Math 40 Probability and Statistical Inference
Winter 2021
Lecture 17 Statistical Properties of An Estimation (6.4)

Yoonsang Lee (yoonsang.lee@dartmouth.edu)

6.4 Statistical properties of an estimator

From data $\{X_i\}_i^n$, we estimate an unknown variable θ

$$\theta \approx \hat{\theta}(X_1, X_2, ..., X_n).$$

Example. Estimate the mean $\theta = \mu$ using the sample mean

$$\mu \approx \hat{\mu}(X_1, X_2, ..., X_n) = \frac{1}{n}\sum_i^n X_i.$$

What can we say about the estimator $\hat{\theta}$ when

1. we have other data sets.
2. the data size $n \to \infty$.

We are going to answer 1 in 6.4.1 while answer 2 in 6.4.4. 6.4.2 and 6.4.3 consider how we define accuracy of the estimator.

6.4.1 Unbiasedness

Every time we have a new data set (of the same size), the estimator $\hat{\theta}$ will be different. Thus, we treat $\hat{\theta}$ as a random variable, and ask its mean, $E(\hat{\theta})$. If

$$E(\hat{\theta}) = \theta,$$

that is, if the mean of your estimator is equal to the true variable, then the estimator is called **unbiased**.

Example 6.12 The sample mean

$$\overline{X} = \frac{1}{n}\sum_{i}^{n} X_i$$

is unbiased. That is, $E(\overline{X}) = E(X)$.

The sample variance

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

is unbiased. That is, $E(\frac{1}{n-1}\sum_i^n (X_i - \overline{X})^2) = Var(X)$.

cf. $E(\frac{1}{n}\sum_i^n (X_i - \mu_X)^2) = Var(X)$ where $\mu_X = E(X)$.

Example 6.13 Let $\{X_i\}_i^n$ be IID from $\mathcal{E}(\lambda)$. The MM estimator of λ, $\hat{\lambda}_{MM} = \frac{1}{\overline{X}}$ is unbiased.

- $n\overline{X}$ follows a Gamma distribution with $\alpha = n$ and λ.
- $E(\frac{1}{\overline{X}}) = n\int_0^\infty \frac{1}{x} f_{gamma}(x;\alpha,\lambda)dx = \frac{n\lambda}{n-1} \neq \lambda$ (check the textbook for calculations).
- Although it is biased, $E(\hat{\lambda}_{MM}) \to \lambda$ as $n \to \infty$. In this case, the estimator is called **asymptotically unbiased** (definition 6.15).

Example 6.14 Let $\{X_i\}_i^n$ be IID from $Uniform(0, \theta)$.

- Let $X_{max} = max(X_1, ..., X_n)$.
- The cdf of X_{max} is given by

$$Pr(X_{max} \leq x) = \prod_i^n Pr(X_i \leq x) = \left(\frac{x}{\theta}\right)^n$$

- By taking the derivative of the cdf,

$$f(x) = \frac{n}{\theta}\left(\frac{x}{\theta}\right)^{n-1}$$

- $E(X_{max}) = \int_0^\theta xf(x) = \frac{n}{n+1}\theta$. Thus, X_{max} is biased but asymptotically unbiased.
- From this calculation, we know that $\hat{\theta} = \frac{n+1}{n}X_{max}$ is an unbiased estimator.

Check R code `6_4_Ex6.14.R` on Canvas.

Example 6.15 A solder saw enemy tank numbers 15, 45 and 38. Assuming that tanks are numbered sequentially 1,2,..., what is an unbiased estimate o the number of tanks in the enemy army?

Solution $45 \times \frac{4}{3} = 60$.

Are you sure this answer is correct?

6.4.2 Mean Square Error (MSE)

For a data set $\{X_i\}$, we can define the error of the estimator $\hat{\theta}$

$$\text{error} = \hat{\theta} - \theta.$$

- As $\hat{\theta}$ changes for a different data set, the estimator $\hat{\theta}$ is a random variable, and thus the error is also a random variable.

- Note that we defined the estimator as unbiased if the mean of the error is zero.

- **Mean square error (MSE)** is the expected squared error

$$MSE = E(error^2) = E\left((\hat{\theta} - \theta)^2\right).$$

- If the estimator is unbiased, MSE is the variance of the error.

- More generally, the decomposition of MSE is

$$\boxed{MSE = Var + Bias^2}$$

  Here

$$Var = E\left((\hat{\theta} - E(\hat{\theta}))^2\right),$$

  and

$$Bias = E(\hat{\theta}) - \theta.$$

- Root MSE (RMSE) is the square root of MSE.

Do you remember the optimal portfolio design (section 3.8)?

There are two stocks with the same mean but with different variances. Then, we were able to find a new portfolio (through a linear combination of the stocks) with a variance smaller than any of the variances of the stocks.

We are going to use a similar idea in example 6.23.

Example 6.23 We have two independent data sets from the same distribution with mean μ and variance σ^2, say $\{X_i\}_i^m$ and $\{Y_j\}_j^n$. Instead of the two estimators $\overline{X}$ and $\overline{Y}$, can you come up with a new estimator with a smaller MSE?

Solution

- First of all, $\overline{X}$ and $\overline{Y}$ are unbiased estimators.
- $Var(\overline{X}) = \frac{\sigma^2}{m}$ and $Var(\overline{Y}) = \frac{\sigma^2}{n}$ (from the independence).
- Thus, MSE of $\overline{X}$ is $Var(\overline{X}) = \frac{\sigma^2}{m}$ while MSE of $\overline{Y}$ is $Var(\overline{Y}) = \frac{\sigma^2}{n}$.
- Let $Z = r\overline{X} + (1-r)\overline{Y}$ for $0 < r < 1$.
- $E(Z) = rE(\overline{X}) + (1-r)E(\overline{Y}) = \mu$, the true mean. That is, Z is also an unbiased estimator.
- MSE of Z$= Var(Z) = \frac{r^2\sigma^2}{m} + \frac{(1-r)^2\sigma^2}{n}$.
- By taking the derivative with respect to r and set it equal to 0, we have $r = \frac{m}{n+m}$ as an optimal value.
- When $r = \frac{m}{n+m}$, that is,

$$Z = \frac{1}{n+m}\left(\sum_i^m X_i + \sum_j^n Y_j\right)$$

  is the unbiased estimator with the smallest MSE.

6.4.3 Multidimensional MSE

We have considered the case when θ is a scalar. What if we are interested in the estimation of more than one variable, say a vector?

- For example, $\boldsymbol{\theta} = (\mu, \sigma^2)$ is a two-dimensional vector.
- In general, for $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_n)$, and its estimator $\hat{\boldsymbol{\theta}}$,
- MSE of $\hat{\boldsymbol{\theta}}$ is defined as

$$E\left((\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})\right),$$

  which is a $n \times n$ matrix.
- The total MSE is the trace of the MSE matrix. That is, it is the sum of the MSE of each θ_i

$$\text{Total MSE } = \sum_i^n \text{MSE of } \theta_i$$

6.4.4 Consistency of Estimators

When we have a data set, which of the following scenarios is more doable?

A: We have other data sets of the same size.

B: We add more data values to the current set.

- The expectation with respect to scenario A is related to bias.
- What can we say about an estimator in scenario B?
- To specify the size of your data, let $\hat{\theta}_n$ be an estimator of your choice using a data set of size n.
- We call the estimator $\hat{\theta}_n$ **consistent** if

$$\hat{\theta}_n \to \theta \text{ as } n \to \infty.$$