Will Monroe
August 4, 2017

with materials by
Mehran Sahami
and Chris Piech

image: Romek

Probability bounds

Announcement: Problem Set #5

Due this coming **Monday, August 7** (before class).

11 problems:

Robot package delivery

Cell reception in the wilderness

Last chance to use late days (no late submissions accepted for PS6)

Announcements: Final exam

Two weeks from tomorrow:

Saturday, August 19, 12:15-3:15pm

Two pages (both sides) of notes

Comprehensive—material that was on the midterm will also be tested

Review session:
Wednesday, August 16, 2:30-3:20pm
in **Gates B03**

Review: Variance of a sum

The **variance of a sum** of random variables is equal to the **sum of pairwise covariances** (*including* variances and double-counted pairs).

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \mathrm{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} X_j\right)$$

$$= \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\sum_{i=1}^{n} \sum_{j=i+1}^{n} \mathrm{Cov}(X_i, X_j)$$

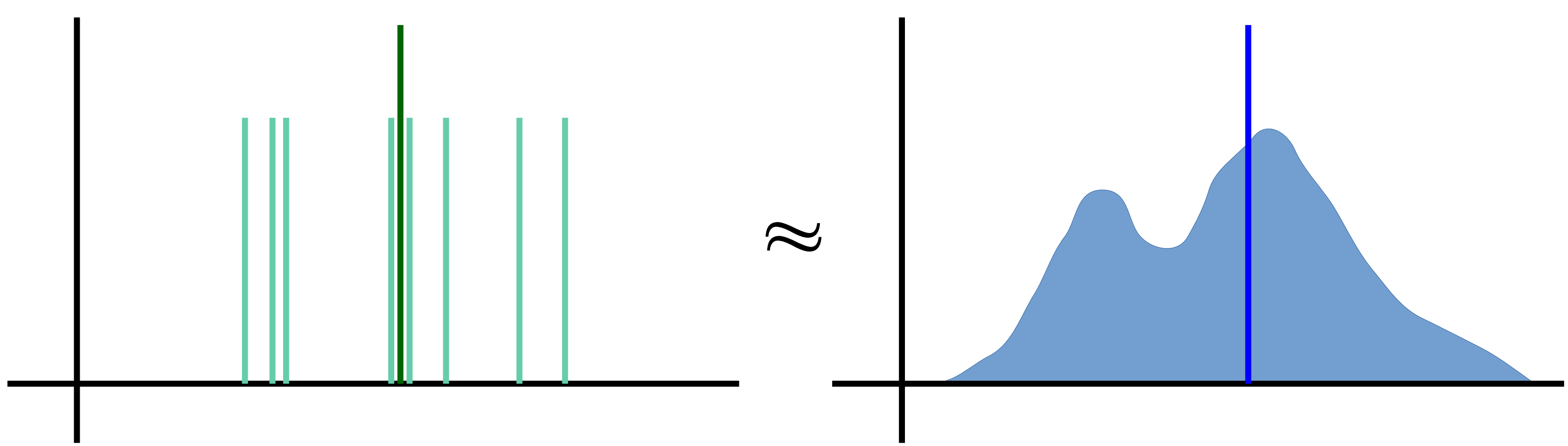
note: independent ⇒ Cov = 0

Review: Parameter estimation

Sometimes we **don't know** things like the expectation and variance of a distribution; we have to **estimate** them from incomplete information.

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \qquad S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

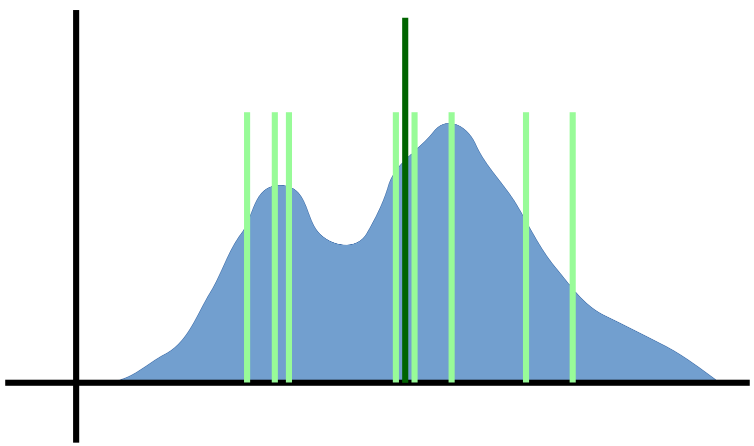
$$\hat{\theta} = \arg\max_{\theta} \mathrm{LL}(\theta)$$



Review: Sample mean

A **sample mean** is an **average** of random variables drawn (usually independently) from the **same distribution**.

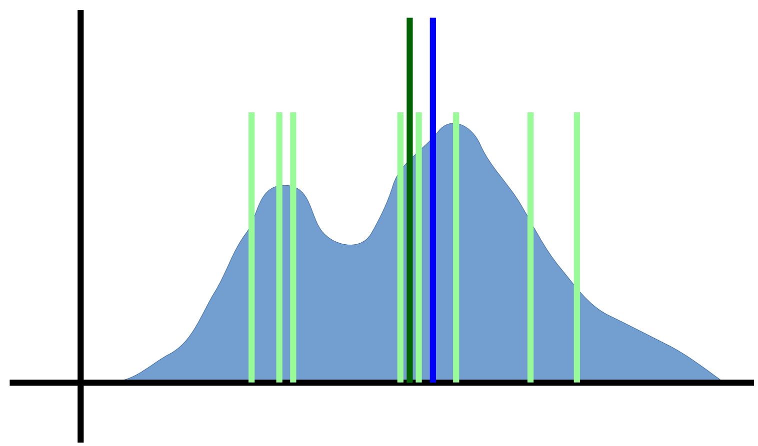
$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$



Review: Unbiased estimator

An **unbiased estimator** is a random variable that has **expectation** equal to the quantity you are estimating.

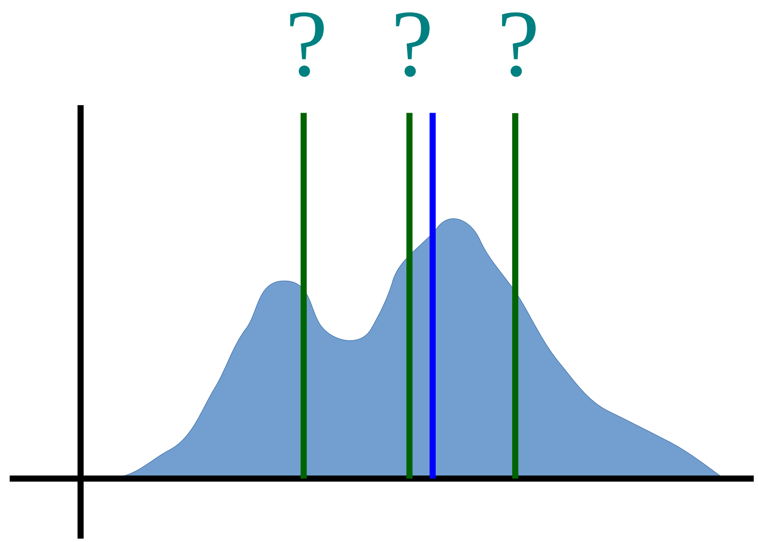
$$E[\bar{X}] = \mu = E[X_i]$$



Review: Variance of the sample mean

The **sample mean** is a random variable; it can differ among samples. That means it has a **variance**.

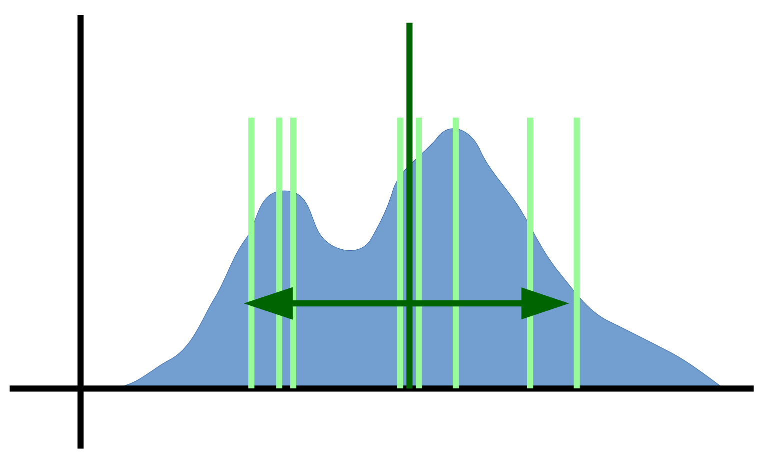
$$\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}$$



Review: Sample variance

Samples can be used to **estimate the variance** of the original distribution.

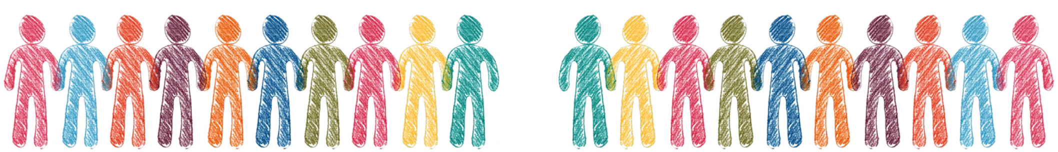
$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X} \right)^2$$



Review: p-values

A ***p*-value** gives the probability of an extreme result, assuming that any extremeness is due to chance.

$$p = P\left(|\bar{X} - \mu| > d \mid H_0\right)$$



Review: Bootstrapping

Bootstrapping allows you to compute complicated statistics from samples using simulation.

```
def bootstrap(sample):
    pmf = sample
    means = []
    for i in range(10000):
        sample = np.random.choice(pmf, len(sample))
        mean = np.mean(sample)
        results.append(mean)
    return means
```

Fun with inequalities

≥ ‿ ≤

>.<

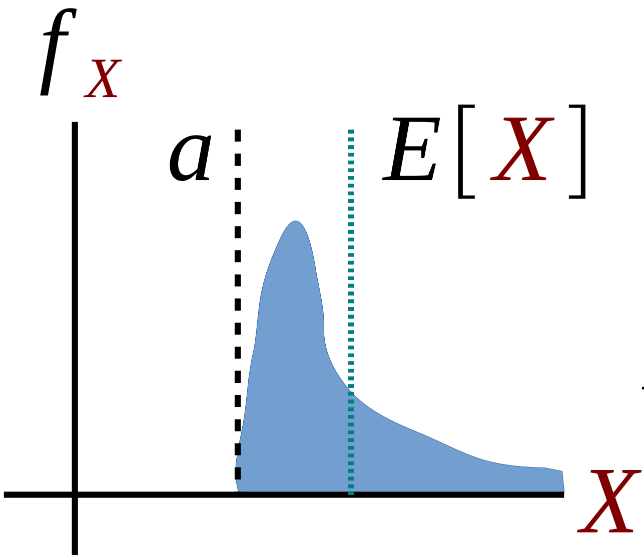
<(^_^<)

Don’t require much knowledge about the full distribution

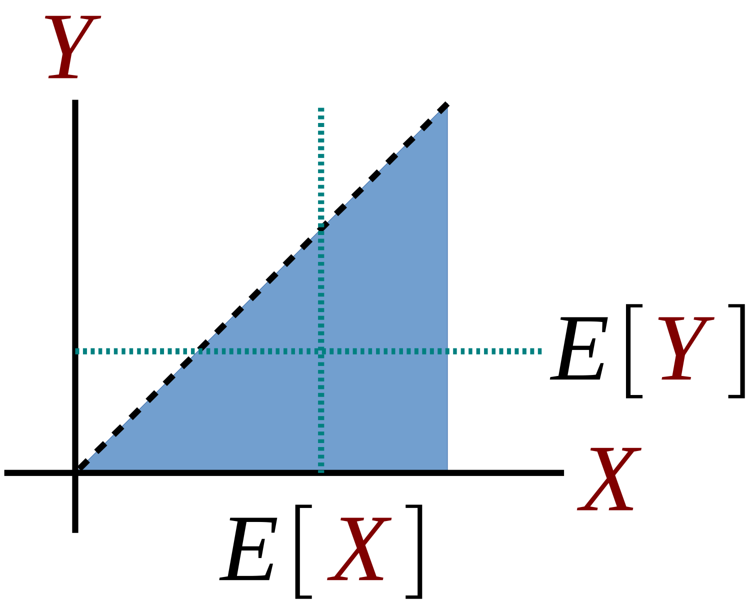
Super useful in proofs

Also super useful for building cute emoticons

Bounding expectation



$$P(a \leq X \leq \infty) = 1 \;\Rightarrow\; a \leq E[X] \leq \infty$$

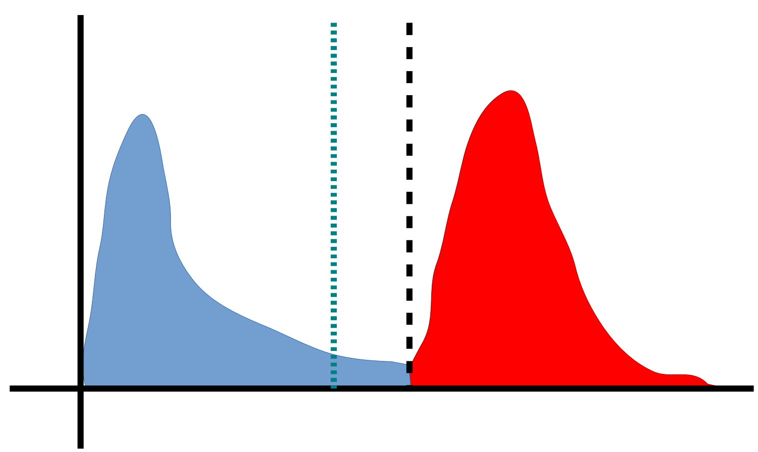


$$P(X \geq Y) = 1 \;\Rightarrow\; E[X] \geq E[Y]$$

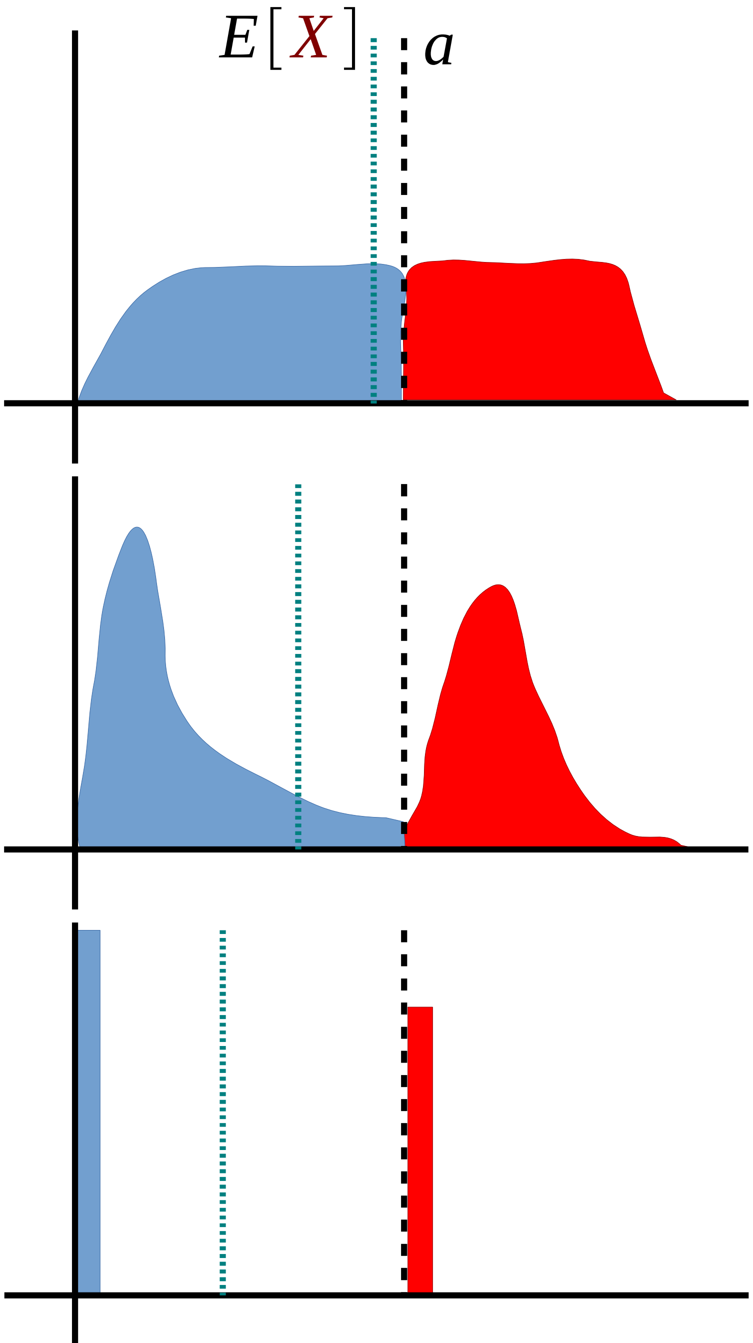
Markov's inequality

Knowing the **expectation** of a **non-negative** random variable lets you bound the probability of **high** values for that variable.

$$X \geq 0 \;\; \Rightarrow \;\; P(X \geq a) \leq \frac{E[X]}{a}$$



Markov's inequality: Intuition



$$P(X \geq a) \leq \frac{E[X]}{a}$$

$$\Updownarrow$$

$$E[X] \geq a \cdot P(X > a)$$

Markov's inequality: Proof

indicator variable $I = \begin{cases} 1 & \text{if } X \geq a \\ 0 & \text{otherwise} \end{cases}$

two cases:

$$X \geq a \Rightarrow \frac{X}{a} \geq 1 = I$$

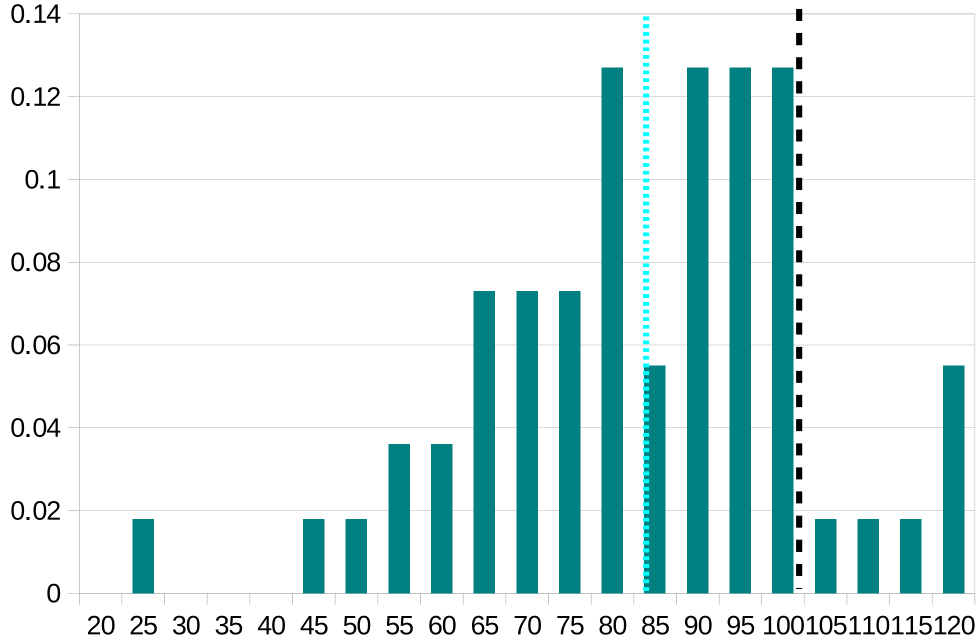
$$X < a \Rightarrow \frac{X}{a} \geq 0 = I$$

$$X \geq 0 \Rightarrow I \leq \frac{X}{a}$$

$$E[I] \leq E\left[\frac{X}{a}\right]$$

$$P(X \geq a) \leq \frac{E[X]}{a}$$

Markov & Midterm



$$E[X]=82.2$$

$$P(X\geq 100)\leq 0.822$$

actual value:

$$P(X\geq 100)=0.109$$

Andrey Andreyevich Markov

Андре́й Андре́евич Ма́рков

Russian mathematician (1856–1922)

Many CS+probability concepts (sharing a common theme) named after him:

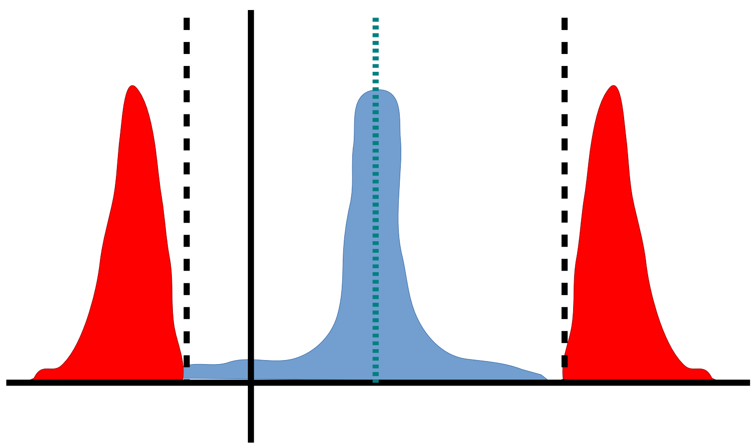
- Hidden Markov model
- Markov decision process
- Markov blanket
- Markov chain

part of the theoretical basis for Google's PageRank algorithm

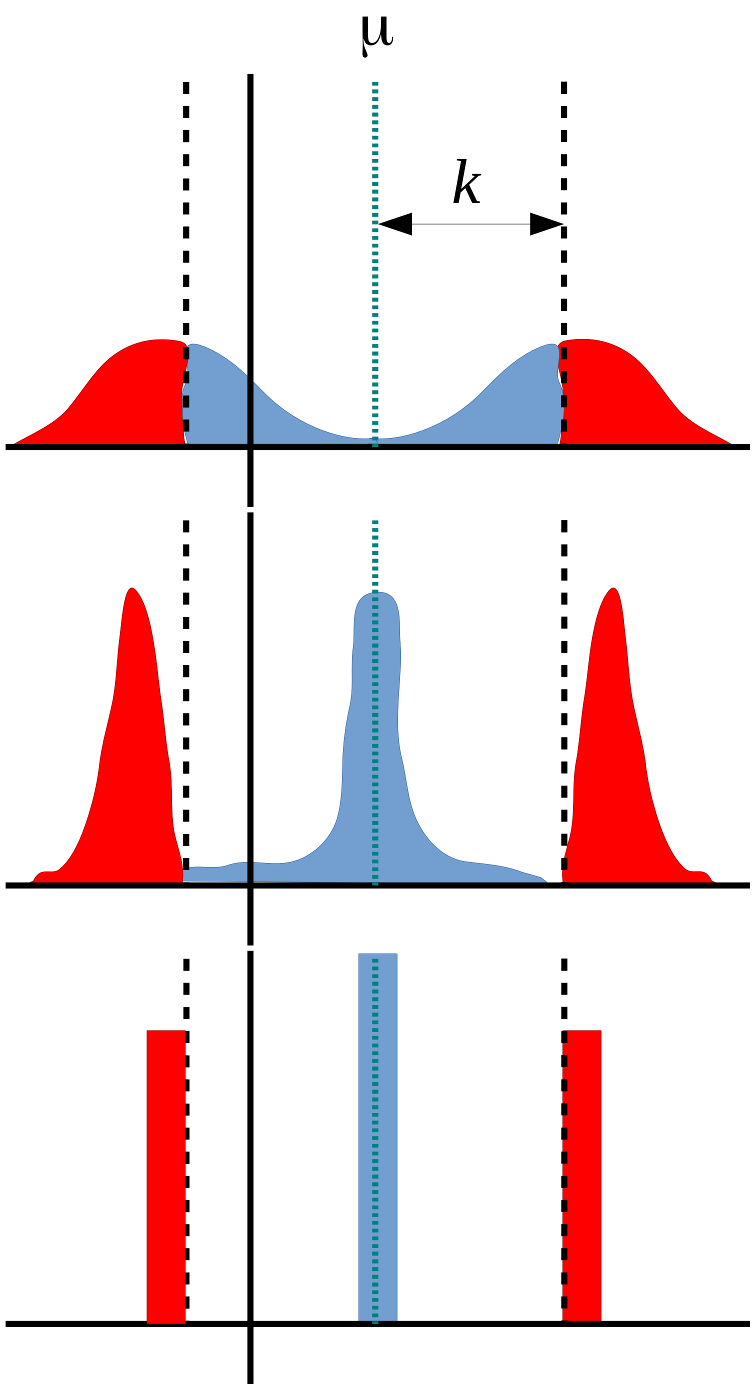
Chebyshev's inequality

Knowing the expectation **and variance** of a random variable lets you bound the probability of **extreme** values for that variable.

$$P\left(|X-\mu|\geq k\right)\leq\frac{\sigma^2}{k^2}$$



Chebyshev's inequality: Intuition



$$P\left(|X-\mu|\geq k\right)\leq\frac{\sigma^2}{k^2}$$

$$\Updownarrow$$

$$\sigma^2\geq k^2\cdot P\left(|X-\mu|\geq k\right)$$

Chebyshev's inequality: Proof

$$Y = (X - \mu)^2 > 0$$

$$\mu = E[X]$$

$$\sigma^2 = \mathrm{Var}(X)$$

$$= E[(X - \mu)^2]$$

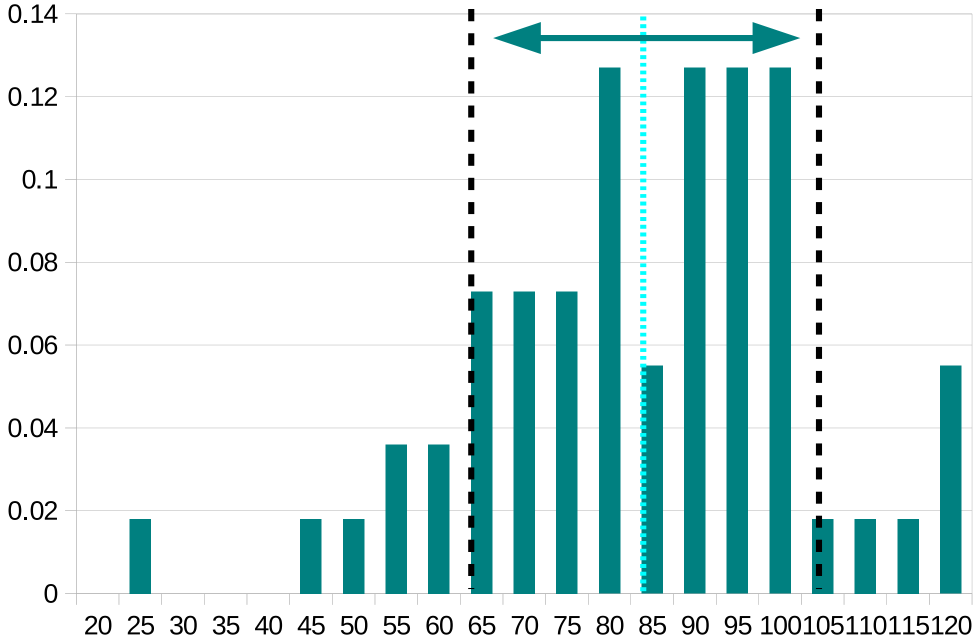
Markov's inequality:

$$P(Y \geq a) \leq \frac{E[Y]}{a}$$

$$P((X - \mu)^2 \geq k^2) \leq \frac{E[(X - \mu)^2]}{k^2}$$

$$P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}$$

Chebyshev takes the midterm



$$\mu = E[X] = 82.2$$
$$\sigma^2 = \mathrm{Var}(X) = 342.5$$
$$\sigma = \mathrm{SD}(X) = 18.5$$

$$P(62.2 \le X \le 102.2) \ge 1 - \frac{342.5}{20^2} \approx 0.144$$

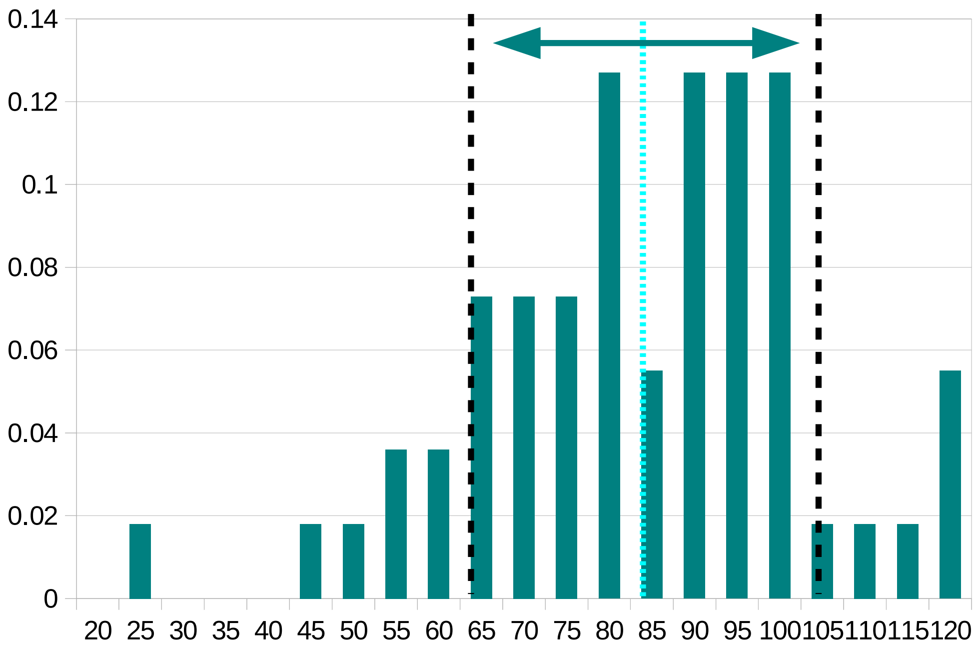
actual value: $P(62.2 \le X \le 102.2) \approx 0.764$

One-sided Chebyshev's inequality

$$P\left(X \geq \mu + a\right) \leq \frac{\sigma^2}{\sigma^2 + a^2}$$

$$P\left(X \leq \mu - a\right) \leq \frac{\sigma^2}{\sigma^2 + a^2}$$

Chebyshev takes the midterm



$$\mu = E[X] = 82.2$$
$$\sigma^2 = \mathrm{Var}(X) = 342.5$$
$$\sigma = \mathrm{SD}(X) = 18.5$$

$$P(X \geq 100 = \mu + 17.8) \leq \frac{342.5}{342.5 + 17.8^2} \approx 0.519$$

Markov's inequality: $P(X \geq 100) \leq 0.822$

actual value: $P(X \geq 100) = 0.109$

Pafnuty Lvovich Chebyshev

Пафну́тий Льво́вич Чебышёв

Russian mathematician (1821–1894)

Chebyshev's inequality is named after him (but actually formulated by colleague Irénée-Jules Bienaymé)

Markov's doctoral advisor (and sometimes credited with first deriving Markov's Inequality)

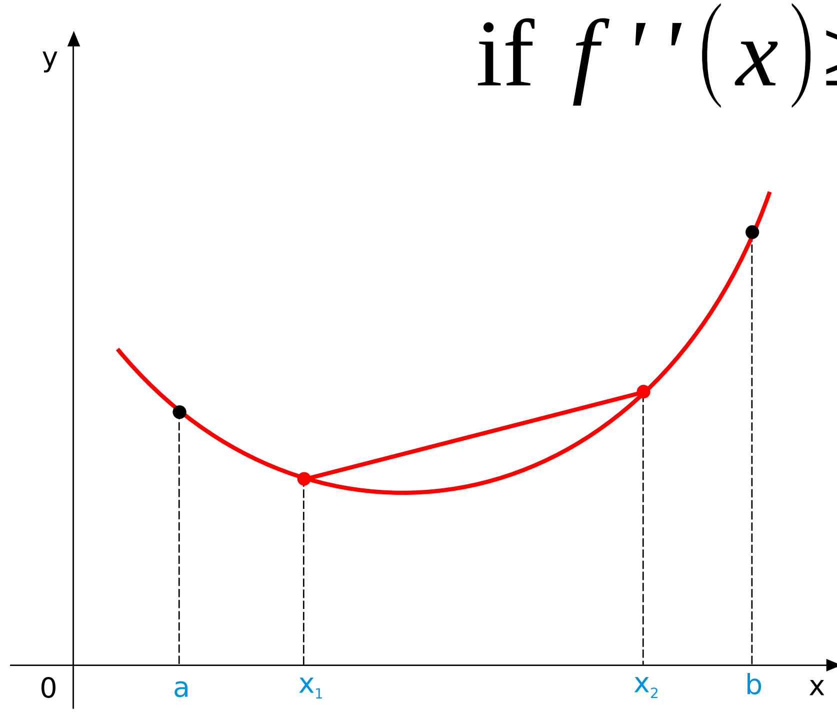
Break time!

Jensen's inequality

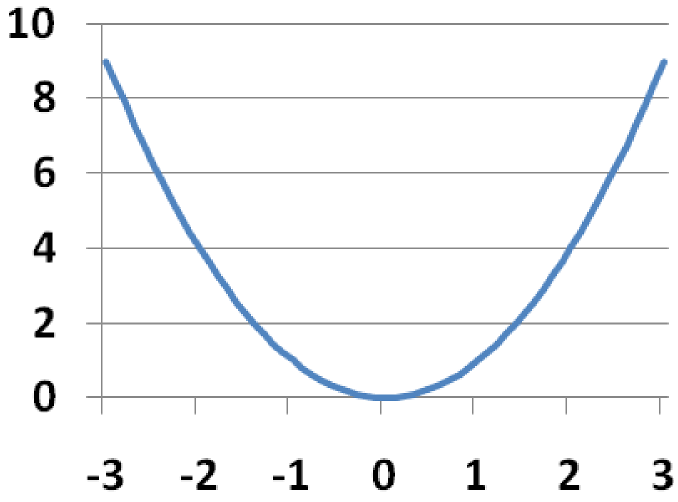
The expectation of a **convex function** of a random variable can't be less than the value of the function applied to the expectation.

$$E[f(X)] \geq f(E[X])$$

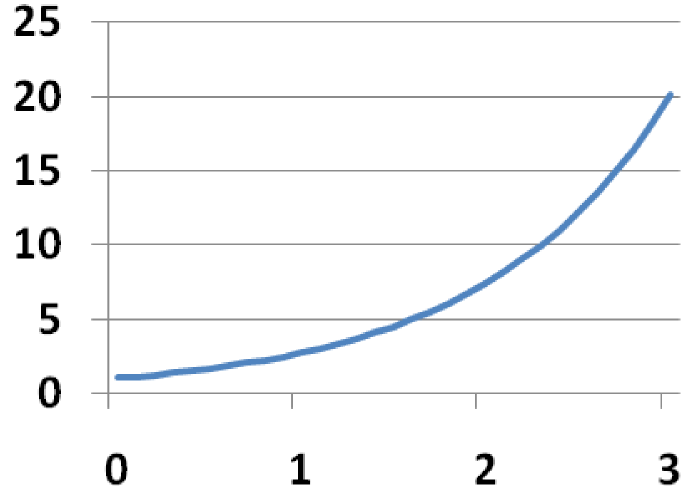
$$\text{if } f''(x) \geq 0$$



Some convex functions

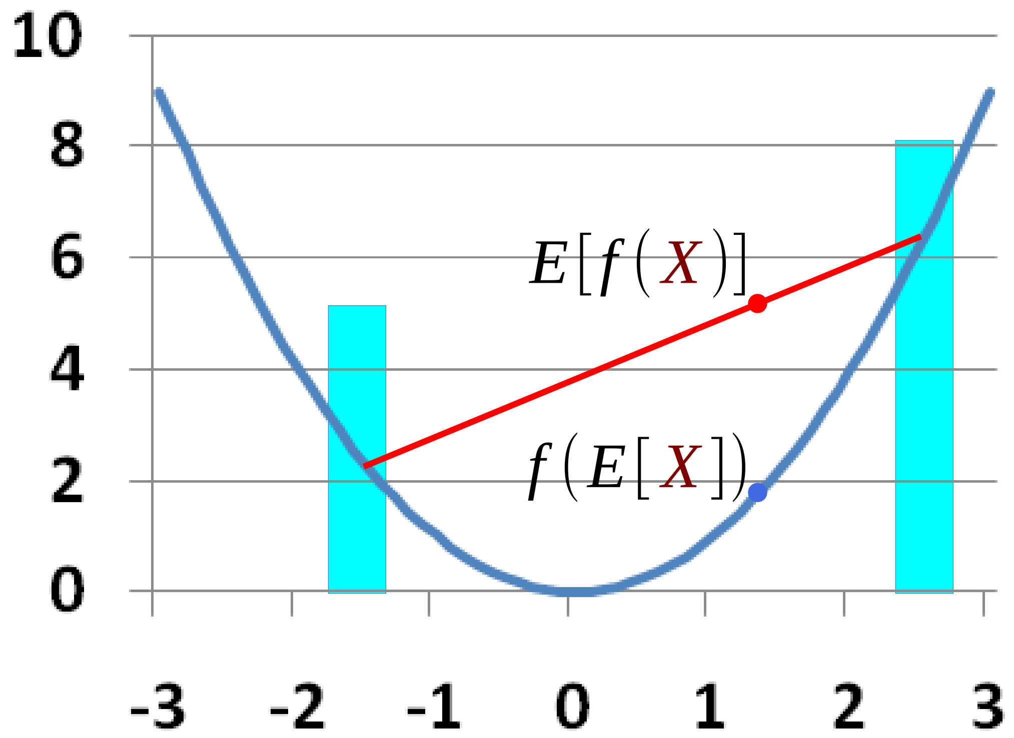


$$f(x) = x^2$$



$$f(x) = e^x$$

Jensen's inequality: Intuition



Johan Ludvig William Valdemar Jensen

Danish mathematician (1859–1925)

Was an engineer in the Copenhagen Telephone Company—did his math in his spare time.

Johan Ludvig William Valdemar Jensen

Danish mathematician (1859–1925)

Was an engineer in the Copenhagen Telephone Company—did his math in his spare time.

image (bottom): Mike Mozart

Law of large numbers

A **sample mean** will **converge to the true mean** if you take a large enough sample.

$$\lim_{n \to \infty} P\left(\left|\bar{X} - \mu\right| \geq \varepsilon\right) = 0$$

$$P\left(\lim_{n \to \infty} \left(\bar{X}\right) = \mu\right) = 1$$

Weak law of large numbers

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\downarrow$$

for any ε > 0: $$\lim_{n \to \infty} P\left(|\bar{X} - \mu| \geq \varepsilon\right) = 0$$

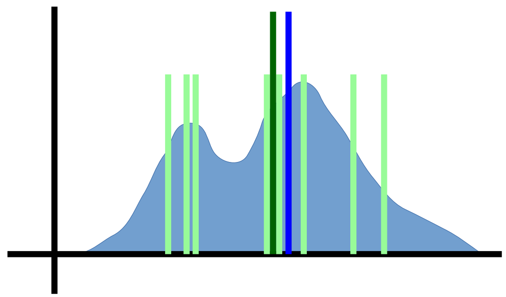
Weak law of large numbers: Proof

$$E[X_i] = \mu$$

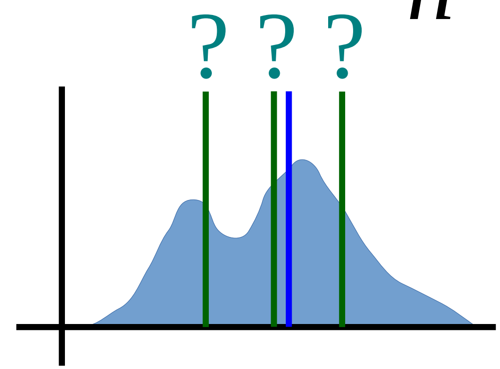
$$\mathrm{Var}(X_i) = \sigma^2$$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$E[\bar{X}] = \mu$$



$$\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}$$



Chebyshev's inequality:

$$P(|\bar{X} - \mu| \geq \varepsilon) \leq \frac{\left(\frac{\sigma^2}{n}\right)}{\varepsilon^2}$$

$$\lim_{n \to \infty} \frac{\left(\frac{\sigma^2}{n}\right)}{\varepsilon^2} = 0$$

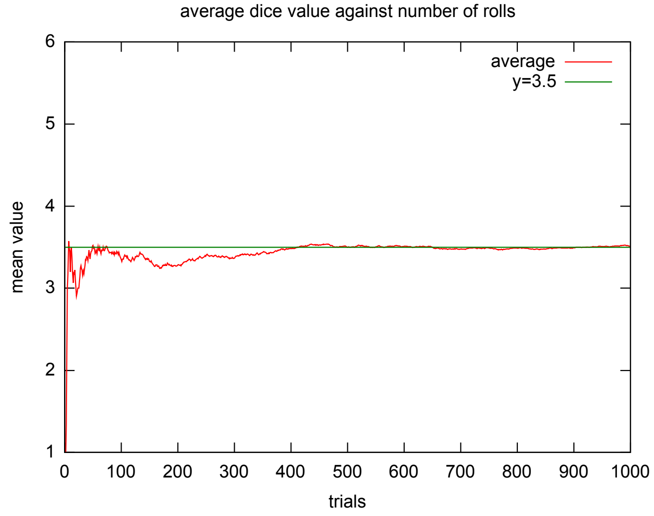
so

$$\lim_{n \to \infty} P(|\bar{X} - \mu| \geq \varepsilon) = 0$$

Consistent estimator

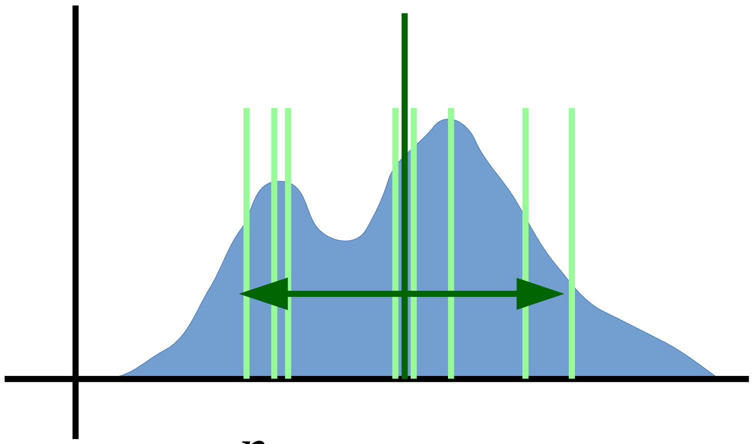
An **consistent estimator** is a random variable that has a **limit** (as number of samples gets large) equal to the quantity you are estimating.

$$\lim_{n \to \infty} P\left(|\hat{\theta} - \theta| < \varepsilon\right) = 1$$



Estimating variance

$$\mathrm{Var}(X) = E[(X-\mu)^2]$$



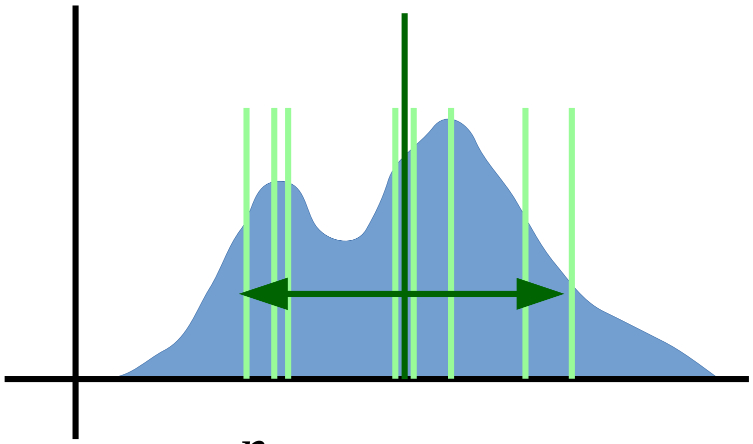
$$Y = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$ is a(n):

A) Unbiased and consistent estimator
B) Biased but consistent estimator
C) Unbiased but not consistent estimator
D) Biased and not consistent estimator

https://bit.ly/1a2ki4G → https://b.socrative.com/login/student/
Room: CS109SUMMER17

Estimating variance

$$\text{Var}(X) = E[(X-\mu)^2]$$



$$Y = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$ is a(n):

B) Biased but consistent estimator

$$E[Y] = \left(\frac{n-1}{n}\right)\sigma^2 \underset{n \to \infty}{\to} \sigma^2$$

https://bit.ly/1a2ki4G → https://b.socrative.com/login/student/
Room: CS109SUMMER17

Strong law of large numbers

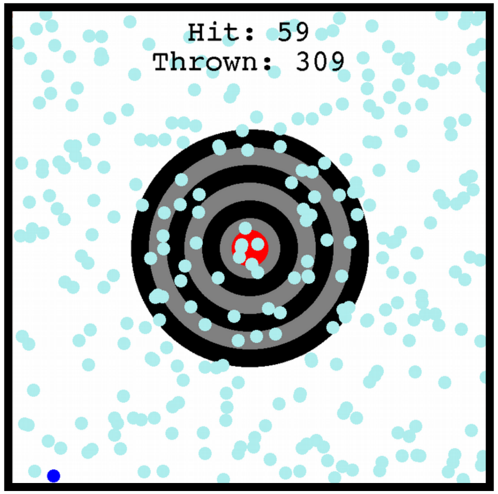
$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\downarrow$$

$$P\left(\lim_{n\to\infty}\left(\bar{X}\right)=\mu\right)=1$$

i.e.: $$P\left(\lim_{n\to\infty}\left(\frac{X_1+X_2+\cdots+X_n}{n}\right)=\mu\right)=1$$

Frequentist probability and LLN

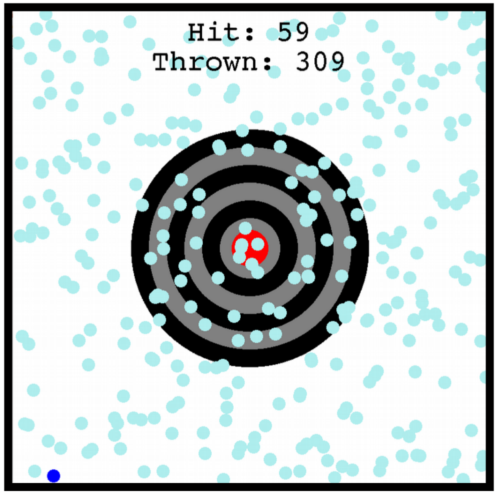


$$P(E) = \lim_{n \to \infty} \frac{\#(E)}{n}$$

indicator variables $X_i = \begin{cases} 1 & \text{if } E \text{ occurs on trial } i \\ 0 & \text{otherwise} \end{cases}$

$$\lim_{n \to \infty} \left(\frac{\overbrace{X_1 + X_2 + \cdots + X_n}^{=\#(E)}}{n} \right) = E[X_i]$$

Frequentist probability and LLN

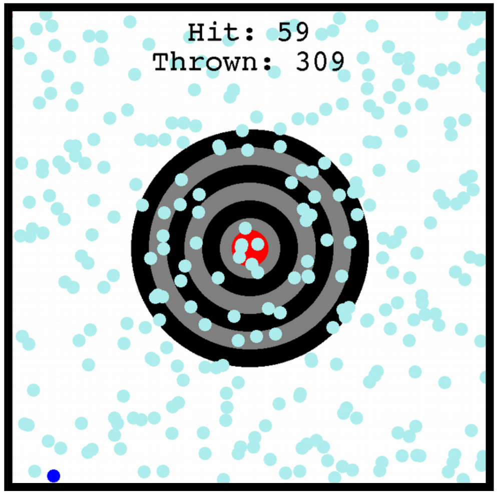


$$P(E) = \lim_{n \to \infty} \frac{\#(E)}{n}$$

indicator variables $X_i = \begin{cases} 1 & \text{if } E \text{ occurs on trial } i \\ 0 & \text{otherwise} \end{cases}$

$$\lim_{n \to \infty} \left(\frac{\overbrace{X_1 + X_2 + \cdots + X_n}^{= \#(E)}}{n} \right) = P(E)$$

Frequentist probability and LLN

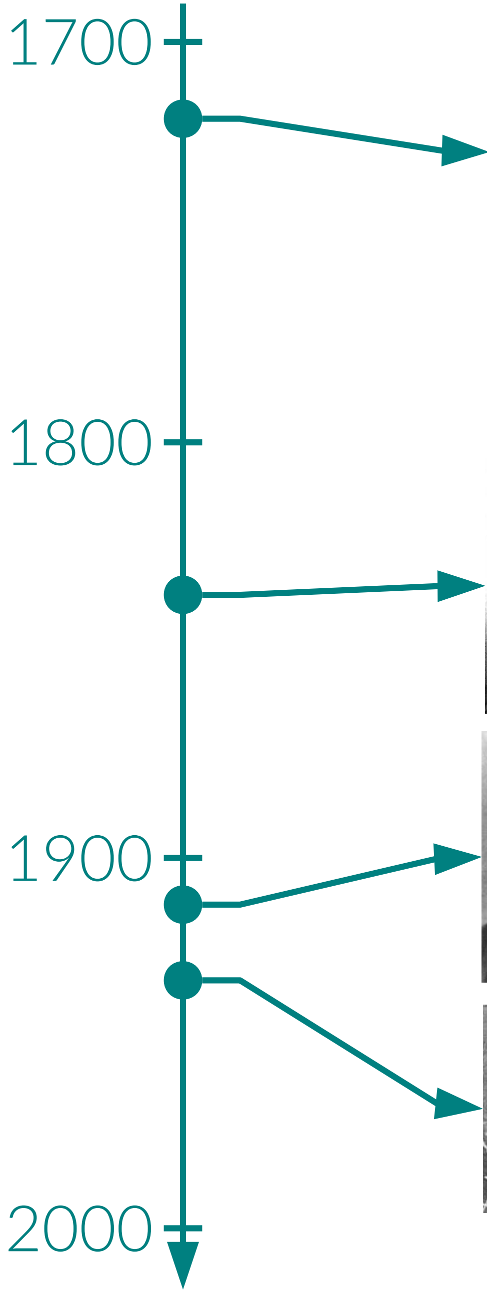


$$P(E) = \lim_{n \to \infty} \frac{\#(E)}{n}$$

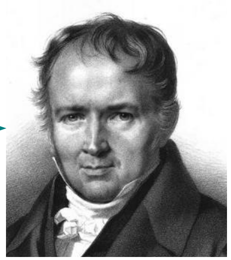
indicator variables $X_i = \begin{cases} 1 & \text{if } E \text{ occurs on trial } i \\ 0 & \text{otherwise} \end{cases}$

$$P\left(\lim_{n \to \infty} \left(\frac{\overbrace{X_1 + X_2 + \cdots + X_n}^{= \#(E)}}{n} \right) = P(E) \right) = 1$$

Law of large numbers: A history



1713: Jacob Bernoulli describes
weak law of large numbers



1835: Siméon Denis Poisson coins name
(*La Loi des Grands Nombres*)

1909: Émile Borel develops strong law for
Bernoulli random variables

1928: Andrei Nikolaevich Kolmogorov shows
strong law only needs expectation to exist

Gambler’s fallacy

“I’m due for a win!”