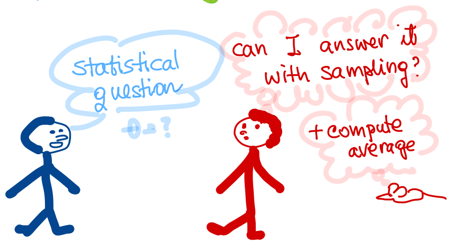
STAT 403 → Resampling Inference = Embracing Randomness

4/4/25



Lecture 3

Expectation, Var

Estimators, their bias and variance

Convergence for R.V.'s → in Probability (like $\bar{X}_n$)

in Distribution (like F_{X_n})

Reminders: OH surveys

HW0

L I posted

L II

Lecture Notes 0 – Probability

Marina Meila
mmp@stat.washington.edu

Department of Statistics
University of Washington

March 2025

Random Variables

Expected Value

Common Distributions

Useful Theorems

Estimators and Estimation Theory

Bias and Var of $\hat{\theta}_n$

Random Variables Reminder from last lecture

$X \sim F$ or $X \sim p$

For two random variables X, Y, their joint CDF is

$$P_{XY}(x,y) = F(x,y) = P(X \leq x, Y \leq y).$$

The corresponding joint PDF is

$$p(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}.$$

The *conditional PDF* of Y given $X = x$ is

$$p(y|x) = \frac{p(x,y)}{p(x)},$$

where $p(x) = \int_{-\infty}^{\infty} p(x,y)dy$

Expected Value

r.v.

R.V. = function on S
X is r.v.

of a fction g(X)

$$\mu_g \equiv \mathbb{E}(g(X)) = \int g(x)dF(x) = \begin{cases} \int_{-\infty}^{\infty} g(x)p(x)dx, & \text{if } X \text{ is continuous} \\ \sum_x g(x)p(x), & \text{if } X \text{ is discrete} \end{cases}.$$

was X

- $\mathbb{E}(\sum_{j=1}^k c_j g_j(X)) = \sum_{j=1}^k c_j \cdot \mathbb{E}(g_j(X_i))$.
- Notation $\mu = \mathbb{E}(X)$
- $\text{Var}(X) = \mathbb{E}((X-\mu)^2)$ is the variance of X. ← $g(X) = (X-\mu)^2$
- If $X_1, \cdots, X_n$ are independent, then

$$\mathbb{E}(X_1 \cdot X_2 \cdots X_n) = \mathbb{E}(X_1) \cdot \mathbb{E}(X_2) \cdots \mathbb{E}(X_n).$$

← Ex.

$E[X_1 + X_2 + \cdots X_n] = \sum_{i=1}^n E[X_i]$ even if $X_{1:n}$ not indep.!

- If $X_1, \cdots, X_n$ are independent, then

$$\text{Var}\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 \cdot \text{Var}(X_i).$$

coefficients

- Covariance

any X, Y (dependent or indep.)

$$\text{Cov}(X,Y) = \mathbb{E}(\underbrace{(X-\mu_x)(Y-\mu_y)}_{g(X,Y)}) = \mathbb{E}(XY) - \mu_x\mu_y$$

Ex

- (Pearson's) correlation (coefficient)

$$\rho(X,Y) = \frac{\text{Cov}(X,Y)}{\sigma_x\sigma_y}.$$

← stdevs

$\in [0,1]$

Conditional Expectation ← VERY IMPORTANT TO UNDERSTAND

The **conditional expectation** of Y given X is the random variable $\mathbb{E}(Y|X) = g(X)$ such that when $X = x$, its value is

obs

$$\mathbb{E}(Y|X = x) = \int y\underbrace{p(y|x)}_{\text{cond. prob of } Y|X} dy,$$

unkn.

where $p(y|x) = p(x, y)/p(x)$.

Theorem ((Weak) Law of Large Numbers)

Let $X_1, \cdots, X_n \sim F$ and $\mu = \mathbb{E}(X_1)$. If $\mathbb{E}|X_1| < \infty$, then the sample average

iid

r.v.

$S_{\bar{X}} = \mathbb{R}$

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

(empirical) mean

sample — n —

(a number)

converges in probability to μ. i.e.,

$$\boxed{\bar{X}_n \xrightarrow{P} \mu.}$$

when $n \to \infty$

random number → fixed num.

Theorem (**Central Limit Theorem**)

Let $X_1, \cdots, X_n \sim F$ and $\mu = \mathbb{E}(X_1)$ and $\sigma^2 = \mathrm{Var}(X_1) < \infty$. Let $\bar{X}_n$ be the sample average. Then

iid

$F_{\bar{X}_n}$ = a distribution

$$\boxed{\sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) \xrightarrow{D} N(0,1).}$$

Note that $N(0,1)$ is also called standard normal random variable.

distribution of $\bar{X}_n$ is ≈ Normal

for $n \to \infty$

random distribution → fixed distrib

Estimators (see next) are r.v.'s

How do they "converge" to the correct answer?

Convergence for random variables

- **random sample** $X_1, \cdots, X_n \sim F$ are IID (independently, identically distributed) from a CDF F.
- For a sequence of random variables $Z_1, \cdots, Z_n, \cdots$, we say Z_n **converges in probability** to a fixed number μ iff for any $\epsilon > 0$,

$$\lim_{n\to\infty} P(|Z_n - \mu| > \epsilon) = 0$$

$Pr[|\bar{X}_n - \mu| \leq \varepsilon] \to 1$ for any ε

Notation $Z_n \xrightarrow{P} \mu$.

- Let $F_1, \cdots, F_n, \cdots$ be the corresponding CDFs of $Z_1, \cdots, Z_n, \cdots$. For a random variable Z with CDF F, we say Z_n **converges in distribution** to Z if for every x,

$$\lim_{n\to\infty} F_n(x) = F(x).$$

$\bar{X}_n$ $N(0,1)$

Notation $Z_n \xrightarrow{D} Z$.

$F, F_n : (-\infty, \infty) \longrightarrow [0,1]$

Estimators and Estimation Theory

(e.g μ)

Let $X_1, \cdots, X_n \sim F$ be a random sample.

- **parameter of interest** $\theta = \theta(F)$ ← statistical question
  - the mean of F, the median of F, standard deviation of F, first quartile of F, ... ← risk
  - μ $P(X \geq t) = 1 - F(t) = S(t)$ survival function ← reliability
  - unknown parameter λ for $\exp(\lambda)$ distribution

think of $\boxed{\bar{X}_n = \text{sample mean}}$

- **statistic** $T_n \equiv T(X_1, \cdots, X_n)$ a function of the random sample

U

- **estimator** $\hat{\theta}_n$ = a statistics we use to estimate $\theta(F)$

$$X_{1:n} \to \boxed{\text{Alg.}} \to \hat{\theta}_n$$ = answer to stat question

$$\Downarrow$$

$$\bar{X}_n$$

Question "*given the parameter of interest, how can I use the random sample to infer it?*"

- How to get $\hat{\theta}_n$? Algorithm
- —"— good is $\hat{\theta}_n$? Theory OR SAMPLING algorithm

How good is an estimator?

$\hat{\theta}_n$ is r.v.

$X_{1:n} \sim F$

$E[\hat{\theta}_n] = \theta$

$\text{Var}\,\hat{\theta}_n \to 0$

wanted

bias: $\text{Bias}(\hat{\theta}_n) = \mathbb{E}(\hat{\theta}_n) - \theta$
variance $\text{Var}(\hat{\theta}_n)$,

$E[\hat{\theta}_n] = \theta \Leftrightarrow \text{Bias} = 0$

Example. 1 $\theta = \mu$

- Let $X_1, \cdots, X_n \sim F$ and $\mu = \mathbb{E}(X_1)$ and $\sigma^2 = \text{Var}(X)$.
- parameter of interest is the population mean μ.
- a natural estimator is the sample average $\hat{\mu}_n = \bar{X}_n$.
- bias$(\hat{\mu}_n) = \mu - \mu = 0$, $\quad \text{Var}(\hat{\mu}_n) = \frac{\sigma^2}{n}$.
- Hence, when $n \to \infty$, both bias and variance converge to 0.
- we say $\hat{\mu}_n$ is a **consistent** estimator of μ.

Biased

$E[\hat{\sigma}^2] < \text{Var}\,X$

$\Uparrow$

$(\hat{\sigma}^2)$

Definition

An estimator $\hat{\theta}_n$ is called a **consistent** *estimator of θ if $\hat{\theta}_n \xrightarrow{P} \theta$.*

$E[\tilde{\sigma}^2] = \text{Var}\,X$ unbiased

corrected variance

$(\hat{\sigma}^2) = \frac{1}{n}\sum_i (X_i - \hat{\mu})^2$ sample var.

$(\tilde{\sigma}^2) = \frac{1}{n-1}\sum_i$ —"—

Lemma

Let $\hat{\theta}_n$ be an estimator of θ. If bias$(\hat{\theta}_n) \to 0$ *and* $\text{Var}(\hat{\theta}_n) \to 0$, *then $\hat{\theta}_n \xrightarrow{P} \theta$. i.e., $\hat{\theta}_n$ is a consistent estimator of θ.*

Example 2 $\theta = \sigma^2 = \text{Var}\,X$