STAT 391

5/9/23

Lecture 13

Part I - Estimating P's

Part II - Model Selection

- Mixture models
- randomness of estimated parameter ✓
- Bayesian estimation ←
- ML estimation ✓

Poll f.b. posted:

- topics for rest of 391

✓ Estimation < discrete, continuous

✓ Model Selection

→ Stat. estimators as R.V

ML ~~Estimation~~

$$
\begin{array}{ccccc}
 & \text{data} & & & \\
x & \{x^{1:n}\} = \mathcal{D} & \xrightarrow{ML} & & \theta^{ML} \\
\cap & \cap & & & \cap \\
S_X & S_X^n & & & S_{\theta^{ML}} \\
P_X & \xrightarrow[\text{iid}]{} P_X^n & \xrightarrow[P_X]{n} & & P_{\theta^{ML}},\ E[\theta^{ML}],\ Var\,\theta^{ML}
\end{array}
$$

assume known

θ^{ML} guess !! $\Longleftrightarrow$ θ^{ML} is a R.V.

Wanted: θ^{true} 0

$\underbrace{\text{for } n \to \infty}_{\text{consistency}}$

• $E_{P_X^n}[\theta^{ML}] = \theta^{true}$ for finite n

Unbiasedness

$E[\theta^{ML}] - \theta^{true}$ = bias

a special form of bias (refers to a parameter)

Normal (μ, σ^2)

- $\boxed{\mu^{ML} = \frac{1}{n} \sum_{i=1}^{n} x^i}$

$$\boxed{E[\mu^{ML}] = E\left[\frac{1}{n} \sum_{i=1}^{n} x^i\right] = \frac{1}{n} \sum_{i=1}^{n} \underbrace{E[\overset{N(\mu,\sigma^2)}{x^i}]}_{\mu} = \mu} \rightarrow \text{unbiased}$$

randomness in sample!!

if σ^2 smaller, n can be smaller, too

- $$\boxed{Var\, \mu^{ML} = Var\left[\frac{1}{n} \sum x^i\right] = \frac{1}{n^2} \sum_{i=1}^{n} \underbrace{Var(x^i)}_{\sigma^2} = \frac{n\sigma^2}{n^2} = \frac{1}{n}\sigma^2}$$

iid (independent!)

Var $\searrow 0$ for $n \to \infty$

In general

$\mu = E[X]$

$P_X \leftarrow$ unknown distribution

$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x^i$, $\quad x^{1:n} \sim P_X$ iid $\quad \Rightarrow E[\hat{\mu}] = \mu$

Estimation of σ^2 for $N(\mu,\sigma^2)$ ← assume $p^{true} = N(\mu,\sigma^2)$

$$(\sigma^2)^{ML} = \frac{1}{n}\sum_{i=1}^{n}(x^i - \mu^{ML})^2 \qquad \frac{\sum x^i}{n}$$

or $\left[Var\, p^{true} = \sigma^2 \quad E_{p^{true}}[x] = \mu \right]$

↑ plug-in estimator

$$E_{(p^{true}_x)^n}\left[(\sigma^2)^{ML}\right] = E\left[\frac{1}{n}\sum_{i=1}^{n}(\underbrace{x^i - \mu}\, + \underbrace{\mu - \mu^{ML}})^2\right]$$

← n samples

$$= \frac{1}{n}\sum_{i=1}^{n} E\left[(x^i-\mu)^2 + (\mu-\mu^{ML})^2 + 2(x^i-\mu)(\mu-\mu^{ML})\right] =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left\{\underbrace{E\left[(x^i-\mu)^2\right]}_{\sigma^2} + \underbrace{E\left[(\mu^{ML}-\mu)^2\right]}_{E\left[(\mu^{ML}-E[\mu^{ML}])^2\right]} + \underbrace{2E\left[(x^i-\mu)(\mu-\mu^{ML})\right]}_{* \ -2\frac{\sigma^2}{n}}\right\}$$

$$E\left[(x-\mu)^2\right] = Var\, x = \sigma^2$$

$$E\left[(\mu^{ML}-E[\mu^{ML}])^2\right] = Var\,\mu^{ML} = \frac{\sigma^2}{n}$$

$$E[\mu^{ML}] = \mu$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left\{\sigma^2 + \frac{\sigma^2}{n} - 2\frac{\sigma^2}{n}\right\}$$

$$= \boxed{\sigma^2\,\frac{n-1}{n}} \neq \sigma^2 \quad \text{BIAS} < 0$$

under estimation

$$* \quad E\left[(x^i-\mu)(\mu^{ML}-\mu)\right]$$

$$\mu^{ML}-\mu = \left(\underbrace{\frac{1}{n}x^i + \frac{1}{n}\sum_{j\neq i} x^j}_{\mu^{ML}} \underbrace{- \frac{1}{n}\mu - \frac{n-1}{n}\mu}_{\mu}\right)$$

$$= E\left[(x^i-\mu)\frac{1}{n}(x^i-\mu)\right] + E\left[(x^i-\mu)\sum_{j\neq i}(x^j-\mu)\cdot\frac{1}{n}\right]$$

x^i, x^j independent

$$\Downarrow$$

$$= \underbrace{E\left[(x^i-\mu)\right]}_{0} \sum_{j\neq i} \underbrace{\overset{E}{\left[(x^j-\mu)\right]}}_{0}\cdot\frac{1}{n}$$

$$= \underbrace{E\left[(x^i-\mu)^2\right]}_{\sigma^2}\cdot\frac{1}{n}$$

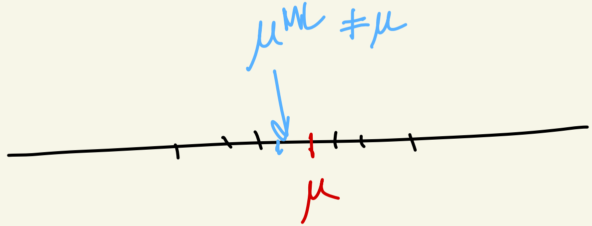
$$E[z^2] = \mu^2 + \sigma^2 \qquad \Rightarrow E\left[(\mu^{ML})^2\right] = \mu^2 + \frac{\sigma^2}{n}$$

z : $E[z] = \mu$

$\mathrm{Var}(z) = \sigma^2$

$$(\sigma^2)^{ML} \neq E\left[\frac{1}{n}\sum_{i=1}^{n}(x^i-\mu)^2\right] = \sigma^2$$

$$E\left[(\sigma^{ML})^2\right] = \sigma^2\left(1 - \underbrace{\frac{1}{n}}_{\mu^{ML} \text{ "causes" bias}}\right)$$



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

$$\boxed{\arg\min_{a} \sum_{i=1}^{n}(x^i - a)^2 = \frac{1}{n}\sum_{i=1}^{n} x^i}$$

Exercise

What's the point

in practice: by CLT

$\mu^{ML} \longrightarrow$ Normal

$(\sigma^2)^{ML} \longrightarrow$ known distribution χ^2

$\underbrace{\text{use } \mu^{ML}, (\sigma^2)^{ML}, n}_{\text{CONFIDENCE INTERVALS}}$

statistical description of uncertainty

α = confidence level $\quad \alpha \lesssim 1$

|| Frequentist statistics ||

Likelihood

$1-\alpha = \Pr[\text{guess is wrong}]$

Statistics

Bayesian Estimation

≠ Phylosophy

Bayes' formula

Data $\mathcal{D} = \{(x^{1:n})\}$

Model family S, $\mathcal{F} = \{ P \text{ on } S \}$

Max likelihood $\Longrightarrow \theta^{ML}$, $P_{\theta^{ML}}$ estimated P

Prior knowledge : $P^0_{(\theta)}$ = distribution over $\mathcal{F}$

"knowledge before $\mathcal{D}$ is observed" ≡ distribution over θ parameters

"prior belief"

Bayesian estimation (≡ A PRINCIPLE)

$\Rightarrow P(\theta)$ = posterior distribution

after observing $\mathcal{D}$ and using $P^0(\theta)$ prior knowledge

"posterior belief"

Example Coin toss

$S = \{0,1\}$

Model Family $\mathcal{F} = \{ \theta_1 \in [0,1],\ \text{Bernoulli}(\theta_1) \}$

Data $x^{1:n} \in S$

Prior p^0 distribution over $\theta_1 \in [0,1]$

1) p^0 uniform $\theta_1 \sim \text{unif}[0,1]$ uninformative prior

wanted $p^{post} = \Pr[\theta_1 \mid \mathcal{D}]$ when $\Pr[\theta_1] = p^0$

$\sim f_{\theta_1}$ $\sim f^0_{\theta_1}$

Bayes' Formula

posterior

$$\boxed{f_{\theta_1}(\theta_1 \mid \mathcal{D})} = \frac{\underbrace{\boxed{f^0_{\theta_1}(\theta_1)}}_{\text{Prior}} \underbrace{\boxed{\prod_{i=1}^{n} \theta_1^{x^i} (1-\theta_1)^{1-x^i}}}_{\text{Likelihood} \equiv \Pr[\mathcal{D} \mid \theta_1]}}{\boxed{\int_0^1 f^0_{\theta_1}(\theta') \prod_{i=1}^{n} (\theta')^{x^i} (1-\theta')^{1-x^i} d\theta'}}$$

normalization $\equiv \int_0^1 \ldots d\theta$

(data) observed

posterior

$$P[A \mid B] = \frac{\boxed{P[A]}\, P[B \mid A]}{\boxed{\sum_a P[a]\, P[B \mid a]}}$$

prior of A

data | θ_1

likelihood what A predicts about B

variable of interest θ_1

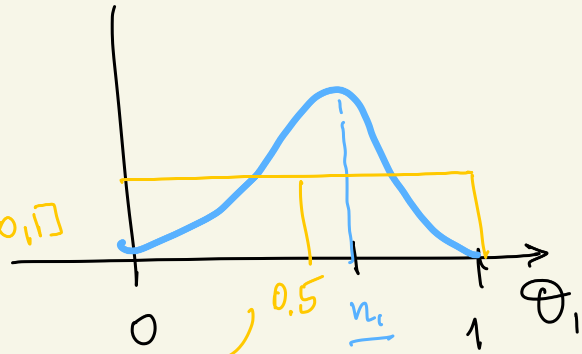
EVIDENCE sum over all possible θ_1's

$\equiv$ total prob

$\equiv$ independent of θ_1

$$f(\theta_1 | \mathcal{D}) \propto \underbrace{1}_{\substack{= \\ \text{unif}[0,1] \\ \text{prior}}} \cdot \underbrace{\theta_1^{n_1}(1-\theta_1)^{n-n_1}}_{\text{likelihood}} = \text{Beta distribution}$$

density of θ



prior expectation

$\frac{n_1}{n} = E[\theta_1 | \mathcal{D}]$

posterior expectation