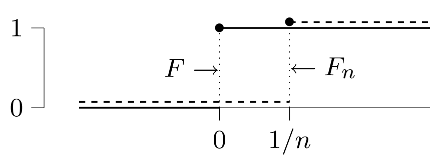
TOPIC. Convergence in distribution and related notions. This section studies the notion of the so-called convergence in distribution of real random variables. This is the kind of convergence that takes place in the central limit theorem, which will be developed in a later section. The basic idea is that the distributions of the random variables involved "settle down." The following definition focuses on the probability contents of intervals.

Definition 1. *Let $X_1, X_2, \dots$ and X be real random variables with respective distribution functions $F_1, F_2, \dots$ and F. One says* ***X_n converges to X in distribution as $n \to \infty$****, and writes* $\boldsymbol{X_n \to_{\mathcal{D}} X}$*, if*

$$\lim_{n\to\infty} F_n(x) = F(x) \quad \text{for all } x \in C_F. \tag{1}$$

Here $C_F := \{x \in \mathbb{R} : F \text{ is continuous at x}\} = \{x \in \mathbb{R} : P[X = x] = 0\}$ *is the set of continuity points of F.*

Example 1. Suppose that the distribution of X_n is a unit mass at the point $1/n$, and the distribution of X is a unit mass at 0. If there is any justice to the definition, we should have $X_n \to_{\mathcal{D}} X$. The distribution functions F_n and F are graphed below; the graph of F_n has been shifted up slightly to distinguish it from the graph of F.



Note that $C_F = \{x \in \mathbb{R} : x \neq 0\}$. It is easy to see that

$$\lim_{n\to\infty} F_n(x) = \begin{cases} 0, & \text{if } x < 0, \\ 1, & \text{if } x > 0, \end{cases} = F(x)$$

for all $x \in C_F$, so we do indeed have $X_n \to_{\mathcal{D}} X$. Note, however, that

$$\lim_n F_n(0) = 0 \neq 1 = F(0);$$

This illustrates why (1) doesn't require $\lim_n F_n(x) = F(x)$ for all x. •

$X_n \to_{\mathcal{D}} X \iff \lim_n F_n(x) = F(x)$ for all $x \in C_F$.

Recall that for a distribution function F, the left-continuous inverse to F is the function F^* mapping $(0,1)$ to $\mathbb{R}$ defined by

$$F^*(u) = \inf\{x : u \leq F(x)\}.$$

$F^*(u)$ is the smallest u^{th}-quantile of F. The following characterization of convergence in distribution in terms of quantiles was part of the first homework (see Exercise 1.10).

Theorem 1. *Let $X_1, X_2, \dots$, and X be real random variables with left-continuous inverse distribution functions $F_1^*, F_2^*, \dots$, and F^* respectively. Then $X_n \to_{\mathcal{D}} X$ if and only if*

$$\lim_{n\to\infty} F_n^*(u) = F^*(u) \quad \text{for all } u \in C_{F^*}. \tag{2}$$

Here $C_{F^*} := \{u \in (0,1) : F^* \text{ is continuous at } u\}$.

Suppose the X_n's and X are random variables for which (2) holds. Let U be a random variable uniformly distributed over $(0,1)$. According to the Inverse Probability Theorem (Theorem 1.5), for each n the random variable $X_n^* := F_n^*(U)$ has the same distribution as X_n; we express this as $X_n^* \sim X_n$. Similarly, $X^* := F^*(U) \sim X$. I claim that X_n^* converges to X^* as $n \to \infty$ in a very strong sense, namely,

$$P[\,\omega : X_n^*(\omega) \to X^*(\omega)\,] = 1. \tag{3}$$

To see this, note that by Exercise 1.9, the set J of jumps of F^* is at most countable; consequently the jumps can be written down in a finite or infinite sequence $u_1, u_2, \dots$. Since $X_n^*(\omega) = F_n^*\big(U(\omega)\big)$ and $X^*(\omega) = F^*\big(U(\omega)\big)$, (2) implies that

$$\begin{aligned} P[\,\omega : X_n^*(\omega) \not\to X^*(\omega)\,] &\leq P[\,\omega : U(\omega) \notin C_{F^*}\,] = P[\,\omega : U(\omega) \in J\,] \\ &= P\big[\textstyle\bigcup_k \{\,\omega : U(\omega) = u_k\}\big] \leq \sum_k P[\,\omega : U(\omega) = u_k\,] = \sum_k 0 = 0. \end{aligned}$$

This proves (3).

Convergence almost everywhere and the representation theorem. The following notion arose in the preceding discussion.

Definition 2. *Let $Y_1, Y_2, \ldots,$ and Y be real random variables, all defined on a common probability space Ω. One says that* ***Y_n converges almost everywhere to Y****, and writes* $\boldsymbol{Y_n \to_{\text{a.e.}} Y}$*, if*

$$P[\,\omega \in \Omega : \lim_n Y_n(\omega) = Y(\omega)\,] = 1. \tag{4}$$

Using this terminology, the result derived on the preceding page can be stated as follows.

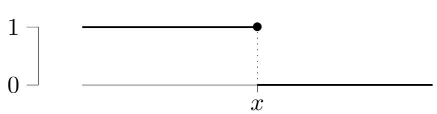
Theorem 2 (The Skorokhod representation theorem). *If $X_n \to_{\mathcal{D}} X$, then there exist on some probability space random variables $X_1^*, X_2^*, \ldots$ and X^* such that*

$$X_n^* \sim X_n \text{ for each } n, \;\; X^* \sim X, \;\text{ and }\; X_n^* \to_{\text{a.e.}} X^*. \tag{5}$$

This is a special case of a result due to the famous Russian probabilist A. V. Skorokhod. It has a number of useful corollaries, which we will turn to after developing some properties of almost everywhere convergence.

The following terminology is used in the next theorem. Let h be a mapping from $\mathbb{R}$ to $\mathbb{R}$ and let X be a real random variable. The ***continuity set of h*** is $C_h := \{\, x : h$ is continuous at $x\,\}$; its complement D_h is called the ***discontinuity set of h***. h is said to be ***X-continuous*** if $P[X \in C_h] = 1$, or, equivalently, $P[X \in D_h] = 0$.

Example 2. Suppose $h = I_{(-\infty,x]}$, as illustrated below:



Graph of $I_{(-\infty,x]}$

Then $D_h = \{x\}$, so h is X-continuous $\iff 0 = P[X \in D_h] = P[X = x]$ $\iff x \in C_F$, F being the distribution function of X. •

$X_n \to_{\text{a.e.}} X$ means $P[\,\omega : X_n(\omega) \to X(\omega)\,] = 1$.
$X_n \to_{\mathcal{D}} X$ means $F_n(x) \to F(x)$ for each $x \in C_F$.

Theorem 3 (The mapping theorem for almost everywhere convergence). *Let $X_1, X_2, \ldots$ and X be real random variables, all defined on a common probability space Ω. If $X_n \to_{\text{a.e.}} X$ as $n \to \infty$ and $h\colon \mathbb{R} \to \mathbb{R}$ is X-continuous, then $h(X_n) \to_{\text{a.e.}} h(X)$.*

Proof Put $Y_n = h(X_n)$ and $Y = h(X)$. If ω is a sample point such that $X_n(\omega) \to X(\omega)$ and $X(\omega) \in C_h$, then $Y_n(\omega) = h\big(X_n(\omega)\big) \to h\big(X(\omega)\big) = Y(\omega)$. Hence

$$\{\omega : Y_n(\omega) \not\to Y(\omega)\} \subset \{\omega : X_n(\omega) \not\to X(\omega)\,\} \cup \{\omega : X(\omega) \notin C_h\}.$$

Since the two events in the union on the right each have probability 0, so does the event on the left. ∎

Theorem 4. *$X_n \to_{\text{a.e.}} X \Longrightarrow X_n \to_{\mathcal{D}} X$.*

Proof Let F_n and F be the distribution functions of X_n and X respectively, and let $x \in C_F$. Since $h := I_{(-\infty,x]}$ is X-continuous (see Example 2), we have

$$Y_n := h(X_n) \to_{\text{a.e.}} Y := h(X).$$

Since $|Y_n| \le 1$ and $E(1) = 1 < \infty$, the DCT implies that

$$E(Y_n) \to E(Y).$$

But $E(Y) = E\big(I_{(-\infty,x]}(X)\big) = P[X \le x] = F(x)$, and similarly, $E(Y_n) = F_n(x)$. Thus $\lim_n F_n(x) = F(x)$, as required. ∎

The converse doesn't hold: in general, $X_n \to_{\mathcal{D}} X$ does not imply $X_n \to_{\text{a.e.}} X$. For example, suppose Y is a random variable taking just the two values ± 1, with probability $1/2$ each. Set $X_n = (-1)^n Y$ for $n = 1, 2, \ldots$. Then $X_n \sim Y$ for each n, so $X_n \to_{\mathcal{D}} Y$. But $X_n \not\to_{\text{a.e.}} Y$, since for each sample point ω, $X_n(\omega)$ alternates between $+1$ and -1, and hence doesn't converge at all, let alone to $Y(\omega)$.

Theorem 5 (The mapping theorem for convergence in distribution). *Suppose $X_n \to_{\mathcal{D}} X$ and $h: \mathbb{R} \to \mathbb{R}$ is X-continuous. Then $h(X_n) \to_{\mathcal{D}} h(X)$.*

Proof Using the Skorokhod representation, construct random variables $X_1^*, X_2^*, \ldots$ and X^* such that $X_n^* \sim X_n$ for each n, $X^* \sim X$, and $X_n^* \to_{\text{a.e.}} X^*$. By Theorem 3, $Y_n^* := h(X_n^*) \to_{\text{a.e.}} Y^* := h(X^*)$, and hence $Y_n^* \to_{\mathcal{D}} Y^*$ by Theorem 4. The claim follows, since $Y_n^* \sim h(X_n)$ for each n and $Y^* \sim h(X)$. ∎

Example 3. Suppose $X_n \to_{\mathcal{D}} Z \sim N(0,1)$. Then $X_n^2 \to_{\mathcal{D}} Z^2 \sim \chi_1^2$. This follows from Theorem 5, by taking $h(x) = x^2$; h is X-continuous since it is continuous — $D_h = \emptyset \Longrightarrow P[X \in D_h] = 0$. •

Theorem 6 (The $\boldsymbol{\Delta}$-method). *Suppose of random variables X_1, $X_2, \ldots$ that*

$$c_n(X_n - x_0) \to_{\mathcal{D}} Y \tag{6}$$

as $n \to \infty$, for some random variable Y, some number $x_0 \in \mathbb{R}$, and some sequence of positive numbers c_n tending to ∞. Suppose also that $g: \mathbb{R} \to \mathbb{R}$ is differentiable at x_0. Then

$$c_n\big(g(X_n) - g(x_0)\big) \to_{\mathcal{D}} g'(x_0)Y. \tag{7}$$

I'll give the formal argument in a moment; here is a heuristic one. Since g is differentiable at x_0, one has

$$g(x_0 + \Delta) \doteq g(x_0) + \Delta g'(x_0) \tag{8}$$

for all small numbers Δ. Put $\Delta_n = X_n - x_0$ and $Y_n = c_n \Delta_n$. Since $Y_n \to_{\mathcal{D}} Y$ and $c_n \to \infty$, for large n it is highly likely that $\Delta_n = Y_n / c_n$ will be close to 0, and hence by (8) that

$$c_n\big[g(X_n) - g(x_0)\big] = c_n\big[\big(g(x_0 + \Delta_n)\big) - g(x_0)\big]$$

will be close to $c_n \Delta_n g'(x_0) = g'(x_0) Y_n$, which will be distributed almost like $g'(x_0)Y$. This method of argument is called the ***$\boldsymbol{\Delta}$-method*** because of the symbol Δ in the seminal approximation (8).

Proof of Theorem 6. By assumption

$$Y_n := c_n(X_n - x_0) \to_{\mathcal{D}} Y.$$

Using the Skorokhod representation theorem, concoct random variables Y_1^*, Y_2^*, … and Y^* having the same marginal distributions as Y_1, Y_2, … and Y respectively, but with $Y_n^* \to_{\text{a.e.}} Y^*$. Since $X_n = x_0 + Y_n / c_n$, it is natural to introduce $X_n^* := x_0 + Y_n^*/c_n$; note that $X_n^* \sim X_n$ for each n and $X_n^* \to_{\text{a.e.}} x_0$ as $n \to \infty$. Now write

$$G_n^* := c_n\big(g(X_n^*) - g(x_0)\big) = c_n(X_n^* - x_0)\Big(\frac{g(X_n^*) - g(x_0)}{X_n^* - x_0}\Big),$$

with the difference quotient taken to be $g'(x_0)$ if $X_n^* = x_0$. Let $n \to \infty$ to get

$$c_n\big(g(X_n^*) - g(x_0)\big) \to G^* := Y^* g'(x_0)$$

almost everywhere, and so also in distribution. This implies $G_n := c_n\big(g(X_n) - g(x_0)\big) \to_{\mathcal{D}} G := Y g'(x_0)$, since $G_n \sim G_n^*$ and $G \sim G^*$. ∎

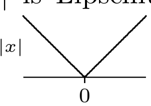
The expectation characterization of convergence in distribution. The following terminology is used in the next theorem. A function $f: \mathbb{R} \to \mathbb{R}$ is said to be ***Lipschitz continuous*** if

$$\|f\|_L := \sup\Big\{\Big|\frac{f(y) - f(x)}{y - x}\Big| : -\infty < x < y < \infty\Big\} < \infty; \tag{9}$$

the quantity $\|f\|_L$ is called the ***Lipschitz norm of $\boldsymbol{f}$***.

Example 4. The absolute value function $f(x) = |x|$ is Lipschitz continuous with norm 1, since

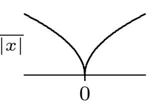
Graph of $f(x) = |x|$



$$\big||y| - |x|\big| \le |y - x|$$

by the triangle inequality. In contrast the function $g(x) = |x|^{1/2}$ is not Lipschitz continuous, since

Graph of $g(x) = \sqrt{|x|}$



$$\lim_{y \downarrow 0} \frac{g(y) - g(0)}{y - 0} = \infty.$$ •

The following result states that $X_n \to_{\mathcal{D}} X$ iff $Ef(X_n) \to Ef(X)$ for all sufficiently smooth bounded functions f. To save space, we write “bd” for “bounded’ in the statement of the theorem.

Theorem 7. *Let $X_1, X_2, \dots, X$ be real random variables. The following are equivalent.*

(a) $X_n \to_{\mathcal{D}} X$.

(b) $Ef(X_n) \to Ef(X)$ *for all bd X-continuous* $f\colon \mathbb{R} \to \mathbb{R}$.

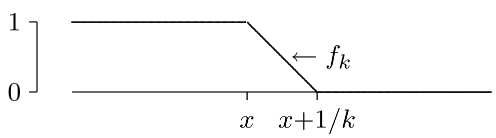
(c) $Ef(X_n) \to Ef(X)$ *for all bd continuous* $f\colon \mathbb{R} \to \mathbb{R}$.

(d) $Ef(X_n) \to Ef(X)$ *for all bd Lipschitz continuous* $f\colon \mathbb{R} \to \mathbb{R}$.

Proof • (a) $\Longrightarrow$ (b): This is rehash of some old ideas. Using the Skorokhod representation, construct random variables $X_1^*, X_2^*, \dots$ and X^* such that $X_n^* \sim X_n$ for each n, $X^* \sim X$, and $X_n^* \to_{\text{a.e.}} X^*$. Since f is X-continuous, $Y_n^* := f(X_n^*) \to_{\text{a.e.}} Y^* := f(X^*)$. Since f is bounded, the DCT implies that $EY_n^* \to EY^*$. Now use $Ef(X_n) = Ef(X_n^*) = EY_n^*$ and $Ef(X) = Ef(X^*) = EY^*$.

• (b) $\Longrightarrow$ (c) $\Longrightarrow$ (d): This is obvious.

• (d) $\Longrightarrow$ (a): Let F_n and F be the distribution functions of X_n and X, respectively. We need to show $\lim_n F_n(x) = F(x)$ for each $x \in C_F$. Fix such an x. For positive integers k, let $f_k\colon \mathbb{R} \to \mathbb{R}$ be the function whose graph is sketched below:



Since $I_{(-\infty,x]} \le f_k$, we have

$$F_n(x) = E(I_{(-\infty,x]}(X_n)) \le E(f_k(X_n))$$

for each n, and since f_k is bounded and Lipschitz continuous, (d) implies

$$\limsup_n F_n(x) \le \limsup_n E(f_k(X_n)) = E(f_k(X)).$$

Since $f_k(X) \to I_{(-\infty,x]}(X)$ as $k \to \infty$, with the convergence being dominated by 1, the DCT implies $\lim_k E(f_k(X)) = E(I_{(-\infty,x]}(X)) =$

$X_n \to X \iff Ef(X_n) \to Ef(X)$ for all Lipschitz continuous $f\colon \mathbb{R} \to \mathbb{R}$.

$F(x)$. The upshot is

$$\limsup_n F_n(x) \le F(x).$$

A similar argument shows that

$$F(x-) = E(I_{(-\infty,x)}(X)) \le \liminf_n F_n(x).$$

Since $x \in C_F \Longrightarrow F(x) = F(x-)$, we have $\lim_n F_n(x) = F(x)$, as required. ∎

Example 5. (A) Let X_n be a random variable taking the two values 0 and n, with probabilities $1 - 1/n$ and $1/n$ respectively. Then $X_n \to_{\mathcal{D}} X$, where $X = 0$. Let $f\colon \mathbb{R} \to \mathbb{R}$ be the identity function: $f(x) = x$ for all x. f is Lipschitz continuous, but $Ef(X_n) = E(X_n) = 0 \times (1 - 1/n) + n \times (1/n) = 1 \not\to 0 = E(X) = Ef(X)$. This doesn't contradict Theorem 7, since f isn't bounded.

(B). Let $t \in \mathbb{R}$ and consider the functions $f_t\colon \mathbb{R} \to \mathbb{R}$ and $g_t\colon \mathbb{R} \to \mathbb{R}$ defined by

$$f_t(x) = \cos(tx) \quad \text{and} \quad g_t(x) = \sin(tx).$$

f_t and g_t are continuous and bounded, so $X_n \to_{\mathcal{D}} X$ implies

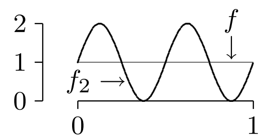
$$\begin{aligned} &E(e^{itX_n}) = E(\cos(tX_n)) + iE(\sin(tX_n)) \\ &\quad \to E(\cos(tX)) + iE(\sin(tX)) = E(e^{itX}). \end{aligned}$$

That is, the characteristic function of X_n converges pointwise to the characteristic function of X. We will see later on that the converse is true too. •

Convergence of densities. Suppose $X_n \to_{\mathcal{D}} X$ and that X_n and X have densities f_n and f respectively. Does f_n converge to f in any nice way? In general, the answer is no.

Example 6. Let X_n and X be random variables taking values in $[0,1]$, with densities

$$f_n(x) = 1 + \sin(2\pi n x)$$
$$f(x) = 1$$



for $0 \le x \le 1$. An easy calculation shows that the corresponding distribution functions satisfy $F_n(x) \to F(x)$ for all $x \in \mathbb{R}$, so $X_n \to_{\mathcal{D}} X$. However, as $n \to \infty$, f_n oscillates ever more wildly, and doesn't converge to f in any nice way; see Exercise 22. •

There is, however, a positive result in the opposite direction: convergence of densities implies convergence in distribution. This fact is a corollary of the next theorem, wherein $\mathcal{B}$ denotes the collection of Borel subsets of $\mathbb{R}$.

Theorem 8. *Let X_1, X_2, … and X be real random variables having corresponding densities f_1, f_2, …, and f with respect to some measure μ. If $\lim f_n(x) = f(x)$ for μ-almost all x, then*

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} |f_n(x) - f(x)|\, \mu(dx) = 0. \tag{10}$$

If (10) *holds, then*

$$\lim_{n\to\infty} \sup\{\, \big|P[X_n \in B] - P[X \in B]\big| : B \in \mathcal{B}\,\} = 0. \tag{11}$$

(11) is a much stronger assertion than $X_n \to_{\mathcal{D}} X$; when it holds, one says that ***the distribution of $\boldsymbol{X_n}$ converges strongly to the distribution of $\boldsymbol{X}$***, and writes $\boldsymbol{X_n \to_s X}$.

Proof of Theorem 8. • (10) holds: Since $|f_n - f| \le U_n := f_n + f \to U := 2f$ and $\int U_n = \int f_n + \int f = 2 \to \int U$, we have $\int |f_n - f| \to 0$ by the Sandwich Theorem (Theorem 10.5).

• (10) $\Longrightarrow$ (11): For each Borel set subset B of $\mathbb{R}$, one has

$$\begin{aligned}\big|P[X_n \in B] - P[X \in B]\big| &= \Big|\textstyle\int_B f_n(x)\,\mu(dx) - \int_B f(x)\,\mu(dx)\Big| \\ &\le \textstyle\int_B \big|f_n(x) - f(x)\big|\,\mu(dx) \le \int_{-\infty}^{\infty} \big|f_n(x) - f(x)\big|\,\mu(dx).\end{aligned}$$ ∎

Example 7. Suppose Y_r is a Gamma random variable with shape parameter r and scale parameter 1. Since Y_r has mean r and variance r, the standardized variable

$$X_r := \frac{Y_r - r}{\sqrt{r}}$$

has mean 0 and variance 1. We are going to show that X_r converges strongly to a standard normal random variable as $r \to \infty$. By the change of variable formula, X_r has density

$$\begin{aligned} f_r(x) &= f_{Y_r}(r + \sqrt{r}x)\sqrt{r} \\ &= \Big[\frac{1}{\Gamma(r)}(r + \sqrt{r}x)^{r-1} e^{-(r+\sqrt{r}x)}\, I_{(0,\infty)}(r + \sqrt{r}x)\Big]\sqrt{r}. \end{aligned}$$

$f_r(x)$ is graphed versus x for $r = 10$, 20, 40, and 80 on the next page; the plot also shows the standard normal density for comparison. By Stirling's formula

$$\begin{aligned} \Gamma(r) &= \frac{\Gamma(r+1)}{r} = (1 + o(1))\frac{\sqrt{2\pi r}\, r^r\, e^{-r}}{r} \\ &= (1 + o(1))\sqrt{2\pi r}\, r^{r-1} e^{-r}, \end{aligned} \tag{12}$$

where $o(1)$ denotes a quantity that tends to 0 as $r \to \infty$. Fix x. For all large r we have $r + \sqrt{r}x > 0$, so

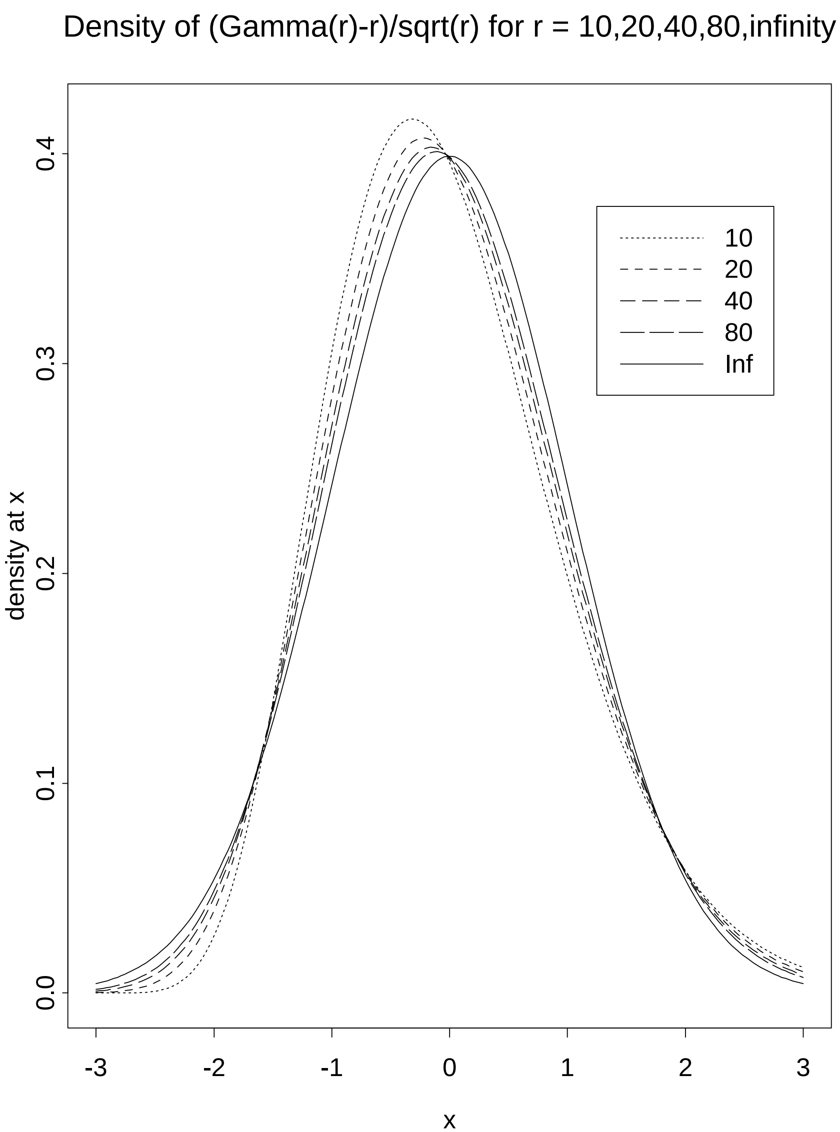
$$f_r(x) = \frac{1}{\sqrt{2\pi}}\,(1 + x/\sqrt{r})^{r-1}\, e^{-\sqrt{r}x}\,(1 + o(1)).$$

Since $\log(1+u) = u - u^2/2 + u^3/3 + \cdots$ for $|u| < 1$,

$$\begin{aligned} \log(f_r(x)) &= C + (r-1)\log(1 + x/\sqrt{r}) - \sqrt{r}x + o(1) \\ &= C + (r-1)\Big[\frac{x}{\sqrt{r}} - \frac{x^2}{2r} + O\Big(\frac{1}{r^{3/2}}\Big)\Big] - \sqrt{r}x + o(1) \\ &= C - x^2/2 + o(1) \end{aligned}$$

where $C = -\log(\sqrt{2\pi})$ and $O(1/r^{3/2})$ denotes a quantity q_r such that $r^{3/2}|q_r|$ remains bounded as $r \to \infty$. Consequently $\lim_{r\to\infty} f_r(x) = \exp(-x^2/2)/\sqrt{2\pi}$. By Theorem 8, $Y_r \to_s Z \sim N(0,1)$. Exercise 25 rearranges this argument into a proof (12). •

Figure 1



Characteristic functions. We are now going to give some criteria for convergence in terms of characteristic functions.

Theorem 9 (The continuity theorem for densities). *Let X_1, X_2, ..., and X be real random variables with corresponding integrable characteristic functions ϕ_1, ϕ_2, ..., and ϕ. Suppose that ϕ_n converges to ϕ in L_1, in the sense that*

$$\|\phi_n - \phi\|_1 := \int_{-\infty}^{\infty} \big|\phi_n(t) - \phi(t)\big|\, dt \to 0 \tag{13}$$

as $n \to \infty$. Then X_1, ..., X have bounded continuous densities f_1, ..., f with respect to Lebesgue measure on $\mathbb{R}$. As $n \to \infty$, f_n converges to f both uniformly and in L_1:

$$\|f_n - f\|_\infty := \sup\{\, |f_n(x) - f(x)| : x \in \mathbb{R}\,\} \to 0, \tag{14}$$

$$\|f_n - f\|_1 := \textstyle\int_{-\infty}^{\infty} |f_n(x) - f(x)|\, dx \to 0. \tag{15}$$

In particular, $X_n \to_s X$.

Proof By the inversion theorem, X_n and X have the bounded continuous densities

$$f_n(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\phi_n(t)\, dt \ \text{ and } \ f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\phi(t)\, dt$$

with respect to Lebesgue measure. (14) follows from

$$\begin{aligned} 2\pi\big|f_n(x) - f(x)\big| &= \left|\int_{-\infty}^{\infty} e^{-itx}\phi_n(t)\, dt - \int_{-\infty}^{\infty} e^{-itx}\phi(t)\, dt\right| \\ &= \left|\int_{-\infty}^{\infty} e^{-itx}\big(\phi_n(t) - \phi(t)\big)\, dt\right| \le \int_{-\infty}^{\infty} \big|\phi_n(t) - \phi(t)\big|\, dt. \end{aligned}$$

(15) follows from (14) and Theorem 8 with $\mu(dx) = dx$. ∎

Now we come to the crucially important

Theorem 10 (The continuity theorem for distribution functions). *Let X_1, X_2, ... and X be real-random variables with corresponding characteristic functions ϕ_1, ϕ_2, ... and ϕ.*

$$X_n \to_{\mathcal{D}} X \iff \phi_n(t) \to \phi(t) \text{ for all } t \in \mathbb{R}. \tag{16}$$

Proof • ($\Longrightarrow$) This was done previously; see Example 5.

• ($\Longleftarrow$) Let $Z \sim N(0,1)$, independently of the X_n's and X. Let k be a (large) positive integer and set $\sigma_k = 1/k$, $Y_{n,k} = X_n + \sigma_k Z$, and $Y_k = X + \sigma_k Z$. I claim that $Y_{n,k} \to_{\mathcal{D}} Y_k$ as $n \to \infty$. To see this, note that $Y_{n,k}$ and Y_k have characteristic functions

$$\psi_{n,k}(t) = E(e^{itY_{n,k}}) = \phi_n(t)e^{-\sigma_k^2 t^2/2}$$
$$\psi_k(t) = E(e^{itY_k}) = \phi(t)e^{-\sigma_k^2 t^2/2}.$$

$\psi_{n,k}$ and ψ_k are integrable, and

$$\int_{-\infty}^{\infty} |\psi_{n,k}(t) - \psi_k(t)|\, dt \le \int_{-\infty}^{\infty} |\phi_n(t) - \phi(t)|\, e^{-\sigma_k^2 t^2/2}\, dt$$

tends to 0 as $n \to \infty$, by virtue of the DCT. Theorem 9 implies that $Y_{n,k}$ converges strongly to Y_k, and hence also in distribution.

We have just shown that for large n, the smoothing $Y_{n,k}$ of X_n has almost the same distribution as the smoothing Y_k of X. The idea now is to show that when k is sufficiently large (and hence the smoothing parameter σ_k is sufficiently small), $Y_{n,k}$ has nearly the same distribution as X_n, and Y_k nearly the same distribution as X; the upshot is that X_n will have nearly the same distribution as X. To make this precise, let $g\colon \mathbb{R} \to \mathbb{R}$ be bounded and Lipschitz continuous. I claim that

$$E\big(g(X_n)\big) \to E\big(g(X)\big) \tag{17}$$

as $n \to \infty$; by criterion (d) of Theorem 7, this is enough to show that $X_n \to_{\mathcal{D}} X$.

(17): $E\big(g(X_n)\big) \to E\big(g(X)\big)$ as $n \to \infty$. $\sigma_k = 1/k$.
$Y_{n,k} := X_n + \sigma_k Z \to_{\mathcal{D}} Y_k := X + \sigma_k Z$ as $n \to \infty$, for each k.

To prove (17), note that by the triangle inequality

$$|Eg(X_n) - Eg(X)| \le I_{n,k} + II_{n,k} + III_k$$

where

$$\begin{aligned} I_{n,k} &= |Eg(X_n) - Eg(Y_{n,k})|, \\ II_{n,k} &= |Eg(Y_{n,k}) - Eg(Y_k)|, \\ III_k &= |Eg(Y_k) - Eg(X)|. \end{aligned}$$

Letting L denote the Lipschitz norm of g, we have

$$|g(X_n) - g(Y_{n,k})| \le L|X_n - Y_{n,k}| = L\sigma_k|Z|$$

since $Y_{n,k} = X_n + \sigma_k Z$. Thus

$$I_{n,k} \le E\big(|g(X_n) - g(Y_{n,k})|\big) \le L\sigma_k E(|Z|).$$

The same bound applies to III_k, for similar reasons. Thus

$$|Eg(X_n) - Eg(X)| \le 2L\sigma_k E(|Z|) + II_{n,k}.$$

This holds for all n and all k. For each k, $\lim_{n\to\infty} II_{n,k} = 0$ since g is continuous and bounded and $Y_{n,k} \to_{\mathcal{D}} Y_k$. (17) follows by taking limits, first as $n \to \infty$ and then as $k \to \infty$. ∎

Remark. With more work, an even better result can be established. Suppose X_1, X_2, ... are random variables with corresponding characteristic functions ϕ_1, ϕ_2, ... If $\phi(t) := \lim_{n\to\infty} \phi_n(t)$ exists for each $t \in \mathbb{R}$ and ϕ is continuous at 0, then ϕ is the characteristic function of some random variable X, and $X_n \to_{\mathcal{D}} X$. •

Example 8. Let ϕ_n be the characteristic function of a random variable X_n having a binomial distribution with parameters n and p_n. If $np_n \to \lambda \in [0,\infty)$ as $n \to \infty$, then for each $t \in \mathbb{R}$

$$\phi_n(t) = (p_n e^{it} + 1 - p_n)^n = \Big(1 + \frac{np_n(e^{it}-1)}{n}\Big)^n$$
$$\to \phi(t) := \exp(\lambda(e^{it}-1))$$

since for complex numbers

$$z_n \to z \Longrightarrow (1 + z_n/n)^n \to e^z \tag{18}$$

(see below). Since ϕ is the characteristic function of a random variable X with a Poisson distribution with parameter λ, this proves that $X_n \to_{\mathcal{D}} X$. •

We need to establish (18).

Theorem 11. (18) *holds.*

Proof For complex numbers ζ satisfying $|\zeta| < 1$, the so-called principal value of $\log(1+\zeta)$ has the power series expansion

$$\log(1+\zeta) = \sum\nolimits_{n=1}^{\infty} (-1)^{n-1}\frac{\zeta^n}{n} = \zeta - \frac{\zeta^2}{2} + \frac{\zeta^3}{3} - \frac{\zeta^4}{4} + \cdots. \tag{19}$$

Consequently for $|\zeta| \le 1/2$

$$\begin{aligned} \big|\log(1+\zeta) - \zeta\big| &\le \frac{|\zeta|^2}{2} + \frac{|\zeta|^3}{3} + \frac{|\zeta|^4}{4} + \cdots \\ &\le \frac{|\zeta|^2}{2}\big(1 + |\zeta| + |\zeta|^2 + \cdots\big) = \frac{|\zeta|^2}{2}\frac{1}{1-|\zeta|} \le |\zeta|^2. \end{aligned} \tag{20}$$

Now suppose $z_n \to z$ in $\mathbb{C}$. Since $z_n/n \to 0$, (20) implies that

$$\begin{aligned} &n\big(\log(1+z_n/n)\big) = n\big((z_n/n) + O(1/n^2)\big) = z_n + O(1/n) \to z \\ &\Longrightarrow (1+z_n/n)^n = \exp\big(n\log(1+z_n/n)\big) \to \exp(z). \end{aligned}$$ ∎

Generalization to random vectors. All of the results of this section carry over to random vectors taking values in $\mathbb{R}^k$ for some integer k. The proofs are straightforward generalizations of the ones for $k = 1$, except for the Skorokhod Representation Theorem, which can no longer be proved using inverse distribution functions.

Exercise 1. For $n = 1, 2, \ldots$, let X_n be a random variable uniformly distributed over the $n+1$ points k/n for $k = 0, 1, \ldots, n$. Show that as $n \to \infty$, $X_n \to_{\mathcal{D}} X$, where $X \sim \text{Uniform}(0,1)$. ◇

Exercise 2. For $n = 1, 2, \ldots$, let G_n be a random variable with a Geometric distribution. Show that if $\mu_n := E(G_n) \to \infty$, then $X_n := G_n/\mu_n \to_{\mathcal{D}} X$, where X is a standard exponential random variable. ◇

Exercise 3. Let $Y_1, Y_2, \ldots$ be iid standard exponential random variables. For each n, set $X_n = \max(Y_1, Y_2, \ldots, Y_n)$. Show that as $n \to \infty$, $X_n - \log(n) \to_{\mathcal{D}} X$, where X is a random variable with distribution function $F(x) = \exp(-\exp(-x))$ for $-\infty < x < \infty$. ◇

In the next 4 exercises, $X_1, X_2, \ldots$, and X are real random variables with corresponding distribution functions $F_1, F_2, \ldots$, and F.

Exercise 4. Show that $X_n \to_{\mathcal{D}} X$ if and only if there exists a dense set D of $\mathbb{R}$ such that $F_n(x) \to F(x)$ for all $x \in D$. ◇

Exercise 5. Suppose $X_n \to_{\mathcal{D}} X$. Show that

$$\lim\nolimits_{c\to\infty} \sup\nolimits_n P[\,|X_n| \ge c\,] = 0. \tag{21}$$ ◇

Exercise 6. Show that $X_n \to_{\mathcal{D}} X$ if and only if

$$\lim\nolimits_n \big(F_n(b) - F_n(a)\big) = F(b) - F(a)$$

for all $a \in C_F$ and $b \in C_F$ with $a < b$. [Hint: for $\Longleftarrow$, first show that (21) holds.] ◇

Exercise 7. Suppose that F is continuous. Show that $X_n \to_{\mathcal{D}} X$ implies that $\lim_n \sup\{\, |F_n(x) - F(x)| : x \in \mathbb{R}\,\} = 0$. ◇

Let $X_1, X_2, \ldots$, and X be real random variables. One says $\boldsymbol{X_n}$ ***converges to X in probability***, and writes $\boldsymbol{X_n \to_P X}$, if

$$\lim_{n\to\infty} P[\,|X_n - X| \geq \epsilon\,] = 0 \text{ for all } \epsilon > 0. \tag{22}$$

Exercise 8. Suppose $X_n \to_P 0$ and there exists a finite constant c such that $|X_n| \leq c$ for all n. Show that $\lim_n E|X_n| = 0$. ◇

Exercise 9. Suppose $X_n \to_{\mathcal{D}} X$ and $Y_n - X_n \to_P 0$. Show that $Y_n \to_{\mathcal{D}} X$. [Hint: use criterion (d) of Theorem 7 in conjunction with the previous exercise.] ◇

Exercise 10. (a) Show that $X_n \to_P X \implies X_n \to_{\mathcal{D}} X$. [Hint: use the preceding exercise.] (b) Show by example that, in general, $X_n \to_{\mathcal{D}} X \not\Longrightarrow X_n \to_P X$. ◇

Exercise 11. Show that if X is constant (with probability one), then $X_n \to_{\mathcal{D}} X \Longrightarrow X_n \to_P X$. ◇

Exercise 12. Suppose $X_1(\omega) \geq X_2(\omega) \geq \cdots \geq X_n(\omega) \geq \cdots \geq 0$ for each sample point ω. Show that $X_n \to_{\text{a.e.}} 0 \Longleftrightarrow X_n \to_P 0$. [Hint: if $A_1, A_2, \ldots$, and A are events such that $A_1 \supset A_2 \supset \cdots \supset A_n \supset \cdots$ and $A = \bigcap_{n=1}^{\infty} A_n$, then $P[A] = \lim_n P[A_n]$; you do not have to prove this fact from measure theory.] ◇

Exercise 13. Show that $X_n \to_{\text{a.e.}} X \Longleftrightarrow \sup_{p\geq n} |X_p - X| \to_P 0$ as $n \to \infty$. [Hint: use the preceding exercise.] ◇

Exercise 14. Suppose $S_1, S_2, \ldots$ are nonnegative variables such that $(S_n - n\mu)/(\sigma\sqrt{n}\,) \to_{\mathcal{D}} Z$, for some positive finite numbers μ and σ and some random variable Z. Show that

$$\sqrt{S_n} - \sqrt{n\mu} \to_{\mathcal{D}} \sigma Z/(2\sqrt{\mu}\,)\,. \tag{23}$$ ◇

Exercise 15 (*The k^{th} order Δ-method*). As in Theorem 6, suppose $X_1, X_2, \ldots$ are random variables such that $c_n(X_n - x_0) \to_{\mathcal{D}} Y$ as $n \to \infty$, for some random variable Y, some number x_0, and some positive numbers c_n tending to ∞. Let k be a positive integer and let $g\colon \mathbb{R} \to \mathbb{R}$ be a function such that

$$g(x_0 + \Delta) - g(x_0) = \gamma\Delta^k + o(\Delta^k) \tag{24}$$

as $\Delta \to 0$. Show that as $n \to \infty$,

$$c_n^k\big(g(X_n) - g(x_0)\big) \to_{\mathcal{D}} \gamma Y^k. \tag{25}$$ ◇

Exercise 16. Let $f\colon [0,1] \to \mathbb{R}$ be a bounded U-continuous function, where $U \sim \text{Uniform}[0,1]$. Show that $(1/n)\sum_{k=0}^{n} f(k/n) \to Ef(U)$ as $n \to \infty$. What does this say about Riemann integration? [Hint: use Theorem 7 and the result of Exercise 1.] ◇

The ***boundary*** ∂B of a subset B of $\mathbb{R}$ is the set of points x for which there exists an infinite sequence $y_1, y_2, \ldots$ of points in B converging to x, and also an infinite sequence $z_1, z_2, \ldots$ of points not in B converging to x; $y_n = x$ or $z_n = x$ for some n's is allowed. For example, the boundary of the interval $B = (a, b]$ is $\partial B = \{a, b\}$.

Exercise 17. Suppose B a (Borel) subset of $\mathbb{R}$. Show that the discontinuity set of the indicator function I_B equals ∂B. Use Theorem 7 to show that if $X_n \to_{\mathcal{D}} X$ and $P[\,X \in \partial B\,] = 0$, then $P[\,X_n \in B\,] \to P[\,X \in B\,]$. How does this relate to (1)? ◇

Exercise 18. Suppose that $X_n \to_{\mathcal{D}} X$ and there exists a strictly positive number ϵ such that $\sup_n E(|X_n|^{1+\epsilon}) < \infty$. Show that: (a) $E(|X|^{1+\epsilon}) < \infty$; (b) X_n and X are integrable; and (c) $E(X_n) \to E(X)$ as $n \to \infty$. [Hint: if $P[\,X = \pm c\,] = 0$, then the function $f_c\colon \mathbb{R} \to \mathbb{R}$ defined by $f_c(x) = xI_{[-c,c]}(x)$ is X-continuous (why?). Moreover, for any random variable Y, one has $E(|Y|I_{\{|Y|\geq c\}}) \leq E(|Y|^{1+\epsilon})/c^{\epsilon}$ (why?).] ◇

Exercise 19. Suppose $X_n \to_{\mathcal{D}} X$. Let M_n and M be the moment generating functions of X_n and X respectively. Show that if $\sup_n M_n(u) < \infty$ for some number $u > 0$, then $\lim_{n\to\infty} M_n(t) = M(t) < \infty$ for all $0 \le t < u$. What happens if $u < 0$? [Hint: use the result of the preceding exercise.] $\diamond$

Exercise 20. Let X_n and X be as in Exercise 1. Show that $X_n \not\to_s X$; specifically, show that for $B = \{\, x \in [0,1] : x \text{ is irrational} \,\}$, one has $P[X_n \in B] = 0$ for all n, whereas $P[X \in B] = 1$. $\diamond$

Exercise 21. Suppose ξ is an irrational number. For each positive integer n, let $x_n = n\xi \bmod 1$. Show that the x_n's are dense in $[0,1)$. [Hint: first show that for each positive integer k, there exist integers i and j with $1 \le i \le k$, $1 \le j \le k$, and $(j-i)\xi \bmod k \le 1/k$.] $\diamond$

Exercise 22. Let f_n and f be as in Example 6. (a) Show that $\limsup_n |f_n(x) - f(x)| = 1$ for each irrational number $x \in [0,1]$. [Hint: use the preceding exercise.] (b) Show that $\int_0^1 |f_n(x) - f(x)|\, dx = 2/\pi$ for all n. $\diamond$

Exercise 23. Suppose that P and Q are two probability distributions on $\mathbb{R}$ having densities p and q with respect to a measure μ. Show that

$$\sup\{\, |Q[B] - P[B]| : B \in \mathcal{B} \,\} = \frac{1}{2}\int_{-\infty}^{\infty} |q(x) - p(x)|\, \mu(dx); \tag{26}$$

here $\mathcal{B}$ denotes the collection of Borel subsets B of $\mathbb{R}$. Deduce that (11) implies (10). [Hint: first show that $\int_{\{q>p\}} (q-p)\, d\mu = \int_{\{p<q\}} (p-q)\, d\mu = 1/2 \int |q-p|\, d\mu$.] $\diamond$

Exercise 24. Suppose that $X_n \to_{\mathcal{D}} X$. Suppose further that X_n and X have densities f_n and f with respect to a measure μ on $\mathbb{R}$ and that there exists a μ-integrable function g such that $f \le g$ and $f_n \le g$ for all n. Show that $P[X_n \in B] \to P[X \in B]$ for each Borel set B (even if f_n does not converge μ-almost everywhere to f). What bearing does this have on Example 6? [Hint: Suppose B is a Borel subset of $\mathbb{R}$ and $\epsilon > 0$. There exist finitely many disjoint subintervals $(a_i, b_i]$ of $\mathbb{R}$ such that $\int_{A \triangle B} g(x)\, \mu(dx) \le \epsilon$; here $A = \bigcup_i (a_i, b_i]$. You do not have to prove this fact from measure theory.] $\diamond$

Exercise 25 (*Stirling's formula for the Gamma function*). Recall that $\Gamma(r+1) = \int_0^\infty y^r e^{-y}\, dy$ for $r \in (0,\infty)$. (a) Use the change of variables $x = (y-r)/\sqrt{r}$ to show that

$$\rho_r := \frac{\Gamma(r+1)}{r^{r+1/2}e^{-r}} = \int_{-\sqrt{r}}^{\infty} e^{-\psi_r(x)}\, dx,$$

where $\psi_r(x) = r\big(\phi(1 + x/\sqrt{r}) - \phi(1)\big)$ with $\phi(u) = u - \log(u)$. (b) Show that $\lim_{r\to\infty} \psi_r(x) = x^2/2$ for each $x \in \mathbb{R}$ and that for all large r, $\psi_r(x) \ge |x| I_{[1,\infty)}(|x|)/4$ for all x. (c) Deduce $\lim_{r\to\infty} \rho_r = \int_{-\infty}^{\infty} \exp(-x^2/2)\, dx = \sqrt{2\pi}$. [Hints: $\phi(u) = (1 + o(1))u^2/2$ as $u \to 0$, and ϕ is convex.] $\diamond$

Exercise 26. Let $X_{\alpha,\beta}$ be a random variable having a Beta distribution with parameters α and β. Show that if α stays fixed and $\beta \to \infty$, then $\beta X_{\alpha,\beta}$ converges strongly to $X \sim G(\alpha, 1)$. $\diamond$

Exercise 27. Let $X_{\alpha,\beta}$ be a random variable having a Beta distribution with parameters $\alpha+1$ and $\beta+1$. Set $\mu_{\alpha,\beta} = \alpha/(\alpha+\beta)$ and $\sigma^2_{\alpha,\beta} = \mu_{\alpha,\beta}(1 - \mu_{\alpha,\beta})/(\alpha+\beta)$. Show if α and β tend to infinity in such a way that $\mu_{\alpha,\beta}$ remains bounded away from 0 and 1, then $(X_{\alpha,\beta} - \mu_{\alpha,\beta})/\sigma_{\alpha,\beta}$ converges strongly to $Z \sim N(0,1)$. $\diamond$

Exercise 28. Let T_n be a random variable having a t distribution with n degrees of freedom. Show that as $n \to \infty$, $T_n \to_s Z \sim N(0,1)$ $\diamond$

Exercise 29. Use Theorem 10 to solve Exercise 2. $\diamond$

Exercise 30. Let X take the values 1 and 0 with probabilities p and $q := 1 - p$ respectively. Show that the characteristic function of $Y := X - E(X) = X - p$ satisfies

$$E(e^{itY}) = (1 + p(e^{it} - 1))e^{-ipt} = 1 - pqt^2/2 + O(t^3) \tag{27}$$

as $t \to 0$. $\diamond$

Exercise 31. Suppose that $X_n \sim \text{Binomial}(n, p)$ for $n = 1, 2, \ldots,$ where $0 < p < 1$. Use Theorem 10 to show that $(X_n - np)/\sqrt{npq} \to_{\mathcal{D}} Z$ as $n \to \infty$. [Hint: use (27).] ⋄

Exercise 32. Show that for complex numbers ζ satisfying $|\zeta| \le c < 1$, the principal value of $\log(1 + \zeta)$ satisfies

$$\Big|\log(1+\zeta) - \sum\nolimits_{j=1}^{k} (-1)^{j-1} \frac{\zeta^j}{j}\Big| \le \frac{|\zeta|^{k+1}}{k+1} \frac{1}{1-c} \tag{28}$$

for each positive integer k. ⋄