Economics Beta

Prove the sample variance is an unbiased estimator

Asked 4 years, 6 months ago Active 4 years, 4 months ago Viewed 10k times

8

I have to prove that the sample variance is an unbiased estimator. What is is asked exactly is to show that following estimator of the sample variance is unbiased:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

I already tried to find the answer myself, however I did not manage to find a complete proof.

2

econometrics statistics self-study

edited Mar 16 '15 at 11:55

asked Mar 15 '15 at 12:11

Andreas Dibiasi
232 3 7

3 Please post what you have accomplished so far -and add the self-study /homework tag. – Alecos Papadopoulos Mar 15 '15 at 15:16

1 @AlecosPapadopoulos Is the homework tag really a thing? I've been removing those where I found them, as I didn't see a value in it. – FooBar Mar 15 '15 at 18:22

@FooBar I am not sure this is a good idea. Our meta-threads indicate a rather strong opinion in favor of explicitly acknowledging homework questions as such, in the tags. – Alecos Papadopoulos Mar 15 '15 at 19:35

@AlecosPapadopoulos could you link to me that discussion? I only found a question without answers: meta.economics.stackexchange.com/questions/1252/… – FooBar Mar

3 Answers

8

I know that during my university time I had similar problems to find a complete proof, which shows exactly step by step why the estimator of the sample variance is unbiased.

The proof I used can be found under http://economictheoryblog.wordpress.com/2012/06/28/latexlatexs2/

The proof itself is not very complicated but rather long. That also the reason why I am not writing it down here and probably it is not fair towards the the person who actually provided it in the first place.

edited Mar 15 '15 at 12:49

answered Mar 15 '15 at 12:20

Thomas Drew

96 4

2 The proof is four-to-five lines *maximum*. I am aware of the link you pointed to, I was always amazed by the unnecessary length of it. – Alecos Papadopoulos Mar 15 '15 at 15:24

11

For a shorter proof, here are a few things we need to know before we start:

$X_1, X_2, \ldots, X_n$ are independent observations from a population with mean μ and variance σ^2

$\mathbb{E}(X_i) = \mu$, $\mathbb{V}\text{ar}(X_i) = \sigma^2$

$\mathbb{E}(X^2) = \sigma^2 + \mu^2$

$\mathbb{V}\text{ar}(X) = \mathbb{E}(X^2) - [E(X)]^2$

$\mathbb{E}(\bar{X}^2) = \frac{\sigma^2}{n} + \mu^2$

Let's try to show that $\mathbb{E}(s^2) = \mathbb{E}\left(\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}\right) = \sigma^2$

To make my life easier, I will omit the limits of summation from now onwards, but let it be known that we are always summing from 1 to n.

$\mathbb{E}\left(\sum (X_i - \bar{X})^2\right) = \mathbb{E}\left(\sum X_i^2 - 2\bar{X}\sum X_i + n\bar{X}^2\right) = \sum \mathbb{E}(X_i^2) - \mathbb{E}\left(n\bar{X}^2\right)$

$\sum \mathbb{E}(X_i^2) - \mathbb{E}\left(n\bar{X}^2\right) = \sum \mathbb{E}(X_i^2) - n\mathbb{E}\left(\bar{X}^2\right) = n\sigma^2 + n\mu^2 - \sigma^2 - n\mu^2$

So far, we have shown that $\mathbb{E}\left(\sum(X_i-\bar{X})^2\right)=(n-1)\sigma^2$

$\mathbb{E}(s^2)=\mathbb{E}\left(\frac{\sum(X_i-\bar{X})^2}{n-1}\right)=\frac{1}{n-1}\mathbb{E}\left(\sum(X_i-\bar{X})^2\right)$

$\mathbb{E}(s^2)=\frac{(n-1)\sigma^2}{n-1}=\sigma^2$

We have now shown that the sample variance is an unbiased estimator of the population variance.

edited Apr 26 '15 at 12:27

answered Mar 16 '15 at 21:58

Five σ

386 1 11

4

Let's improve the "answers per question" metric of the site, by providing a variant of @FiveSigma 's answer that uses visibly the i.i.d. assumption (showing also its necessity).

We want to prove the unbiasedness of the sample-variance estimator,

$$s^2 \equiv \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

using an i.i.d. sample of size n, from a distribution having variance σ^2,

$$E(s^2) =?\ \sigma^2$$

First, write

$$s^2 \equiv \frac{n}{n-1}\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

Then

$$\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2=\frac{1}{n}\left(\sum_{n=1}^{n}(x_i^2-2\bar{x}x_i+\bar{x}^2)\right)=\frac{1}{n}\sum_{n=1}^{n}x_i^2-2\bar{x}\frac{1}{n}\sum_{n=1}^{n}x_i+\bar{x}^2$$

Since $\bar{x}=\frac{1}{n}\sum_{n=1}^{n}x_i$ we get

$$\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2=\frac{1}{n}\sum_{n=1}^{n}x_i^2-\bar{x}^2$$

$$E\left(\frac{1}{n}\sum_{n=1} x_i^2\right) = \frac{1}{n}\sum_{n=1} E(x_i^2) = E(X^2)$$

since the variables are identically distributed.

Also

$$\bar{x}^2 = \left(\frac{1}{n}\sum_{n=1}^{n} x_i\right)^2 = \frac{1}{n^2}\left(\sum_{n=1}^{n} x_i^2 + \sum_{i\neq j} x_i x_j\right)$$

the second sum having $n^2 - n$ elements. So

$$E(\bar{x}^2) = \frac{1}{n^2}(nE(X^2)) + \frac{1}{n^2}\left[(n^2-n)E(x_i)E(x_j)\right]$$

We were able to write $E(x_i x_j) = E(x_i)E(x_j)$ because the sample is comprised of independent RVs. More over they are identical so $E(x_i)E(x_j) = [E(X)]^2$. Therefore

$$E(\bar{x}^2) = \frac{1}{n}E(X^2) + \frac{n-1}{n}[E(X)]^2$$

Bringing it all together,

$$E(s^2) = \frac{n}{n-1}\cdot\left[E(X^2) - \frac{1}{n}E(X^2) - \frac{n-1}{n}[E(X)]^2\right]$$

$$= \frac{n}{n-1}\cdot\left[\frac{n-1}{n}E(X^2) - \frac{n-1}{n}[E(X)]^2\right]$$

$$\implies E(s^2) = E(X^2) - [E(X)]^2 \equiv \mathrm{Var}(X)$$

answered Mar 17 '15 at 0:47

Alecos Papadopoulos

26.4k 1 28 88