No.

Date

IV Point estimation

Aim: infer expressions for some parameters θ from a random sample $\underline{X}$ or from a realized measurement $\underline{x}$.

2 parts: 1) finding estimation for θ (several methods)

2) evaluation of this estimations

Framework: $\underline{X} = (X_1, \cdots, X_N)$ a r. s. with $X_j = X$ a pdf or pmf $f(\cdot|\theta)$.

$f_{\underline{X}}(\cdot|\theta)$ is the joint pdf or pmf.

1.1 Method of moments ($k \in \mathbb{N}$)

$$\mu_k := E(X^k) = \int_{\mathbb{R}} x^k f(x|\theta)\,dx$$

Set $m_k := \frac{1}{N}\sum_{j=1}^{N} X_j^k$ and solve the system of equation

$$\begin{cases} m_1 = \mu_1 \\ m_2 = \mu_2 \\ \quad \vdots \\ m_K = \mu_K \leftarrow \text{depending on the size of } \theta \end{cases}$$

Example (used for estimating the crime rate)

$X_j = X \sim \text{binomial}(k, p)$ with $k \in \mathbb{N}$, $p \in (0,1)$

X: number of crimes reported to the police every day; p: reporting rate of crime; k: "real" number of crime every day

$\mu_1 = kp$; $\mu_2 = kp(1-p) + kp^2$ (from table)

$$\begin{cases} m_1 = \overline{X} = kp \\ m_2 = \frac{1}{N}\sum_{j=1}^{N} X_j^2 = kp(1-p) + kp^2 \end{cases}$$

The solution is

$$k = \frac{\overline{X}^2}{\overline{X} - \frac{1}{N}\sum_j (X_j - \overline{X})^2} \; ; \; p = \frac{\overline{X}}{k}$$

1.2 Maximum likelihood estimator (MLE)

Recall $L(\theta|\underline{x}) = f_{\underline{X}}(\underline{x}|\theta)$

Def: For any fixed sample point $\underline{x}$ let $\hat{\theta}(\underline{x})$ be the point where $\theta \longmapsto L(\theta|\underline{x})$ takes its maximal value. There might be more than one. Then the maximum likelihood estimator for θ on a sample $\underline{X}$ is $\hat{\theta}(\underline{X})$.

Then given $\underline{x}$ we estimate θ by $\hat{\theta}(\underline{x})$.

Justification: The maximum likelihood estimator is the parameter point which is observed most likely by the sample.

Drawbacks:

- requires heavy computation (compute derivatives and study the Hessian matrix)
- not always unique
- the maximum can be on the boundary of the parameter space
- problems with the discrete parameters.

Remark: since $s \mapsto \ln(s)$ is strictly increasing, one can also study

$\theta \mapsto \ln L(\theta | \underline{x})$ and get the same maximum.

One speaks about log likelihood functions.

Example

$$X_j \sim n(0,1)$$

$$L(\theta|\underline{x}) = \frac{1}{(2\pi)^{N/2}} e^{-\frac{1}{2}\sum_{j=1}^{N}(x_j-\theta)^2}$$

One has

$$\frac{\partial L}{\partial \theta}(\theta|\underline{x}) = 0 \Leftrightarrow \sum_{j=1}^{N}(x_j\ \theta) = 0 \Leftrightarrow \theta = \frac{1}{N}\sum_{j=1}^{N} x_j$$

and it turns out to be a maximum. Thus

$$\hat{\theta}(\underline{x}) = \frac{1}{N}\sum_{j=1}^{N} x_j$$

1.3 Bayes estimator

Idea: we suppose that θ follows a certain prob. distribution (called prior distribution) which is going to be updated with the sample to a posterior distribution. The update is based on Bayes formula.

Recall that Bayes formula reads:

If $\{C_j\}$ is a partition of the sample space S $(:\Leftrightarrow \bigsqcup_j C_j = S)$ then

$$P(C_j|B) = \frac{P(B|C_j)P(C_j)}{\sum_k P(B|C_k)P(C_k)}$$

Example

Let $\underline{X} = (X_1, \cdots, X_N)$ a r.s. with $X_j \sim$ Bernoulli (p) and suppose $p \sim \mathrm{beta}(\alpha,\beta)$

$p = \theta \in (0,1)$ — prior distribution — α, β fixed

Set $Y = \sum_{j=1}^{N} X_j \sim$ binomial (N, p)

$P(C_j, B) \rightsquigarrow f(\theta | Y = y)$ posterior distribution

$$\frac{P(B|C_j)P(C_j)}{\sum_k P(B|C_k)P(C_k)} \rightsquigarrow \frac{\mathrm{binomial}(y|N,p)\,\mathrm{beta}(p|\alpha,\beta)}{\int \mathrm{binomial}(y|N,s)\,\mathrm{beta}(s|\alpha,\beta)\,ds}$$

This computation will be finished next week.