Mathematical statistics
Week 2/b: Finite sample properties of estimators

Péter Elek and Ádám Reiff

26th September, 2013

1. Unbiasedness, relative efficiency and MSE-criterion

Unbiasedness.

- Consider an estimator $\widehat{\theta}$ that is used to estimate a population parameter θ.
  - $\widehat{\theta}$ is a random variable and its distribution depends on the true value of θ.
- $\widehat{\theta}$ is said to be an *unbiased estimator* of θ if $E(\widehat{\theta}) = \theta$ for all possible values of θ.
- If $E(\widehat{\theta}) \neq \theta$ for *at least one* value of θ, then $\widehat{\theta}$ is a biased estimator of θ.
- Bias of $\widehat{\theta}$ is $Bias_{\widehat{\theta}}(\theta) = E\left(\widehat{\theta}\right) - \theta$.
  - The bias is also a function of the true parameter value θ.

Example.

- Example (from previous lecture): the (true) sampling distribution of the sample mean and sample median are:

	0	1	2	3	4	5	6	8	9	12
$\overline{x}$	$\frac{1}{27}$	$\frac{3}{27}$	$\frac{3}{27}$	$\frac{1}{27}$	$\frac{3}{27}$	$\frac{6}{27}$	$\frac{3}{27}$	$\frac{3}{27}$	$\frac{3}{27}$	$\frac{1}{27}$
m	$\frac{7}{27}$	0	0	$\frac{13}{27}$	0	0	0	0	0	$\frac{7}{27}$

- Which is an unbiased estimator of the population mean?

Relative efficiency.

- Consider $\widehat{\theta}_1$ and $\widehat{\theta}_2$, which are two unbiased estimators of a population parameter θ.
- Estimator $\widehat{\theta}_1$ is said to be *relatively more efficient* than $\widehat{\theta}_2$ (to estimate θ) if $Var(\widehat{\theta}_1) \leq Var(\widehat{\theta}_2)$ for all θ, with strict inequality for at least one value of θ.
  - i.e. the variance of its sampling distribution is smaller

Efficiency (best unbiased estimator).

- The unbiased estimator $\widehat{\theta}$ is said to be *efficient* (or *best unbiased estimator*) if it has the smallest sampling variance among all unbiased estimators.
  - That is, for any other $\widehat{\theta}_2$ unbiased estimator, $Var(\widehat{\theta}) \leq Var(\widehat{\theta}_2)$ for all possible values of θ.
  - Note: efficiency is sometimes defined in a different way.
- Since the sampling variances may depend on the true value of θ (and hence functions of θ are compared in the definition), a best unbiased estimator does not always exist.

Mean squared error (MSE) criterion.

- Comparing variances is useful only for unbiased estimators.
- To compare more general estimators, we can use the mean squared error.
- $MSE\left(\widehat{\theta}\right) = E\left[\left(\widehat{\theta} - \theta\right)^2\right]$.
- MSE criterion: we choose the estimator that has a smaller MSE.
- Proposition: $MSE\left(\widehat{\theta}\right) = Var\left(\widehat{\theta}\right) + \left[Bias\left(\widehat{\theta}\right)^2\right]$

Proof of $MSE\left(\widehat{\theta}\right) = Var\left(\widehat{\theta}\right) + \left[Bias\left(\widehat{\theta}\right)^2\right]$.

- We have:

$$\begin{aligned} MSE\left(\widehat{\theta}\right) &= E\left[\left(\widehat{\theta} - \theta\right)^2\right] = E\left[\left((\widehat{\theta} - E(\widehat{\theta})) - (\theta - E(\widehat{\theta}))\right)^2\right] \\ &= E\left[\left(\widehat{\theta} - E(\widehat{\theta})\right)^2\right] - 2E\left[\left(\widehat{\theta} - E(\widehat{\theta})\right)\left(\theta - E(\widehat{\theta})\right)\right] \\ &\quad + E\left[\left(\theta - E(\widehat{\theta})\right)^2\right]. \end{aligned}$$

- Here the first term is $Var(\widehat{\theta})$, the second term is zero, and the third term is not a random variable, so it is $\left(\theta - E(\widehat{\theta})\right)^2 = Bias^2(\widehat{\theta})$.
- Therefore $MSE\left(\widehat{\theta}\right) = Var\left(\widehat{\theta}\right) + Bias^2\left(\widehat{\theta}\right)$.

Example (cont.).

- Which of the previous two sample statistics (sample mean and sample median) has smaller MSE at the particular true parameter value?
  - "True parameter value" now means: the three possible outcomes occur with equal probability.
- $Var(\overline{x}) = \frac{1}{27}(0-5)^2 + \frac{3}{27}(1-5)^2 + \ldots + \frac{1}{27}(12-5)^2 = 8.6667$ and $MSE(\overline{x}) = Var(\overline{x})$ because of unbiasedness.
- $Var(m) = \frac{7}{27}(0-4.5556)^2 + \frac{13}{27}(3-4.5556)^2 + \frac{7}{27}(12-4.5556)^2 = 20.9136$ and $MSE(m) > Var(m)$ because of the bias.
- Hence $\overline{x}$ is better than m in terms of the MSE-criterion. (And also the former is unbiased, while the latter is not.)

2. Finite sample properties of sample mean and sample variance

2.1. Properties of the sample mean

Properties of the sample mean of an i.i.d. sample.

- Suppose $X_1, X_2, \ldots, X_n$ is an i.i.d. sample from a distribution with unknown population mean (expected value) μ and variance σ^2.
- Then the sample mean is unbiased for μ.
  - $E(\overline{X}) = E\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{1}{n}E\left(\sum_{i=1}^n X_i\right) = \frac{1}{n}\sum_{i=1}^n E(X_i) = \frac{1}{n}\sum_{i=1}^n \mu = \frac{1}{n}(n\mu) = \mu$.
- Its standard deviation is proportional to $1/\sqrt{n}$.
  - $Var(\overline{X}) = Var\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{1}{n^2}Var\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2}\sum_{i=1}^n Var(X_i) = \frac{1}{n^2}\sum_{i=1}^n \sigma^2 = \frac{1}{n^2}\left(n\sigma^2\right) = \frac{\sigma^2}{n}$
  - $sd(\overline{X}) = \sigma/\sqrt{n}$

Best linear unbiased estimator (BLUE).

- *Best linear unbiased estimator (BLUE)*: an unbiased estimator is BLUE if it has the smallest variance among all unbiased estimators, which are linear combinations of the sample elements.
- The sample mean is not always the best unbiased estimator for μ. (There are "weird" counter-examples.)
- But it is the best *linear* unbiased estimator for μ in the case of an i.i.d. sample.

Proof of BLUE-property of the sample mean in an i.i.d. sample.

- Let $\widehat{\theta}$ be an arbitrary unbiased linear estimator: $\widehat{\theta} = \sum_{i=1}^{n} a_i X_i$, with $E\left(\widehat{\theta}\right) = \mu$.
- Then $E\left(\widehat{\theta}\right) = E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E\left(X_i\right) = \mu$, so by unbiasedness we have $\sum_{i=1}^{n} a_i = 1$.
- Also, $Var\left(\widehat{\theta}\right) = Var\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 Var\left(X_i\right) = \sigma^2 \sum_{i=1}^{n} a_i^2$.
- Hence, since $Var\left(\overline{x}\right) = \frac{\sigma^2}{n}$, we only have to prove that $\sum_{i=1}^{n} a_i^2 \geq \frac{1}{n}$ if $\sum_{i=1}^{n} a_i = 1$.
- But this is true because of the inequality between the quadratic and arithmetic mean. So $\overline{X}$ is indeed the Best Linear Unbiased Estimator of μ.

Properties of the sample mean in a normal sample.

- Suppose $X_1, X_2, \ldots, X_n$ is an i.i.d. sample from a $N(\mu, \sigma^2)$ distribution with unknown parameters.
- Then the sample mean is not only BLUE but it is the best unbiased estimator for μ.
  - i.e. it is the best among all (even nonlinear) estimators. We do not prove this.
- Moreover, as the linear combination of normal random variables is also normally distributed, $\overline{X}$ is also normally distributed: $\overline{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$.

Exercise 1.

- A filling machine is set to pour 500 g-s of cereal into a box container. Denote the actual weight of cereal filled into the container by X, and assume that $X \sim N\left(500, 20^2\right)$.
- A random sample of $n = 25$ boxes is selected (i.e. $x_1, \ldots, x_{25}$ is drawn), and the plant manager stops the process if $\overline{x} > 510$ or $\overline{x} < 490$.
- What is the probability of stopping?

Solution.

- Since $X \sim N\left(500, 20^2\right)$ and $n = 25$, $\overline{X} \sim N(500, \frac{20^2}{25} = 16)$, and $\frac{\overline{X}-500}{20/\sqrt{25}} \sim N(0,1)$.
- Therefore $\Pr(stop) = 1 - \Pr(notstop) = 1 - \Pr(490 < \overline{X} < 510) = 1 - \Pr\left(\frac{490-500}{20/\sqrt{25}} < \frac{\overline{X}-500}{20/\sqrt{25}} < \frac{510-500}{20/\sqrt{25}}\right) = 1 - \Pr\left(-2.5 < \frac{\overline{X}-500}{20/\sqrt{25}} < 2.5\right) = 1 - [\Phi(2.5) - \Phi(-2.5)] = 2 - 2\Phi(2.5) = 0.0124$.

2.2. Properties of the sample variance

"Ideal" sampling variance.

- The "ideal" sample variance is obtained when we treat the expected value μ as known in the variance formula: $s^2_{ideal} = \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{n}$.
  - "Ideal" because we have the generally unknown μ instead of $\overline{X}$ in the expression.
- In an i.i.d. sample, s^2_{ideal} is an unbiased estimator of σ^2.
  - Proof: $E\left(s^2_{ideal}\right) = \frac{\sum_{i=1}^{n} E\left[(X_i-\mu)^2\right]}{n} = \frac{n\sigma^2}{n} = \sigma^2$
- Moreover, in a normally distributed sample, $\frac{ns^2_{ideal}}{\sigma^2}$ follows a chi-squared distribution with n degrees of freedom.
  - Proof: $s^2_{ideal} = \frac{\sigma^2}{n}\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2$. Hence $\frac{ns^2_{ideal}}{\sigma^2}$ is the sum of squares of n independent standard normal variables.

Exercise 2.

- Assume that the size of output of some production process is distributed normally: $X \sim N(10, 0.1^2)$ (i.e. the true variance of the size is known).
- We draw a sample of $n = 25$ observations from a large number of outputs.
- What is the probability that the ideal sample variance will exceed 0.014?

Solution.

- $\Pr(s^2_{ideal} > 0.014) = \Pr\left(\frac{ns^2_{ideal}}{\sigma^2} > \frac{25 \cdot 0.014}{0.01}\right) = \Pr\left(\frac{ns^2_{ideal}}{\sigma^2} > 35\right) = 1 - \Pr\left(\frac{ns^2_{ideal}}{\sigma^2} < 35\right) \approx 0.1$
- from the table of chi-squared distribution with 25 degrees of freedom.

"Actual" uncorrected sample variance.

- "Actual" uncorrected sample variance is: $s^2 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n}$
  - Actual because μ is replaced by $\overline{x}$
- $E(s^2) = \frac{n-1}{n}\sigma^2$, hence it is a biased estimator of σ^2.
- Proof:
  - $s^2 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n} = \frac{\sum_{i=1}^{n}\left[(x_i-\mu)-\left(\overline{X}-\mu\right)\right]^2}{n} = \ldots = \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{n} - \left(\overline{X} - \mu\right)^2$.
  - In the expected value we have $E\left[(X_i - \mu)^2\right] = Var(X_i) = \sigma^2$, and also we have $E\left[\left(\overline{X} - \mu\right)^2\right] = E\left[\left(\overline{X} - E(\overline{X})\right)^2\right] = Var(\overline{X}) = \frac{\sigma^2}{n}$.
  - Therefore $E(s^2) = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2$, so s^2 is a biased estimator of σ^2.

Corrected sample variance.

- $s^{*2} = \frac{n}{n-1}\frac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}{n} = \frac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}{n-1}$ is unbiased for σ^2, and it is called the *corrected sample variance*.
- Moreover, in a normally distributed sample, $s^{*2} = \frac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}{n-1} = \ldots = \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{n-1} - \frac{n}{n-1}\left(\overline{X} - \mu\right)^2$, so $\frac{(n-1)s^{*2}}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 - \left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right)^2$ is distributed as χ^2_{n-1}.
  - We "lose" one degree of freedom because μ was replaced by $\overline{X}$.
  - Correct proof is based on induction (not covered in class).

Further properties in case of a normal sample.

- Since $(n-1)s^{*2} = ns^2$, therefore $\frac{ns^2}{\sigma^2} = \frac{(n-1)s^{*2}}{\sigma^2}$ also follows a χ^2_{n-1} distribution.
- Further, $Var\left(\frac{(n-1)s^{*2}}{\sigma^2}\right) = 2(n-1)$ (why?).
- Therefore $Var\left(s^{*2}\right) = \frac{\sigma^4}{(n-1)^2} Var\left(\frac{(n-1)s^{*2}}{\sigma^2}\right) = \frac{\sigma^4}{(n-1)^2} 2(n-1) = \frac{2\sigma^4}{n-1}$.
- By similar arguments it is easy to see that $Var(s^2) = \frac{2(n-1)\sigma^4}{n^2}$.

"Normalized" sample mean in a normal sample.

- "Normalized" sample mean: $\frac{\overline{X}-\mu}{s^*/\sqrt{n}}$
- Suppose $X_1, X_2, \ldots, X_n$ is an i.i.d. sample from a $N(\mu, \sigma^2)$ distribution with unknown parameters.
- We know that
  - $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ and
  - $\sqrt{\frac{(n-1)s^{*2}/\sigma^2}{n-1}} = \frac{s^*}{\sigma} \sim \sqrt{\frac{\chi^2_{n-1}}{n-1}}$.
- Further, these two random variables are independent.
  - We do not prove this.
- So $\frac{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}}{s^*/\sigma} = \frac{\overline{X}-\mu}{s^*/\sqrt{n}}$ follows a t_{n-1}-distribution.

Exercise 3.

- Assume that the test-scores at the CEU math entry test are distributed normally.
- From a random sample of 25 students we find that $\overline{x} = 70$ and $s^{*2} = 400$.
- Find the "probability" that the expected value of the results lies in the $[60, 80]$ interval!

Solution.

- We know that $\frac{\overline{X}-\mu}{s^*/\sqrt{n}} \sim t_{24}$,
- so $\Pr(60 < \mu < 80) = \Pr(-80 < -\mu < -60) = \Pr\left(\frac{\overline{X}-80}{20/\sqrt{25}} < \frac{\overline{X}-\mu}{s^*/\sqrt{n}} < \frac{\overline{X}-60}{20/\sqrt{25}}\right) = \Pr\left(-2.5 < \frac{\overline{X}-\mu}{s^*/\sqrt{n}} < 2.5\right) = 0.98$
- where the last result is taken from the table of the t-distribution with 24 degrees of freedom.

Material.

- Further exercises are on a separate sheet (with solutions).
- Wooldridge Appendix C.1-C.2
- Casella-Berger 5.1-5.3 (except for Theorem 5.2.11), 7.1, 7.3.1-7.3.2.
  - only to the extent covered in the course