Section 5.4. Extensions to Multivariate Distributions

Note. In this section, we consider convergence in probability and distribution of sequences of random vectors. The ideas are briefly discussed and some results are stated without proof. Recall that random vectors were introduced in our Section 2.6. Extension to Several Random Variables.

Note. Recall that a *norm* $\|\cdot\|$ on $\mathbb{R}^p$ is a mapping $\|\cdot\| : \mathbb{R}^p \to \mathbb{R}$ such that

(a) For all $\mathbf{v} \in \mathbb{R}^p$ we have $\|\mathbf{v}\| \geq 0$, and $\|\mathbf{v}\| = 0$ if and only if $\mathbf{v} = \mathbf{0}$.

(b) For all $\mathbf{v} \in \mathbb{R}^p$ and $a \in \mathbb{R}$ we have $\|a\mathbf{v}\| = |a|\|\mathbf{v}\|$.

(c) For all $\mathbf{v}, \mathbf{u} \in \mathbb{R}^p$, we have $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$ (the Triangle Inequality).

These ideas are covered in Linear Algebra (MATH 2010); see my online Linear Algebra notes on Section 1.2. The Norm and Dot Product (notice "Theorem 1.2. Properties of the Norm in $\mathbb{R}n$"). The Euclidean norm on $\mathbb{R}^n$ is given by

$$\|\mathbf{v}\| = \sqrt{\sum_{i=1}^{p} v_i^2},$$

where $\mathbf{v} \in \mathbb{R}^p$ is of the form $\mathbf{v}' = (v_1, v_2, \ldots, v_p)$. Here, we treat vectors in $\mathbb{R}^p$ as column vectors and we represent the transpose of $\mathbf{v}$ as $\mathbf{v}'$ (a row vector); similarly, the transpose of matrix $\mathbf{A}$ is $\mathbf{A}'$.

Note. Recall that the standard basis vectors for $\mathbb{R}^p$ are the vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_p$ where for $\mathbf{e}_i$ all components are 0 except the ith component which is 1. So for $\mathbf{v}' = (v_1, v_2, \ldots, v_p)$, we have $\mathbf{v} = \sum_{i=1}^{p} v_i \mathbf{e}_i$.

Lemma 5.4.1. Let $\mathbf{v}' = (v_1, v_2, \ldots, v_p) \in \mathbb{R}^p$. Then $|v_j| \le \|\mathbf{v}\| \le \sum_{i=1}^p |v_i|$ for all $j = 1, 2, \ldots, p$.

Definition 5.4.1. Let $\{\mathbf{X}_n\}$ be a sequence of p-dimensional vectors and let $\mathbf{X}$ be a random vector, all defined on the same sample space. We say that $\{\mathbf{X}_n\}$ *converges in probability* to $\mathbf{X}$ if $\lim_{n\to\infty} P(\|\mathbf{X}_n - \mathbf{X}\| \ge \varepsilon) = 0$, for all $\varepsilon > 0$. As in the univariate case, we write $\mathbf{X}_n \xrightarrow{P} \mathbf{X}$.

Note. The next result shows that convergence in probability of vectors is equivalent to componentwise convergence in probability (though this is only try for finite dimensional vectors).

Theorem 5.4.1. Let $\{\mathbf{X}_n\}$ be a sequence of p-dimensional vectors and let $\mathbf{X}$ be a random vector, all defined on the same sample space. Then

$$\mathbf{X}_n \xrightarrow{P} \mathbf{X} \text{ if and only if } X_{nj} \xrightarrow{P} X_j \text{ for all } j = 1, 2, \ldots, p.$$

Note. Recall that a statistic T_n based on a sample $X_1, X_2, \ldots, X_n$ from the distribution of random variable X is a *consistent estimator* of parameter θ is $T_n \xrightarrow{P} \theta$. In Section 5.1 we say that is $X_1, X_2, \ldots, X_n$ is a random sample from the distribution of random variable X with mean μ and variance σ^2, then $\overline{X}_n$ and S_n^2 are consistent estimators of μ and σ^2 (see Note 5.1.A and Theorem 5.1.B). So by Theorem 5.4.1, the vector $(\overline{X}_n, S_n^2)$ is a consistent estimator of (μ, σ^2).

Note. Similar to the previous, but in the setting of random vectors, suppose $\{\mathbf{X}_n\}$ is a sequence of identical in distribution random vectors with common mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$. Let $\overline{\mathbf{X}}_n$ be defined as $\overline{\mathbf{X}}_n = \dfrac{1}{n}\sum_{i=1}^n \mathbf{X}_i$. Then $\overline{\mathbf{X}}_n$ is the vector of sample means $\overline{\mathbf{X}}_n = (\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_p)'$. By the Weak Law of Large Numbers (Theorem 5.1.1), $\overline{X}_j \xrightarrow{P} \mu_j$ for each $j = 1, 2, \ldots, p$. So by Theorem 5.4.1, $\overline{\mathbf{X}}_n \xrightarrow{P} \boldsymbol{\mu}$. With $\mathbf{X}_i = (X_{i1}, X_{i2}, \ldots, X_{ip})'$, define the sample variances and covariances

$$S_{n,j}^2 = \frac{1}{n-1}\sum_{i=1}^n (X_{ij} - \overline{X}_j)^2 \text{ for } j = 1, 2, \ldots, p$$

$$S_{n,jk} = \frac{1}{n-1}\sum_{i=1}^n (X_{ij} - \overline{X}_j)(X_{ik} - \overline{X}_k) \text{ for } j, k = 1, 2, \ldots, p \text{ where } j \neq k.$$

Assuming $E(X_i^4)$ is finite for $i = 1, 2, \ldots, p$, then by Theorem 5.1.B (which follows from the Weak Law of Large Numbers), we have $S_{n,j}^2 \xrightarrow{P} \sigma_j^2$ for $j = 1, 2, \ldots, p$ and $S_{n,jk}^2 \xrightarrow{P} \mathrm{Cov}(X_j, X_k)$ for $j, k = 1, 2, \ldots, p$ where $j \neq k$. If we define the $p \times p$ matrix $\mathbf{S}$ to be the jth diagonal entry $S_{n,j}^2$ and the (j,k)th entry $S_{n,jk}^2$, then $\mathbf{S} \xrightarrow{P} \boldsymbol{\Sigma}$ by Theorem 5.4.1.

Definition 5.4.2. Let $(\mathbf{X}_n)$ be a sequence of random vectors with $\mathbf{X}_n$ having distribution function $F_n(\mathbf{x})$ and $\mathbf{X}$ be a random vector with distribution function $F(\mathbf{x})$. Then $(\mathbf{X}_n)$ *converges in distribution* to $\mathbf{X}$ if $\lim_{n\to\infty} F_n(\mathbf{x}) = F(\mathbf{x})$ for all points $\mathbf{x}$ at which $F(\mathbf{x})$ is continuous. This is denoted $\mathbf{X}_n \xrightarrow{D} \mathbf{X}$.

Note. Since the definition of convergence in distribution here is virtually identical to the case of a single variable ("univariate"), then several results from that setting (given in Section 5.2. Convergence in Distribution) carry over to this setting. So we now state two theorems, but omit the proofs (the corresponding results from Section 5.2 that have proofs "beyond the scope of this book"). The first is a generalization of Theorem 5.2.10 (and includes a converse of Theorem 5.2.10).

Theorem 5.4.2. Let $(\mathbf{X}_n)$ be a sequence of random vectors that converges in distribution to a random vector $\mathbf{X}$ and let $g(\mathbf{X})$ be a function that is continuous on the support of $\mathbf{X}$. Then $g(\mathbf{X}_n)$ converges in distribution to $g(\mathbf{X})$.

Note. Theorem 5.4.2 can by used to prove that convergence in distribution implies "marginal convergence" (that is, convergence in distribution of each of the marginal distributions).

Theorem 5.4.3. Let $(\mathbf{x}_n)$ be a sequence of random vectors with $\mathbf{X}_n$ having distribution function $F_n(\mathbf{x})$ and moment generating function $M_n(\mathbf{t})$. Let $\mathbf{X}$ be a random vector with distribution function $f(\mathbf{x})$ and moment generating function $M(\mathbf{t})$. Then $(\mathbf{X}_n)$ converges in distribution to $\mathbf{X}$ if and only if, for some $h > 0$, $\lim_{n\to\infty} M_n(\mathbf{t}) = M(\mathbf{t})$, for all $\mathbf{t}$ such that $\|\mathbf{t}\| < h$.

Note. We now state and prove the main result of this section. It is really just a corollary of the Central Limit Theorem (Theorem 5.3.1).

Theorem 5.4.4. Multivariate Central Limit Theorem.

Let $(\mathbf{X}_n)$ be a sequence of independent and identically distributed ("iid") random vectors with common mean vector μ and variance-covariance matrix $\mathbf{\Sigma}$ which is positive definite. Assume that the common moment generating function $M(\mathbf{t})$ exists in an open neighborhood of $\mathbf{0}$. Let

$$\mathbf{Y}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\mathbf{X}_i - \mu) = \sqrt{n}(\sqrt{\mathbf{X}} - \mu).$$

Then $\mathbf{Y}_n$ converges in distribution to a $N_p(\mathbf{0}, \mathbf{\Sigma})$ distribution.

Note. We conclude by stating, without proof, results concerning convergence in distribution. The first concerns a linear transformation of a random variable. The second is a generalization of the Δ-method (see Theorem 5.2.9)

Theorem 5.4.5. Let $(\mathbf{X}_n)$ be a sequence of p-dimensional random vectors. Suppose $\mathbf{X}_m \xrightarrow{D} N(\boldsymbol{\mu}, \mathbf{\Sigma})$. Let $\mathbf{A}$ be an $m \times p$ matrix of constants and let $\mathbf{b}$ be an m-dimensional vector of constants. Then $\mathbf{A}\mathbf{X}_n + \mathbf{b} \xrightarrow{D} N(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\mathbf{\Sigma}\mathbf{A}')$.

Theorem 5.4.6. Let $(\mathbf{X}_n)$ be a sequence of p-dimensional random vectors. Suppose $\sqrt{n}(\mathbf{X}_n - \boldsymbol{\mu}_0) \xrightarrow{D} N_p(\mathbf{0}, \mathbf{\Sigma})$. Let $\mathbf{g}$ be a transformation

$$\mathbf{g}(\mathbf{x}) = (g_1(\mathbf{x}), g_2(\mathbf{x}), \ldots, g_k(\mathbf{x}))'$$

such that $1 \leq k \leq p$ and the $k \times p$ matrix of partial derivatives,

$$\mathbf{B} = \left[\frac{\partial g_i}{\partial \mu_j}\right] \text{ for } i = 1, 2, \ldots, k,\ j = 1, 2, \ldots, p$$

are continuous and do not vanish in a neighborhood of μ_0. Let $\mathbf{B}_0 = \mathbf{B}$ at $\boldsymbol{\mu}_0$. Then

$$\sqrt{n}(\mathbf{g}(\mathbf{X}_n) - \mathbf{g}(\boldsymbol{\mu}_0)) \xrightarrow{D} N_k(\mathbf{0}, \mathbf{B}_0\mathbf{\Sigma}\mathbf{B}_0').$$

Revised: 7/21/2021