma^2$, $\text{cov}[(X_1-\mu),(X_1-\mu)^2]=\mathcal{E}[(X_1-\mu)^3]=\mu_3$ and $\text{var}[(X_1-\mu)^2]=\mu_4-\sigma^4$. Note that the covariance of the asymptotic distribution is μ_3/n, which is the exact covariance of $\overline{X}_n$ and S_n^2. Note also that $\overline{X}_n$ and S_n^2 are *asymptotically independent* if $\mu_3=0$.

We can use the asymptotic distribution of $\overline{X}_n$ and S_n^2 and Corollary 5.4.3 to derive an asymptotic distribution for functions of $\overline{X}_n$ and S_n^2. For example, if we sought the asymptotic distribution of the sample mean and sample standard deviation, we take $g(\underline{z})=\begin{pmatrix}g_1(\underline{z})\\g_2(\underline{z})\end{pmatrix}=\begin{pmatrix}z_1\\\sqrt{z_2}\end{pmatrix}$, implying $J_g(\underline{z})=\begin{pmatrix}1&0\\0&1/2\sqrt{z_2}\end{pmatrix}$ so that $J_g(\underline{a})=\begin{pmatrix}1&0\\0&1/2\sigma\end{pmatrix}$. Then Corollary 5.4.3 says $\sqrt{n}\left(\begin{pmatrix}\overline{X}_n\\S_n\end{pmatrix}-\begin{pmatrix}\mu\\\sigma\end{pmatrix}\right)\xrightarrow{d}N_{\underline{0},J_g(\underline{a})\underline{\Sigma}J_g'(\underline{a})}$. But $J_g(\underline{a})\underline{\Sigma}J_g'(\underline{a})=\begin{pmatrix}1&0\\0&1/2\sigma\end{pmatrix}\begin{pmatrix}\sigma^2&\mu_3\\\mu_3&\mu_4-\sigma^4\end{pmatrix}\begin{pmatrix}1&0\\0&1/2\sigma\end{pmatrix}=\begin{pmatrix}\sigma^2&\mu_3/2\sigma\\\mu_3/2\sigma&(\mu_4-\sigma^4)/4\sigma^2\end{pmatrix}$, so

$$\begin{pmatrix} \overline{X}_n \\ S_n \end{pmatrix} \overset{\text{asymp}}{\sim} \text{BVN} \left(\begin{matrix} \mu \\ \sigma \end{matrix} ; \frac{1}{n} \begin{pmatrix} \sigma^2 & \mu_3/2\sigma \\ \mu_3/2\sigma & (\mu_4 - \sigma^4)/4\sigma^2 \end{pmatrix} \right).$$

The asymptotic distributions of other functions of sample moments are given in the Problems, including the asymptotic joint distribution of the sample coefficient of variation and sample skewness coefficient alluded to in the introduction of this Section 5.4.

5.5 Asymptotics of Sample Quantiles

We restrict to the case when $X_1, \ldots, X_n, \ldots, \overset{iid}{\sim} F(\cdot)$ and the X_i's are random variables. Recall that *sample quantiles* are *order statistics*, denoted $X_{j:n}$, where $X_{j:n} = j$th smallest of $X_1, \ldots, X_n$. We consider two types of limiting results of sample quantiles, depending on the behavior of j/n as n approaches infinity. The first case considers $j/n \longrightarrow q$, where $0 < q < 1$, as $n \longrightarrow \infty$; this says that j and n approach infinity together in such a way that the ratio j/n approaches q. In essence, the *relative* position of the sample quantile $X_{j:n}$ remains fixed as the sample size n goes to infinity. For example, if $q = 1/2$, then the sample quantile of interest is the sample median as n approaches infinity.

For the other case, the ratio j/n approaches either zero or one. In this case, one can think that the *absolute* position remains fixed as $n \longrightarrow \infty$. For instance, one can think of the smallest order statistic, in which case $j/n = 1/n \longrightarrow 0$; or think of the largest order statistic, in which case $j/n = n/n \longrightarrow 1$. One could also think of the kth smallest or kth largest as $n \longrightarrow \infty$ for fixed k. These latter type order statistics are *extremes* as $n \longrightarrow \infty$ and we delay our introduction to the *theory of extremes* until the next section. Here we consider the asymptotic distribution of those sample quantiles whose relative position is fixed as $n \longrightarrow \infty$.

THEOREM 5.5.1. Consider a sequence $\{q_n\}$ where $0 < q_n < 1$, satisfying $q_n \longrightarrow q$ as $n \longrightarrow \infty$ such that nq_n is an integer and $n|q_n - q|$ is bounded as $n \longrightarrow \infty$. If $X_1, \ldots, X_n, \ldots \overset{iid}{\sim}$ with probability density $f(\cdot)$ and cumulative distribution function $F(\cdot)$, where $F(\cdot)$ is strictly monotone for $0 < F(x) < 1$, then

$$X_{nq_n:n} \overset{\text{asymp}}{\sim} N(\xi_q, q(1-q)/nf^2(\xi_q)), \qquad 5.5.1$$

where ξ_q is the qth population quantile.

PROOF Omitted ////

Remark The asymptotic distribution of Eq. 5.5.1 really comes from the limiting result that says

$$\sqrt{n}(X_{nq_n:n} - \xi_q) \xrightarrow{d} N_{0,q(1-q)/f^2(\xi_q)}.$$

Remark $\sqrt{n}(X_{nq_n:n} - \xi_q) \xrightarrow{d} N$ implies (by Slutzky) $X_{nq_n:n} \xrightarrow{P} \xi_q$ which supports the use of a sample quantile to estimate a population quantile.

PROOF $1/\sqrt{n} \xrightarrow{P} 0$ and $\sqrt{n}(X_{nq_n:n} - \xi_q) \xrightarrow{d} N$ implies $(1/\sqrt{n}) \cdot \sqrt{n}(X_{nq_n:n} - \xi_q) \xrightarrow{d} 0 \cdot N = 0$. ////

EXAMPLE 21 Assume $X_1, \ldots, X_n, \ldots \overset{iid}{\sim} N(\mu, \sigma^2)$. We know the sample mean $\sim N(\mu, \sigma^2/n)$. Theorem 5.5.1 says the sample median $\overset{\text{asymp}}{\sim} N(\mu, \pi\sigma^2/2n)$, since for $q = 1/2$, $\xi_q = \mu$, and $f(\mu) = 1/\sqrt{2\pi}\sigma$.

Hence, the sample mean has variance smaller than the asymptotic variance of the sample median. ////

EXAMPLE 22 Assume $X_1, \ldots, X_n, \ldots \overset{iid}{\sim}$ Cauchy (α, β), so that $f(x) = 1/\beta\pi\,\{1 + [(x-\alpha)/\beta]^2\}$. For $q = 1/2$, $\xi_{1/2} = \alpha$, so $f(\alpha) = 1/\beta\pi$; hence the sample median $\overset{\text{asymp}}{\sim} N(\alpha, \beta^2\pi^2/4n)$. Recall $\overline{X}_n \sim$ Cauchy (α, β). (See Example 13.) Here the sample median has a much smaller asymptotic variance than the variance, which is infinite, of the sample mean! ////

There is a higher dimensional version of Theorem 5.5.1; we state it in the bivariate case, with the belief that the higher dimensional version ought then be evident.

THEOREM 5.5.2 *For $0 < q' < q'' < 1$ and sequences $\{q'_n\}$ and $\{q''_n\}$ satisfying $q'_n < q''_n$, both nq'_n and nq''_n integers, and $q'_n \longrightarrow q'$ and $q''_n \longrightarrow q''$ so that $n|q'_n - q'|$ and $n|q''_n - q''|$ are bounded; if $X_1, \ldots, X_n, \ldots \overset{iid}{\sim} f(\cdot)$ and $F(\cdot)$ with $F(\cdot)$ strictly monotone for $0 < F(x) < 1$, then*

$$\sqrt{n}\begin{pmatrix} X_{nq'_n:n} - \xi_{q'} \\ X_{nq''_n:n} - \xi_{q''} \end{pmatrix} \xrightarrow{d} N_{\underline{0},\underline{\Sigma}}, \qquad \text{where}$$

$$\underline{\Sigma} = \begin{pmatrix} \frac{q'(1-q')}{f^2(\xi_{q'})} & \frac{q'(1-q'')}{f(\xi_{q'})f(\xi_{q''})} \\ \frac{q'(1-q'')}{f(\xi_{q'})f(\xi_{q''})} & \frac{q''(1-q'')}{f^2(\xi_{q''})} \end{pmatrix}.$$

PROOF Omitted. ////

Remark The correlation coefficient of the asymptotic (bivariate) distribution of $X_{nq'_n:n}$ and $X_{nq''_n:n}$, written $\text{acorr}(X_{nq'_n:n}, X_{nq''_n:n})$, equals $\sqrt{\frac{q'(1-q'')}{(1-q')q''}}$, which is distribution free. For example, $q' = 1/4$ and $q'' = 3/4$ gives an asymptotic correlation of $1/3$; that is, acorr(1st sample quartile, 3rd sample quartile) $= 1/3$. ////