Tufts UNIVERSITY

Bruce M. Boghosian

Properties of Estimators

Sufficiency

Bruce M. Boghosian

Tufts UNIVERSITY | School of Arts and Sciences

Department of Mathematics

Tufts University

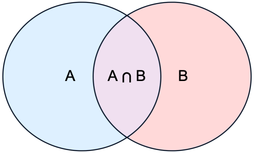
Outline

Bruce M. Boghosian

Joint and conditional probabilities

- *Joint prob., "A and B"*, denoted $P(A \cap B)$ or $P(A, B)$
- *Conditional prob., "A given B"*, denoted $P(A \mid B)$
- Demand $P(A, B) = P(A \mid B)P(B) = P(B \mid A)P(A)$, so

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



Sufficient estimators – Example 1

- You roll pair of dice, but 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?

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

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

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

Sufficient estimators – Example 2

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

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

- Maximum likelihood estimator $\hat{p}(\vec{X}) = \frac{1}{n}\sum_{j=1}^{n} X_j$.
- Maximum likelihood estimate $p_e = \frac{1}{n}\sum_{j=1}^{n} k_j$.
- Is $\hat{p}$ a *sufficient estimator* for p?

Sufficient estimators – Example 2

- Is $\hat{p}$ a *sufficient estimator* for p? What would this mean?
- It means that
  - conditional probability of the 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 data can tell us about p contained in est. p_e.
  - prob. of sample can be determined without knowing p.
- Returning to example, is the following independent of p?

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

Sufficient estimators – Example 2

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

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

- 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 $\overline{k} := \frac{1}{n}\sum_{j=1}^n k_j$.
- Find the conditional probability of the particular observation, given the estimate

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

- 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 $\hat{\theta}$ were sufficient, any two random samples, with same value of θ_e should yield exactly same information about θ.

Sufficient estimators – Example 4

- If $\hat{\theta}$ were sufficient, any two random samples, with same value of θ_e should yield exactly 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:
  - Case 1: true θ **could** be $= 7$, because $y_1, y_2, y_3 < 7$.
  - Case 2: true θ **could not** be $= 7$, because $y_3 = 8 > 7$.
- Hence, without even calculating the conditional probability, we know $\hat{\theta}$ is not sufficient.

Formal definition of sufficiency

Bruce M. Boghosian

- 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:
  - $L(p) = P(X_1 = k_1, \ldots, X_n = k_n) = p^K (1-p)^{n-K}$
  - $f_{\hat{p}}(p) = \binom{n}{np_e} p^K (1-p)^{n-K}$
  - 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.

The idea behind sufficiency

Bruce M. Boghosian

- Estimators, by their very nature, discard data, $\hat{\theta}(\vec{X})$
- In doing so, they accomplish a kind of *data reduction*.
- For example, if you can reduce 10^6 normally distributed numbers to a mean and a variance, you have accomplished substantial data reduction.
- You need *all* 10^6 numbers to estimate mean and variance, since you want to estimate those as accurately as possible.
- Once you have $\hat{\mu}(\vec{X})$ and $\hat{\sigma^2}(\vec{X})$, you don't need $\vec{X}$.
- A sufficient estimator does not *needlessly* discard data.
- If estimator $\hat{\theta}$ is sufficient, everything that can be known about the parameter θ has been extracted from the data, and nothing has been left behind.

Relation between sufficiency and other properties

Bruce M. Boghosian

- Given n pieces of data $\vec{X} = \langle X_1, \ldots, X_n \rangle$
- An estimator can be unbiased, but not sufficient
- $\hat{\mu}_n(\vec{X}) = X_1$ is unbiased,but not sufficient
  - Unbiased because $E(\hat{\mu}_n) = E(X_1) = \mu$
  - Not sufficient because it wastes $n-1$ of the numbers in the sample $\vec{X}$
- An estimator can be sufficient, but not unbiased
  - If $\hat{\theta}$ sufficient for θ, any invertible function of $\hat{\theta}$ is likewise.
  - e.g., $\hat{\theta}_2 = \hat{\theta}^3$ has the same information content as $\hat{\theta}$.
  - One would not expect $E\left(\hat{\theta}^3\right) = E\left[\left(\hat{\theta}\right)\right]^3$ so not unbiased

Summary

- We have defined and studied sufficiency of estimators.
- We have looked at examples of estimators that are sufficient and not sufficient.
- We have discussed the relationship between sufficiency and other properties of estimators.