The Derivation of
Sample Variance's unbiasedness

An article in LaTeX2e

Maiar

January 23, 2013

Contents

Chapter 1

Question

1.1 Variance

Well, and here begins my first article (actually a book, that's not a joke. . .) [1] written in LaTeX2e.

It's defined as follows:

$$p_i = P(X = x_i) \tag{1.1}$$

$$E(X) = \sum_{i=1}^{n} x_i p_i \tag{1.2}$$

$$D(X) = E\Big(\big(X - E(X)\big)^2\Big) = \sum_{i=1}^{n} \big(x_i - E(X)\big)^2 p_i \tag{1.3}$$

1.2 Sample Variance

We introduce a new variable $\overline{X}$ as equation(1.4) on page 1 shows:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{1.4}$$

Then

$$E(\overline{X}) = \frac{1}{n}E\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \mu \tag{1.5}$$

$$D(\overline{X}) = \frac{1}{n^2}D\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} D(X_i) = \frac{\sigma^2}{n} \tag{1.6}$$

---

[1] 这是我的第一个注脚。

Sample Variance is defined as equation(1.7):

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \tag{1.7}$$

1.3 Unbiasedness

The unbiasedness of Sample Mean and Sample Variance is as follows:

$$\begin{cases} E(\overline{X}) = E(X) \\ E(S^2) = D(X) \end{cases} \tag{1.8}$$

Now we get to solve it.

This page is intended to be blank!!

Chapter 2

Answer

2.1 Derivation

Firstly, we all know:

$$D(X) = E(X^2) - (E(X))^2$$
$$E(X^2) = D(X) + (E(X))^2$$

And the means and variances:

$$E(X_i) = E(\overline{X}) = \mu$$
$$D(X_i) = \sigma^2, D(\overline{X}) = \frac{\sigma^2}{n}$$

So we get

$$E(X_i^2) = D(X_i) + \mu^2 = \sigma^2 + \mu^2$$
$$E(\overline{X}^2) = D(\overline{X}) + \mu^2 = \frac{\sigma^2}{n} + \mu^2$$
$$\begin{aligned} E(\overline{X}X_i) &= E\Big(\frac{X_1X_i + X_2X_i + \cdots + X_i^2 + \cdots + X_nX_i}{n}\Big) \\ &= \frac{1}{n}\Big(\sum_{\substack{k=0 \\ k\neq i}}^{n} E(X_kX_i) + E(X_i^2)\Big) \\ &= \frac{1}{n}((n-1)\mu^2 + (\sigma^2 + \mu^2)) \\ &= \frac{\sigma^2}{n} + \mu^2 \\ &= E(\overline{X}^2) \end{aligned}$$

Now let's begin the derivation:

$$
\begin{aligned}
E(S^2) &= \frac{1}{n-1} E\left(\sum_{i=1}^{n} (X_i - \overline{X})^2\right) \\
&= \frac{1}{n-1} E\left(\sum_{i=1}^{n} \left(X_i^2 - 2\overline{X}X_i + \overline{X}^2\right)\right) \\
&= \frac{1}{n-1} \sum_{i=1}^{n} \left(E\left(X_i^2 - 2\overline{X}X_i + \overline{X}^2\right)\right) \\
&= \frac{1}{n-1} \left(2n\mu^2 + (n+1)\sigma^2 - 2\sum_{i=1}^{n} \left(E(\overline{X}X_i)\right)\right) \\
&= \frac{1}{n-1} \left(2n\mu^2 + (n+1)\sigma^2 - 2n(\frac{\sigma^2}{n} + \mu^2)\right) \\
&= \sigma^2 \\
&= D(X)
\end{aligned}
$$

That's all.
Hope you enjoy it.

2.2 Tips

What's the difference between x_i and X_i??
Think about it, when you get an answer, you can go back to click **the blank page** 3 and comment on my article
. . . and here it ends.