Almost Sure Convergence

Let X_n $(X_1, \ldots)$ be random variables in a sequence (e.g. $X_n \sim N(0, \frac{1}{n})$) of random variables defined on a common probability space. Let X be another r.v. on that probability space. We say X_n <u>converges almost surely</u> <u>a.s.</u> to X (or converges to X with probability 1 or <u>WP1</u>, or strongly, or almost everywhere, a.e.) if

$$\lim_{n \to \infty} P(|X_m - X| < \varepsilon, \text{ all } m \geq n) = 1 \text{ every } \varepsilon > 0$$

If $X_n \neq X$ is not true infinitely often, there is almost sure convergence

for example, if $X_1, X_2, \ldots$ is a sequence of iid Unif $(0,1)$ and $X_{(n)}$ is a sequence of the maximums of the r.v.s $\leq n$, then

$$P(|X_{(n)} - 1| < \varepsilon, \forall n \geq m) = P(X_{(n)} \geq 1 - \varepsilon, \forall n \geq m)$$
$$= P(X_{(m)} \geq 1 - \varepsilon) = 1 - (1 - \varepsilon)^m \to 1 \text{ as } m \to \infty.$$

The maximum strictly increases, and always increases, the probability it doesn't surpass any $1 - \varepsilon$ is zero, so the probability the max is less than 1 i.o. is zero.

Convergence In Probability

Let X_n $(X_1, X_2, \ldots)$ be a sequence of random variables defined on a common probability space. Let X be another r.v. also defined on that probability space. X_n <u>converges in probability</u> to X if, for any ε greater than 0, it is true that

$$\forall \varepsilon > 0 \quad \lim_{n \to \infty} P(|X_n - X| < \varepsilon) = 1$$

In other words, the probability that the difference between X_n and X is larger than any fixed ε goes to zero as $n \to \infty$

For example, $X_n = N(\mu, \frac{\sigma^2}{n})$, $X = \mu$

$$X_n \xrightarrow{P} X$$

Because $\forall \varepsilon > 0 \ \lim_{n \to \infty} P(|\mu - N(\mu, \frac{\sigma^2}{n})| < \varepsilon)$ approaches 1 as $n \to \infty$ and the variance shrinks to zero

Example

Let X_n $(X_1, \dots)$ be a sequence of random variables such that $X_n = 1$ with probability $\frac{1}{n}$ and $X_n = 0$ with probability $1 - \frac{1}{n}$

First, note that $\sum_{i=1}^{n} \frac{1}{n} = \infty$, so $\sum_{i=1}^{n} P(X_n \neq 0) = \infty$

By the Borel-Cantelli lemma, this means $X_n = 1$ occurs infinitely often, so $X_n \overset{a.s.}{\not\to} 0$

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

$\lim_{n\to\infty} (1 - \frac{1}{n}) = 1 \quad \forall \varepsilon > 0$, so $X_n \xrightarrow{P} 0$

So X_n converge's to zero in probability, but not almost surely.

Convergence In Distribution

Let X_n $(X_1, X_2, \dots)$ be r.v.s, Consider their distribution functions F_{X_n} and the distribution function of a r.v. X, F_X, We say that X_n <u>converges in distribution</u> to X if

$$\lim_{n\to\infty} F_{X_n}(t) = F_X(t) \text{ at every point } t$$

continuous in F_X

let $X_n \sim N\left(u, \sigma^2 + \frac{1}{n}\right)$

$$\lim_{n\to\infty} \frac{1}{\sqrt{2\pi\left(\sigma^2+\frac{1}{n}\right)}} e^{\frac{(x-u)^2}{-2\left(\sigma^2+\frac{1}{n}\right)}} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-u)^2}{2\sigma^2}} \quad \text{So}$$

$$X_n \xrightarrow{d} N(u, \sigma^2)$$

Note that if $\lim_{n\to\infty} E[e^{X_n t}] = E[e^{Xt}]$

This convergence in MGF implies convergence in distribution

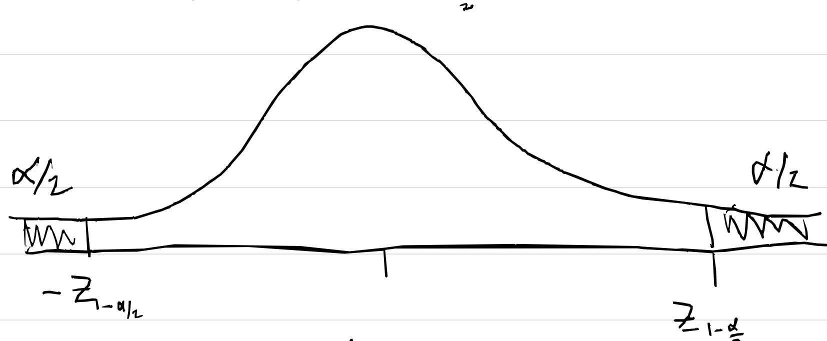
Bounded In Probability

A sequence of random variables X_n is said to be **bounded in probability** if, for any $\varepsilon > 0$, there exists a constant K such that $P(|X_n| > K) \leq \varepsilon \ \forall n$

In other words, the sequence is bounded in probability when the probability that $|X_n|$ is larger than some constant K doesn't "blow up", it's fixed below ε

For example, $X_1, X_2, \ldots$ iid $N(0,1)$

X_n is bounded in probability, since
$\forall n \quad P(|X_n| > Z_{1-\frac{\alpha}{2}}) \leq \alpha$ since



The sequence X_n $N(0, 1+n)$ wouldn't be bounded in Prob

O_p and o_p

For notational purposes the statement "$X_n \xrightarrow{P} X$" and "X_n is bounded in probability" are often abbreviated with "Big Oh" and "little Oh" notation

$$X_n = o_p(g(n)) \text{ means } \frac{X_n}{g(n)} \xrightarrow{P} 0$$

$$X_n = O_p(g(n)) \text{ means } \frac{X_n}{g(n)} \text{ is bounded in probability}$$

For example, CLT says that $\bar{X}_n \xrightarrow{d} N(\mu, \frac{\sigma^2}{n})$. As a result, we can say, using o_p notation that

$$\bar{X} = \mu + o_p(1) \text{ since } \bar{X}_n \xrightarrow{P} \mu \Leftrightarrow \frac{\bar{X} - \mu}{1} \xrightarrow{P} 0$$

We can also say with O_p notation that

$$\frac{\bar{X} - \mu}{\sigma} = O_p(n^{-1/2}) \text{ since } \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \text{ is bounded in}$$

probability (since it converges to the normal distribution)

Convergence Continued

Let sequence X_n and X be r.v.s defined on common probability space. For $r > 0$, we say X_n converges to X in rth mean, $X_n \xrightarrow{rth} X$, if

$$\lim_{n \to \infty} E[|X_n - X|^r] = 0$$

By Jensen's inequality, $X_n \xrightarrow{rth} X \Rightarrow X_n \xrightarrow{sth} X$, $s<r$

(Jensen's: $g(\cdot)$ convex, X & g integrable, $g(E[X]) \leq E[g(X)]$)

The strengths of convergence types are as follows:

$$\left.\begin{array}{l} X_n \xrightarrow{wp1} X \\ X_n \xrightarrow{rth} X, \ r>0 \end{array}\right\} \Rightarrow X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{d} X$$

Also, if, $\forall \varepsilon > 0$, $\sum_{i=1}^{\infty} P(|X_n - X| > \varepsilon) < \infty$, then $X_n \xrightarrow{wp1} X$

Markov Inequality

Let X be a r.v., then for $r>0$ $P(|X|>\varepsilon)\leq \frac{E[|X|^r]}{\varepsilon^r}$

Pf: $E[|X|^r]=\int |X|^r \, dP$

$$=\int_{|X|^r\leq\varepsilon^r}|X|^r\,dP+\int_{|X|^r>\varepsilon^r}|X|\,dP$$

$$\geq \int_{|X|^r\leq\varepsilon^r}|X|^r\,dP+\int_{|X|^r>\varepsilon^r}\varepsilon^r\,dP$$

$$=\int_{|X|^r\leq\varepsilon^r}|X|^r\,dP+\varepsilon^r\int_{|X|^r>\varepsilon^r}1\,dP$$

$$=\int_{|X|^r\leq\varepsilon^r}|X|^r\,dP+\varepsilon^r P(|X^r|>\varepsilon^r)$$

$$\geq \varepsilon^r P(|X|^r>\varepsilon^r)$$

Since in this part $|X|^r>\varepsilon^r$ we are replacing $|X|^r$ with something strictly smaller

$$E[|X|^r]\geq\varepsilon^r P(|X|^r>\varepsilon^r)$$

$$\frac{E[|X^r|]}{\varepsilon^r}\geq P(|X|>\varepsilon)$$

Chebychev's Inequality

Consider Markov's Inequality where $r = 2$ and let $X = Y_n - c$

$$P(|Y_n - c| > \varepsilon) \leq \frac{E[(Y_n - c)^2]}{\varepsilon^2}$$

$$= \frac{Var(Y_n) + (E[Y_n] - c)^2}{\varepsilon^2}$$

As a result, if some X is an integrable r.v. with mean u and variance $\sigma^2 < \infty$, then for any real number $K > 0$,

$$P(|X - u| \geq K\sigma) \leq \frac{1}{K^2}$$

Since the top becomes the variance which cancels with σ^2 on the bottom.
Alternatively

$$P(|X - u| \geq K) \leq \frac{\sigma^2}{K^2}$$

Laws of Large Numbers

The **Weak Law of Large Numbers** says that if $X_1, \dots, X_n$ are i.i.d. with finite mean $\mu = E[X_1]$, then $\bar{X}_n \xrightarrow{P} \mu$

The **Strong Law of Large Numbers** says that if $X_1, \dots, X_n$ are i.i.d with finite mean $\mu = E[X_1]$, then $\bar{X}_n \xrightarrow{a.s.} \mu$

Pf of WLLN: Use chebychev's + def of conv. in prob.

$$\forall \varepsilon > 0,\ P(|\bar{X}_n - \mu| > \varepsilon) \le \frac{E[(\bar{X}_n - \mu)^2]}{\varepsilon^2}$$

$$\lim_{n\to\infty} \frac{\underbrace{E[(\bar{X}_n-\mu)^2]}}{\varepsilon^2} = \lim_{n\to\infty} \frac{Var[\bar{X}_n]}{\varepsilon^2} = \lim_{n\to\infty} \frac{\frac{1}{n^2} Var(\Sigma X_i)}{\varepsilon^2},$$ since

X_i are i.i.d., $Var(\Sigma X_i) = \Sigma Var(X_i) = \Sigma \sigma^2 = n\sigma^2$

$$\lim_{n\to\infty} \frac{\frac{1}{n^2} n\sigma^2}{\varepsilon^2} = \lim_{n\to\infty} \frac{\sigma^2}{n\varepsilon^2} = 0,$$ thus $\forall \varepsilon$

$\lim_{n\to\infty} P(|\bar{X}_n - \mu| > \varepsilon) = 0$, by def this means $\bar{X}_n \xrightarrow{P} \mu$

Central Limit Theorem

Suppose $X_1, \ldots$ is a sequence of i.i.d. random variables with $E[X]=\mu$ and $Var(X)=\sigma^2<\infty$.

Then $\sqrt{n}\left(\frac{\bar{X}_n-\mu}{\sigma}\right) \xrightarrow{d} \text{Normal}(0,1)$

Alternatively, $\sqrt{n}(\bar{X}_n-\mu) \xrightarrow{d} \text{Normal}(0,\sigma^2)$, this implies $\sqrt{n}(\bar{X}_n-\mu)=O_p(1)$ and $(\bar{X}_n-\mu)=o_p(1)$

Prequel for later notes:
Further, if $\hat{\theta}$ is an estimator of some quantity θ, and certain "regularity conditions" are met, we will often be able to use CLT to show

$$\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{d} N(0, \sigma^2_\infty(\hat{\theta}))$$

$\hookrightarrow \lim_{n\to\infty} n\,var(\hat{\theta})$

When $\hat{\theta}$ is unbiased and asymptotically eff., often when $\hat{\theta}$ is based on MLE, we'll have

$$\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{d} N(0, I^{-1}(\theta)), \quad (\hat{\theta}-\theta) = O_p(n^{-1/2})$$

CRLB

"root-n" convergence

Slutsky's Thm & O_p rules

If $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{p} c$, then $X_n + Y_n \xrightarrow{d} X + c$. If $X_n \xrightarrow{d} X$, $Y_n \xrightarrow{p} c$, then $X_n Y_n \xrightarrow{d} cX$. If $X_n \xrightarrow{d} X$, $Y_n \xrightarrow{p} c$ and $c \neq 0$, then $\frac{X_n}{Y_n} \xrightarrow{d} \frac{X}{c}$.

Results of operations on O_p and o_p terms:

$$o_p(1) + o_p(1) = o_p(1)$$

addition rules

$$o_p(1) + O_p(1) = O_p(1)$$

$$O_p(1)\, o_p(1) = o_p(1)$$

multiplication rule

$$(1 + o_p(1))^{-1} = O_p(1)$$

inverse rule

$$o_p(R_n) = R_n\, o_p(1)$$

Factor out rule

$$O_p(R_n) = R_n\, O_p(1)$$

$$o_p(O_p(1)) = o_p(1)$$

o of O = o rule

Delta Method

Let $\hat{\theta}_n$ be a seauence of statistics such that $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \sigma_\infty^2(\hat{\theta}_n))$

Let $g: \mathbb{R} \to \mathbb{R}$ be a once differentiable function at θ with $g'(\theta) \neq 0$. Then

$$\sqrt{n}\left[g(\hat{\theta}_n) - g(\theta)\right] \xrightarrow{d} N\left(0, [g'(\theta)]^2 \sigma_\infty^2(\hat{\theta}_n)\right)$$

In general, if $a_n(\hat{\theta}_n - \theta) \xrightarrow{d} Y$ then

$$a_n[g(T_n) - g(\theta)] \xrightarrow{d} [g'(\theta)]\, Y$$

Pf: Use taylor exp. of $g(\hat{\theta}_n)$ at θ,
$g(\hat{\theta}_n) = g(\theta) + (\hat{\theta}_n - \theta) g'(\theta) + o_p(\hat{\theta}_n - \theta)$,
Since all remaining terms have $(\hat{\theta}_n - \theta)^q$ $q > 2$, and $\hat{\theta}_n - \theta = o_p(1)$, thus the terms after $g'(\theta)$ are $o_p(1)$ after dividing by $\hat{\theta}_n - \theta$. Some algebra gives

$$\sqrt{n}(g(\hat{\theta}_n) - g(\theta)) = \underbrace{\sqrt{n}(\hat{\theta}_n - \theta)}\, g'(\theta) + \sqrt{n}\, o_p(\hat{\theta}_n - \theta)$$

$$\sqrt{n}(g(\hat{\theta}_n) - g(\theta)) \xrightarrow{d} N(0, \sigma_\infty^2(\hat{\theta}_n))\, g'(\theta) + \sqrt{n}(\hat{\theta}_n - \theta)\, o_p(1) \to O_p(1) o_p(1) \to o_p(1)$$

$$\sqrt{n}(g(\hat{\theta}_n) - g(\theta)) \xrightarrow{d} N(0, \sigma_\infty^2(\hat{\theta}_n)(g'(\theta))^2)$$

Extensions of Delta Methods

Again, let $\hat{\theta}_n$ be a sequence of statistics s.t.
$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \text{var}(\hat{\theta}_1))$

Let g be a real valued function differentiable $k (\geq 1)$ at θ w/ $g^{(k)}(\theta) \neq 0$ but $g^{(j)}(\theta) = 0$ for $j < k$. Then

$$(\sqrt{n})^k [g(\hat{\theta}_n) - g(\theta)] \xrightarrow{d} \frac{1}{k!} [g^{(k)}(\theta)] (N(0, \sigma_\infty^2(\hat{\theta})))^k$$

i.e. delta method can use higher orders

Multivariate Delta Method: Let $\vec{\hat{\theta}}_n$ be a sequence of k dimensional random vectors s.t. $\sqrt{n}(\vec{\hat{\theta}}_n - \vec{\theta}) \xrightarrow{d} N(\vec{0}, \Sigma(\vec{\hat{\theta}}_1))$
Let $g: \mathbb{R}^k \to \mathbb{R}^m$ be once differentiable at θ with the gradient matrix $\nabla g(\theta)$, then

$$\sqrt{n}(g(\vec{\hat{\theta}}_n) - g(\vec{\theta})) \xrightarrow{d} N_m(0, (\nabla g(\vec{\theta}))^T \Sigma(\vec{\hat{\theta}}_1) \nabla g(\vec{\theta}))$$

if $(\nabla g(\vec{\theta}))^T \Sigma(\vec{\hat{\theta}}_1) \nabla g(\vec{\theta}))$ is P.S.D. (Positive Semi-definite)