
Will Monroe
August 2, 2017



with materials by
Mehran Sahami
and Chris Piech


image: Pexels

# Samples and bootstrapping

# Announcement: Problem Set #5

Due **Monday, August 7** before class.

11 problems:

Robot package delivery

Cell reception
in the wilderness

# Review: Conditional expectation

One can compute the **expectation** of a random variable while **conditioning** on the values of other random variables.

$$E[X|Y=y]=\sum_x x\, p_{X|Y}(x|y)$$

$$E[X|Y=y]=\int_{-\infty}^{\infty} dx\, x\, f_{X|Y}(x|y)$$

# Review: Quicksort

Let X = $\boxed{\text{number of comparisons}}$ to the pivot.
What is E[X]? expected number of events = indicator variables!

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|

$Y_1$ $Y_2$ … $Y_n$

Define $Y_1 \ldots Y_n$ = elements in sorted order.

Indicator variables $I_{ab}$ = 1 if $Y_a$ and $Y_b$ are ever compared.

$$E[X] = E\left[\underbrace{\boxed{\sum_{a=1}^{n-1}\sum_{b=a+1}^{n}}}_{\text{unique pairs}} I_{ab}\right] = \sum_{a=1}^{n-1}\sum_{b=a+1}^{n} E[I_{ab}]$$

$$= \sum_{a=1}^{n-1}\sum_{b=a+1}^{n} P(Y_a \text{ and } Y_b \text{ ever compared})$$

# Review: Quicksort

P( $Y_a$ and $Y_b$ ever compared) = ?

Case 1:

| | | $Y_a$ | | | | $Y_b$ | |
|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |

yes!

Case 2:

| | | $Y_a$ | | | | $Y_b$ | |
|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |

no!

Case 3:

| | | $Y_a$ | | | | $Y_b$ | |
|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |

recursive call

$$\therefore \quad P\left(Y_a \text{ and } Y_b \text{ ever compared}\right) = \frac{2}{b-a+1}$$

# Review: Quicksort

$$E[X]=\sum_{a=1}^{n-1}\sum_{b=a+1}^{n} P(Y_a \text{ and } Y_b \text{ ever compared})$$

$$=\sum_{a=1}^{n-1}\sum_{b=a+1}^{n}\frac{2}{b-a+1}$$

$$\approx\sum_{a=1}^{n-1} 2\ln(n-a+1)$$

$$\approx\int_{a=1}^{n-1} da\, 2\ln(n-a+1)$$

$$\sum_{b=a+1}^{n}\frac{2}{b-a+1}\approx\int_{b=a+1}^{n} db\,\frac{2}{b-a+1}$$

$$=\left[2\ln(b-a+1)\right]_{b=a+1}^{n}$$

$$=2\ln(n-a+1)-2\ln 2$$

$$\approx 2\ln(n-a+1)$$

for large $n$

$$=-2\int_{y=n}^{2} dy\,\ln y$$

$$=-2\left[y\ln y-y\right]_{y=n}^{2}$$

$$=-2\left[\cancel{(2\ln 2-2)}-(n\ln n-\cancel{n})\right]$$

constants

lower-order term

$$=O(n\ln n)$$

$$u=\ln y \qquad v=y$$

$$du=\frac{1}{y}\,dy \qquad dv=dy$$

$$\int u\,dv=u\,v-\int v\,du$$

$$\int \ln y\,dy=y\ln y-\int \cancel{y}\,\frac{1}{\cancel{y}}\,dy$$

$$=y\ln y-y+C$$

# Review: Variance of a linear function

Adding a <u>constant</u>? Variance **doesn't change**.
Multiplying by a <u>constant</u>? **Multiply** the variance by the **square** of the constant.

$$
\begin{aligned}
\mathrm{Var}(aX+b) &= E[(aX+b)^2]-(E[aX+b])^2 \\
&= E[a^2X^2+2abX+b^2]-(aE[X]+b)^2 \\
&= a^2E[X^2]+2ab\,E[X]+b^2 \\
&\quad -[a^2(E[X])^2+2abE[X]+b^2] \\
&= a^2E[X^2]-a^2(E[X])^2 \\
&= a^2[E[X^2]-(E[X])^2] \\
&= a^2\,\mathrm{Var}(X)
\end{aligned}
$$

# 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).

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

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

# Proof: Variance of a sum

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

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

$\mathrm{Cov}(X, X) = \mathrm{Var}(X)$

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

$\mathrm{Cov}(X_i, X_j) = \mathrm{Cov}(X_j, X_i)$

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

# 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).

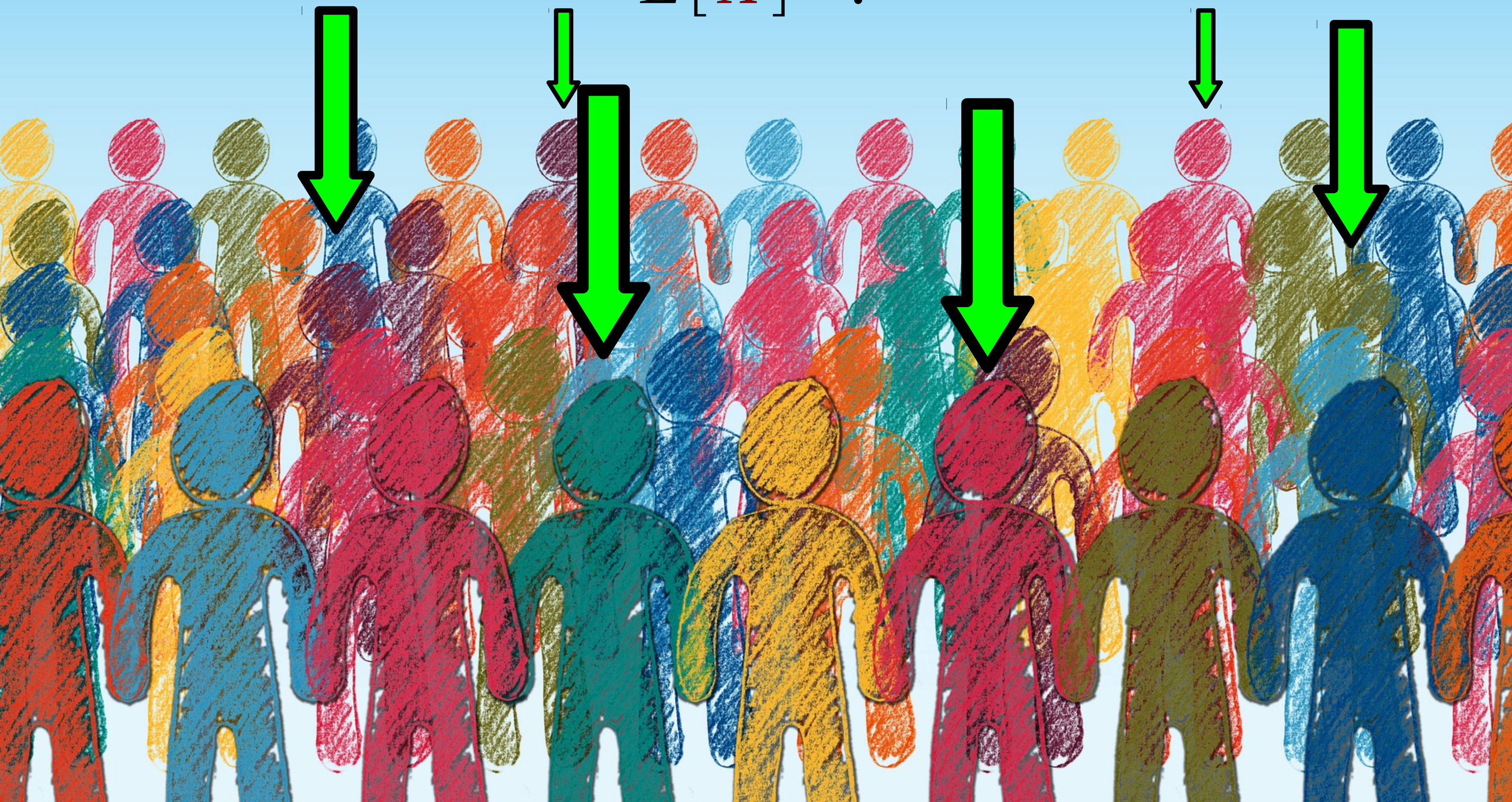
$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \mathrm{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} X_j\right)$$

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

note: independent ⇒ Cov = 0

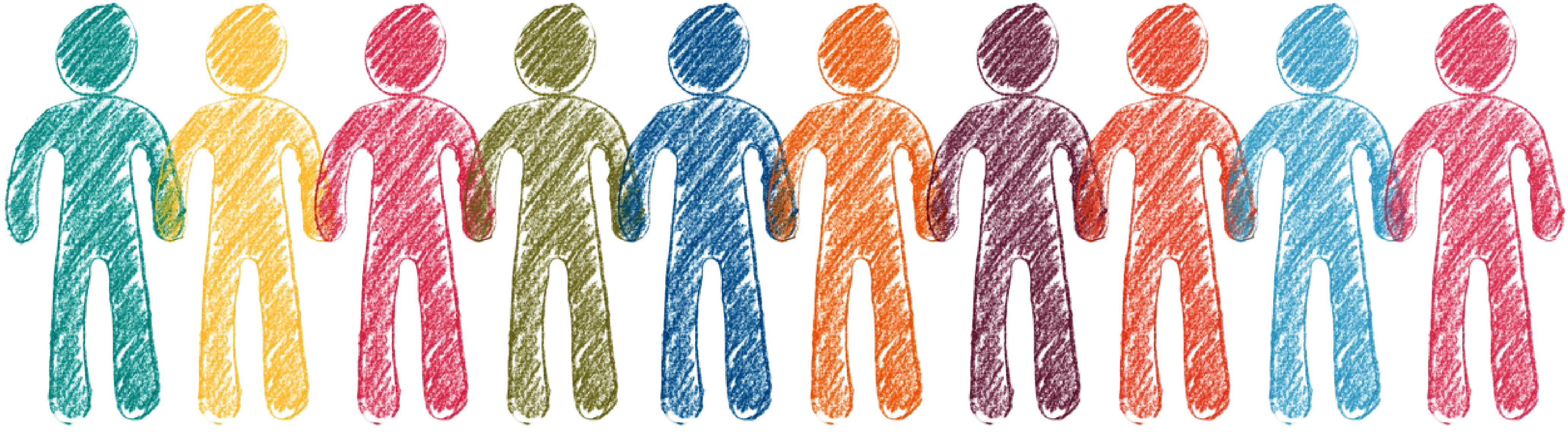
# Sampling from a large population

$$E[X]=?$$



# Sampling from a large population

$$E[X] \approx \frac{1}{n} \sum ($$

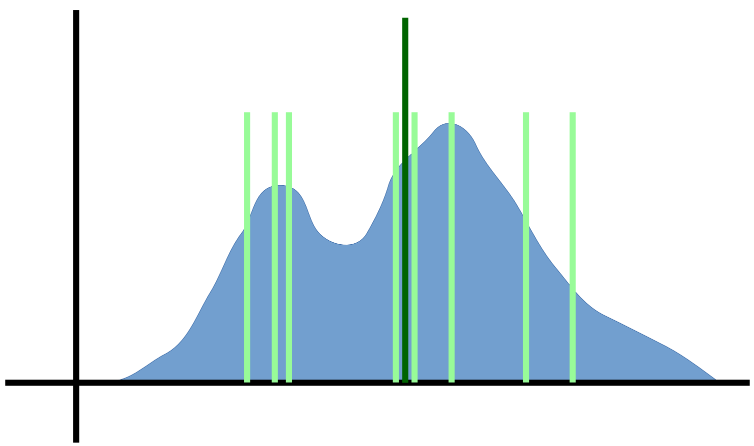


$$)$$

# Sample mean

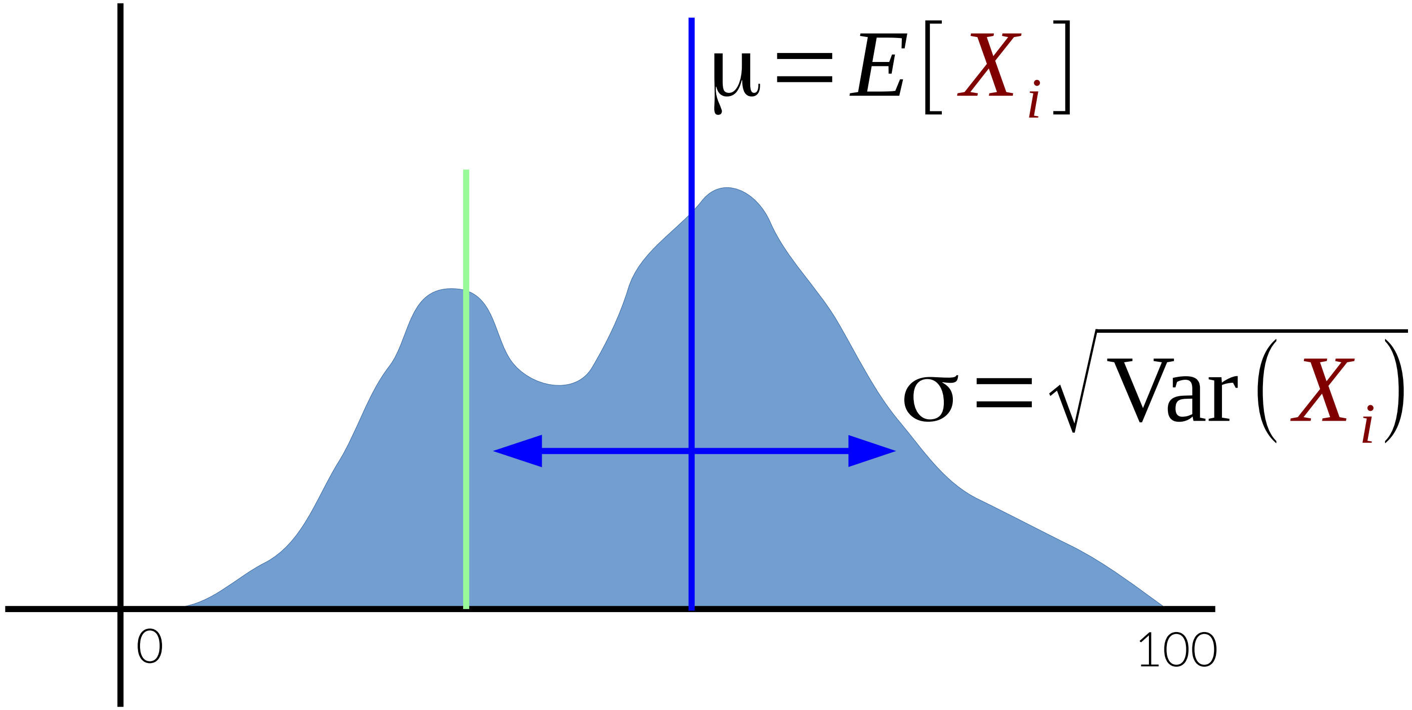
A **sample mean** is an **average** of random variables drawn (usually independently) from the **same distribution**.

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



# Samples = random variables

$X_1 = 37$



# Samples = random variables

$X_1 = 37$

$X_2 = 53$

$X_3 = 34$

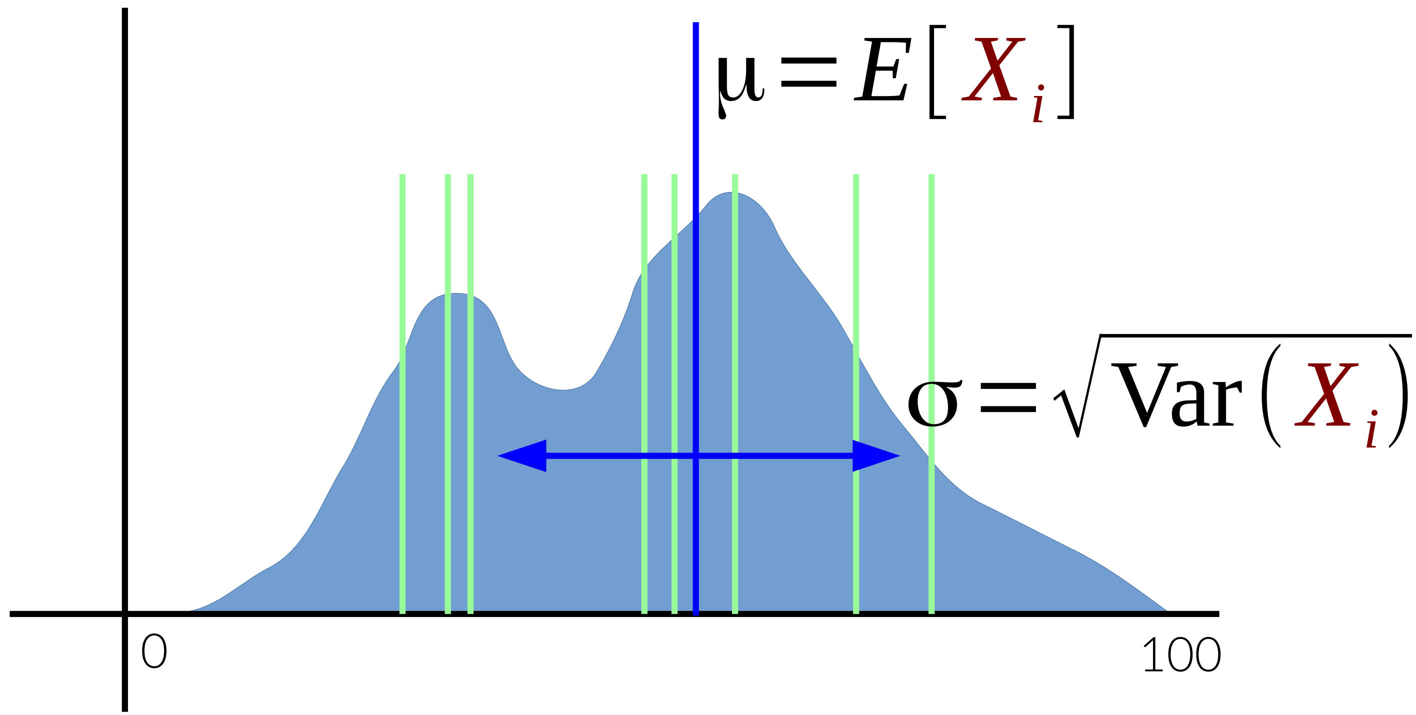
$X_4 = 70$

$X_5 = 59$

$X_6 = 29$

$X_7 = 48$

$X_8 = 81$




"independent and identically distributed"
(I.I.D.)

# Taking an average

$X_1 = 37$

$X_2 = 53$

$X_3 = 34$

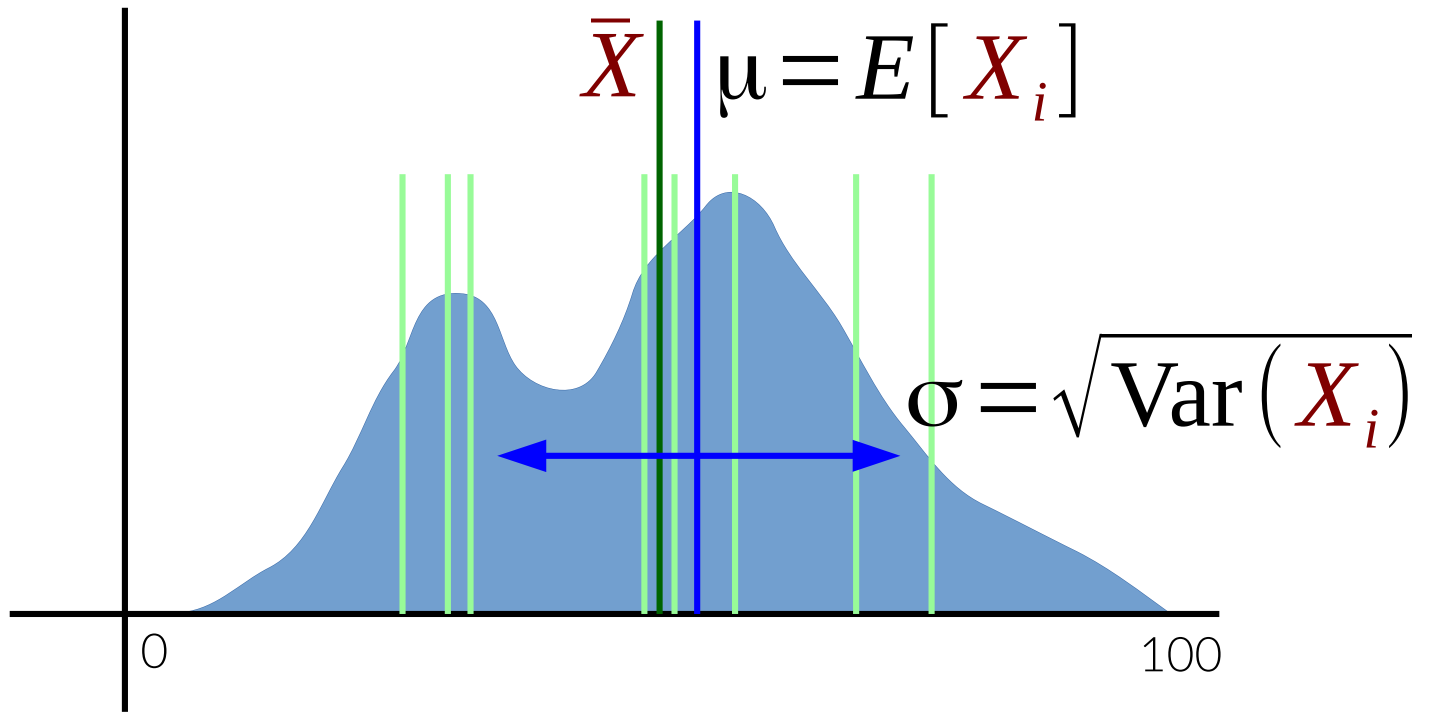
$X_4 = 70$

$X_5 = 59$

$X_6 = 29$

$X_7 = 48$

$X_8 = 81$



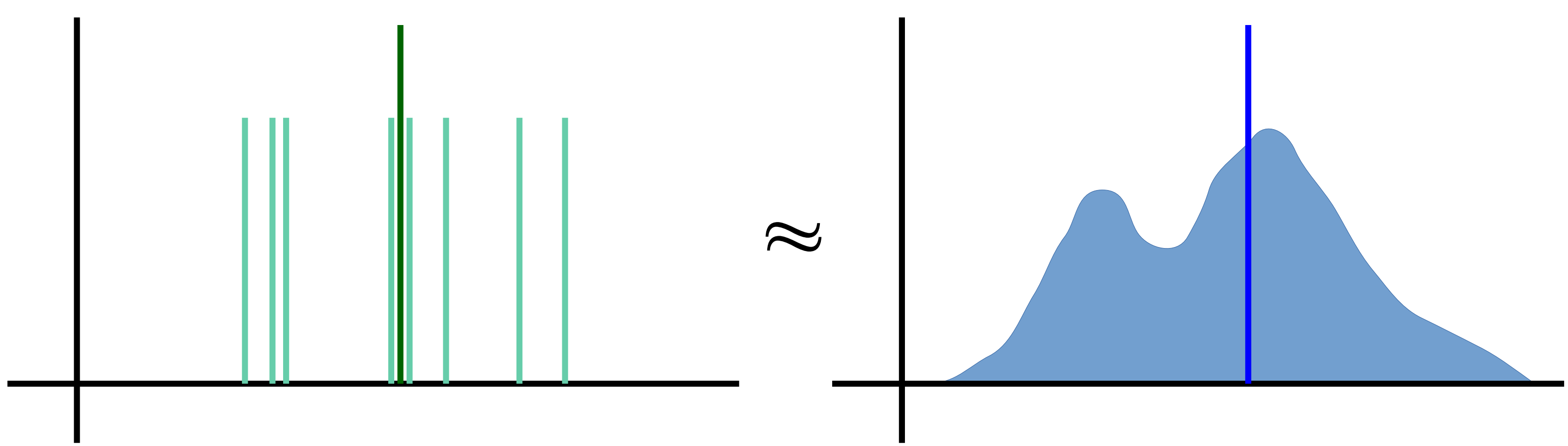

$$\bar{X} = \frac{1}{8}\sum_{i=1}^{8} X_i \approx 51.4$$

# Parameter estimation

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

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

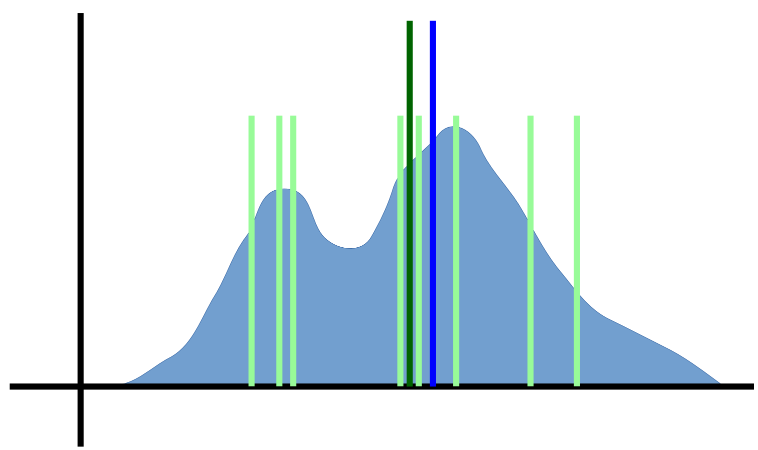
$$\hat{\theta} = \arg\max_{\theta} \mathrm{LL}(\theta)$$



# Unbiased estimator

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

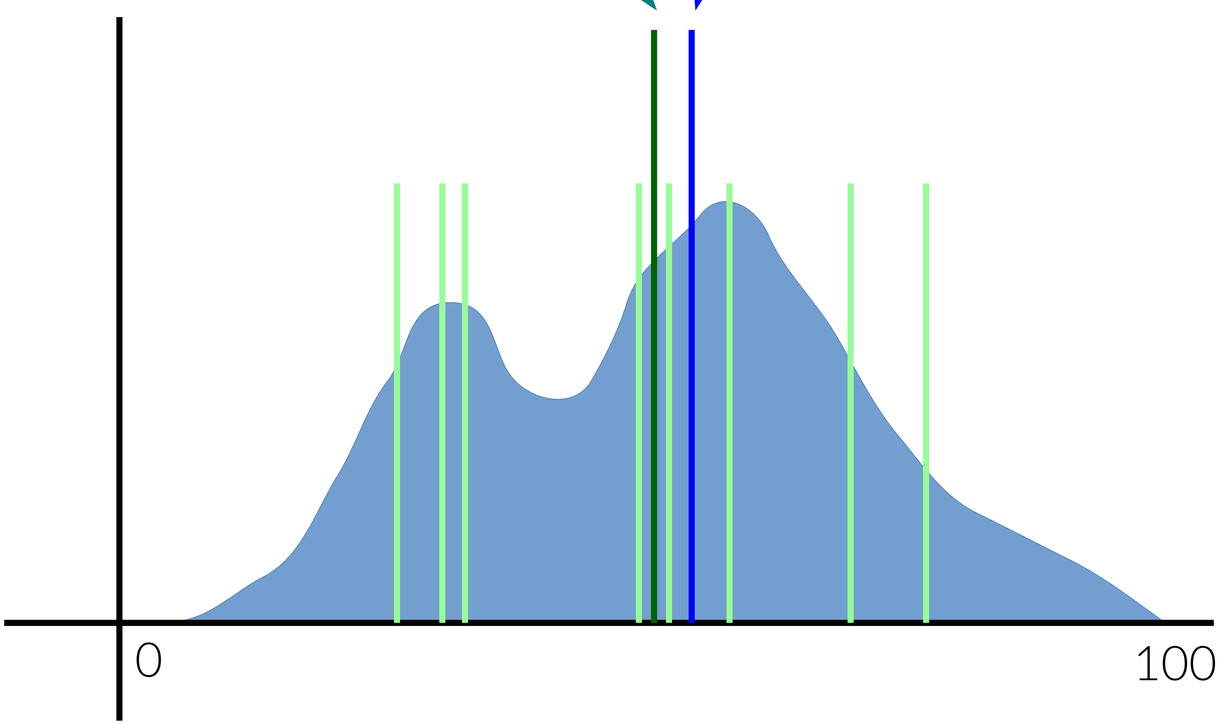
$$E[\bar{X}] = \mu = E[X_i]$$



# Sample mean is unbiased

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

$$\mu = E[X_i]$$



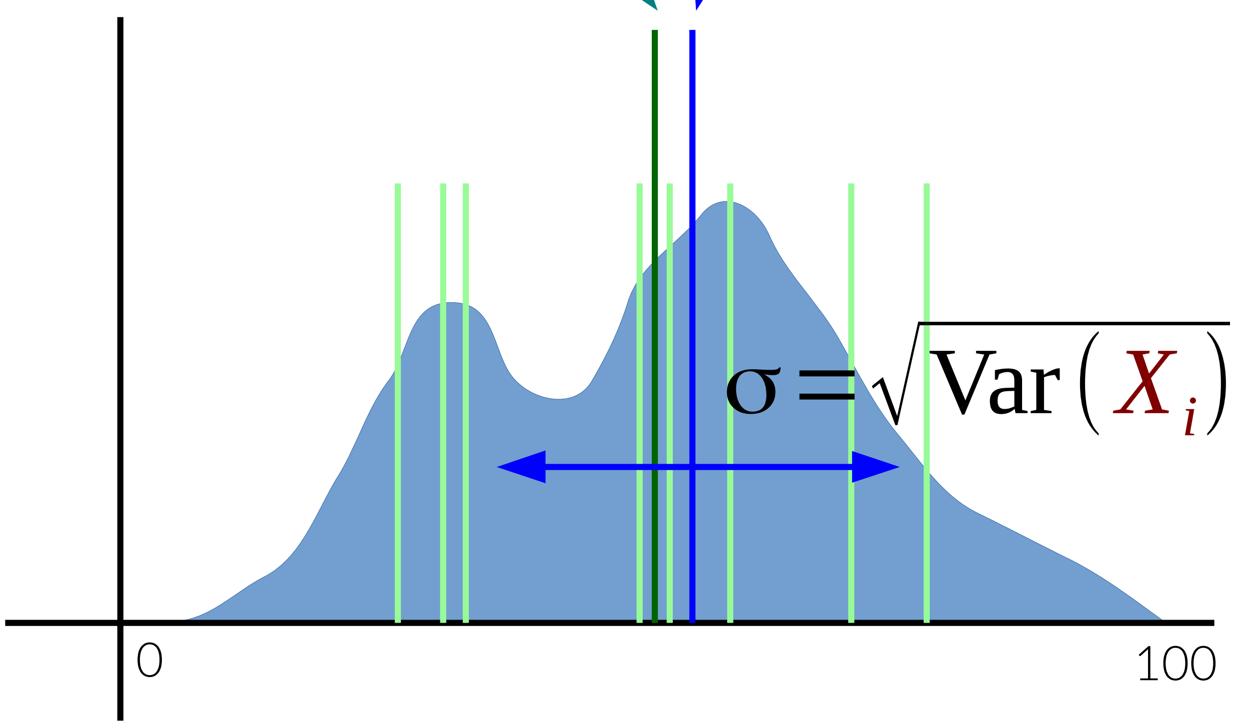

$$\begin{aligned} E[\bar{X}] &= E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \\ &= \frac{1}{n}\sum_{i=1}^{n} E[X_i] \\ &= \frac{1}{n}\sum_{i=1}^{n} \mu \\ &= \frac{1}{n}\cdot n\mu \\ &= \mu \end{aligned}$$

# How volatile is our estimate?

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

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

$$\mu = E[X_i]$$



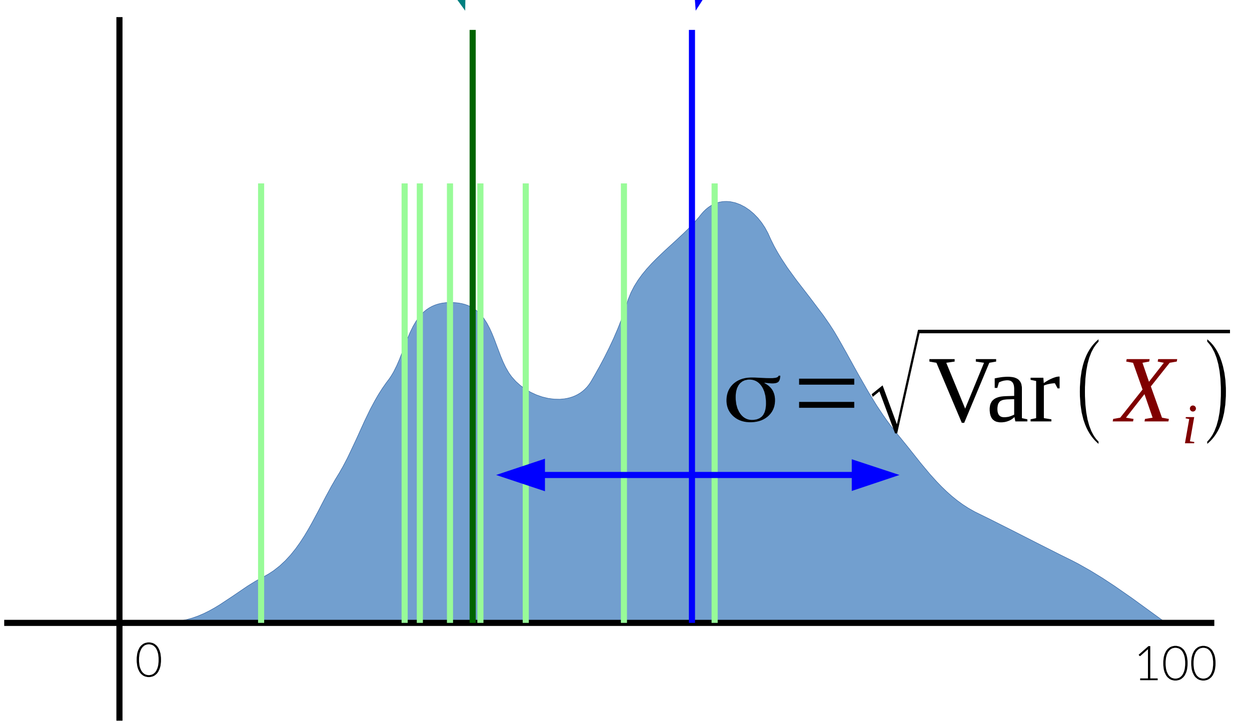

$$\sigma = \sqrt{\mathrm{Var}(X_i)}$$

# How volatile is our estimate?

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

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

$$\mu = E[X_i]$$




$$\sigma = \sqrt{\mathrm{Var}(X_i)}$$

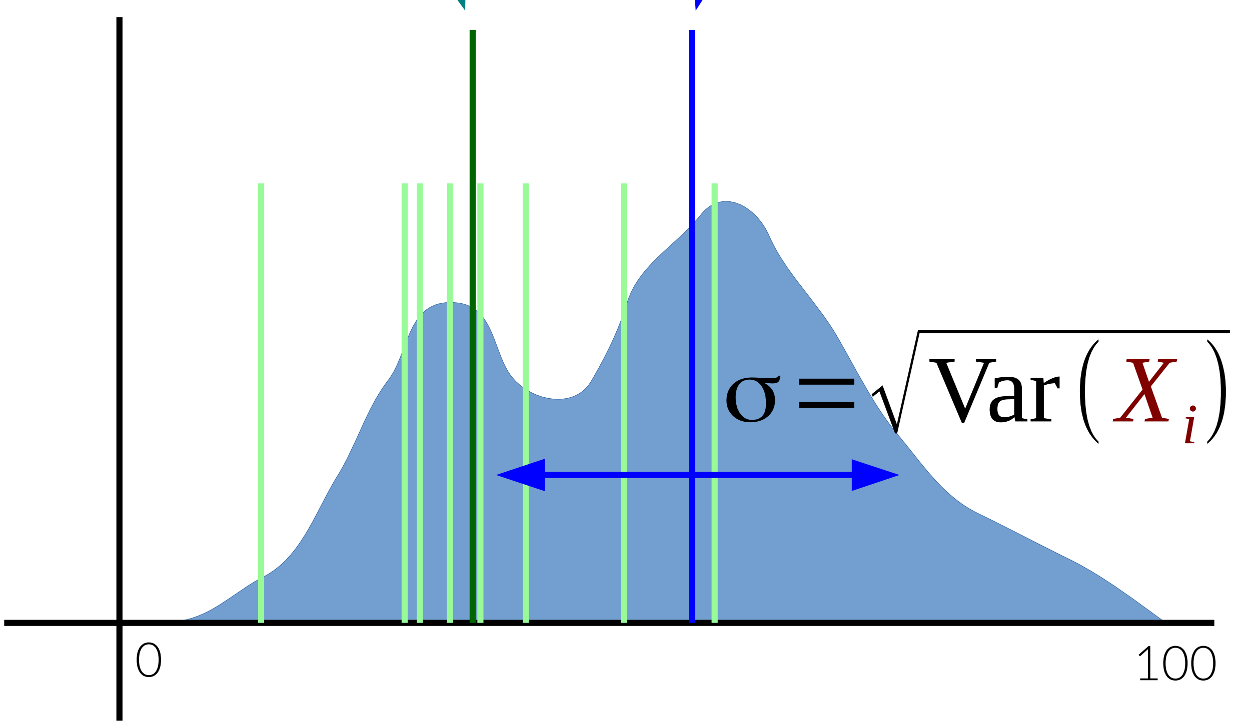
# How volatile is our estimate?

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

$$\mu = E[X_i]$$

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

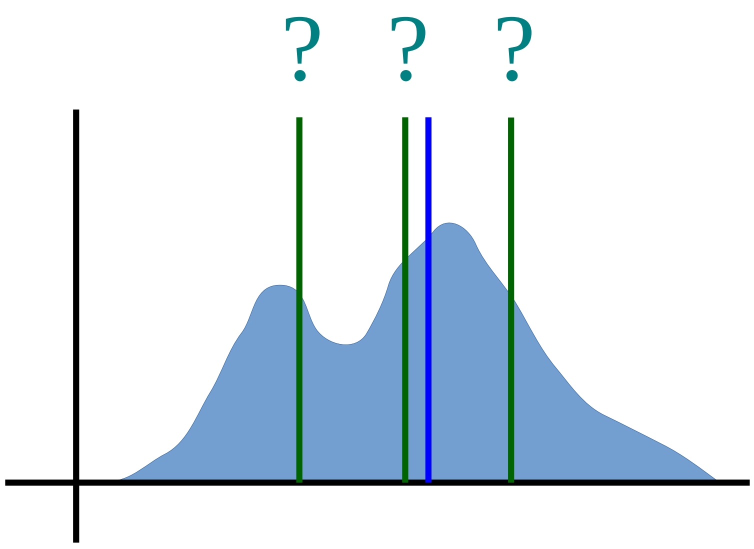
$$\mathrm{Var}(\bar{X}) = ?$$



# Variance of the sample mean

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

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

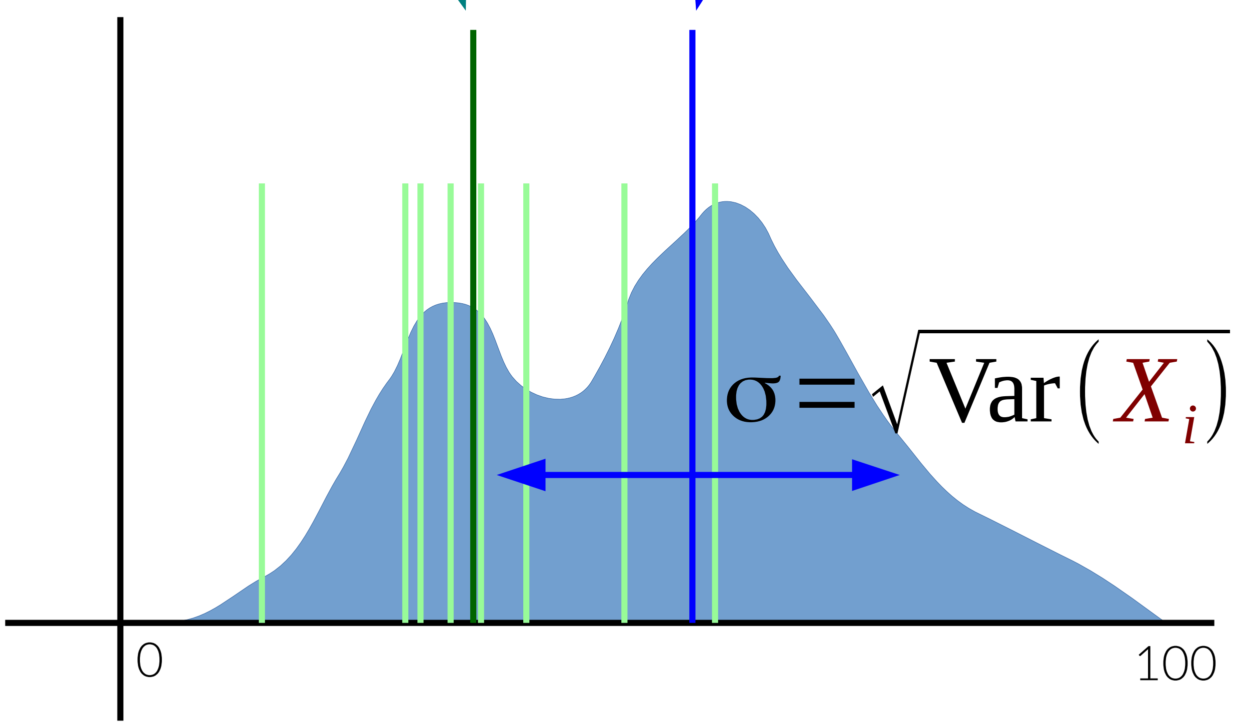


# How volatile is our estimate?

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

$$\mu = E[X_i]$$

$$\sigma = \sqrt{\mathrm{Var}(X_i)}$$



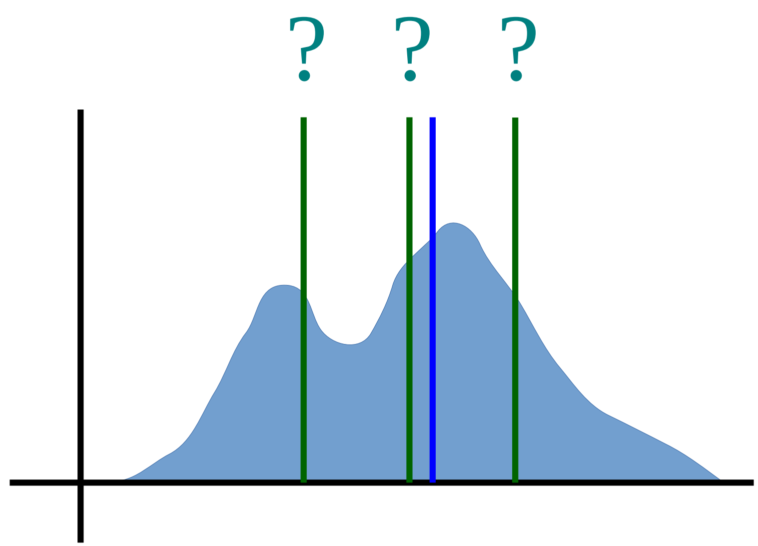

$$E[\bar{X}] = \mu$$

$$\begin{aligned}
\mathrm{Var}(\bar{X}) &= \mathrm{Var}\left(\sum_{i=1}^{n} \frac{X_i}{n}\right) \\
&= \left(\frac{1}{n}\right)^2 \mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(X_i) \\
&= \frac{1}{n^2} n \cdot \sigma^2 = \frac{\sigma^2}{n}
\end{aligned}$$

# Variance of the sample mean

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

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



# Teaser

Next week: **Central limit theorem**

(arguably "the greatest result in probability theory")—
lets you prove many statements about sample means

Later today: **Bootstrapping**

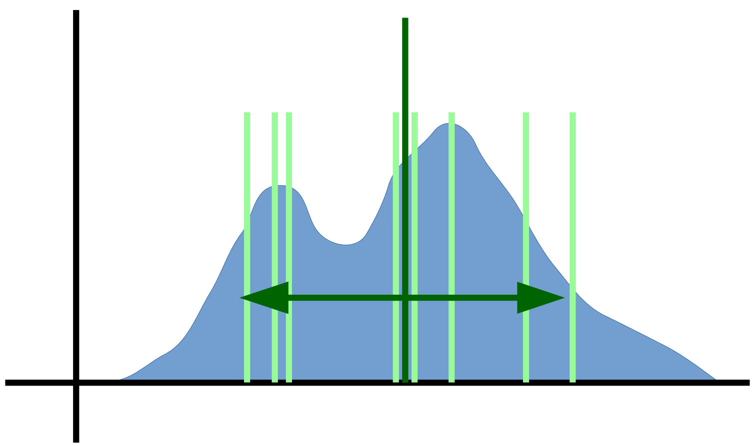
For when things are hard to derive analytically—
make the computer do the work for you!

Break time!

# Sample variance

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

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



# Estimating variance from samples

$$\operatorname{Var}(X_i) = E[(X_i - \mu)^2]$$

$$\approx E[(X_i - \bar{X})^2]$$

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

Unbiased? Nope!

$$E[S^2] = \boxed{\left(\frac{n-1}{n}\right)}\sigma^2$$

(algebra skipped—see lecture notes)

# Estimating variance from samples

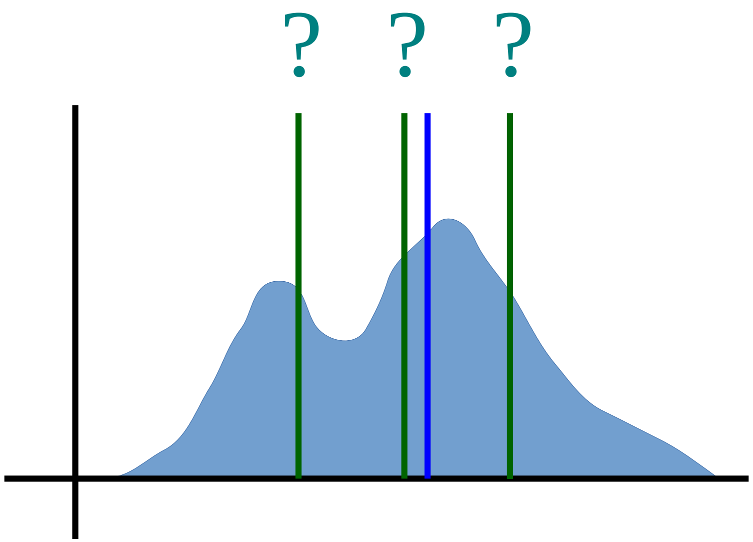
$$\begin{aligned} \mathrm{Var}(X_i) &= E[(X_i - \mu)^2] \\ &\approx E[(X_i - \bar{X})^2] \\ &\approx \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = S^2 \end{aligned}$$

Unbiased? Yes!

$$E[S^2] = \sigma^2$$

(algebra skipped—see lecture notes)

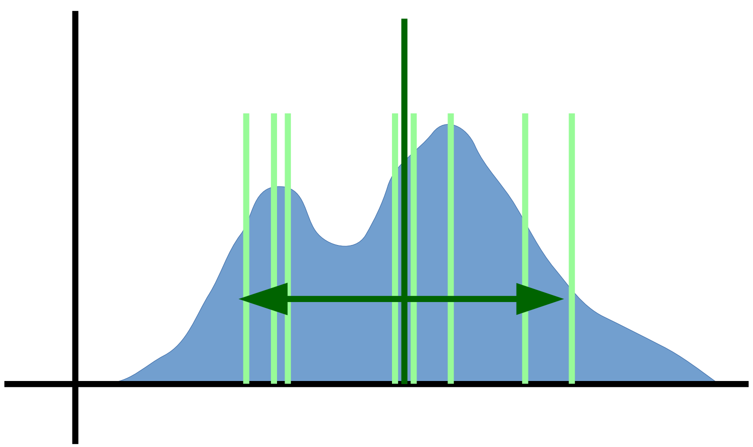
# Variance of the sample mean




- Is a single number
- Shrinks with number of samples $\left(=\dfrac{\sigma^2}{n}\right)$
- Measures the stability of an estimate

vs.

# Sample variance

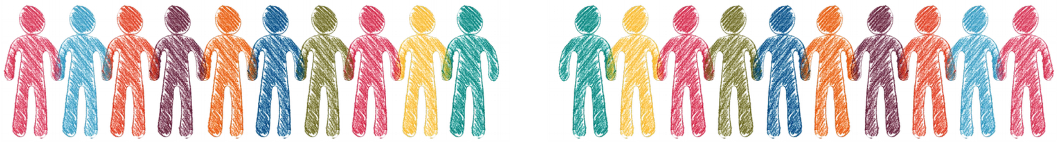


- Is a random variable
- Constant with number of samples $\left(\approx\sigma^2\right)$
- Is an estimate (of a variance) itself

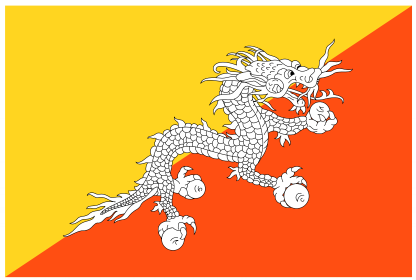
# p-values

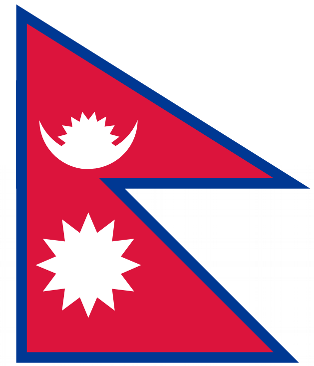
A ***p*-value** gives the probability of an extreme result, assuming that any extremeness is due to chance.

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



# Comparing two samples





$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \approx 87.1$$

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \approx 87.6$$

# Is it a fluke?

Sample means have
random fluctuations.
What's the probability that
we see the difference we
found if any differences
are due to chance alone?

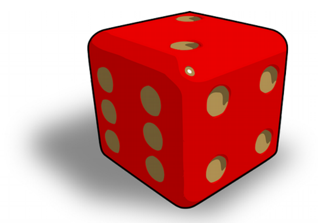
(Yes!)

# Is it a fluke?

Sample means have
random fluctuations.
What's the probability that
we see the difference we
found if any differences
are due to chance alone?

**null hypothesis** ($H_0$):
the assumption that any extreme
result happens by chance alone

# Suspicious dice



Roll a 6 on two out of three rolls of one die.

How likely is this by chance?

$H_0$ = die is fair, all extreme values are by chance

$X$ = number of 6's on three rolls

$$p = P(X \geq 2 | H_0) = P(X = 2 | H_0) + P(X = 3 | H_0)$$

$$= \binom{3}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right) + \left(\frac{1}{6}\right)^3$$

$$\approx 0.074$$

# Interpreting *p*-values

✗ Suppose I got this result. How likely is it to be a fluke?

✓ Suppose this result is a fluke. How unlikely is the result?

# Bootstrapping

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

# Bootstrapping motivation

Computers can **simulate** taking samples from many distributions.

What if we try to reverse-engineer the distribution from the sample we have, then simulate new samples?

# The “original” bootstrap

```
def bootstrap(sample):
    pmf = fancy_estimate_distribution(sample)
    results = []
    for i in range(10000):
        sample = pmf.sample(size=len(sample))
        stat = compute_stat(sample)
        results.append(stat)
    return results
```

# The "original" bootstrap

Also next week: **parameter estimation**
= how to write this function

```
def bootstrap(sample):
    pmf = fancy_estimate_distribution(sample)
    means = []
    for i in range(10000):
        sample = pmf.sample(size=len(sample))
        mean = np.mean(sample)
        means.append(mean)
    return means
```

Now you have a bunch of means.

Can answer questions like: what is P(mean is between 40 and 60)?

# Empirical distribution

$$X \sim \mathcal{E}:$$

$$P(X = x) = \text{fraction of values in the sample equal to } x$$

# Easy bootstrap

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

Draw a bunch of points from data we already have (with replacement)

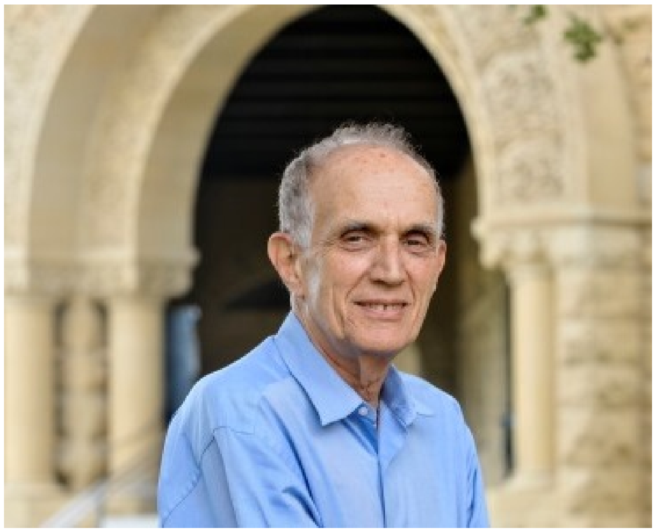
Now you have a bunch of means.

Can answer questions like: what is P(mean is between 40 and 60)?

# Bootstrap for p-values

```
def pvalue_bootstrap(sample1, sample2):
    n = len(sample1)
    m = len(sample2)
    observed_diff = abs(np.mean(sample2) -
                        np.mean(sample1))
    universal_pmf = sample1 + sample2
    count_extreme = 0
    for i in range(10000):
        resample1 = np.random.choice(universal_pmf, n)
        resample2 = np.random.choice(universal_pmf, m)
        new_diff = abs(np.mean(resample2) -
                       np.mean(resample1))
        if new_diff >= observed_diff:
            count_extreme += 1
    return count_extreme / 10000.
```

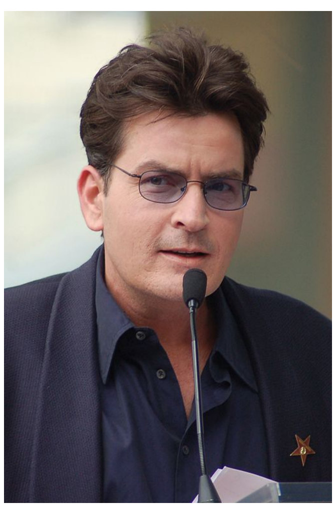
# You're in the right place



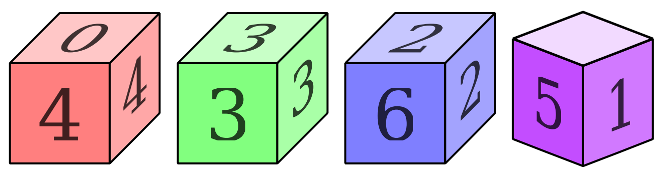
Bradley Efron (1938–)

Published paper proposing bootstrapping in 1979

At Stanford, still teaching as recently as 2015 (STATS 306A)!



(nope)




"Efron's dice"—

4 dice ($A$, $B$, $C$, $D$) such that:

$P(A > B) = P(B > C) = P(C > D) = P(D > A) = 2/3$