ECON 6200: Section 3

2/7/25

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

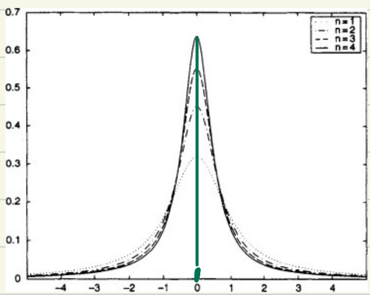
The sequence of RV $X_1, \dots, X_n$ converges in probability to a RV X*, denoted $X_n \xrightarrow{P} X$, if for all $\varepsilon > 0$,

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

or equivalently,

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

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



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

⊛ Convergence in Distribution

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

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

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

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

⊛ Things to Know

- Convg in Prob <u>STRONGER</u> than Convg 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

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

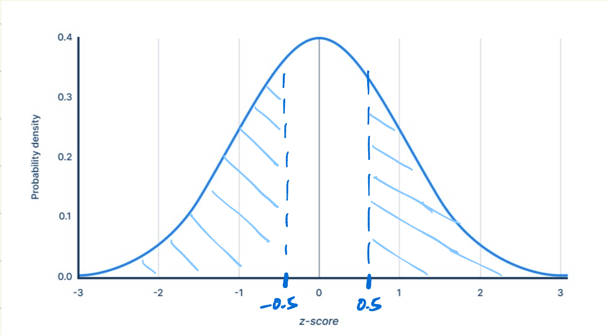
$$P\left\{\left|\frac{X_n}{f_n}\right| > M_\varepsilon\right\} < \varepsilon, \quad \forall n \geq N_{\varepsilon,M}$$

↳ "X_n is stochastically bounded"

Intuition for Big-oh-p

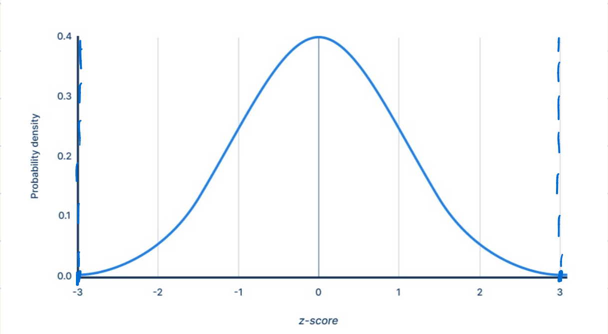
(What does it mean to be bounded in probability?)

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

⇒ 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 $1/\sqrt{n}$

⇒ 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:

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

$\Rightarrow X_n = O_p(a_n)$ 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 O 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 $\frac{a_n}{b_n} \to 0$, then $X_n = o_p(b_n)$

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

Linear Model: Asymptotic Distribution

* Consistency

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

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

Then

$$\hat{\beta} \equiv \mathbb{E}_n(XX')^{-1}\mathbb{E}_n(XY) \xrightarrow{P} b^*$$

where $b^* = E(XX')^{-1}EX'Y$

* Asymptotic Distribution

Assume:

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

Then

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

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 $\mathbb{E}\varepsilon = \operatorname{var}\varepsilon = \lambda$).

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?

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

- (Y, X) iid
- $X_i \in \mathbb{R}$
- ε indep X, $\varepsilon \sim Pois(\lambda = 1)$

$E[\varepsilon] = var(\varepsilon) = \lambda = 1$

Issue: ε are <u>not</u> mean zero

↳ if I can change this $\Rightarrow$ OLS!

$$Y = \alpha + \beta X + \varepsilon + \lambda - \lambda^{=1}$$

$$Y = \underbrace{(\alpha + \lambda)}_{\equiv \tilde{\alpha}} + \beta X + \underbrace{(\varepsilon - \lambda)}_{\equiv \tilde{\varepsilon} \text{ (mean zero)}}$$

$\Rightarrow \hat{\beta}$ is unbiased

$\Rightarrow \hat{a}$ biased

2) $\hat{\beta}$ is BLUE by Gauss - Markov Thm

bc $\hat{\beta} = (X'X)^{-1} X'Y$