Lecture 6: Bayes Intro & Parameter Estimation

Thursday, February 3, 2022 9:32 PM

$x_1, \dots, x_n$: obs data

θ: unknown, we're interested

Microwave A — 3 pos. reviews, 0 neg. reviews

Microwave B — 19 pos rev, 1 neg rev

$x_1, \dots, x_n$: reviews, 0 or 1

θ: "goodness" of microwave : prob of pos. review

"is distributed as"

$x_i \mid \theta \sim \text{Bernoulli}(\theta)$

likelihood

$$p(x_i \mid \theta) = \begin{cases} \theta & \text{if } x_i = 1 \\ 1-\theta & \text{if } x_i = 0 \end{cases} = \theta^{x_i}(1-\theta)^{1-x_i}$$

FREQUENTIST

θ: fixed

Goal: to find "best" $\hat{\theta}$ from obs $x_1, \dots, x_n$

MLE: Maximum Likelihood Estimator

↳ find val of θ that maximizes (log) likelihood

"conditionally iid"

$$p(x_1, \dots, x_n \mid \theta) = \prod_{i=1}^{n} p(x_i \mid \theta)$$

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

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

$$\log p(x_1, \dots, x_n \mid \theta) = \underbrace{\sum x_i}_{=k} \cdot \log\theta + \sum(1-x_i)\log(1-\theta)$$

$$= k\log\theta + (n-k)\log(1-\theta)$$

Take derivative

$$k \cdot \frac{1}{\hat{\theta}} + (n-k) \cdot \frac{1}{1-\hat{\theta}} \cdot -1 = 0$$

an estimator for θ

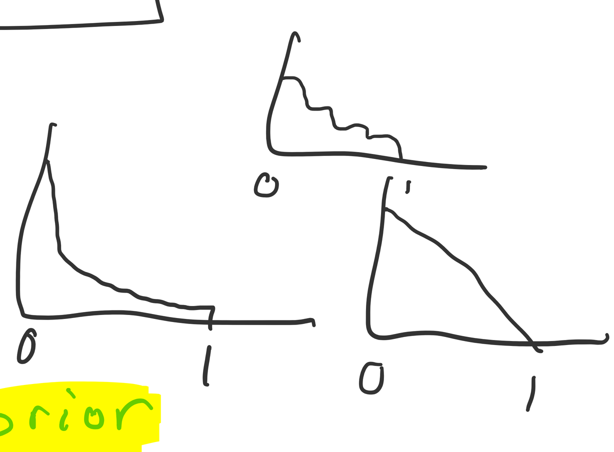
$$k(1-\hat{\theta}) - (n-k)\cdot\hat{\theta} = 0$$

$$k - \cancel{k\theta} = n\theta - \cancel{k\theta}$$

$$\hat{\theta} = \frac{k}{n}$$

$$\boxed{\hat{\theta}_A = \frac{3}{3} = 1} \qquad \boxed{\hat{\theta}_B = \frac{19}{20} = .95}$$

BAYESIAN



θ is random (still unknown)

$x_1, \dots, x_n$

$$\underbrace{p(\theta \mid x)}_{\text{posterior}} = \frac{\overbrace{p(x \mid \theta)}^{\text{likelihood}}\, p(\theta)}{p(x)}$$

prior — belief about θ before any obs.

$p(x)$ ← avoid computing

$\int_\theta p(x \mid \theta) p(\theta) d\theta$ — a "constant"

posterior

"is proportional to"

$$p(\theta \mid x) \propto p(x \mid \theta)\, p(\theta)$$

A convenient choice for $p(\theta)$ is the Beta distribution:

$$p(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

0–1

Beta(2, 1): $p(\theta) \propto \theta$ for $\theta \in [0,1]$

$\alpha = 8$, $\beta = 23$

$$p(\theta) \propto \theta^{7}(1-\theta)^{22} \quad (\text{for } \theta \in [0,1])$$

- Likelihood is Bernoulli
- Prior is Beta
- What is posterior?

$k = \sum x_i$

$x_1, \dots, x_n$

$$p(\theta \mid x) \propto p(k \mid \theta)\, p(\theta)$$

$$\propto \theta^{k}(1-\theta)^{n-k} \cdot \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$$\propto \theta^{k+\alpha-1}(1-\theta)^{n-k+\beta-1}$$

$$p(\theta \mid x) = \text{Beta}(k+\alpha,\ n-k+\beta)$$

posterior is also Beta

"Point estimates": from dist. to a $\hat{\theta}$

Maximum a posteriori (MAP) estimate

↳ val of θ that maximizes posterior

for Beta(α,β) dist, mode is $\dfrac{\alpha-1}{\alpha+\beta-2}$

Suppose we choose $\theta \sim \text{Beta}(1, 5)$ as our prior.

$$\theta_A \mid x \sim \text{Beta}(\cancel{3}\,4,\ 5) \qquad \theta_B \mid x \sim \text{Beta}(20, 6)$$

$$\boxed{\hat{\theta}_A = \frac{\cancel{4}\,3}{6} = \frac{1}{\cancel{3}\,2}} \qquad \boxed{\hat{\theta}_B = \frac{19}{24} = .76}$$