Chapter 5

Multiple Random Variables

Convergence in Didtribution

Outline

Convergence in Distribution

We say that X_n converges to X in distribution ($X_n \to^{\mathcal{D}} X$ or $X_n \Rightarrow X$) if, for every bounded continuous function $h : \mathbb{R} \to \mathbb{R}$,

$$\lim_{n\to\infty} Eh(X_n) = Eh(X).$$

Convergence in distribution differs from the other modes of convergence in that it is based not on a direct comparison of the random variables X_n with X but rather on a comparison of the distributions $P\{X_n \in A\}$ and $P\{X \in A\}$. Using the change of variables formula, convergence in distribution can be written

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} h(x)\, dF_{X_n}(x) = \int_{-\infty}^{\infty} h(x)\; dF_X(x).$$

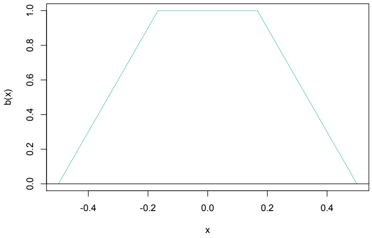
In this case, we may also write $F_{X_n} \to^{\mathcal{D}} F_X$

Convergence in Distribution

Let X_n be uniformly distributed on the points $\{1/n, 2/n, \cdots n/n = 1\}$.Then, using the convergence of a Riemann sum to a Riemann integral, we have as $n \to \infty$,

$$Eh(X_n) = \sum_{i=1}^{n} h\left(\frac{i}{n}\right)\frac{1}{n} \to \int_0^1 h(x)\,dx = Eh(X)$$

where X is a uniform random variable on the interval $[0,1]$.



Define the bump function $b(x)$ with support $[-1/2, 1/2]$, ramping up from 0 to 1, taking the value 1 at $x = 0$ and than ramping back to zero. In addition, define the shift $b_{x_0}(x) = b(x - x_0)$

Convergence in Distribution

For X_n and X , integer-valued random variables, then

$$\lim_{n\to\infty} P\{X_n = x_0\} = \lim_{n\to\infty} Eb_{x_0}(X_n) = Eb_{x_0}(X) = P\{X = x_0\}$$

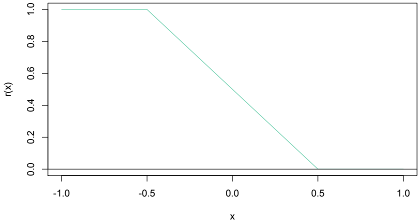
Thus, convergence in distribution for integer-valued random variables is the same is the convergences of the mass function.

Example. Let $p \in (0,1)$ and let $X_n \sim Hyper([np], n-[np], k)$. Then,

$$\begin{aligned} P\{X_n = x_0\} \quad &= \quad \binom{k}{x_0} \frac{([np])_{x_0}(n-[np])_{k-x_0}}{(n)_k} = \binom{k}{x_0} \frac{([np])_{x_0}}{(n)_{x_0}} \cdot \frac{(n-[np])_{k-x_0}}{(n-x_0+1)_{k-x_0}} \\ &\to \binom{k}{x_0} p^{x_0}(1-p)^{k-x_0} \end{aligned}$$

and the limiting distribution is $Bin(k,p)$.

Convergence in Distribution



Define the ramp function $r(x)$ ramping down from 1 to 0 on $[-1/2, 1/2]$, continuous and flat elsewhere, In addition, define $r_{x_0,\epsilon}(x) = r((x - x_0)/\epsilon)$.

With the choice $r_{x_0,\epsilon}(x)$, taking the limit as $\epsilon \to 0$, we show that $X_n \to^{\mathcal{D}} X$ if and only if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

for all points x that are continuity points of F_X.

Example. For X_n be uniformly distributed on the points $\{1/n, 2/n, \cdots n/n = 1\}$

$$P\{X_n \le x\} = \frac{[nx]}{n} \to x = P\{X \le x\}$$

Convergence in Distribution

Example. Let $X_i, 1 \leq i \leq n$, be independent uniform random variable in the interval $[0,1]$ and let $Y_n = n(1 - X_{(n)})$. Then,

$$F_{Y_n}(y) = P\{n(1 - X_{(n)}) \leq y\} = P\left\{1 - \frac{y}{n} \leq X_{(n)}\right\} = 1 - \left(1 - \frac{y}{n}\right)^n \to 1 - e^{-y}.$$

Thus, the magnified gap between the highest order statistic and 1 converges in distribution to an exponential random variable, parameter 1.

Example. Let X_p be $Geo(p)$. Then $P\{X_p > n\} = (1-p)^n$. $EX_p = (1-p)/p$, $E[pX_p] = (1-p) \sim 1$ for p near 0. Then,

$$P\{pX_p > x\} = P\{X_p > x/p\} = (1-p)^{[x/p]} \to \exp(-x) \quad \text{as} \quad p \to 0.$$

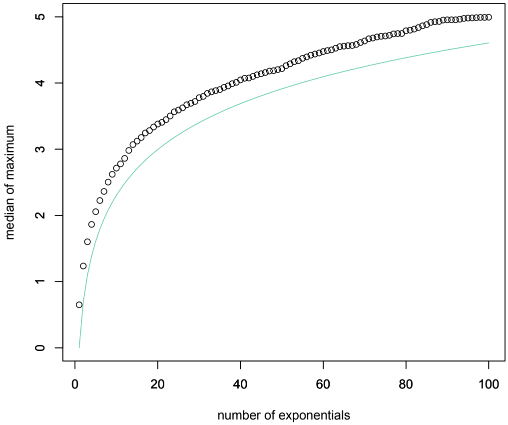
Therefore pX_p converges in distribution to an $Exp(1)$ random variable.

Convergence in Distribution

Example. Let $X_i, i \geq 1$, be independent $Exp(1)$ random variables. Define $M_n = \max_{1 \leq i \leq n} X_i$. To see how quickly M_n grows, we simulate.

```
> simexp<-matrix(rexp(100000),ncol=100)
> cmax<-matrix(numeric(100000),ncol=100)
> medmax<-numeric(100)
> for (i in 1:1000){cmax[i,]<-cummax(simexp[i,])}
> for (j in 1:100){medmax[j]<-median(cmax[,j])}
> plot(1:100,medmax,xlim=c(1,100),ylim=c(0,5),
  xlab="number of exponentials",ylab="median of maximum")
> par(new=TRUE)
```

Convergence in Distribution




```
> curve(log(x),1,100,ylim=c(0,5),col="aquamarine3",xlab="",ylab="")
```

Convergence in Distribution

Thus, define $Y_n = \max_{1\leq i\leq n} X_i - \ln n$. Then,

$$
\begin{aligned}
P\{Y_n \leq y\} &= P\{X_1 \leq y + \ln n, \ldots, X_n \leq y + \ln n\} = P\{X_1 \leq y + \ln n\}^n \\
&= (1 - e^{-(y+\ln n)})^n = (1 - \frac{e^{-y}}{n})^n \\
&\to \exp(e^{-y}) \quad \text{as} \quad n \to \infty
\end{aligned}
$$

This is called a Gumbel distribution and is an example of an extreme value distribution.

Separating and Convergence Determining

1. A collection $\mathcal{H}$ of continuous and bound functions is called separating if for any two distribution functions F, G,

$$\int h \, dF = \int h \, dG \text{ for all } h \in \mathcal{H}$$

   implies $F = G$.

2. A collection $\mathcal{H}$ of continuous and bound functions is called convergence determining if for any sequence distribution functions $\{F_n; n \geq 1\}$ and a distribution F,

$$\lim_{n\to\infty} \int h \, dF_n = \int h \, dF \text{ for all } h \in \mathcal{H}$$

   implies $F_n \to^{\mathcal{D}} F$.

Convergence determining sets are separating.

Separating and Convergence Determining

Example. For integer-valued random variables, then by the uniqueness of power series, the collection $\mathcal{H} = \{z^x; , 0 \leq z \leq 1\}$ is separating. Take the mass functions $f_{X_k} = I_{\{k\}}$ to see that it is not convergence determining.

If we want to use a separating collection for convergence in distribution, we will need an additional requirement on the distribution functions to ensure that "mass does not run off to infinity."

Definition. A collection $\mathcal{F}$ of distribution functions is called tight if for each $\epsilon > 0$, then exists $M > 0$ such that

$$F(M) - F(-M) \geq 1 - \epsilon, \text{ for all } F \in \mathcal{F}.$$

Exercise. Any finite collection of distribution functions is tight.

Separating and Convergence Determining

Theorem. Let $\{F_n; n \geq 1\}$ be a tight family of distribution functions and let $\mathcal{H}$ be separating. Then $F_n \to^{\mathcal{D}} F$ if and only if

$$\lim_{n\to\infty} \int h \, dF_n$$

exists for all $h \in \mathcal{H}$. In this case, the limit is $\int h \, dF$.

Proof. Cantor diagonalization argument.

The goal, for any given separating class, is to find a sufficient condition to ensure that the distributions in the approximating sequence of distributions are tight. For example,

Theorem. Let $\{X_n; n \geq 1\}$ be $\mathbb{N}$-valued random variables having respective probability generating functions $\rho_n(z) = Ez^{X_n}$. If

$$\lim_{n\to\infty} \rho_n(z) = \rho(z),$$

and ρ is continuous at $z = 1$, then X_n converges in distribution to a random variable X with generating function ρ.

Discrete Random Variables

Let X_n be a $Bin(n, p)$ random variable. Then

$$\rho_{X_n}(z) = Ez^{X_n} = ((1-p) + pz)^n = ((1 + p(z-1))^n.$$

Set $\lambda = np$, then

$$\lim_{n\to\infty} Ez^{X_n} = \lim_{n\to\infty} (1 + \frac{\lambda}{n}(z-1))^n = \exp\lambda(z-1),$$

the generating function of a Poisson random variable. The convergence of the distributions of $\{X_n; n \geq 1\}$ follows from the fact that the limiting probability generating function is continuous at $z = 1$.

Discrete Random Variables

Let $z \in [0,1)$ and choose $\zeta \in (z,1)$. Then for each n and k,

$$P\{X_n = k\}z^k < \zeta^k.$$

Thus, by the Weierstrass M-test, ρ_n converges uniformly to an analytical function $\tilde{\rho}$ on $[0,\zeta]$ and thus $\tilde{\rho}$ is continuous at z.

$$\begin{aligned}
\lim_{n\to\infty} P\{X_n > x\} &= \lim_{n\to\infty}\lim_{z\to 1}\left(\rho_n(z) - \sum_{k=1}^{x} P\{X_n = k\}z^k\right) \\
&= \lim_{z\to 1}\lim_{n\to\infty}\left(\rho_n(z) - \sum_{k=1}^{x} P\{X_n = k\}z^k\right) = \lim_{z\to 1}\left(\tilde{\rho}(z) - \sum_{k=1}^{x} \tilde{\rho}^{(k)}(0)z^k\right) \\
&= \tilde{\rho}(1) - \sum_{k=1}^{x} \tilde{\rho}^{(k)}(0) < \epsilon
\end{aligned}$$

by choosing x sufficiently large. Thus, we have that $\{X_n; n \geq 1\}$ is tight.

Discrete Random Variables

Consider families of discrete random variables and let $\{F_X(\cdot|\theta_n); n \geq 1\}$ be a sequence of distributions from that family. Then

$$F_X(\cdot|\theta_n) \to^{\mathcal{D}} F_X(\cdot|\theta) \quad \text{if and only if} \quad \theta_n \to \theta.$$

This applies to binomial, geometric, negative binomial (p), and Poisson (λ) families of random variables.

In each case,

$$\lim_{n\to\infty} E_{\theta_n} z^{X_n} = \lim_{n\to\infty} \rho_{\theta_n}(z) = \rho_\theta(z).$$

and $\rho_\theta(z)$ is continuous at $z = 1$.

We now move on to continuous random variables.

Characteristic Functions

The characteristic function or the Fourier transform for a probability distribution F on $\mathbb{R}^d$ is

$$\phi(\theta) = \int e^{i\langle \theta, x\rangle}\, dF(x) = Ee^{i\langle \theta, X\rangle}$$

where X is a random variable with distribution function F. Because the Fourier transform is one-to-one $\{e^{i\langle \theta, x\rangle}; \theta \in \mathbb{R}^d\}$ is a separating class of functions. Here is the main result.

Continuity Theorem. Let $\{F_n; n \geq 1\}$ be a sequence of distributions on $\mathbb{R}$ with corresponding characteristic function $\{\phi_n; n \geq 1\}$ satisfying

1. $\lim_{n\to\infty} \phi_n(\theta)$ exists for all $\theta \in \mathbb{R}$, and
2. $\lim_{n\to\infty} \phi_n(\theta) = \phi(\theta)$ is continuous at $\theta = 0$.

Then there exists a distribution function F with characteristic function ϕ and $F_n \to^{\mathcal{D}} F$.

Continuous Random Variables

Consider families of continuous random variables and let $\{F_X(\cdot|\theta_n); n \geq 1\}$ be a sequence of distributions from that family. Then

$$F_X(\cdot|\theta_n) \to^{\mathcal{D}} F_X(\cdot|\theta) \quad \text{if and only if} \quad \theta_n \to \theta.$$

This applies to beta, gamma, Pareto (α, β), chi-square (ν), exponential (λ), chi-square, t (ν), exponential (λ), F (ν_1, ν_2), normal, log-normal (μ, σ), logistic (μ, s) and uniform (a, b) families of random variables.

In each case,

$$\lim_{n\to\infty} \phi_{\theta_n}(z) = \phi_\theta(z).$$

and $\phi_\theta(\theta)$ is continuous at $\theta = 0$.

Moment Generating Functions

In using the characteristic function to establish convergence in distribution, we must work with the issues of the logarithm on the complex plane $\mathbb{C}$. In particular, no continuous definition for the logarithm exists whose domain is all of $\mathbb{C}$.

A alternative is the moment generating function or the Laplace transform. For a probability distribution F on $\mathbb{R}^d$,

$$M(t) = \int e^{\langle t,x \rangle} \, dF(x) = Ee^{\langle t,X \rangle}$$

where X is a random variable with distribution function F. Because the Laplace transform is one-to-one $\{e^{i\langle \theta,x \rangle}; \theta \in \mathbb{R}^d\}$ is a separating class of functions. Here is the corresponding result.

Moment Generating Functions

Theorem. Let $\{F_n; n \geq 1\}$ be a sequence of distributions on $\mathbb{R}$ with corresponding moment generating function $\{M_n; n \geq 1\}$ satisfying

1. $\lim_{n\to\infty} M_n(t)$ exists for all $t \in (-\delta, \delta), \delta > 0$, and
2. $\lim_{n\to\infty} M_n(t) = M(t)$ is continuous at $t = 0$.

Then there exists a distribution functtion F with moment generating function M and $F_n \to^{\mathcal{D}} F$.

This is not as general as the theorem using characteristic functions. However, taking the logarithm $K_n(t) = \ln M_n(t)$ is straightforward and we can replace the moment generating function with the cumulant generating function in the theorem above.