Convergence of Sequences of Random Variables[1]

1. Definitions (All quantities in boldface are vectors in $\mathbb{R}^m$ unless otherwise stated)

   - $\star$ $\mathbf{X}_n \stackrel{a.s.}{\rightarrow} \mathbf{X}$ means $P\{s : \lim_{n\rightarrow\infty} \mathbf{X}_n(s) = \mathbf{X}(s)\} = 1$.
   - $\star$ $\mathbf{X}_n \stackrel{P}{\rightarrow} \mathbf{X}$ means $\forall\epsilon > 0$, $\lim_{n\rightarrow\infty} P\{|\mathbf{X}_n - \mathbf{X}| < \epsilon\} = 1$.
   - $\star$ $\mathbf{X}_n \stackrel{d}{\rightarrow} \mathbf{X}$ means for every continuity point $\mathbf{x}$ of $F_{\mathbf{X}}$, $\lim_{n\rightarrow\infty} F_{\mathbf{X}_n}(\mathbf{x}) = F_{\mathbf{X}}(\mathbf{x})$.

2. $\mathbf{X}_n \stackrel{a.s.}{\rightarrow} \mathbf{X} \Rightarrow \mathbf{X}_n \stackrel{P}{\rightarrow} \mathbf{X} \Rightarrow \mathbf{X}_n \stackrel{d}{\rightarrow} \mathbf{X}$.

3. If $\mathbf{a}$ is a vector of constants, $\mathbf{X}_n \stackrel{d}{\rightarrow} \mathbf{a} \Rightarrow \mathbf{X}_n \stackrel{P}{\rightarrow} \mathbf{a}$.

4. Strong Law of Large Numbers (SLLN): Let $\mathbf{X}_1, \ldots \mathbf{X}_n$ be i.i.d. random vectors with finite first moment. Then $\overline{\mathbf{X}}_n \stackrel{a.s.}{\rightarrow} E(\mathbf{X}_1)$.

5. Central Limit Theorem: Let $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be i.i.d. random vectors with expected value vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Then $\sqrt{n}(\overline{\mathbf{X}}_n - \boldsymbol{\mu})$ converges in distribution to a multivariate normal with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$.

6. Slutsky Theorems for Convergence in Distribution:

   (a) If $\mathbf{X}_n \in \mathbb{R}^m$, $\mathbf{X}_n \stackrel{d}{\rightarrow} \mathbf{X}$ and if $f : \mathbb{R}^m \rightarrow \mathbb{R}^q$ (where $q \leq m$) is continuous except possibly on a set C with $P(\mathbf{X} \in C) = 0$, then $f(\mathbf{X}_n) \stackrel{d}{\rightarrow} f(\mathbf{X})$.

   (b) If $\mathbf{X}_n \stackrel{d}{\rightarrow} \mathbf{X}$ and $(\mathbf{X}_n - \mathbf{Y}_n) \stackrel{P}{\rightarrow} 0$, then $\mathbf{Y}_n \stackrel{d}{\rightarrow} \mathbf{X}$.

   (c) If $\mathbf{X}_n \in \mathbb{R}^d$, $\mathbf{Y}_n \in \mathbb{R}^k$, $\mathbf{X}_n \stackrel{d}{\rightarrow} \mathbf{X}$ and $\mathbf{Y}_n \stackrel{d}{\rightarrow} \mathbf{c}$, then

   $$\begin{pmatrix} \mathbf{X}_n \\ \mathbf{Y}_n \end{pmatrix} \stackrel{d}{\rightarrow} \begin{pmatrix} \mathbf{X} \\ \mathbf{c} \end{pmatrix}$$

7. Slutsky Theorems for Convergence in Probability:

   (a) If $\mathbf{X}_n \in \mathbb{R}^m$, $\mathbf{X}_n \stackrel{P}{\rightarrow} \mathbf{X}$ and if $f : \mathbb{R}^m \rightarrow \mathbb{R}^q$ (where $q \leq m$) is continuous except possibly on a set C with $P(\mathbf{X} \in C) = 0$, then $f(\mathbf{X}_n) \stackrel{P}{\rightarrow} f(\mathbf{X})$.

   (b) If $\mathbf{X}_n \stackrel{P}{\rightarrow} \mathbf{X}$ and $(\mathbf{X}_n - \mathbf{Y}_n) \stackrel{P}{\rightarrow} 0$, then $\mathbf{Y}_n \stackrel{P}{\rightarrow} \mathbf{X}$.

   (c) If $\mathbf{X}_n \in \mathbb{R}^d$, $\mathbf{Y}_n \in \mathbb{R}^k$, $\mathbf{X}_n \stackrel{P}{\rightarrow} \mathbf{X}$ and $\mathbf{Y}_n \stackrel{P}{\rightarrow} \mathbf{Y}$, then

   $$\begin{pmatrix} \mathbf{X}_n \\ \mathbf{Y}_n \end{pmatrix} \stackrel{P}{\rightarrow} \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix}$$

8. Delta Method (Theorem of Cramér, Ferguson p. 45): Let $g : \mathbb{R}^d \rightarrow \mathbb{R}^k$ be such that the elements of $\dot{\mathrm{g}}(\mathbf{x}) = \left[\frac{\partial g_i}{\partial x_j}\right]_{k\times d}$ are continuous in a neighborhood of $\boldsymbol{\theta} \in \mathbb{R}^d$. If $\mathbf{X}_n$ is a sequence of d-dimensional random vectors such that $\sqrt{n}(\mathbf{X}_n - \boldsymbol{\theta}) \stackrel{d}{\rightarrow} \mathbf{X}$, then $\sqrt{n}(g(\mathbf{X}_n) - g(\boldsymbol{\theta})) \stackrel{d}{\rightarrow} \dot{\mathrm{g}}(\boldsymbol{\theta})\mathbf{X}$. In particular, if $\sqrt{n}(\mathbf{X}_n - \boldsymbol{\theta}) \stackrel{d}{\rightarrow} \mathbf{X} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, then $\sqrt{n}(g(\mathbf{X}_n) - g(\boldsymbol{\theta})) \stackrel{d}{\rightarrow} \mathbf{Y} \sim N(\mathbf{0}, \dot{\mathrm{g}}(\boldsymbol{\theta})\boldsymbol{\Sigma}\dot{\mathrm{g}}(\boldsymbol{\theta})')$.

---

[1]This material can be found in many texts. I took it from T. S. Ferguson's *A course in large sample theory:* Chapman and Hall, 1996