Module – 6.1

STATISTICAL ESTIMATION

S. Lakshmivarahan

School of Computer Science

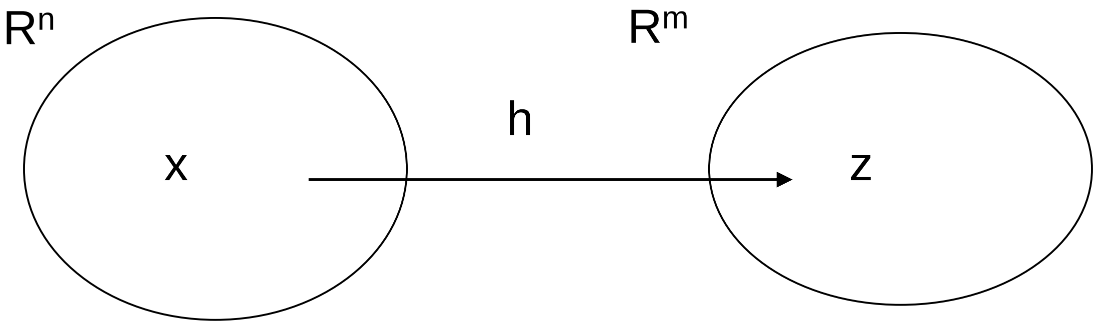
University of Oklahoma

Norman, Ok – 73069, USA

varahan@ou.edu

ESTIMATION PROBLEM

- $x \in R^n$ is the unknown to be estimated – “state” / “true state”
- x is not directly observable but a function of x is.
    - i.e.,
- z is called observation $z = h(x)$



- h – measurement system
- h – linear $z = Hx$
- h – nonlinear

PROBLEM: KNOWING Z, FIND THE BEST ESTIMATE $\hat{\mathbf{x}}$ OF X

- z is modeled by $z = h(x) + v$
  - x and v are not correlated
  - v : noise $v \sim N(0,R)$, $E(v) = 0$, $E(v^T v) = R$

Model for x

- <u>Fisher</u>
- x is a constant, μ
- Max. likelihood
- Point estimation

- <u>Bayesian</u>
- x is random with prior distribution
- Given z, obtain a posterior distribution
- $E(x) = \mu$

- Given h(·), z, assumptions about x and v
- Let $\Phi: R^m \rightarrow R^n$ where $\hat{x} = \Phi(z)$
  - Φ(·) called estimator
- Example: Given the reflectivity, find the rain
- Since z is random, so is $\hat{x}$
- Goal: To obtain the probabilistic characterization of the estimate
- If Φ(·) is linear => $\hat{x}$ is a linear estimate, otherwise, it is nonlinear

TWO APPROACHES

Fisher's

$z = h(x) + v$

$v \sim N(0,R)$

$\therefore p(z|x)$ is known

$z \sim N(h(x), R)$

Max. Likelihood   Least Squares

Bayesian

p(x) – prior given

p(z|x) - conditional distribution known

$p(x,z) = p(z|x)p(x) = p(x|z)p(z)$

$$p(\mathbf{x}|\mathbf{z}) = \frac{p(\mathbf{z}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{z})} = \frac{p(\mathbf{z}|\mathbf{x})p(\mathbf{x})}{\int_{-\infty}^{\infty} p(\mathbf{x},\mathbf{z})d\mathbf{x}}$$

$$= \frac{p(\mathbf{z}|\mathbf{x})p(\mathbf{x})}{\int_{-\infty}^{\infty} p(\mathbf{z}|\mathbf{x})p(\mathbf{x})d\mathbf{x}}$$

Note: When p(x|z) is computed, we could this in a variety of ways

PROPERTIES OF ESTIMATES

- Unbiasedness
- Relative Efficiency
- Efficient Estimate
- Consistency
- Sufficiency

UNBIASEDNESS

- Unbiasedness: Relates to the relative location of the mean of $p(\hat{x}|x)$ – Sampling distribution
- It stands to reason to expect that:
  - $E[\hat{x}|x] = x$ if x is a constant
  - $E_x[E(\hat{x}|x)] = E(\hat{x}) = E(x)$ if x is random

w.r.to prior

- $(E(\hat{x}) - x)$ or $(E(\hat{x}) - E(x)$ is called the bias

EXAMPLE 13.2.1 (LLD (2006))

Coin Toss

Event	H	T
Prob.	p	1-p = q

p –a constant

- Given the results of m (independent) tosses of coin
- E(z) = p, var(z) = pq, z → 1 – H → p; z → 0 – T → q

EXAMPLE 13.2.1 (CONT'D)

- In our notation:
  - $Z_i = p + v_i$
  - $v_i =$ (1-p) with prob. p
    -p with prob. q
  - $E(v_i) = (1-p)p - p(1-p) = 0$
  - $var(v_i) = (1-p)^2 p + p^2(1-p) = pq$
  - $E(z_i) = p$
  - $var(z_i) = pq$

EXAMPLE 13.2.1 (CONT'D)

- An estimate is the sample mean $\hat{p} = \frac{1}{m}\sum_{i=1}^{m} z_i.$

=> $$E(\hat{p}) = \frac{1}{m}\sum_{i=1}^{m} E(z_i) = p$$

$$\begin{aligned} VAR(\hat{p}) &= E[\frac{1}{m}\sum_{i=1}^{m} z_i - p]^2 \\ &= \frac{1}{m^2}\sum_{i=1}^{m} E(z_i - p)^2 \\ &= \frac{pq}{m} \end{aligned}$$

- Distribution of $\widehat{\mathrm{P}}$ has mean p and $\lim_{m\to\infty}$ var = pq/m = 0
- $\widehat{\mathrm{P}}$ is an unbiased estimate of p

EXAMPLE 13.2.1 (CONT'D)

- Why unbiasedness? Consider M.S. error in $\widehat{X}$
- Let x be a constant, then

$$
\begin{aligned}
E(\hat{x}-x)^2 &= E[\hat{x}-E(\hat{x})+E(\hat{x})-x]^2 \\
&= E(\hat{x}-E(\hat{x}))^2+E(E(\hat{x})-x)^2 \\
&\quad +2E[(\hat{x}-E(\hat{x}))(E(\hat{x})-x)]
\end{aligned}
$$

- Since $(E(\hat{x})-x)$ is a constant, $2[E(\hat{x})-x][E(\hat{x})-E(\hat{x})]=0$
- Then $MSE(\hat{x})=E(\hat{x}-x)^2=VAR(\hat{x})+[Bias(\hat{x})]^2$
- M.S.E. = Variance if bias is zero
- Minimizing MSE is equivalent to minimizing Variance

(B) RELATIVE EFFICIENCY

- Let $\hat{X}_a$ and $\hat{X}_b$ be two estimates of the unknown x. We say $\hat{X}_a$ is more efficient than $\hat{X}_b$ if

$$VAR(\hat{\mathbf{x}}_a) \leq VAR(\hat{\mathbf{x}}_b)$$

- The ratio $\frac{VAR(\hat{\mathbf{x}}_b)}{VAR(\hat{\mathbf{x}}_a)}$ the relative efficiency
- Example: Coin tossing $\hat{X}_a = \hat{P}$, $\hat{X}_a = z_i$

$$\mathrm{var}(\hat{P}) = pq/m < \mathrm{var}(z_i) = pq$$

=> Mean is more efficient than a single realization

(B) RELATIVE EFFICIENCY (CONT'D)

- Question 1: Is there a most efficient estimate?
  - YES. This is obtained using maximum likelihood estimate which we will see in a future class
- Question 2: Could it happen that a biased estimate may be more efficient than unbiased estimate? – Yes

(C) CONSISTENCY

- $\hat{X}$ is said to be consistent if $Prob[\ |\hat{\mathbf{x}} - \mathbf{x}| > \epsilon] \longrightarrow 0$ as $m \longrightarrow \infty$

- (ie) $\hat{X}$ converges in probability to x as $m \longrightarrow \infty$

(D) SUFFICIENCY

- Conditions under which the chosen random sample has enough information to obtain the required estimates
- It is more technical – refer to C. R. Rao (1973) Linear Statistical Inference and its Applications, *Wiley*

EXAMPLE 13.2.2 (LLD (2006))

- $z_i = \mu + v_i$ $v_i \sim (0, \sigma^2)$ iid (independent, identically distributed)
- Sample mean: $\bar{z} = \frac{1}{m}\sum_{i=1}^{m} z_i$

$E(\bar{z}) = \mu$ unbiased

$$VAR(\bar{z}) = VAR\left(\frac{1}{m}\sum_{i=1}^{m} z_i\right) = E\left[\frac{1}{m}\sum_{i=1}^{m}(z_i - \mu)\right]^2 = \frac{\sigma^2}{m}$$

- $\bar{Z}$ is consistent since $\frac{\sigma^2}{m}$ → 0 as $m \rightarrow \infty$

EXAMPLE 13.2.2 (CONT'D) ESTIMATE σ^2

- μ is known

$$\hat{\sigma}^2 = \frac{1}{m}\sum_{i=1}^{m}(z_i - \mu)^2$$

$$E(\hat{\sigma}^2) = \frac{1}{m}\sum_{i=1}^{m}E(z_i - \mu)^2 = \sigma^2$$ - unbiased -> (2)

$$VAR(\hat{\sigma}^2) = \frac{2\sigma^4}{m}$$ - consistent -> (3)

EXAMPLE 13.2.2 (CONT'D) ESTIMATE σ^2

- μ is not known - $\overline{Z}$ is used in place of μ

$s^2 = \frac{1}{m}\sum_{m=1}^{m}(z_i - \overline{z})^2$, $E\left(z_i^2\right) = \sigma^2 + \mu^2, E\left(\overline{z}^2\right) = var(\overline{z}) + [E(\overline{z})]^2$ ->(4)

$E(s^2) = \frac{1}{m}[m\sigma^2 + m\mu^2 - \sigma^2 - m\mu^2] = \frac{\sigma^2}{m} + \mu^2$ -> (5)

- => s^2 is biased with bias = $E(s^2) - \sigma^2 = -\frac{\sigma^2}{m}$
- $VAR(s^2) = \frac{2(m-1)\sigma^4}{m^2}$ -> (6)

EXAMPLE 13.2.2 (CONT'D) ESTIMATE σ^2

- Since $$VAR(\sigma^2) = \frac{2\sigma^4}{m} > \frac{2\sigma^4}{m-1}(\frac{m-1}{m})^2 = VAR(s^2)$$ -> (7)

s^2 is more efficient than $\hat{\sigma}^2$

EXERCISES

- Verify the relations (1) through (7)

REFERENCES

- J. L. Melsa and D. L. Cohn (1978) Decision and Estimation Theory, *McGraw Hill*
- A. P. Sage and J. L. Melsa (1971) Estimation Theory and its Application to Communication and Control, *McGraw Hill*
- Also refer to Chapter 13 in LLD (2006)