Estimation II: Sufficiency

Stat 3202 @ OSU, Autumn 2018

Dalpiaz

The Main Idea

Suppose we have a random sample $Y_1, \ldots, Y_n$ from a $N(\mu, \sigma^2)$ population, with mean μ (unknown) and variance σ^2 (known).

To estimate μ, we have proposed using the sample mean $\bar{Y}$. This is a nice, intuitive, *unbiased* estimator of μ – but we could ask: does it *encode all the information we can glean from the data about the parameter* μ?

- Another way of asking this question: if I collected the data and calculated $\bar{Y}$, and I kept the data secret and only told you $\bar{Y}$, do I have *any more* information than you do about where μ is?

In this model, the answer is: $\bar{Y}$ *does* encode all the information in the data about the location of μ – there is nothing more we can get from the actual data values $Y_1, \ldots, Y_n$.

- We will call $\bar{Y}$ a **sufficient statistic** for μ.

Multivariate Probability Distributions

Last semester you learned a lot about probability (mass) functions (pmfs) for discrete random variables X, and probability density functions (pdfs) for continuous random variables X.

- For discrete variables, the pmf gives you the probability of observing a particular value:

$$p(x) = P(X = x)$$

- For two discrete variables, the *joint* pmf gives the probability of observing particular values for each variable:

$$p(x, y) = P(X = x \text{ and } Y = y)$$

Multivariate Probability Distributions

If you have an iid sample $X_1, \ldots, X_n$, from a population with pmf $f(x \mid \theta)$, then the *joint* pmf is the function that gives the probability of observing a particular array of values $x_1, \ldots, x_n$:

$$f(x_1, x_2, \ldots, x_n \mid \theta) = \prod_{i=1}^{n} f(x_i \mid \theta)$$

Later, we will consider the $x_1, \ldots, x_n$ known and instead consider θ as unknown, we would call this function the **likelihood**:

$$\mathcal{L}(\theta \mid x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f(x_i \mid \theta)$$

Conditional Distributions

You can also construct *conditional* pmfs, which give the probability of observing $X = x$ given that you have already observed $Y = y$:

$$p(x \mid y) = P(X = x \mid Y = y) = \frac{P(X = x \text{ and } Y = y)}{P(Y = y)}$$

Definition of Sufficiency

Let $Y_1, \ldots, Y_n$ denote a random sample from a probability distribution with unknown parameter θ. Then a statistic $U = g(Y_1, \ldots, Y_n)$ is said to be **sufficient** for θ if the conditional distribution of $Y_1, \ldots, Y_n$ given U, does not depend on θ.

- The intuition is that the statistic U contains all the information in the sample that is relevant for estimating θ.
- What is the conditional distribution of $Y_1, Y_2, \ldots, Y_n$ given U?

Example: Using the Definition of Sufficiency

Let $Y_1, Y_2, \ldots, Y_n$ be iid observations from a Poisson distribution with parameter λ. Show that $U = \sum_{i=1}^{n} Y_i$ is sufficient for λ.

An Anti-Example

If $X_1, X_2, X_3 \sim$ iid Bernoulli(θ), show that $Y = X_1 + 2X_2 + X_3$ is **not** a sufficient statistic for θ.

The Factorization Theorem

Let U be a statistic based on a random sample $Y_1, Y_2, \ldots, Y_n$. Then U is a sufficient statistic for θ if and only if the joint probability distribution or density function can be factored into two nonnegative functions,

$$f(y_1, y_2, \ldots, y_n|\theta) = g(u, \theta) \cdot h(y_1, y_2, \ldots, y_n),$$

where $g(u, \theta)$ is a function only of u and θ and $h(y_1, y_2, \ldots, y_n)$ is not a function of θ.

Poission Example, Again

Let $Y_1, Y_2, \ldots, Y_n$ be iid observations from a Poisson distribution with parameter λ. Show that $U = \sum_{i=1}^{n} Y_i$ is sufficient for λ.

Another Example

Let $X_1, X_2, \ldots, X_n$ be iid observations from a distribution

$$f(x \mid \theta) = \frac{\theta}{(1+x)^{\theta+1}}, \quad 0 < \theta < \infty, 0 < x < \infty$$

Find a sufficient statistic for θ.

Another Anti-Example

Let $X_1, X_2, \ldots, X_n$ be iid observations from a distribution

$$f(x \mid \theta) = \frac{1}{\pi(1 + (x - \theta)^2)}$$

Can you find a sufficient statistic for θ?

One-To-One Functions of Sufficient Statistics

Any one-to-one function of a sufficient statistic is sufficient.

Example: We found $U = \sum_{i=1}^{n} X_i$ is sufficient for λ in the Poisson example.

Thus, $\bar{X}$ is also sufficient for λ.

Bernoulli Example

Let $X_1, X_2, \ldots, X_n$ be iid from a Bernoulli distribution, such that $P(X_i = 1) = p$ and $P(X_i = 0) = 1 - p$, for each i. Note that $\mathrm{E}[X_i] = p$ and $\mathrm{E}[X_i] = p(1-p)$.

- Show that $\sum_{i=1}^{n} X_i$ is a sufficient statistic for p.
- Find an unbiased estimator of p that is also a sufficient statistic.
- We might try to estimate the variance of X_i by using the statistic $V = \bar{X}(1 - \bar{X})$. Is V a sufficient statistic for p?

Where We Are

So far, we have looked at some properties of estimators which we may use to judge estimators. Specifically, if we have an estimator, $\hat{\theta}$, of θ, it'd be nice if:

- $\hat{\theta}$ was unbiased for θ, or had low bias
- $\hat{\theta}$ had low variance
- $\hat{\theta}$ had low MSE
- $\hat{\theta}$ was consistent for theta
- $\hat{\theta}$ was sufficient for theta

These properties are useful to use to evaluate a single estimator $\hat{\theta}$, or to *compare* two estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ to decide which one is "better."

Next up, we will discuss methods for *finding* estimators for parameters.