Convergence: Suppose that $a_1, \dots, a_n, \dots$ are a sequence of real numbers.

$a_n \to \alpha$ as $n \to \infty$ ← Definition

For any $\varepsilon > 0$ there is some N such that $|a_n - \alpha| < \varepsilon$ for $n \geq N$.

Limiting Distributions

We introduce the mode of convergence for a sequence of random variables, and discuss the convergence in probability and in distribution. The concept of convergence leads us to the two fundamental results of probability theory: Law of large number and central limit theorem.

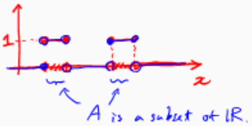
Convergence: Suppose that $X_1, X_2, \dots, X_n, \dots$ are a sequence of random variables.

$X_n \xrightarrow{?} X$ or $X_n \to \alpha$ as $n \to \infty$?

Indicator function.

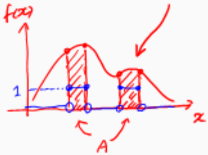
We can define an *indicator function* I_A of a subset A of real line as

$$I_A(x) = \begin{cases} 1 & x \in A; \\ 0 & \text{otherwise.} \end{cases}$$



Let X be a random variable having the pdf $f(x)$. Then we can express the probability

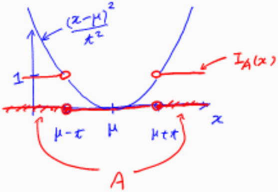
$$P(X \in A) = \int_A f(x)\,dx = \int_{-\infty}^{\infty} I_A(x) f(x)\,dx$$



Chebyshev's inequality.

Suppose that X has the mean $\mu = E[X]$ and the variance $\sigma^2 = \mathrm{Var}(X)$, and that we choose the particular subset

$$A = \{x : |x - \mu| > t\}.$$



Then we can observe that

$$I_A(x) \le \frac{1}{t^2}(x - \mu)^2 \quad \text{for all } x, \qquad (*)$$

which can be used to derive

$$P(|X - \mu| > t) = P(X \in A) = \int_{-\infty}^{\infty} I_A(x) f(x)\, dx$$

$$\le \frac{1}{t^2} \int_{-\infty}^{\infty} (x - \mu)^2 f(x)\, dx = \frac{1}{t^2} \mathrm{Var}(X) = \frac{\sigma^2}{t^2}.$$

The inequality $P(|X - \mu| > t) \le \sigma^2/t^2$ is called the *Chebyshev's inequality*.

$$E[(X-\mu)^2] = E[(X-E(X))^2] = \mathrm{Var}(X)$$

Convergence in probability.

Let $Y_1, Y_2, \ldots$ be a sequence of random variables. Then we can define the event $A_n = \{|Y_n - \alpha| > \varepsilon\}$ and the probability $P(A_n) = P(|Y_n - \alpha| > \varepsilon)$. We say that the sequence Y_n converges to α in probability if we have

$$\lim_{n\to\infty} P(|Y_n - \alpha| > \varepsilon) = 0,$$

for any choice of $\varepsilon > 0$. In general, the convergent value can be a random variable Y, and we say that Y_n converges to Y in probability if it satisfies

$$\lim_{n\to\infty} P(|Y_n - Y| > \varepsilon) = 0,$$

for any choice of $\varepsilon > 0$.

Laws of large numbers.

Now let $X_1, X_2, \dots$ be a sequence of iid random variables with mean μ and variance σ^2. Then for each $n = 1, 2, \dots$, the sample mean

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

has the mean μ and the variance σ^2/n. Given a fixed $\varepsilon > 0$, we can apply the Chebyshev's inequality to obtain

$$P(|\bar{X}_n - \mu| > \varepsilon) \leq \frac{\sigma^2/n}{\varepsilon^2} \to 0 \quad \text{as } n \to \infty.$$

Laws of large numbers, continued.

We have established the assertion

"the sample mean $\bar{X}_n$ converges to μ in probability"

which is called the *weak law of large numbers*. A stronger result of the above statement, "the sample mean $\bar{X}_n$ converges to μ almost surely," was established by Kolmogorov, and is called the *strong law of large numbers*.

Convergence in distribution.

Let $X_1, X_2, \ldots$ be a sequence of random variables having the cdf's $F_1, F_2, \ldots$, and let X be a random variable having the cdf F. Then we say that the sequence X_n converges to X in distribution (in short, X_n converges to F) if

$$\lim_{n\to\infty} F_n(x) = F(x) \quad \text{for every } x \text{ at which } F(x) \text{ is continuous.} \tag{7.1}$$

Convergence in distribution, continued.

Convergence in distribution does not necessarily imply convergence in probability.

Example

Let Y be a uniform random on $[-1/2, 1/2]$, and let $X_1 = Y, X_2 = -Y, X_3 = Y, X_4 = -Y, \cdots$. Then we have

$$F_n(t) = F_Y(t) = \begin{cases} 0 & \text{if } t < -1/2; \\ t + 1/2 & \text{if } -1/2 \leq t \leq 1/2; \\ 1 & \text{if } t > 1/2. \end{cases}$$

Then clearly X_n converges to Y in distribution. Does X_n converges to Y in probability?

Solution to Example 1.

Let $0 < \varepsilon < 1$ be arbitrary. Since we have

$$P(|X_n - Y| > \varepsilon) = \begin{cases} 0 & \text{if } n \text{ is odd;} \\ P(|2Y| > \varepsilon) = 1 - \varepsilon & \text{if } n \text{ is even,} \end{cases}$$

$\lim_{n\to\infty} P(|X_n - X| > \varepsilon)$ does not exist. Thus, X_n does not converge to Y in probability. In fact, the limit in probability does not exist.

Convergence in distribution, continued.

Convergence in distribution does not necessarily establish $\lim_{n\to\infty} F_n(x) = F(x)$ for every point.

Example

Let X_n be a uniform random on $[-1/n, 1/n]$, and let $X \equiv 0$ be the constant.

$$F_n(t) = \begin{cases} 0 & \text{if } t < -1/n; \\ (nt+1)/2 & \text{if } -1/n \le t \le 1/n; \\ 1 & \text{if } t > 1/n. \end{cases}$$

Does X_n converges to X in distribution? In probability?

Solution to Example 2.

Observe that the constant $X \equiv 0$ has the cdf

$$F(t) = \begin{cases} 0 & \text{if } t < 0; \\ 1 & \text{if } t \geq 0. \end{cases}$$

If $t < 0$ then $\lim_{n\to\infty} F_n(t) = 0 = F(t)$; if $t > 0$ then $\lim_{n\to\infty} F_n(t) = 1 = F(t)$. However, we can observe

$$\lim_{n\to\infty} F_n(0) = 1/2 \neq F(0).$$

Since $F(t)$ is not continuous at $t = 0$, X_n converges to X in distribution. Since for any $\varepsilon > 0$ we have

$$\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = \lim_{n\to\infty} P(|X_n| > \varepsilon) = 0,$$

it converges in probability.

Alternative criterion for convergence in distribution.

Let $Y_1, Y_2, \ldots$ be a sequence of random variables, and let Y be a random variable. Then the sequence Y_n converges to Y in distribution if and only if

$$\lim_{n \to \infty} E[g(Y_n)] = E[g(Y)]$$

for every continuous function g satisfying $E[|g(Y_n)|] < \infty$ for all n. By setting $g(y) = e^{ty}$, we obtain the moment generating function $M_Y(t) = E[g(Y)]$. Thus, if Y_n converges to Y in distribution then

$$\lim_{n \to \infty} M_{Y_n}(t) = M_Y(t)$$

when $M_{Y_n}(t)$ exists for all n.

Alternative criterion, continued.

Convergence in (7.1) is completely characterized in terms of the distributions $F_1, F_2, \ldots$ and F. Recall that the distributions $F_1(x), F_2(x), \ldots$ and $F(x)$ are uniquely determined by the respective moment generating functions, say $M_1(t), M_2(t), \ldots$ and $M(t)$. Furthermore, we have an "equivalent" version of the convergence in terms of the m.g.f's.

Theorem

If

$$\lim_{n \to \infty} M_n(t) = M(t) \quad \textit{for every } t \textit{ (around } 0\textit{)}, \tag{7.2}$$

then the corresponding distributions F_n's converge to F in the sense of (7.1).

Slutsky's theorem.

Although the following theorem is technical, it plays a critical role later in asymptotic theory of statistics.

Theorem

Let $Z_1, Z_2, \ldots$ and $W_1, W_2, \ldots$ be two sequences of random variables, and let c be a constant value. Suppose that the sequence Z_n converges to Z in distribution, and that the sequence W_n converges to c in probability. Then

1. *The sequence $Z_n + W_n$ converges to $Z + c$ in distribution.*
2. *The sequence $Z_n W_n$ converges to cZ in distribution.*
3. *Assuming $c \neq 0$, the sequence Z_n / W_n converges to Z / c in distribution.*

Central limit theorem.

Central limit theorem (CLT) is the most important theorem in probability and statistics.

Theorem

Let $X_1, X_2, \ldots$ be a sequence of iid random variables having the common distribution F with mean 0 and variance 1. Then

$$Z_n = \sum_{i=1}^{n} X_i \Big/ \sqrt{n} \quad n = 1, 2, \ldots$$

converges to the standard normal distribution.

A sketch of proof.

Let M be the m.g.f. for the distribution F, and let M_n be the m.g.f. for the random variable Z_n for $n = 1, 2, \ldots$. Since X_i's are iid random variables, we have

$$M_n(t) = E\left[e^{tZ_n}\right] = E\left[e^{\left(\frac{t}{\sqrt{n}}\right)(X_1+\cdots+X_n)}\right]$$

$$= E\left[e^{\left(\frac{t}{\sqrt{n}}\right)X_1}\right] \times \cdots \times E\left[e^{\left(\frac{t}{\sqrt{n}}\right)X_n}\right] = \left[M\left(\frac{t}{\sqrt{n}}\right)\right]^n$$

A sketch of proof, continued.

Observing that $M'(0) = E[X_1] = 0$ and $M''(0) = E[X_1^2] = 1$, we can apply a Taylor expansion to $M\left(\frac{t}{\sqrt{n}}\right)$ around 0.

$$
\begin{aligned}
M\left(\frac{t}{\sqrt{n}}\right) &= M(0) + \frac{t}{\sqrt{n}}M'(0) + \frac{t^2}{2n}M''(0) + \varepsilon_n(t) \\
&= 1 + \frac{t^2}{2n} + \varepsilon_n(t).
\end{aligned}
$$

Recall that $\lim_{n\to\infty}\left(1 + \frac{a}{n}\right)^n = e^a$; in a similar manner, we can obtain

$$
\lim_{n\to\infty} M_n(t) = \lim_{n\to\infty}\left[1 + \frac{t^2/2}{n} + \varepsilon_n(t)\right]^n = e^{t^2/2},
$$

which is the m.g.f. for the standard normal distribution.

Central limit theorem, modified.

Mean and variance for a sequence can be arbitrary in general, and CLT can be modified for them.

Theorem

Let $X_1, X_2, \ldots$ be a sequence of iid random variables having the common distribution F with mean μ and variance σ^2. Then

$$Z_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} = \frac{\sum_{i=1}^{n} (X_i - \mu)/\sigma}{\sqrt{n}} \quad n = 1, 2, \ldots$$

converges to the standard normal distribution.

Normal approximation.

Let $X_1, X_2, \ldots, X_n$ be finitely many iid random variables with mean μ and variance σ^2. If the size n is adequately large, then the distribution of the sum

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

can be approximated by the normal distribution with parameter $(n\mu, n\sigma^2)$. A general rule for "adequately large" n is about $n \geq 30$, but it is often good for much smaller n.

Normal approximation, continued.

Similarly the sample mean

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

has approximately the normal distribution with parameter $(\mu, \sigma^2/n)$. Then the following application of normal approximation to probability computation can be applied.

Random variable	Approximation	Probability
$Y = X_1 + \cdots + X_n$	$N(n\mu, n\sigma^2)$	$\Phi\left(\frac{b-n\mu}{\sigma\sqrt{n}}\right) - \Phi\left(\frac{a-n\mu}{\sigma\sqrt{n}}\right)$
$\bar{X} = \frac{X_1+\cdots+X_n}{n}$	$N\left(\mu, \frac{\sigma^2}{n}\right)$	$\Phi\left(\frac{b-\mu}{\sigma/\sqrt{n}}\right) - \Phi\left(\frac{a-\mu}{\sigma/\sqrt{n}}\right)$

Example

An insurance company has 10,000 automobile policyholders. The expected yearly claim per policyholder is \$240 with a standard deviation of \$800. Find approximately the probability that the total yearly claim exceeds \$2.7 million. Can you say that such event is statistically highly unlikely?

Example

An insurance company has 10,000 automobile policyholders. The expected yearly claim per policyholder is \$240 with a standard deviation of \$800. Find approximately the probability that the total yearly claim exceeds \$2.7 million. Can you say that such event is statistically highly unlikely?

Let $X_1, \ldots, X_{10000}$ be normally distributed with mean $\mu = 240$ and $\sigma = 800$. Then $Y = X_1 + \cdots + X_{10000}$ is approximately normally distributed with mean $n\mu = 2,400,000$ and standard deviation $\sqrt{n}\sigma = 80,000$. Thus, we obtain

$$P(Y > 2,700,000) \approx 1 - \Phi(3.75) \approx 0,$$

which is highly unlikely.

Normal approximation to binomial distribution.

Suppose that $X_1, \ldots, X_n$ are iid Bernoulli random variables with the mean $p = E(X)$ and the variance $p(1-p) = \mathrm{Var}(X)$. If the size n is adequately large, then the distribution of the sum

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

can be approximated by the normal distribution with parameter $(np, np(1-p))$. Thus, the normal distribution $N(np, np(1-p))$ approximates the binomial distribution $B(n, p)$. A general rule for "adequately large" n is to satisfy $np \geq 5$ and $n(1-p) \geq 5$.

Normal approximation to binomial distribution, continued.

Let Y be a binomial random variable with parameter (n, p), and let Z be a normal random variable with parameter $(np, np(1-p))$. Then the distribution of Y can be approximated by that of Z. Since Z is a continuous random variable, the approximation of probability should improve when the following formula of *continuity correction* is considered.

$$P(i \le Y \le j) \approx P(i - 0.5 \le Z \le j + 0.5)$$
$$= \Phi\left(\frac{j + 0.5 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{i - 0.5 - np}{\sqrt{np(1-p)}}\right).$$

Exercises

Problem

Let S_n^2 denote the sample variance of random sample of size n from normal distribution with mean μ and variance σ^2.

1. *Find $E[S_n^2]$ and $\mathrm{Var}(S_n^2)$.*
2. *Show that S_n^2 converges to σ^2 in probability*

Problem

Let $X_1, \ldots, X_n$ be a random sample of size n from the common pdf $f(x) = e^{-(x-\theta)}$ if $\theta \leq x < \infty$; otherwise, $f(x) = 0$, and let $X_{(1)}$ be the first order statistic, and $Z_n = n(X_{(1)} - \theta)$.

1. *Find the pdf for $X_{(1)}$, and the mgf $M(t)$.*
2. *Find the mgf $M_n(t)$ for Z_n.*
3. *Does Z_n converge in distribution? If so, find the limiting distribution.*

Problem

Suppose that Z_n has a Poisson distribution with parameter $\lambda = n$. Then show that $(Z_n - n)/\sqrt{n}$ converges to the standard normal distribution.

Problem

An actual voltage of new a 1.5*-volt battery has the probability density function*

$$f(x) = 5, \quad 1.4 \leq x \leq 1.6.$$

Estimate the probability that the sum of the voltages from 120 *new batteries lies between* 170 *and* 190 *volts.*

Problem

The germination time in days of a newly planted seed has the probability density function

$$f(x) = 0.3e^{-0.3x}, \quad x \geq 0.$$

If the germination times of different seeds are independent of one another, estimate the probability that the average germination time of 2000 *seeds is between* 3.1 *and* 3.4 *days.*

Problem

Calculate the following probabilities by using normal approximation with continuity correction.

1. *Let X be a binomial random variable with $n = 10$ and $p = 0.7$. Find $P(X \geq 8)$.*
2. *Let X be a binomial random variable with $n = 15$ and $p = 0.3$. Find $P(2 \leq X \leq 7)$.*
3. *Let X be a binomial random variable with $n = 9$ and $p = 0.4$. Find $P(X \leq 4)$.*
4. *Let X be a binomial random variable with $n = 14$ and $p = 0.6$. Find $P(8 \leq X \leq 11)$.*

Optional problems

Convergence in distribution.

Definition

A function $f(x)$ is said to be *bounded* if there exists some $M > 0$ such that $|f(x)| \leq M$ for all x. A function $f(x)$ is called *uniformly continuous* if for any $\varepsilon > 0$ there exists some $\delta > 0$ such that $|f(x) - f(y)| < \varepsilon$ whenever $|x - y| < \delta$.

Let $X_1, X_2, \ldots$ be a sequence of random variables, and let X be a random variable. In order to see that X_n converges to X in distribution it suffices to show that

$$\lim_{n \to \infty} E[g(X_n)] = E[g(X)]$$

for every bounded and uniformly continuous function f.

Problem

Choose a bounded and uniformly continuous function f so that $|f(x)| \leq M$ for all x and $|f(x) - f(y)| < \varepsilon$ whenever $|x - y| < \delta$. Let X be a random variable, and let $A_\delta = \{y : |y - c| < \delta\}$. Suppose that a sequence Y_n of random variables converges to a constant c in probability.

1. *Show that $|f(X + Y_n) - f(X + c)| \leq 2M$.*
2. *Show that $|f(X + Y_n) - f(X + c)| I_A(Y_n) \leq \varepsilon$.*
3. *Show that $|E[f(X + Y_n)] - E[f(X + c)]| \leq 2MP(|Y_n - c| \geq \delta) + \varepsilon$*
4. *The above result implies that $\lim_{n\to\infty} |E[f(X + Y_n)] - E[f(X + c)]| \leq \varepsilon$ for every $\varepsilon > 0$, and therefore, that $\lim_{n\to\infty} |E[f(X + Y_n)] - E[f(X + c)]| = 0$. Then argue that $X + Y_n$ converges to $X + c$ in distribution.*

Problem

Let $Z_1, Z_2, \ldots$ and $W_1, W_2, \ldots$ be two sequences of random variables, and let c be a constant value. Suppose that Z_n converges to Z in distribution, and that W_n converges to c in probability.

1. *Choose a bounded and uniformly continuous function f so that $|f(x)| \leq M$ for all x and $|f(x) - f(y)| < \varepsilon$ whenever $|x - y| < \delta$. Then show that*

$$\begin{aligned} &|E[f(Z_n + W_n)] - E[f(Z + c)]| \\ &\leq |E[f(Z_n + c)] - E[f(Z + c)]| + |E[f(Z_n + W_n)] - E[f(Z_n + c)]| \\ &\leq |E[f(Z_n + c)] - E[f(Z + c)]| + 2MP(|W_n - c| \geq \delta) + \varepsilon \end{aligned}$$

2. *The above result implies that $\lim_{n\to\infty} |E[f(Z_n + W_n)] - E[f(Z + c)]| \leq \varepsilon$, and therefore, that $\lim_{n\to\infty} |E[f(Z_n + W_n)] - E[f(Z + c)]| = 0$. Then argue that $Z_n + W_n$ converges to $Z + c$ in distribution (Slutsky's theorem).*

Answers to exercises

Problem 8.

Recall that $W_{n-1} = (n-1)S_n^2/\sigma^2$ is a chi-square distribution with $(n-1)$ degrees of freedom, and that $E[W_{n-1}] = (n-1)$ and $\mathrm{Var}(W_{n-1}) = 2(n-1)$.

(1) $E[S_n^2] = E\left[\frac{\sigma^2 W_{n-1}}{(n-1)}\right] = \frac{\sigma^2}{n-1}E[W_{n-1}] = \sigma^2$, and

$$\mathrm{Var}(S_n^2) = \mathrm{Var}\left(\frac{\sigma^2 W_{n-1}}{(n-1)}\right) = \left(\frac{\sigma^2}{n-1}\right)^2 \mathrm{Var}(W_{n-1}) = \frac{2\sigma^4}{n-1}.$$

(2) By Chebyshev's inequality we obtain

$$P\left(|S_n^2 - \sigma^2| > \varepsilon\right) \leq \frac{2\sigma^4}{(n-1)\varepsilon^2} \to 0 \text{ as } n \to \infty.$$

Problem 9.

(1) $X_{(1)}$ has the pdf $f_{(1)}(x) = ne^{-n(x-\theta)}$ if $x \geq \theta$; otherwise, $f_{(1)}(x) = 0$. Then we obtain

$$M(t) = \int_{\theta}^{\infty} e^{tx} ne^{-n(x-\theta)} dx = ne^{t\theta} \int_{0}^{\infty} e^{-(n-t)u} du = \frac{ne^{t\theta}}{n-t}$$

for $t < n$.

(2) We obtain

$$M_n(t) = E\left[e^{tn(X_{(1)}-\theta)}\right] = e^{nt\theta} M(nt) = \frac{1}{1-t}$$

for $t < 1$.

(3) Since $\lim_{n\to\infty} M_n(t) = \frac{1}{1-t}$, the limiting distribution is an exponential with $\lambda = 1$.

Problem 10.

Let $X_1, X_2, \ldots$ be iid random variables having a Poisson distribution with $\lambda = 1$. Then we have $Z_n = X_1 + \cdots + X_n$. By CLT we obtain

$$Z_n = \frac{\sum_{i=1}^{n} X_i - n}{\sqrt{n}}$$

converges to the standard normal distribution.

Problem 11.

Let $X_1, \dots, X_{120}$ be the voltage of each battery, having the pdf $f(x)$. Since $\mu = E[X_i] = 1.5$ and $\sigma^2 = E[X_i^2] - \mu^2 = 0.0033$, the sum $Y = \sum_{i=1}^{120} X_i$ is approximately distributed as $N(120\mu, 120\sigma^2) = N(180, 0.396)$. Thus, we obtain

$$P(170 \le Y \le 190) = \Phi\left(\frac{190 - 180}{\sqrt{0.396}}\right) - \Phi\left(\frac{170 - 180}{\sqrt{0.396}}\right) \approx 1$$

Problem 12.

Let $X_1, \ldots, X_{2000}$ be the germination time of individual seed, having the exponential distribution with $\lambda = 0.3$. Since $\mu = E[X_i] = 1/0.3 \approx 3.33$ and $\sigma^2 = \mathrm{Var}(X_i) = 1/0.3^2 \approx 11.11$, the sample mean $\bar{X}$ is approximately distributed as $N(\mu, \sigma^2/n) = N(3.33, 0.0056)$. Thus,

$$P(3.1 \leq \bar{X} \leq 3.4) = \Phi\left(\frac{3.4 - 3.33}{\sqrt{0.0056}}\right) - \Phi\left(\frac{3.1 - 3.33}{\sqrt{0.0056}}\right) \approx 0.825$$

Problem 13.

(1) $P(X \geq 8) = 1 - \Phi\left(\frac{7.5 - (10)(0.7)}{\sqrt{(10)(0.7)(0.3)}}\right) \approx 0.365$

(2) $P(2 \leq X \leq 7) = \Phi\left(\frac{7.5 - (15)(0.3)}{\sqrt{(15)(0.3)(0.7)}}\right) - \Phi\left(\frac{1.5 - (15)(0.3)}{\sqrt{(15)(0.3)(0.7)}}\right) \approx$
0.909

(3) $P(X \leq 4) = \Phi\left(\frac{4.5 - (9)(0.4)}{\sqrt{(9)(0.4)(0.6)}}\right) \approx 0.7299$

(4) $P(8 \leq X \leq 11) =$
$\Phi\left(\frac{11.5 - (14)(0.6)}{\sqrt{(14)(0.6)(0.4)}}\right) - \Phi\left(\frac{7.5 - (14)(0.6)}{\sqrt{(14)(0.6)(0.4)}}\right) \approx 0.6429$