Grundkurs i statistisk teori, del 2

Räkneövning 1 - Egenskaper hos estimatorer, 20.03.2015

1. Låt μ och σ^2 beteckna väntevärdet respektive variansen för tre i.i.d. stokastiska variabler $X = (X_1, X_2, X_3)$. För att skatta μ skapas följande tre estimatorer:

$$\hat{\mu}_1(X) = \frac{1}{3}X_1 + \frac{1}{3}X_2 + \frac{1}{3}X_3$$
$$\hat{\mu}_2(X) = \frac{1}{4}X_1 + \frac{1}{2}X_2 + \frac{1}{4}X_3$$
$$\hat{\mu}_3(X) = \frac{1}{2}X_1 + \frac{1}{4}X_2 + \frac{1}{8}X_3$$

   (a) Vilka av ovanstående estimatorer är väntevärdesriktiga?

   $\mathbb{E}(\hat{\mu}_1(X)) = \mu \Rightarrow$ VVR
   $\mathbb{E}(\hat{\mu}_2(X)) = \mu \Rightarrow$ VVR
   $\mathbb{E}(\hat{\mu}_3(X)) = \frac{7}{8}\mu \Rightarrow$ EJ VVR

   (b) Vilken av de väntevärdesriktiga estimatorerna är effektivast?

   $Var(\hat{\mu}_1(X)) = \frac{\sigma^2}{3}$
   $Var(\hat{\mu}_2(X)) = \frac{3}{8}\sigma^2$
   $Var(\hat{\mu}_1(X)) < Var(\hat{\mu}_2(X)) \Rightarrow \hat{\mu}_1(X)$ är effektivare än $\hat{\mu}_2(X)$

2. Låt $X = (X_1, \ldots, X_n)$ vara n i.i.d. stokastiska variabler med $X_j \sim Bernoulli(\theta)$, dvs

$$f(x_j;\theta) = \theta^{x_j}(1-\theta)^{1-x_j} \qquad \Omega_{X_j} = \{0,1\} \qquad \theta \in (0,1) \qquad \mathbb{E}(X_j) = \theta$$

   För att skatta θ skapas följande två estimatorer:

$$\hat{\theta}_1(X) = X_1 \qquad\qquad \hat{\theta}_2(X) = \frac{1}{n}\sum_{j=1}^{n} X_j$$

   (a) Vilka värden kan respektive estimator anta?

   $\Omega_{\hat{\theta}_1(X)} = \{0,1\}$
   $\Omega_{\hat{\theta}_2(X)} = \{0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1\}$

   (b) Visa att estimatorerna är väntevärdesriktiga.

   Visa att $\mathbb{E}(\hat{\theta}(X)) = \theta$.

   (c) Undersök om estimatorerna är konsistenta.

   (i) $\lim_{n\to\infty} (\mathbb{E}(\hat{\theta}_1(X))) = \theta$

   (ii) $\lim_{n\to\infty} (Var(\hat{\theta}_1(X))) = \theta(1-\theta) \Rightarrow$ Ej konsistent

   (i) $\lim_{n\to\infty} (\mathbb{E}(\hat{\theta}_2(X))) = \theta$

   (ii) $\lim_{n\to\infty} (Var(\hat{\theta}_2(X))) = 0$

   (i) och (ii) ger att $\hat{\theta}_2(X)$ är konsistent.

3. Låt $X = (X_1, \ldots, X_n)$ vara n i.i.d. stokastiska variabler med väntevärdet μ och variansen σ^2. En väntevärdesriktig estimator av σ^2 fås av

$$\hat{\sigma}^2(X) = \frac{1}{n}\sum_{j=1}^{n}(X_j - \mu)^2.$$

För att använda estimatorn ovan krävs det att vi känner till μ. Om vi inte känner till μ kan vi skatta σ^2 med

$$s^2(X) = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \overline{X})^2 \quad \text{där} \quad \overline{X} = \frac{1}{n}\sum_{j=1}^{n} X_j.$$

Visa att $s^2(X)$ är väntevärdesriktig. Tips: $\sum_{j=1}^{n}(X_j - \overline{X})^2 = \sum_{j=1}^{n}[(X_j - \mu) - (\overline{X} - \mu)]^2$

Visa att $\mathbb{E}(s^2(X)) = \sigma^2$.

4. Avståndet mellan två punkter mäts fem gånger och följande resultat erhålls:

101.23	100.73	98.93	99.35	99.59

Mätningarna kan betraktas som ett stickprov från en fördelning där väntevärdet μ motsvarar det korrekta avståndet och standardavvikelsen σ motsvarar precisionen av mätmetoden. Beräkna ett väntevärdesriktigt estimat av variansen σ^2 då

(a) $\mu = 100$.

$\hat{\sigma}^2 = 0.76$ (0.7563)

(b) μ är okänd.

$s^2 = 0.94$ (0.9439)

5. För att skatta en kvadrats yta mäter man dess sida n gånger. De n mätningarna kan betraktas som n i.i.d. stokastiska variabler $X = (X_1, \ldots, X_n)$ för vilka väntevärdet μ motsvarar den korrekta längden på kvadratens sida och standardavvikelsen σ motsvarar precisionen av mätmetoden. För att skatta ytan $A = \mu^2$ skapas följande två estimatorer:

$$\hat{A}_1(X) = \left(\frac{1}{n}\sum_{j=1}^{n} X_j\right)^2 \qquad\qquad \hat{A}_2(X) = \frac{1}{n}\sum_{j=1}^{n} X_j^2$$

Vilken av estimatorerna har lägre bias?

Tips: $Var(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$

$\mathbb{E}(\hat{A}_1(X)) = \sigma^2/n + \mu^2 \;\Rightarrow\; b(\hat{A}_1(X)) = \mathbb{E}(\hat{A}_1(X)) - \mu^2 = \sigma^2/n$

$\mathbb{E}(\hat{A}_2(X)) = \sigma^2 + \mu^2 \;\Rightarrow\; b(\hat{A}_2(X)) = \mathbb{E}(\hat{A}_2(X)) - \mu^2 = \sigma^2$

$b(\hat{A}_1(X)) \leq b(\hat{A}_2(X)) \;\Rightarrow\; \hat{A}_1(X)$ har lägre bias (=systematiskt fel).