8. Convergence in Distribution

Basic Theory

Definition

Suppose that $X_n,\ n \in \mathbb{N}_+$ and X are real-valued random variables with distribution functions $F_n,\ n \in \mathbb{N}_+$ and F, respectively. We say that the distribution of X_n **converges** to the distribution of X as $n \to \infty$ if

$$F_n(x) \to F(x) \text{ as } n \to \infty$$

for all x at which F is continuous. The first fact to notice is that convergence in distribution, as the name suggests, only involves the *distributions* of the random variables. Thus, the random variables need not even be defined on the same probability space (that is, they need not be defined for the same random experiment). This is in sharp contrast to the other modes of convergence we have studied:

- Convergence with probability 1
- Convergence in probability
- Convergence in k^{th} mean

We will show, in fact, that convergence in distribution is the weakest of all of these modes of convergence. It is nonetheless very important. The central limit theorem, one of the two fundamental theorems of probability, is a theorem about convergence in distribution.

Preliminary Examples

The examples below show why the definition is given in terms of distribution functions, rather than density functions, and why convergence is only required at the points of continuity of the limiting distribution function.

1. Let $X_n = \frac{1}{n}$ for $n \in \mathbb{N}_+$ and let $X = 0$. Let f_n and f be the corresponding density functions and let F_n and F be the corresponding distribution functions. Show that

a. $f_n(x) \to 0$ as $n \to \infty$ for all $x \in \mathbb{R}$
b. $F_n(x) \to \begin{cases} 0, & x \le 0 \\ 1, & x > 0 \end{cases}$ as $n \to \infty$
c. $F_n(x) \to F(x)$ as $n \to \infty$ for all $x \ne 0$

2. Suppose that X_n has the discrete uniform distribution on $\left\{\frac{1}{n}, \frac{2}{n}, ..., \frac{n-1}{n}, 1\right\}$ for each $n \in \mathbb{N}_+$. Let X have the continuous uniform distribution on the interval $[0, 1]$

a. Show that the distribution of X_n converges to the distribution of X as $n \to \infty$.
b. Recall that $\mathbb{Q}$ denotes the set of rational numbers. Show that $\mathbb{P}(X_n \in \mathbb{Q}) = 1$ for each n but $\mathbb{P}(X \in \mathbb{Q}) = 0$.

c. Let f_n denote the (discrete) probability density function of X_n. Show that $f_n(x) \to 0$ as $n \to \infty$ for all $x \in [0, 1]$

Probability Density Functions

As Exercise 2 shows, it is quite possible to have a sequence of discrete distributions converge to a continuous distribution (or the other way around). Recall that probability density functions have very different meanings in the discrete and continuous cases: density with respect to counting measure in the first case, and density with respect to Lebesgue measure in the second case. This is another indication that distribution functions, rather than density functions, are the correct objects of study. However, if probability density functions of a fixed type converge then the distributions converge. The following results are a consequence of Scheffe's theorem, which is given in advanced topics below.

Suppose that $f_n, \; n \in \mathbb{N}_+$ and f are probability density functions for discrete distributions on a countable set S, and that $f_n(x) \to f(x)$ as $n \to \infty$ for each $x \in S$. Then the distribution defined by f_n converges to the distribution defined by f as $n \to \infty$. Similarly, suppose that $f_n, \; n \in \mathbb{N}_+$ and f are probability density functions for continuous distributions on $\mathbb{R}$, and that $f_n(x) \to f(x)$ as $n \to \infty$ for all $x \in \mathbb{R}$ (except perhaps on a set with Lebesgue measure 0). Then the distribution defined by f_n converges to the distribution defined by f as $n \to \infty$.

Relation to Convergence in Probability

3. Suppose that $(X_1, X_2, \ldots)$ and X are random variables (defined on the same probability space) with distribution functions $(F_1, F_2, \ldots)$ and F, respectively. Show that if $X_n \to X$ as $n \to \infty$ in probability, then the distribution of X_n converges to the distribution of X as $n \to \infty$.

a. Show that $\mathbb{P}(X_n \le x) = \mathbb{P}(X_n \le x, \; X \le x + \varepsilon) + \mathbb{P}(X_n \le x, \; X > x + \varepsilon)$ for $\varepsilon > 0$.
b. Conclude that $F_n(x) \le F(x + \varepsilon) + \mathbb{P}(|X_n - X| > \varepsilon)$ for $\varepsilon > 0$,
c. Show that $\mathbb{P}(X \le x - \varepsilon) = \mathbb{P}(X \le x - \varepsilon, \; X_n \le x) + \mathbb{P}(X \le x - \varepsilon, \; X_n > x)$ for $\varepsilon > 0$.
d. Conclude that $F(x - \varepsilon) \le F_n(x) + \mathbb{P}(|X_n - X| > \varepsilon)$ for $\varepsilon > 0$.
e. Combine (b) and (d) to conclude that $F(x - \varepsilon) - \mathbb{P}(|X_n - X| > \varepsilon) \le F_n(x) \le F(x + \varepsilon) + \mathbb{P}(|X_n - X| > \varepsilon)$ for $\varepsilon > 0$.
f. Let $n \to \infty$ in (e) to show that $F(x - \varepsilon) \le \liminf_{n \to \infty} F_n(x) \le \limsup_{n \to \infty} F_n(x) \le F(x + \varepsilon)$ for $\varepsilon > 0$.
g. Let $\varepsilon \downarrow 0$ in (f) to show that if F is continuous at x then $F_n(x) \to F(x) \;\text{ as }\; n \to \infty$

Our next example shows that even when the variables are defined on the same probability space, a sequence can converge in distribution, but not in any other way.

4. Let X be an indicator variable with $\mathbb{P}(X = 1) = \mathbb{P}(X = 0) = \frac{1}{2}$, and let $X_n = X$ for $n \in \mathbb{N}_+$. Show that

a. $1 - X$ has the same distribution as X.
b. The distribution of X_n converges to the distribution of $1 - X$ as $n \to \infty$.
c. $|X_n - (1 - X)| = 1$ for any $n \in \mathbb{N}_+$
d. $\mathbb{P}(X_n \;\text{ does not converge to }\; 1 - X \;\text{ as }\; n \to \infty) = 1$
e. $\mathbb{P}\left(|X_n - (1 - X)| > \frac{1}{2}\right) = 1$ for each $n \in \mathbb{N}_+$ so X_n does not converge to $1 - X$ as $n \to \infty$ in probability.

f. $\mathbb{E}(|X_n - (1 - X)|) = 1$ for each $n \in \{1, 2, \ldots\}$ so X_n does not converge to $1 - X$ as $n \to \infty$ in mean.

To summarize, we have the following implications for the various modes of convergence; no other implications hold in general.

- Convergence with probability 1 implies convergence in probability.
- Convergence in mean implies convergence in probability.
- Convergence in probability implies convergence in distribution.

However, the following exercise gives an important converse to the last implication in the summary above, when the limiting variable is a constant. Of course, a constant can be viewed as a random variable defined on any probability space.

5. Suppose that $(X_1, X_2, \ldots)$ is a sequence of random variables (defined on the same probability space) and that the distribution of X_n converges to the distribution of the constant c as $n \to \infty$. Show that $X_n \to c$ as $n \to \infty$ in probability:

a. $\mathbb{P}(X_n \le x) \to \begin{cases} 0, & x < c \\ 1, & x > c \end{cases}$ as $n \to \infty$
b. $\mathbb{P}(|X_n - c| \le \varepsilon) \to 1$ as $n \to \infty$ for any $\varepsilon > 0$

Examples and Applications

There are several important cases where a special distribution converges to another special distribution as a parameter approaches a limiting value. Indeed, such convergence results are part of the reason why such distributions are *special* in the first place.

Convergence of the Hypergeometric Distribution to the Binomial

Recall that the hypergeometric distribution with parameters m, r, and n is the distribution that governs the number of type 1 objects in a sample of size n, drawn without replacement from a population of m objects with r of type 1. It has discrete probability density function

$$f(k) = \binom{n}{k} \frac{r^{(k)} \, (m - r)^{(n-k)}}{m^{(n)}}, \quad k \in \{0, 1, \ldots, n\}$$

6. Suppose that r_m depends on m, and that $\frac{r_m}{m} \to p$ as $m \to \infty$. Show that for fixed n, the hypergeometric distribution with parameters m, r_m, and n converges to the binomial distribution with parameters n and p as $m \to \infty$.

From a practical point of view, the result in the last exercise means that if the population size m is “large” compared to sample size n, then the hypergeometric distribution with parameters m, r, and n (which corresponds to sampling without replacement) is well approximated by the binomial distribution with parameters n and $p = \frac{r}{m}$ (which corresponds to sampling with replacement). This is often a useful result, because the binomial distribution has fewer parameters than the hypergeometric distribution (and often in real problems, the parameters may only be known approximately). Specifically, in the limiting binomial distribution, we do not need to know the population size m and the number of type 1 objects r

individually, but only in the *ratio* $\frac{r}{m}$.

7. In the ball and urn experiment, set $m = 100$ and $r = 30$ For each of the following values of n (the sample size), switch between *sampling without replacement* (the hypergeometric distribution) and *sampling with replacement* (the binomial distribution). Note the difference in the density function. Run the simulation 1000 times with an update frequency of 10 for each sampling mode.

a. 10
b. 20
c. 30
d. 40
e. 50

Convergence of the Binomial Distribution to the Poisson

Recall that the binomial distribution with parameters $n \in \mathbb{N}_+$ and $p \in [0, 1]$ is the distribution of the number successes in n Bernoulli trials, when p is the probability of success on a trial. This distribution has probability density function

$$f(k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k \in \{0, 1, \ldots, n\}$$

Recall also that the Poisson distribution with rate parameter $r > 0$ has probability density function;

$$g(k) = e^{-r} \frac{r^k}{k!}, \quad k \in \mathbb{N}$$

8. Suppose now that the success parameter p in the binomial distribution depends on the trial parameter n and that $n\, p_n \to r$ as $n \to \infty$ where $r > 0$. Show that this binomial distribution converges to the Poisson distribution with parameter r as $n \to \infty$.

From a practical point of view, the result in the last exercise means that if the number of trials n is "large" and the probability of success p "small", so that the product $r = n\, p$ is of moderate size, then the binomial distribution with parameters n and p is well approximated by the Poisson distribution with parameter r. This is often a useful result, because the Poisson distribution has fewer parameters than the binomial distribution (and often in real problems, the parameters may only be known approximately). Specifically, in the limiting Poisson distribution, we do not need to know the number of trials n and the probability of success p *individually*, but only in the *product* $n\, p$.

The Geometric Distribution

9. Suppose that U is a random variable with probability density function $\mathbb{P}(U = k) = p\,(1 - p)^{k-1}, \ k \in \mathbb{N}_+$, where $p \in (0, 1]$ is a parameter. Thus, U has the geometric distribution on $\mathbb{N}_+$ with parameter p. Random variable U can be interpreted as the trial number of the first success in a sequence of Bernoulli trials.

a. Find the conditional density function of U given $U \le n$.
b. Show that the distribution in (a) converges to the uniform distribution on $\{1, 2, \ldots, n\}$ as $p \downarrow 0$.

10. Suppose that U_n has the geometric distribution on $\mathbb{N}_+$ with success parameter p_n. Moreover, suppose that $n\,p_n \to r$ as $n \to \infty$ where $r > 0$. Show that the distribution of $\frac{U_n}{n}$ converges to the exponential distribution with parameter r as $n \to \infty$.

Note that the limiting condition on n and p in the last exercise is precisely the same as the condition for the convergence of the binomial distribution to the Poisson discussed above. For a deeper interpretation of both of these results, see the section on the Bernoulli trials and the Poisson process.

Convergence of the Matching Distribution to the Poisson

Consider a random permutation $(X_1, X_2, \ldots, X_n)$ of the elements in the set $\{1, 2, \ldots, n\}$. We say that a **match** occurs at position i if $X_i = i$.

11. Show that $\mathbb{P}(X_i = i) = \frac{1}{n}$ for each $i \in \{1, 2, \ldots, n\}$. Thus, the matching events all have the same probability, which varies inversely with the number of trials.

12. Show, however, that $\mathbb{P}\left(X_i = i,\ X_j = j\right) = \frac{1}{n\,(n-1)}$ for each $i \in \{1, 2, \ldots, n\}$, $j \in \{1, 2, \ldots, n\}$ with $i \neq j$. Thus, the matching events are dependent, and in fact are positively correlated. In particular, the matching events do not form a sequence of Bernoulli trials.

The matching problem is studied in detail in the chapter on Finite Sampling Models. In particular, the number of matches N has the following density function

$$\mathbb{P}(N = k) = \frac{1}{k!} \sum_{j=0}^{n-k} \frac{(-1)^j}{j!}, \quad k \in \{0, 1, \ldots, n\}$$

13. Show that the distribution of N converges to the Poisson distribution with parameter 1 as $n \to \infty$.

The Extreme Value Distribution

Suppose that $(X_1, X_2, \ldots)$ is a sequence of independent random variables, each with the standard exponential distribution. Thus, the common distribution function is

$$G(x) = 1 - e^{-x}, \quad x \geq 0$$

14. Find the distribution function of

a. $V_n = \max\{X_1, X_2, \ldots, X_n\}$
b. $Y_n = V_n - \ln(n)$

15. Show that the distribution of Y_n converges to the distribution with the following distribution function as $n \to \infty$:

$$F(x) = e^{-e^{-x}}$$

The limiting distribution in the last exercise is the **type 1 extreme value distribution**, also known as the **Gumbel distribution** in honor of Emil Gumbel. Extreme value distributions are studied in detail in the chapter on Special Distributions.

The Pareto Distribution

16. Suppose that X_n takes values in $[1, \infty)$ for $n \in \mathbb{N}_+$ with distribution function $F_n(x) = 1 - \frac{1}{x^n}$, $x \ge 1$. Thus, X_n has the **Pareto distribution** with shape parameter n, named for Vilfredo Pareto. The Pareto distribution is studied in more detail in the chapter on Special Distributions.

a. Find the limiting distribution of X_n as $n \to \infty$.
b. Find the limiting distribution of $Y_n = n X_n - n$ as $n \to \infty$.

Fundamental Theorems

The two fundamental theorems of basic probability theory, the law of large numbers and the central limit theorem, are studied in detail in the chapter on Random Samples. For this reason we will simply state the results in this section.

Suppose that $(X_1, X_2, \ldots)$ is a sequence of independent, identically distributed, real-valued random variables (defined on the same probability space) with mean μ and standard deviation σ. Let

$$Y_n = \sum_{i=1}^n X_i$$

denote the sum of the first n variables. A weak version of the law of large numbers states that the distribution of the average $M_n = \frac{1}{n} Y_n$ converges to the point mass distribution at μ as $n \to \infty$. From Exercise 5, the convergence is also in probability. In fact the convergence is with probability 1 (much stronger), assuming that σ is finite. The central limit theorem states that the distribution of the standard score

$$Z_n = \frac{Y_n - n\mu}{\sqrt{n}\,\sigma}$$

converges to the standard normal distribution as $n \to \infty$.

Advanced Topics

The Skorohod Representation

Suppose that $(F_1, F_2, \ldots)$ and F are distribution functions, and that $F_n \to F$ as $n \to \infty$ in the sense of convergence of distribution. In this subsection we will prove the **Skorohod representation theorem**: there exist random variables $(X_1, X_2, \ldots)$ and X (defined on the same probability space) such that

- X_n has distribution function F_n for each $n \in \mathbb{N}_+$.
- X has distribution F,
- $X_n \to X$ as $n \to \infty$ with probability 1.

17. Prove the Skorohod representation theorem using the following steps:

a. Let U be uniformly distributed on the interval $(0, 1)$.

b. Define $X_n = F_n^{-1}(U)$ and $X = F^{-1}(U)$ where F_n^{-1} and F^{-1} are the quantile functions of F_n and F respectively.
c. Recall that X_n has distribution function F_n for each $n \in \mathbb{N}_+$, and X has distribution function F.
d. Let $\varepsilon > 0$ and let $u \in (0, 1)$. Pick a continuity point x of F such that $F^{-1}(u) - \varepsilon < x < F^{-1}(u)$.
e. Show that $F(x) < u$.
f. Show that $F_n(x) < u$ for n sufficiently large.
g. Conclude that $F^{-1}(u) - \varepsilon < x < F_n^{-1}(u)$ for n sufficiently large.
h. Let $n \to \infty$ and $\varepsilon \downarrow 0$ to conclude that $F^{-1}(u) \le \liminf_{n\to\infty} F_n^{-1}(u)$ for any $u \in (0, 1)$.
i. Next, let v satisfy $0 < u < v < 1$ and let $\varepsilon > 0$. Pick a continuity point x of F such that $F^{-1}(v) < x < F^{-1}(v) + \varepsilon$
j. Show that $u < v < F(x)$.
k. Show that $u < F_n(x)$ for n sufficiently large.
l. conclude that $F_n^{-1}(u) \le x < F^{-1}(v) + \varepsilon$ for n sufficiently large,
m. Let $n \to \infty$ and $\varepsilon \downarrow 0$ to conclude that $\limsup_{n\to\infty} F_n^{-1}(u) \le F^{-1}(v)$ for any u, v with $0 < u < v < 1$.
n. Let $v \uparrow u$ to show that $\limsup_{n\to\infty} F_n^{-1}(u) \le F^{-1}(u)$ if u is a point of continuity of F^{-1}.
o. Conclude that $F_n^{-1}(u) \to F^{-1}(u)$ as $n \to \infty$ if u is a point of continuity of F^{-1}.
p. Recall from analysis that since F^{-1} is increasing, the set $D \subseteq (0, 1)$ of discontinuities of F^{-1} is countable.
q. Conclude that $\mathbb{P}(U \in D) = 0$
r. Conclude that $\mathbb{P}(X_n \to X \text{ as } n \to \infty) = 1$

The following important result illustrates the value of the Skorohod representation.

18. Suppose that $(X_1, X_2, \ldots)$ and X are real-valued random variables such that the distribution of X_n converges to the distribution of X as $n \to \infty$. If g is a continuous function from $\mathbb{R}$ into $\mathbb{R}$, then the distribution of $g(X_n)$ converges to the distribution of $g(X)$ as $n \to \infty$.

a. Let $Y_n,\ n \in \mathbb{N}_+$ and Y be random variables, defined on the same probability space, such that Y_n has the same distribution as X_n for each $n \in \mathbb{N}_+$, Y has the same distribution as X, and $Y_n \to Y$ as $n \to \infty$ with probability 1.
b. Argue that $g(Y_n) \to g(Y)$ as $n \to \infty$ with probability 1.
c. Argue that the distribution of $g(Y_n)$ converges to the distribution of $g(Y)$ as $n \to \infty$.
d. Argue that $g(Y_n)$ has the same distribution as $g(X_n)$ and that $g(Y)$ has the same distribution as $g(X)$.

Scheffé's Theorem

The following exercises gives **Scheffé's theorem**, named after Henry Scheffé.

19. Suppose that f_n is a probability density function for a continuous distribution $\mathbb{P}_n$ on $\mathbb{R}$ for each $n \in \mathbb{N}_+$, and that f is a probability density function for a continuous distribution $\mathbb{P}$ on $\mathbb{R}$. Suppose that $f_n(x) \to f(x)$ as $n \to \infty$ for all $x \in \mathbb{R}$, except perhaps on a set of Lebesgue measure 0. Then $\mathbb{P}_n(A) \to \mathbb{P}(A)$ as $n \to \infty$ uniformly in (measurable) $A \subseteq \mathbb{R}$.

a. Use basic properties of integrals to show that $|\mathbb{P}(A) - \mathbb{P}_n(A)| \le \int_\mathbb{R} |f(x) - f_n(x)| dx$
b. Let $g_n = f - f_n$, and let g_n^+ denote the positive part of g_n and g_n^- the negative part of g_n. Show that $g_n^+ \le f$.
c. Show that $g_n^+(x) \to 0$ as $n \to \infty$ for all $x \in \mathbb{R}$ (except perhaps on a set of Lebesgue measure 0).
d. Use part (b) and (c) and the dominated convergence theorem to conclude that $\int_\mathbb{R} g_n^+(x) dx \to 0$ as $n \to \infty$.
e. Show that $\int_\mathbb{R} g_n(x) dx = 0$.
f. Conclude that $\int_\mathbb{R} g_n^+(x) dx = \int_\mathbb{R} g_n^-(x) dx$.
g. Conclude that $\int_\mathbb{R} |g_n(x)| dx = 2 \int_\mathbb{R} g_n^+(x) dx$.
h. Use the results of parts (a), (d), and (g) to finish the proof.

Scheffé's theorem is true if the functions are probability density functions with respect to an arbitrary positive measure on $\mathbb{R}$, not just Lebesgue measure. The proof is essentially the same. In particular, if we use counting measure, we get the version of Scheffé's theorem for discrete distributions.

Generating Functions

Generating functions are studied in the chapter on Expected Value. In part, the importance of generating functions stems from the fact that ordinary (pointwise) convergence of a sequence of generating functions corresponds to the convergence of the distributions in the sense of this section. Often it is easier to show convergence in distribution using generating functions than directly from the definition.