9/15/2013, Sunday

- ~~Point estimat~~

MSE & biase - var tradeoff

example

$$MSE = E|\theta - \hat{\theta}|^2$$

(fact: ~~EX~~ $VAR(X) = EX^2 - (EX)^2$

$\Rightarrow EX^2 = VAR(X) + (EX)^2$)

$$= VAR(\theta - \hat{\theta}) + \left[E[\theta - \hat{\theta}]\right]^2$$

$$= VAR(\hat{\theta}) + \left[BIASE(\hat{\theta}, \theta)\right]^2$$

Example biased and variance of normal estimator:

let $x_1, \dots, x_n$ iid $\sim N(\mu, \sigma^2)$ (example 1: digital thermometer)

$\bar{x}$ = estimator for the mean

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

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

now: $E\bar{x} = E\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n}\sum_{i=1}^{n} Ex_i = \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{n\mu}{n} = \mu$

$$ES^2 = E\left[\frac{\sigma^2}{n-1}\frac{(n-1)S^2}{\sigma^2}\right] = \frac{\sigma^2}{n-1}\cdot n-1 = \sigma^2$$

fact: $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ & mean: $(n-1)$ var: $2(n-1)$

~~$MSE(S^2, \sigma^2) =$~~

~~$MSE(S^2) = ES^2 + (VAR(S^2) + [BIASE(S^2, \sigma^2)]^2$~~

=

~~$MSE(\bar{x}) = E\bar{x}^2$~~

$BIASE(\bar{x}) = E\bar{x} - \mu = 0$

$BIASE(S^2) = ES^2 - \sigma^2 = 0$

$$VAR(\bar{x}) = VAR\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)$$
$$= \frac{1}{n^2} VAR\left(\sum_{i=1}^{n} x_i\right)$$
$$= \frac{1}{n^2} \sum_{i=1}^{n} VAR(x_i)$$
$$= \frac{1}{n^2} \cdot \sum_{i=1}^{n} \sigma^2 = \frac{n\sigma^2}{n^2} = \boxed{\frac{\sigma^2}{n}}$$

$$VAR(S^2) = VAR\left(\frac{1}{n-1} \cdot \frac{(n-1)}{\sigma^2} \cdot \frac{\sigma^2}{n-1} \cdot \boxed{\frac{(n-1)S^2}{\sigma^2}}\right)$$
$$= \left[\frac{\sigma^2}{n-1}\right]^2 \cdot 2(n-1) \qquad W \sim \chi^2_{n-1}$$
$$= \frac{2\sigma^4}{n-1}$$

$$MSE(\bar{x}) = BIASE(\bar{x}) + VAR(\bar{x}) = 0 + \frac{\sigma^2}{n} = \frac{\sigma^2}{n}$$
$$MSE(S^2) = BIASE(S^2) + VAR(S^2) = 0 + \frac{2\sigma^2}{n-1} = \frac{2\sigma^2}{n-1}$$

Estimator # 2

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{n-1}{n} S^2$$
$$E\hat{\sigma}^2 = E\left(\frac{n-1}{n}\right)S^2 = E\left[\frac{\sigma^2}{n} \cdot \frac{(n-1)S^2}{\sigma^2}\right]$$
$$= \frac{\sigma^2}{n} E\left[\underbrace{\frac{(n-1)S^2}{\sigma^2}}_{W \sim \chi^2_{n-1}}\right]$$
$$= \frac{\sigma^2}{n} \cdot (n-1)$$
$$= \frac{n-1}{n}\sigma^2.$$
$$VAR(\hat{\sigma}^2) = VAR\left(\frac{n-1}{n} S^2\right) = VAR\left(\frac{\sigma^2}{n} \cdot \frac{(n-1)S^2}{\sigma^2}\right)$$
$$= VAR\left(\frac{n-1}{n}\right)^2 = \frac{\sigma^4}{n^2} \cdot VAR\left(\frac{(n-1)S^2}{\sigma^2}\right)$$
$$= \frac{\sigma^4}{n^2} \cdot 2(n-1) = \frac{2(n-1)\sigma^4}{n^2}$$

smaller variance ✓ $\quad \frac{n-1}{n^2} < \frac{n-1}{(n-1)^2} = \frac{1}{n-1}$

$$MSE(\hat{\sigma}^2) = BIASE(\hat{\sigma}^2) + VAR(\hat{\sigma}^2)$$

$$= \left(E(\hat{\sigma}^2) - \sigma^2\right)^2 + VAR(\hat{\sigma}^2)$$

$$= \left[\frac{n-1}{n}\sigma^2 - \sigma^2\right]^2 + \frac{2(n-1)\sigma^4}{n^2}$$

$$= \left[\frac{-1}{n}\sigma^2\right]^2 + \frac{2(n-1)\sigma^4}{n^2}$$

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

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

$$< \left(\frac{2}{n-1}\right)\sigma^4 = MSE(S^2)$$

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

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

$$= \frac{1-3n}{n^2(n-1)} < 0 \quad (n > 2)$$

$n = 1, 2 \ldots$

$$\Rightarrow \frac{2n-1}{n^2} < \frac{2}{n-1}$$

$\hat{\sigma}^2$ is a biased estimator, but with smaller variance, than S^2 and also smaller MSE.

methods for finding estimators

① method of moments (MOM)

- perhaps the oldest.
- date back to Karl Pearson in late 1800
- simple to use.
- usually not the best.

$X_1 \cdots X_n$ samples from population $\sim f(x|\theta_1 \cdots \theta_k)$

- MOM estimator found by equating first k sample moments to the corresponding k "theoretical" moments calculated from pdf.

sample moments → a number

"theoretical" moments → equations involving k unknowns.

$$m_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = EX$$

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

$$\vdots$$

EX 1 (normal) $X_1, \cdots, X_n$ iid $N(\mu, \sigma^2)$

then to estimate μ and σ^2

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \hat{\mu}$$

$$(\approx EX^2)\quad \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \mathrm{VAR}(X_i) + (EX_i)^2 = \hat{\sigma}^2 + \hat{\mu}^2$$

$$\Rightarrow \boxed{\begin{aligned} \hat{\mu} &= \frac{1}{n}\sum_{i=1}^{n} X_i \\ \hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2 \end{aligned}}$$

Ex 2 Binomial.

$X_1, \cdots X_n$ iid Binomial (k, p)

$P(X_i = x \mid k, p) = \binom{k}{x} p^x (1-p)^{k-1}$, $x = 0, 1, \cdots, k$.

to estimate p

$\frac{1}{n} \sum_{i=1}^{n} X_i = k\hat{p} \Rightarrow \hat{p} = \frac{\frac{1}{n} \sum_{i=1}^{n} X_i}{k}$

(if $n = 1$, only count once.

$X = k\hat{p} \Rightarrow \hat{p} = \frac{X}{k}$)

Ex 3 $X_1 \cdots X_n \sim exp(\lambda)$

$EX = 1/\lambda = \frac{1}{n} \sum_{i=1}^{n} X_i \quad \lambda = n / \left(\sum_{i=1}^{n} X_i \right)$

(2) Max likelihood

- most popular for deriving estimators

*1

- if $X_1 \cdots X_n$ iid $f(x_1 \mid \theta_1 \cdots \theta_K)$

likelihood function

$L(\theta \mid X) = \prod_{i=1}^{n} f(x_i \mid \theta_1 \cdots \theta_K)$

$\hat{\theta} = \underset{\theta \in (H)}{\operatorname{argmax}} L(\theta \mid X)$

↓

m L estimator.

(parameter point where the observed samples is most likely)

eg ~~f(x)~~ $X_1 \dots X_n$ Bernoulli(p)

$$f(x;p) = \begin{cases} p^x (1-p)^{1-x} & , x = 1, 0 \\ = 0 \end{cases}$$

$$L(p) = \prod_{i=1}^{n} f(x_i|p)$$

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

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

~~$\frac{\partial}{\partial p} L(p)$~~ $\log L(p) = \left(\sum_{i=1}^{n} x_i\right) \log p + \left(n - \sum_{i=1}^{n} x_i\right) \log(1-p)$

$$\frac{\partial}{\partial p} \log L(p) = \frac{1}{p}\left(\sum_{i=1}^{n} x_i\right) + - \frac{1}{1-p}\left(n - \sum_{i=1}^{n} x_i\right) = 0$$

$$\frac{1-p}{p} = \frac{n - \sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i} = \frac{1 - \bar{x}}{\bar{x}}$$

$$\Rightarrow \bar{x} - p\bar{x} = p(1 - \bar{x})$$

$$= p - p\bar{x}$$

$$\Rightarrow \bar{x} = \hat{p}$$

$$\boxed{\hat{p} = \frac{1}{n} \sum_{i=1}^{n} x_i}$$

e.g. 2 $X_1, \ldots X_n \sim N(\mu, \sigma^2)$

$L(\mu, \sigma^2) = f_{X_i}(x \mid \theta) \overset{\mu, \sigma^2}{=} \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$

$\log L(\mu, \sigma^2) = \prod_{i=1}^{n} f(x_i \mid \mu, \sigma^2)$

$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$

$\log L(\mu, \sigma^2) = \sum_{i=1}^{n} \left[-\frac{1}{2} \log(2\pi\sigma^2) - \frac{(x_i-\mu)^2}{2\sigma^2} \right]$

$\frac{\partial}{\partial \mu} \log L(\mu, \sigma^2) = \sum_{i=1}^{n} -\frac{2(x_i-\mu)}{2\sigma^2}(-1)$

$= \sum_{i=1}^{n} \frac{(x_i-\mu)}{\sigma^2} = 0 \Rightarrow \boxed{\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i}$

$\frac{\partial}{\partial \sigma^2} \log L(\mu, \sigma^2) = \sum_{i=1}^{n} \left[-\frac{1}{2} \cdot \frac{1}{2\pi\sigma^2} \cdot 2\pi - (-1) \frac{(x_i-\mu)^2}{2\sigma^4} \right] = 0$

$\Rightarrow \sum_{i=1}^{n} \left[-\frac{1}{2} \frac{1}{\sigma^2} + \frac{(x_i-\mu)^2}{2\sigma^4} \right] = 0$

$\sum_{i=1}^{n} \frac{(x_i-\mu)^2}{\sigma^4} = \frac{n}{\sigma^2}$

$\Rightarrow \boxed{\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{\mu})^2}$

- Why Maximum likelihood estimator ?

MLE property :

when n is large , and $\hat{\theta}$ is the ML estimator for θ, then

(i) $\hat{\theta}$ is an approximately unbiased estimator

$$E(\hat{\theta}) \approx \theta$$

(ii) $VAR(\hat{\theta})$ is approx smallest among all estimators

(iii) $\hat{\theta}$ has approximate normal distribution.

e.g. $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{X})^2$

- Drawback :
  - have to find

$$\underset{\theta}{\operatorname{argmax}}\ L(\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$

not always an easy task

- method 1 : set $\frac{\partial L(\theta)}{\partial \theta} = 0$

↓

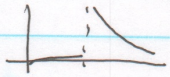
Solve the equation, not always easy

- example : $X \sim$ uniform over 0 to a

$$f(x) = 1/a \text{ , for } 0 \le x \le a$$

$$L(a) = \prod_{i=1}^{n} \frac{1}{a} = \frac{1}{a^n} \quad \text{for } 0 \le x_i \le a$$

$$\frac{dL(a)}{da} = -n a^{-(n+1)}$$



Calculus method doesn't work here because $L(a)$ is maximized at discontinuity