# Chapter 7 point estimate

Thursday, May 30, 2024 07:40

point estimator: any function of $X_1, X_2 \dots X_n$ (i.e., any stats) is a point estimator
- no mention of correspondence b/w estimator & param)
- no mention of the range of W. usually support of W and support of param coincide but not always.

## Methods of finding estimators

**方法① MoM (method of moment)**

sample moment = population moment

$\frac{1}{n}\sum X_i = EX$

$\frac{1}{n}\sum X_i^2 = EX^2$

example: Satterthwaite approximation.

$Y_1, Y_2 \dots Y_n$ iid $\chi^2_{r}$ $\Rightarrow \sum Y_i \sim \chi^2_{\sum r_i}$, then.

$\sum a_i Y_i \sim \frac{\chi^2(\nu)}{\nu}$ (approximately)

$E(\frac{\chi^2(\nu)}{\nu}) = 1$

$\frac{1}{n}\sum a_i Y_i$

**方法② MLE** $L(\boldsymbol\theta \mid \mathbf{X}) = \prod_{i=1}^{n} f(\mathbf{x} \mid \boldsymbol\theta) = \prod_{i=1}^{n} f(x_i \mid \theta_1, \theta_2 \dots \theta_k)$

$\hat\theta(\mathbf{X})$: param value that maximize $L(\boldsymbol\theta \mid \mathbf{X})$ → a fun of $\boldsymbol\theta$

: mle estimator of the param $\theta$ based on a sample $\mathbf{X}$

How to find the MLE estimator?

↳ 方法①

step 1: score = 0

step 2: second derivative < 0 (concave)

for multivariate:
- at least one second-order partial derivatives is negative.
- Jocabian of the second-order partial derivatives at $\hat\theta$ is positive

$$\begin{vmatrix} \frac{\partial^2 \ell(\theta_1,\theta_2)}{\partial\theta_1^2} & \frac{\partial \ell(\theta_1,\theta_2)}{\partial\theta_1\partial\theta_2} \\ \frac{\partial \ell(\theta_1,\theta_2)}{\partial\theta_1\partial\theta_2} & \frac{\partial^2 \ell(\theta_1,\theta_2)}{\partial\theta_2^2} \end{vmatrix}_{\theta_1=\hat\theta_1,\ \theta_2=\hat\theta_2} > 0$$

step 3) check the values at boundary of the support of $\theta$.

↳ 方法② 找到一个 global upper bound.

example: $X_i \sim N(\theta, 1)$

$L(\theta \mid \mathbf{X}) = \frac{1}{\sqrt{2\pi}} \cdot \exp\{-\frac{1}{2}\sum(x_i-\theta)^2\}$

we know that for any number a

$\sum(x_i - a)^2 \geq \sum(x_i - \bar{x}_n)^2$

$\therefore \exp\{-\frac{1}{2}\sum(x_i-\theta)^2\} \leq \exp\{-\frac{1}{2}\sum(x_i-\bar{x}_n)^2\}$

$\therefore$ global upper bound $\bar{x}_n$. $\therefore \bar{x}_n$ is an MLE.

nice properties of MLE

(i) invariance: if $\hat\theta_{MLE}$ of $\theta$, then for any function $\tau(\theta)$ the MLE of $\tau(\theta)$ is $\tau(\hat\theta)$

$X^2$, $\sqrt{X(1-X)}$, ....

**方法③ Baye's estimator**

$f(\theta \mid \mathbf{X}) = \frac{f(\mathbf{x}\mid\theta)f(\theta)}{f(\mathbf{x})}$ — sampling dist; prior — $\theta$ is a RV.

$f(\mathbf{X})$: marginal dist of the data

Example: $X \mid p \sim Bin(n,p)$ p with prior $p \sim beta(\alpha,\beta)$

(A) numerator: Joint PDF of x, p is $f(x\mid p)\cdot f(p)$

$f(x,p) = \binom{n}{x} p^x (1-p)^{n-x} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}$

$= \binom{n}{x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1+x}(1-p)^{n-x+\beta-1}$

(B) denominator: marginal PDF of x.

$f(x) = \int_p f(x,p)\, \cancel{dx}\, dp$ 错误: 应该对 p 求 integral

$= \int_0^1 \binom{n}{x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot p^{\alpha-1+x}(1-p)^{n-x+\beta-1}\, \cancel{dx}\, dp$

$= \binom{n}{x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 \binom{n+\alpha+\beta-1}{\alpha-1+x} \cdot p^{\alpha+x}(1-p)^{n-x+\beta-1}\, dp \Big/ \binom{n+\alpha+\beta-1}{\alpha-1+x}$ 得到落回的! exercise 4.34 kernel of binomial

$= \binom{n}{x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(\alpha+x)\Gamma(n+\beta-x)}{\Gamma(n+\alpha+\beta)}$ 棒棒! → known as beta-binomial dist

(C) $\therefore$ posterior $f(p\mid x)$ dist. of p given data (of interest is p, now p is a dist instead of a #.

$f(p\mid x) = \frac{f(x\mid p)f(p)}{f(x)} = \binom{n}{x} p^x(1-p)^{n-x} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1} \Big/ \binom{n}{x}\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \frac{\Gamma(\alpha+x)\Gamma(n+\beta-x)}{\Gamma(n+\alpha+\beta)}$

$= \frac{\Gamma(n+\alpha+\beta)}{\Gamma(\alpha+x)\Gamma(n+\beta-x)} \cdot p^{\alpha+x-1}(1-p)^{n-x+\beta-1}$

$\therefore$ posterior is a beta $(\alpha+x, \beta+n-x)$

(D) $\hat p_{Bayes} = \frac{\alpha+x}{n+\alpha+\beta}$ (mean of $f(p\mid x)$)

$= \left(\frac{y}{n}\right)\left(\frac{n}{\alpha+\beta+n}\right) + \left(\frac{\alpha}{\alpha+\beta}\right)\cdot\left(\frac{\alpha+\beta}{\alpha+\beta+n}\right)$ → sample mean; prior mean; weight

Bayes estimator $\hat p_{Bayes}$ is a linear combination of the prior and sample means.

Conjugate prior (family): if the prior and posterior in the same pdf family.

↳ beta family is conjugate for the binomial family $f(x\mid\theta)$ $f(\theta\mid x), f(\theta)$

**方法④ EM algorithm.** skip for now.

## methods of evaluating estimators

Finite-sample measures

**方法① MSE** mean squared error of an estimator

MSE of an estimator W of a param $\theta$ is a function of $\theta$

$E_\theta(W-\theta)^2$

bias: $E_\theta(W-\theta) = E_\theta W - \theta$

↳ unbiased estimator $E_\theta(W) = \theta$

$E_\theta(W-\theta)^2 = E_\theta(W - E_\theta(W) + E_\theta(W) - \theta)^2$ Tricks: identify what's random (the data) what's fixed: E(stuff) / true param value

$= E_\theta\{[W-E_\theta(W)]^2 + [E_\theta(W)-\theta]^2 - 2(W-E_\theta(W))(E_\theta(W)-\theta)\}$

$= E_\theta(W-E_\theta(W))^2 + E_\theta(E_\theta(W)-\theta)^2 - 2E_\theta(W)-\theta\, E_\theta(W - E_\theta(W))$ — E(constant) = constant; $=(E_\theta(W)-\theta)^2$

$= Var_\theta(W) + (Bias_\theta(W))^2$

Example: $X_1, X_2 \dots X_n$ iid $N(\mu, \sigma^2)$

estimator set 1. $\hat\mu = \bar{X}_n$. $\hat\sigma = S_n^2$

(A) bias $E(\bar{X}) = \mu$, $E(S_n^2) = E(\frac{1}{n-1}\sum(X_i-\bar{X}_n)^2)$

$= \frac{1}{n-1}E(\sum(X_i^2 + \bar{X}_n^2 - 2X_i\bar{X}_n))$

$= \frac{1}{n-1}E(\sum X_i^2 - n\bar{X}_n^2)$

$= \frac{1}{n-1}(\sum E(X_i^2) - nE(\bar{X}_n^2))$

$= \frac{1}{n-1}\{n[(EX_i)^2 + Var(X_i)] - n[(E\bar{X}_n)^2 + Var(\bar{X}_n)]\}$

$= \frac{n}{n-1}\{\mu^2 + \sigma^2 - \mu^2 - \frac{\sigma^2}{n}\}$

$= \frac{n}{n-1}\cdot\frac{(n-1)}{n}\sigma^2 = \sigma^2$

$\therefore \bar{X}_n, S_n^2$ unbiased (true without normal assumptions)

(B) variance $E(\bar{X}_n - \mu)^2 = var\bar{X} = \frac{\sigma^2}{n}$

$E(S_n - \sigma^2)^2 = var S_n^2 = \frac{2\sigma^4}{n-1}$

对于服从的 $S_n^2$, we have: $\frac{n-1}{\sigma^2}S_n^2 \sim \chi^2(n-1)$, $Var(\frac{n-1}{\sigma^2}S_n^2) = 2(n-1)$, $Var(S_n^2) = \frac{\sigma^4 2(n-1)}{(n-1)^2}$

estimator set 2: MLE estimates. $\hat\mu = \bar{X}_n$. $\hat\sigma^2 = \frac{1}{n}\sum(X_i-\bar{X}_n)^2 = \frac{1}{n}\cdot\frac{n-1}{n-1}\sum(X_i-\bar{X}_n)^2 = \frac{n-1}{n}S_n^2$

$E(\frac{n-1}{n}S_n^2) = \frac{n-1}{n}\sigma^2 \neq \sigma^2$ biased bias $= \frac{n-1}{n}\sigma^2 - \sigma^2 = -\frac{1}{n}\sigma^2$

$Var(\frac{n-1}{n}S_n^2) = (\frac{n-1}{n})^2 \cdot Var S_n^2 = (\frac{n-1}{n})^2\cdot\frac{2\sigma^4}{n-1} = \frac{n-1}{n^2}\cdot 2\sigma^4$

MSE of $\hat\sigma^2 = (\frac{\sigma^2}{n})^2 + \frac{n-1}{n^2}\cdot 2\sigma^4 = (\frac{2n-1}{n^2})\sigma^4$

comparison of $S^2$ and $\hat\sigma^2_{MLE}$ MSE

MSE $\hat\sigma^2_{MLE} = \frac{2n-1}{n^2}\sigma^4 < (\frac{2}{n-1})\sigma^4$

$\therefore$ MLE gives smaller MSE, though MLE estimator for $\sigma^2$ is biased (underestimate). MLE reasonable for location param, but not for scale params.

Example: MSE of binomial p. $X_1, X_2 \dots X_n$ iid Bin(p)

MLE $\hat p_{MLE} = \frac{x}{n}$

bias$(\hat p) = E(\hat p_{MLE} - p) = E(\frac{x}{n}) - p = \frac{1}{n}E(x) - p = \frac{1}{n}np - p = 0$

$var(\hat p_{MLE}) = E(\hat p_{MLE} - E(\hat p_{MLE}))^2$

$= E\hat p_{MLE}^2 - (E\hat p_{MLE})^2$

$= E((\frac{x}{n})^2) - p^2$

$= \frac{1}{n^2}EX^2 - p^2$

$= \frac{1}{n^2}\{var X + (EX)^2\} - p^2$

$= \frac{1}{n^2}\{np(1-p) + (np)^2\} - p^2$

$= \frac{np}{n^2} - \frac{np^2}{n^2} + \frac{n^2p^2}{n^2} - \frac{n^2p^2}{n^2}$

$= \frac{p(1-p)}{n}$

更简单的解法 $var(\hat p_{MLE}) = var(\frac{x}{n}) = \frac{1}{n^2}var(x) = \frac{1}{n^2}\cdot np(1-p) = \frac{p(1-p)}{n}$

$MSE(\hat p_{MLE}) = Var(\hat p_{MLE}) = \frac{p(1-p)}{n}$ (bias = 0)

$f(x;p) = \binom{n}{x}p^x(1-p)^{n-x}$

$\log\frac{n!}{x!(n-x)!} + x\log p + (n-x)\log(1-p)$

$\frac{x}{p} - \frac{n-x}{1-p} = 0$

$\frac{x(1-p)-(n-x)p}{p(1-p)} = 0$

$x - xp - np + xp = 0$

$x = np \Rightarrow \hat p = \frac{x}{n}$

Bayes estimator $\hat p_B = \frac{x+\alpha}{\alpha+\beta+n}$

[illegible]

## Cramér-Rao Inequality

[illegible]

## Fisher information

[illegible]