Properties of Estimators III: Sufficiency and Consistency

Bruce M. Boghosian

Department of Mathematics

Tufts University

Outline

Bruce M. Boghosian

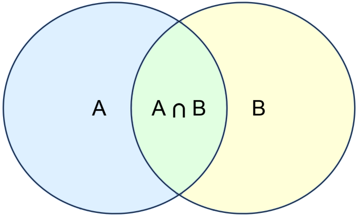
Properties studied to date

- Unbiasedness
- Efficiency
  - Relative efficiency
  - Absolute efficiency, Cramér-Rao bound

Recall joint and conditional probabilities

- *Joint probability, "A" and "B"*, is denoted $P(A \cap B)$ or $P(A, B)$
- *Conditional probability, "A" given "B"*, is denoted $P(A \mid B)$
- We must have $P(A, B) = P(A \mid B)P(B) = P(B \mid A)P(A)$, so

$$\boxed{P(A \mid B) = \frac{(A, B)}{P(B)} \quad \text{and} \quad P(B \mid A) = \frac{(A, B)}{P(A)}}$$



Sufficient estimators – Example 1

Bruce M. Boghosian

- You rolled a pair of dice, but are not allowed to see the outcome.
- You would like to know if the sum is an even number.
- Two people see the outcome, and they each give you information about it.
    - Person A tells you that the sum is ≤ 7.
    - Person B tells you that the sum is an odd number.
- Whose information was more helpful?

Sufficient estimators – Example 1

- Whose information was more helpful?

$$\begin{aligned} P(\text{Sum is even} \mid \text{Sum} \leq 7) &= \frac{P(2)+P(4)+P(6)}{P(2)+P(3)+P(4)+P(5)+P(6)+P(7)} \\ &= \frac{\frac{1}{36}+\frac{3}{36}+\frac{5}{36}}{\frac{1}{36}+\frac{2}{36}+\frac{3}{36}+\frac{4}{36}+\frac{5}{36}+\frac{6}{36}} \\ &= \frac{9}{21} \end{aligned}$$

$$P(\text{Sum is even} \mid \text{Sum is odd}) = 0.$$

- Clearly, Person B's information was more helpful.
- In fact, Person B's information was *sufficient*, whereas Person A's was not.

Sufficient estimators – Example 2

- Recall Bernoulli PDF: $p_X(k;p) = p^k(1-p)^k$ where $k = 0, 1$
- Sample n Bernoulli-distributed random numbers and find

$$X_1 = k_1, \quad X_2 = k_2, \quad X_3 = k_3, \quad \dots \quad , X_n = k_n$$

- Maximum likelihood estimator $\hat{p}(\vec{X}) = \frac{1}{n}\sum_{j=1}^{n} X_j$.
- Maximum likelihood estimate $p_e = \frac{1}{n}\sum_{j=1}^{n} k_j$.
- What does it mean to say that $\hat{p}$ is a *sufficient estimator* for p?

Sufficient estimators – Example 2

Bruce M. Boghosian

- What does it mean to say that $\hat{p}$ is a *sufficient estimator* for p?
- It means that
  - the conditional probability of the particular observation, given the estimate, is independent of p.
  - $P(X_1 = k_1, \ldots, X_n = k_n \mid \hat{p} = p_e)$ *is independent of* p.
  - while the joint probability $P(X_1 = k_1, \ldots, X_n = k_n)$ may depend on $\vec{k}$ and p, when conditioned on the observation made, namely $\hat{p} = p_e$, the dependence on p disappears.
  - everything the data can tell us about p is contained in the estimate p_e.
  - that the probability of the sample can be determined without knowing p.
- Returning to example, is $P(X_1 = k_1, \ldots, X_n = k_n \mid \hat{p} = p_e)$ independent of p?

Sufficient estimators – Example 2

Bruce M. Boghosian

Recap

Sufficiency

Consistency

Summary

- Is $P(X_1 = k_1, \ldots, X_n = k_n \mid \hat{p} = p_e)$ independent of p?
- We have

$$P(X_1 = k_1, \ldots, X_n = k_n \mid \hat{p} = p_e) = \frac{P(X_1 = k_1, \ldots, X_n = k_n, \hat{p} = p_e)}{P(\hat{p} = p_e)}$$

$$= \frac{P(X_1 = k_1, \ldots, X_n = k_n)}{P(\hat{p} = p_e)}$$

- Let's calculate the numerator and denominator separately.

Sufficient estimators – Example 2

- Numerator:

$$\begin{aligned} P(X_1 = k_1, \ldots, X_n = k_n) =& p^{k_1}(1-p)^{1-k_1} \cdots p^{k_n}(1-p)^{1-k_n} \\ =& p^{k_1+\cdots+k_n}(1-p)^{(1-k_1)+\cdots+(1-k_n)} \\ =& p^{np_e}(1-p)^{n-np_e} \end{aligned}$$

- Denominator:

$$P(\hat{p} = p_e) = P\left(\sum_{j=1}^{n} X_j = np_e\right) = \binom{n}{np_e} p^{np_e}(1-p)^{n-np_e}$$

- Quotient is independent of n so estimator $\hat{p}$ is sufficient:

$$P(X_1 = k_1, \ldots, X_n = k_n \mid \hat{p} = p_e) = \frac{p^{np_e}(1-p)^{n-np_e}}{\binom{n}{np_e} p^{np_e}(1-p)^{n-np_e}} = \frac{1}{\binom{n}{np_e}}$$

Sufficient estimators – Example 3

- Redo Example 2 with estimator $\hat{p}^*(\vec{X}) = X_1$, so $p_e^* = k_1$.
- Define $K := \sum_{j=1}^n k_j$.
- Find the conditional probability of the particular observation, given the estimate

$$
\begin{aligned}
P\left(X_1 = k_1, \ldots, X_n = k_n \mid \hat{p}^* = k_1\right) \\
&= \frac{p^K (1-p)^{n-K}}{p^{k_1}(1-p)^{1-k_1}} \\
&= p^{K-k_1}(1-p)^{n-1-K+k_1}.
\end{aligned}
$$

- This depends on p, so the estimator $\hat{p}$ is not sufficient.

Sufficient estimators – Example 4

- Recall the PDF:

$$f_Y(y) = \begin{cases} \frac{2y}{\theta^2} & \text{if } 0 \leq y \leq \theta \\ 0 & \text{otherwise} \end{cases}$$

- The method of moments estimator for this is $\hat{\theta} = \frac{3}{2}\overline{Y} = \frac{3}{2n}\sum_{j=1}^{n} Y_j$
- If this were sufficient, any two random samples, having the same value of θ_e should yield exactly the same information about θ.
- To demonstrate that is not the case, consider:
  - Case 1: $\vec{y} = \{3, 4, 5\}$ so that $\theta_e = \frac{3}{2} \cdot \frac{1}{3}(3 + 4 + 5) = 6$
  - Case 2: $\vec{y} = \{1, 3, 8\}$ so that $\theta_e = \frac{3}{2} \cdot \frac{1}{3}(1 + 3 + 8) = 6$
- In spite of the fact that $\theta_e = 6$ for both cases, note:
  - Based on Case 1, true θ **could** be equal to 7, because $y_1, y_2, y_3 < 7$.
  - Based on Case 2, true θ **could not** be equal to 7, because $y_3 = 8 > 7$.
- Hence, without even calculating the conditional probability, we know $\hat{\theta}$ is not sufficient.

Formal definition of sufficiency

- Let $X_1 = k_1, \ldots, X_n = k_n$ be a random sample of size n from $p_X(k;\theta)$. The statistic $\hat{\theta}(X_1, \ldots, X_n)$ is *sufficient* for θ if the likelihood function, $L(\theta)$, factors into the product of the pdf for $\hat{\theta}$ and a constant that does not involve θ,

$$L(\theta) = \prod_{j=1}^{n} p_X(k_j;\theta) = p_{\hat{\theta}}(\theta_e;\theta) b(k_1, \ldots, k_n).$$

- Example 1 again:
  - $P(X_1 = k_1, \ldots, X_n = k_n) = L(p) = p^K (1-p)^{n-K}$
  - $f_{\hat{p}}(p) = \binom{n}{np_e} p^K (1-p)^{n-K}$
  - Hence $L(p) = f_{\hat{p}}(p) \left[\binom{n}{np_e}\right]^{-1}$ where $p_e = \frac{1}{n}\sum_{j=1}^{n} k_j$.

Why is sufficiency desirable

- For any unknown pdf parameter, there will be an infinity of unbiased estimators.
- Some subset of these will be sufficient estimators, or functions of sufficient estimators.
- The variance of any unbiased estimator based on a sufficient estimator will be lower than that of any unbiased estimator that is not based on a sufficient estimator.
- Hence, sufficient estimators tend to be more efficient.

Consistency

- Recall from our calculation of the sample variance of the normal distribution

$$E\left[\frac{1}{n}\sum_{j=1}^{n}\left(Y_j-\overline{Y}\right)^2\right]=\frac{n-1}{n}\sigma^2\neq\sigma^2.$$

- As $n\to\infty$, this approaches σ^2, so we say that the estimator is *asymptotically unbiased*.
- For any estimator with finite n, however, say $\hat{\theta}_n$, we may also worry about the shape of the distribution of $\hat{\theta}_n$ in the vicinity of the actual value of θ.
- **Definition:** An estimator at finite n, call it $\hat{\theta}_n=h(W_1,\ldots,W_n)$ is said to be *consistent* if it *converges in probability* to θ,

$$\lim_{n\to\infty}P\left(\left|\hat{\theta}_n-\theta\right|<\epsilon\right)=1.$$

Using Chebyshev's Theorem to establish consistency

- Let W be any random variable with mean μ and variance σ^2. For any $\epsilon > 0$,

$$P(|W-\mu| < \epsilon) \geq 1 - \frac{\sigma^2}{\epsilon^2} \quad \text{or equivalently} \quad P(|W-\mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}.$$

- Suppose $X_1, \ldots, X_n$ is a random sample of size n from a discrete pdf with theoretical mean μ and variance $\sigma^2 < \infty$. Let $\hat{\mu}_n = \frac{1}{n}\sum_{j=1}^n X_j$. Is $\hat{\mu}_n$ consistent?
- By Chebyshev's inequality $P(|\hat{\mu}_n - \mu| < \epsilon) > 1 - \frac{\text{Var}(\hat{\mu})}{\epsilon^2}$
- But $\text{Var}(\hat{\mu}) = \frac{\sigma^2}{n}$, so $P(|\hat{\mu}_n - \mu| < \epsilon) > 1 - \frac{\sigma^2}{n\epsilon^2}$.
- For any ϵ and δ, an n can be found that makes $\frac{\sigma^2}{n\epsilon^2} < \delta$, so

$$\lim_{n\to\infty} P(|\hat{\mu}_n - \mu| < \epsilon) = 1 \qquad \text{(Weak law of large numbers)}$$

Summary

Bruce M. Boghosian

- We have defined and studied sufficiency of estimators.
- We have looked at examples of estimators that are sufficient and not sufficient.
- We have defined and studied consistency.
- We have used Chebyshev's Theorem to show that the sample mean is always a consistent estimator for the mean.