STAT 811 Probability Theory II

Chapter 8: Convergence in Distribution

Dr. Dewei Wang
Associate Professor
Department of Statistics
University of South Carolina
deweiwang@stat.sc.edu

This chapter discusses the basic notions of convergence in distribution. Given a sequence of random variables, when do their distributions converge in a useful way to a limit?

In statisticians' language, given a random sample $X_1, \ldots, X_n,$ the sample mean $\bar{X}_n$ is CAN; that is, Consistent and Asymptotically Normal. This means that $\bar{X}$ has an approximately normal distribution as the sample size grows. What exactly does this mean?

Recall our notation that *df* stands for distribution function. For the time being, we will understand this to correspond to a probability measure on $\mathbb{R}$. Recall that F is a *df* if

(i) $0 \leq F(x) \leq 1$;

(ii) F is non-decreasing;

(iii) $F(x+) = F(x), \forall x \in \mathbb{R},$ where

$$F(x+) = \lim_{\epsilon > 0, \epsilon \downarrow 0} F(x + \epsilon)$$

that is, F is right continuous.

$$F(\infty) := \lim_{y \uparrow \infty} F(y) \quad F(-\infty) := \lim_{y \downarrow \infty} F(y).$$

F is a *probability distribution function* if

$$F(-\infty) = 0, \quad F(+\infty) = 1$$

In this case, F is **proper** or *non-defective*. If $F(x)$ is a df, set

$$\mathcal{C}(F) = \{x \in \mathbb{R} : F \text{ is continuous at } x\}$$

A finite interval I with endpoints $a < b$ is called an *interval of continuity* for F if both $a, b \in \mathcal{C}(F)$. We know that

$$(\mathcal{C}(F))^c = \{x : F \text{ is discontinuous at } x\}$$

is at most countable. For an interval $I = (a, b]$, we write, as usual, $F(I) = F(b) - F(a)$. If $a, b \in \mathcal{C}(F)$, then $F((a, b)) = F((a, b])$. For non-interval set B, we write $F(B) = P(X \in B)$ where $X \sim F$.

Lemma 8.1.1

A distribution function $F(x)$ is determined on a dense set. Let D be dense in R. Suppose $F_D(\cdot)$ is defined on D and satisfies the following:

(a) $F_D(\cdot)$ is non-decreasing on D.

(b) $0 \leq F_D(x) \leq 1$, for all $x \in D$.

(c) $\lim_{x \in D, x \to +\infty} F_D(x) = 1, \quad \lim_{x \in D, x \to -\infty} F_D(x) = 0$

Define for all $x \in \mathbb{R}$

$$F(x) := \inf_{\substack{y > x \\ y \in D}} F_D(y) = \lim_{\substack{y \downarrow x \\ y \in D}} F_D(y)$$

Then F is a right continuous probability df. Thus, any two right continuous df's agreeing on **a dense set** will agree everywhere.

Proof of Lemma 8.1.1.

Four definitions.

We now consider four definitions related to weak convergence of probability measures. Let $\{F_n, n \geq 1\}$ be probability distribution functions and let F be a distribution function which is not necessarily proper.

(1) **Vague convergence.** The sequence $\{F_n\}$ converges vaguely to F, written $F_n \xrightarrow{v} F$, if for every finite interval of continuity I of F, we have

$$F_n(I) \to F(I)$$

(See Chung (1968), Feller (1971).)

(2) **Proper convergence.** The sequence $\{F_n\}$ converges properly to F, written $F_n \to F$ if $F_n \xrightarrow{v} F$ and F is a proper df; that is $F(\mathbb{R}) = 1$. (See Feller (1971).)

(3) **Weak convergence.** The sequence $\{F_n\}$ converges weakly to F, written $F_n \xrightarrow{w} F$, if

$$F_n(x) \to F(x)$$

for all $x \in \mathcal{C}(F)$. (See Billingsley (1968, 1995).)

(4) **Complete convergence.** The sequence $\{F_n\}$ converges completely to F, written $F_n \xrightarrow{c} F$, if $F_n \xrightarrow{w} F$ and F is proper. (See Loève (1977).)

Example.

Define

$$F_n(x) := F\left(x + (-1)^n n\right)$$

Theorem 8.1 .1 (Equivalence of the Four Definitions)

If F is proper, then the four definitions (1),(2),(3),(4) are equivalent.

Proof.

8.1 Basic Definitions

8.1 Basic Definitions

Notation: If $\{F, F_n, n \geq 1\}$ are probability distributions, write $F_n \Rightarrow F$ to mean any of the equivalent notions given by $(1)-(4)$. If X_n is a random variable with distribution F_n and X is a random variable with distribution F, we write $X_n \Rightarrow X$ to mean $F_n \Rightarrow F$. This is read X_n converges in distribution to X or F_n converges weakly to F. Notice that unlike almost sure, in probability, or L_p convergence, convergence in distribution says **nothing** about the behavior of the random variables themselves and only comments on the behavior of the distribution functions of the random variables.

Example 8.1.1.

Let N be an $N(0,1)$ random variable so that the distribution function is symmetric. Define for $n \geq 1$

$$X_n = (-1)^n N$$

Then $X_n \stackrel{d}{=} N$, so automatically $X_n \Rightarrow N$. But of course $\{X_n\}$ neither converges almost surely nor in probability.

Example 8.1.2.

Let $\{X_n, n \geq 1\}$ be iid with common unit exponential distribution

$$P[X_n > x] = e^{-x}, \quad x > 0$$

Set $M_n = \max_{i=1}^n X_i$ for $n \geq 1$. Then $M_n - \log n \Rightarrow Y$, where

$$P[Y \leq x] = \exp\left\{-e^{-x}\right\}, \quad x \in \mathbb{R}$$

Remark 8.1.2

Weak limits are unique. If $F_n \xrightarrow{w} F$, and also $F_n \xrightarrow{w} G$, then $F = G$.

Why?

Consider the following modes of convergence that are stronger than weak convergence.

(a) $F_n(A) \to F(A), \forall A \in \mathcal{B}(\mathbb{R})$
(b) $\sup_{A \in \mathcal{B}(\mathbb{R})} |F_n(A) - F(A)| \to 0$.

Definition (a) (and hence (b)) would rule out many circumstances we would like to fall under weak convergence. Two examples illustrate the point of this remark.

Example 8.2.1 (i). Suppose $X_n \sim F_n$ puts mass $\frac{1}{n}$ at points $\left\{\frac{1}{n}, \frac{2}{n}, \ldots, \frac{n}{n}\right\}$., and let

$$X_n \sim F(x) = x, \quad 0 \leq x \leq 1$$

where F is the uniform distribution on $[0, 1]$.

Example 8.2.1 (ii). *DeMoivre-Laplace central limit theorem*: This is a situation similar to what was observed in (i). Suppose $\{X_n, n \geq 1\}$ are iid, with

$$P[X_n = 1] = p = 1 - P[X_n = 0]$$

Set $S_n = \sum_{i=1}^{n} X_i$, which has a binomial distribution with parameters n, p. Then the DeMoivre-Laplace central limit theorem states

$$F_n(x) = P\left[\frac{S_n - np}{\sqrt{npq}} \leq x\right] \to F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$$

Weak convergence, because of its connection to continuous functions, is more useful than the convergence notions (a) or (b).

8.2 Scheffé's lemma

The convergence definition (b) is called total variation convergence and has connections to density convergence through Scheffé's lemma.

Lemma 8.2.1 (Scheffé's lemma)

Suppose $\{F, F_n, n \geq 1\}$ are probability distributions with densities $\{f, f_n, n \geq 1\}$. Then

$$\sup_{B \in \mathcal{B}(\mathrm{R})} |F_n(B) - F(B)| = \frac{1}{2} \int |f_n(x) - f(x)| \, dx$$

If $f_n(x) \to f(x)$ almost everywhere (that is, for all x except a set of Lebesgue measure 0), then

$$\int |f_n(x) - f(x)| \, dx \to 0$$

and thus $F_n \to F$ in total variation (and hence weakly).

Proof of Scheffés lemma.

Proof of Scheffés lemma.

Note that if $F_n \stackrel{w}{\to} F$ and F_n and F have densities f_n, f, it does not necessarily follow that $f_n(x) \to f(x)$.

Example.

Proposition 8.2.1 Suppose $\{U_n, n \geq 1\}$ are iid $U(0,1)$ random variables and suppose

$$U_{(1,n)} \leq U_{(2,n)} \leq \cdots \leq U_{(n,n)}$$

are the order statistics. Assume $k = k(n)$ is a function of n satisfying $k(n) \to \infty$ and $k/n \to 0.$ as $n \to \infty$. Let

$$\xi_n = \frac{U_{(k,n)} - \frac{k}{n}}{\sqrt{\frac{k}{n}\left(1 - \frac{k}{n}\right)\frac{1}{n}}}$$

Then the density of ξ_n converges to a standard normal density and hence by Scheffe's lemma, as $n \to \infty$,

$$\sup_{B \in \mathcal{B}(\mathbb{R})} \left| P\left[\xi_n \in B\right] - \int_B \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du \right| \to 0$$

A proof proceeds by playing with the density of $U_{(k,n)}$

$$f_n(x) = \frac{n!}{(k-1)!(n-k)!} x^{k-1}(1-x)^{n-k}, \quad 0 < x < 1.$$

8.3 The Baby Skorohod Theorem Skorohod's theorem is a conceptual aid which makes certain weak convergence results easy to prove by continuity arguments. The theorem is true in great generality. We only consider the result for real valued random variables and hence the name Baby Skorohod Theorem.
We begin with a brief discussion of the relationship of almost sure convergence and weak convergence.

Proposition 8.3.1

Suppose $\{X, X_n, n \geq 1\}$ are random variables.

$$\text{If} \quad X_n \stackrel{a.s.}{\to} X, \quad \text{then} \quad X_n \Rightarrow X$$

Proof.

The converse if false: Recall Example 8.1.1 .

Despite the fact that convergence in distribution does not imply a.s. convergence, Skorohod's theorem provides a partial converse.

Theorem 8.3.2 (Baby Skorohod Theorem)

Suppose $\{X_n, n \geq 0\}$ are random variables defined on the probability space $(\Omega, \mathcal{B}, P)$ such that

$$X_n \Rightarrow X_0$$

Then there exist random variables $\left\{X_n^\#, n \geq 0\right\}$ defined on the Lebesgue probability space $([0,1], \mathcal{B}([0,1]), \lambda =$ Lebesgue measure $)$ such that for each fixed $n \geq 0$

$$X_n \stackrel{d}{=} X_n^\# \quad \text{and} \quad X_n^\# \stackrel{a.s.}{\to} X_0^\#$$

where a.s. means almost surely with respect to λ.

The proof of Skorohod's theorem requires the following result.

Lemma 8.3.1

Suppose F_n is the distribution function of X_n so that $F_n \Rightarrow F_0$.

$$\text{If} \quad t \in (0,1) \cap \mathcal{C}\left(F_0^{\leftarrow}\right), \quad \text{then} \quad F_n^{\leftarrow}(t) \to F_0^{\leftarrow}(t).$$

Proof.

This lemma only guarantees convergence of $F_n^{\leftarrow}$ to $F_0^{\leftarrow}$ at continuity points of the limit. However, convergence could take place on more points. For instance, if $F_n = F_0$ for all n, $F_n^{\leftarrow} = F_0^{\leftarrow}$ and convergence would be everywhere.

Proof of the Baby Skorohod Theorem.

The next corollary looks tame when restricted to $\mathbb{R}$, but its multidimensional generalizations have profound consequences. For a map $h : \mathbb{R} \mapsto \mathbb{R}$, define

$$\mathrm{Disc}(h) = \{x : h \text{ is not continuous at } x\} = (\mathcal{C}(h))^c$$

Corollary 8.3.1 (Continuous Mapping Theorem)

Let $\{X_n, n \geq 0\}$ be a sequence of random variables such that

$$X_n \Rightarrow X_0$$

For $n \geq 0$, assume F_n is the distribution function of X_n. Let $h : \mathbb{R} \mapsto \mathbb{R}$ satisfy $P\left[X_0 \in \mathrm{Disc}(h)\right] = 0$. Then

$$h\left(X_n\right) \Rightarrow h\left(X_0\right)$$

and if h is bounded, dominated convergence implies

$$Eh\left(X_n\right) = \int h(x)F_n(dx) \to Eh(x) = \int h(x)F_0(dx).$$

Proof.

8.3.1 The Delta Method

The delta method allows us to take a basic convergence, for instance to a limiting normal distribution, and apply smooth functions and conclude that the functions are asymptotically normal as well.

In statistical estimation we try to estimate a parameter θ from a parameter set Θ based on a random sample of size n with a statistic

$$T_n = T_n(X_1, \ldots, X_n)$$

This means we have a family of probability models

$$\{(\Omega, \mathcal{B}, P_\theta), \theta \in \Theta\}$$

and we are trying to choose the correct model. The estimator T_n is consistent if

$$T_n \stackrel{P_\theta}{\to} \theta$$

for every θ. The estimator T_n is CAN, if for all $\theta \in \Theta$

$$\lim_{n\to\infty} P_\theta [\sigma_n (T_n - \theta) \leq x] = N(0, 1, x)$$

for some $\sigma_n \to \infty$

8.3.1 The Delta Method

The delta method allows us to take a basic convergence, for instance to a limiting normal distribution, and apply smooth functions and conclude that the functions are asymptotically normal as well.

In statistical estimation we try to estimate a parameter θ from a parameter set Θ based on a random sample of size n with a statistic

$$T_n = T_n(X_1, \ldots, X_n)$$

This means we have a family of probability models

$$\{(\Omega, \mathcal{B}, P_\theta), \theta \in \Theta\}$$

and we are trying to choose the correct model. The estimator T_n is consistent if

$$T_n \xrightarrow{P_\theta} \theta$$

for every θ. The estimator T_n is CAN, if for all $\theta \in \Theta$

$$\lim_{n\to\infty} P_\theta[\sigma_n(T_n - \theta) \leq x] = N(0, 1, x)$$

for some $\sigma_n \to \infty$.

Suppose we have a CAN estimator of θ, but we need to estimate a smooth function $g(\theta)$. For example, in a family of exponential densities, θ may represent the mean but we are interested in the variance θ^2. We see from the delta method that $g\left(T_n\right)$ is also CAN for $g(\theta)$.

We illustrate the method using the central limit theorem (CLT) to be proved in the next chapter. Let $\{X_j, j \geq 1\}$ be iid with $E\left(X_n\right) = \mu$ and $\operatorname{Var}\left(X_n\right) = \sigma^2$. From the CLT we get in terms of $\bar{X} = \sum_{i=1}^{n} X_i / n$ that

$$\sqrt{n}\left(\frac{\bar{X}-\mu}{\sigma}\right) \Rightarrow N(0,1)$$

So $\bar{X}$ is CAN for μ. The delta method asserts that if $g(x)$ has a non-zero derivative $g'(\mu)$,

$$\sqrt{n}\left(\frac{g(\bar{X})-g(\mu)}{\sigma g'(\mu)}\right) \Rightarrow N(0,1) \qquad (*)$$

So $g(\bar{X})$ is CAN for $g(\mu)$.

Proof of $(*)$.

Remark. Suppose $\{X_n, n \geq 0\}$ is a sequence of random variables such that

$$X_n \Rightarrow X_0$$

Suppose further that

$$h : \mathbb{R} \mapsto \mathbb{S}$$

where $\mathbb{S}$ is some nice metric space, for example, $\mathbb{S} = \mathbb{R}^2$. Then if

$$P\left[X_0 \in \text{Disc}(h)\right] = 0$$

Skorohod's theorem suggests that it should be the case that

$$h\left(X_n\right) \Rightarrow h(X)$$

in $\mathbb{S}$. But what does weak convergence in $\mathbb{S}$ mean?

8.4 Weak Convergence Equivalences; Portmanteau Theorem

In this section we discuss several conditions which are equivalent to weak convergence of probability distributions. Some of these are of theoretical use and some allow easy generalization of the notion of weak convergence to higher dimensions and even to function spaces. The definition of weak convergence of distribution functions on $\mathbb{R}$ is notable for not allowing easy generalization to more sophisticated spaces. The modern theory of weak convergence of stochastic processes rests on the equivalences to be discussed next. We nead the following definition. For $B \in \mathcal{B}(\mathbb{R})$, let

$$\begin{aligned}
\partial(B) &= \text{ the boundary of } B \\
&= \bar{B} \backslash B^{\circ} = \text{ the closure of } B \text{ minus the interior of } B \\
&= \{x : \exists y_n \in B, y_n \to x \text{ and } \exists z_n \in B^c, z_n \to x\} \\
&= \text{ points reachable from both outside and inside } B.
\end{aligned}$$

Theorem 8.4.1 (Portmanteau Theorem) Alexandrov

Let $\{F_n, n \geq 0\}$ be a family of proper distributions with random vectors $X_n \sim F_n$ and $X \sim F_0$. The following are equivalent.

(i) $F_n \Rightarrow F_0$.

(ii) For all $f : \mathbb{R} \mapsto \mathbb{R}$ which are bounded and continuous,

$$\int f dF_n \to \int f dF_0 \quad \text{or} \quad Ef(X_n) \to Ef(X_0).$$

(iii) $Ef(X_n) \to Ef(X_0)$ for all bounded, Lipschitz functions f.

(iv) $\liminf_{n\to\infty} P(X_n \in G) \geq P(X \in G)$ for every open set G.

(v) $\limsup_{n\to\infty} P(X_n \in F) \leq P(X \in F)$ for every closed set F.

(vi) $P(X_n \in B) \to P(X \in B)$ for all $B \in \mathcal{B}(\mathbb{R})$ satisfies $P(X \in \partial(B)) = 0$.

(vii) $Ef(X_n) \to Ef(X_0)$ for all bounded, uniformly continuous functions f.

Proof.

8.4 Weak Convergence Equivalences; Portmanteau Theorem

8.4 Weak Convergence Equivalences; Portmanteau Theorem

Example 8.4.1 The discrete uniform distribution is close to the continuous uniform distribution.

Suppose F_n has atoms at i/n, $1 \leq i \leq n$ of size $1/n$. Let F_0 be the uniform distribution on $[0, 1]$. Then $F_n \Rightarrow F_0$.

Proof.

Proposition 8.5.1

Let $\{X, X_n, n \geq 1\}$ be random variables on the same probability space (Ω, B, P)

(i) If $X_n \stackrel{a.s.}{\to} X$ then $X_n \stackrel{P}{\to} X$.

(ii) If $X_n \stackrel{P}{\to} X$ then $X_n \Rightarrow X$.

All the converses are false.

Proof.

There is one special case where convergence in probability and convergence in distribution are the same.

Proposition 8.5.2

Suppose $\{X_n, n \geq 1\}$ are random variables. If c is a constant, then

$$X_n \xrightarrow{P} c \quad \Leftrightarrow \quad X_n \Rightarrow c.$$

Proof.

Theorem 8.6.1 (Slutsky's Theorem)

Suppose $\{X, X_n, Y_n, \xi_n, n \geq 1\}$ are random variables.

(a) If $X_n \Rightarrow X$, and $X_n - Y_n \xrightarrow{P} 0$ then $Y_n \Rightarrow X$.

(b) Equivalently, if $X_n \Rightarrow X$, and $\xi_n \xrightarrow{P} 0$, then $X_n + \xi_n \Rightarrow X$.

Proof.

Theorem 8.6.2 (Second Converging Together Theorem)

Let us suppose that $\{X_{un}, X_u, Y_n, X; n \geq 1, u \geq 1\}$ are random variables such that for each n $Y_n, X_{un}, u \geq 1$ are defined on a common domain. Assume for each u, as $n \to \infty$

$$X_{un} \Rightarrow X_u$$

and as $u \to \infty$

$$X_u \Rightarrow X$$

Suppose further that for all $\epsilon > 0$,

$$\lim_{u \to \infty} \limsup_{n \to \infty} P\left[|X_{un} - Y_n| > \epsilon\right] = 0.$$

Then we have

$$Y_n \Rightarrow X$$

as $n \to \infty$.

Proof.

8.7 The Convergence to Types Theorem

Many convergence in distribution results in probability and statistics are of the following form: Given a sequence of random variables $\{\xi_n, n \geq 1\}$ and $a_n > 0$ and $b_n \in \mathbb{R}$, we prove that

$$\frac{\xi_n - b_n}{a_n} \Rightarrow Y$$

where Y is a non-degenerate random variable; that is, Y is not a constant a.s. This allows us to write

$$P\left[\frac{\xi_n - b_n}{a_n} \leq x\right] \approx P[Y \leq x] =: G(x)$$

or by setting $y = a_n x + b_n$

$$P\left[\xi_n \leq y\right] \approx G\left(\frac{y - b_n}{a_n}\right)$$

This allows us to approximate the distribution of ξ_n with a location-scale family.

Example. suppose $\{X_n, n \geq 1\}$ are iid with $E(X_n) = \mu$ and $\mathrm{Var}(X_n) = \sigma^2$. The Central Limit Theorem: for $x \in \mathbb{R}$

$$P\left[\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right] \to P[Y \leq x] = \int_{-\infty}^{x} \frac{e^{-u^2/2}}{\sqrt{2\pi\sigma}} du$$

so that

$$P[S_n \leq y] \approx N\left(\frac{y - n\mu}{\sigma\sqrt{n}}\right)$$

Definition. Two distribution functions $U(x)$ and $V(x)$ are of the same type if there exist constants $A > 0$ and $B \in \mathbb{R}$ such that

$$V(x) = U(Ax + B)$$

In terms of random variables, if X has distribution U and Y has distribution V, then

$$Y \stackrel{d}{=} \frac{X - B}{A}$$

For example, if $X_{0,1}$ has $N(0,1,x)$ as its distribution and $X_{\mu,\sigma}$ has $N(\mu, \sigma^2)$ as its distribution, then $X_{\mu,\sigma} \stackrel{d}{=} \sigma X_{0,1} + \mu$.

Theorem 8.7.1 (Convergence to Types Theorem)

We suppose $\mathcal{U}(x)$ and $\mathcal{V}(x)$ are two proper distributions, neither of which is concentrated at a point. Suppose $X_n \sim F_n$ and the $U \sim \mathcal{U}$, $V \sim \mathcal{V}$. We have constants $a_n > 0, \alpha_n > 0, b_n \in \mathbb{R}, \beta_n \in \mathbb{R}$
(a) If $F_n\left(a_n x + b_n\right) \xrightarrow{w} \mathcal{U}(x), \quad F_n\left(\alpha_n x + \beta_n\right) \xrightarrow{w} \mathcal{V}(x)$ or equivalently

$$\frac{X_n - b_n}{a_n} \Rightarrow U, \quad \frac{X_n - \beta_n}{\alpha_n} \Rightarrow V \quad (*1)$$

then there exist constants $A > 0$, and $B \in \mathbb{R}$ such that as $n \to \infty$,

$$\frac{\alpha_n}{a_n} \to A > 0, \quad \frac{\beta_n - b_n}{a_n} \to B \quad (*2)$$

and $\mathcal{V}(x) = \mathcal{U}(Ax + B), \quad V \stackrel{d}{=} (U - B)/A \quad (*3)$.
(b) Conversely, if $(*2)$ holds, then either of the relations in $(*1)$ implies the other and $(*3)$ holds.

Proof.

8.7 The Convergence to Types Theorem

8.7 The Convergence to Types Theorem

8.7.1 Application of Convergence to Types: Limit Distributions for Extremes

A beautiful example of the use of the convergence to types theorem is the derivation of the extreme value distributions. These are the possible limit distributions of centered and scaled maxima of iid random variables.

Suppose $\{X_n, n \geq 1\}$ is an iid sequence of random variables with common distribution F. The extreme observation among the first n is

$$M_n := \bigvee_{i=1}^{n} X_i$$

8.7.1 Application of Convergence to Types: Limit Distributions for Extremes

Theorem 8.7.2

Suppose there exist normalizing constants $a_n > 0$ and $b_n \in \mathbb{R}$ such that

$$F^n(a_n x + b_n) = P\left[\frac{M_n - b_n}{a_n} \leq x\right] \xrightarrow{w} G(x)$$

where the limit distribution G is proper and non-degenerate. Then G is the type of one of the following extreme value distributions:

(i) $\Phi_\alpha(x) = \exp\{-x^{-\alpha}\}, \quad x > 0, \quad \alpha > 0$

(ii)

$$\Psi_\alpha(x) = \begin{cases} \exp\{-(x)^\alpha\}, & x < 0, \quad \alpha > 0 \\ 1 & x > 0 \end{cases}$$

(iii) $\Lambda(x) = \exp\{-e^{-x}\}, \quad x \in \mathbb{R}$

8.7.1 Application of Convergence to Types: Limit Distributions for Extremes

The statistical significance is the following. The types of the three extreme value distributions can be united as a one parameter family indexed by a shape parameter $\gamma \in \mathbb{R}$:

$$G_\gamma(x) = \exp\left\{-(1+\gamma x)^{-1/\gamma}\right\}, \quad 1+\gamma x > 0$$

where we interpret the case of $\gamma = 0$ as

$$G_0 = \exp\left\{-e^{-x}\right\}, \quad x \in \mathbb{R}$$

Often in practical contexts the distribution F is unknown and we must estimate the distribution of M_n or a quantile of M_n. For instance, we may wish to design a dam so that in 10,000 years, the probability that water level will exceed the dam height is 0.001 . If we assume F is unknown but satisfies (8.24) with some G_γ as limit, then we may write

$$P\left[M_n \le x\right] \approx G_\gamma\left(a_n^{-1}\left(x - b_n\right)\right)$$

and now we have a three parameter estimation problem since we

Proof.

8.7.1 Application of Convergence to Types: Limit Distributions for Extremes

8.7.1 Application of Convergence to Types: Limit Distributions for Extremes

8.7.1 Application of Convergence to Types: Limit Distributions for Extremes

8.7.1 Application of Convergence to Types: Limit Distributions for Extremes

8.7.1 Application of Convergence to Types: Limit Distributions for Extremes

8.7.1 Application of Convergence to Types: Limit Distributions for Extremes

8.7.1 Application of Convergence to Types: Limit Distributions for Extremes

8.7.1 Application of Convergence to Types: Limit Distributions for Extremes