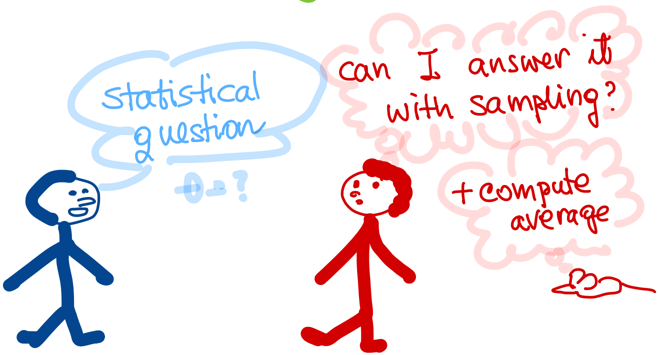
STAT 403 2/4/25

Lecture 2

Resampling Inference = Embracing Randomness

- Web site
- HW ϕ $\geq 50\%$
- Labs



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

R.V. = distribution P

S = sample space $\quad S \subseteq (-\infty, \infty)$

$X \in S$ → discrete
- finite $S = \{0,1\}$ Bernoulli, $S = \{1,2,3,\dots k\}$ Categorical distribution
- countable $S = \{0,1,2,\cdots\}$

→ continuous

Ex: Geometric

$P(x) \propto \lambda^x \qquad \lambda \in (0,1)$

"proportional to" $\qquad$ parameter

$P(x) = \frac{1}{Z_\lambda} \lambda^x$ → normalization constant

Ex: Poisson

$$P(x) = e^{-\lambda} \frac{\lambda^x}{x!}$$

$$= \frac{1}{e^\lambda} = \frac{1}{Z_\lambda}$$

- describe discrete P
  - P(x) for all x = PMF prob. Mass Function
  - CDF Cumulative Distrib. Function

- Binomial (p, n) Family of distributions

$p \in (0,1)$ $\qquad$ $0,1,2,\dots$

p, n fixed: $S = \{0, 1, \dots n\}$ ← finite for any n

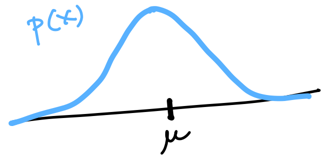
X = # successes in n trials

Continuous S

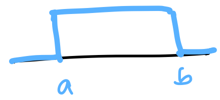
S = interval $\subseteq (-\infty, \infty)$

Ex: Normal (μ, σ^2) : $S = (-\infty, \infty)$

Uniform (a, b) : $S = [a, b]$



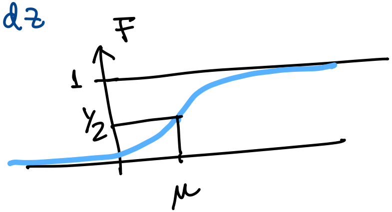
$$p(x) = \begin{cases} 0 & \text{if } x \notin [a,b] \\ \frac{1}{b-a} & \text{if } x \in [a,b] \end{cases}$$

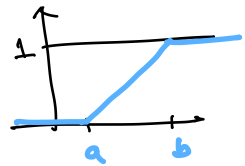


Describing P : PDF = density

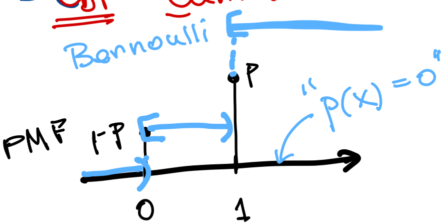
CDF $\quad F(x) = P(-\infty, x] \equiv P[X \leq x]$

$$= \int_{-\infty}^{x} p(z)\,dz$$

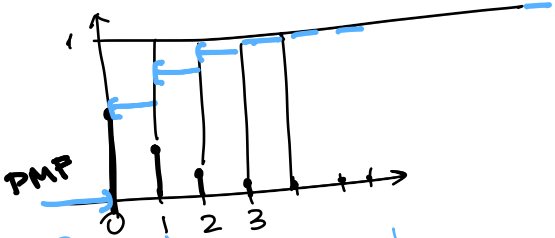




- CDF Cumulative Distrib. Function



CDF : $P[X \leq x]$



Geometric CDF $\quad \lambda = \frac{1}{2}$

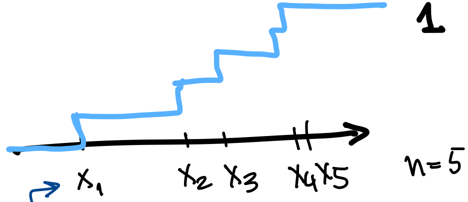
$$F(x) = P[X \leq x]$$

$$\lambda^0 + \lambda^1 + \cdots \lambda^n + \cdots =$$

$$= \frac{1}{1-\lambda}$$

$$P(x) = (1-\lambda)\lambda^x$$

$$\lambda = \frac{1}{2} \longrightarrow = \frac{1}{2} \cdot \frac{1}{2^x}$$



$n=5$ a sample

$\hat{P}$ has $S = \{x_1, \dots x_n\}$

Multiple variables X, Y $\quad P_{XY}$ $\qquad S = S_X \times S_Y$

Data $= \{(x_1, y_1), \dots (x_n, y_n)\}$

PDF (if $\exists$) $\quad p_{XY} = \dfrac{\partial^2 F(x,y)}{\partial x \, \partial y}$

CDF $\quad F(x,y) = P_{XY}[X \le x, Y \le y]$

Conditional density $\quad p_{Y|X}(y|x) = \dfrac{p_{XY}(x,y)}{p_X(x)}$

Marginal density $\quad p_X(x) = \displaystyle\int_{-\infty}^{\infty} p_{XY}(x,y)\, dy$

Random Variables

$X \sim F$ 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

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

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

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

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

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

- Covariance

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

- (Pearson's) correlation

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

P distribution $\quad x \in S \subseteq (-\infty, \infty)$

$$E[X] \equiv \mu_P = \begin{cases} \sum_S x\,P(x) & S \text{ discrete} \\ \int_{-\infty}^{\infty} x\,p(x)\,dx & S \text{ continuous} \end{cases}$$

Properties

- $E[X]$ Linear

$X, X' \in S$

$Z = aX + bX'$ $\quad \Rightarrow \quad \boxed{E[Z] = a\,E[X] + b\,E[X']}$

NO MATTER if X, X' dependent or NOT

$P_{XY}(x,y)$

$$CDF = P[X \le x, Y \le y] = \int_{-\infty}^{x}\int_{-\infty}^{y} P_{XY}(X,Y)\,dX\,dY = F(x,y)$$