7. Exercise: Strengths and Weaknesses of Estimators

You observe samples $X_1, \ldots, X_n \stackrel{iid}{\sim} \mathrm{Ber}(\theta)$ where $\theta \in (0,1)$ is an unknown parameter. Suppose that n is much larger than 1 so we have access to many samples from the specified distribution. Consider three candidate estimators for θ.

- $\overline{X}_n = \dfrac{1}{n}\sum_{i=1}^{n} X_i$
- 0.5

- X_1.

In the next three questions, you will consider potential strengths and weaknesses of these estimators.

In this particular section (just for now), the phrase “efficiently computable" refers to the existence of an explicit formula. More precisely, here we say that an estimator $\hat{\theta}_n$ is efficiently computable if there's a formula that takes as input $X_1, \ldots, X_n$ whose output is $\hat{\theta}_n$. (Not all estimators are efficiently computable, in this sense of the word. We may encounter examples of such estimators in a later unit.)

Strengths and Weaknesses of Estimators I

1/1 point (graded)

Which of the following is a potential **disadvantage** of using $\hat{\theta}_n = .5$ as an estimator for θ? (Choose all that apply.)

- [x] Unless $\theta = .5$, this estimator is biased.
- [x] Unless $\theta = .5$, this estimator is not consistent.
- [x] This estimator does not use any of the samples.
- [ ] This estimator is efficiently computable.

✔

Solution:

- Biased estimators inherently introduce errors in parameter estimation. Hence, having non-zero bias is a potential disadvantage of an estimator, so the first choice is correct.

- If the estimator does not converge to the true parameter as the sample size n grows very large, this is also a potential disadvantage. Estimators that are not consistent have inherently limited accuracy in approximating the true parameter, regardless of how many samples are taken. Hence, the second choice, "Unless $\theta = .5$, this estimator is not consistent," is also correct.

- The third choice, "This estimator does not depend on the sample, which is the only information that is given to us related to the distribution.", is correct. We haven't used any information given to us so we can't expect to learn anything new about the true parameter with this estimator.

- The fourth choice is not a disadvantage. The given estimator is indeed efficiently computable, and this is in general an *advantage* of certain estimators. In contrast, there are estimators that often don't have an explicit formula (for example, we may encounter such estimators in a later unit on Generalized Linear Models), and often requires approximate computation via a computer program.

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

Strengths and Weaknesses of Estimators II

1/1 point (graded)

Which of the following is a potential **disadvantage** of using $\hat{\theta}_n = X_1$ as an estimator for θ? (Choose all that apply.)

- [ ] This estimator is unbiased.
- [x] This estimator is not consistent.
- [x] This estimator uses only information given by only one sample, even though we have access to many samples.
- [x] The quadratic risk of this estimator does not tend to 0 as the sample size $n \to \infty$.

✔

Solution:

- The first option is incorrect: unbiasedness is in general an advantage of an estimator, and the estimator X_1 is unbiased because $\mathbb{E}[X_1] = \theta$.

- Since X_1 is Bernoulli, it is either 0 or 1, it is not equal to the true parameter θ which is assumed to lie in $(0,1)$. Hence, $\hat{\theta}_n = X_1$ is not consistent. This is a potential disadvantage of the estimator, so the second choice, "This estimator is not consistent.", is correct. Intuitively, inconsistent estimators inherently have limited accuracy in approximating the true parameter, no matter how many samples are collected. Hence, inconsistency is in general a disadvantage.

- As a statistician, it is generally best to use all information that is given about a distribution for parameter estimation. One sample (which we know is either 0 or 1) does not tell us much about the underlying distribution. Hence, it is a potential disadvantage that the samples $X_2, \ldots, X_n$ were not used to construct the estimator $\hat{\theta}_n = X_1$. The third choice, "This estimator uses only information given by one sample, even though we have access to many samples.", is thus correct.

- Note that $\mathrm{Var}[X_1] = \theta(1-\theta)$. Since

$$\text{quadratic risk} = \text{variance} + \text{bias}^2.$$

and X is unbiased, the quadratic risk is equal to the variance: $\theta(1-\theta)$. This estimator does not tend to θ as $n \to \infty$. This is a potential disadvantage, because if the quadratic risk does not go to 0, then $\hat{\theta}_n$ does not converge to θ in L^2. Intuitively, an estimator that does not converge in L^2 inherently has limited accuracy no matter how many samples are collected. Thus the fourth choice, "The quadratic risk of this estimator does not tend to 0 as the sample size $n \to \infty$.", is correct.

Remark: Convergence in L^2 is, mathematically speaking, a stronger guarantee than convergence in probability. That is, if $\hat{\theta}_n \xrightarrow{L^2} \theta$, then also, $\hat{\theta}_n \xrightarrow{\text{prob}} \theta$. Refer to Chapter 1 to review the different notions of convergence and their properties.

Submit

You have used 2 of 3 attempts

Answers are displayed within the problem

Strengths and Weaknesses of Estimators III

1/1 point (graded)

Which of the following are potential **advantages** of using $\hat{\theta}_n = \overline{X}_n$ as an estimator for θ? (Choose all that apply.)

- [x] This estimator is unbiased.
- [x] This estimator is consistent.

- [x] This estimator is efficiently computable.
- [x] The quadratic risk of this estimator tends to 0 as the sample size $n \to \infty$.

✔

Solution:

All of the above properties are potential advantages.

- Unbiased estimators avoid inherent inaccuracies of approximation that result from biasedness, so the first choice is correct.
- Consistent estimators become better and better approximations to the true parameter as the sample size increases. This is an advantage so the second choice is correct.
- Averages can be computed in linear time from the data, so is a computationally efficient estimator. In general, for applications, we need to work with estimators that can be computed efficiently, so the third choice is correct.
- The estimator is unbiased, so its variance is the same as its quadratic risk. By independence and the iid assumption,

$$\mathrm{Var}\left(\overline{X}_n\right) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}\left(X_1\right) = \frac{\theta\left(1-\theta\right)}{n}$$

  which tends to 0 as $n \to \infty$. Thus, $\hat{\theta}_n \overset{L^2}{\sim} \theta$, which ensures consistency. Hence, the last choice, "The quadratic risk of this estimator tends to 0 as the sample size $n \to \infty$." , is an advantage of this estimator.

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

Discussion

Hide Discussion

Topic: Unit 2 Foundation of Inference:Lecture 4: Parametric Estimation and Confidence Intervals / 7. Exercise: Strengths and Weaknesses of Estimators

Add a Post

Show all posts ▼ by recent activity ▼

There are no posts in this topic yet.

✖

Learn About Verified Certificates