AMCS 143

Lect # 11

IV - Parameter Estimation Using the Sample Mean.

Goal: We consider experiments performed to obtain information about a probability model (i.e. estimation of a specific parameter $\hat{\lambda}$

$\Rightarrow$ Two types of estimates:
- point estimate
- Confidence interval estimate $\lambda \in [a, b]$ with a specific probability

1/ Point Estimation

1_a Properties of Estimators

1. Assume that we have n observations =

$$X_1, X_2, \ldots, X_n$$

2. Need to estimate a specific parameter r

3. We use the observations $X_1, X_2, \ldots$ to produce a sequence of estimates of r:

$\hat{R}_1$ function of X_1

$\hat{R}_2$ function of X_1 and X_2

$\hat{R}_3$ " " $X_1, X_2,$ and X_3

$\vdots$

Property 1: **Consistent Estimator**

The sequence of estimates $\hat{R}_1, \hat{R}_2, \ldots$ of the parameter r is consistent if for any $\varepsilon > 0$

$$\lim_{n \to +\infty} \Pr\left[\,|\hat{R}_n - r| \geq \varepsilon\,\right] = 0$$

Property 2 **Unbiased Estimator.**

An estimate $\hat{R}$ of a parameter r is unbiased if $E[\hat{R}]$ is equal to r (otherwise $\hat{R}$ is biased)

Property 3 Asymptotically Unbiased Estimator

The sequence of estimators $\hat{R}_n$ of a parameter r is asymptotically unbiased if

$$\lim_{n \to +\infty} E[\hat{R}_n] = r$$

1.b Estimation of the Expected Value of a RV.

We use the sample mean $M_n(X) = \frac{\sum_{i=1}^{n} X_i}{n}$

Theorem: If X has a finite variance, then the sample mean $M_n(X)$ is an unbiased and consistent estimator of $E(X)$.

Proof:

$$\underline{E[M_n(X)]} = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} \underbrace{E(X_i)}_{=E(X)} = \frac{1}{\cancel{n}}\,\cancel{n}\cdot E(X) = \underline{E(X)}$$

$\Downarrow$

unbiased

1.c Estimation of the variance of a RV

Let $r = Var[X]$ and our goal is to find an estimate of $r = \hat{R}$ using samples $X_1, X_2, \ldots, X_n$.

<u>Case 1</u> $E(X)$ is known

* If $E(X) = 0 \Rightarrow$ $Var(X) = E(X^2) - (\underset{0}{\underset{\shortparallel}{E(X)}})^2$

$$= E(X^2)$$

<u>Thus</u> $\hat{R} = \dfrac{\sum_{i=1}^{n} X_i^2}{n}$

again we have a consistent and unbiased estimator.

* If $E(X) = \mu$

$$\hat{R} = \frac{\sum_{i=1}^{n} (X_i - \mu)^2}{n}$$

<u>Case 2</u> $E(X)$ is unkown

- In this case, a natural approach is to replace the unkown mean $\underset{E(X)}{\underset{\parallel}{\mu}}$ by the corresponding sample mean $M_n(X)$

- In other words we define the estimator:

$$V_n(X) = \frac{1}{n} \sum_{i=1}^{n} (X_i - M_n(X))^2.$$

Theorem:

$$E[V_n(X)] = \frac{n-1}{n} Var[X]$$

Implications

1- $V_n(X)$ is biased since $E[V_n(X)] \neq Var(X)$

2- $\lim_{n \to +\infty} E[V_n(X)] = Var[X]$. Thus $V_n(X)$ is asymptotically unbiased.

3- $$V_n'(X) = \frac{n}{n-1} V_n(X)$$

$$= \frac{\cancel{n}}{n-1} \cdot \frac{1}{\cancel{n}} \sum_{i=1}^{n} (X_i - M_n(X))^2$$

$$= \boxed{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - M_n(X))^2}$$ is an unbiased estimator of $Var(X)$.

Proof of Theorem

$$V_n[X] = \frac{1}{n} \sum_{i=1}^{n} \left(X_i^2 + M_n^2(X) - 2 X_i M_n(X) \right)$$

$$= \frac{1}{n} \sum_{i=1}^{n} X_i^2 + M_n^2(X) - 2 M_n(X) \sum_{i=1}^{n} \frac{X_i}{n}$$

recall $\quad M_n(X) = \frac{1}{n} \left(\sum_{j=i}^{n} X_j \right)$

$$= \frac{1}{n} \sum_{i=1}^{n} X_i^2 + \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} X_i X_j - \frac{2}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} X_i X_j$$

$$= \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} X_i X_j$$

$(\mathrm{Cov}[X_i, X_j] + \mu_X$

$$E[V_n(X)] = \frac{1}{n} \sum_{i=1}^{n} \underbrace{E(X_i^2)}_{E(X^2)} - \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} E[X_i X_j]$$

$$= \frac{1}{n} \cdot n\, E[X^2] - \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\mathrm{Cov}[X_i X_j] + \mu_X^2 \right)$$

$$= \frac{1}{n} \cdot n \, E[X^2] - \frac{1}{n^2}\left(n \, Var[X] + n^2 \underset{E(x)}{\mu_X^2}\right)$$

$$= E[X^2] - \frac{1}{n} Var(X) \quad E[X]^2$$

$$= Var(X) - \frac{1}{n} Var(X) = \frac{n-1}{n} \, Var(X)$$

QED

I-B <u>Confidence Interval Estimation.</u>

In this case, we for look for

$$Pr[A \leq r \leq B] \geq 1 - \alpha$$

* $B-A$ is called the Confidence Interval
* $1-\alpha$ is the confidence coefficient.

→ An accurate estimate of r is reflected in a low value of $B-A$ and a high value of $1-\alpha$ (as close as possible to 1)

Application to Gaussian RV

Let $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$. If the sample mean $M_n(X)$ satisfies

$$\Pr\left[M_n(X) - d \le \mu_X \le M_n(X) + d \right] = 1 - \alpha$$

then $\alpha = 2\,Q\left(\frac{d\sqrt{n}}{\sigma_X}\right)$

Proof:

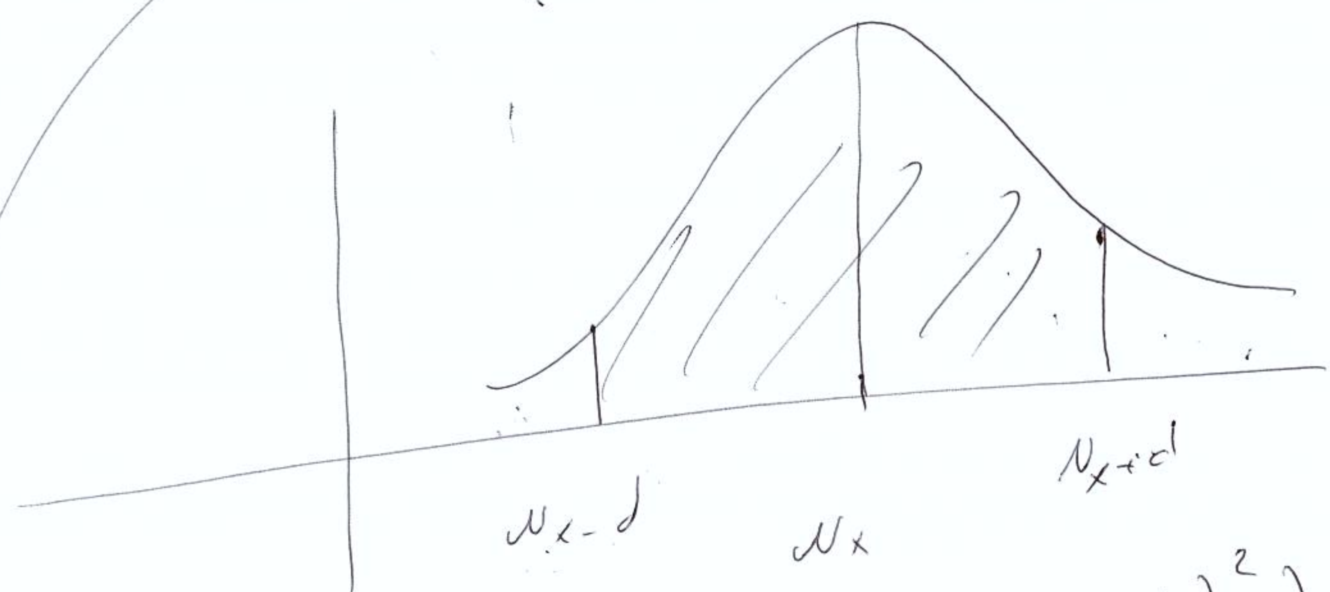
$$\Pr\left[M_n(X) - d \le \mu_X \le M_n(X) + d \right] =$$

$$\Pr\left[\mu_X - d \le M_n(X) \le \mu_X + d \right]$$

$$\Pr\left[\mu_X - d \le \underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i}_{M_n(X)} \le \mu_X + d \right]$$

Given that $X \sim \mathcal{N}(\mu_x, \sigma_x^2)$

$$M_n(X) = \frac{1}{n} \sum_{i=1}^{n} X_i \sim \mathcal{N}\left(\mu_x, \frac{1}{n} \sigma_x^2\right)$$



$$f_{M_n(X)}(m) = \frac{1}{\sqrt{2\pi \frac{\sigma_x^2}{n}}} \exp\left(- \frac{(m - \mu_x)^2}{2 \, \sigma_x^2 / n}\right)$$

$$\Pr\left[\mu_x - d \le M_n(x) \le \mu_x + d\right] = \int_{\mu_x - d}^{\mu_x + d} f_{M_n(x)}(m)\, dm$$

$$= 1 - 2\int_{\mu_x + d}^{+\infty} f_{M_n(x)}(m)\, dm$$

$$= 1 - 2\, Q\left(\frac{d\sqrt{n}}{\sigma_x}\right)$$