Section 5.2. Convergence in Distribution

Note. In this section we define convergence in distribution by considering the limit of a sequence of cumulative distribution functions. We relate convergence in probability and convergence in distribution (see Example 5.2.B and Theorem 5.2.1). We state several theorems concerning convergence in distribution of sequences of random variables. We define "bounded in probability" and show that convergence in distribution implies bounded in probability (see Theorem 5.2.6). We consider a sequence of moment generating functions (see Theorem 5.2.10, a proof of which is "beyond the scope of this book"). After completing this material, we will have the background to address the Central Limit Theorem in the next section.

Definition 5.2.1. Let (X_n) be a sequence of random variables and let X be a random variable. Let F_{X_n} and F_X be, respectively, the cumulative distribution functions of X_n and X. Let $C(F_X)$ denote the set of all points where F_X is continuous Then X_n *converges in distribution* to X if

$$\lim_{n\to\infty} F_n(x) = F_X(x) \text{ for all } x \in C(F_X).$$

When this holds, we denote it as $X_n \xrightarrow{D} X$.

Note. Common terminology for convergence in distribution is to say that X is the "asymptotic distribution" or "limiting distribution" of the sequence (of random variables) (X_n). A common notation to represent $X_n \xrightarrow{D} X$ where X is $N(0,1)$ is to write "$X_n \xrightarrow{D} N(0,1)$"; but this is subtly incorrect since $N(0,1)$ is not a random variable but instead a distribution.

Example 5.2.A. Notice that the definition of convergence in distribution only involves limits for x values where F_X is continuous (i.e., $x \in C(F_X)$). We motivate this by considering random variable X_n with all its mass at $1/n$ (for each $n \in \mathbb{N}$) and random variable X with all its mass at 0. Then the graph of F_X is as given in Figure 5.2.1 (for insight, see Theorem 1.5.1 and Figure 1.5.1 where the cumulative distribution function associated with rolling a 6-sided die is given).

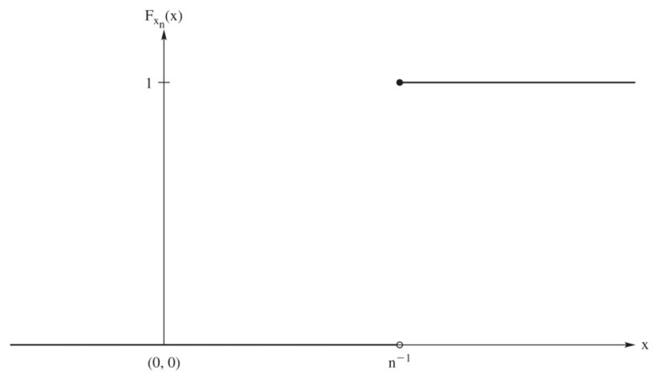


Figure 5.2.1. The Cumulative distribution of X_n.

Since $F_{X_n}(0) = 0$ for all $n \in \mathbb{N}$ then $\lim_{n\to\infty} F_{X_n}(0) = 0 \neq 1 = F_X(0)$. But F_X is not continuous at $x = 0$. For any $x \neq 0$, we have $\lim_{n\to\infty} F_x(x) = F(x)$ (since $F(x) = 0$ if $x < 0$, and $F(x) = 1$ if $x > 0$), and so by definition we have $X_n \xrightarrow{D} X$. Certainly this is the desired case for the sequence (X_n).

Example 5.2.B. Convergence in distribution does not imply convergence in probability, as we now show with an example. Consider X, a continuous random variable with a probability density function $f_X(x)$ that is symmetric about $x = 0$ (i.e., $f_X(-x) = f_X(x)$). Then random variable $-X$ also has probability density function $f_X(x)$. So X and $-X$ have thee same distributions. Define the sequence (X_n)

where

$$X_n \begin{cases} X & \text{if } n \text{ is odd} \\ -X & \text{if } n \text{ is even.} \end{cases}$$

Since $f_X(x) = f_{-X}(x)$ then $F_{X_n}(x) = F_X(x)$ so that $X_n \xrightarrow{D} X$. However, $\lim_{n\to\infty} X_n(x)$ does not exist for any $x \in \mathbb{R}$ such that $X(x) > 0$, hence X_n does not converge to X in probability. We'll see in Theorem 5.2.1 that convergence of (X_n) in probability implies convergence of (X_n) in distribution.

Example 5.2.2. Let discrete random variable X_n have probability mass function

$$p_n(x) = \begin{cases} 1 & \text{for } x = 2 + 1/n \\ 0 & \text{elsewhere.} \end{cases}$$

Notice that $\lim_{n\to\infty} p_n(x) = 0$ for all $x \in \mathbb{R}$. The cumulative distribution function of X_n is (for insight, see Theorem 1.5.1 and Figure 1.5.1, as mentioned above)

$$F_n(x) = \begin{cases} 0 & \text{for } x < 2 + 1/n \\ 1 & \text{for } x \geq 2 + 1/n, \end{cases} \quad \text{so } \lim_{n\to\infty} F_n(x) = \begin{cases} 0 & \text{for } x \leq 2 \\ 1 & \text{for } x > 2. \end{cases}$$

Notice that $\lim_{n\to\infty} F_n(x)$ is not continuous from the right at $x = 2$ (that is, $\lim_{x\to 2^+} F_n(x) = 2 \neq 0 = F(2)$), so by Theorem 1.5.1(d), $\lim_{n\to\infty} F_n(x)$ is not a cumulative distribution function. However,

$$F(x) = \begin{cases} 0 & \text{for } x \ < 2 \\ 1 & \text{for } x \geq 2 \end{cases}$$

is a cumulative distribution function and $\lim_{n\to\infty} F_n(x) = F(x)$ for all x at which F is continuous. So $X_n \xrightarrow{D} X$ where X has cumulative distribution function $F(x)$. This example shows that we cannot determine limiting distributions form probability mass functions. We now turn our attention to relationships between convergence in probability and convergence in distribution. □

Example 5.2.3. Let T_n have a t-distribution with n degrees of freedom where $n \in \mathbb{N} = \{1, 2, \ldots\}$. Then, as seen in Section 3.6, the probability density function is

$$f_n(y) = \int_{-\infty}^{t} \frac{\Gamma((n+1/2)}{\sqrt{\pi n}\Gamma(n/2)} \frac{1}{(1+y^2/n)^{(n+1)/2)}}\, dy$$

and so the cumulative density function is

$$F_n(t) = \int_{-\infty}^{t} f_n(y)\, dy = \int_{-\infty}^{t} \frac{\Gamma((n+1/2)}{\sqrt{\pi n}\Gamma(n/2)} \frac{1}{(1+y^2/n)^{(n+1)/2)}}\, dy.$$

We now show that $T_n \xrightarrow{D} N(0,1)$. So we consider

$$\lim_{n\to\infty} F_n(t) = \lim_{n\to\infty} \int_{-\infty}^{t} f_n(y)\, dy. \qquad (*)$$

We need to pass the limit inside the integral. If we consider Riemann integrals, then we need f_n to converge uniformly to a limit function in order to pass the limit inside the integral (see Theorem 8.3 of my online notes for Analysis 2 [MATH 4227/5227] on Section 8.1. Sequences of Functions). Alternatively, we can appeal to Lebesgue integrals, as Hogg, McKean, and Craig do. By the Lebesgue Dominated Convergence Theorem (see my online notes of Real Analysis 1 [MATH 5210] on Section 4.4. The General Lebesgue Integral), if $|f_n| \le g$ on measurable set E and $\int_E g < \infty$, then

$$\lim_{n\to\infty} \left(\int_E f_n \right) = \int_E \left(\lim_{n\to\infty} f_n \right)$$

(here, all integrals are Lebesgue integrals, through Lebesgue and Riemann integrals coincide when Riemann integrals exist; see Theorem 4.3 of my online notes for Real Analysis 1 on Section 4.2. Lebesgue Integration of a Bounded Measurable Function over a Set of Finite Measure). So we need an integrable function (that is, a function for which the integral exists and is finite) that dominates each f_n. We have $|f_n(y)| = f_n(y) \le 10 f_1(y)$ by Exercise 5.2.A.

$$\begin{aligned}
\int_{-\infty}^t 10 f_1(y)\,dy &= 10\int_{-\infty}^t \frac{\Gamma(1)}{\sqrt{\pi}\Gamma(1/2)}\frac{1}{1+y^2}\,dy \\
&= \frac{10}{\sqrt{\pi}}\frac{0!}{\sqrt{\pi}}\int_{-\infty}^t \frac{1}{1+y^2}\,dy \text{ since } \Gamma(n) = (n-1)! \text{ and } \Gamma(1/2) = \sqrt{\pi} \\
&= \frac{10}{\pi}\lim_{a\to-\infty}\int_1^t \frac{1}{1+y^2}\,dy = \frac{10}{\pi}\lim_{a\to\infty}\left(\tan^{-1}(y)|_1^t\right) \\
&= \frac{1}{\pi}\lim_{a\to-\infty}(\tan^{-1}(y) - \tan^{-1}(a)) = \frac{10}{\pi}(\tan^{-1}(t) - (-\pi/2)) \\
&= 5 + \frac{10}{\pi}\tan^{-1}(t) < \infty.
\end{aligned}$$

So we can apply the Lebesgue Dominated Convergence Theorem to $(*)$ to get

$$\begin{aligned}
\lim_{n\to\infty} F_n(t) &= \lim_{n\to\infty}\int_{-\infty}^t f_n(y)\,dy \\
&= \int_{-\infty}^t \lim_{n\to\infty} f_n(y)\,dy \text{ by the Lebesgue Dominated Convergence Theorem} \\
&= \int_{-\infty}^t \lim_{n\to\infty}\frac{\Gamma((n+1/2)}{\sqrt{\pi n}\Gamma(n/2)}\frac{1}{(1+y^2/n)^{(n+1)/2)}}\,dy \\
&= \int_{-\infty}^t \lim_{n\to\infty}\frac{\Gamma((n+1/2)}{\sqrt{\pi n}\Gamma(n/2)}\lim_{n\to\infty}\frac{1}{(1+y^2/n)^{n/2)}}\lim_{n\to\infty}\frac{1}{\sqrt{2\pi}}\left(1+\frac{y^2}{n}\right)^{-n/2}\,dy \\
&= \int_{-\infty}^t (1)(1)\lim_{n\to\infty}\frac{1}{\sqrt{2\pi}}\left(1+\frac{y^2}{n}\right)^{-n/2}\,dy \text{ since the first limit is 1} \\
&\qquad \text{and by Exercise 5.2.21 } \lim_{n\to\infty}\frac{1}{(1+y^2/n)^{1/2}} = 1 \text{ for all } y\in\mathbb{R} \\
&= \int_{-\infty}^t \frac{1}{\sqrt{2\pi}}(e^{y^2})^{-1/2}\,dy \text{ since } \lim_{n\to\infty}\left(1+\frac{y}{n}\right)^n = e^y \text{ and } \lim_{n\to\infty}\left(1+\frac{y^2}{n}\right)^n = e^{y^2} \\
&= \int_{-\infty}^t e^{-y^2/2}\,dy.
\end{aligned}$$

Since the last integral is the cumulative distribution function of $N(0,1)$, then $T_n \xrightarrow{D} N(0,1)$. $\square$

Note. In "Remark 5.2.3," Hogg, McKean, and Craig briefly review the ideas of limit superior, $\overline{\lim}$, and limit inferior, $\underline{\lim}$, of a sequence of real numbers. These ideas covered in Analysis 1 (MATH 4217/5217) in Section 2.3. Bolzano-Weierstrass Theorem.

Theorem 5.2.1. If sequence of random variables (X_n) converges to X in probability, then (X_n) converges to X in distribution.

Note. We saw in Example 5.2.B that there is a sequence of random variables (X_n) such that $X_n \xrightarrow{D} X$ but $X_n \overset{P}{\not\to} X$. Therefore, the converse of Theorem 5.2.1 does not (in general) hold. However, in some special cases, the converse does hold.

Theorem 5.2.2. If sequence of random variables (X_n) converges to constant b in distribution, then (X_n) converges to b in probability.

Note. The proof of the next theorem is similar to that of Theorem 5.2.2 and is to be given in Exercise 5.2.13.

Theorem 5.2.3. Suppose sequence of random variables (X_n) converges to X in distribution and sequence of random variables (Y_n) converges in probability to 0. Then $X_n + Y_n$ converges to X in distribution.

Note. Theorem 5.2.3 has the following application.

Lemma 5.2.A. Suppose sequence of random variables (Y_n) converges to X in distribution and that sequence of random variables $(Y_n - X_n)$ converges in probability to 0. Then $(X_n) = (Y_n + (X_n - Y_n))$ converges to X in distribution.

Note. A proof of the next theorem "can be found in a more advanced text. . . " (see Hogg, McKean, and Craig, page 332). A proof is given in *Probability and Measure Theory* 2nd Edition, by Robert B. Ash with contributions from Catherine Doleans-Dade (Academic Press, 2000); see "Theorem 7.7.3. Convergence of Transformed Sequences" (see part (c)) in "Chapter 7. The Central Limit Theorem." Notice that the result also hold if convergence "in distribution" is replaced with convergence "in probability" (see part (b)) or with convergence "almost everywhere" (see part (a)). The proof uses Slutsky's Theorem (Theorem 7.7.1 of *Probability and Measure Theory*, to be stated below) and Skorokhod's Theorem (Theorem 7.7.2 of *Probability and Measure Theory*).

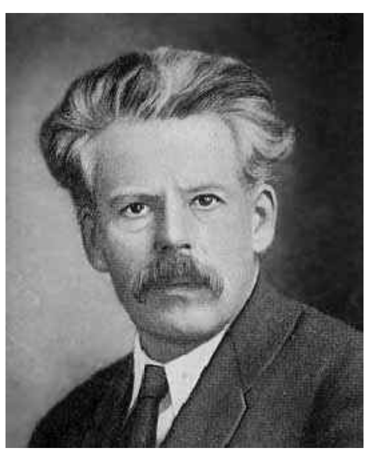
Theorem 5.2.4. Suppose sequence of random variables (X_n) converges to X in distribution and g is a continuous function on the support of X. Then $g(X_n)$ converges to $g(X)$ in distribution.

Note. The proof of the next result is similar to that of Theorem 5.2.1 and is to be given in Exercise 5.2.B. A proof can also be found in *Probability and Measure Theory*: see Theorem 7.7.1 in "Chapter 7. The Central Limit Theorem."

Theorem 5.2.5. Slutsky's Theorem.

Let X_n, X, A_n, and B_n be random variables (where $n \in \mathbb{N}$) and let a and b constants. If $X_n \xrightarrow{D} X$, $A_n \xrightarrow{P} a$, and $B_n \xrightarrow{P} b$, then $A_n + B_n X_n \xrightarrow{D} a + bX$.

Note. Evgeny Slutsky (April 19, 1880–March 10, 1948) was born in Russia and grew up in Kiev in Ukraine. He attended the University of Kiev but, following political turmoil, transfered to Munich Polytechnikum in 1902 and complete a degree in engineering there. He did a degree in political economics at the University of Kiev in 1911. From 1913 to 1926 he taught at the Kiev Institute of Commerce. In 1926 he started working for the government statistics office in Moscow. He worked in the foundations of probability theory, a safe topic in Stalin's Soviet Union. He joined the Central Institute of Meteorology in 1931 and later joined the University of Moscow. In 1938 he started working for the USSR Academy of Sciences. Slutsky was influenced by the work of Karl Pearson. His interests included the mathematical foundations of statistical methods and applications of statistics to economics and the natural sciences. This biographical information and the image below are from the MacTutor History of Mathematics Archive on Evgeny Slutsky.



Evgeny Evgenievich Slutsky (1880–1948)

The theorem named for Slutsky appeared in "Über stochastische Asymptoten und Grenzwerte," *Metron*, **5**(3), 3–89 (1925); for a review of the paper, see the Zentralblatt MATH page (the paper and review are both in German).

Note. Let X be any random variable X with cumulative distribution function $F_X(x)$. Then $\lim_{x\to-\infty} F_X(x) = 0$ and $\lim_{x\to\infty} F_X(x) = 1$. So for any given $\varepsilon > 0$, there are $\eta_1, \eta_2 \in \mathbb{R}$ such that $F_X(x) < \varepsilon/2$ for $x \le \eta_1$, and $F_X(x) > 1 - \varepsilon/2$ for $x \ge \eta_2$. Let $\eta = \max\{|\eta_1|, |\eta_2|\}$. Then

$$P(|X| \le \eta) = P(-\eta \le X \le \eta) = P(X \le \eta) - P(X < -\eta)$$

$$= P(X \le \eta) - P(X \le -\eta) + P(X = -\eta) = F_X(\eta) - F_X(-\eta) + P(X = -\eta)$$

$$> \left(1 - \frac{\varepsilon}{2}\right) - \left(\frac{\varepsilon}{2}\right) + 0 = 1 - \varepsilon. \tag{5.2.7}$$

So even if random variable X is unbounded then it is bounded in probability in this sense. We now state a definition related to this idea, but in the setting of a sequence of random variables.

Definition 5.2.2. The sequence of random variables (X_n) is *bounded in probability* if, for all $\varepsilon > 0$, there exists a constant $B_\varepsilon > 0$ and $N_\varepsilon \in \mathbb{N}$ such that

$$\eta \ge N_\varepsilon \text{ implies } P(|X_n| \le B_\varepsilon) \ge 1 - \varepsilon.$$

Note. The next Theorem shows that a sequence which converges in distribution is bounded in probability.

Theorem 5.2.6. Let (X_n) be a sequence of random variables and let X be a random variable. If $X_n \to X$ in distribution, then (X_n) is bounded in probability.

Note. The converse of Theorem 5.2.6 does not hold, as we show in the following example. That is, there exists a sequence (X_n) that is bounded in probability but (X_n) does not converge in distribution.

Example 5.2.5. Let the sequence of random variables (X_n) be defined as $X_n = 2+1/n$ for n even (with probability 1) and $X_n = 1+1/(n+1)$ for n odd (with probability 1). The sequence $(X_2, X_4, X_6, \ldots)$ converges in distribution to the random variable $Y = 2$, and the sequence $(X_1, X_3, X_5, \ldots)$ converges in distribution to the random variable $W = 1$. Since $Y \neq W$ then the sequence (X_n) does not converge in distribution. Since all X_n lie in the interval $[1, 5/2]$ then with $B_\varepsilon = 5/2$ and $N_\varepsilon = 1$, we have for all $\varepsilon > 0$ that $n \geq N_\varepsilon = 1$ implies $P(|X_n| \leq B_\varepsilon = 5/2) = 1 \geq 1 - \varepsilon$. So (X_n) is bounded in probability.

Theorem 5.2.7. Let (X_n) be a sequence of random variables which is bounded in probability and let (Y_n) be a sequence of random variables that converges to 0 in probability. Then $X_n Y_n \xrightarrow{P} 0$.

Note. We now turn our attention to the "Δ method" (given in Theorem 5.2.9 below). But first, we need to explore Taylor's Theorem from the theory of power series (Hogg, McKean, and Craig correctly refer to this as "the mean value theorem with reminder" but mistakenly claim that it is sometimes called "Young's Theorem"; they give two references, "Hardy (1992)" and "Lehmann (1999)," but both refer to this as Taylor's Theorem). E. L. Lehmann, *Elements of Large-Sample Theory*, NY: Springer-Verlag (1999) [accessed 5/28/2021] on page 85 states:

Theorem 2.5.1 of Lehmann. Suppose that $f(x)$ has r derivatives at the point a. Then

$$f(a+b) = f(a) + bf'(a) + \frac{b^r}{r!} f^{(r)}(a) + o(b^r).$$

We describe the "little o" notation below. With $f(x) = g(x)$, $y = a+b$, $x = a$, and $r = 1$ this implies

Theorem 5.2.A. A General Mean Value Theorem.

Suppose that $g(x)$ is differentiable at x. Then

$$g(y) = g(x) + (y-x)g'(x) + o(y-x).$$

This is the statement that Hogg, McKean, and Craig give in (5.2.9). However, Lehman does not give a proof and defers to G. H. Hardy, who states on page 320 of *A Course of Pure Mathematics*, Third Edition, Cambridge University Press (1921):

Taylor's or the General Mean Value Theorem.

If $f(x)$ is a function of x which has derivatives of the first n orders throughout the interval $[a,b]$, then

$$f(b) = f(a) + (b-a)f'(a) + \frac{(b-a)^2}{2!} f''(a) + \cdots$$
$$+ \frac{(b-a)^{n-1}}{(n-1)!} f^{(n-1)}(a) + \frac{(b-a)^n}{n!} f^{(n)}(\xi),$$

where $a < \xi < b$.

Notice that this version requires a level of differentiability on $[a,b]$, whereas Theorem 5.2.A only requires differentiability at a point. Hardy *does* gives a proof of

"Taylor's or the General Mean Value Theorem" (as we do in Analysis 2 [MATH 4227/5227]; see "Corollary 8.13(c). Taylor's Law of the Mean" in my online Analysis 2 notes on Section 8.3. Taylor Series in which differentiability on an open interval containing $[a, b]$ is required). You also see it in Calculus 2 (MATH 1920); see "Theorem. Taylors Formula" in my online Calculus 2 notes on Section 10.9. Convergence of Taylor Series; Error Estimates. Finally, Theorem 5.2.A is also given as Exercise 5.2.15(b) (where differentiability on an interval is required) in James R. Kirkwood's *An Introduction to Analysis*, Second Edition, Waveland Press (2002) (this is the book I use in Analysis 1 and 2). So we take Lehmann's Theorem 2.5.1 and Theorem 5.2.A as correct, though we have some lingering concerns about the set on which we impose the differentiability hypothesis (this arises again in the Δ method in Theorem 5.2.9).

Note. Recall from Calculus 2 (MATH 1920) that: A function f is *of smaller order than* g as $x \to a$ (where $a \in \mathbb{R} \cup \{\pm\infty\}$) if $\lim_{x \to a} f(x)/g(x) = 0$, denoted $f = o(g)$ (read "f is little-oh of g"). Notice that $\displaystyle \lim_{b-a \to 0} \frac{(b-a)^n}{n!} f^{(n)}(\xi) = 0$ so that "Taylor's of the General Mean Value Theorem" implies Theorem 5.2.A. Also, f is *of at most the order of* g as $x \to a$ (where $a \in \mathbb{R} \cup \{\pm\infty\}$) if there is a neighborhood of a for which $f(x)/g(x) \leq M$, denoted $f = O(g)$. See my online Calculus 2 notes on Section 7.4. Relative Rates of Growth (where the definitions are given for $a = \infty$ only). We now define a similar concept concerning convergence-in and boundedness-in probability.

Definition. Consider two sequences of random variables, (X_n) and (Y_n). Sequence (Y_n) is *little-o_p* of (X_n), denoted $Y_n = o_p(X_n)$, if $Y_n/X_n \xrightarrow{P} 0$ as $n \to \infty$. Sequence (Y_n) is *big-O_p* of (X_n), denoted $Y_n = O_p(X_n)$, if sequence (Y_n/X_n) is bounded in probability as $n \to \infty$.

Theorem 5.2.8. Suppose sequence of random variables (Y_n) is bounded in probability. Suppose $X_n = o_p(Y_n)$. Then $X_n \xrightarrow{P} 0$ as $n \to \infty$.

Note. The result in the next theorem is called the "Δ method." Notice that we hypothesize the differentiability of g at a single point (and still have concerns over this hypothesis, as mentioned above).

Theorem 5.2.9. (The Δ Method). Let (X_n) be a sequence of random variables which that $\sqrt{n}(X_n - \theta) \xrightarrow{D} N(0, \sigma^2)$. Suppose the function $g(x)$ is differentiable at θ and $g'(\theta) \neq 0$. Then

$$\sqrt{n}(g(X_n) - g(\theta)) \xrightarrow{D} N(0, \sigma^2(g'(\theta))^2).$$

Note. To show that $X_n \xrightarrow{D} X$ using the definition, we need the cumulative distribution functions F_{X_n} and F_X. The next theorem shows that we can conclude $X_n \xrightarrow{D} X$ based on the use of moment generating functions M_{X_n} and M, instead of cumulative distribution functions.

Theorem 5.2.10. Let (X_n) be a sequence of random variables with moment generating function $M_{X_n}(t)$ that exists for $-h < t < h$ for all $n \in \mathbb{N}$. Let X be a random variable with moment generating function $M(t)$, which exists for $|t| \le h_1 \le h$. If $\lim_{n\to\infty} M_n(t) = M(t)$ for $|t| \le h_1$, then $X_n \xrightarrow{D} X$.

Note. The proof of Theorem 5.2.10 is "beyond the scope of this book," and Hogg, McKean, and Craig mention Leo Breiman's *Probability* (London: Addison-Wesley, 1968) page 171 as a reference. My preferred reference on measure theory based probability, Robert B. Ash's (with contributions from Catherine Doleans-Dade) *Probability and Measure Theory* 2nd Edition (Academic Press, 2000), addresses convergence in distribution in "Chapter 7. The Central Limit Theorem" but it does not seem to specifically address Theorem 5.2.10. We need the following limit result in some of our convergence exercises and examples.

Theorem 5.2.B. Let $b, c \in \mathbb{R}$ and suppose $\lim_{n\to\infty} \psi(n) = 0$. Then

$$\lim_{n\to\infty}\left(1 + \frac{b}{n} + \frac{\psi(n)}{n}\right)^{cn} = \lim_{n\to\infty}\left(1 + \frac{b}{n}\right) = e^{bc}.$$

Note. We take Theorem 5.2.B as given (from Advanced Calculus, as Hogg, McKean, and Craig say). However, if we add the hypotheses that $\psi(x)$ and $\psi'(x)$ are defined for all $x \in \mathbb{R}$ sufficiently large, $\lim_{x\to\infty} \psi(x) = 0$, and $\lim_{x\to\infty} x\psi'(x) = 0$, then we can prove Theorem 5.2.B using L'Hôpital's Rule as is to be done in Exercise 5.2.C.

Example 5.2.7. Let Z_n be $\chi^2(n)$. Then the moment generating function is $(1-2t)^{-n/2}$ for $t < 1/2$ (by definition). The mean and variance are n and $2n$, respectively (as explained in Section 3.3). Consider the sequence of ransom variables $Y_n = (Z_n - n)/\sqrt{2n}$. By defining the moment generating function of Y_n is

$$\begin{aligned}
M_{Y_n}(t) &= E\left(\exp\left(t\frac{Z_n - n}{\sqrt{2n}}\right)\right) = E(\exp(tZ_n/\sqrt{2n})\exp(-tn/\sqrt{2n})) \\
&= \exp(-tn/\sqrt{2n})E(\exp(tZ_n/\sqrt{2n})) \text{ since expectation is linear} \\
&= \exp(-tn/\sqrt{2n})\left(1 - 2\frac{t}{\sqrt{2n}}\right)^{-n/2} \text{ for } t < \frac{\sqrt{2n}}{2} \\
&\quad \text{since } E(tZ_n) = (1-2t)^{-n/2} \text{ for } t < 1/2 \text{ implies} \\
&\quad E\left(\frac{t}{\sqrt{2n}}Z_n\right) = \left(1 - 2\frac{t}{\sqrt{2n}}\right)^{-n/2} \text{ (replacing } t \text{ with } t/\sqrt{2n}) \\
&\quad \text{for } \frac{t}{\sqrt{2n}} < \frac{1}{2} \text{ or } t < \frac{\sqrt{2n}}{2} \\
&= \exp\left(-t\sqrt{\frac{2}{n}\frac{n}{2}}\right)\left(1 - 2\frac{t}{\sqrt{2n}}\right)^{-n/2} \text{ for } t < \frac{\sqrt{2n}}{2} \\
&= (e^{t\sqrt{2/n}})^{-n/2}\left(1 - 2\frac{t}{\sqrt{2n}}\right)^{-n/2} \text{ for } t < \frac{\sqrt{2n}}{2} \\
&= \left(e^{t\sqrt{2/n}}\left(1 - 2\frac{t}{\sqrt{2n}}\right)\right)^{-n/2} \text{ for } t < \frac{\sqrt{2n}}{2} \\
&= \left(e^{t\sqrt{2/n}} - \sqrt{\frac{2}{n}}te^{t\sqrt{2/n}}\right)^{-n/2} \text{ for } t < \frac{\sqrt{2n}}{2}. \qquad (*)
\end{aligned}$$

Applying "Taylor's or the General Mean Value Theorem" (stated after Theorem 5.2.A) applied to $f(x) = e^x$ with $n = 3$ implies that

$$f(b) = f(a) + (b-a)f'(a) + \frac{(b-a)^2}{2!}f''(a) + \frac{(b-a)^3}{3!}f'''(\xi)$$

for some $\xi \in (a,b)$. With $b = t\sqrt{2/n}$ and $a = 0$ this gives

$$e^{t\sqrt{2/n}} = 1 + t\sqrt{\frac{2}{n}} + \frac{1}{2}\left(t\sqrt{\frac{2}{n}}\right)^2 + \frac{1}{6}\left(t\sqrt{\frac{2}{n}}\right)^3 e^{\xi(n)}$$

for some $\xi(n) \in (0, t\sqrt{2/n})$. Substituting this into $(*)$ gives

$$\begin{aligned} M_{Y_n}(t) &= \left(e^{t\sqrt{2/n}} - \sqrt{\frac{2}{n}}te^{t\sqrt{2/n}}\right)^{-n/2} \\ &= \left(1 + t\sqrt{\frac{2}{n}} + \frac{t^2}{n} + \frac{\sqrt{2}t^3e^{\xi(n)}}{2n\sqrt{n}} - \sqrt{\frac{2}{n}}t\left(1 + t\sqrt{\frac{2}{n}} + \frac{t^2}{n} + \frac{\sqrt{2}t^3e^{\xi(n)}}{3n\sqrt{n}}\right)\right)^{-n/2} \\ &= \left(1 - \frac{t^2}{n} + \frac{\sqrt{2}t^3e^{\xi(n)}}{3n\sqrt{n}} - \frac{\sqrt{2}t^3}{n\sqrt{n}} - \frac{2t^4e^{\xi(n)}}{3n^2}\right)^{-n/2} \\ &= \left(1 - \frac{t^2}{n} + \frac{\psi(n)}{n}\right)^{-n/2} \text{ where } \psi(n) = \frac{\sqrt{2}t^3e^{\xi(n)}}{2\sqrt{n}} - \frac{\sqrt{2}t^3}{\sqrt{n}} - \frac{2t^4e^{\xi(n)}}{3n}. \end{aligned}$$

Since $0 < \xi(n) < t\sqrt{2/n}$ then $\lim_{n\to\infty} \xi(n) = 0$ and so $\lim_{n\to\infty} \psi(n) = 0$. So by Theorem 5.2.B, we have

$$M_Y(t) = \lim_{n\to\infty} M_{Y_n}(t) = \lim_{n\to\infty}\left(1 - \frac{t^2}{n} + \frac{\psi(n)}{n}\right)^{-n/2} = e^{-t^2/n}.$$

By Note 3.4.A, the moment generating function of $N(0,1)$ is $e^{t^2/2}$, so we can conclude from Theorem 5.2.10 that $Y_n \xrightarrow{D} N(0,1)$. This is illustrated in Figure 5.2.2 (below) in which histograms of random samples of size 1000 from $\chi^2(n)$ distributions are given for $n = 5, 10, 20, 50$. A standard normal distribution is also given in each case. Notice that as n increases, the histograms better approximate $N(0,1)$.

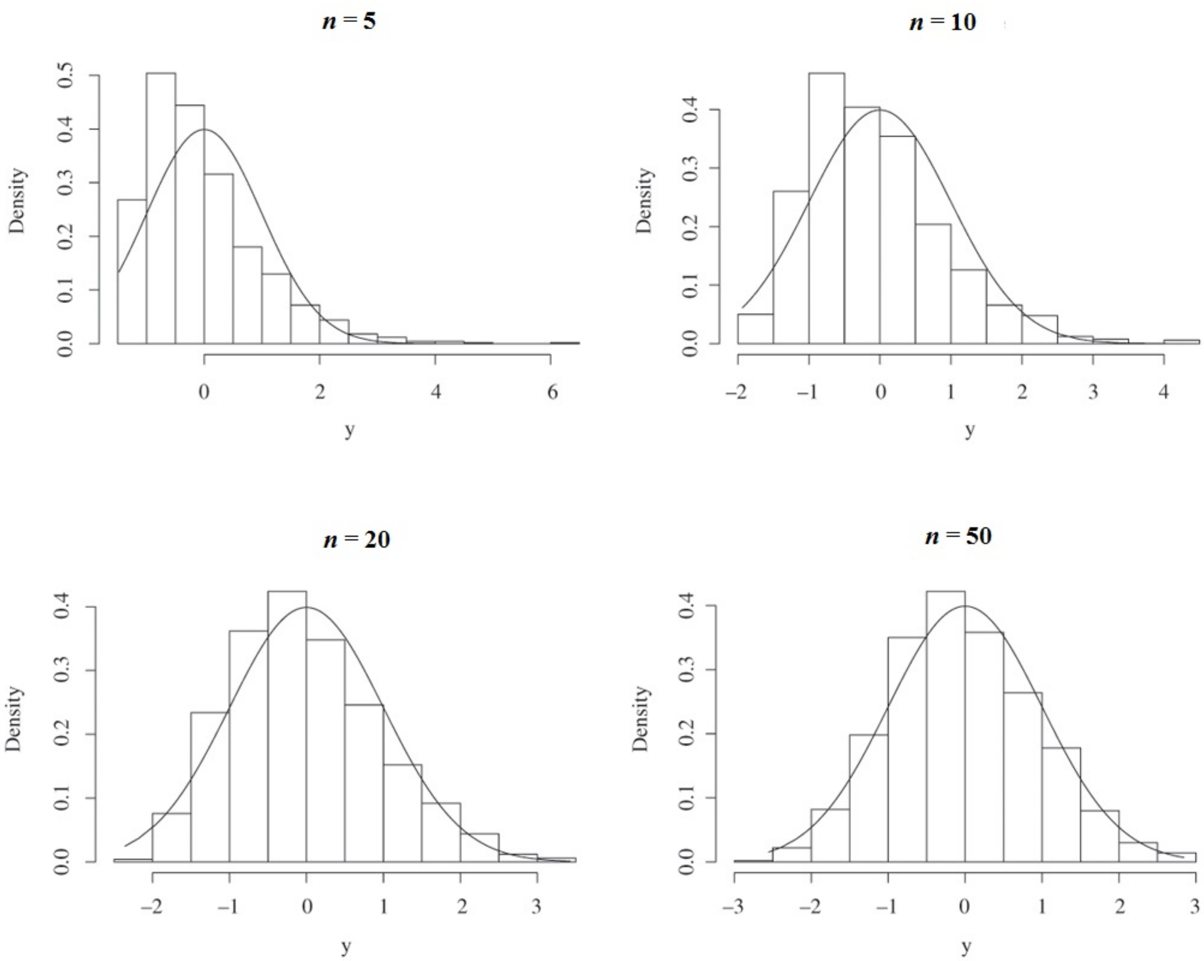


Figure 5.2.2. (Corrected)

Example 5.2.8 (Example 5.2.7 continued). In the previous example, we showed that for $Z_n = \chi^2(n)$, we have

$$Y_n = (Z_n - n)/\sqrt{2n} = \sqrt{n}\left(\frac{1}{\sqrt{2n}}Z_n - \frac{1}{\sqrt{2}}\right) \xrightarrow{D} N(0,1).$$

We now illustrate the Δ model by letting $g(t) = \sqrt{t}$ and let $W_n = g(Z_n/\sqrt{2n}) = (Z_n/\sqrt{2n})^{1/2}$. Now $g(1/\sqrt{2}) = 1/2^{1/4}$ and

$$g'(1/\sqrt{2}) = \frac{1}{2}\frac{1}{\sqrt{1/\sqrt{2}}} = 1/2^{3/4} = 2^{-3/4}.$$

By Theorem 5.2.9 ("the Δ method") we have $\sqrt{n}(g(X_n) - g(\theta)) \xrightarrow{D} N(0, \sigma^2(g'(\theta))^2)$ where $g(X_n) = g(Z_n/\sqrt{2n}) = W_n$, $g(\theta) = g(1/\sqrt{2}) = 1/2^{1/4}$, $\sigma^2 = 1$, and $g'(\theta) = g'(1/\sqrt{2}) = 2^{-3/4}$, or $\sqrt{n}(W_n - 1/2^{1/4}) \xrightarrow{D} N(0, 2^{-3/2})$.

Revised: 6/6/2021