# The Law of Large Numbers and Deviation Inequalities

Probability and Statistics for Data Science

Carlos Fernandez-Granda

NYU COURANT INSTITUTE OF MATHEMATICAL SCIENCES

NYU DATA SCIENCE

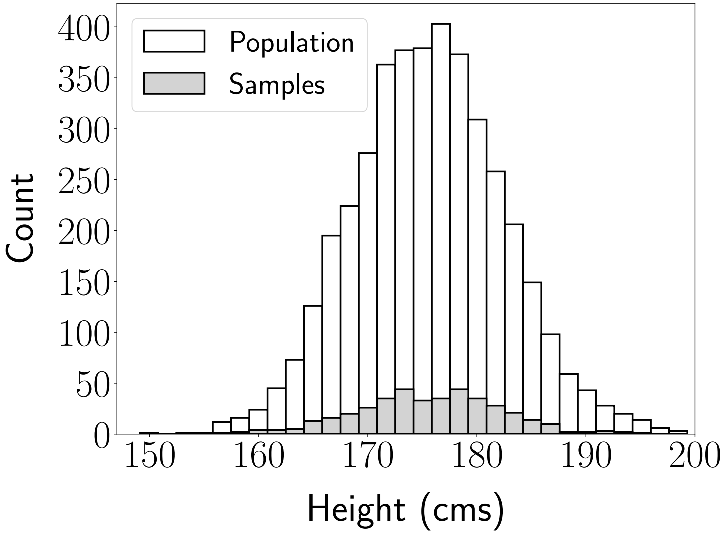
These slides are based on the book Probability and Statistics for Data Science by Carlos Fernandez-Granda, available for purchase here. A free preprint, videos, code, slides and solutions to exercises are available at https://www.ps4ds.net

# Estimation of population parameters

Simple idea: Choose a random subset of the population

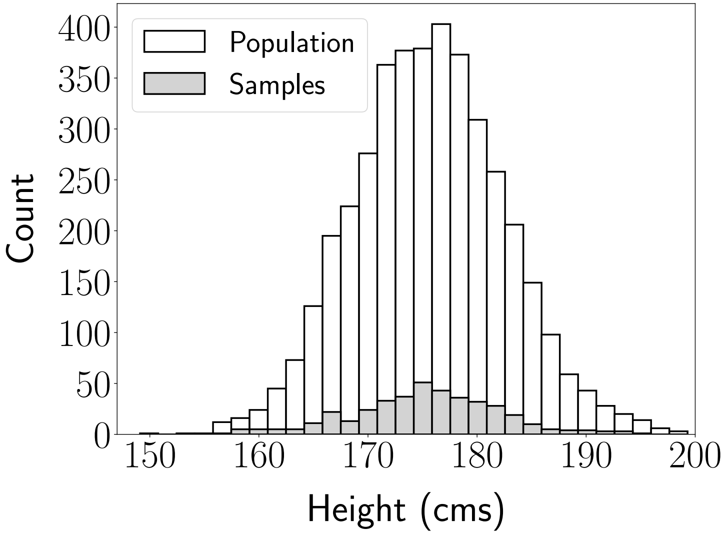
# Random sampling

Sample mean = 175.5 ($\mu_{\text{pop}} = 175.6$)



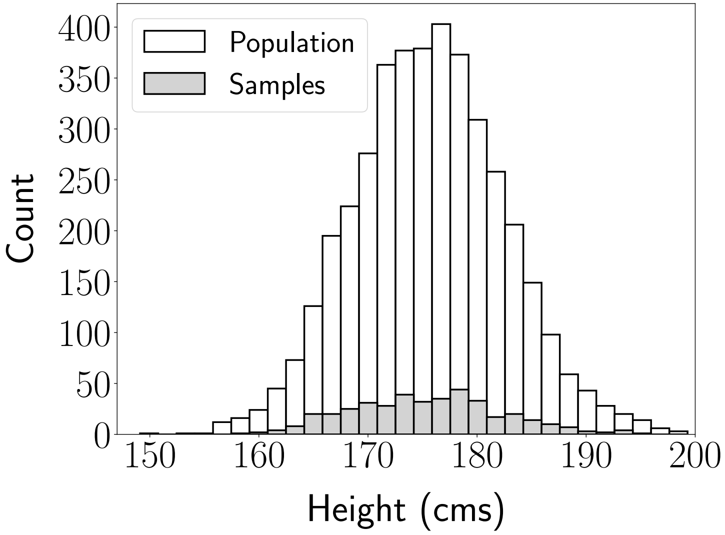
# 400 random samples

Sample mean = 175.2 ($\mu_{\text{pop}} = 175.6$)

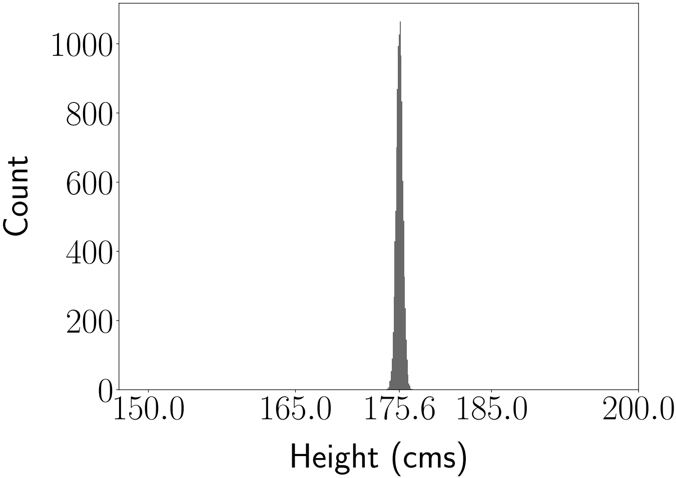


# 400 random samples

Sample mean = 176.1 ($\mu_{\text{pop}} = 175.6$)



# Sample means of 10,000 subsets of size 400



# Sample mean

Population mean: $\mu_{\text{pop}}$

Population variance: $\sigma^2_{\text{pop}}$

Random samples selected independently and uniformly at random with replacement: $\tilde{x}_1$, $\tilde{x}_2$, ..., $\tilde{x}_n$

$$\tilde{m} := \frac{1}{n}\sum_{i=1}^{n} \tilde{x}_i$$

$$\mathrm{E}\left[\tilde{m}\right] = \mu_{\text{pop}}$$

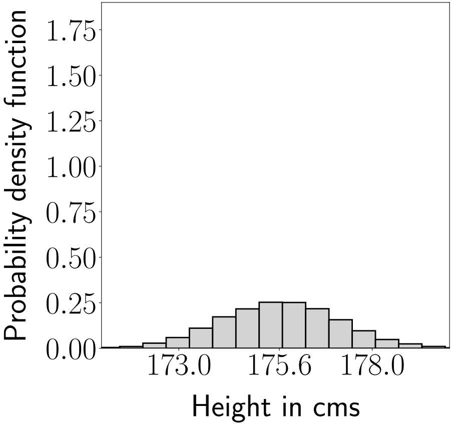
$$\text{se}\left[\widetilde{m}\right] = \frac{\sigma_{\text{pop}}}{\sqrt{n}}$$

# Height data: $n = 20$

$\mu_{\text{pop}} := 175.6$ cm, $\sigma_{\text{pop}} = 6.85$ cm

Total population $N := 4{,}082$

$10^4$ sample means

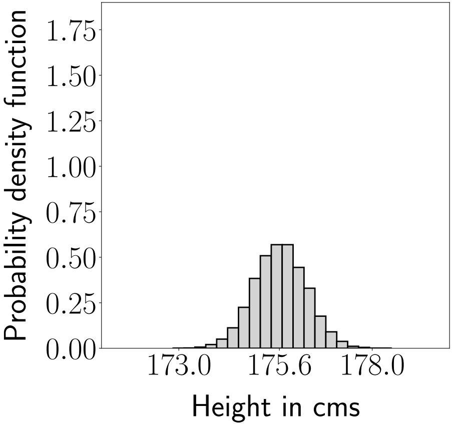


# $n = 100$

$\mu_{\text{pop}} := 175.6$ cm, $\sigma_{\text{pop}} = 6.85$ cm

Total population $N := 4{,}082$

$10^4$ sample means

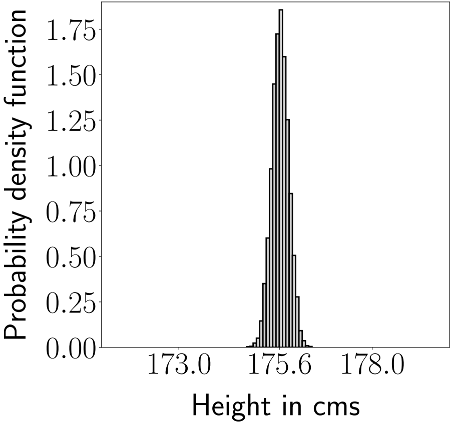


# $n = 1{,}000$

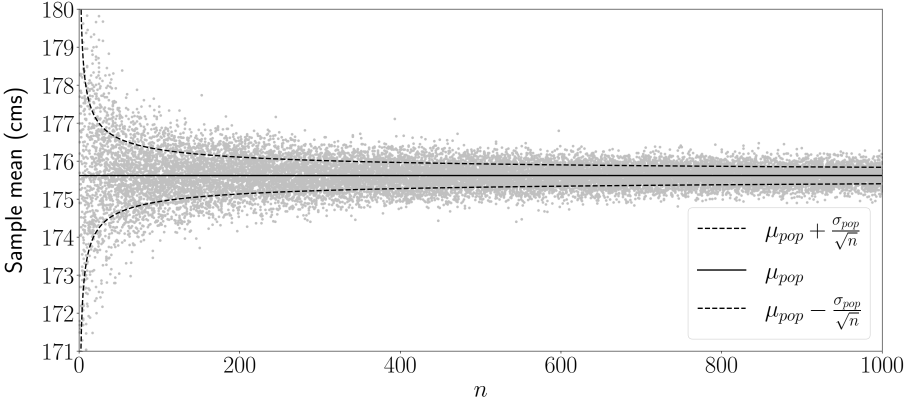
$\mu_{\text{pop}} := 175.6$ cm, $\sigma_{\text{pop}} = 6.85$ cm

Total population $N := 4{,}082$

$10^4$ sample means



# Height data



180
179
178
177
176
175
174
173
172
171
Sample mean (cms)
0
200
400
600
800
1000
$n$
$\mu_{pop} + \frac{\sigma_{pop}}{\sqrt{n}}$
$\mu_{pop}$
$\mu_{pop} - \frac{\sigma_{pop}}{\sqrt{n}}$

# Convergence in mean square

Independent random variables with mean $\mu$ and variance $\sigma^2$
$\tilde{x}_1$, $\tilde{x}_2$, $\tilde{x}_3$, ...

$$\widetilde{m}_n := \frac{1}{n}\sum_{j=1}^{n}\tilde{x}_j$$

$$\mathrm{E}\left[\widetilde{m}_n\right] = \mathrm{E}\left[\frac{1}{n}\sum_{j=1}^{n}\tilde{x}_j\right] = \frac{1}{n}\sum_{j=1}^{n}\mathrm{E}\left[\tilde{x}_j\right] = \mu$$

$$\begin{aligned}\mathrm{MSE}_n = \mathrm{E}\left[\left(\widetilde{m}_n - \mu\right)^2\right] &= \mathrm{Var}\left[\frac{1}{n}\sum_{j=1}^{n}\tilde{x}_j\right] \\ &= \frac{1}{n^2}\sum_{j=1}^{n}\mathrm{Var}\left[\tilde{x}_j\right] = \frac{\sigma^2}{n}\end{aligned}$$

$\mathrm{MSE}_1$, $\mathrm{MSE}_2$, $\mathrm{MSE}_3$, ... converges to 0, $\lim_{n\to\infty}\mathrm{MSE}_n = 0$

# Convergence in probability

Probability of deviating by $\epsilon$

$$p_n := \mathrm{P}\left(|\widetilde{m}_n - \mu| > \epsilon\right)$$

$$p_1,\ p_2,\ p_3,\ p_4,\ \ldots$$

# Deviation inequalities

Goal: Bound probabilities using the mean and variance

# Markov's inequality

A nonnegative random variable with a small mean cannot take large values with high probability

# Markov's inequality

Discrete random variable $\tilde{a}$ with range $A$

For any $c > 0$

Goal: Bound $\mathrm{P}(\tilde{a} \geq c)$ using $\mathrm{E}[\tilde{a}]$

$$
\begin{aligned}
\mathrm{E}[\tilde{a}] &= \sum_{a \in A} a p_{\tilde{a}}(a) \\
&= \sum_{a < c} a p_{\tilde{a}}(a) + \sum_{a \geq c} a p_{\tilde{a}}(a) \\
&\geq \sum_{a < c} a p_{\tilde{a}}(a) + c \sum_{a \geq c} p_{\tilde{a}}(a) \\
&\geq c \sum_{a \geq c} p_{\tilde{a}}(a)
\end{aligned}
$$

$$
\mathrm{P}(\tilde{a} \geq c) = \sum_{a \geq c} p_{\tilde{a}}(a) \leq \frac{\mathrm{E}[\tilde{a}]}{c}
$$

# Markov's inequality

Let $\tilde{a}$ be a nonnegative random variable

For any $c > 0$

$$\mathrm{P}\left(\tilde{a} \geq c\right) \leq \frac{\mathrm{E}\left[\tilde{a}\right]}{c}$$

# Age of students

The mean age of NYU students is 20 years, bound fraction that is above 30

$$\mathrm{P}(\tilde{a} \geq 30) \leq \frac{\mathrm{E}[\tilde{a}]}{30} = \frac{2}{3}$$

At most two thirds

# Chebyshev's inequality

A random variable with small variance cannot be far from its mean $\mu$ with high probability

For any $c > 0$ and any random variable $\tilde{a}$ with bounded variance,

$$\begin{aligned} \mathrm{P}\left( (\tilde{a} - \mu)^2 \geq c \right) &\leq \frac{\mathrm{E}\left[ (\tilde{a} - \mu)^2 \right]}{c} \\ &= \frac{\mathrm{Var}\left[\tilde{a}\right]}{c} \end{aligned}$$

# Zero variance

What happens when a random variable has zero variance?

# Detour: Zero variance

Random variable $\tilde{a}$ with mean $\mu$ and $\mathrm{Var}[\tilde{a}] = 0$

Take any $\epsilon > 0$

$$\mathrm{P}(|\tilde{a} - \mu| \geq \epsilon) \leq \frac{\mathrm{Var}[\tilde{a}]}{\epsilon^2} = 0$$

If $\mathrm{Var}[\tilde{a}] = 0$ then $\mathrm{P}(\tilde{a} = \mu) = 1$

If $\mathrm{E}\left[\tilde{a}^2\right] = 0$ then $\mathrm{P}(\tilde{a} = 0) = 1$

# Age of students

Mean is 20 years, standard deviation is 3

Bound fraction above 30

$$\begin{aligned} \mathrm{P}(\tilde{a} \geq 30) &\leq \mathrm{P}\left(|\tilde{a} - \mathrm{E}\left[\tilde{a}\right]| \geq 10\right) \\ &\leq \frac{\mathrm{Var}\left[\tilde{a}\right]}{100} \\ &= \frac{9}{100} \end{aligned}$$

Much better bound than Markov's inequality (9% vs 2/3)

# Law of large numbers

If $\tilde{x}_1, \tilde{x}_2, \ldots$ are independent random variables with mean $\mu$ and variance $\sigma^2$

$$\tilde{m}_n := \frac{1}{n} \sum_{i=1}^{n} \tilde{x}_i$$

$$\begin{aligned}
\mathrm{P}\left(|\tilde{m}_n - \mu| > \epsilon\right) &\leq \frac{\mathrm{Var}\left[\tilde{m}_n\right]}{\epsilon^2} \\
&= \frac{1}{\epsilon^2} \mathrm{Var}\left[\frac{1}{n} \sum_{j=1}^{n} \tilde{x}_j\right] \\
&= \frac{1}{n^2 \epsilon^2} \sum_{j=1}^{n} \mathrm{Var}\left[\tilde{x}_j\right] = \frac{\sigma^2}{n \epsilon^2}
\end{aligned}$$

Converges to zero for any $\epsilon$!

# Consistency

Random measurements: $\tilde{x}_1$, $\tilde{x}_2$, ..., $\tilde{x}_n$

Deterministic parameter of interest: $\gamma$

An estimator $h(\tilde{x}_1, \ldots, \tilde{x}_n)$ is consistent if for any $\epsilon > 0$

$$\lim_{n \to \infty} \mathrm{P}\left(|h(\tilde{x}_1, \ldots, \tilde{x}_n) - \gamma| > \epsilon\right) = 0$$

# The sample mean is consistent

Data: $a_1, a_2, \ldots, a_N$

Population mean: $\mu_{\text{pop}}$

Population variance: $\sigma^2_{\text{pop}}$

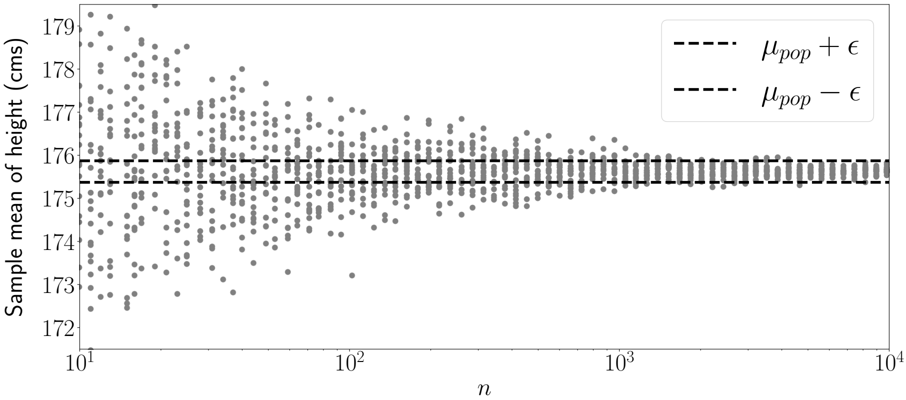
Random independent samples: $\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n$

$$\mathrm{E}\left[\tilde{x}_j\right] = \mu_{\text{pop}}$$

$$\mathrm{Var}\left[\tilde{x}_j\right] = \sigma^2_{\text{pop}}$$

Sample mean converges in probability to $\mu_{\text{pop}}$

# Height data



Sample mean of height (cms)
179
178
177
176
175
174
173
172
$10^1$
$10^2$
$10^3$
$10^4$
$n$
$\mu_{pop} + \epsilon$
$\mu_{pop} - \epsilon$

# What have we learned

To control probabilities using deviation inequalities

Law of large numbers

The sample mean and sample proportion are consistent

# Key inequality

$$\mathrm{P}\left(|\tilde{m}_n - \mu_{\mathsf{pop}}| > \epsilon\right) \leq \frac{\sigma_{\mathsf{pop}}^2}{n\epsilon^2}$$

How tight is this bound?

# Pretty loose...