

Agenda

- Convergence in probability & distribution
- $o(p)$ and $O(p)$
- Connection between $o(p)$ and convergence
- Linear model asymptotic distribution
- Past exam problem

Convergence

* Convergence in Probability

* also applies to
a constant c

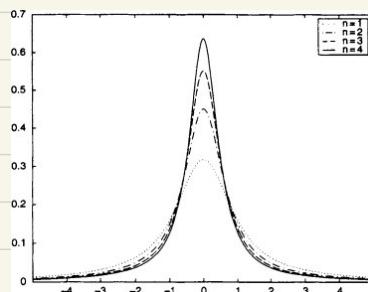
The sequence of RV X_1, \dots, X_n converges in probability to a RV X , denoted $X_n \xrightarrow{P} X$, if for all $\varepsilon > 0$,

$$\lim_{n \rightarrow \infty} P[|X_n - X| < \varepsilon] = 1$$

or equivalently,

$$\lim_{n \rightarrow \infty} P[|X_n - X| \geq \varepsilon] = 0$$

- ie: As $n \rightarrow \infty$, the probability that X_n is far from X gets very small
- Pointwise notion of convergence: individual values of X_n cluster around X as $n \rightarrow \infty$



As $n \rightarrow \infty$, we can expect the distribution to converge to the point 0.

(*) Convergence in Distribution

A sequence of RV X_1, X_2, X_3, \dots converges in distribution to an RV X , denoted by $X_n \xrightarrow{d} X$, if

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

at which cdf $F_X(x)$ is continuous.

- ie: distribution of X_n approaches the shape of distribution for X

(*) Things to Know

- Conv in Prob STRONGER than Conv in Dist

$$X_n \xrightarrow{P} x \Rightarrow X_n \xrightarrow{d} x$$

Stochastic Orders

Let X_n and f_n , $n=1, \dots$, be a sequence of RVs and constants respectively,

① Small oh-p

$$X_n = o_p(f_n) \text{ if } \frac{X_n}{f_n} \xrightarrow{P} 0$$

↳ " X_n is of smaller stochastic order than f_n "

② Big oh-p

$$X_n = O_p(f_n) \text{ if } \forall \varepsilon > 0, \exists M_\varepsilon < \infty \text{ and } N_{\varepsilon, M} \text{ st.}$$

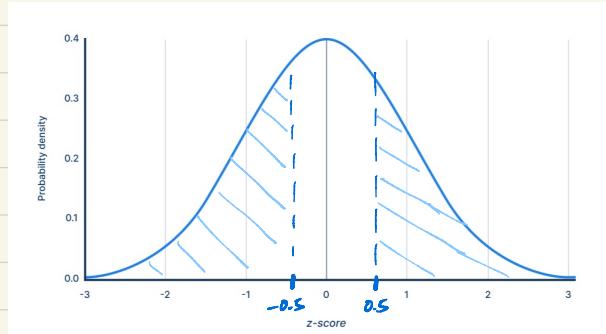
$$P \left\{ \left| \frac{X_n}{f_n} \right| > M_\varepsilon \right\} < \varepsilon, \quad \forall n \geq N_{\varepsilon, M}$$

↳ " X_n is stochastically bounded"

Intuition for Big-oh-p

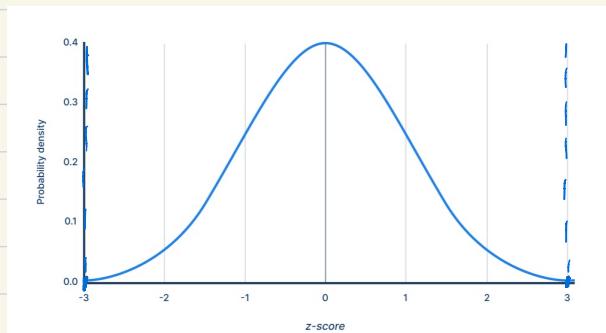
(What does it mean to be bounded in probability?)

Suppose we're working with $N(0,1)$. Let's consider $P(|Z| > 0.5)$



If I set $M = 0.5$, then it's obvious that the distribution is not bounded.

However, if I increase $M = 3$, then only ϵ of the probability mass will be beyond these bounds



So $Z \sim N(0,1)$ is bounded in probability when we set $M=3$
↳ ie: we have a density in which most of its range is within $[-3, 3]$

Source: Bounded in Probability (deeper dive) by StatsDojo (Youtube)

(What does it mean when we talk about rate of convergence?)

We can think of O_p as the "rate at which we zoom in" relative to the axis"

Ex: If we know that $\bar{X}_n \xrightarrow{P} \mu$, it doesn't give us a lot of info
↳ only tells us the point \bar{X}_n will converge to

\Rightarrow We want to know how fast is our sequence of RVs converging?

↳ ie: if we keep zooming in on our axis, at what rate will we be able to capture our density within our bounds?

Turns out, from CLT, we find that \bar{X}_n converges at rate $\frac{1}{\sqrt{n}}$

\Rightarrow if we "zoom in" by multiplying \sqrt{n} , we get:

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$$

Things to Note w/ $O_p(\cdot)$, $o_p(\cdot)$

a) If $X_n = O_p(a_n)$ does NOT necessarily mean that

$$\frac{X_n}{a_n} \xrightarrow{d} X$$

- Big $O_p(\cdot)$ only tells you that $\frac{X_n}{a_n}$ is bounded in probability
- However, $X_n = O_p(a_n)$ where $\frac{X_n}{a_n} \xrightarrow{d} Z$ is a common case

Ex:

$$X_n = \begin{cases} a_n & \text{w prob } \frac{1}{2} \\ -a_n & \text{w prob } \frac{1}{2} \end{cases}$$

$\Rightarrow X_n = O_p(a_n)$ because

$$P\left(\left|\frac{X_n}{a_n}\right| > c\right) < 0 \quad \forall c \geq 1$$

b) $O_p(1)$ is a weaker notion than $o_p(1)$

↳ ie:

$X_n = o_p(1) \Rightarrow X_n = O_p(1)$ but reverse not true

c) \sqrt{n} - consistency

An estimation error $(\hat{\theta} - \theta_0)$ is said to be of (stochastic) order $n^{-1/2}$ if

$$\hat{\theta} - \theta_0 = O_p(n^{-1/2}) \iff \sqrt{n}(\hat{\theta} - \theta_0) = O_p(1)$$

Algebra of Stochastic Orders

① If $X_n = O_p(a_n)$, $Y_n = O_p(b_n)$

- $X_n Y_n = O_p(a_n b_n)$
- $X_n + Y_n = O_p(\max\{a_n, b_n\})$

② Replace 0 with o everywhere in ①, still holds

③ If $X_n = O_p(a_n)$, $Y_n = o_p(b_n)$, then $X_n Y_n = o_p(a_n b_n)$

④ If $X_n = O_p(a_n)$, and $\frac{a_n}{b_n} \rightarrow 0$, then $X_n = o_p(b_n)$

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

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

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

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

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

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

Linear Model : Asymptotic Distribution

* Consistency

Assumptions

- (X, Y) are iid
- $E(Y^2) < \infty, E(X^2) < \infty$
- $E(XX')$ is positive definite

Note: $E(X\epsilon) = 0$
necessary for
causal interpretation

Then

$$\hat{\beta} = E_n(XX')^{-1}E_n(XY) \xrightarrow{b} b^*$$

$$\text{where } b^* = E(XX')^{-1}EXY$$

* Asymptotic Distribution

Assume:

- ① (X, Y) are i.i.d.
- ② $\mathbb{E}Y^4 < \infty, \mathbb{E}\|X\|^4 < \infty$.
- ③ $E(XX')$ is positive definite.

Then

$$\begin{aligned} \sqrt{n}(\hat{\beta} - b^*) &\xrightarrow{d} N(0, \text{avar}(\hat{\beta})), \\ \text{avar}(\hat{\beta}) &= (\mathbb{E}XX')^{-1}\Omega(\mathbb{E}XX')^{-1} \\ &[= Q_{XX}^{-1}\Omega Q_{XX}^{-1} = \Sigma_{XX}^{-1}\Omega\Sigma_{XX}^{-1}] \end{aligned}$$

3 Consider the model

$$Y = \alpha + \beta X + \varepsilon,$$

where:

- the scalars (Y, X) are i.i.d.,
- X is supported on the entire real line, and
- ε is distributed independently from X according to the Poisson distribution with parameter $\lambda = 1$ (thus $\mathbb{E}\varepsilon = \text{var } \varepsilon = \lambda$).

A researcher attempts to estimate (α, β) by computing $(\hat{\alpha}, \hat{\beta})$, the OLS regression of Y on X .

3.1 Argue that $\hat{\beta}$ is unbiased. Is $\hat{\alpha}$ unbiased?

3.2 Is $\hat{\beta}$ the best (=minimum variance) linear unbiased estimator of β ?