More MLE

CSE 312 Summer 25
Lecture 22

Biased

One property we might want from an estimator is for it to be **unbiased**.

An estimator $\hat{\theta}$ is "unbiased" if

$$\mathbb{E}[\hat{\theta}] = \theta$$

The expectation is taken over the randomness in the samples we drew.

The formula is fixed, the data we draw to evaluate the formula becomes the source of the randomness.

So we're not consistently overestimating or underestimating.

If an estimator isn't unbiased then it's **biased**.

Are our MLEs unbiased?

$$\widehat{\theta_\mu} = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$\mathbb{E}\left[\widehat{\theta_\mu}\right] = \frac{1}{n}\mathbb{E}[\sum X_i] = \frac{1}{n}\sum \mathbb{E}[X_i] = \frac{1}{n} \cdot n \cdot \mu = \mu$$

Unbiased!

Here, think of the X_i not as data points, but as random variables.

Made them capital letters so we see the randomness again.

Are our MLEs biased?

Our estimate for the coin-flips (if we generalized a bit) would be

$$\frac{\text{num heads}}{\text{total flips}}$$

Is this biased or unbiased?

Are our MLEs biased?

Our estimate for the coin-flips (if we generalized a bit) would be

$$\frac{\text{num heads}}{\text{total flips}}$$

What is $\mathbb{E}\left[\frac{\text{num heads}}{\text{total flips}}\right]$?

$$\mathbb{E}\left[\frac{\text{num heads}}{\text{total flips}}\right] = \frac{\theta \cdot n}{n} = \theta$$

Unbiased!

Summary

If you get independent samples $x_1, x_2, \ldots, x_n$ from a $\mathcal{N}(\mu, \sigma^2)$ where μ and σ^2 are unknown, the maximum likelihood estimates of the normal is:

$\widehat{\theta_\mu} = \frac{\sum_{i=1}^{n} x_i}{n}$ and $\widehat{\theta_{\sigma^2}} = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \widehat{\theta_\mu}\right)^2$

The maximum likelihood estimator of the mean is the **sample mean** that is the estimate of μ is the average value of all the data points.

The MLE for the variance is: the variance of the experiment "choose one of the x_i at random"

Unbiased?

$\mathbb{E}[\theta_{\sigma^2}] = \mathbb{E}[\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \widehat{\theta_\mu}\right)^2]$

$= \frac{1}{n}\mathbb{E}\left[\sum\left(x_i - \widehat{\theta_\mu}\right)^2\right]$

$= \frac{1}{n}\mathbb{E}\left[\sum x_i^2 - 2x_i\widehat{\theta_\mu} + \widehat{\theta_\mu}^2\right]$

…

Then an algebraic miracle occurs…

$= \frac{n-1}{n} \cdot \sigma^2$ where $\sigma^2 = \mathbb{E}[(x_i - \mathbb{E}[x_i])\text{^}2]$

Intuition for the algebra miracle:
$\widehat{\theta_\mu} = \sum x_i/n$. So when that gets squared, there are terms that have $x_i x_j$ terms and $x_i \cdot x_i$ terms.
The $1/n$ fraction of terms that are $x_i x_i$ decrease the variance because you can't deviate from yourself.

Optional: Algebra

Showing MLE of Variance is biased

That Algebraic Miracle

$$= \frac{1}{n}\mathbb{E}\left[\sum x_i^2 - 2x_i\widehat{\theta_\mu} + \widehat{\theta_\mu}^2\right]$$

$$= \frac{1}{n}\mathbb{E}\left[\sum x_i^2\right] - \frac{1}{n}\mathbb{E}\left[\sum 2x_i\widehat{\theta_\mu} - \sum\widehat{\theta_\mu}^2\right]$$

$$= \frac{1}{n}n\mathbb{E}[x_1^2] - \frac{1}{n}\mathbb{E}\left[2\widehat{\theta_\mu}\sum x_i - \sum\widehat{\theta_\mu}^2\right]$$

$$= \mathbb{E}[x_1^2] - \frac{1}{n}\mathbb{E}\left[2n\widehat{\theta_\mu}^2 - n\widehat{\theta_\mu}^2\right]$$

$$\widehat{\theta_\mu} = \sum x_i/n$$

$$= \mathbb{E}[x_1^2] - \frac{1}{n}\mathbb{E}\left[n\widehat{\theta_\mu}^2\right]$$

$$= \mathbb{E}[x_1^2] - \mathbb{E}\left[\widehat{\theta_\mu}^2\right]$$

More of That Algebraic Miracle

$$\mathbb{E}\left[\widehat{\theta_\mu}^2\right] = \mathbb{E}\left[\left(\frac{\sum x_i}{n}\right)\left(\frac{\sum x_i}{n}\right)\right]$$

These are the $x_i x_i$ terms.

$$= \frac{1}{n^2}\mathbb{E}\left[\sum_{i \neq j} x_i \cdot x_j + \sum_i x_i^2\right]$$

$$= \frac{1}{n^2}\mathbb{E}\left[\sum_{i \neq j} x_i \cdot x_j\right] + \frac{1}{n^2}\mathbb{E}\left[\sum_i x_i^2\right]$$

$$= \frac{1}{n^2} \cdot n(n-1)\mathbb{E}[x_1 \cdot x_2] + \frac{1}{n^2} n\mathbb{E}[x_1^2]$$

$$= \frac{n-1}{n}\mathbb{E}[x_1]\mathbb{E}[x_1] + \frac{1}{n}\mathbb{E}[x_1^2]$$

This is where the $x_i x_i$ terms end up

Wrapping Up the Algebraic Miracle

$\mathbb{E}[\theta_{\sigma^2}] = \mathbb{E}[x_1^2] - \mathbb{E}\left[\widehat{\theta_\mu}^2\right]$

Plugging in $\mathbb{E}\left[\widehat{\theta_\mu}^2\right] = \frac{n-1}{n}\mathbb{E}[x_1]\mathbb{E}[x_1] + \frac{1}{n}\mathbb{E}[x_1^2]$ we get:

$\mathbb{E}[\theta_{\sigma^2}] = \mathbb{E}[x_1^2] - \left(\frac{n-1}{n}\mathbb{E}[x_1]\mathbb{E}[x_1] + \frac{1}{n}\mathbb{E}[x_1^2]\right)$

$= \mathbb{E}[x_1^2] - \frac{n-1}{n}\mathbb{E}[x_1]^2 - \frac{1}{n}\mathbb{E}[x_1^2]$

$= \frac{n-1}{n}\mathbb{E}[x_1^2] - \frac{n-1}{n}\mathbb{E}[x_1]^2$

$= \frac{n-1}{n}\mathrm{Var}(x_1)$

Non-optional Takeaways

Not Unbiased

$$\mathbb{E}\left[\widehat{\theta_{\sigma^2}}\right] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \widehat{\theta_\mu}\right)^2\right]$$

$$= \frac{n-1}{n}\sigma^2$$

Which is not what we wanted. This is a biased estimator. But it's not too biased…

An estimator $\hat{\theta}$ is "consistent" if

$$\lim_{n\to\infty} \mathbb{E}\left[\hat{\theta}\right] = \theta$$

The MLE is consistent (under some very mild assumptions),
but it can be biased or unbiased.

Correction

The MLE slightly underestimates the true variance.

You could correct for this! Just multiply by $\frac{n}{n-1}$.

This would give you a formula of:

$\frac{n}{n-1} \cdot \frac{1}{n} \sum_{i=1}^{n} \left(x_i - \widehat{\theta_\mu}\right)^2$

$= \frac{1}{n-1} \sum_{i=1}^{n} \left(x_i - \widehat{\theta_\mu}\right)^2$ where $\widehat{\theta_\mu}$ is the sample mean.

Called the "sample variance" because it's the variance you estimate if you want an (unbiased) estimate of the variance given only a sample.

If you took a statistics course, you probably learned the square root of this as the definition of standard deviation.

Are our MLE’s accurate?

Confidence for MLEs

We said our MLE for "probability of heads on a flip" is $\widehat{p} = \frac{\text{num heads}}{\text{num flips}}$

And $\mathbb{E}[\hat{p}] = p$. (where p is the true probability of heads).

But how close is it to the true value? What if on-average it's correct, but it's often very far away.

If only we had a tool…one that would describe the probability of being far from your expectation…