unbiased estimate of the covariance

Asked 2 years, 10 months ago Active 1 year, 5 months ago Viewed 8k times

7

1

How can I prove that

$$\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})$$

is an unbiased estimate of the covariance $\operatorname{Cov}(X,Y)$ where $\bar{X}=\dfrac{1}{n}\sum_{i=1}^{n}X_i$ and $\bar{Y}=\dfrac{1}{n}\sum_{i=1}^{n}Y_i$ and $(X_1,Y_1),\ldots,(X_n,Y_n)$ an independent sample from random vector (X,Y)?

statistics covariance

edited Nov 18 '16 at 0:13
Michael Hardy
214k 23 210 491

asked Nov 17 '16 at 23:00
Mama Bullki
86 1 6

1 Probably the reason why someone down-voted this question and someone voted to close it is that questions posted here should not be phrased in language suitable for assigning homework. – Michael Hardy Nov 18 '16 at 0:09

Do the multiplication, and deal with expectations of the resulting terms. – BruceET Nov 18 '16 at 0:58

2 One cannot show that it is an "unbiased estimate of the covariance". Perhaps you intend: `unbiased estimator of the covariance` – wolfies Nov 18 '16 at 3:26

@BruceET : Would you do something substantially different from what is in my answer posted below? – Michael Hardy Nov 18 '16 at 23:04

That seems to work nicely. Proof that $E(S^2)=\sigma^2$ is similar, but easier. Perhaps my clue was too simplistic (omitting the $-\mu+\mu=0$ trick). – BruceET Nov 18 '16 at 23:20

2 Answers

Let $\mu=\operatorname{E}(X)$ and $\nu=\operatorname{E}(Y)$. Then

$$\sum_{i=1}^{n}(X_i-\bar X)(Y_i-\bar Y)$$

$$=\sum_{i=1}^{n}\Big((X_i-\mu)+(\mu-\bar X)\Big)\Big((Y_i-\nu)+(\nu-\bar Y)\Big)$$

$$=\left(\sum_i (X_i-\mu)(Y_i-\nu)\right)+\left(\sum_i (X_i-\mu)(\nu-\bar Y)\right)$$
$$+\left(\sum_i (\mu-\bar X)(Y_i-\nu)\right)+\left(\sum_i (\mu-\bar X)(\nu-\bar Y)\right).$$

The expected value of the first of the four terms above is

$$\sum_i \operatorname{E}\Big((X_i-\mu)(Y_i-\mu)\Big)=\sum_i \operatorname{cov}(X_i,Y_i)=n\operatorname{cov}(X,Y).$$

The expected value of the second term is

$$\sum_i -\operatorname{cov}(X_i,\bar Y)=\sum_i -\operatorname{cov}\left(X_i,\frac{Y_1+\cdots+Y_n}{n}\right)$$

$$=\ -n\operatorname{cov}\left(X_1,\frac{Y_1+\cdots+Y_n}{n}\right)=-\operatorname{cov}(X_1,Y_1+\cdots+Y_n)$$

$$=-\operatorname{cov}(X_1,Y_1)+0+\cdots+0=-\operatorname{cov}(X,Y).$$

The third term is similarly that same number.

The fourth term is

$$\sum_i \overbrace{\operatorname{cov}(\bar X,\bar Y)}^{\text{No ``}i\text{'' appears here.}}=n\operatorname{cov}(\bar X,\bar Y)=n\operatorname{cov}\left(\frac1n\sum_i X_i,\frac1n\sum_i Y_i\right)$$

$$=n\cdot\frac{1}{n^2}\Big(\underbrace{\cdots+\operatorname{cov}(X_i,Y_j)+\cdots}_{n^2\text{ terms}}\Big).$$

This last sum is over all pairs of indices i and j. But the covariances are 0 except the ones in which $i=j$. Hence there are just n nonzero terms, and we have

I leave the rest as an exercise.

edited Apr 13 '18 at 20:39
Community ♦
1

answered Nov 18 '16 at 1:09
Michael Hardy
214k 23 210 491

6

Additional Comment, after some thought, following an exchange of Comments with @MichaelHardy:

His answer closely parallels the usual demonstration that $E(S^2) = \sigma^2$ and is easy to follow. However, the proof below, in abbreviated notation I hope is not too cryptic, may be more direct.

$$(n-1)S_{xy} = \sum (X_i - \bar{X})(Y_i - \bar{Y}) = \sum X_i Y_i - n\bar{X}\bar{Y} = \sum X_i Y_i - \frac{1}{n} \sum X_i \sum Y_i.$$

Hence,

$$\begin{aligned}(n-1)E(S_{xy}) &= E\left(\sum X_i Y_i\right) - \frac{1}{n} E\left(\sum X_i \sum Y_i\right) \\ &= n\mu_{xy} - \frac{1}{n}[n\mu_{xy} + n(n-1)\mu_x\mu_y] \\ &= (n-1)[\mu_{xy} - \mu_x\mu_y] = (n-1)\sigma_{xy},\end{aligned}$$

So the expectation of the sample covariance S_{xy} is the population covariance $\sigma_{xy} = \mathrm{Cov}(X, Y)$, as claimed.

Note that $\mathrm{E}(\sum X_i \sum Y_i)$ has n^2 terms, among which $\mathrm{E}(X_i Y_i) = \mu_{xy}$ and $\mathrm{E}(X_i Y_j) = \mu_x \mu_y$.

edited Nov 19 '16 at 16:23
Michael Hardy
214k 23 210 491

answered Nov 19 '16 at 6:39
BruceET
38k 7 17 43