Example

Let $W_0, W_1, \ldots$ be iid normal, $N(0,1)$

Let $X_{-1}=0$, $X_n = 0.9X_{n-1} + W_n$, $n \geq 0$ ⇒ Solution next page

In what sense does X_n converge, if at all?

Hard to guess limit, if one even exists.

Cauchy criteria; Cauchy sequence $\{x_n\}$: $\lim_{m,n\to\infty} |x_m - x_n| = 0$

Since $\mathbb{R}$ is complete, every Cauchy sequence converges to a finite limit as $n \to \infty$

A convergent sequence is Cauchy: if $\{x_n\}$ has limit x, the triangle inequality gives

$$|x_n - x_m| \leq \underbrace{|x_n - x|}_{\to 0} + \underbrace{|x_m - x|}_{\to 0} \Rightarrow |x_n - x_m| \to 0$$

Same result carries over to RVs

example

$$X_n \xrightarrow{P} X \text{ iff } \forall \varepsilon > 0$$
$$\lim_{m,n\to\infty} P\{|X_m - X_n| \geq \varepsilon\} = 0$$

Proof

$X_m \quad X \quad X_n \quad X$

$\geq 2\varepsilon$

Suppose $X_n \xrightarrow{P} X_\infty$

$$P\{|X_m - X_n| \geq 2\varepsilon\} \leq P\{|X_m - X_\infty| \geq \varepsilon \text{ or } |X_n - X_\infty| \geq \varepsilon\}$$
$$\leq \underbrace{P\{|X_m - X_\infty| \geq \varepsilon\}}_{\to 0} + \underbrace{P\{|X_n - X_\infty| \geq \varepsilon\}}_{\to 0}$$

Suppose $\lim_{m,n\to\infty} P\{|X_m - X_n| \geq \varepsilon\} = 0$

select an increasing sequence of indices k_i s.t.

$$P\{|X_{k_{i+1}} - X_{k_i}| \geq 2^{-i}\} \leq 2^{-i}$$

$$\sum_i 2^{-i} < \infty$$

Borel-Cantelli gives

$$P\{|X_{k_{i+1}} - X_{k_i}| \leq 2^{-i} \text{ for all } i \geq \text{some threshold}\} = 1$$

So, we have convergence as $\mathbb{R}$ is complete: $X_{k_i} \xrightarrow{a.s.} X_\infty$.

Corollary If $X_n \xrightarrow{P} X_\infty$, then there exists a subsequence $(X_{k_i} : i \geq 1)$ such that $\lim_{i\to\infty} X_{k_i} = X_\infty$ a.s.

$$P\{|X_n - X_{n-1}| \geq 2\} = P\{|0.9X_{n-1} + W_n - X_{n-1}| \geq 2\} = P\{|-0.1X_{n-1} + W_n| \geq 2\}$$

$$= P\{-0.1X_{n-1} + W_n \geq 2,\ -0.1X_{n-1} + W_n \leq -2\}$$

$$\geq P\{X_{n-1} \leq 0, W_n \geq 2\} + P\{X_{n-1} \geq 0, W_n \leq -2\}$$

$$= P\{X_{n-1} \leq 0\} \underline{P\{W_n \geq 2\}} + P\{X_{n-1} \geq 0\} \underline{P\{W_n \leq -2\}}$$

$$= (\underbrace{P\{X_{n-1} \leq 0\} + P\{X_{n-1} \geq 0\}}_{1}) P\{W_n \geq 2\}$$

$$= P\{W_n \geq 2\} \geq 0.02$$

But $P\{|X_n - X| \geq 1\} + P\{|X_{n-1} - X| \geq 1\}$

$$\geq P\{|X_n - X| \geq 1 \text{ or } |X_{n-1} - X| \geq 1\}$$

$$\geq P\{|X_n - X_{n-1}| \geq 2\} \geq 0.02$$

$P\{|X_n - X| \geq \varepsilon\}$ does not converge to zero

$X_0 = W_0 \quad \mathcal{N}(0,1)$

$X_1 = 0.9X_0 + W_1 \sim \mathcal{N}(0, 1.81) \qquad \mathrm{var}(X_1) = (0.9)^2 \mathrm{var}(X_0) + 1$

$X_2 = 0.9X_1 + W_2 \sim \mathcal{N}(0, 0.81^2 \cdot 1.81 + 1) \qquad \mathrm{var}(X_2) = (0.9)^2 \mathrm{var}(X_1) + 1$

$X_n \xrightarrow{d} X \qquad X \sim \mathcal{N}(0, 5.263)$

Interesting result

Suppose $\{X_n\}$ is a sequence of Gaussian RVs and $X_n \to X$ as $n \to \infty$ in any of the four manners discussed. Then X_∞ is also a Gaussian RV

Correlation version of the Cauchy criteria

$\{X_n\}$ s.t. $E[X_n^2] < \infty \ \forall n$

$\exists$ a RV X s.t. $X_n \xrightarrow{m.s.} X$ iff $\lim_{m,n\to\infty} E[X_n X_m]$ exists and is finite. If $X_n \xrightarrow{m.s.} X$, then $\lim_{m,n\to\infty} E[X_n X_m] = E[X^2]$

Proof: Suppose that $\lim_{m,n\to\infty} E[X_n X_m] = c$, c finite const.

$$E[(X_n-X_m)^2] = E[X_n^2 - 2X_nX_m + X_m^2] \underset{n,m\to\infty}{\to} c - 2c + c = 0$$

X_n is Cauchy in the MS sense,
so $X_n \underset{m.s}{\to} X$ as $\mathbb{R}$ is complete

$X_n \overset{m.s.}{\to} X$ assumed next

$$E[X_m X_n] = E[(X+(X_m-X))(X+(X_n-X))]$$
$$= E[X^2 + X(X_m-X) + X(X_n-X) + (X_m-X)(X_n-X)]$$

Cauchy-Schwarz gives

$$E[(X_m-X)X] \le \cancel{E[(X_m-X)^2]^{1/2}}^{\,0}\ \overbrace{E[X^2]^{1/2}}^{\text{finite}} \to 0$$

$E[(X_n-X)X] \to 0$ according to same trick

$E[(X_m-X)(X_n-X)] \to 0$ same as above

Thus $E[X_mX_n] \to E[X^2]$

<u>The Law of Large Numbers</u>

a) Let $X_1, X_2, \ldots,$ be a sequence of RVs with finite mean m,
Let $S_n = X_1 + X_2 + \cdots + X_n$

Then $\frac{S_n}{n} \overset{m.s.}{\to} m$ if $\exists c$ s.t. $Var(X_i) \le c\ \forall i$, and $Cov(X_i, X_j) = 0$ for $i \ne j$

b) $\frac{S_n}{n} \overset{P}{\to} m$ if X_i's are iid

c) $\frac{S_n}{n} \overset{a.s.}{\to} m$ if X_i's are iid (Hard to prove)

a) $E[(\frac{S_n}{n} - m)^2] = var(\frac{S_n}{n}) = \frac{1}{n^2} var(S_n) = \frac{1}{n^2} \sum_i \sum_j Cov(X_i, X_j)$

$= \frac{1}{n^2} \sum_i var(X_i) \le \frac{const.}{n} \to 0$ as $n \to \infty$

b) Let us find the characteristic function of $\frac{X_i}{n}$ $\forall i$

$$E[\exp(ju\tfrac{X_i}{n})] = \phi_X(\tfrac{u}{n})$$

so that $\phi_{\frac{S_n}{n}}(u) = (\phi_X(\frac{u}{n}))^n$

Since $E[X_1]=m$, Φ is differentiable with $\Phi_X(0)=1$, $\Phi_X'(0)=jm$

Taylor's expansion gives

$$\Phi_X\left(\frac{u}{n}\right) = 1+\frac{u}{n}\left(Re\left(\Phi_X'(u_n)\right)+ j\, Im\left(\Phi_X'(v_n)\right)\right)$$

(Lagrange theorem, also known as Remainder theorem)
for some $u_n, v_n \in [0, \frac{u}{n}]$, $\forall n$

Suppose that $f(x)$ is $(n+1)$-diff
Then, $R_n = f - T_n \Rightarrow$
T_n - Taylor series at a order n,
$* = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{n+1}$
where $c \in [a,x]$

Hence, $Re(\Phi_X'(u_n)) + j\, Im(\Phi_X'(v_n)) \to jm$ as $n \to \infty$

Hence $\Phi_X^n\left(\frac{u}{n}\right) \to \lim_{n\to\infty}\left(1+\frac{u}{n}(jm)\right)^n = \exp(jum)$

$\exp(jum)$ is the MGF of a degenerate RV equal to m w.p. 1

Hence $\frac{S_n}{n} \xrightarrow{d} m$

But convergence in distribution to a const. implies convergence in probability.

Observe that if we had $E[X_i^2]<\infty$ in addition to $E[|X_i|]<\infty$ we could have applied Chebyshev's inequality to get the following:

X_i mean m, variance δ^2

$\frac{1}{n}S_n$ has mean m, $var\left(\frac{1}{n}S_n\right) = \frac{1}{n^2}n\delta^2 = \frac{\delta^2}{n}$

so that $P\left\{\left|\frac{1}{n}S_n - m\right| \geq \varepsilon\right\} \leq \frac{\delta^2}{n\varepsilon^2} \to 0$ as $n\to\infty$

Can get rid of finite variance requirement by using <u>truncation</u> argument (Khinchin)

Central limit theorem

Suppose that $X_1, X_2, \ldots$ are iid with finite mean μ and variance δ^2.

Then, $\frac{S_n - n\mu}{\sqrt{n}}$ converges in distribution to $\mathcal{N}(0,\delta^2)$ as $n\to\infty$

Similar argument as above, based on the characteristic function

What the CLT says is that, for the case $\mu = 0$, $\sigma^2 = 1$, is that

$$P\{a < \frac{S_n}{\sqrt{n}} < b\} \to \Phi(b) - \Phi(a)$$
$$= \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\, dx$$

Moivre–Laplace : Normal distribution may be used as an approximation to the binomial distribution for certain parameter choices

As $n \to \infty$, for k close to np

$$\binom{n}{k} p^k q^{n-k} \sim \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}}, \quad \begin{matrix} p+q=1 \\ p,q \geq 0 \end{matrix}$$

ratio tends to 1 as $n \to \infty$

Berry–Esseen theorem gives the rate at which convergence takes place (i.e. how large of an n is needed)

Let $X_1, \ldots X_n, \ldots$ be iid, $E[X_i] = 0$ $\forall i$, $E[X_i^2] = \sigma^2 > 0$, $E[|X_1|^3] = \rho < \infty$

$$\Sigma_n = \frac{X_1 + \cdots + X_n}{\sigma\sqrt{n}} = \frac{S_n}{\sigma\sqrt{n}}$$

then there exists a const. C so that

$$|F_n(x) - \Phi(x)| \leq \frac{C\rho}{\sigma^3\sqrt{n}} \quad \forall x, n$$

CDF of Σ_n CDF of standard normal

estimate of C : $C \leq 0.47$

Chernoff Bound and Large Deviations

The WLLNs implies that $P\{\frac{S_n}{n} \geq a\} \underset{n\to\infty}{\to} 0$ for fixed $a > \mu$

Large deviation results show how quickly

$P\{\frac{S_n}{n} \geq a\}$ converges to zero as $n \to \infty$

Note characteristic function, moment generating function: difference in j in the exponent (Laplace versus Fourier)

$$M(\theta) = E[e^{\theta X_1}]$$

$\log M(\theta)$ = log moment generating function

(Characteristic function always exists, moment generating function requires the moments to exist; think of Laplace vs. Fourier tra

$S_n = X_1 + \cdots + X_n$

$$P\left\{\frac{S_n}{n} \geq a\right\} = P\{S_n \geq na\} = P\{S_n - na \geq 0\} = P\{\theta(S_n - na) \geq 0\}\Big|_{\theta>0}$$

$$= P\{e^{\theta(S_n-na)} \geq 1\}$$

$$\leq E[e^{\theta(X_1+\cdots+X_n-na)}] = e^{-na\theta} E[e^{\theta X_1}]$$

$$= \exp(-n[\theta a - \ln M(\theta)])$$

Select the "best" $\theta > 0$ to maximize

$$\theta a - \ln M(\theta)$$

Observations about the moment generating function

$\ln M(\theta)$ is convex

$\ln M(0) = 0$ as $M(0) = 1$

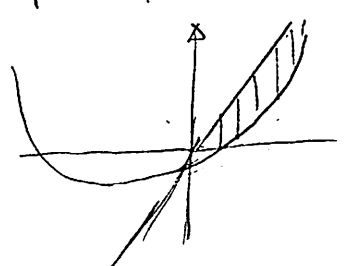
Assume that $M(\theta)$ is finite for some $\theta > 0$.

Then

$$\left.\frac{d \ln M(\theta)}{d\theta}\right|_{\theta=0} = \left.\frac{1}{E[e^{\theta X_1}]} E[X_1 e^{\theta X_1}]\right|_{\theta=0} = E[X_1]$$

slope at $\theta = 0$

slope of $\ln M(\theta)$ at $\theta = 0$



$$l(a) = \sup_{-\infty<\theta<\infty} \theta a - \ln M(\theta)$$

so that $P\left\{\frac{S_n}{n} \geq a\right\} \leq \exp(-n\,l(a))$

$a > E[X_1]$

Cramer's theorem

Assume that $E[X_1] < a < \infty$. For all $\varepsilon > 0$, $\exists\, n_\varepsilon$ s.t.

$$P\{\frac{S_n}{n} \geq a\} \geq \exp(-n\ell(a) + \varepsilon)$$

$\forall\, n \geq n_\varepsilon$. This implies that

$$\lim_{n\to\infty} \frac{1}{n} \ln P\{\frac{S_n}{n} \geq a\} = -\ell(a)$$

Examples

Let $X_1, X_2, \ldots$ be iid exponential RVs with parameter $\lambda = 1$. Then,

$$\ln M(\theta) = \ln \int_0^{+\infty} e^{\theta x} e^{-x}\, dx = \ln \int_0^{+\infty} e^{-x(1-\theta)}\, dx \Big|_{\theta<1} = \ln \Big[\frac{1}{1-\theta} \int_{-\infty}^{0} e^{-x} dx \Big]_{\theta<1}$$

$$= \ln \Big[\frac{1}{1-\theta} \Big] \Big|_{\theta<1} = -\ln(1-\theta) \quad \text{for } \theta < 1$$

For $\theta \geq 1$, we clearly have $\ln M(\theta) = \infty$

Therefore,

$$\ell(a) = \max_{\theta} \{ a\theta - \ln M(\theta) \}$$

$$= \max_{\theta<1} \{ a\theta + \ln(1-\theta) \}$$

For $a > 0$, have $0 = (a\theta + \ln(1-\theta))' = a + \frac{1}{1-\theta}(-1)$

$$\therefore\ a = \frac{1}{1-\theta} \Rightarrow \theta = 1 - \frac{1}{a}$$

Some remarks

Strong law of large numbers was proved by Kolmogorov in a slightly stronger form where the variables are independent, but not identically distributed

$$X_1, X_2, \ldots$$
$$\swarrow \quad \downarrow$$
$$\mu_1, \sigma_1^2 \quad \mu_2, \sigma_2^2, \ldots$$

$$m_n = \mu_1 + \cdots + \mu_n, \quad s_n^2 = \sigma_1^2 + \cdots + \sigma_n^2$$

Result relies on Kolmogorov's maximal inequality

$$P\{\max_{1 \le j \le n} |S_j - m_j| \ge t\} \le \frac{s_n^2}{t^2}$$

This gives rise to Kolmogorov's criterion:

If $\sum_{j=1}^{\infty} \frac{\sigma_j^2}{j^2}$ converges then $\frac{1}{n}(S_n - m_n) \xrightarrow{a.s.} 0$

<u>Convergence in distribution to a const $\Rightarrow$ convergence in prob to a constant</u>

$Y_n \xrightarrow{d} c \quad \Rightarrow \quad X_n \xrightarrow{P} c$ provided that c is a constant

Fix $\varepsilon > 0$. Invoke the portmanteau lemma

$\{X_n\} \xrightarrow{d} X$ iff $\limsup P\{X_n \in C\} \le P\{X \in C\}$ for all closed sets C

Let $B_\varepsilon(c)$ = open ball of radius ε around c

$$= \{x : |x - c| < \varepsilon\}$$

$B_\varepsilon^c(c)$ = complement of $B_\varepsilon(c)$, a closed set

$$P\{|X_n - c| \ge \varepsilon\} = P\{X_n \in B_\varepsilon^c(c)\}$$

Since we have convergence in distribution,

$$\lim_{n \to \infty} P\{|X_n - c| \ge \varepsilon\} \le \limsup_{n \to \infty} P\{X_n \in B_\varepsilon^c(c)\} \le P\{c \in B_\varepsilon^c(c)\} = 0$$

so that $X_n \xrightarrow{P} c$