Frequentist Estimation

Jacob M. Montgomery

2018

Frequentist point estimation (and more)

Overview

- Last we talked about
  - What is an MLE estimate?
  - What are the properties of these estimators?
  - How could we actually go about estimating them?
- Today we are are going to talk about frequentist statistics
  - "Simple" methods to make inferences using this approach
  - Some advanced approaches applicable both here and in MLE (the delta method and the parametric bootstrap)

Frequentist statistics

- The key here is to see that we have a *realized* quantity calculated from a *theoretical* distribution.

Frequentist statistics

- The key here is to see that we have a *realized* quantity calculated from a *theoretical* distribution.
- The weirdness is that we make inferences not based on the *realized* quantity but from the *theoretical* distribution which we cannot and do not know.

- In essence:
  1. We collect a sample and construct a sample statistic $\hat{\theta} = t(\mathbf{x})$.

- In essence:
  1. We collect a sample and construct a sample statistic $\hat{\theta} = t(\mathbf{x})$.
  2. We use our knowledge of probability theory (the CLT) to define the known distribution of $\hat{\Theta} = t(\mathbf{X})$, which is the function $t()$ applied to a theoretical sample of $\mathbf{X}$ from $\mathcal{F}$.

- In essence:
    1. We collect a sample and construct a sample statistic $\hat{\theta} = t(\mathbf{x})$.
    2. We use our knowledge of probability theory (the CLT) to define the known distribution of $\hat{\Theta} = t(\mathbf{X})$, which is the function $t()$ applied to a theoretical sample of $\mathbf{X}$ from $\mathcal{F}$.
    3. We try to make statements about how accurate our estimate will be given "an infinite sequence of future trials."

- In essence:
  1. We collect a sample and construct a sample statistic $\hat{\theta} = t(\mathbf{x})$.
  2. We use our knowledge of probability theory (the CLT) to define the known distribution of $\hat{\Theta} = t(\mathbf{X})$, which is the function $t()$ applied to a theoretical sample of $\mathbf{X}$ from $\mathcal{F}$.
  3. We try to make statements about how accurate our estimate will be given "an infinite sequence of future trials."

In essence, frequentists ask themsleves, "What would I see if I reran the same situation again (and again and again)?"
- Efron and Hastie

Discussion for frequentist inference

- Note that the inference here is not based *just* on the sample or the sample statistic we calculated.
- Rather, we are going to make statements about how often this procedure $t()$ will be accurate given repeated sampling that we will not be doing.

Example: Your undergraduate teaching hell

- Let $X_1, X_2, \ldots, X_n$ be an iid random sample from a population with mean μ and variance σ^2 where n is "large."

Example: Your undergraduate teaching hell

- Let $X_1, X_2, \ldots, X_n$ be an iid random sample from a population with mean μ and variance σ^2 where n is "large."
- We know from the central limit theorem that *in general* it will be true that

$$\bar{X} \sim N\left(\mu, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right).$$

Example: Your undergraduate teaching hell

- Let $X_1, X_2, \ldots, X_n$ be an iid random sample from a population with mean μ and variance σ^2 where n is "large."
- We know from the central limit theorem that *in general* it will be true that
$$\bar{X} \sim N\left(\mu, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right).$$
- Of course, we do not know either μ or σ^2.

Example: Your undergraduate teaching hell

- Let $X_1, X_2, \ldots, X_n$ be an iid random sample from a population with mean μ and variance σ^2 where n is "large."
- We know from the central limit theorem that *in general* it will be true that

$$\bar{X} \sim N\left(\mu, \left(\frac{\sigma}{\sqrt{n}}\right)^2\right).$$

- Of course, we do not know either μ or σ^2.
- So we simply "plug in" our unbiased estimates calculated from the sample:

$$\bar{X} \sim N\left(\bar{X}, \left(\frac{S}{\sqrt{n}}\right)^2\right).$$

Discussion: Your undergraduate teaching hell

- "The true parameter will fall in this interval 95% of the time."

Discussion: Your undergraduate teaching hell

- "The true parameter will fall in this interval 95% of the time." WRONG!
- "I am 95% sure that the true (unknown) parameter falls in this interval."

Discussion: Your undergraduate teaching hell

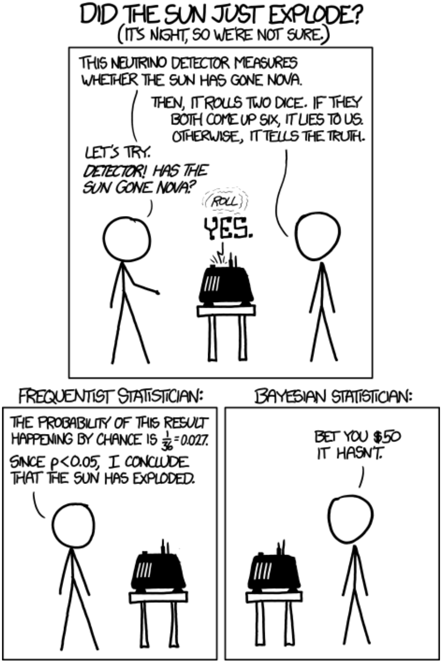
- "The true parameter will fall in this interval 95% of the time." WRONG!
- "I am 95% sure that the true (unknown) parameter falls in this interval." WRONG!
- If we followed the same procedure as above over, and over, and over again, 95% of the time the true parameter would fall within the confidence interval we constructed.

Discussion: Your undergraduate teaching hell

- "The true parameter will fall in this interval 95% of the time." WRONG!
- "I am 95% sure that the true (unknown) parameter falls in this interval." WRONG!
- If we followed the same procedure as above over, and over, and over again, 95% of the time the true parameter would fall within the confidence interval we constructed.
- The "confidence coefficient" represents the probability that the interval will capture the true parameter value in repeated samples,

Discussion: Your undergraduate teaching hell

- "The true parameter will fall in this interval 95% of the time." WRONG!
- "I am 95% sure that the true (unknown) parameter falls in this interval." WRONG!
- If we followed the same procedure as above over, and over, and over again, 95% of the time the true parameter would fall within the confidence interval we constructed.
- The "confidence coefficient" represents the probability that the interval will capture the true parameter value in repeated samples, even though we will only collect one sample.


DID THE SUN JUST EXPLODE?
(IT'S NIGHT, SO WE'RE NOT SURE.)
THIS NEUTRINO DETECTOR MEASURES WHETHER THE SUN HAS GONE NOVA.
THEN, IT ROLLS TWO DICE. IF THEY BOTH COME UP SIX, IT LIES TO US. OTHERWISE, IT TELLS THE TRUTH.
LET'S TRY.
DETECTOR! HAS THE SUN GONE NOVA?
(ROLL)
YES.
FREQUENTIST STATISTICIAN:
THE PROBABILITY OF THIS RESULT HAPPENING BY CHANCE IS 1/36 = 0.027.
SINCE p<0.05, I CONCLUDE THAT THE SUN HAS EXPLODED.
BAYESIAN STATISTICIAN:
BET YOU $50 IT HASN'T.

Class example: Difference of means with pooled variance

Let $X_{11}, X_{12}, \ldots, X_{1n}$ be an iid random sample from a normal population with mean μ_1 and variance σ^2. Let $X_{21}, X_{22}, \ldots, X_{2n}$ be an iid random sample from a normal population with mean μ_2 and variance σ^2. We want to create a 95% confidence interval for the difference in means, $\mu_2 - \mu_1$.

- An unbiased estimator of $\mu_2 - \mu_1$ is $\bar{X}_2 - \bar{X}_1$.
- An unbiased estimator of σ is

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n_1}(X_{1i} - \bar{X}_1)^2 + \sum_{i=1}^{n_1}(X_{2i} - \bar{X}_2)^2}{n_1 + n_2 - 2}$$

.

You don't know how to do this. But help me work out how to calculate a 95% CI for this quantity of interest.

The Delta method

- Sometimes we want to talk about the distribution of a statistic that is a **function** of some variable whose limiting distribution we know.

The Delta method

- Sometimes we want to talk about the distribution of a statistic that is a **function** of some variable whose limiting distribution we know.
- In this case, we cannot use the "plug in" method directly but must instead use local linear approximations.

The Delta method

- Sometimes we want to talk about the distribution of a statistic that is a **function** of some variable whose limiting distribution we know.
- In this case, we cannot use the "plug in" method directly but must instead use local linear approximations.
- This is known as the "Delta method", and is based on the Taylor-series approximation.

The Delta method in broad strokes

- Let's say that we are trying to estimate some population parameter that is a function of the standard parameter of interest.

$$\tau = g(\theta)$$

  , where $g(\theta)$ is some smooth function.

The Delta method in broad strokes

- Let's say that we are trying to estimate some population parameter that is a function of the standard parameter of interest.

  $$\tau = g(\theta)$$

  , where $g(\theta)$ is some smooth function.
- We want to find the limiting distribution of τ.

The Delta method in broad strokes

- Let's say that we are trying to estimate some population parameter that is a function of the standard parameter of interest.

  $$\tau = g(\theta)$$

  , where $g(\theta)$ is some smooth function.
- We want to find the limiting distribution of τ.
- We want to do the following:
  - Show that the MLE for $\hat{\tau} = g(\hat{\theta})$
  - Find the variance for $\hat{\tau}$.

Equivariance of the MLE

- One of the properties of the MLE that we did not discussed is **equivariance**.
- Let $\tau = g(\theta)$ be a (smooth) function of θ. Let $\hat{\theta}$ be the MLE of θ. Then $\hat{\tau} = g(\hat{\theta})$ is the MLE of τ.
- See proof of Wasserman pg. 128

Example: Equivariance of the MLE

Let $X_1, \ldots, X_n \sim N(\theta, 1)$. The MLE for θ is $\hat{\theta} = \bar{X}$. Find the MLE for $\tau = \theta^2$.

Now we just need to find the variance for our estimate of τ.

- Let T be random variables (statistics) with mean θ.

Now we just need to find the variance for our estimate of τ.

- Let T be random variables (statistics) with mean θ.
- For concreteness say T is the MLE for θ, although this is not required below.

Now we just need to find the variance for our estimate of τ.

- Let T be random variables (statistics) with mean θ.
- For concreteness say T is the MLE for θ, although this is not required below.
- To get the variance

$$Var(\tau) \approx E([g(T) - g(\theta)]^2)$$

- We cannot calculate this directly. But we can assume that $g(T)$ is *near* $g(\theta)$ and use something called a Taylor polynomial to approximate this quantity.

Taylor polynomials

If a function $g(\theta)$ has derivatives of order r, then for any constant a^{th} Taylor polynomial of order r about a is

$$Taylor(\theta) = \sum_{i=0}^{r} \frac{g^{(i)}(a)}{i!}(\theta - a)^i$$

Taylor polynomials

If a function $g(\theta)$ has derivatives of order r, then for any constant a^{th} Taylor polynomial of order r about a is

$$Taylor(\theta) = \sum_{i=0}^{r} \frac{g^{(i)}(a)}{i!}(\theta - a)^i$$

- In one dimension, it turns out that you can approximate any function evaluated *near* (but not at) point θ as:

$$g(\theta) \approx \sum_{i=0}^{n} \frac{f^{(i)}(a)}{i!}(\theta - a)^i$$

- Example:

$$g(0+\theta) \approx g(0) + g'(0)(\theta - 0) + g''(0)(\theta - 0)^2 + g'''(0)(\theta - 0)^3 \ldots$$

- Taylor's theorem shows that these higher order terms always tends towards 0 fast so that we can ignore them.
- $$g(0+\theta) \approx g(0) + g'(0)(\theta - 0) + R$$

- The idea is that we have some function we want to evaluate at a point a.
- We do not know the behavior of the function at that point. But we do know how to evaluate at the function at some point close to a, which is θ.
- We can just approximate the evaluation this way:

$$g(a) = g(\theta) + g'(\theta)(a - \theta) + R$$

- Note that, if a depends on the data and θ does not and $E(T) = \theta$,

$$E(g(a)) = g(\theta) + g'(\theta)E((a - \theta))$$

- $$E(g(a)) = g(\theta)$$

Back to the variance

- Now we know that

$$Var(\tau) \approx E([g(T) - g(\theta)]^2)$$

- We also know that

$$g(T) \approx g(\theta) + g'(\theta)(T - \theta)$$

- So we know that

$$Var(\tau) = E[\left(g'(\theta)(T - \theta)\right)^2]$$

- Which can be re-written as

$$Var(\tau) = g'(\hat{\theta})^2 Var(\hat{\theta})$$

- This implies that $\tau \sim N(g(\theta), g'(\theta)^2 Var(\theta))$
- Using the plug in method, this implies that

$$\tau \sim N(g(\hat{\theta}), g'(\hat{\theta})^2 Var(\hat{\theta}))$$

Example 1: Inference on the odds

Suppose we observe $X_1, \ldots, X_n$ iid Bernoulli(p) random variables. We might be interested in $\tau = \frac{p}{1-p}$, which is the odds of success. So if $p = 2/3$ then the event has a 2:1 odds of happening. Find the asymptotic distribution of for τ.

Example 2: Inference on the inverse mean

Suppose that $X_1, \ldots, X_n$ are iid Normal data with mean μ and variance σ^2. Say we wish to make an inverse on $\tau = \frac{1}{\mu}$. Find the approximation for the asymptotic distribution of τ.

Parametric bootstrap

- Sometimes we might know $\hat{\theta}$, but be unable to (or unwilling to) calculate the asymptotic variance
- In this case we can use the **parametric bootstrap** to estimate the asyptotic variance.

The procedure is rather simple:

1. Find $\hat{\theta}$
2. Generate a new sample based on $\hat{\theta}$ assuming that our paramteric model is correct.
3. Based on this simulated sample, calculate our quantities of interest
   - We can calculate $\hat{\mu}^*$
   - We can calculate $\hat{\sigma}^*$
   - We can even calculate $\hat{\tau}^* = g(\hat{\mu}^*, \hat{\sigma}^*)$
4. Repeat this B times and calculate

$$\hat{se}_{boot} = \sqrt{\frac{\sum_{b=1}^{B}(\hat{\tau}^* - \hat{\tau})^2}{B}}$$

Example 1: Inference on the odds

Suppose we observe $X_1, \dots, X_n$ iid Bernoulli(p) random variables. We might be interested in $\tau = \frac{p}{1-p}$, which is the odds off success. So if $p = 2/3$ then the even has a 2:1 odds of happening. Use the parametric bootstrap to estimate the asymptotic distribution of τ.

Example 2: Inference on the inverse mean

Suppose that $X_1, \ldots, X_n$ are iid Normal data with mean μ and variance σ^2. Say we wish to make an inverse on $\tau = \frac{1}{\mu}$. Use the parametric bootstrap to estimate the asymptotic distribution of τ.

Class business

- For next chapter, read Wasserman Chapter 11.
- Problem set forthcoming. Due one week from distribution. Complaints welcome.