Mathematical Statistics

Anna Janicka

Lecture V, 22.03.2021

PROPERTIES OF ESTIMATORS, PART I

Plan for today

1. Maximum likelihood estimation examples – cont.
2. Basic estimator properties:
   - estimator bias
   - unbiased estimators
3. Measures of quality: comparing estimators
   - mean square error
   - incomparable estimators
   - minimum-variance unbiased estimator

MLE – Example 1.

- Quality control, cont. We maximize

$$L(\theta) = P_\theta(X = x) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

or equivalently maximize

$$l(\theta) = \ln\binom{n}{x} + \ln(\theta^x) + \ln((1-\theta)^{n-x}) = \ln\binom{n}{x} + x\ln(\theta) + (n-x)\ln(1-\theta)$$

i.e. solve $\quad l'(\theta) = \frac{x}{\theta} - \frac{n-x}{1-\theta} = 0$

solution: $\quad MLE(\theta) = \hat{\theta}_{ML} = \frac{x}{n}$

MLE – Example 3.

- Normal model: $X_1, X_2, \ldots, X_n$ are a sample from N(μ, σ^2). μ, σ unknown.

$$l(\mu,\sigma) = \ln\left(\left(\frac{1}{\sqrt{2\pi\sigma}}\right)^n \exp\left(-\frac{1}{2\sigma^2}\sum(x_i-\mu)^2\right)\right)$$

$$= -\frac{n}{2}\ln(2\pi) - n\ln\sigma - \frac{1}{2\sigma^2}\left(\Sigma x_i^2 - 2\mu\Sigma x_i + n\mu^2\right)$$

we solve

$$\begin{cases} \frac{\partial l}{\partial\sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}(\Sigma x_i^2 - 2\mu\Sigma x_i + n\mu^2) = 0 \\ \frac{\partial l}{\partial\mu} = \frac{1}{\sigma^2}\Sigma x_i - \frac{n\mu}{\sigma^2} = 0 \end{cases}$$

we get:

$$\hat{\mu}_{ML} = \overline{X}, \quad \hat{\sigma}^2_{ML} = \frac{1}{n}\sum(x_i - \overline{X})^2$$

Estimator properties

- Aren't the errors too large? Do we estimate what we want?
- $\hat{\theta}$ is supposed to approximate θ.

  In general: $\hat{g}(X)$ is to approximate $g(\theta)$.
- What do we want? Small error. But:
  - errors are random variables (data are RV)

  $\rightarrow$ we can only control the expected value
  - the error depends on the unknown θ.

  $\rightarrow$ *we can't do anything about it...*

Estimator bias

If $\hat{\theta}(X)$ is an estimator of θ:

bias of the estimator is equal to

$$b(\theta) = E_\theta(\hat{\theta}(X) - \theta) = E_\theta \hat{\theta}(X) - \theta$$

If $\hat{g}(X)$ is an estimator of $g(\theta)$:

bias of the estimator is equal to

$$b(\theta) = E_\theta(\hat{g}(X) - g(\theta)) = E_\theta \hat{g}(X) - g(\theta)$$

$\hat{\theta}(X) / \hat{g}(X)$ is **unbiased**, if

$$b(\theta) = 0 \quad \text{for } \forall \theta \in \Theta$$

other notations, e.g.:

$B_\theta(\hat{g})$

The normal model: reminder

Normal model: $X_1, X_2, \ldots, X_n$ are a sample from distribution $N(\mu, \sigma^2)$. μ, σ unknown.

Theorem. In the normal model, $\bar{X}$ and S^2 are independent random variables, such that

$$\bar{X} \sim N(\mu, \sigma^2/n)$$

$$\frac{n-1}{\sigma^2} S^2 \sim \chi^2(n-1)$$

In particular:

$$E_{\mu,\sigma} \bar{X} = \mu, \text{ and } \mathrm{Var}_{\mu,\sigma} \bar{X} = \sigma^2/n$$

$$E_{\mu,\sigma} S^2 = \sigma^2, \text{ and } \mathrm{Var}_{\mu,\sigma} S^2 = 2\sigma^4/(n-1)$$

Estimator bias – Example 1

In a normal model:

- $\hat{\mu} = \overline{X}$
- $\hat{\mu}_1 = X_1$
- $\hat{\mu}_2 = 5$

Estimator bias – Example 1

In a normal model:

- $\hat{\mu} = \overline{X}$ is an unbiased estimator of μ:

$$E_{\mu,\sigma}\hat{\mu}(X) = E_{\mu,\sigma}\overline{X} = E_{\mu,\sigma}\tfrac{1}{n}\sum_{i=1}^{n} X_i = \tfrac{1}{n} n\mu = \mu$$

- $\hat{\mu}_1 = X_1$ is an unbiased estimator of μ:

$$E_{\mu,\sigma}\hat{\mu}_1(X) = E_{\mu,\sigma}X_1 = \mu$$

- $\hat{\mu}_2 = 5$ is biased:

$$E_{\mu,\sigma}\hat{\mu}_2(X) = E_{\mu,\sigma}5 = 5 \neq \mu \quad \text{eg for } \mu = 2$$

bias:

$$b(\mu) = 5 - \mu$$

Estimator bias – Example 1

In ~~a normal~~ model: any model with unknown mean μ:

- $\hat{\mu} = \overline{X}$ is an unbiased estimator of μ:

$$E_{\mu,\sigma}\hat{\mu}(X) = E_{\mu,\sigma}\overline{X} = E_{\mu,\sigma}\tfrac{1}{n}\sum_{i=1}^{n} X_i = \tfrac{1}{n} n\mu = \mu$$

- $\hat{\mu}_1 = X_1$ is an unbiased estimator of μ:

$$E_{\mu,\sigma}\hat{\mu}_1(X) = E_{\mu,\sigma}X_1 = \mu$$

- $\hat{\mu}_2 = 5$ is biased:

$$E_{\mu,\sigma}\hat{\mu}_2(X) = E_{\mu,\sigma}5 = 5 \neq \mu \quad \text{eg for } \mu = 2$$

bias:

$$b(\mu) = 5 - \mu$$

Estimator bias – Example 1 cont.

- $\hat{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ is a *biased* estimator of σ^2:

$$E_{\mu,\sigma}\hat{S}^2(X) = E_{\mu,\sigma}\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n}E_{\mu,\sigma}\left(\sum X_i^2 - n\overline{X}^2\right)$$
$$= \frac{1}{n}\left(n(\mu^2+\sigma^2) - n(\mu^2 + \sigma^2/n)\right) = \sigma^2 - \sigma^2/n \neq \sigma^2$$

- $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ is an *unbiased* estimator of σ^2 :

$$E_{\mu,\sigma}S^2(X) = E_{\mu,\sigma}\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n-1}E_{\mu,\sigma}\left(\sum X_i^2 - n\overline{X}^2\right)$$
$$= \frac{1}{n-1}\left(n(\mu^2+\sigma^2) - n(\mu^2 + \sigma^2/n)\right) = \frac{1}{n-1}\left(\sigma^2(n-1)\right) = \sigma^2$$

Estimator bias – Example 1 cont.

- $\hat{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ is a *biased* estimator of σ^2:

$$E_{\mu,\sigma}\hat{S}^2(X) = E_{\mu,\sigma}\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n}E_{\mu,\sigma}\left(\sum X_i^2 - n\overline{X}^2\right)$$

$$= \frac{1}{n}\left(n(\mu^2+\sigma^2) - n(\mu^2 + \sigma^2/n)\right) = \sigma^2 - \sigma^2/n \neq \sigma^2$$

- $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ is an *unbiased* estimator of σ^2:

$$E_{\mu,\sigma}S^2(X) = E_{\mu,\sigma}\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n-1}E_{\mu,\sigma}\left(\sum X_i^2 - n\overline{X}^2\right)$$

$$= \frac{1}{n-1}\left(n(\mu^2+\sigma^2) - n(\mu^2 + \sigma^2/n)\right) = \frac{1}{n-1}\left(\sigma^2(n-1)\right) = \sigma^2$$

not necessarily the normal model!

Estimator bias – Example 1 cont. (2)

Bias of estimator $\hat{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$

is equal to

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

for $n \to \infty$, bias tends to 0, so this estimator is also OK for large samples

for any distribution with a variance

Asymptotic unbiased estimator

- An estimator $\hat{g}(X)$ of $g(\theta)$ is **asymptotically unbiased**, if

$$\forall \theta \in \Theta : \quad \lim_{n \to \infty} b(\theta) = 0$$

How to compare estimators?

- We want to minimize the error of the estimator; the estimator which makes smaller mistakes is *better*.
- The error may be either + or -, so usually we look at the square of the error (the mean difference between the estimator and the estimated value)

Mean Square Error

If $\hat{\theta}(X)$ is an estimator of θ:

Mean Square Error of estimator $\hat{\theta}(X)$ is the function

$$MSE(\theta,\hat{\theta}) = E_{\theta}(\hat{\theta}(X) - \theta)^2$$

If $\hat{g}(X)$ is an estimator of $g(\theta)$:

MSE of estimator $\hat{g}(X)$ is the function

$$MSE(\theta,\hat{g}) = E_{\theta}(\hat{g}(X) - g(\theta))^2$$

We will only consider the MSE. Other measures are also possible (eg with absolute value)

Properties of the MSE

We have:

$$MSE(\theta,\hat{g}) = b^2(\theta) + \mathrm{Var}(\hat{g})$$

For unbiased estimators, the MSE is equal to the variance of the estimator

MSE – Example 1

$X_1, X_2, ..., X_n$ are a sample from a distribution with mean μ, and variance σ^2. μ, σ unknown.

- MSE of $\hat{\mu} = \overline{X}$ (unbiased):

$$MSE(\mu,\sigma,\overline{X}) = E_{\mu,\sigma}(\overline{X}-\mu)^2 = Var_{\mu,\sigma}\overline{X} = \frac{\sigma^2}{n}$$

- MSE of $\hat{\mu}_1 = X_1$ (unbiased):

$$MSE(\mu,\sigma,X_1) = E_{\mu,\sigma}(X_1-\mu)^2 = Var_{\mu,\sigma}X_1 = \sigma^2$$

- MSE of $\hat{\mu}_2 = 5$ (biased):

$$MSE(\mu,\sigma,5) = E_{\mu,\sigma}(5-\mu)^2 = (5-\mu)^2$$

MSE – Example 2
Normal model

- MSE of $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$

$$MSE(\mu,\sigma,S^2) = E_{\mu,\sigma}(S^2 - \sigma^2)^2 = Var_{\mu,\sigma}S^2 = \frac{2\sigma^4}{n-1}$$

- MSE of $\hat{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$

$$MSE(\mu,\sigma,\hat{S}^2) = E_{\mu,\sigma}(\hat{S}^2 - \sigma^2)^2 = b^2(\sigma) + Var_{\mu,\sigma}\hat{S}^2$$

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

$$MSE(\mu,\sigma,S^2) > MSE(\mu,\sigma,\hat{S}^2)$$

MSE – Example 2
Normal model

- MSE of $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$

$$MSE(\mu,\sigma,S^2) = E_{\mu,\sigma}(S^2-\sigma^2)^2 = Var_{\mu,\sigma}S^2 = \frac{2\sigma^4}{n-1}$$

- MSE of $\hat{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$

$$MSE(\mu,\sigma,\hat{S}^2) = E_{\mu,\sigma}(\hat{S}^2-\sigma^2)^2 = b^2(\sigma) + Var_{\mu,\sigma}\hat{S}^2$$

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

in any model: similarly, just with different expressions

$$MSE(\mu,\sigma,S^2) > MSE(\mu,\sigma,\hat{S}^2)$$

MSE and bias – Example 2.

Poisson Model: $X_1, X_2, \ldots, X_n$ are a sample from a Poisson distribution with unknown parameter θ.

$$\hat{\theta}_{ML} = \ldots = \overline{X}$$

$$b(\theta) = 0$$

$$MSE(\theta, \overline{X}) = \mathrm{Var}_\theta \overline{X} = \mathrm{Var}_\theta \tfrac{1}{n} \sum_{i=1}^{n} X_i = \frac{\theta}{n}$$

Comparing estimators

$\hat{g}_1(X)$ is **better** than (dominates) $\hat{g}_2(X)$, if

$$\forall \theta \in \Theta \qquad MSE(\theta, \hat{g}_1) \leq MSE(\theta, \hat{g}_2)$$

and

$$\exists \theta \in \Theta \qquad MSE(\theta, \hat{g}_1) < MSE(\theta, \hat{g}_2)$$

an estimator will be better than a different estimator only if its plot of the MSE never lies above the MSE plot of the other estimator; if the plots intersect, estimators are **incomparable**

MSE – Example 1 again

$X_1, X_2, \ldots, X_n$ are a sample from a distribution with mean μ, and variance σ^2. μ, σ unknown.

- $\hat{\mu} = \overline{X}$ (unbiased)
- $\hat{\mu}_1 = X_1$ (unbiased)
- $\hat{\mu}_2 = 5$ (biased)

- S^2 (biased)
- $\hat{S}^2$ (unbiased)

Comparing estimators – Example 1 cont.

We have

- From among $\hat{\mu} = \overline{X}$ and $\hat{\mu}_1 = X_1$

  $\hat{\mu}$ is better (for *n*>1)
- $\hat{\mu} = \overline{X}$ and $\hat{\mu}_2 = 5$ are incomparable, just like $\hat{\mu}_1 = X_1$ and $\hat{\mu}_2 = 5$
- From among S^2 and $\hat{S}^2$

  $\hat{S}^2$ is better

Comparing estimators – cont.

A lot of estimators are incomparable $\rightarrow$ comparing any old thing is pointless; we need to constrain the class of estimators

If we compare two unbiased estimators, the one with the smaller variance will be better

Minimum-variance unbiased estimator

We constrain comparisons to the class of unbiased estimators. In this class, one can usually find the best estimator:

$g^*(X)$ is a **minimum-variance unbiased estimator (MVUE)** for $g(\theta)$, if

- $g^*(X)$ is an unbiased estimator of $g(\theta)$,
- for any unbiased estimator $\hat{g}(X)$ we have

$$Var_{\theta} g^*(X) \leq Var_{\theta} \hat{g}(X) \quad \text{for } \theta \in \Theta$$

How can we check if the estimator has a minimum variance?

- In general, it is not possible to freely minimize the variance of unbiased estimators – for many statistical models there exists a limit of variance minimization. It depends on the distribution and on the sample size.