Feb-01-16 / Lecture-1 / 1

X is a r.v.

Suppose $(X) \subseteq \mathbb{R}$

$x=1$ (realization) (data)

Flipping coin $H=1$ $T=0$

$|supp(X)| > 1$ (not random)

$|supp(X)| \leq |\mathbb{N}|$

X is a discrete r.v.

$|supp(X)| = |\mathbb{R}|$

X is a continous r.v.

If it is discrete

$p(x)$ is p.m.f

$:= P(X=x)$

$p.\ supp(X) \to (0,1)$ not $\boxed{\cancel{X}}$

If it is continous, X has

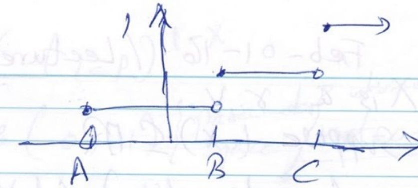
$f(x)$ PDF (Density)

All r.v.'s have $F(x)$ C.D.F

$:= P(X \leq x)$

$P(X \in [a,b]) = F(b) - F(a)$

2



(p.m.f)

C.D.F

$$:= \frac{dF}{dx} = \lim_{x \to \delta} \frac{F(x+\delta) - F(x)}{\delta}$$

$$\approx P(X \in [x, x+\delta])$$

$$P(x) \approx \delta f(x)$$

$$X \sim \text{Bernoulli}(p) := p^{x}(1-p)^{1-x}$$

$supp(X) = \{0,1\}$

$$X \sim \text{Binom}(n,p) := \binom{n}{x} p^{x} (1-p)^{n-x}$$

$supp(X) = \{0,1,\ldots,n\}$

$$X \sim \text{Exp}(\overset{\text{rate}}{\lambda}) := \lambda e^{-\lambda x}$$

$supp(X) = [0,\infty)$

$$X \sim N(\mu, \sigma) := \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

$supp[X] = \mathbb{R}$

3

$p^x \quad (1-p)^{1-x}$

$p \in (0, 1)$ { parameter space } "where parameter lives" not $= 0 = 1$

parameters belong to parametric statistical models.

$\theta^x (1-\theta)^{1-x}$

$\theta \in \Theta = (0,1)$

we rename parameter $(p, \lambda, \mu, \sigma)$ into θ

$\vec{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}$

$\dim[\vec{\theta}] = 2 < \infty$

$X \sim N(\theta_1, \theta_2)$

These are called parametric models.

Generally, you can't write non-parametric functions.

4

$$\mathcal{F} = \{ p(x) = \theta^x (1-\theta)^{1-x} : \theta \in (0,1) = \Theta \}$$

(a set of models)

$$\mathcal{F} = \{ p(x; \theta) : \theta \in \Theta \}$$

$p(x; \theta)$, $f(x; \theta)$

$\rightarrow$; probability of x assuming $\theta = \ldots$

$\rightarrow$; prob of x assuming a parametric value of θ.

$\underbrace{X_1, \ldots, X_n}_{}$ r.v.s i.i.d

$\underbrace{x_1, \ldots, x_n}_{\text{Corresponding data}}$

$p(x_1, \ldots, x_n ; \theta)$

$= p(x_1; \theta) \, p(x_2; \theta) \ldots p(x_n; \theta)$

$= \prod_{i=1}^{n} p(x_i; \theta)$ Likelihood Principle

$\langle \underset{x_1}{0}, \underset{x_2}{0}, \underset{x_3}{1}, \underset{x_4}{0}, \underset{x_5}{1}, \underset{x_6}{0} \rangle$

Pick an $\mathcal{F}$ — $x_1 \ldots x_6$

Statistical Inference:

1. Point Estimation $\hat{\theta}$
2. Confidence set $[\hat{\theta} \pm .]$
3. Testing

$H_0 : \theta \in \Theta_a$ s.t.

$H_a : \theta \in \Theta_b$ $\quad \Theta_a \cap \Theta_b = \phi$

$p(x; \theta = 0.5) = (0.5)^6 = 0.0156$

$$= \begin{cases} p(0; 0.5) = (0.5)^0 (1-0.5)^{1-0} = \underline{0.5} \\ p(0; 0.5) = 0.5^0 (1-0.5)^{1-0} = \underline{0.5} \\ p(1; 0.5) = (0.5)^1 (1-0.5)^{1-1} = \underline{0.5} \\ p(0; 0.5) = 0.5^0 (1-0.5)^{1-0} = \underline{0.5} \\ p(1; 0.5) = \underline{0.5} \quad p(0; 0.5) = \underline{0.5} \end{cases}$$

$p(x; \theta = 0.25) = 0.75^4 \cdot 0.25^2 = 0.0198$

"more likely": $\theta = 0.25$

$\mathcal{L}(\theta; x_1, \dots, x_n)$ likelihood

$\langle 0, 1 \rangle$

$p(0, 1; \theta) = (1-\theta)\,\theta$

$\mathcal{L}(\theta; 0, 1)$ {same}

$$\hat{\theta}_{MLE} = \arg\max_{\theta \in \Theta} \{ \mathcal{L}(\theta; x_1, \dots x_n) \}$$

Maximum Likelihood Estimator

What θ maximizes the likelihood?

$$\{ \ln \mathcal{L}(\theta; x_1, \dots x_n) \}$$ log-likelihood

If we take log everything under '1' becomes negative. p.d.f is >0 no

7

$F = \{ p(x) = \theta^x (1-\theta)^{1-x} ; \theta \in (0,1) = \Theta \}$

$\ln \mathcal{L}(\theta; x_1, \ldots, x_n)$

$= \ln \prod_{i=1}^{n} \theta^{x_i} (1-\theta)^{1-x_i}$

$= \ln \theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{n - \sum_{i=1}^{n} x_i}$

Let $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} x_i$

$(1-\theta)^{1-x_1} (1-\theta)^{1-x_2} = (1-\theta)^{(1-x_1)+(1-x_2)} = 2 -$

$= \ln \theta^{\sum x_i} (1-\theta)^{n - \sum x_i}$

$= \ln \theta^{n\bar{X}} (1-\theta)^{n - n\bar{X}}$

Across multiply

$\sum_{i=1}^{n} x_i = n\bar{X}$

$n\bar{X} \ln(\theta) + (n - n\bar{X}) \ln(1-\theta)$

$\frac{d}{d\theta} (\quad) = \frac{n\bar{X}}{\theta} - \frac{(n - n\bar{X})}{1-\theta} = 0$

$= n\left(\frac{\bar{X}}{\theta} - \frac{1-\bar{X}}{1-\theta}\right) = 0$

$\frac{\bar{X}}{\theta} = \frac{1-\bar{X}}{1-\theta}$

fn: adn: of θ and x is given

8

$\bar{X}(1-\theta) = (1-\bar{X})\theta$

$\bar{X} - \cancel{\bar{X}\theta} = \theta - \cancel{\bar{X}\theta}$

$\hat{\theta}_{MLE} = \bar{X}$

$E[\bar{X}] = p = \theta$

$\bar{X}$ is best for family F (Bern)

In the long run, average is the best.

MLE's properties

① $\hat{\theta}_{MLE} \xrightarrow{P} \theta$ { depends on data set }

② $\frac{\hat{\theta}_{MLE} - \theta}{SE[\hat{\theta}_{MLE}]} \longrightarrow N(0,1)$ Asymptotic Normality

③ $SE(\hat{\theta}) < SE(\hat{\theta}_{others})$

optimal

9

from ② $\hat{\theta}_{MLE} \sim N(\theta, SE(\hat{\theta}_{MLE}$

proposes

$\bar{X} \sim N(\theta, (\sqrt{\theta(1-\theta)})^2$