

3.8: Convergence in Distribution

This section is concerned with the convergence of probability distributions, a topic of basic importance in probability theory. Since we will be almost exclusively concerned with the convergences of sequences of various kinds, it's helpful to introduce the notation $\mathbb{N}_+^* = \mathbb{N}_+ \cup \{\infty\} = \{1, 2, \dots\} \cup \{\infty\}$.

Distributions on $(\mathbb{R}, \mathcal{R})$

Definition

We start with the most important and basic setting, the measurable space $(\mathbb{R}, \mathcal{R})$, where \mathbb{R} is the set of real numbers of course, and \mathcal{R} is the Borel σ -algebra of subsets of \mathbb{R} . Recall that if P is a probability measure on $(\mathbb{R}, \mathcal{R})$, then the function $F : \mathbb{R} \rightarrow [0, 1]$ defined by $F(x) = P(-\infty, x]$ for $x \in \mathbb{R}$ is the (cumulative) distribution function of P . Recall also that F completely determines P . Here is the definition for convergence of probability measures in this setting:

Suppose P_n is a probability measure on $(\mathbb{R}, \mathcal{R})$ with distribution function F_n for each $n \in \mathbb{N}_+^*$. Then P_n converges (weakly) to P_∞ as $n \rightarrow \infty$ if $F_n(x) \rightarrow F_\infty(x)$ as $n \rightarrow \infty$ for every $x \in \mathbb{R}$ where F_∞ is continuous. We write $P_n \Rightarrow P_\infty$ as $n \rightarrow \infty$.

Recall that a distribution function F is continuous at $x \in \mathbb{R}$ if and only if $\mathbb{P}(X = x) = 0$, so that x is not an *atom* of the distribution (a point of positive probability). We will see shortly why this condition on F_∞ is appropriate. Of course, a probability measure on $(\mathbb{R}, \mathcal{R})$ is usually associated with a real-valued random variable for some random experiment that is modeled by a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. So to review, Ω is the set of outcomes, \mathcal{F} is the σ -algebra of events, and \mathbb{P} is the probability measure on the sample space (Ω, \mathcal{F}) . If X is a real-valued random variable defined on the probability space, then the *distribution* of X is the probability measure P on $(\mathbb{R}, \mathcal{R})$ defined by $P(A) = \mathbb{P}(X \in A)$ for $A \in \mathcal{R}$, and then of course, the distribution function of X is the function F defined by $F(x) = \mathbb{P}(X \leq x)$ for $x \in \mathbb{R}$. Here is the convergence terminology used in this setting:

Suppose that X_n is a real-valued random variable with distribution P_n for each $n \in \mathbb{N}_+^*$. If $P_n \Rightarrow P_\infty$ as $n \rightarrow \infty$ then we say that X_n converges in distribution to X_∞ as $n \rightarrow \infty$. We write $X_n \rightarrow X_\infty$ as $n \rightarrow \infty$ in distribution.

So if F_n is the distribution function of X_n for $n \in \mathbb{N}_+^*$, then $X_n \rightarrow X_\infty$ as $n \rightarrow \infty$ in distribution if $F_n(x) \rightarrow F_\infty(x)$ at every point $x \in \mathbb{R}$ where F_∞ is continuous. On the one hand, the terminology and notation are helpful, since again most probability measures are associated with random variables (and every probability measure can be). On the other hand, the terminology and notation can be a bit misleading since the random variables, as functions, do not converge in any sense, and indeed the random variables need not be defined on the same probability spaces. It is only the *distributions* that converge. However, often the random variables *are* defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$, in which case we can compare convergence in distribution with the other modes of convergence we have or will study:

- Convergence with probability 1
- Convergence in probability
- Convergence in mean

We will show, in fact, that convergence in distribution is the weakest of all of these modes of convergence. However, strength of convergence should not be confused with importance. Convergence in distribution is one of the most important modes of convergence; 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 probability density functions, and why convergence is only required at the points of continuity of the limiting distribution function. Note that the distributions considered are probability measures on $(\mathbb{R}, \mathcal{R})$, even though the support of the distribution may be a much smaller subset. For the first example, note that if a deterministic sequence converges in the ordinary calculus sense, then naturally we want the sequence (thought of as random variables) to converge in distribution. Expand the proof to understand the example fully.

Suppose that $x_n \in \mathbb{R}$ for $n \in \mathbb{N}_+^*$. Define random variable $X_n = x_n$ with probability 1 for each $n \in \mathbb{N}_+^*$. Then $x_n \rightarrow x_\infty$ as $n \rightarrow \infty$ if and only if $X_n \rightarrow X_\infty$ as $n \rightarrow \infty$ in distribution.

Proof

For $n \in \mathbb{N}_+^*$, the CDF F_n of X_n is given by $F_n(x) = 0$ for $x < x_n$ and $F_n(x) = 1$ for $x \geq x_n$.

1. Suppose that $x_n \rightarrow x_\infty$ as $n \rightarrow \infty$. If $x < x_\infty$ then $x < x_n$, and hence $F_n(x) = 0$, for all but finitely many $n \in \mathbb{N}_+$, and so $F_n(x) \rightarrow 0$ as $n \rightarrow \infty$. If $x > x_\infty$ then $x > x_n$, and hence $F_n(x) = 1$, for all but finitely many $n \in \mathbb{N}_+$, and so $F_n(x) \rightarrow 1$ as $n \rightarrow \infty$. Nothing can be said about the limiting behavior of $F_n(x_\infty)$ as $n \rightarrow \infty$ without more information. For example, if $x_n \leq x_\infty$ for all but finitely many $n \in \mathbb{N}_+$ then $F_n(x_\infty) \rightarrow 1$ as $n \rightarrow \infty$. If $x_n > x_\infty$ for all but finitely many $n \in \mathbb{N}_+$ then $F_n(x_\infty) \rightarrow 0$ as $n \rightarrow \infty$. If $x_n < x_\infty$ for infinitely many $n \in \mathbb{N}_+$ and $x_n > x_\infty$ for infinitely many $n \in \mathbb{N}_+$ then $F_n(x_\infty)$ does not have a limit as $n \rightarrow \infty$. But regardless, we have $F_n(x) \rightarrow F_\infty(x)$ as $n \rightarrow \infty$ for every $x \in \mathbb{R}$ except perhaps x_∞ , the one point of discontinuity of F_∞ . Hence $X_n \rightarrow X_\infty$ as $n \rightarrow \infty$ in distribution.
2. Conversely, suppose that $X_n \rightarrow X_\infty$ as $n \rightarrow \infty$ in distribution. If $x < x_\infty$ then $F_n(x) \rightarrow 0$ as $n \rightarrow \infty$ and hence $x < x_n$ for all but finitely many $n \in \mathbb{N}_+$. If $x > x_\infty$ then $F_n(x) \rightarrow 1$ as $n \rightarrow \infty$ and hence $x \geq x_n$ for all but finitely many $n \in \mathbb{N}_+$. So, for every $\epsilon > 0$, $x_n \in (x_\infty - \epsilon, x_\infty + \epsilon)$ for all but finitely many $n \in \mathbb{N}_+$, and hence $x_n \rightarrow x_\infty$ as $n \rightarrow \infty$.

The proof is finished, but let's look at the probability density functions to see that these are not the proper objects of study. For $n \in \mathbb{N}_+^*$, the PDF f_n of X_n is given by $f_n(x_n) = 1$ and $f_n(x) = 0$ for $x \in \mathbb{R} \setminus \{x_n\}$. Only when $x_n = x_\infty$ for all but finitely many $n \in \mathbb{N}_+$ do we have $f_n(x) \rightarrow f(x)$ for $x \in \mathbb{R}$.

For the example below, recall that \mathbb{Q} denotes the set of rational numbers. Once again, expand the proof to understand the example fully

For $n \in \mathbb{N}_+$, let P_n denote the discrete uniform distribution on $\left\{\frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1\right\}$ and let P_∞ denote the continuous uniform distribution on the interval $[0, 1]$. Then

1. $P_n \Rightarrow P_\infty$ as $n \rightarrow \infty$
2. $P_n(\mathbb{Q}) = 1$ for each $n \in \mathbb{N}_+$ but $P_\infty(\mathbb{Q}) = 0$.

Proof

As usual, let F_n denote the CDF of P_n for $n \in \mathbb{N}_+^*$.

1. For $n \in \mathbb{N}_+$ note that F_n is given by $F_n(x) = \lfloor nx \rfloor / n$ for $x \in [0, 1]$. But $nx - 1 \leq \lfloor nx \rfloor \leq nx$ so $\lfloor nx \rfloor / n \rightarrow x$ as $n \rightarrow \infty$ for $x \in [0, 1]$. Of course, $F_n(x) = 0$ for $x < 0$ and $F_n(x) = 1$ for $x > 1$. So $F_n(x) \rightarrow F_\infty(x)$ as $n \rightarrow \infty$ for all $x \in \mathbb{R}$.
2. Note that by definition, so $P_n(\mathbb{Q}) = 1$ for $n \in \mathbb{N}_+$. On the other hand, P_∞ is a continuous distribution and \mathbb{Q} is countable, so $P_\infty(\mathbb{Q}) = 0$.

The proof is finished, but let's look at the probability density functions. For $n \in \mathbb{N}_+$, the PDF f_n of P_n is given by $f_n(x) = \frac{1}{n}$ for $x \in \left\{\frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1\right\}$ and $f_n(x) = 0$ otherwise. Hence $0 \leq f_n(x) \leq \frac{1}{n}$ for $n \in \mathbb{N}_+$ and $x \in \mathbb{R}$, so $f_n(x) \rightarrow 0$ as $n \rightarrow \infty$ for every $x \in \mathbb{R}$.

The point of the example is that it's reasonable for the discrete uniform distribution on $\left\{\frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, 1\right\}$ to converge to the continuous uniform distribution on $[0, 1]$, but once again, the probability density functions are evidently not the correct objects of study.

Probability Density Functions

As the [previous example](#) 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. Recall again that we are thinking of our probability distributions as measures on $(\mathbb{R}, \mathcal{B})$ even when supported on a smaller subset.

Convergence in distribution in terms of probability density functions.

1. Suppose that f_n is a probability density function for a discrete distribution P_n on a countable set $S \subseteq \mathbb{R}$ for each $n \in \mathbb{N}_+^*$. If $f_n(x) \rightarrow f_\infty(x)$ as $n \rightarrow \infty$ for each $x \in S$ then $P_n \Rightarrow P_\infty$ as $n \rightarrow \infty$.
2. Suppose that f_n is a probability density function for a continuous distribution P_n on \mathbb{R} for each $n \in \mathbb{N}_+^*$. If $f_n(x) \rightarrow f(x)$ as $n \rightarrow \infty$ for all $x \in \mathbb{R}$ (except perhaps on a set with Lebesgue measure 0) then $P_n \Rightarrow P_\infty$ as $n \rightarrow \infty$.

Proof

1. Fix $x \in \mathbb{R}$. Then $P_n(-\infty, x] = \sum_{y \in S, y \leq x} f(y)$ for $n \in \mathbb{N}_+$ and $P(-\infty, x] = \sum_{y \in S, y \leq x} f(y)$. It follows from [Scheffé's theorem](#) with the measure space $(S, \mathcal{P}(S), \#)$ that $P_n(-\infty, x] \rightarrow P(-\infty, x]$ as $n \rightarrow \infty$.
2. Fix $x \in \mathbb{R}$. Then $P_n(-\infty, x] = \int_{-\infty}^x f(y) dy$ for $n \in \mathbb{N}_+$ and $P(-\infty, x] = \int_{-\infty}^x f(y) dy$. It follows from [Scheffé's theorem](#) with the measure space $(\mathbb{R}, \mathcal{B}, \lambda)$ that $P_n(-\infty, x] \rightarrow P(-\infty, x]$ as $n \rightarrow \infty$.

Convergence in Probability

Naturally, we would like to compare convergence in distribution with other modes of convergence we have studied.

Suppose that X_n is a real-valued random variable for each $n \in \mathbb{N}_+^*$, all defined on the same probability space. If $X_n \rightarrow X_\infty$ as $n \rightarrow \infty$ in probability then $X_n \rightarrow X_\infty$ as $n \rightarrow \infty$ in distribution.

Proof

Let F_n denote the distribution function of X_n for $n \in \mathbb{N}_+^*$. Fix $\epsilon > 0$. Note first that $\mathbb{P}(X_n \leq x) = \mathbb{P}(X_n \leq x, X_\infty \leq x + \epsilon) + \mathbb{P}(X_n \leq x, X_\infty > x + \epsilon)$. Hence $F_n(x) \leq F_\infty(x + \epsilon) + \mathbb{P}(|X_n - X_\infty| > \epsilon)$. Next, note that $\mathbb{P}(X_\infty \leq x - \epsilon) = \mathbb{P}(X_\infty \leq x - \epsilon, X_n \leq x) + \mathbb{P}(X_\infty \leq x - \epsilon, X_n > x)$. Hence $F_\infty(x - \epsilon) \leq F_n(x) + \mathbb{P}(|X_n - X_\infty| > \epsilon)$. From the last two results it follows that

$$F_\infty(x - \epsilon) - \mathbb{P}(|X_n - X_\infty| > \epsilon) \leq F_n(x) \leq F_\infty(x + \epsilon) + \mathbb{P}(|X_n - X_\infty| > \epsilon) \quad (3.8.1)$$

Letting $n \rightarrow \infty$ and using convergence in probability gives

$$F_\infty(x - \epsilon) \leq \liminf_{n \rightarrow \infty} F_n(x) \leq \limsup_{n \rightarrow \infty} F_n(x) \leq F_\infty(x + \epsilon) \quad (3.8.2)$$

Finally, letting $\epsilon \downarrow 0$ we see that if F_∞ is continuous at x then $F_n(x) \rightarrow F_\infty(x)$ as $n \rightarrow \infty$.

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.

Let X be an indicator variable with $\mathbb{P}(X = 0) = \mathbb{P}(X = 1) = \frac{1}{2}$, so that X is the result of tossing a fair coin. Let $X_n = 1 - X$ for $n \in \mathbb{N}_+$. Then

1. $X_n \rightarrow X$ as $n \rightarrow \infty$ in distribution.
2. $\mathbb{P}(X_n \text{ does not converge to } X \text{ as } n \rightarrow \infty) = 1$.
3. X_n does not converge to X as $n \rightarrow \infty$ in probability.
4. X_n does not converge to X as $n \rightarrow \infty$ in mean.

Proof

1. This trivially holds since $1 - X$ has the same distribution as X .
2. This follows since $|X_n - X| = 1$ for every $n \in \mathbb{N}_+$.
3. This follows since $\mathbb{P}(|X_n - X| > \frac{1}{2}) = 1$ for each $n \in \mathbb{N}_+$.
4. This follows since $\mathbb{E}(|X_n - X|) = 1$ for each $n \in \mathbb{N}_+$.

The critical fact that makes this counterexample work is that $1 - X$ has the same distribution as X . Any random variable with this property would work just as well, so if you prefer a counterexample with continuous distributions, let X have probability density function f given by $f(x) = 6x(1 - x)$ for $0 \leq x \leq 1$. The distribution of X is an example of a beta distribution.

The following summary gives the implications for the various modes of convergence; no other implications hold in general.

Suppose that X_n is a real-valued random variable for each $n \in \mathbb{N}_+^*$, all defined on a common probability space.

1. If $X_n \rightarrow X_\infty$ as $n \rightarrow \infty$ with probability 1 then $X_n \rightarrow X_\infty$ as $n \rightarrow \infty$ in probability.
2. If $X_n \rightarrow X_\infty$ as $n \rightarrow \infty$ in mean then $X_n \rightarrow X_\infty$ as $n \rightarrow \infty$ in probability.

3. If $X_n \rightarrow X_\infty$ as $n \rightarrow \infty$ in probability then $X_n \rightarrow X_\infty$ as $n \rightarrow \infty$ in distribution.

It follows that convergence with probability 1, convergence in probability, and convergence in mean all imply convergence in distribution, so the latter mode of convergence is indeed the weakest. However, our next theorem gives an important converse to part (c) in (7), when the limiting variable is a constant. Of course, a constant can be viewed as a random variable defined on any probability space.

Suppose that X_n is a real-valued random variable for each $n \in \mathbb{N}_+$, defined on the same probability space, and that $c \in \mathbb{R}$. If $X_n \rightarrow c$ as $n \rightarrow \infty$ in distribution then $X_n \rightarrow c$ as $n \rightarrow \infty$ in probability.

Proof

Assume that the probability space is $(\Omega, \mathcal{F}, \mathbb{P})$. Note first that $\mathbb{P}(X_n \leq x) \rightarrow 0$ as $n \rightarrow \infty$ if $x < c$ and $\mathbb{P}(X_n \leq x) \rightarrow 1$ as $n \rightarrow \infty$ if $x > c$. It follows that $\mathbb{P}(|X_n - c| \leq \epsilon) \rightarrow 1$ as $n \rightarrow \infty$ for every $\epsilon > 0$.

The Skorohod Representation

As noted in the [summary above](#), convergence in distribution does not imply convergence with probability 1, even when the random variables are defined on the same probability space. However, the next theorem, known as the *Skorohod representation theorem*, gives an important partial result in this direction.

Suppose that P_n is a probability measure on $(\mathbb{R}, \mathcal{B})$ for each $n \in \mathbb{N}_+^*$ and that $P_n \Rightarrow P_\infty$ as $n \rightarrow \infty$. Then there exist real-valued random variables X_n for $n \in \mathbb{N}_+^*$, defined on the same probability space, such that

1. X_n has distribution P_n for $n \in \mathbb{N}_+^*$.
2. $X_n \rightarrow X_\infty$ as $n \rightarrow \infty$ with probability 1.

Proof

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and U a random variable defined on this space that is uniformly distributed on the interval $(0, 1)$. For a specific construction, we could take $\Omega = (0, 1)$, \mathcal{F} the σ -algebra of Borel measurable subsets of $(0, 1)$, and \mathbb{P} Lebesgue measure on (Ω, \mathcal{F}) (the uniform distribution on $(0, 1)$). Then let U be the identity function on Ω so that $U(\omega) = \omega$ for $\omega \in \Omega$, so that U has probability distribution \mathbb{P} . We have seen this construction many times before.

1. For $n \in \mathbb{N}_+^*$, let F_n denote the distribution function of P_n and define $X_n = F_n^{-1}(U)$ where F_n^{-1} is the quantile functions of F_n . Recall that X_n has distribution function F_n and therefore X_n has distribution P_n for $n \in \mathbb{N}_+^*$. Of course, these random variables are also defined on $(\Omega, \mathcal{F}, \mathbb{P})$.
2. Let $\epsilon > 0$ and let $u \in (0, 1)$. Pick a continuity point x of F_∞ such that $F_\infty^{-1}(u) - \epsilon < x < F_\infty^{-1}(u)$. Then $F_\infty(x) < u$ and hence $F_n(x) < u$ for all but finitely many $n \in \mathbb{N}_+$. It follows that $F_\infty^{-1}(u) - \epsilon < x < F_n^{-1}(u)$ for all but finitely many $n \in \mathbb{N}_+$. Let $n \rightarrow \infty$ and $u \downarrow 0$ to conclude that $F_\infty^{-1}(u) \leq \liminf_{n \rightarrow \infty} F_n^{-1}(u)$. Next, let v satisfy $0 < u < v < 1$ and let $\epsilon > 0$. Pick a continuity point x of F_∞ such that $F_\infty^{-1}(v) < x < F_\infty^{-1}(v) + \epsilon$. Then $u < v < F_\infty(x)$ and hence $u < F_n(x)$ for all but finitely many $n \in \mathbb{N}_+$. It follows that $F_n^{-1}(u) \leq x < F_\infty^{-1}(v) + \epsilon$ for all but finitely many $n \in \mathbb{N}_+$. Let $n \rightarrow \infty$ and $\epsilon \downarrow 0$ to conclude that $\limsup_{n \rightarrow \infty} F_n^{-1}(u) \leq F_\infty^{-1}(v)$. Letting $v \downarrow u$ it follows that $\limsup_{n \rightarrow \infty} F_n^{-1}(u) \leq F_\infty^{-1}(u)$ if u is a point of continuity of F_∞^{-1} . Therefore $F_n^{-1}(u) \rightarrow F_\infty^{-1}(u)$ as $n \rightarrow \infty$ if u is a point of continuity of F_∞^{-1} . Recall from analysis that since $F_\infty^{-1}(u)$ is increasing, the set $D \subseteq (0, 1)$ of discontinuities of F_∞^{-1} is countable. Since U has a continuous distribution, $\mathbb{P}(U \in D) = 0$. Finally, it follows that $\mathbb{P}(X_n \rightarrow X_\infty \text{ as } n \rightarrow \infty) = 1$.

The following theorem illustrates the value of the Skorohod representation and the usefulness of random variable notation for convergence in distribution. The theorem is also quite intuitive, since a basic idea is that *continuity* should preserve *convergence*.

Suppose that X_n is a real-valued random variable for each $n \in \mathbb{N}_+^*$ (not necessarily defined on the same probability space). Suppose also that $g: \mathbb{R} \rightarrow \mathbb{R}$ is measurable, and let D_g denote the set of discontinuities of g , and P_∞ the distribution of X_∞ . If $X_n \rightarrow X_\infty$ as $n \rightarrow \infty$ in distribution and $P_\infty(D_g) = 0$, then $g(X_n) \rightarrow g(X_\infty)$ as $n \rightarrow \infty$ in distribution.

Proof

By Skorohod's theorem, there exists random variables Y_n for $n \in \mathbb{N}_+^*$, defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$, such that Y_n has the same distribution as X_n for $n \in \mathbb{N}_+^*$, and $Y_n \rightarrow Y_\infty$ as $n \rightarrow \infty$ with probability 1. Since

$\mathbb{P}(Y_\infty \in D_g) = P_\infty(D_g) = 0$ it follows that $g(Y_n) \rightarrow g(Y_\infty)$ as $n \rightarrow \infty$ with probability 1. Hence by the [theorem above](#), $g(Y_n) \rightarrow g(Y_\infty)$ as $n \rightarrow \infty$ in distribution. But $g(Y_n)$ has the same distribution as $g(X_n)$ for each $n \in \mathbb{N}_+^*$.

As a simple corollary, if X_n converges X_∞ as $n \rightarrow \infty$ in distribution, and if $a, b \in \mathbb{R}$ then $a + bX_n$ converges to $a + bX$ as $n \rightarrow \infty$ in distribution. But we can do a little better:

Suppose that X_n is a real-valued random variable and that $a_n, b_n \in \mathbb{R}$ for each $n \in \mathbb{N}_+^*$. If $X_n \rightarrow X_\infty$ as $n \rightarrow \infty$ in distribution and if $a_n \rightarrow a_\infty$ and $b_n \rightarrow b_\infty$ as $n \rightarrow \infty$, then $a_n + b_n X_n \rightarrow a + bX_\infty$ as $n \rightarrow \infty$ in distribution.

Proof

Again by [Skorohod's theorem](#), there exist random variables Y_n for $n \in \mathbb{N}_+^*$, defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that Y_n has the same distribution as X_n for $n \in \mathbb{N}_+^*$ and $Y_n \rightarrow Y_\infty$ as $n \rightarrow \infty$ with probability 1. Hence also $a_n + b_n Y_n \rightarrow a_\infty + b_\infty Y_\infty$ as $n \rightarrow \infty$ with probability 1. By the [result above](#), $a_n + b_n Y_n \rightarrow a_\infty + b_\infty Y_\infty$ as $n \rightarrow \infty$ in distribution. But $a_n + b_n Y_n$ has the same distribution as $a_n + b_n X_n$ for $n \in \mathbb{N}_+^*$.

The definition of convergence in distribution requires that the sequence of probability measures converge on sets of the form $(-\infty, x]$ for $x \in \mathbb{R}$ when the limiting distribution has probability 0 at x . It turns out that the probability measures will converge on lots of other sets as well, and this result points the way to extending convergence in distribution to more general spaces. To state the result, recall that if A is a subset of a topological space, then the *boundary* of A is $\partial A = \text{cl}(A) \setminus \text{int}(A)$ where $\text{cl}(A)$ is the *closure* of A (the smallest closed set that contains A) and $\text{int}(A)$ is the *interior* of A (the largest open set contained in A).

Suppose that P_n is a probability measure on $(\mathbb{R}, \mathcal{R})$ for $n \in \mathbb{N}_+^*$. Then $P_n \Rightarrow P_\infty$ as $n \rightarrow \infty$ if and only if $P_n(A) \rightarrow P_\infty(A)$ as $n \rightarrow \infty$ for every $A \in \mathcal{R}$ with $P(\partial A) = 0$.

Proof

Suppose that $P_n \Rightarrow P_\infty$ as $n \rightarrow \infty$. Let X_n be a random variable with distribution P_n for $n \in \mathbb{N}_+^*$. (We don't care about the underlying probability spaces.) If $A \in \mathcal{R}$ then the set of discontinuities of $\mathbf{1}_A$, the indicator function of A , is ∂A . So, suppose $\mathbb{P}_\infty(\partial A) = 0$. By the [continuity theorem above](#), $\mathbf{1}_A(X_n) \rightarrow \mathbf{1}_A(X_\infty)$ as $n \rightarrow \infty$ in distribution. Let G_n denote the CDF of $\mathbf{1}_A(X_n)$ for $n \in \mathbb{N}_+^*$. The only possible points of discontinuity of G_∞ are 0 and 1. Hence $G_n(\frac{1}{2}) \rightarrow G_\infty(\frac{1}{2})$ as $n \rightarrow \infty$. But $G_n(\frac{1}{2}) = P_n(A^c)$ for $n \in \mathbb{N}_+^*$. Hence $P_n(A^c) \rightarrow \mathbb{P}_\infty(A^c)$ and so also $P_n(A) \rightarrow P_\infty(A)$ as $n \rightarrow \infty$.

Conversely, suppose that the condition in the theorem holds. If $x \in \mathbb{R}$, then the boundary of $(-\infty, x]$ is $\{x\}$, so if $P_\infty\{x\} = 0$ then $P_n(-\infty, x] \rightarrow P_\infty(-\infty, x]$ as $n \rightarrow \infty$. So by definition, $P_n \Rightarrow P_\infty$ as $n \rightarrow \infty$.

In the context of this result, suppose that $a, b \in \mathbb{R}$ with $a < b$. If $P\{a\} = P\{b\} = 0$, then as $n \rightarrow \infty$ we have $P_n(a, b) \rightarrow P(a, b)$, $P_n[a, b) \rightarrow P[a, b)$, $P_n(a, b] \rightarrow P(a, b]$, and $P_n[a, b] \rightarrow P[a, b]$. Of course, the limiting values are all the same.

Examples and Applications

Next we will explore several interesting examples of the convergence of distributions on $(\mathbb{R}, \mathcal{R})$. 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.

The Hypergeometric Distribution

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 objects of type 1. It has discrete probability density function f given by

$$f(k) = \frac{\binom{r}{k} \binom{m-r}{n-k}}{\binom{m}{n}}, \quad k \in \{0, 1, \dots, n\} \quad (3.8.3)$$

The parameters m , r , and n are positive integers with $n \leq m$ and $r \leq m$. The hypergeometric distribution is studied in more detail in the chapter on Finite Sampling Models

Recall next that *Bernoulli trials* are independent trials, each with two possible outcomes, generically called *success* and *failure*. The probability of success $p \in [0, 1]$ is the same for each trial. The *binomial distribution* with parameters $n \in \mathbb{N}_+$ and p is the distribution of the number successes in n Bernoulli trials. This distribution has probability density function g given by

$$g(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k \in \{0, 1, \dots, n\} \quad (3.8.4)$$

The binomial distribution is studied in more detail in the chapter on Bernoulli Trials. Note that the binomial distribution with parameters n and $p = r/m$ is the distribution that governs the number of type 1 objects in a sample of size n , drawn *with replacement* from a population of m objects with r objects of type 1. This fact is motivation for the following result:

Suppose that $r_m \in \{0, 1, \dots, m\}$ for each $m \in \mathbb{N}_+$ and that $r_m/m \rightarrow p$ as $m \rightarrow \infty$. For fixed $n \in \mathbb{N}_+$, the hypergeometric distribution with parameters m , r_m , and n converges to the binomial distribution with parameters n and p as $m \rightarrow \infty$.

Proof

Recall that for $a \in \mathbb{R}$ and $j \in \mathbb{N}$, we let $a^{(j)} = a(a-1) \cdots [a-(j-1)]$ denote the falling power of a of order j . The hypergeometric PDF can be written as

$$f_m(k) = \binom{n}{k} \frac{r_m^{(k)} (m-r_m)^{(n-k)}}{m^{(n)}}, \quad k \in \{0, 1, \dots, n\} \quad (3.8.5)$$

In the fraction above, the numerator and denominator both have n factors. Suppose that we group the k factors in $r_m^{(k)}$ with the first k factors of $m^{(n)}$ and the $n-k$ factors of $(m-r_m)^{(n-k)}$ with the last $n-k$ factors of $m^{(n)}$ to form a product of n fractions. The first k fractions have the form $(r_m - j)/(m - j)$ for some j that does not depend on m . Each of these converges to p as $m \rightarrow \infty$. The last $n-k$ fractions have the form $(m - r_m - j)/(m - k - j)$ for some j that does not depend on m . Each of these converges to $1-p$ as $m \rightarrow \infty$. Hence

$$f_m(k) \rightarrow \binom{n}{k} p^k (1-p)^{n-k} \text{ as } m \rightarrow \infty \text{ for each } k \in \{0, 1, \dots, n\} \quad (3.8.6)$$

The result now follows from the [theorem above](#) on density functions.

From a practical point of view, the last result 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 $p = r/m$ (which corresponds to sampling with replacement). This is often a useful result, not computationally, but rather 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* r/m .

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 probability density functions. Run the simulation 1000 times for each sampling mode and compare the relative frequency function to the probability density function.

1. 10
2. 20
3. 30
4. 40
5. 50

The Binomial Distribution

Recall again 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 given by

$$f(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k \in \{0, 1, \dots, n\} \quad (3.8.7)$$

Recall also that the *Poisson distribution* with parameter $r \in (0, \infty)$ has probability density function g given by

$$g(k) = e^{-r} \frac{r^k}{k!}, \quad k \in \mathbb{N} \quad (3.8.8)$$

The distribution is named for Simeon Poisson and governs the number of “random points” in a region of time or space, under certain ideal conditions. The parameter r is proportional to the size of the region of time or space. The Poisson distribution is studied in more detail in the chapter on the Poisson Process.

Suppose that $p_n \in [0, 1]$ for $n \in \mathbb{N}_+$ and that $np_n \rightarrow r \in (0, \infty)$ as $n \rightarrow \infty$. Then the binomial distribution with parameters n and p_n converges to the Poisson distribution with parameter r as $n \rightarrow \infty$.

Proof

For $k, n \in \mathbb{N}$ with $k \leq n$, the binomial PDF can be written as

$$f_n(k) = \frac{n^{(k)}}{k!} p_n^k (1 - p_n)^{n-k} = \frac{1}{k!} (np_n) [(n-1)p_n] \cdots [(n-k+1)p_n] (1 - p_n)^{n-k} \quad (3.8.9)$$

First, $(n-j)p_n \rightarrow r$ as $n \rightarrow \infty$ for $j \in \{0, 1, \dots, n-1\}$. Next, by a famous limit from calculus, $(1 - p_n)^n = (1 - np_n/n)^n \rightarrow e^{-r}$ as $n \rightarrow \infty$. Hence also $(1 - p_n)^{n-k} \rightarrow e^{-r}$ as $n \rightarrow \infty$ for fixed $k \in \mathbb{N}_+$. Therefore $f_n(k) \rightarrow e^{-r} r^k / k!$ as $n \rightarrow \infty$ for each $k \in \mathbb{N}_+$. The result now follows from the [theorem above](#) on density functions.

From a practical point of view, the convergence of the binomial distribution to the Poisson means that if the number of trials n is “large” and the probability of success p “small”, so that np^2 is small, then the binomial distribution with parameters n and p is well approximated by the Poisson distribution with parameter $r = np$. This is often a useful result, again not computationally, but rather 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 approximating 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* np . As we will see in the next chapter, the condition that np^2 be small means that the variance of the binomial distribution, namely $np(1 - p) = np - np^2$ is approximately $r = np$, which is the variance of the approximating Poisson distribution.

In the binomial timeline experiment, set the parameter values as follows, and observe the graph of the probability density function. (Note that $np = 5$ in each case.) Run the experiment 1000 times in each case and compare the relative frequency function and the probability density function. Note also the successes represented as “random points” in discrete time.

1. $n = 10, p = 0.5$
2. $n = 20, p = 0.25$
3. $n = 100, p = 0.05$

In the Poisson experiment, set $r = 5$ and $t = 1$, to get the Poisson distribution with parameter 5. Note the shape of the probability density function. Run the experiment 1000 times and compare the relative frequency function to the probability density function. Note the similarity between this experiment and the one in the previous exercise.

The Geometric Distribution

Recall that the *geometric distribution* on \mathbb{N}_+ with success parameter $p \in (0, 1]$ has probability density function f given by

$$f(k) = p(1 - p)^{k-1}, \quad k \in \mathbb{N}_+ \quad (3.8.10)$$

The geometric distribution governs the trial number of the first success in a sequence of Bernoulli trials.

Suppose that U has the geometric distribution on \mathbb{N}_+ with success parameter $p \in (0, 1]$. For $n \in \mathbb{N}_+$, the conditional distribution of U given $U \leq n$ converges to the uniform distribution on $\{1, 2, \dots, n\}$ as $p \downarrow 0$.

Proof

The CDF F of U is given by $F(k) = 1 - (1 - p)^k$ for $k \in \mathbb{N}_+$. Hence for $n \in \mathbb{N}_+$, the conditional CDF of U given $U \leq n$ is

$$F_n(k) = \mathbb{P}(U \leq k | U \leq n) = \frac{\mathbb{P}(U \leq k)}{\mathbb{P}(U \leq n)} = \frac{1 - (1-p)^k}{1 - (1-p)^n}, \quad k \in \{1, 2, \dots, n\} \quad (3.8.11)$$

Using L'Hospital's rule, gives $F_n(k) \rightarrow k/n$ as $p \downarrow 0$ for $k \in \{1, 2, \dots, n\}$. As a function of k this is the CDF of the uniform distribution on $\{1, 2, \dots, n\}$.

Next, recall that the *exponential distribution* with rate parameter $r \in (0, \infty)$ has distribution function G given by

$$G(t) = 1 - e^{-rt}, \quad 0 \leq t < \infty \quad (3.8.12)$$

The exponential distribution governs the time between “arrivals” in the Poisson model of random points in time.

Suppose that U_n has the geometric distribution on \mathbb{N}_+ with success parameter $p_n \in (0, 1]$ for $n \in \mathbb{N}_+$, and that $np_n \rightarrow r \in (0, \infty)$ as $n \rightarrow \infty$. The distribution of U_n/n converges to the exponential distribution with parameter r as $n \rightarrow \infty$.

Proof

Let F_n denote the CDF of U_n/n . Then for $x \in [0, \infty)$

$$F_n(x) = \mathbb{P}\left(\frac{U_n}{n} \leq x\right) = \mathbb{P}(U_n \leq nx) = \mathbb{P}(U_n \leq \lfloor nx \rfloor) = 1 - (1-p_n)^{\lfloor nx \rfloor} \quad (3.8.13)$$

We showed in the proof of the [convergence of the binomial distribution](#) that $(1-p_n)^n \rightarrow e^{-r}$ as $n \rightarrow \infty$, and hence $(1-p_n)^{nx} \rightarrow e^{-rx}$ as $n \rightarrow \infty$. But by definition, $\lfloor nx \rfloor \leq nx < \lfloor nx \rfloor + 1$ or equivalently, $nx - 1 < \lfloor nx \rfloor \leq nx$ so it follows from the squeeze theorem that $(1-p_n)^{\lfloor nx \rfloor} \rightarrow e^{-rx}$ as $n \rightarrow \infty$. Hence $F_n(x) \rightarrow 1 - e^{-rx}$ as $n \rightarrow \infty$. As a function of $x \in [0, \infty)$, this is the CDF of the exponential distribution with parameter r .

Note that the limiting condition on n and p in the last result is precisely the same as the condition for the [convergence of the binomial distribution](#) to the Poisson distribution. For a deeper interpretation of both of these results, see the section on the Poisson distribution.

In the negative binomial experiment, set $k = 1$ to get the geometric distribution. Then decrease the value of p and note the shape of the probability density function. With $p = 0.5$ run the experiment 1000 times and compare the relative frequency function to the probability density function.

In the gamma experiment, set $k = 1$ to get the exponential distribution, and set $r = 5$. Note the shape of the probability density function. Run the experiment 1000 times and compare the empirical density function and the probability density function. Compare this experiment with the one in the previous exercise, and note the similarity, up to a change in scale.

The Matching Distribution

For $n \in \mathbb{N}_+$, consider a random permutation (X_1, X_2, \dots, X_n) of the elements in the set $\{1, 2, \dots, n\}$. We say that a *match* occurs at position i if $X_i = i$.

$$\mathbb{P}(X_i = i) = \frac{1}{n} \text{ for each } i \in \{1, 2, \dots, n\}.$$

Proof

The number of permutations of $\{1, 2, \dots, n\}$ is $n!$. For $i \in \{1, 2, \dots, n\}$, the number of such permutations with i in position i is $(n-1)!$. Hence $\mathbb{P}(X_i = i) = (n-1)!/n! = 1/n$. A more direct argument is that i is no more or less likely to end up in position i as any other number.

So the matching events all have the same probability, which varies inversely with the number of trials.

$$\mathbb{P}(X_i = i, X_j = j) = \frac{1}{n(n-1)} \text{ for } i, j \in \{1, 2, \dots, n\} \text{ with } i \neq j.$$

Proof

Again, the number of permutations of $\{1, 2, \dots, n\}$ is $n!$. For distinct $i, j \in \{1, 2, \dots, n\}$, the number of such permutations with i in position i and j in position j is $(n-2)!$. Hence $\mathbb{P}(X_i = i, X_j = j) = (n-2)!/n! = 1/n(n-1)$.

So 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 that section we show that the number of matches N_n has probability density function f_n given by:

$$f_n(k) = \frac{1}{k!} \sum_{j=0}^{n-k} \frac{(-1)^j}{j!}, \quad k \in \{0, 1, \dots, n\} \quad (3.8.14)$$

The distribution of N_n converges to the Poisson distribution with parameter 1 as $n \rightarrow \infty$.

Proof

For $k \in \mathbb{N}$,

$$f_n(k) = \frac{1}{k!} \sum_{j=0}^{n-k} \frac{(-1)^j}{j!} \rightarrow \frac{1}{k!} \sum_{j=0}^{\infty} \frac{(-1)^j}{j!} = \frac{1}{k!} e^{-1} \quad (3.8.15)$$

As a function of $k \in \mathbb{N}$, this is the PDF of the Poisson distribution with parameter 1. So the result follows from the [theorem above](#) on density functions.

In the matching experiment, increase n and note the apparent convergence of the probability density function for the number of matches. With selected values of n , run the experiment 1000 times and compare the relative frequency function and the probability density function.

The Extreme Value Distribution

Suppose that (X_1, X_2, \dots) is a sequence of independent random variables, each with the standard exponential distribution (parameter 1). Thus, recall that the common distribution function G is given by

$$G(x) = 1 - e^{-x}, \quad 0 \leq x < \infty \quad (3.8.16)$$

As $n \rightarrow \infty$, the distribution of $Y_n = \max\{X_1, X_2, \dots, X_n\} - \ln n$ converges to the distribution with distribution function F given by

$$F(x) = e^{-e^{-x}}, \quad x \in \mathbb{R} \quad (3.8.17)$$

Proof

Let $X_{(n)} = \max\{X_1, X_2, \dots, X_n\}$ and recall that $X_{(n)}$ has CDF G^n . Let F_n denote the CDF of Y_n . For $x \in \mathbb{R}$

$$F_n(x) = \mathbb{P}(Y_n \leq x) = \mathbb{P}(X_{(n)} \leq x + \ln n) = G^n(x + \ln n) = \left[1 - e^{-(x + \ln n)}\right]^n = \left(1 - \frac{e^{-x}}{n}\right)^n \quad (3.8.18)$$

By our famous limit from calculus again, $F_n(x) \rightarrow e^{-e^{-x}}$ as $n \rightarrow \infty$.

The limiting distribution in Exercise (27) is the *standard extreme value distribution*, also known as the *standard Gumbel distribution* in honor of Emil Gumbel. Extreme value distributions are studied in detail in the chapter on Special Distributions.

The Pareto Distribution

Recall that the *Pareto distribution* with shape parameter $a \in (0, \infty)$ has distribution function F given by

$$F(x) = 1 - \frac{1}{x^a}, \quad 1 \leq x < \infty \quad (3.8.19)$$

The Pareto distribution, named for Vilfredo Pareto, is a heavy-tailed distribution sometimes used to model financial variables. It is studied in more detail in the chapter on Special Distributions.

Suppose that X_n has the Pareto distribution with parameter n for each $n \in \mathbb{N}_+$. Then

1. $X_n \rightarrow 1$ as $n \rightarrow \infty$ in distribution (and hence also in probability).
2. The distribution of $Y_n = nX_n - n$ converges to the standard exponential distribution as $n \rightarrow \infty$.

Proof

1. The CDF of X_n is $F_n(x) = 1 - 1/x^n$ for $x \geq 1$. Hence $F_n(x) = 0$ for $n \in \mathbb{N}_+$ and $x \leq 1$ while $F_n(x) \rightarrow 1$ as $n \rightarrow \infty$ for $x > 1$. Thus the limit of F_n agrees with the CDF of the constant 1, except at $x = 1$, the point of discontinuity.
2. Let G_n denote the CDF of Y_n . For $x \geq 0$,

$$G_n(x) = \mathbb{P}(Y_n \leq x) = \mathbb{P}(X_n \leq 1 + x/n) = 1 - \frac{1}{(1 + x/n)^n} \quad (3.8.20)$$

By our famous theorem from calculus again, it follows that $G_n(x) \rightarrow 1 - 1/e^x = 1 - e^{-x}$ as $n \rightarrow \infty$. As a function of $x \in [0, \infty)$, this is the CDF of the standard exponential distribution.

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. So suppose that (X_1, X_2, \dots) is a sequence of independent, identically distributed, real-valued random variables (defined on the same probability space) with mean $\mu \in (-\infty, \infty)$ and standard deviation $\sigma \in (0, \infty)$. For $n \in \mathbb{N}_+$, let $Y_n = \sum_{i=1}^n X_i$ denote the sum of the first n variables, $M_n = Y_n/n$ the average of the first n variables, and $Z_n = (Y_n - n\mu)/\sqrt{n}\sigma$ the standard score of Y_n .

The fundamental theorems of probability

1. $M_n \rightarrow \mu$ as $n \rightarrow \infty$ with probability 1 (and hence also in probability and in distribution). This is the *law of large numbers*.
2. The distribution of Z_n converges to the standard normal distribution as $n \rightarrow \infty$. This is the *central limit theorem*.

In part (a), convergence with probability 1 is the *strong law of large numbers* while convergence in probability and in distribution are the *weak laws of large numbers*.

General Spaces

Our next goal is to define convergence of probability distributions on more general measurable spaces. For this discussion, you may need to refer to other sections in this chapter: the integral with respect to a positive measure, properties of the integral, and density functions. In turn, these sections depend on measure theory developed in the chapters on Foundations and Probability Measures.

Definition and Basic Properties

First we need to define the type of measurable spaces that we will use in this subsection.

We assume that (S, d) is a complete, separable metric space and let \mathcal{S} denote the Borel σ -algebra of subsets of S , that is, the σ -algebra generated by the topology. The standard spaces that we often use are special cases of the measurable space (S, \mathcal{S}) :

1. *Discrete*: S is countable and is given the discrete metric so \mathcal{S} is the collection of all subsets of S .
2. *Euclidean*: \mathbb{R}^n is given the standard Euclidean metric so \mathcal{B}_n is the usual σ -algebra of Borel measurable subsets of \mathbb{R}^n .

Additional details

Recall that the metric space (S, d) is *complete* if every Cauchy sequence in S converges to a point in S . The space is *separable* if there exists a countable subset that is dense. A complete, separable metric space is sometimes called a *Polish space* because such spaces were extensively studied by a group of Polish mathematicians in the 1930s, including Kazimierz Kuratowski.

As suggested by our setup, the definition for convergence in distribution involves both measure theory and topology. The motivation is the [theorem above](#) for the one-dimensional Euclidean space $(\mathbb{R}, \mathcal{B})$.

Convergence in distribution:

1. Suppose that P_n is a probability measure on (S, \mathcal{S}) for each $n \in \mathbb{N}_+$. Then P_n *converges (weakly)* to P_∞ as $n \rightarrow \infty$ if $P_n(A) \rightarrow P_\infty(A)$ as $n \rightarrow \infty$ for every $A \in \mathcal{S}$ with $P_\infty(\partial A) = 0$. We write $P_n \Rightarrow P_\infty$ as $n \rightarrow \infty$.

2. Suppose that X_n is a random variable with distribution P_n on (S, \mathcal{S}) for each $n \in \mathbb{N}_+^*$. Then X_n converges in distribution to X_∞ as $n \rightarrow \infty$ if $P_n \Rightarrow P_\infty$ as $n \rightarrow \infty$. We write $X_n \rightarrow X_\infty$ as $n \rightarrow \infty$ in distribution.

Notes

1. The definition makes sense since $A \in \mathcal{S}$ implies $\partial A \in \mathcal{S}$. Specifically, $\text{cl}(A) \in \mathcal{S}$ because $\text{cl}(A)$ is closed, and $\text{int}(A) \in \mathcal{S}$ because $\text{int}(A)$ is open.
2. The random variables need not be defined on the same probability space.

Let's consider our two special cases. In the discrete case, as usual, the measure theory and topology are not really necessary.

Suppose that P_n is a probability measures on a discrete space (S, \mathcal{S}) for each $n \in \mathbb{N}_+^*$. Then $P_n \Rightarrow P_\infty$ as $n \rightarrow \infty$ if and only if $P_n(A) \rightarrow P_\infty(A)$ as $n \rightarrow \infty$ for every $A \subseteq S$.

Proof

This follows from the definition. Every subset is both open and closed so $\partial A = \emptyset$ for every $A \subseteq S$.

In the Euclidean case, it suffices to consider distribution functions, as in the one-dimensional case. If P is a probability measure on $(\mathbb{R}^n, \mathcal{B}_n)$, recall that the *distribution function* F of P is given by

$$F(x_1, x_2, \dots, x_n) = P((-\infty, x_1] \times (-\infty, x_2] \times \dots \times (-\infty, x_n]), \quad (x_1, x_2, \dots, x_n) \in \mathbb{R}^n \quad (3.8.21)$$

Suppose that P_n is a probability measures on $(\mathbb{R}^n, \mathcal{B}_n)$ with distribution function F_n for each $n \in \mathbb{N}_+^*$. Then $P_n \Rightarrow P_\infty$ as $n \rightarrow \infty$ if and only if $F_n(\mathbf{x}) \rightarrow F_\infty(\mathbf{x})$ as $n \rightarrow \infty$ for every $\mathbf{x} \in \mathbb{R}^n$ where F_∞ is continuous.

Convergence in Probability

As in the case of $(\mathbb{R}, \mathcal{B})$, convergence in probability implies convergence in distribution.

Suppose that X_n is a random variable with values in S for each $n \in \mathbb{N}_+^*$, all defined on the same probability space. If $X_n \rightarrow X_\infty$ as $n \rightarrow \infty$ in probability then $X_n \rightarrow X_\infty$ as $n \rightarrow \infty$ in distribution.

Notes

Assume that the common probability space is $(\Omega, \mathcal{F}, \mathbb{P})$. Recall that convergence in probability means that $\mathbb{P}[d(X_n, X_\infty) > \epsilon] \rightarrow 0$ as $n \rightarrow \infty$ for every $\epsilon > 0$,

So as before, convergence with probability 1 implies convergence in probability which in turn implies convergence in distribution.

Skorohod's Representation Theorem

As you might guess, [Skorohod's theorem](#) for the one-dimensional Euclidean space $(\mathbb{R}, \mathcal{B})$ can be extended to the more general spaces. However the proof is not nearly as straightforward, because we no longer have the quantile function for constructing random variables on a common probability space.

Suppose that P_n is a probability measures on (S, \mathcal{S}) for each $n \in \mathbb{N}_+^*$ and that $P_n \Rightarrow P_\infty$ as $n \rightarrow \infty$. Then there exists a random variable X_n with values in S for each $n \in \mathbb{N}_+^*$, defined on a common probability space, such that

1. X_n has distribution P_n for $n \in \mathbb{N}_+^*$
2. $X_n \rightarrow X_\infty$ as $n \rightarrow \infty$ with probability 1.

One of the main consequences of Skorohod's representation, the preservation of convergence in distribution under continuous functions, is still true and has essentially the same proof. For the general setup, suppose that (S, d, \mathcal{S}) and (T, e, \mathcal{T}) are spaces of the type [described above](#).

Suppose that X_n is a random variable with values in S for each $n \in \mathbb{N}_+^*$ (not necessarily defined on the same probability space). Suppose also that $g: S \rightarrow T$ is measurable, and let D_g denote the set of discontinuities of g , and P_∞ the distribution of X_∞ . If $X_n \rightarrow X_\infty$ as $n \rightarrow \infty$ in distribution and $P_\infty(D_g) = 0$, then $g(X_n) \rightarrow g(X_\infty)$ as $n \rightarrow \infty$ in distribution.

Proof

By Skorohod's theorem, there exists random variables Y_n with values in S for $n \in \mathbb{N}_+^*$, defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$, such that Y_n has the same distribution as X_n for $n \in \mathbb{N}_+^*$, and $Y_n \rightarrow Y_\infty$ as $n \rightarrow \infty$ with probability 1. Since $\mathbb{P}(Y_\infty \in D_g) = P_\infty(D_g) = 0$ it follows that $g(Y_n) \rightarrow g(Y_\infty)$ as $n \rightarrow \infty$ with probability 1. Hence $g(Y_n) \rightarrow g(Y_\infty)$ as $n \rightarrow \infty$ in distribution. But $g(Y_n)$ has the same distribution as $g(X_n)$ for each $n \in \mathbb{N}_+^*$.

A simple consequence of the continuity theorem is that if a sequence of random vectors in \mathbb{R}^n converge in distribution, then the sequence of each coordinate also converges in distribution. Let's just consider the two-dimensional case to keep the notation simple.

Suppose that (X_n, Y_n) is a random variable with values in \mathbb{R}^2 for $n \in \mathbb{N}_+^*$ and that $(X_n, Y_n) \rightarrow (X_\infty, Y_\infty)$ as $n \rightarrow \infty$ in distribution. Then

1. $X_n \rightarrow X_\infty$ as $n \rightarrow \infty$ in distribution.
2. $Y_n \rightarrow Y_\infty$ as $n \rightarrow \infty$ in distribution.

Scheffé's Theorem

Our next discussion concerns an important result known as *Scheffé's theorem*, named after Henry Scheffé. To state our theorem, suppose that (S, \mathcal{S}, μ) is a measure space, so that S is a set, \mathcal{S} is a σ -algebra of subsets of S , and μ is a positive measure on (S, \mathcal{S}) . Further, suppose that P_n is a probability measure on (S, \mathcal{S}) that has density function f_n with respect to μ for each $n \in \mathbb{N}_+$, and that P is a probability measure on (S, \mathcal{S}) that has density function f with respect to μ .

If $f_n(x) \rightarrow f(x)$ as $n \rightarrow \infty$ for almost all $x \in S$ (with respect to μ) then $P_n(A) \rightarrow P(A)$ as $n \rightarrow \infty$ uniformly in $A \in \mathcal{S}$.

Proof

From basic properties of the integral it follows that for $A \in \mathcal{S}$,

$$|P(A) - P_n(A)| = \left| \int_A f d\mu - \int_A f_n d\mu \right| = \left| \int_A (f - f_n) d\mu \right| \leq \int_A |f - f_n| d\mu \leq \int_S |f - f_n| d\mu \quad (3.8.22)$$

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 . Note that $g_n^+ \leq f$ and $g_n^+ \rightarrow 0$ as $n \rightarrow \infty$ almost everywhere on S . Since f is a probability density function, it is trivially integrable, so by the dominated convergence theorem, $\int_S g_n^+ d\mu \rightarrow 0$ as $n \rightarrow \infty$. But $\int_{\mathbb{R}} g_n d\mu = 0$ so $\int_{\mathbb{R}} g_n^+ d\mu = \int_{\mathbb{R}} g_n^- d\mu$. Therefore $\int_S |g_n| d\mu = 2 \int_S g_n^+ d\mu \rightarrow 0$ as $n \rightarrow \infty$. Hence $P_n(A) \rightarrow P(A)$ as $n \rightarrow \infty$ uniformly in $A \in \mathcal{S}$.

Of course, the most important special cases of Scheffé's theorem are to discrete distributions and to continuous distributions on a subset of \mathbb{R}^n , as in the [theorem above](#) on density functions.

Expected Value

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.

In addition, convergence in distribution has elegant characterizations in terms of the convergence of the expected values of certain types of functions of the underlying random variables.

This page titled [3.8: Convergence in Distribution](#) is shared under a [CC BY 2.0](#) license and was authored, remixed, and/or curated by [Kyle Siegrist \(Random Services\)](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.