Point Estimates and Sampling Variability

Parameter estimation

- We are often interested in *population parameters*.
- Since complete populations are difficult (or impossible) to collect data on, we use *sample statistics* as *point estimates* for the unknown population parameters of interest.

Suppose we are interested in estimating the distribution of heights of adults in the US. Suppose we assume that adult height follows a Normal distribution with unknown mean μ and standard deviation $10cm$

Suppose we randomly sample 1,000 adults from each state in the US. What is the best estimate for μ?

Maximum Likelihood estimation

Assume that the first person we sample is 176cm.

If $\mu = 170$, $f(x_1; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{1}{2}\frac{(176-170)^2}{10}} =$

If $\mu = 150$, $f(x_1; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{1}{2}\frac{(176-150)^2}{10}} =$

For every possible $\mu \in (-\infty, \infty)$

$$f(x_1; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{1}{2}\frac{(176-\mu)^2}{10}} =$$

Maximum Likelihood estimation

Assume that the first person we sample is 176cm.

If $\mu = 170$, $f(x_1; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(176-170)^2}{10}} =$

If $\mu = 150$, $f(x_1; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(176-150)^2}{10}} =$

For every possible $\mu \in (-\infty, \infty)$

$$f(x_1; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(176-\mu)^2}{10}} =$$

Maximum Likelihood estimation

Assume that the first person we sample is 176cm.

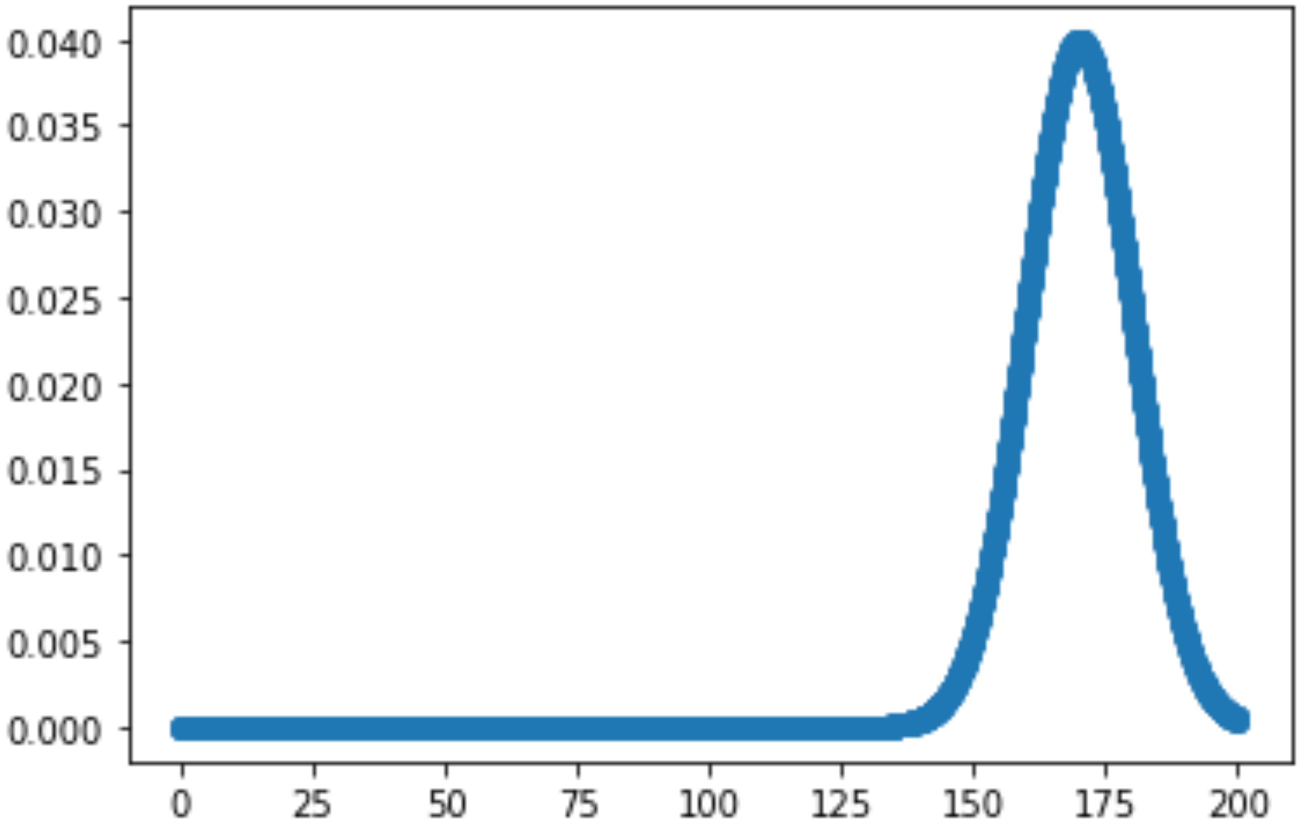
If $\mu = 170$, $f(x_1; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(176-170)^2}{10}} = 0.033$

If $\mu = 150$, $f(x_1; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{(176-150)^2}{10}} = 0.0013$

For every possible $\mu \in (-\infty, \infty)$

$$f(x_1; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{1}{2}\frac{(176-\mu)^2}{10}}$$

Likelihood for a single sample



Likelihood for multiple samples

For multiple samples $\{x_1, \ldots, x_n\}$

$$f(x_1, \ldots, x_n | \mu, \sigma^2) = f(x_1 | \mu, \sigma^2) \ldots f(x_n | \mu, \sigma^2) =$$

$$= \prod_{i=1}^{n} f(x_i | \mu, \sigma^2)$$

Likelihood for multiple samples

For multiple samples $\{x_1, \ldots, x_n\}$

$$f(x_1, \ldots, x_n | \mu, \sigma^2) = f(x_1|\mu, \sigma^2) \ldots f(x_n|\mu, \sigma^2) =$$

$$= \prod_{i=1}^{n} f(x_i|\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{1}{2}\frac{(x_i - 170)^2}{10}}$$

Find $\hat{\mu} = argmax_\mu f(x_1, \ldots, x_n|\mu)$,

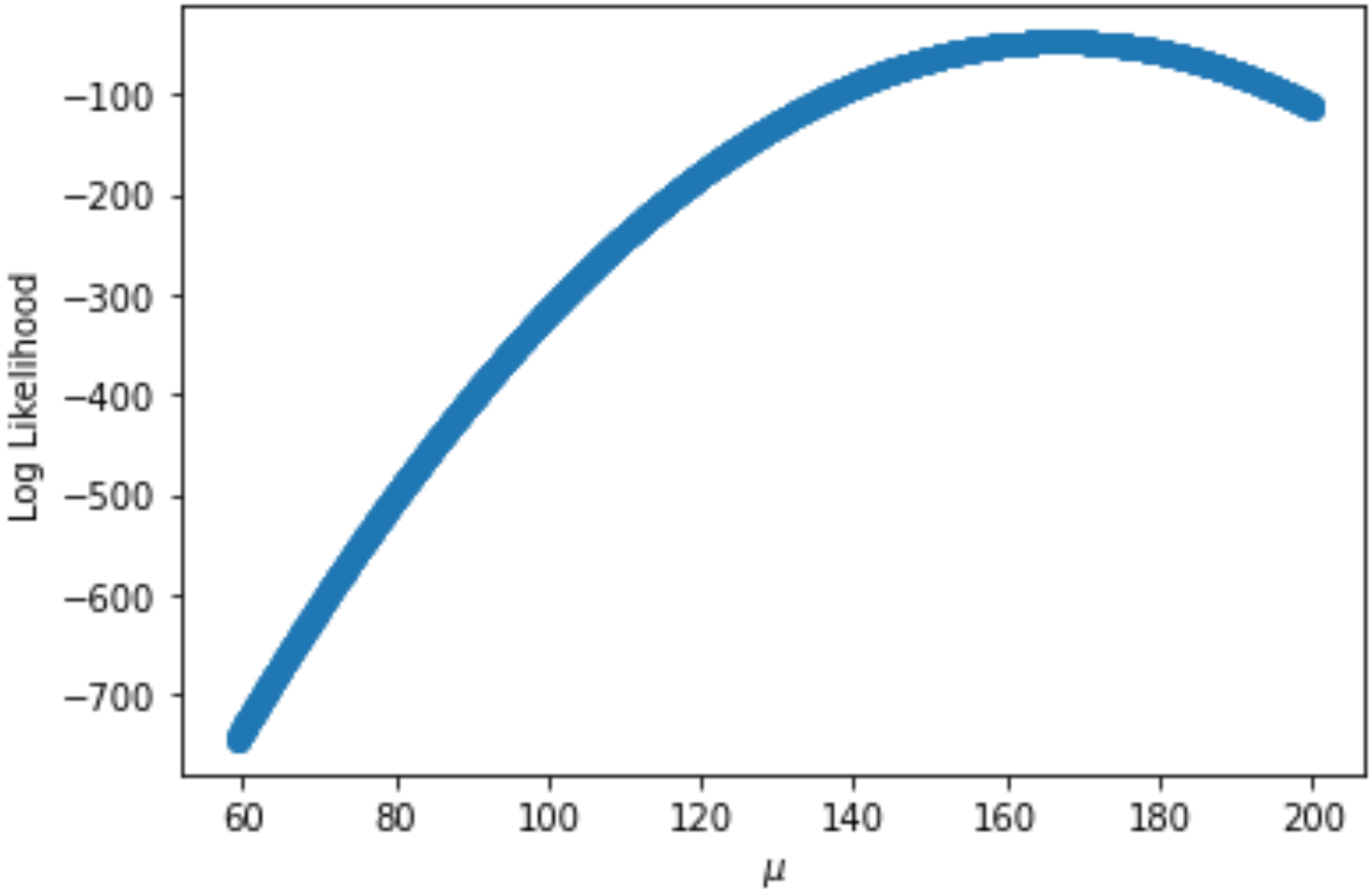
Steps to MLE

- Compute the likelihood $L(x;\theta)$
- Usually, it is more convenient to work with $LL(x;\theta) = \log\big(L(x;\theta)\big)$
- The log is monotonic:

  if $\theta^* = argmax_\theta \log f(\theta)$, then $\theta^* = argmax_\theta f(\theta)$

- Maximize $LL(x;\theta)$:
  - Find θ^* for which $dLL(x;\theta)/d\theta$ is zero, check that it is a maximum.
  - If there is no closed form for $dLL(x;\theta)/d\theta$, use numerical methods.

MLE for normal with known σ^2



MLE for Uniform

- Let $X_1, \dots, X_n$ be i.i.d. Uniform $[0, \theta]$, where $\theta > 0$. Find $\hat{\theta}$
- Let $X_1, \dots, X_n$ be i.i.d. Uniform $[\theta, \theta + 1]$. Find $\hat{\theta}$

MLE for Uniform

- Let $X_1, \dots, X_n$ be i.i.d. Uniform $[0, \theta]$, where $\theta > 0$. Find $\hat{\theta}$
- Let $X_1, \dots, X_n$ be i.i.d. Uniform $[\theta, \theta + 1]$. Find $\hat{\theta}$

MLE estimator may not exist
MLE estimator may not be unique

MLE for Bernoulli

Life rating in Greece. Greece has faced a severe economic crisis since the end of 2009. A Gallup poll surveyed 1,000 randomly sampled Greeks in 2011 and found that 25% of them said they would rate their lives poorly enough to be considered suffering.

Describe the parameter of interest. What is the value of the point estimate of this parameter?

Data: Would you rate your life poorly enough to be considered suffering?

X_1	X_2	X_3	…	X_{999}	X_{1000}
Yes	No	No	…	No	Yes

Point estimators are RVs

An estimator (MLE or otherwise) is an RV and a statistic (a function of the sample)

The estimator for the Bernoulli parameter is the sample mean $\overline{X_n} = \frac{\sum_i X_i}{n}$

Let's talk about the distribution of this RV

Desired properties of estimators

- Unbiased: $E\left[\widehat{\Theta}\right] = \theta$
  - Not always true for MLE estimator

- Consistent: $\Theta_n \rightarrow \theta$
  - True under mild conditions for MLE estimators

- Low Mean Squared Error
  - $E\left[\left(\widehat{\Theta} - \theta\right)^2\right] = \mathrm{Var}\left(\widehat{\Theta}\right) + Bias\left(\widehat{\Theta}\right)^2$ is low.

Let’s poll!

- We want to poll how many drinks University students drink every week.

- Go to https://polyhedron.math.uoc.gr/2223/moodle/mod/data/view.php?id=428

- Fill in the number of drinks you drink per week.