2. Biased and unbiased estimation for variance of Bernoulli variables

(a)

2/2 points (graded)

Let $X_1, \ldots, X_n$ be i.i.d. Bernoulli random variables, with unknown parameter $p \in (0,1)$. The aim of this exercise is to estimate the common variance of the X_i.

First, recall what $\mathsf{Var}(X_i)$ is for Bernoulli random variables.

$\mathsf{Var}(X_i)$ = p*(1-p) ✔ **Answer:** p*(1-p)

$p \cdot (1-p)$

Let $\overline{X}_n$ be the sample average of the X_i,

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

We are interested in finding an estimator for $\mathsf{Var}(X_i)$, and propose to use

$$\hat{V} = \overline{X}_n (1 - \overline{X}_n).$$

Check the correct statement that applies to $\hat{V}$:

- ○ $\hat{V}$ is not consistent because $\mathsf{Var}(X_i)$ is not linear in p
- ◉ $\hat{V}$ is consistent because of the Law of Large Numbers ✔
- ○ $\hat{V}$ is consistent because of the Central Limit Theorem

STANDARD NOTATION

Solution:

Let us compute the variance of a Bernoulli random variable:

$$\begin{aligned}\mathsf{Var}(X_i) &= \mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2 \\ &= \mathbb{E}[X_i] - \mathbb{E}[X_i^2] \qquad (X_i^2 = X_i) = \qquad p - p^2 \\ &= p(1-p)\end{aligned}$$

Now, we know that by the Law of Large Numbers, X_n is a consistent estimator for p, hence so is $\overline{X}_n(1 - \overline{X}_n)$ for $p(1-p) = \mathsf{Var}(X_i)$. There are no assumptions about the specific form of the variance in the LLN, except that the variable needs to have a mean. The Central Limit Theorem on the other hand tells us something about a rescaled version, but all we need here is the Law of Large Numbers.

提交 你已经尝试了2次（总共可以尝试2次）

Answers are displayed within the problem

(b)

1/2 points (graded)

Now, we are interested in the bias of $\hat{V}$. Compute:

$\mathbb{E}[\hat{V}] - \mathsf{Var}(X_i) =$ -1 * p*(1-p)/n ✔ **Answer:** -p*(1-p)/n

$-\frac{1\cdot p\cdot(1-p)}{n}$

Using this, find an unbiased estimator $\hat{V}'$ for $p(1-p)$ if $n \geq 2$.

Write barX_n for $\overline{X}_n$.

$\hat{V}' =$ barX_n*(1-barX_n) + bar ✖ **Answer:** n/(n-1)*barX_n*(1-barX_n)

" _ "

STANDARD NOTATION

Solution:

To compute the bias of the estimator, compute

$$\mathbb{E}\left[\overline{X}_n\left(1-\overline{X}_n\right)\right] = \mathbb{E}\left[\overline{X}_n\right] - \mathbb{E}\left[\overline{X}_n^2\right],$$

$$\begin{aligned}\mathbb{E}\left[\overline{X}_n\right] = &\ \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right]\\ = &\ \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \ p\end{aligned}$$

$$\begin{aligned}\mathbb{E}\left[\overline{X}_n^2\right] = &\ \mathsf{Var}\left(X_n\right) + \mathbb{E}[X_n]^2 = \ \frac{1}{n^2}\sum_{i=1}^{n}\mathsf{Var}\left(X_i\right) + p^2 && (X_i \text{ independent})\\ = &\ \frac{1}{n^2}\sum_{i=1}^{n} p(1-p) + p^2 && (X_i \text{ independent})\\ = &\ \frac{1}{n}p(1-p) + p^2\end{aligned}$$

Combined, we get

$$\begin{aligned}\mathbb{E}[\hat{V}] = &\ p - p^2 - \frac{1}{n}p(1-p)\\ = &\ \frac{n-1}{n}p(1-p)\end{aligned}$$

and therefore the bias is

$$\mathbb{E}[\hat{V}] - \mathsf{Var}(X_i) = \ -\frac{1}{n}p(1-p)$$

From the previous calculation, we observe that we can obtain an unbiased estimator if we compensate the multiplicative bias in $\hat{V}$, so set

$$\hat{V}' = \frac{n}{n-1}\overline{X}_n\left(1-\overline{X}_n\right).$$