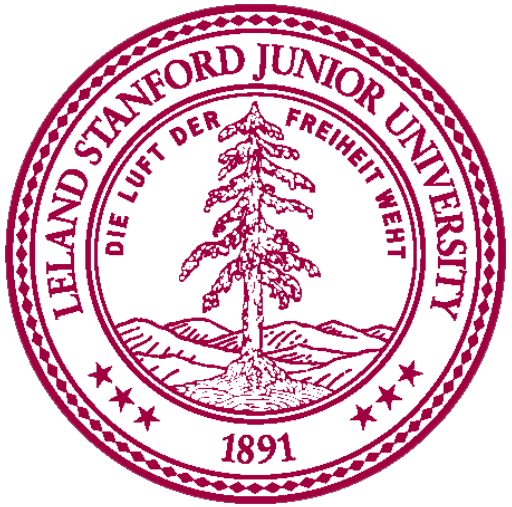
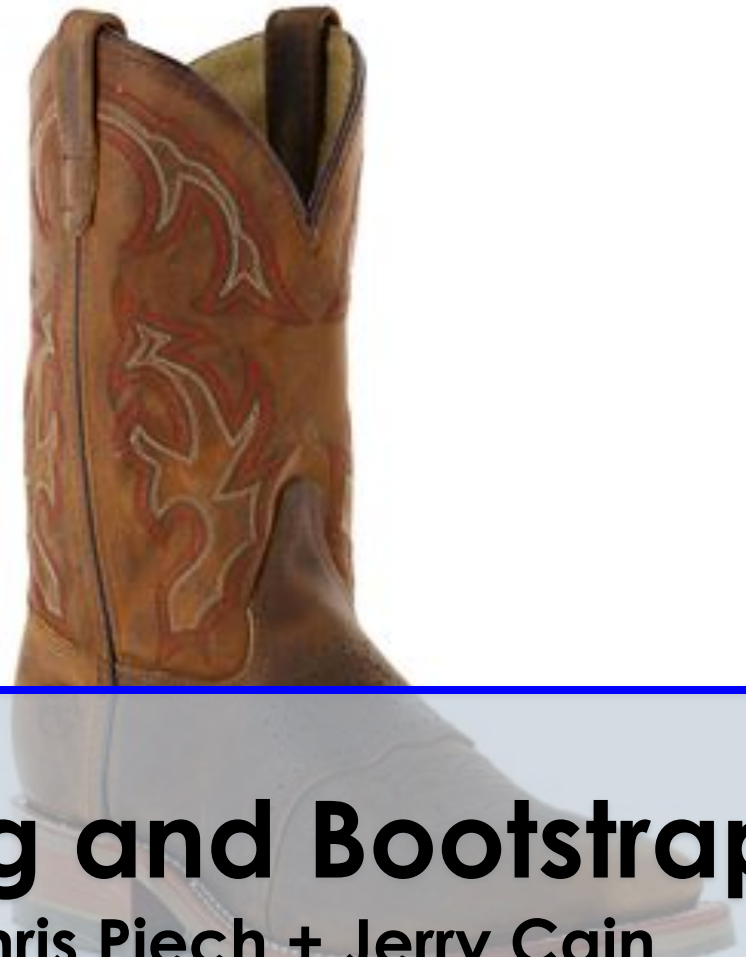




Sampling and Bootstrapping

Chris Piech + Jerry Cain
CS109, Stanford University

A real difference?

Learning in Context A	Learning in Context B
4.44	2.15
3.36	3.01
5.87	2.02
2.31	1.43
...	...
3.70	1.83

18 students

23 students

$\mu_1 = 3.1$

$\mu_2 = 2.4$

Claim: Group 1 and Group 2 are samples from different distributions with a 0.7 difference of means.

How confident are you in this claim?

The Classic Science Test

Group 1	Group 2
4.44	2.15
3.36	3.01
5.87	2.02
2.31	1.43
...	...
3.70	1.83
$\mu_1 = 3.1$	$\mu_2 = 2.4$

Claim: Group 1 and Group 2 are samples from different distributions with a 0.7 difference of means.

How confident are you in this claim?

<review>

Independent, Identically Distributed

Consider *n* random variables X_1, X_2, ... X_n

- X_i are all independently and identically distributed (I.I.D.)
- All have the same PMF (if discrete) or PDF (if continuous)
- All have the same expectation
- All have the same variance

IID iid

The sum of independent, identically distributed variables:

$$Y = \sum_{i=0}^{n} X_i$$

Is normally distributed:

$$Y \sim N(n\mu, n\sigma^2)$$

---

where $\mu = E[X_i]$

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

By the Central Limit Theorem, the average of IID variables is distributed normally:

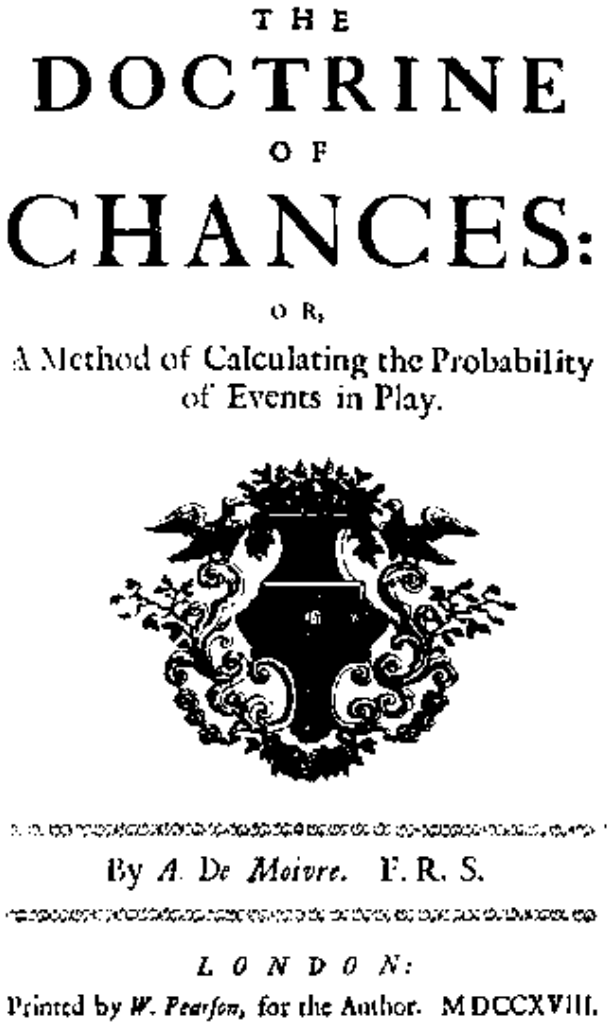
$$\frac{1}{n}\sum_{i=0}^{n} X_i \sim N(\mu, \frac{\sigma^2}{n})$$

---

where $\mu = E[X_i]$

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

How did Probability become English Language Heavy?

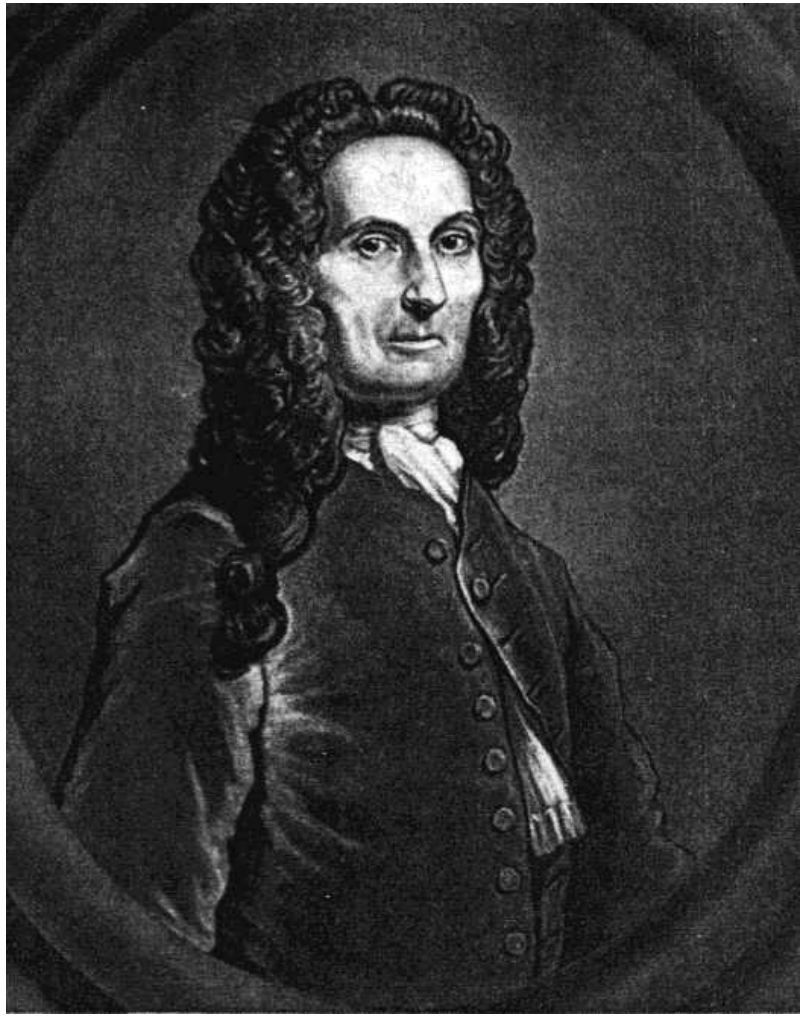


THE
DOCTRINE
OF
CHANCES:
OR,
A Method of Calculating the Probability
of Events in Play.

By *A. De Moivre.* F. R. S.

LONDON:
Printed by *W. Pearson,* for the Author. MDCCXVIII.

Abraham De Moivre



</review>

Sampling definitions

Motivating example

You want to know the true mean and variance of happiness in Bhutan.

- But you can't ask everyone.
- You poll 200 random people.
- Your data looks like this:

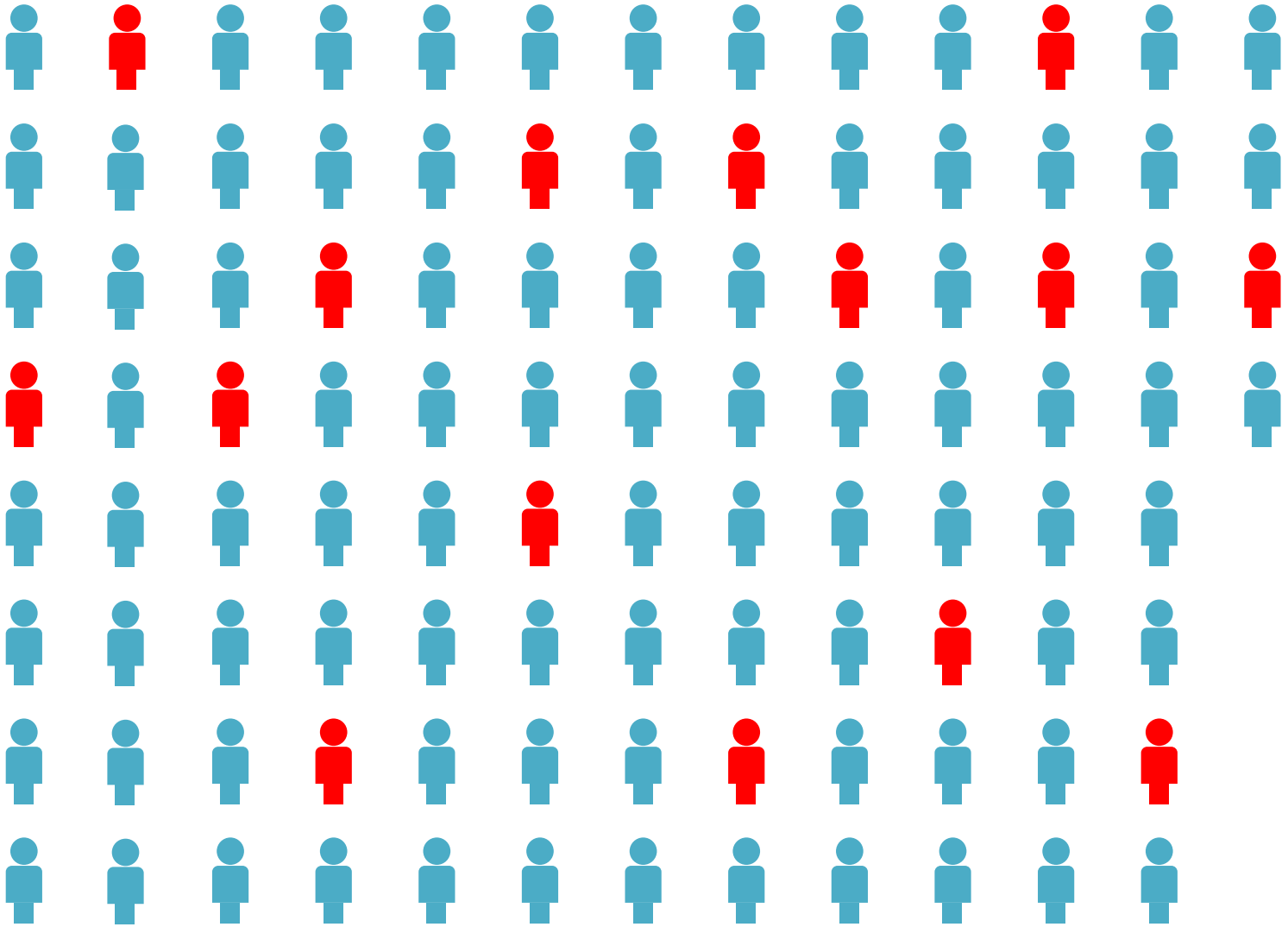
  Happiness = {72, 85, 79, 91, 68, ..., 71}

- The mean of all these numbers is 83.

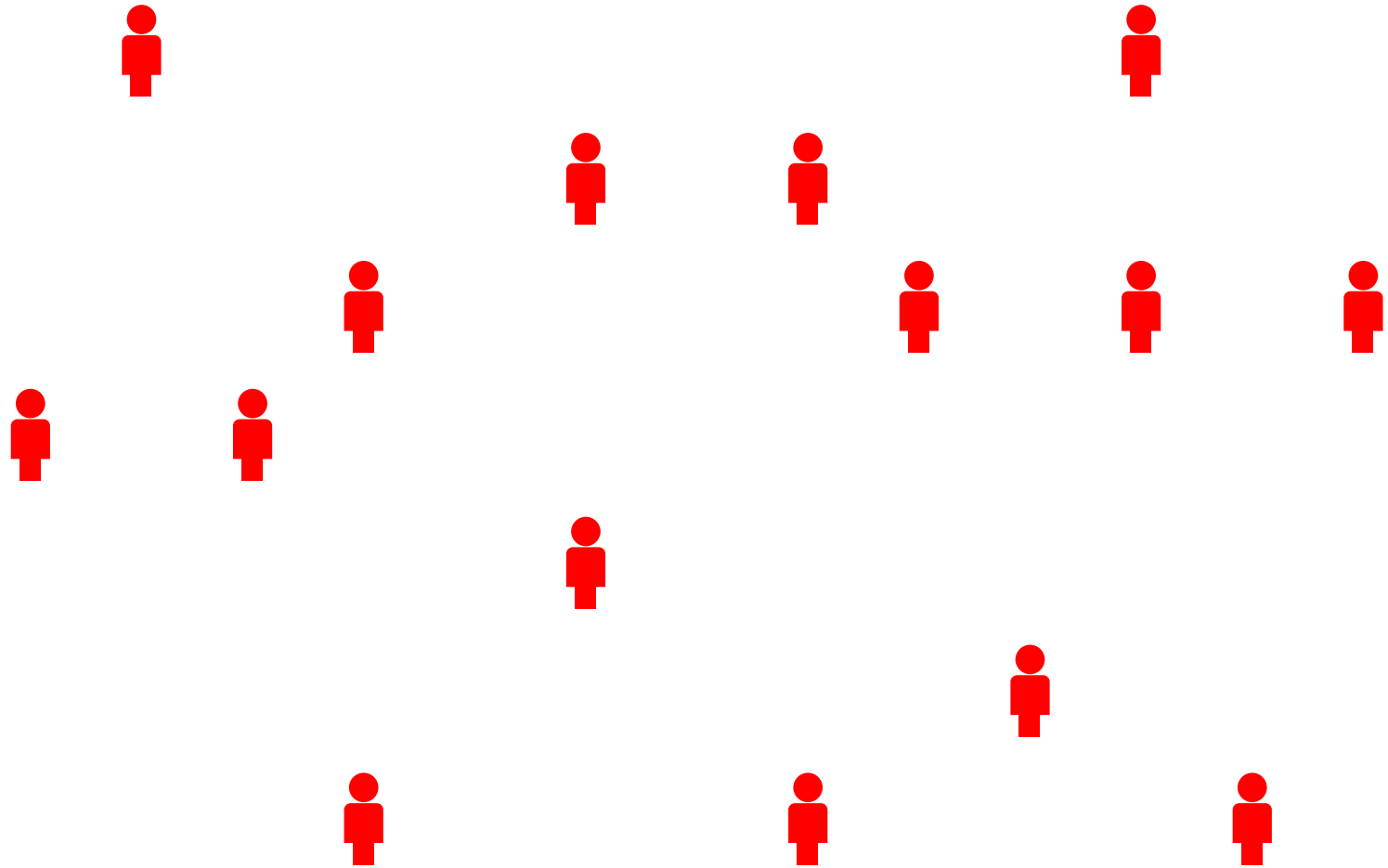
Is this the true mean happiness of Bhutanese people?

Population

Sample



Sample



Collect one (or more) numbers from each person

Population

This is a population.

Sample

A sample is selected from a population.

Sample










Collect **one (or more)** numbers from each person

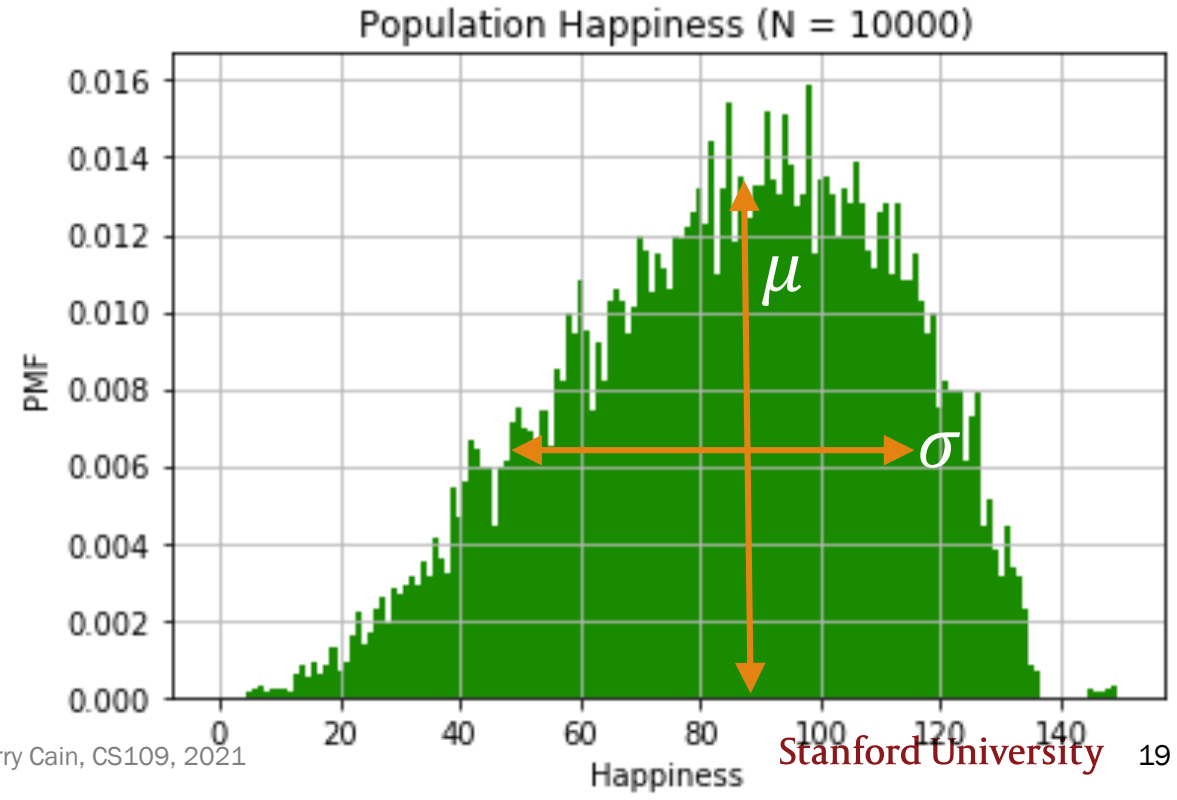
Sample

A sample, mathematically

Consider n random variables $X_1, X_2, \ldots, X_n$.

The sequence $X_1, X_2, \ldots, X_n$ is a **sample** from distribution F if:

- X_i are all independent and identically distributed (i.i.d.)
- X_i all have same distribution function F (the **underlying distribution**), where $E[X_i] = \mu$, $\text{Var}(X_i) = \sigma^2$



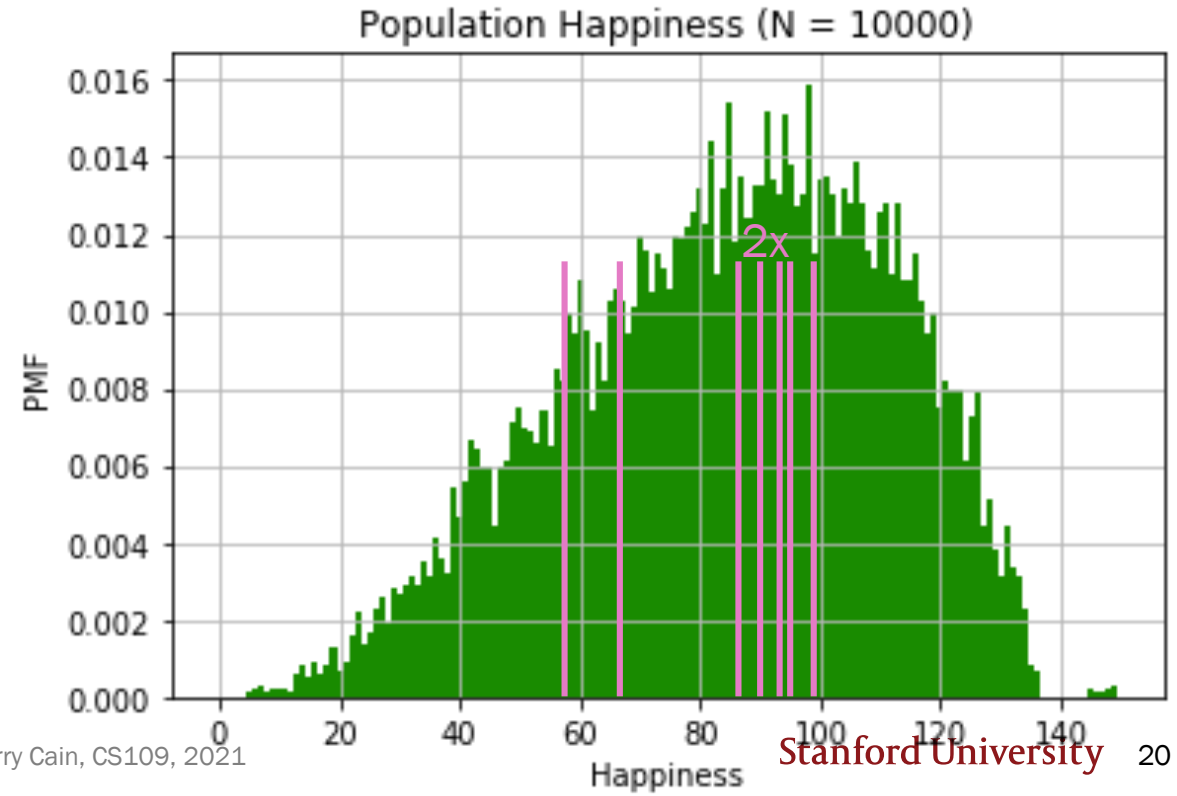
A sample, mathematically

A sample of **sample size** 8:

$$(X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8)$$

A **realization** of a sample of size 8:

$$(59, 87, 94, 99, 87, 78, 69, 91)$$



A single sample

A happy
Bhutanese person

If we had a distribution F of our entire population, we could compute exact statistics about about happiness.

But we only have 200 people (a sample).

Today: If we only have a single sample,

- How do we report *estimated* statistics?
- How do we report estimated error of these estimates?
- How do we perform hypothesis testing?

Estimating Core Statistics (Mean + Var)

A single sample

A happy
Bhutanese person

If we had a distribution F of our entire population, we could compute exact statistics about about happiness.

But we only have 200 people (a sample).

So these population statistics are unknown:

- μ, the population mean
- σ^2, the population variance

A single sample

A happy
Bhutanese person

If we had a distribution F of our entire population, we could compute exact statistics about about happiness.

But we only have 200 people (a sample).

- From these 200 people, what is our best estimate of population mean and population variance?
- How do we define best estimate?

Estimating the Mean

Consider *n* random variables X_1, X_2, ... X_n

- X_i are all independently and identically distributed (I.I.D.)
- Have same distribution function *F* and $E[X_i] = \mu$
- We call sequence of X_i a **sample** from distribution *F*
- *How would you estimate the population mean??*

$$\text{Estimate} = \frac{1}{n}\sum_{i=0}^{n} X_i$$

Sample Mean: This is a fancy way of saying "your estimate of the mean"

$$\bar{X} = \frac{1}{n}\sum_{i=0}^{n} X_i$$

Is that estimate any good?

$$\bar{X} = \frac{1}{n}\sum_{i=0}^{n} X_i$$

Consider *n* random variables X_1, X_2, ... X_n

- Have same distribution function *F* and $E[X_i] = \mu$
- *Is our estimate of mean any good??*

$$E[\bar{X}] = E\left[\sum_{i=1}^{n}\frac{X_i}{n}\right] = \frac{1}{n}E\left[\sum_{i=1}^{n} X_i\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{1}{n}n\mu = \mu$$

Estimating the population mean

1. What is our best estimate of μ, the **mean happiness** of Bhutanese people?

If we only have a sample, $(X_1, X_2, \ldots, X_n)$:

The best estimate of μ is the **sample mean**:

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

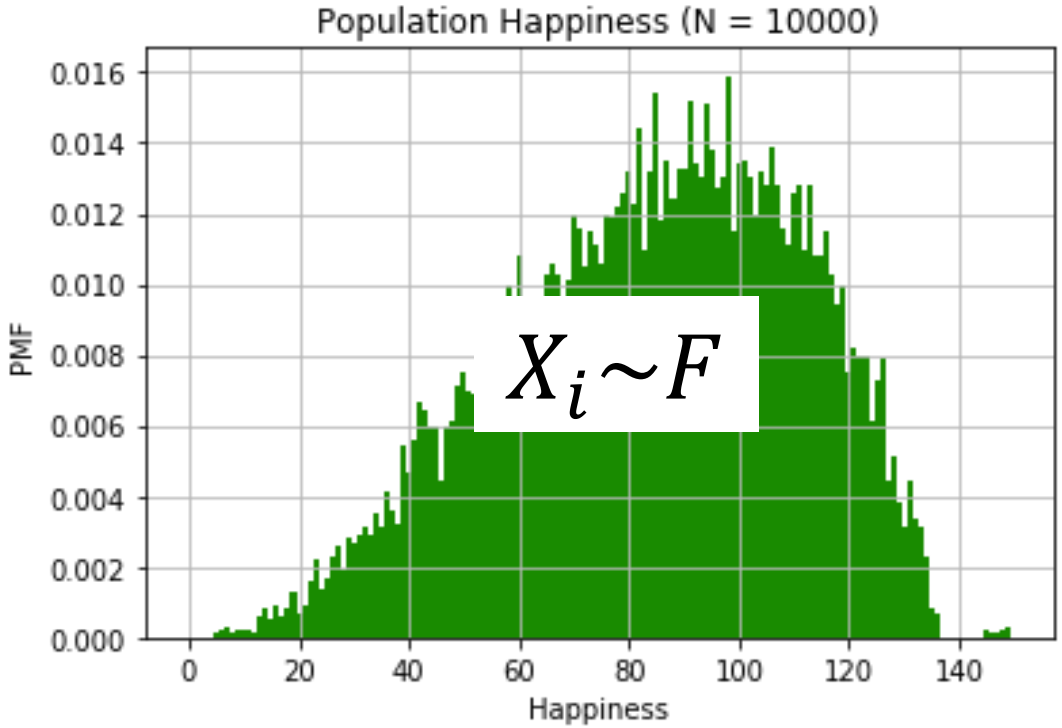
$\bar{X}$ is an **unbiased estimator** of the population mean μ. $E[\bar{X}] = \mu$

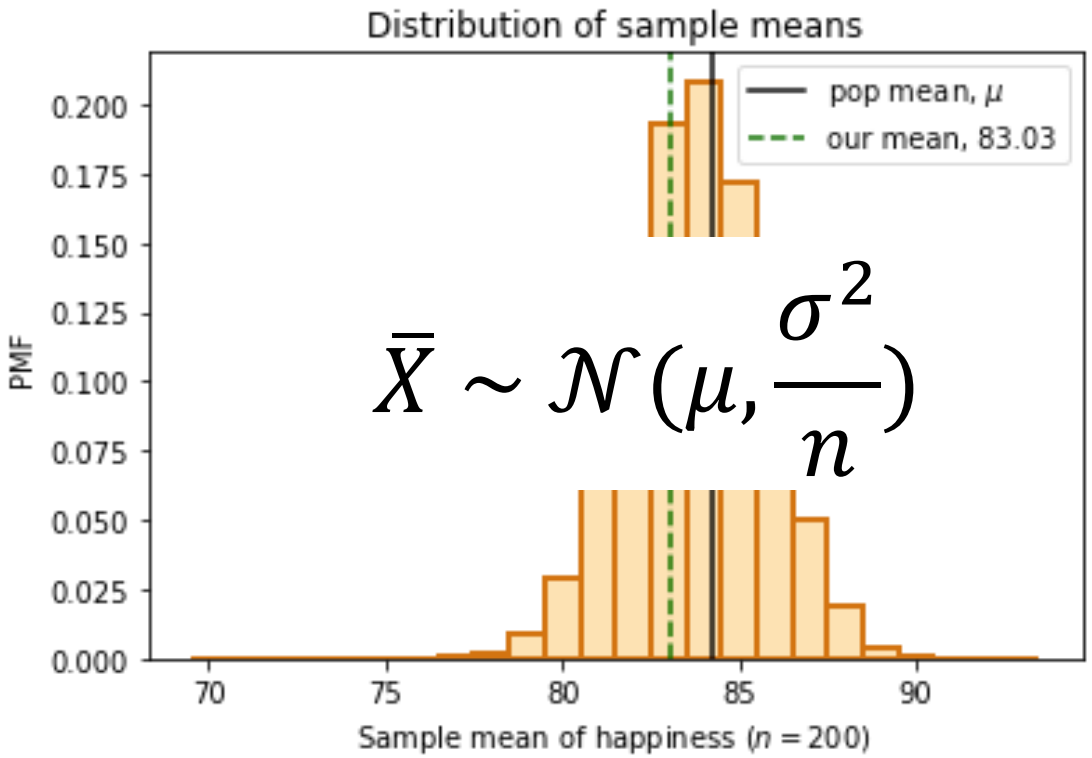
Intuition: By the CLT, $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$

If we could take *multiple* samples of size n:

1. For each sample, compute sample mean
2. On average, we would get the population mean

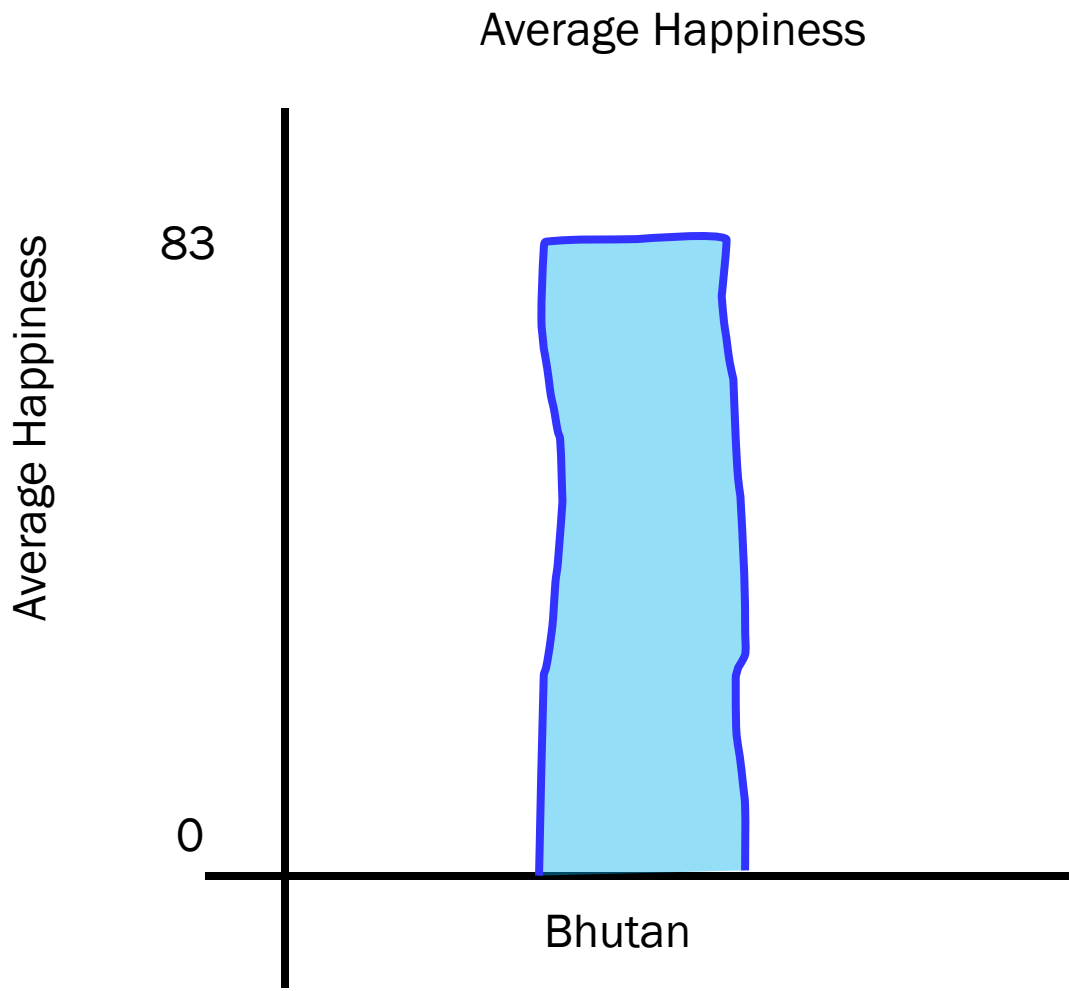
Sample mean





Even if we can't report μ, we can report our sample mean 83.03, which is an unbiased estimate of μ.

Our Report to Bhutan Government



Sample Mean:

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

ith sample

Size of the sample

Estimating the population variance

2. What is σ^2, the variance of happiness of Bhutanese people?

If we knew the entire population $(x_1, x_2, \ldots, x_N)$:

population mean

population variance $\sigma^2 = E[(X - \mu)^2] = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$

If we only have a sample, $(X_1, X_2, \ldots, X_n)$:

sample mean

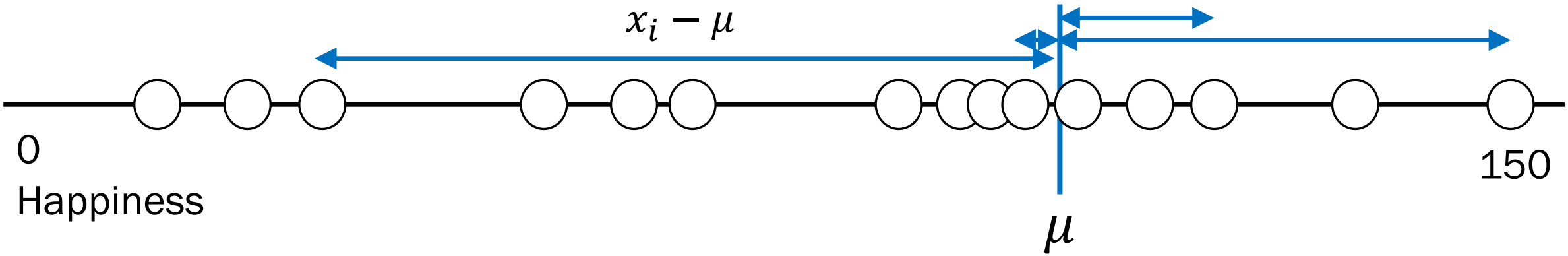
sample variance $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

Intuition about the sample variance, S^2

Actual, σ^2

population mean

population variance $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$



Population size, N

Calculating population statistics **exactly** requires us knowing all N datapoints.

Intuition about the sample variance, S^2

Actual, σ^2

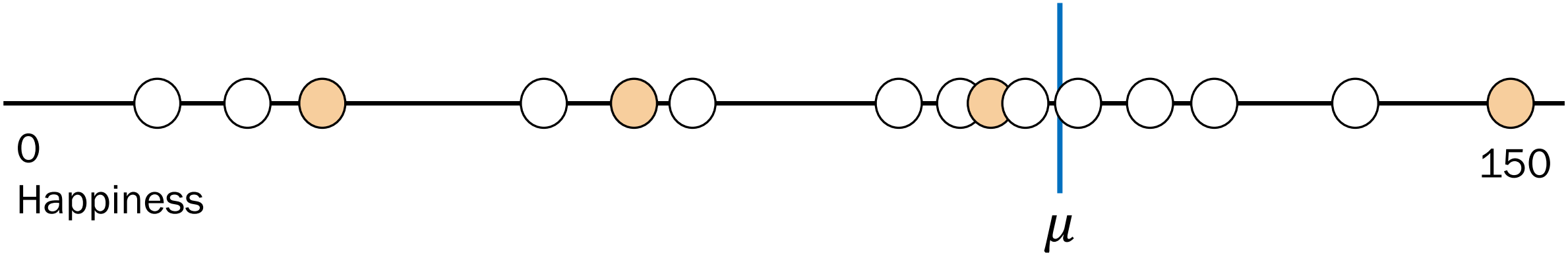
population mean

population variance $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$

Estimate, S^2

sample mean

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



Population size, N

Intuition about the sample variance, S^2

Actual, σ^2

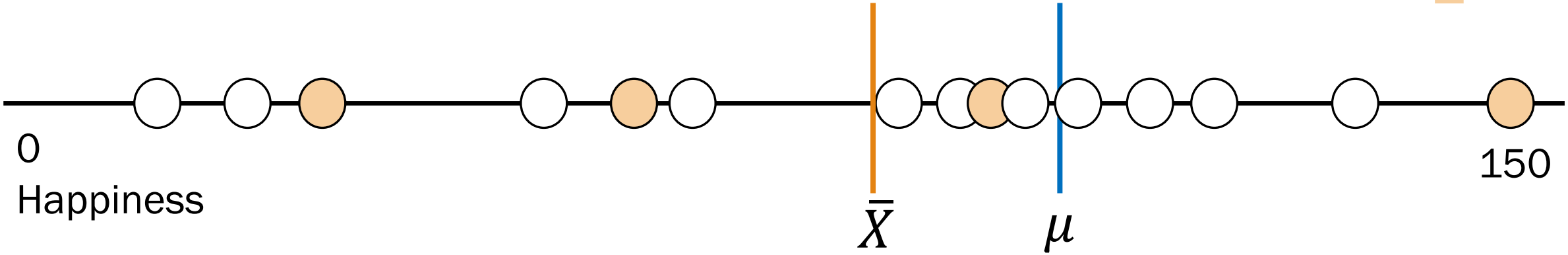
population mean

population variance $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$

Estimate, S^2

sample mean

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



Population size, N

Intuition about the sample variance, S^2

Actual, σ^2

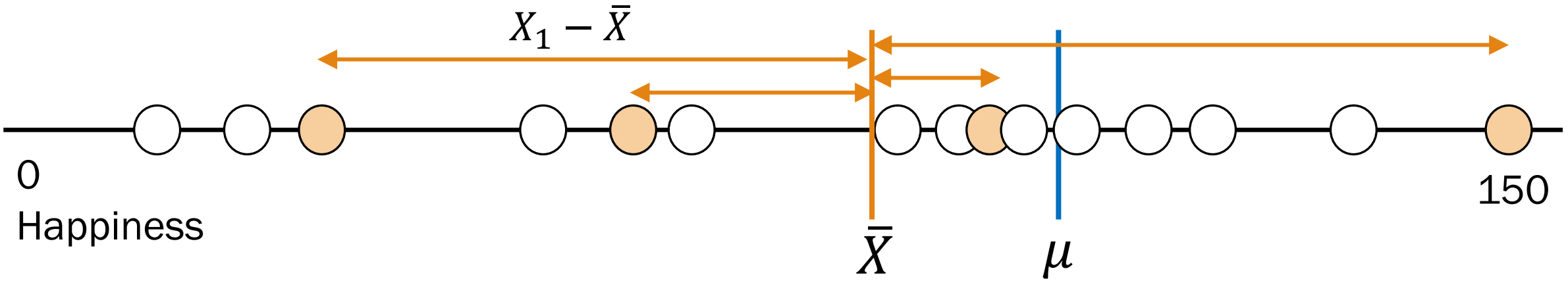
population mean

population variance $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$

Estimate, S^2

sample mean

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



Population size, N

Sample variance is an estimate using an estimate, so it needs additional scaling.

Proof that S^2 is unbiased (just for reference)

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

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

$$(n-1)E[S^2] = E\left[\sum_{i=1}^{n}\big((X_i - \mu) + (\mu - \bar{X})\big)^2\right] \qquad \text{(introduce } \mu - \mu\text{)}$$

$$= E\left[\sum_{i=1}^{n}(X_i - \mu)^2 + \sum_{i=1}^{n}(\mu - \bar{X})^2 + 2\sum_{i=1}^{n}(X_i - \mu)(\mu - \bar{X})\right]$$

$$= E\left[\sum_{i=1}^{n}(X_i - \mu)^2 + n(\mu - \bar{X})^2 - 2n(\mu - \bar{X})^2\right]$$

$$= E\left[\sum_{i=1}^{n}(X_i - \mu)^2 - n(\mu - \bar{X})^2\right] = \sum_{i=1}^{n} E[(X_i - \mu)]^2 - nE[(\bar{X} - \mu)^2]$$

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

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

$$2(\mu - \bar{X})n(\bar{X} - \mu)$$

$$-2n(\mu - \bar{X})^2$$

$$= n\sigma^2 - n\text{Var}(\bar{X}) = n\sigma^2 - n\frac{\sigma^2}{n} = n\sigma^2 - \sigma^2 = (n-1)\sigma^2$$

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

Estimating the population variance

2. What is σ^2, the variance of happiness of Bhutanese people?

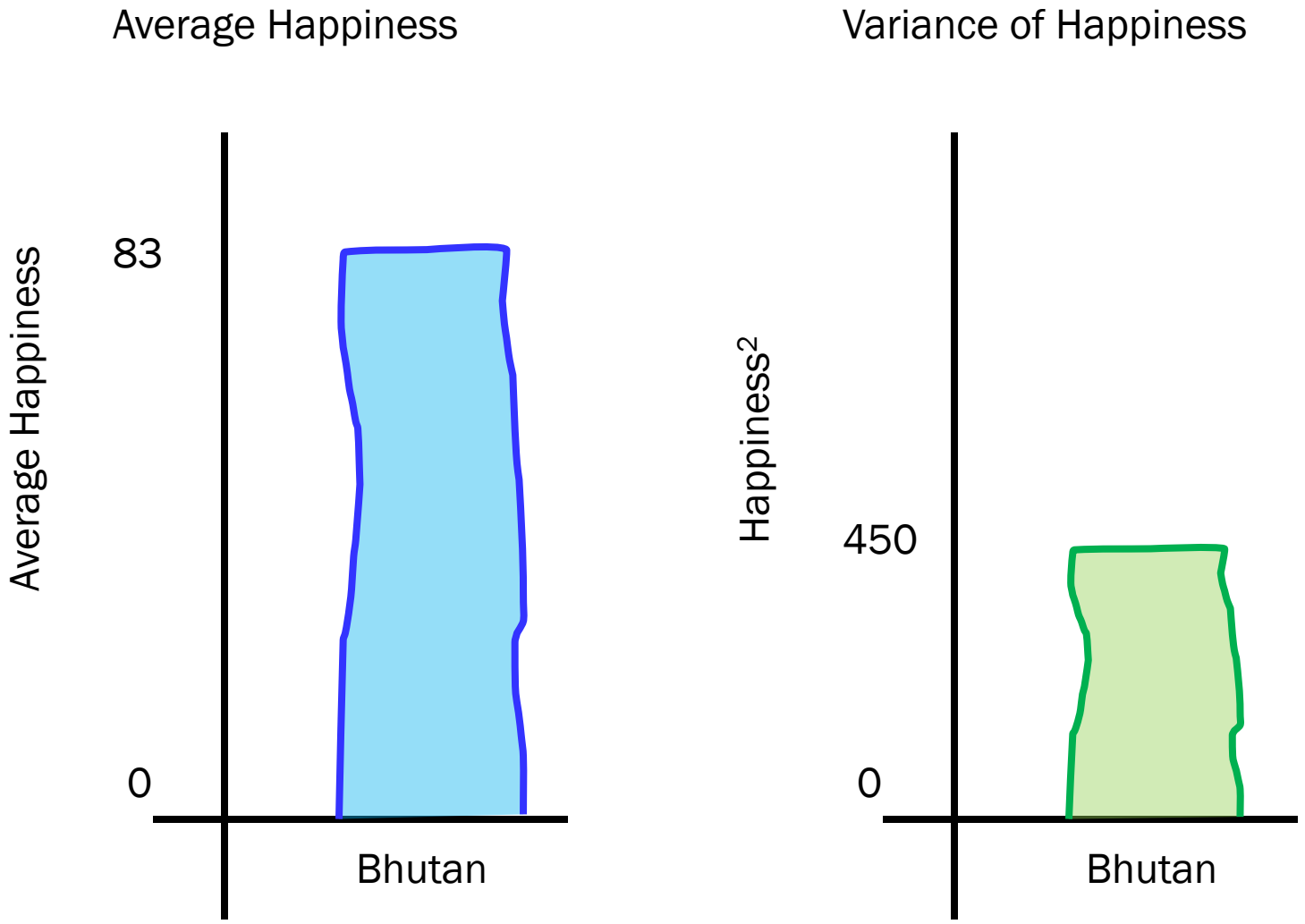
If we only have a sample, $(X_1, X_2, \ldots, X_n)$:

The best estimate of σ^2 is the **sample variance**:

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

S^2 is an **unbiased estimator** of the population variance, σ^2. $E[S^2] = \sigma^2$

Our Report to Bhutan Government



Sample Variance:

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

Sample mean

Makes it "unbiased"

No Error Bars ☹

Review: Estimating population statistics

A particular outcome

1. Collect a sample, $X_1, X_2, \ldots, X_n$. $(72, 85, 79, 79, 91, 68, \ldots, 71)$, $n = 200$

2. Compute **sample mean**, $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. $\bar{X} = 83$

3. Compute sample deviation, $X_i - \bar{X}$. $(-11, 2, -4, -4, 8, -15, \ldots, -12)$

4. Compute **sample variance**, $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$. $S^2 = 793$

How "close" are our estimates $\bar{X}$ and S^2?

Quick check

1. μ, the population mean
2. $(X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8)$, a sample
3. σ^2, the population variance
4. $\bar{X}$, the sample mean
5. $\bar{X} = 83$
6. $(X_1 = 59, X_2 = 87, X_3 = 94, X_4 = 99,$
   $X_5 = 87, X_6 = 78, X_7 = 69, X_8 = 91)$

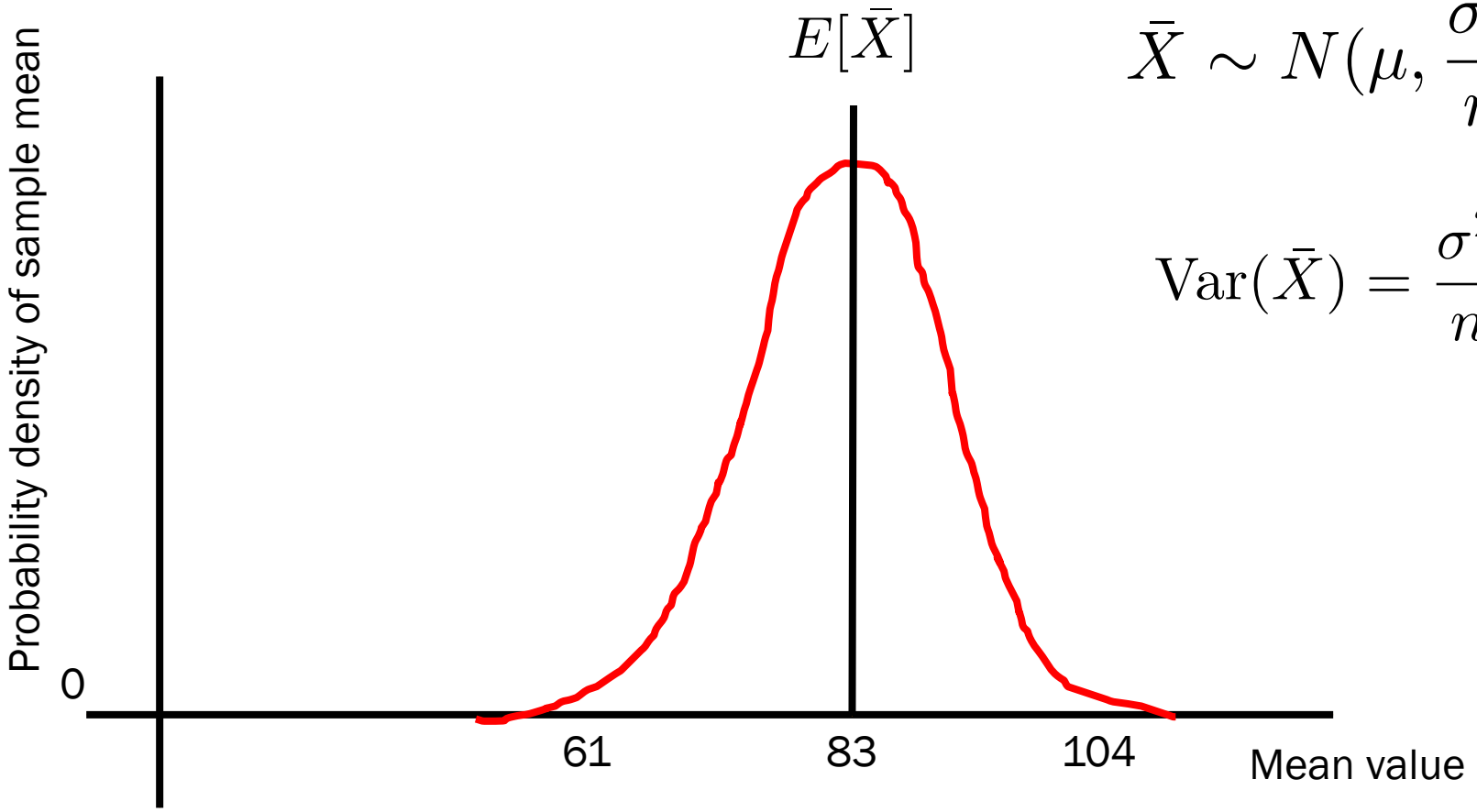
A. Random variable(s)
B. Value
C. Event

Insight: Sample Mean is an RV with known Var

By central limit theorem:

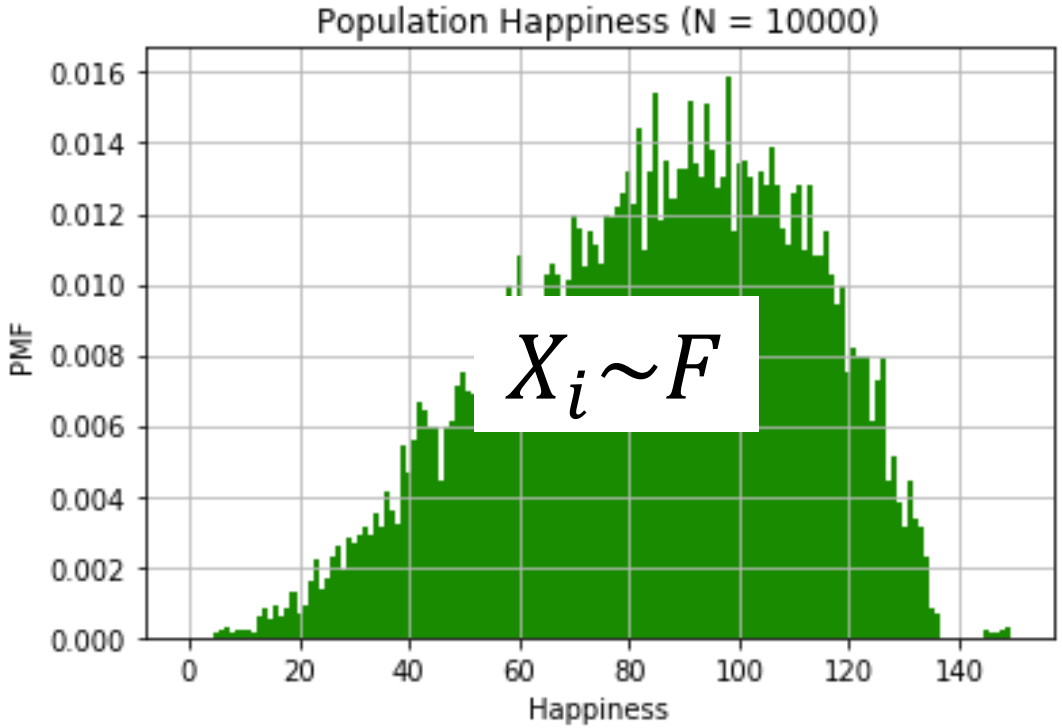
$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

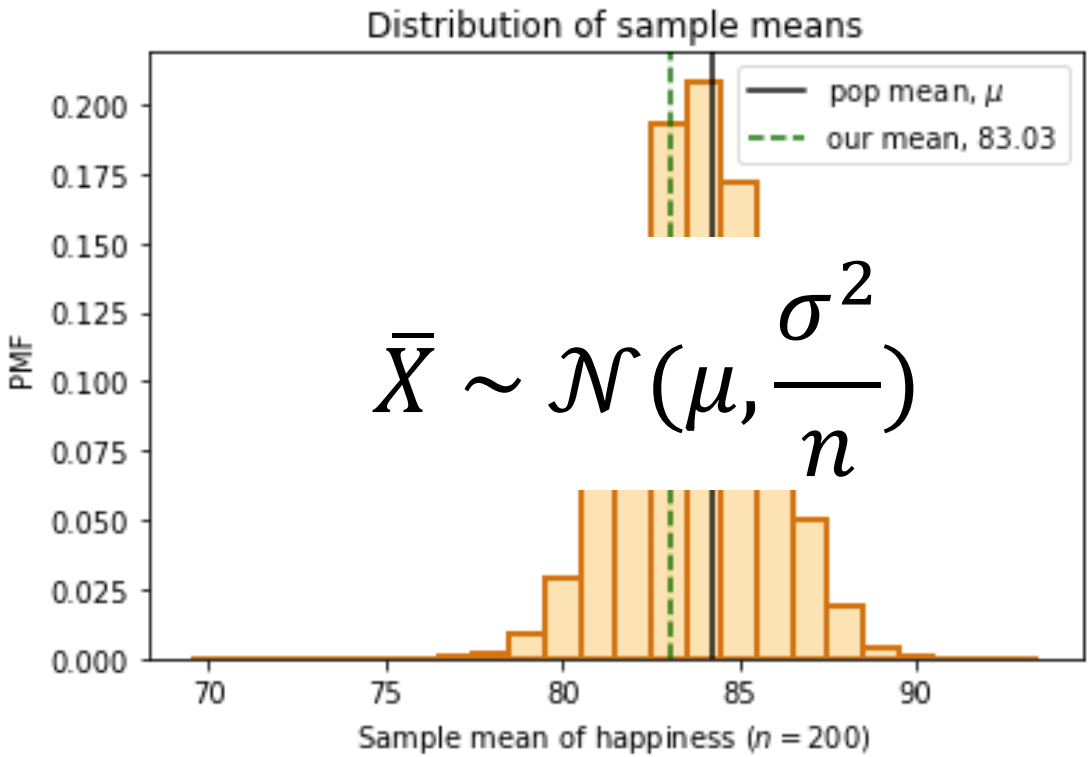
$$\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$$



Standard error of the mean

Sample mean





- $\text{Var}(\bar{X})$ is a measure of how "close" $\bar{X}$ is to μ.
- **How do we estimate $\text{Var}(\bar{X})$?**

Standard Error of the Mean

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

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

We want to estimate this

def The **standard error** of the mean is an estimate of the standard deviation of $\bar{X}$.

$$SE = \sqrt{\frac{S^2}{n}}$$

Intuition:

- S^2 is an unbiased estimate of σ^2
- S^2/n is an unbiased estimate of $\sigma^2/n = \text{Var}(\bar{X})$
- $\sqrt{S^2/n}$ can estimate $\sqrt{\text{Var}(\bar{X})}$

More info on bias of standard error: wikipedia

Standard Error of the Mean

$$\text{Var}(\bar{X}) = \text{Var}\left(\sum_{i=1}^{n} \frac{X_i}{n}\right) = \left(\frac{1}{n}\right)^2 \text{Var}\left(\sum_{i=1}^{n} X_i\right) = \frac{\sigma^2}{n}$$

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

$$= \frac{S^2}{n}$$

Since S_2 is an unbiased estimate

$$\text{Std}(\bar{X}) = \sqrt{\frac{S^2}{n}}$$

Change variance to standard deviation

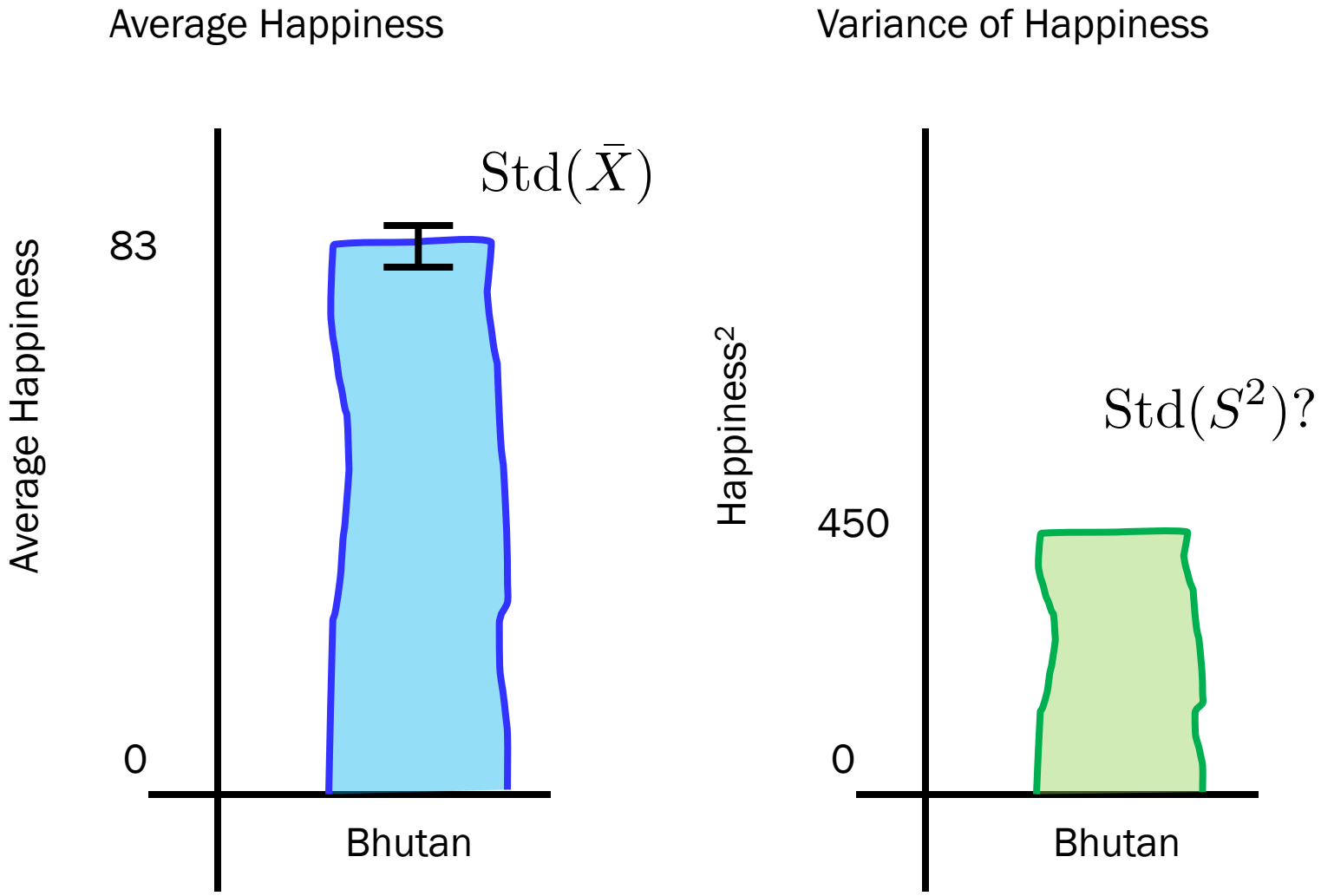
$$= \sqrt{\frac{450}{200}}$$

The numbers for our Bhutanese poll

$$= 1.5$$

Bhutanese standard error of the mean

Our Report to Bhutan Government



Claim: The average happiness of Bhutan is 83 ± 2

Bootstraping

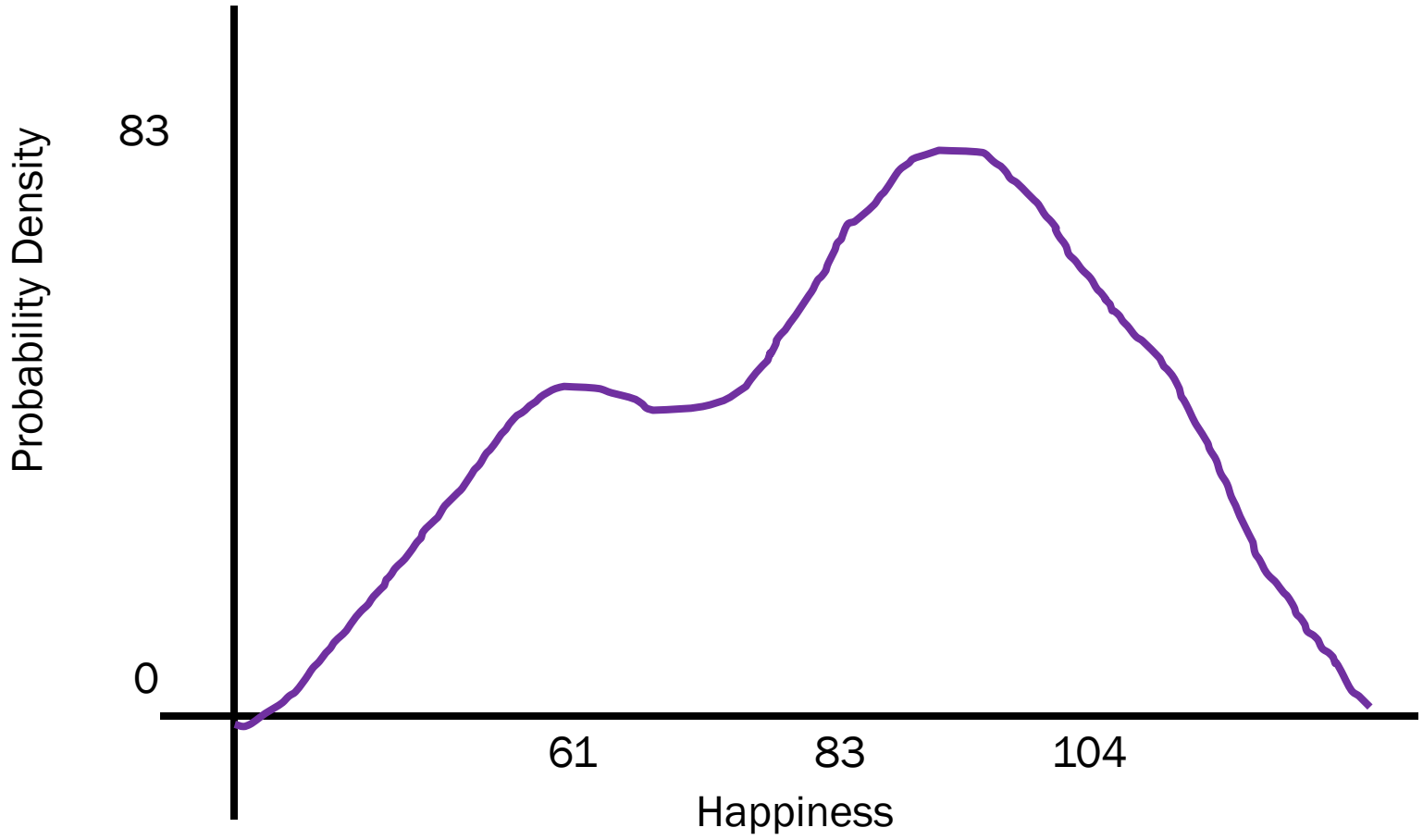
Bootstrap:
Probability for Computer Scientists

Bootstraping allows you to:

- Know the **distribution of statistics**
- Calculate **p values**

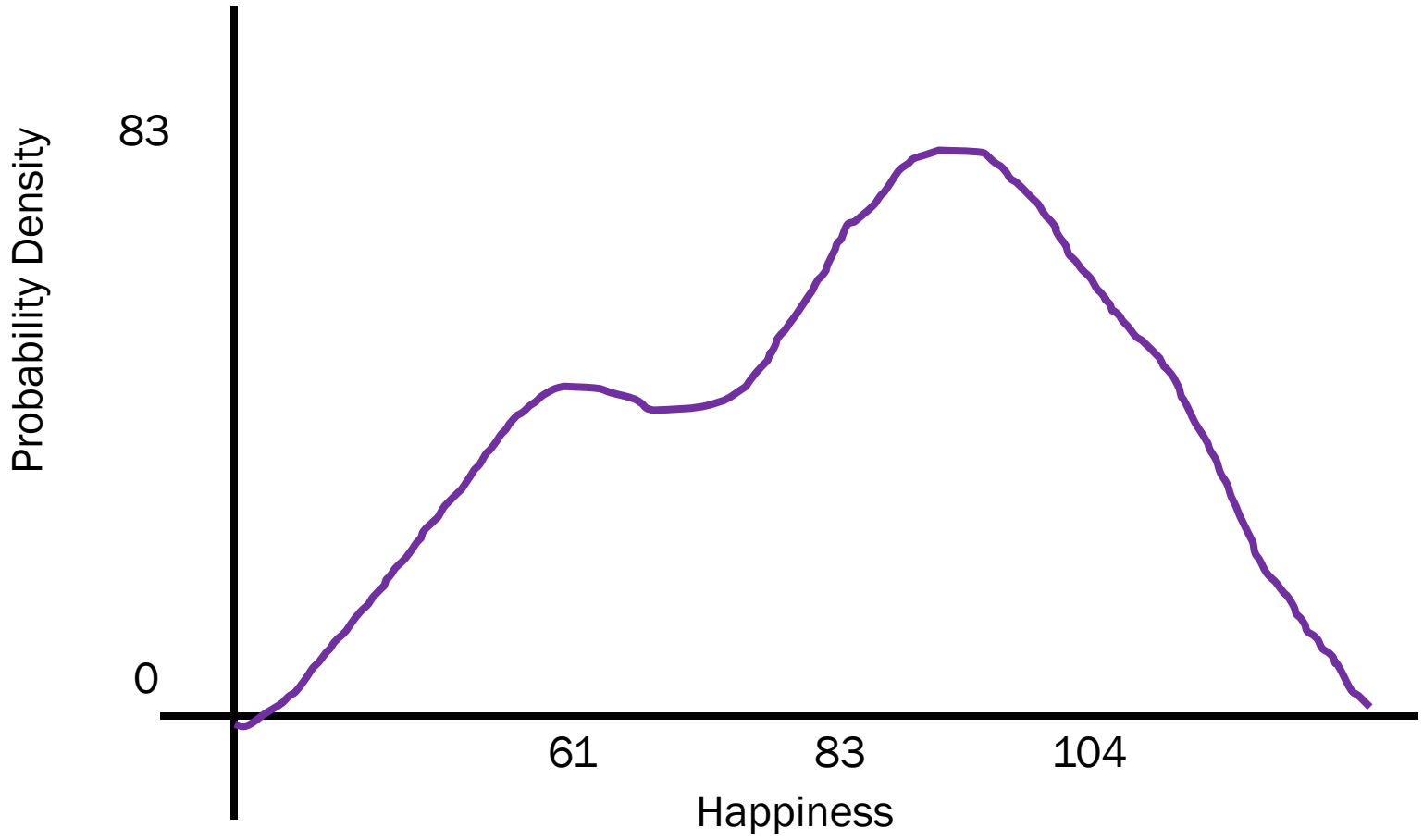
Hypothetical

What is the probability that a Bhutanese peep is just straight up loving life?



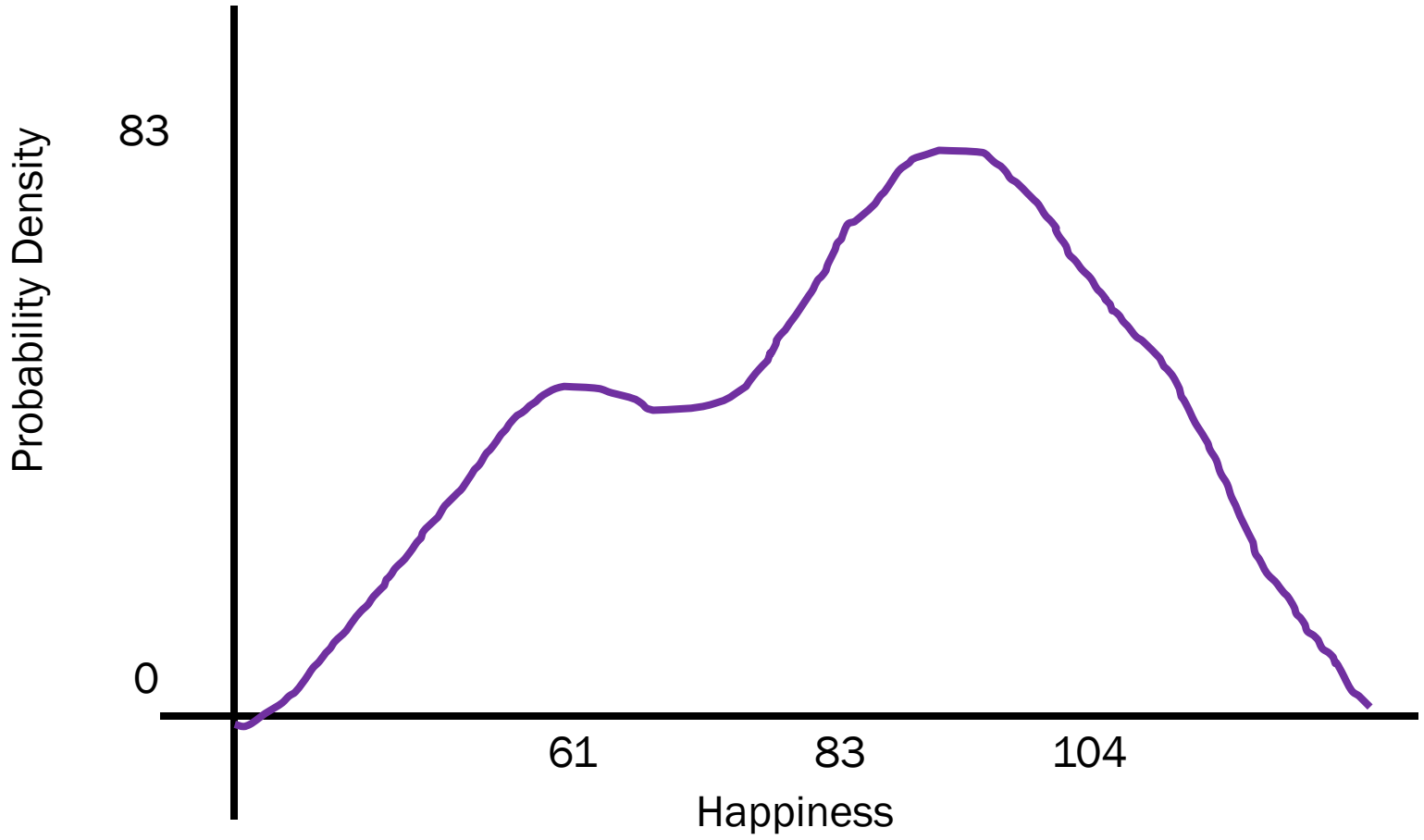
Hypothetical

What is the probability that the mean of a sample of 200 people is within the range 81 to 85?



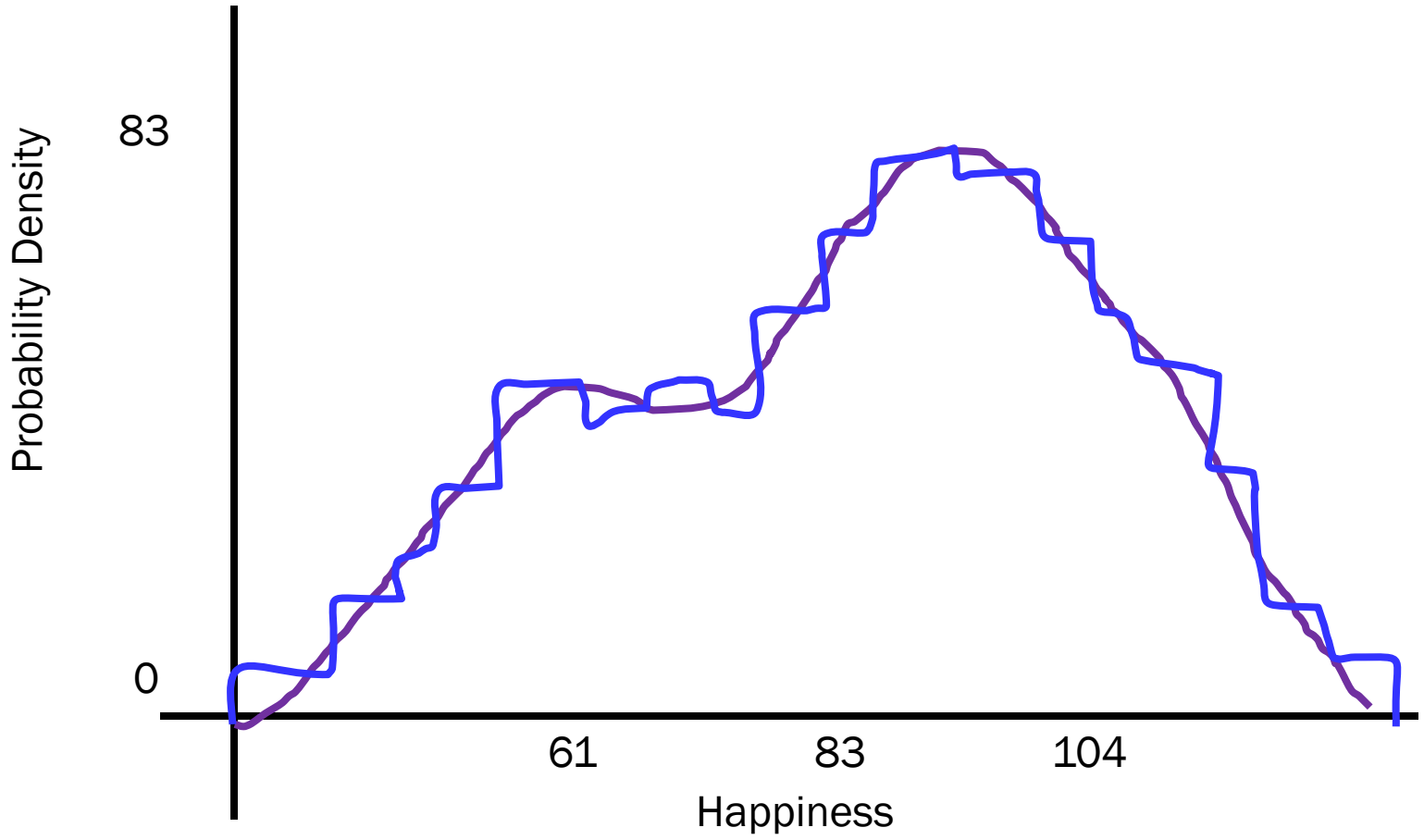
Hypothetical

What is the variance of the sample variance of subsamples of 200 people?



Key Insight

You can estimate the PMF of the underlying distribution, using your sample.*



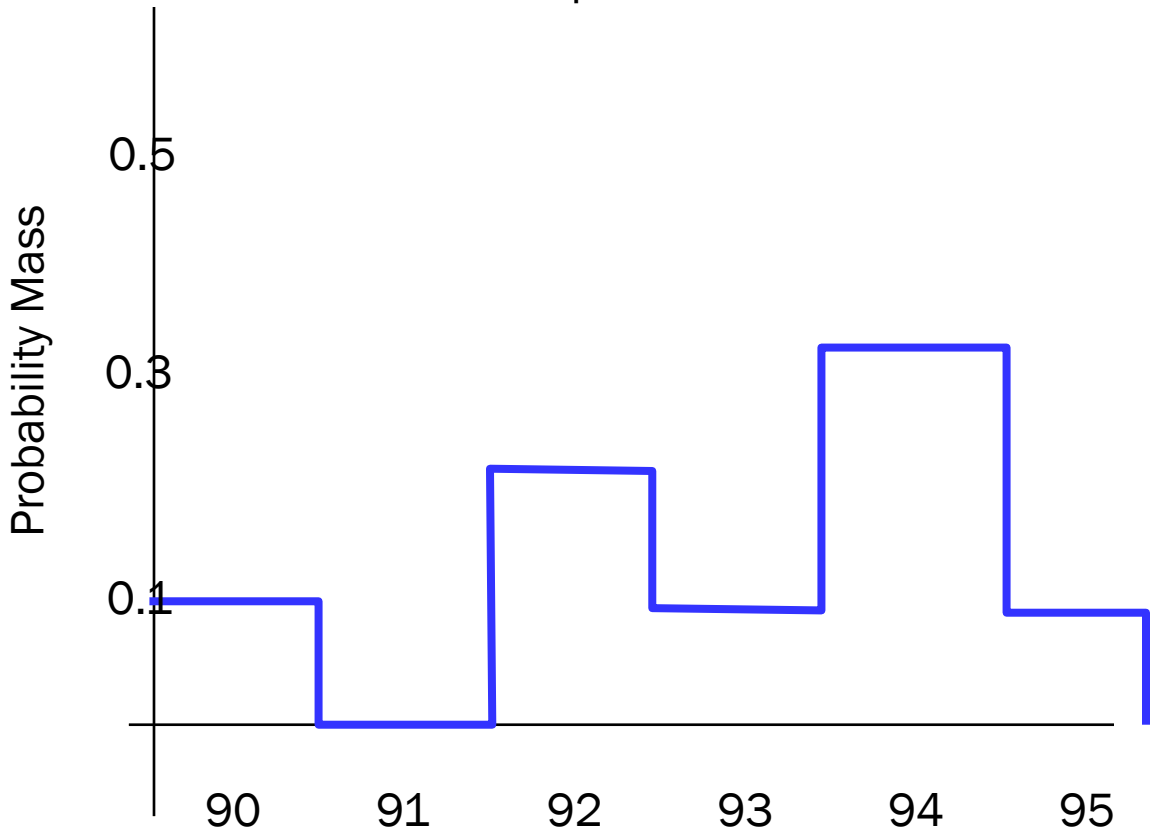
* This is just a histogram of your data!!

Key Insight

IID Samples

90,
92,
92,
93,
94,
94,
94,
95,

Sample Distribution



Bootstrapping Assumption

$$F \approx \hat{F}$$

The underlying distribution

The sample distribution

(aka the histogram of your data)

Algorithm

Bootstrap Algorithm (sample):

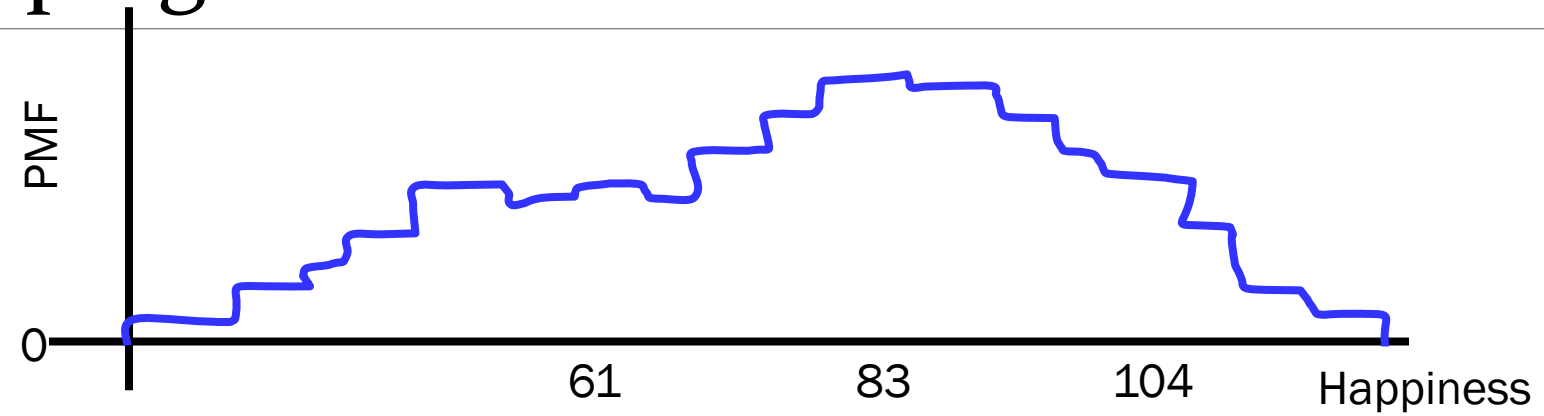
1. Estimate the **PMF** using the sample
2. Repeat **10,000** times:
   a. Resample **sample.size()** from PMF
   b. Recalculate the stat on the resample
3. You now have a **distribution of your stat**

Bootstrapping of Means (we could do this with CLT)

Bootstrap Algorithm (sample):

1. Estimate the **PMF** using the sample
2. Repeat **10,000** times:
   a. Draw **sample.size()** new samples from PMF
   b. **Recalculate the mean on the resample**
3. You now have a **distribution of your means**

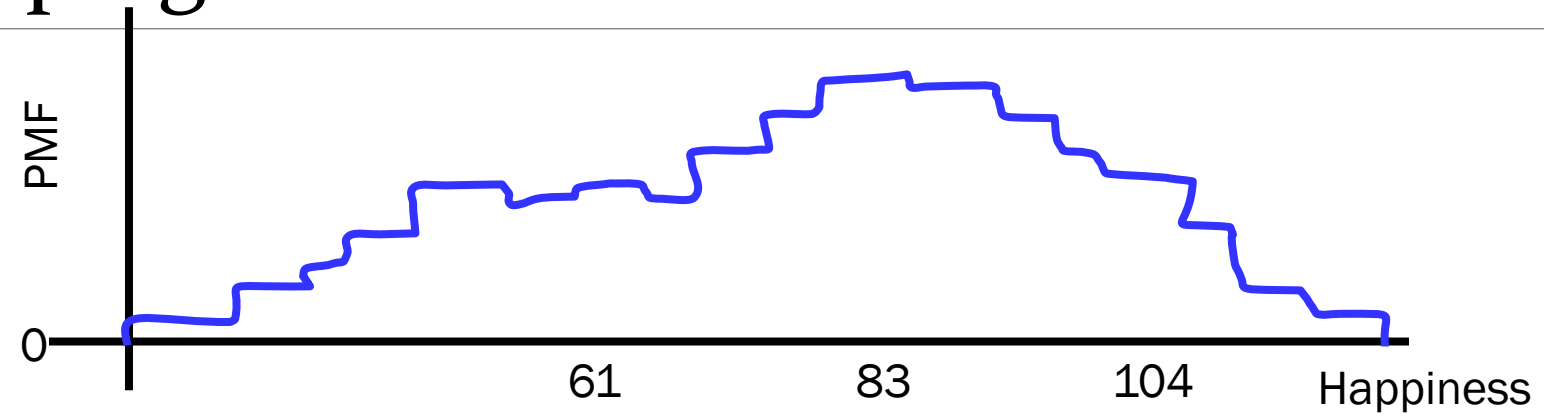
Bootstrapping of Means



Bootstrap Algorithm (sample):

1. Estimate the **PMF** using the sample
2. Repeat **10,000** times:
   a. Draw **sample.size()** new samples from PMF
   b. **Recalculate the mean on the resample**
3. You now have a **distribution of your means**

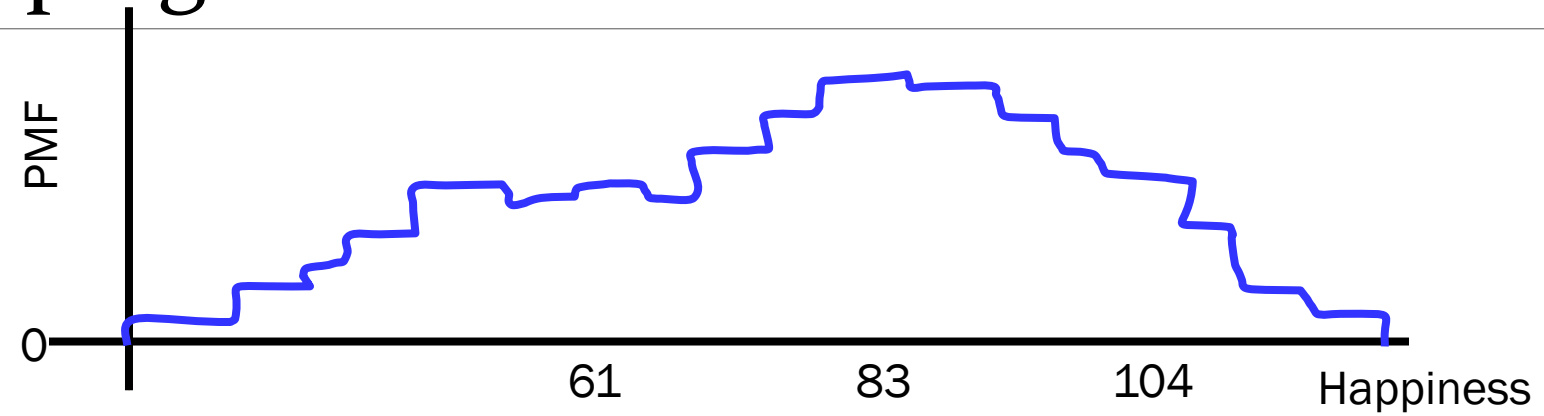
Bootstrapping of Means



Bootstrap Algorithm (sample):

1. Estimate the **PMF** using the sample
2. Repeat **10,000** times:
   a. Draw **sample.size()** new samples from PMF
   b. **Recalculate the mean on the resample**
3. You now have a **distribution of your means**

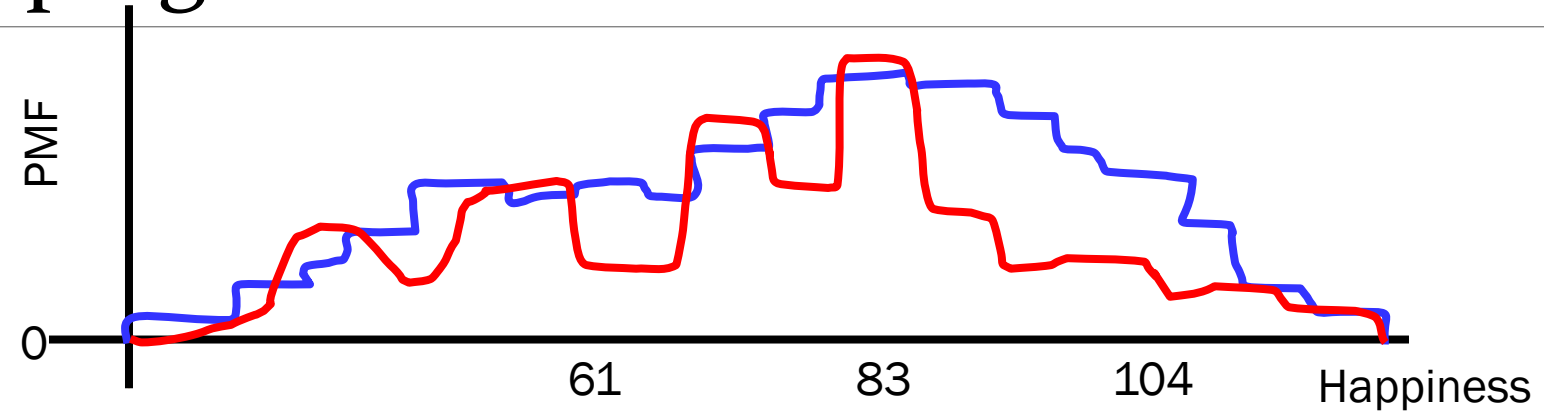
Bootstrapping of Means



Bootstrap Algorithm (sample):

1. Estimate the **PMF** using the sample
2. Repeat **10,000** times:
   a. Draw **sample.size()** new samples from PMF
   b. **Recalculate the mean on the resample**
3. You now have a **distribution of your means**

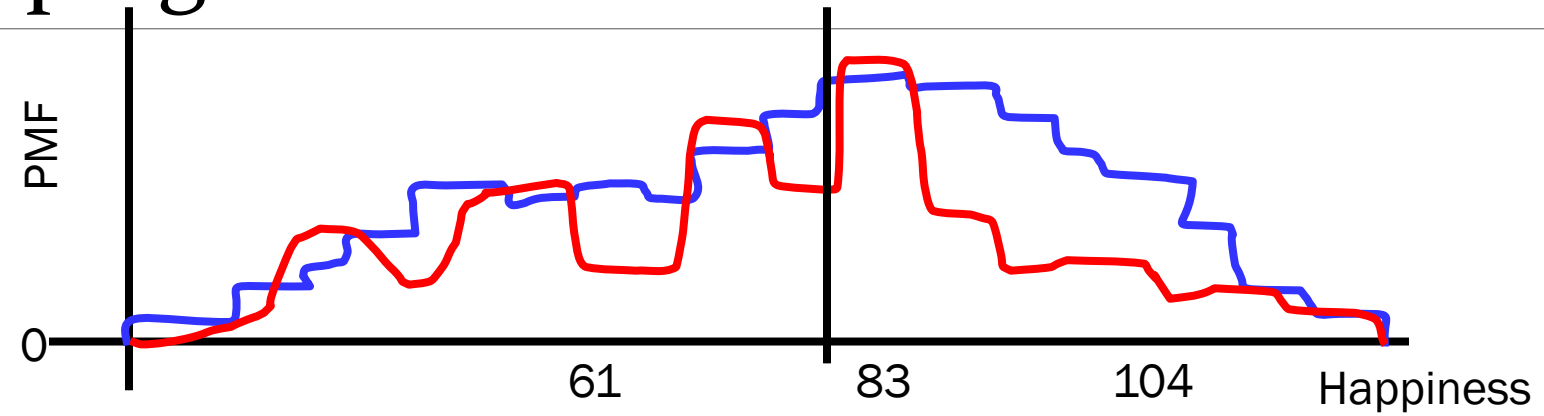
Bootstrapping of Means



Bootstrap Algorithm (sample):

1. Estimate the **PMF** using the sample
2. Repeat **10,000** times:
   a. Draw **sample.size()** new samples from PMF
   b. **Recalculate the mean on the resample**
3. You now have a **distribution of your means**

Bootstrapping of Means

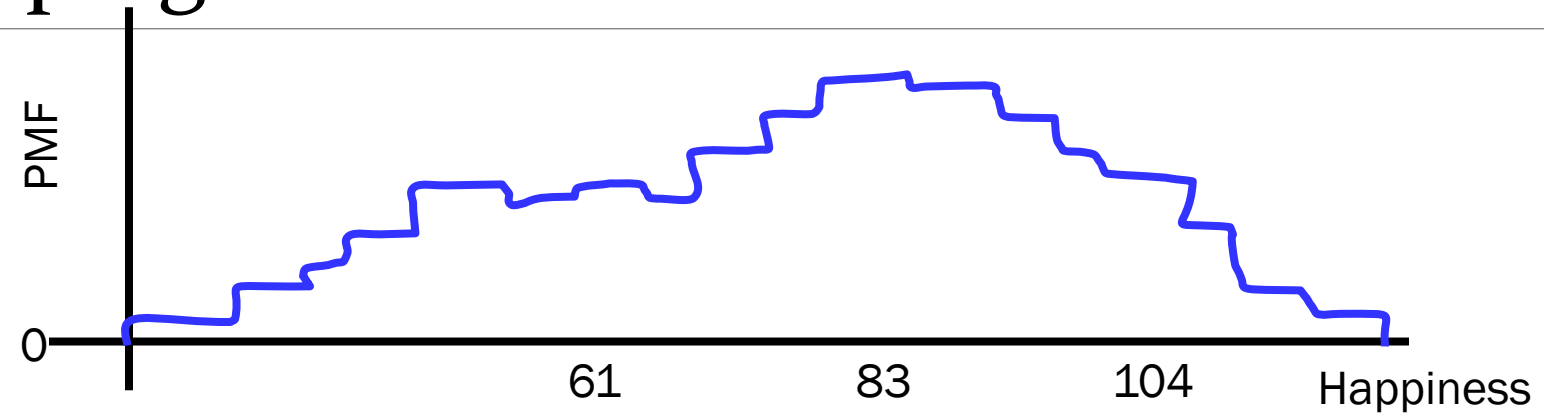


Bootstrap Algorithm (sample):

1. Estimate the **PMF** using the sample
2. Repeat **10,000** times:
   a. Draw **sample.size()** new samples from PMF
   b. **Recalculate the mean on the resample**
3. You now have a **distribution of your means**

Means = [82.7]

Bootstrapping of Means

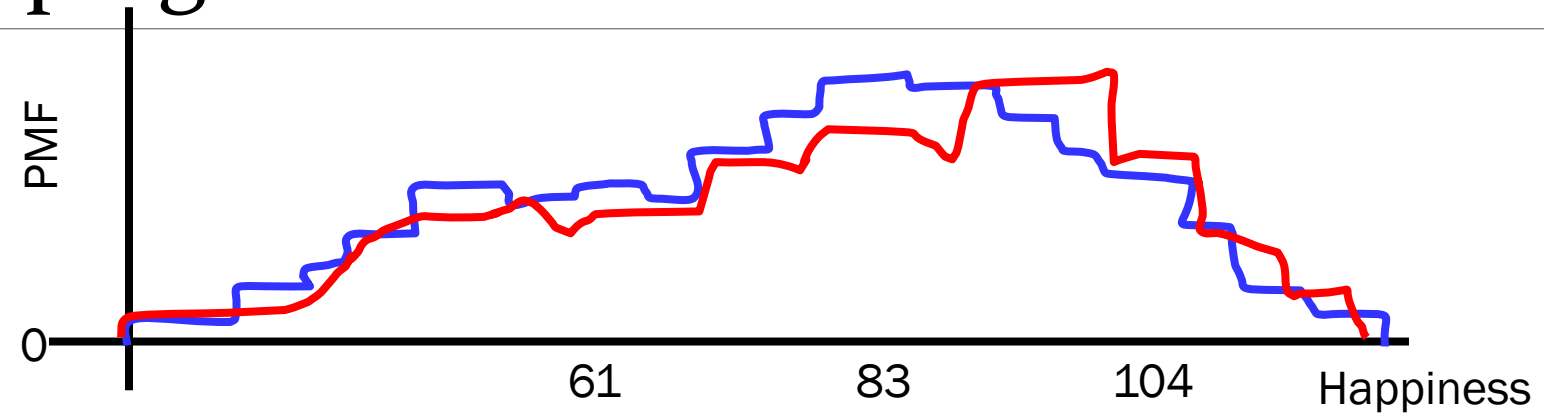


Bootstrap Algorithm (sample):

1. Estimate the **PMF** using the sample
2. Repeat **10,000** times:
   a. Draw **sample.size()** new samples from PMF
   b. **Recalculate the mean** on the resample
3. You now have a **distribution of your means**

Means = [82.7]

Bootstrapping of Means

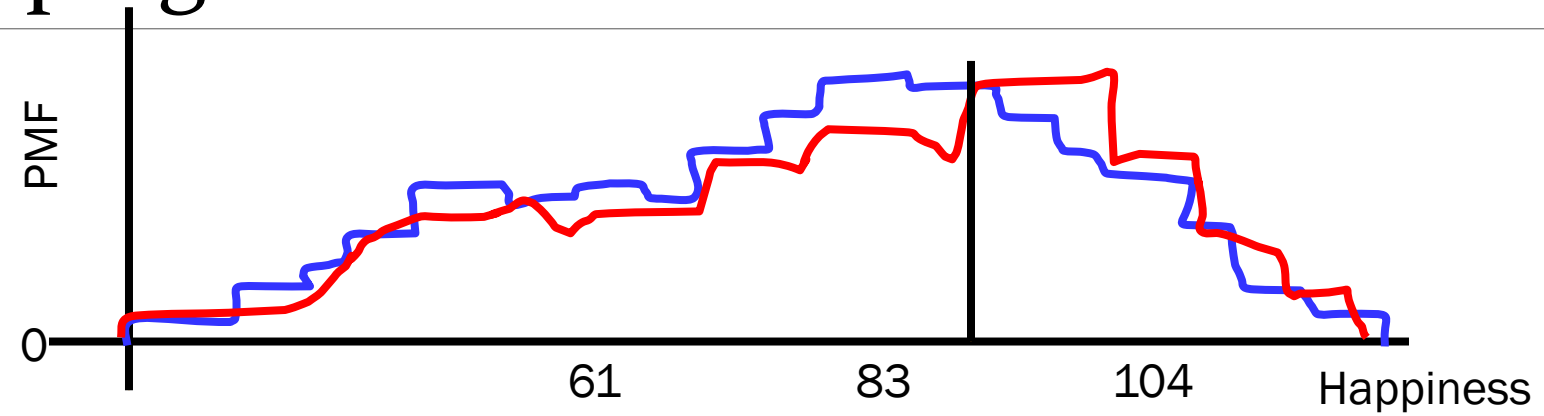


Bootstrap Algorithm (sample):

1. Estimate the **PMF** using the sample
2. Repeat **10,000** times:
   a. Draw **sample.size()** new samples from PMF
   b. **Recalculate the mean on the resample**
3. You now have a **distribution of your means**

Means = [82.7]

Bootstrapping of Means

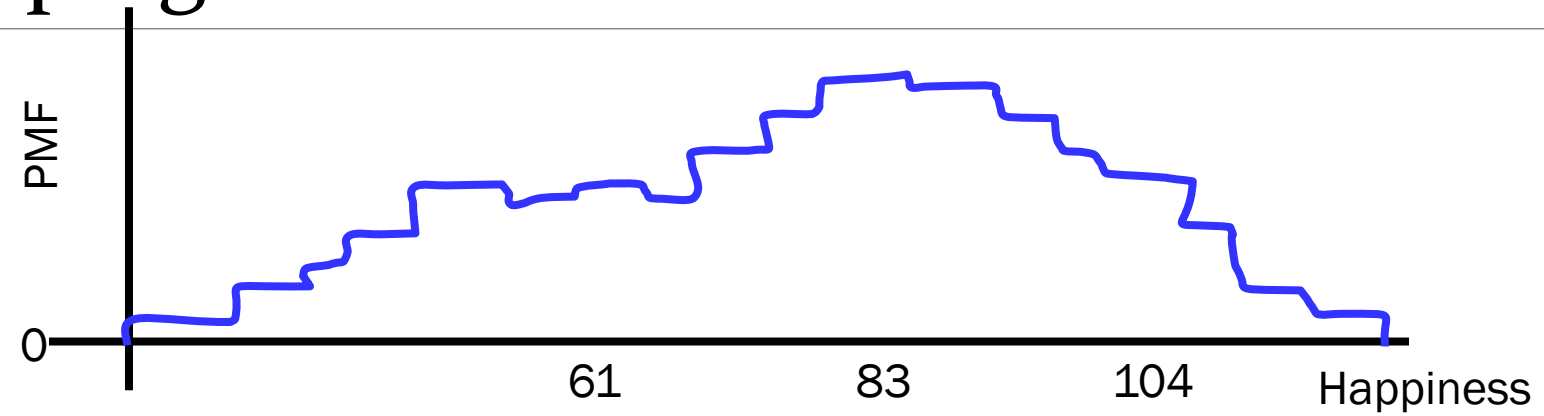


Bootstrap Algorithm (sample):

1. Estimate the **PMF** using the sample
2. Repeat **10,000** times:
   a. Draw **sample.size()** new samples from PMF
   b. **Recalculate the mean on the resample**
3. You now have a **distribution of your means**

Means = [82.7, 83.4]

Bootstrapping of Means

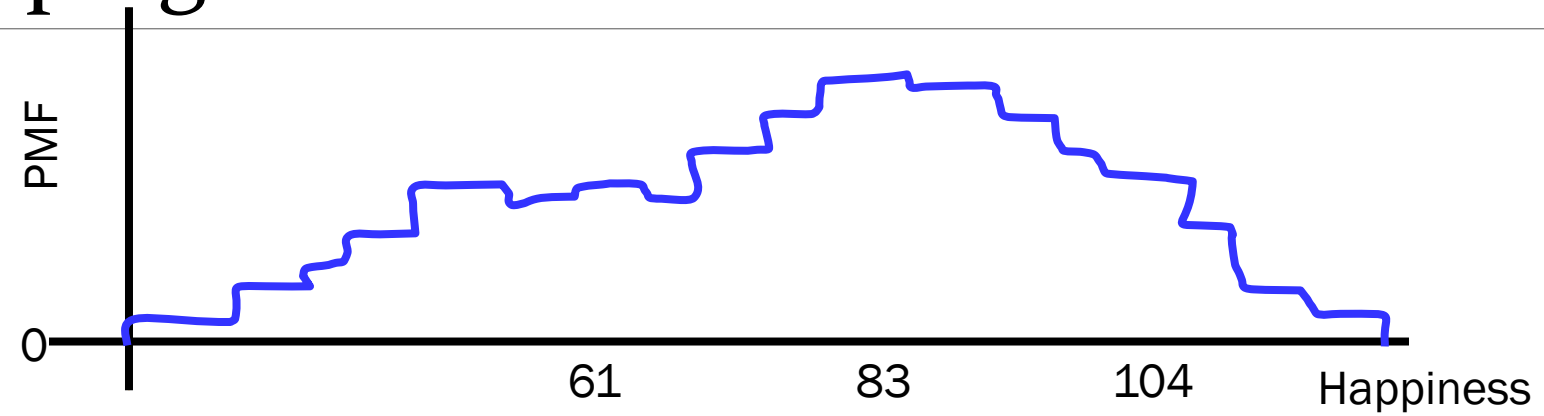


Bootstrap Algorithm (sample):

1. Estimate the **PMF** using the sample
2. Repeat **10,000** times:
   a. Draw **sample.size()** new samples from PMF
   b. **Recalculate the mean on the resample**
3. You now have a **distribution of your means**

Means = [82.7, 83.4]

Bootstrapping of Means



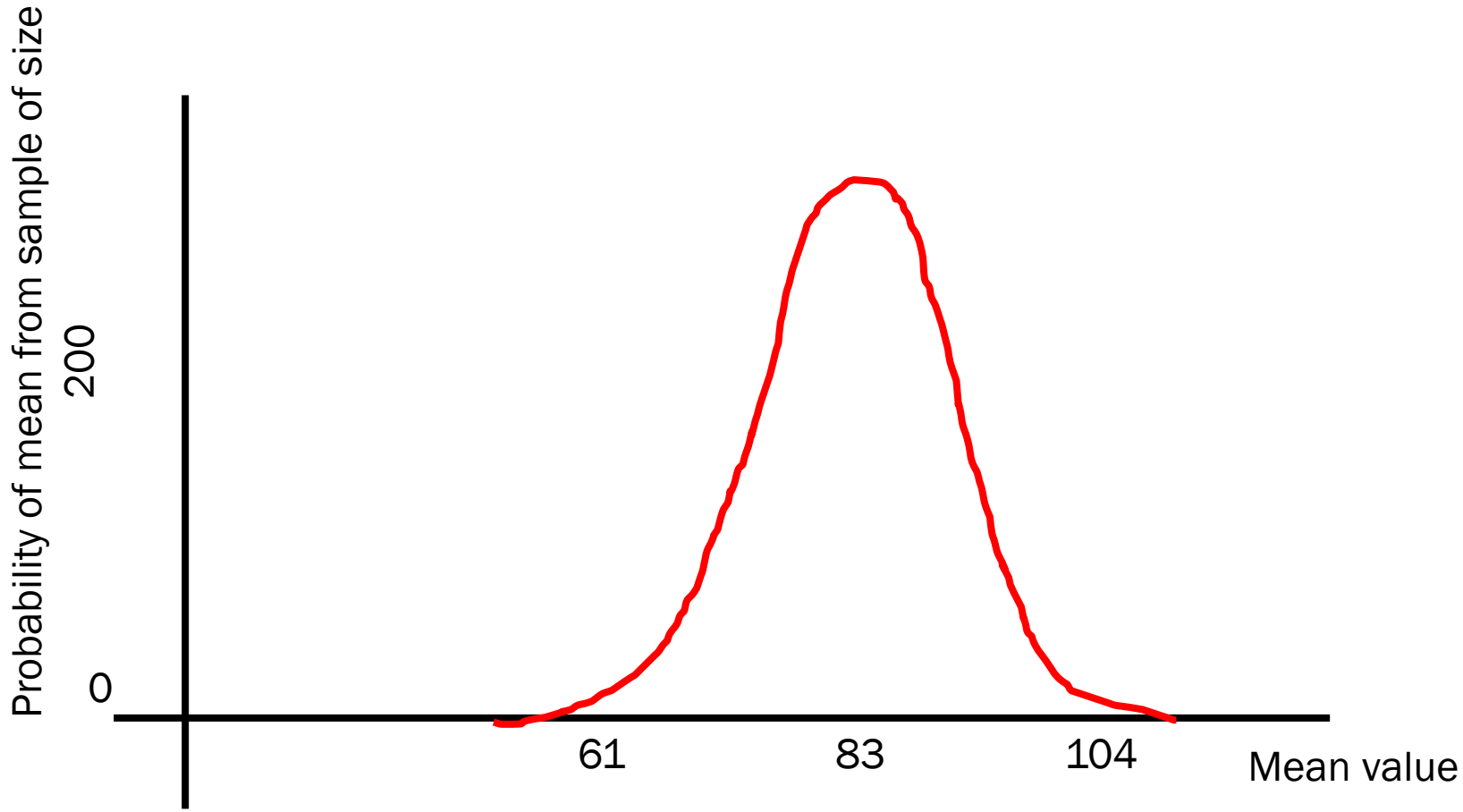
Bootstrap Algorithm (sample):

1. Estimate the **PMF** using the sample
2. Repeat **10,000** times:
   a. Draw **sample.size()** new samples from PMF
   b. **Recalculate the mean** on the resample
3. You now have a **distribution of your means**

Means = [82.7, 83.4, 82.9, 91.4, 79.3, 82.1, ..., 81.7]

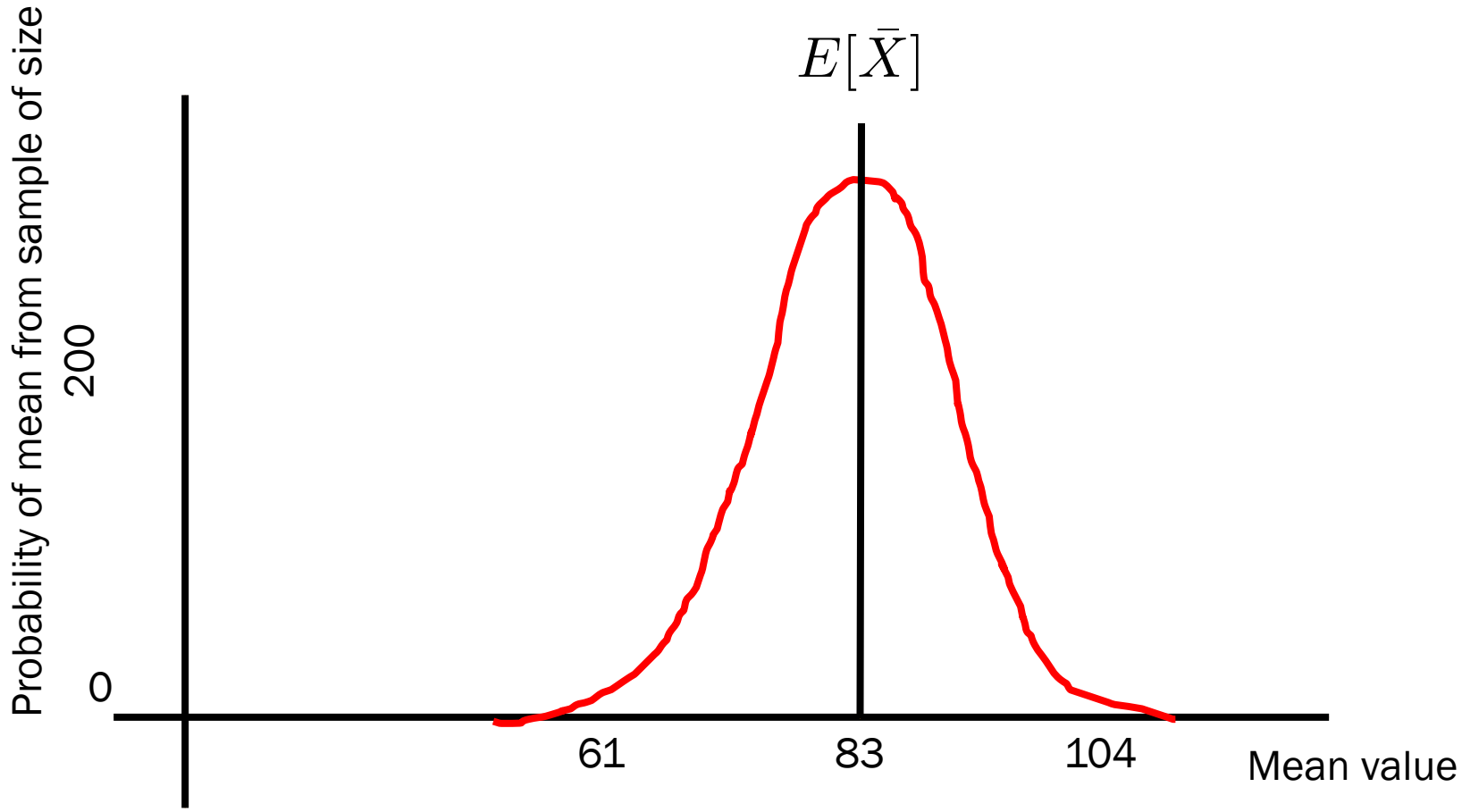
Bootstrapping of Means

Means = [82.7, 83.4, 82.9, 91.4, 79.3, 82.1, ..., 81.7]



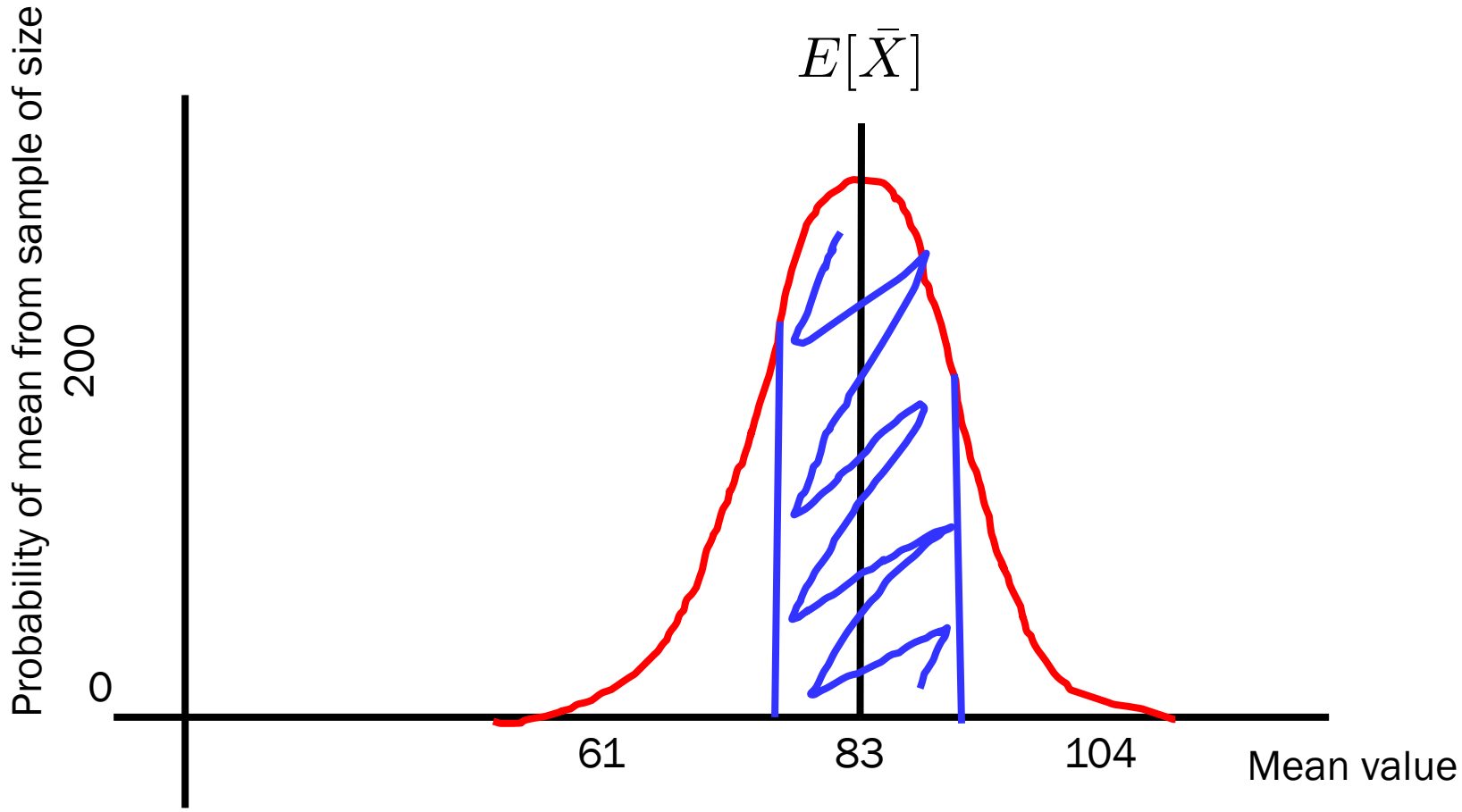
Bootstrapping of Means

Means = [82.7, 83.4, 82.9, 91.4, 79.3, 82.1, ..., 81.7]

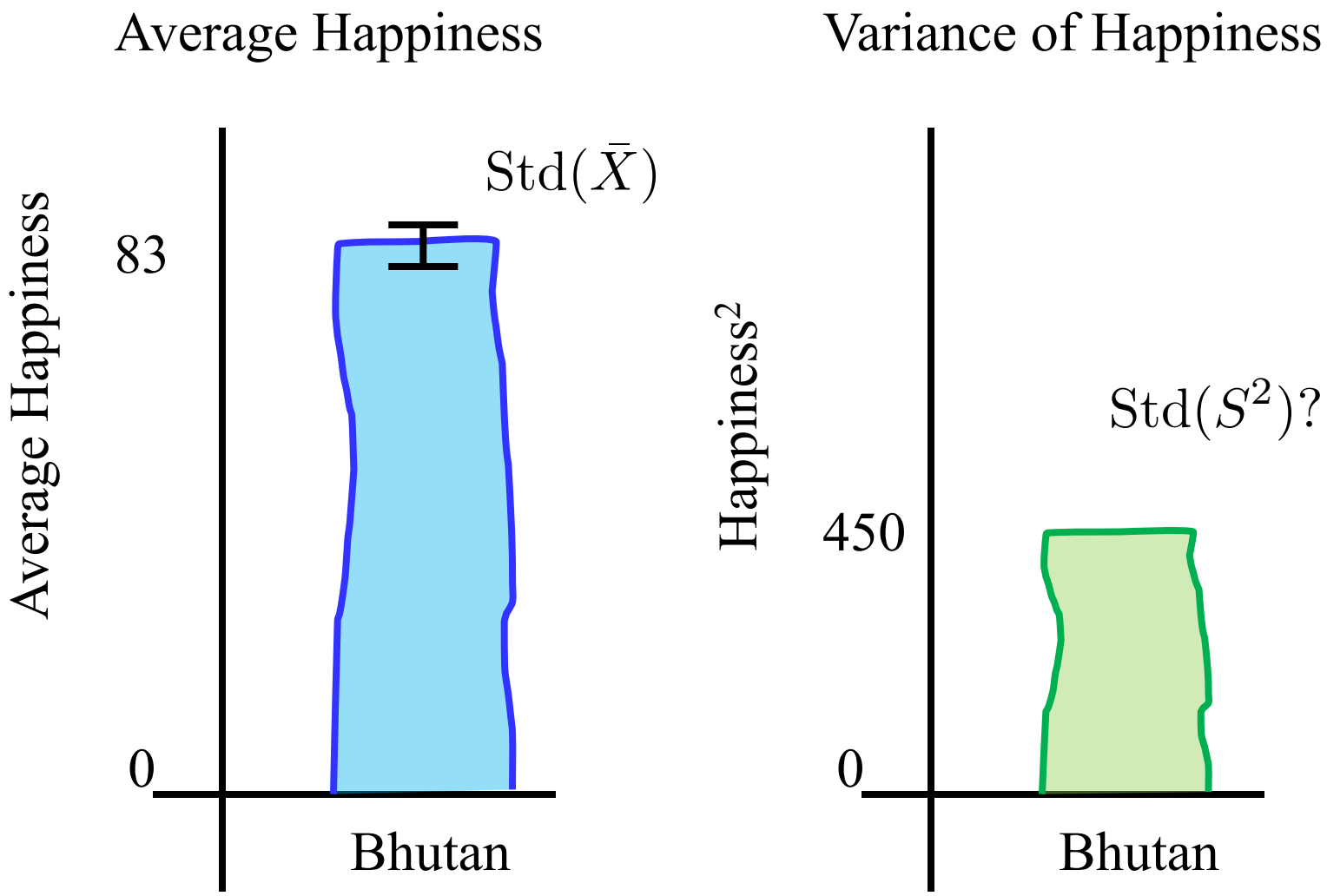


Bootstrapping of Means

What is the probability that the mean is in the range 81 to 85?



But What about Standard Deviation of Variance Estimate?



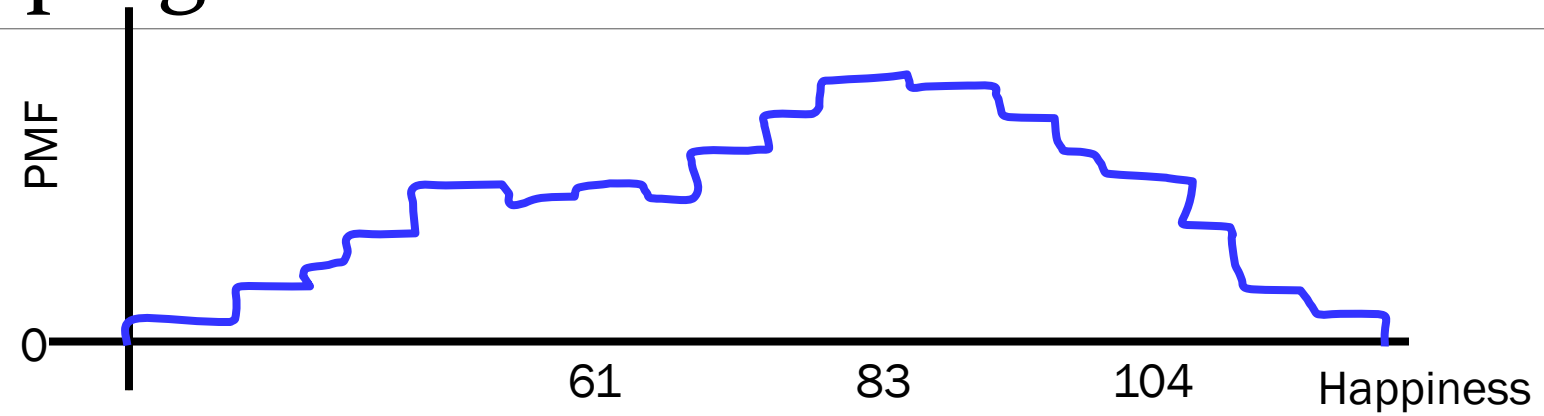
Claim: The average happiness of Bhutan is 83 ± 2

Bootstrapping of Variance

Bootstrap Algorithm (sample):

1. Estimate the **PMF** using the sample
2. Repeat **10,000** times:
   a. Draw **sample.size()** new samples from PMF
   b. Recalculate the variance on the resample
3. You have a **distribution of your variances**

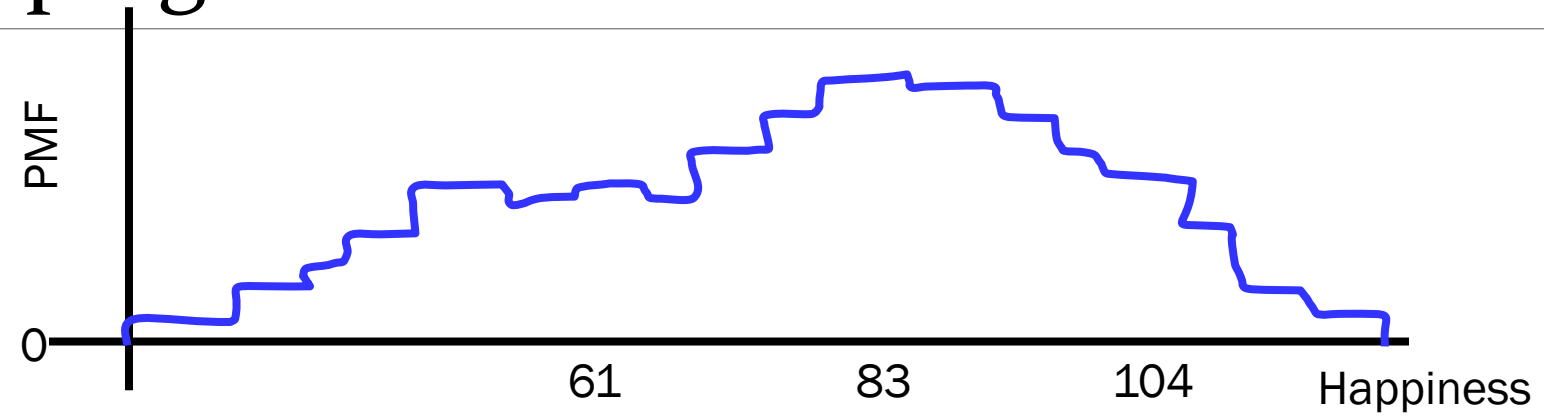
Bootstrapping of Variance



Bootstrap Algorithm (sample):

1. Estimate the **PMF** using the sample
2. Repeat **10,000** times:
   a. Draw **sample.size()** new samples from PMF
   b. **Recalculate the var on the resample**
3. You now have a **distribution of your vars**

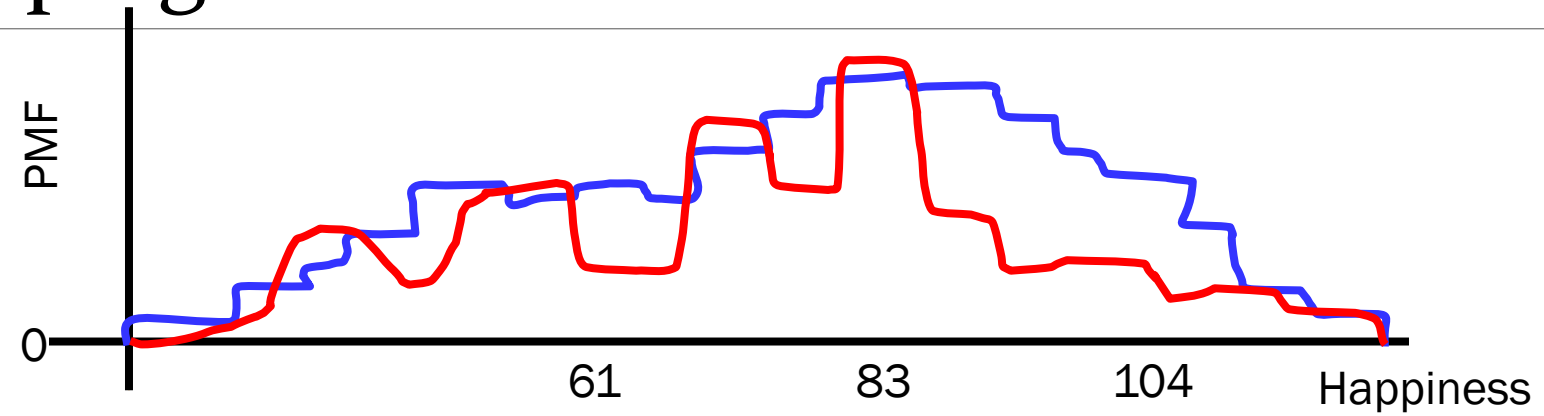
Bootstrapping of Variance



Bootstrap Algorithm (sample):

1. Estimate the **PMF** using the sample
2. Repeat **10,000** times:
   a. Draw **sample.size()** new samples from PMF
   b. **Recalculate the var on the resample**
3. You now have a **distribution of your vars**

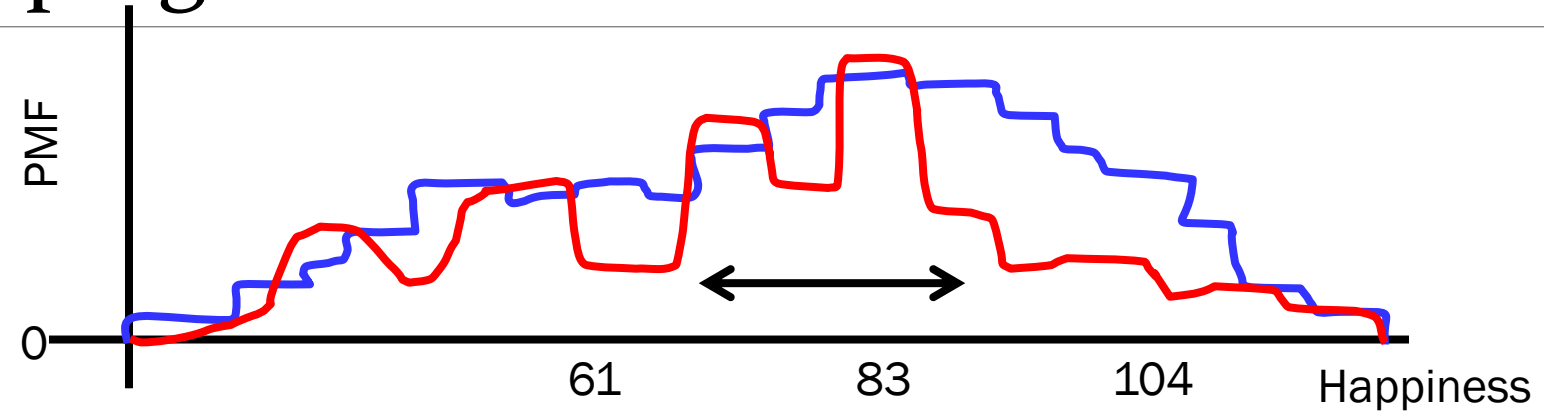
Bootstrapping of Variance



Bootstrap Algorithm (sample):

1. Estimate the **PMF** using the sample
2. Repeat **10,000** times:
   a. Draw **sample.size()** new samples from PMF
   b. **Recalculate the var** on the resample
3. You now have a **distribution of your vars**

Bootstrapping of Variance

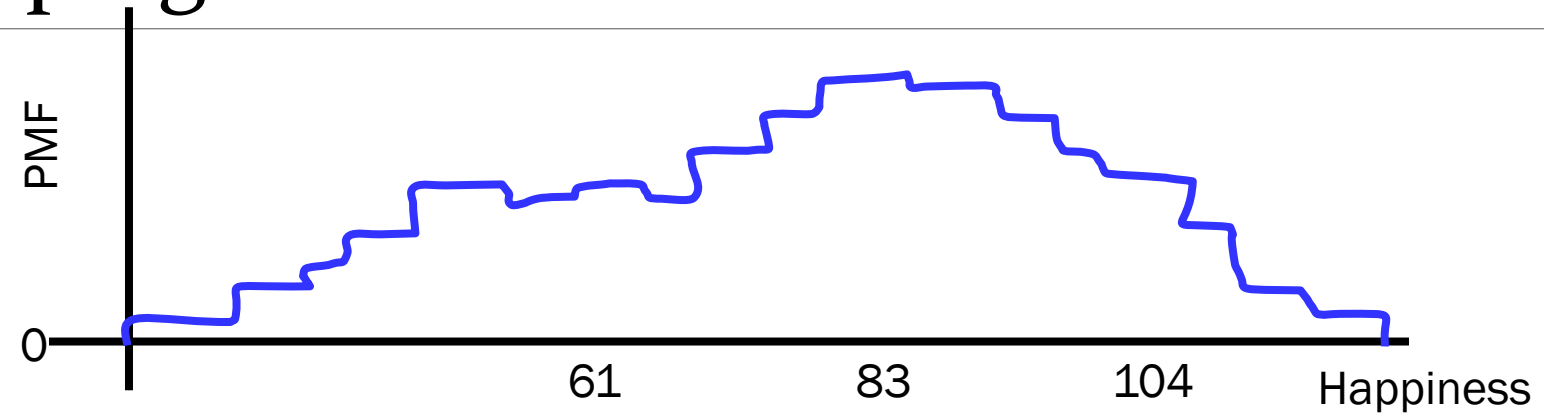


Bootstrap Algorithm (sample):

1. Estimate the **PMF** using the sample
2. Repeat **10,000** times:
   a. Draw **sample.size()** new samples from PMF
   b. **Recalculate the vars on the resample**
3. You now have a **distribution of your vars**

Vars = [472.7]

Bootstrapping of Variance

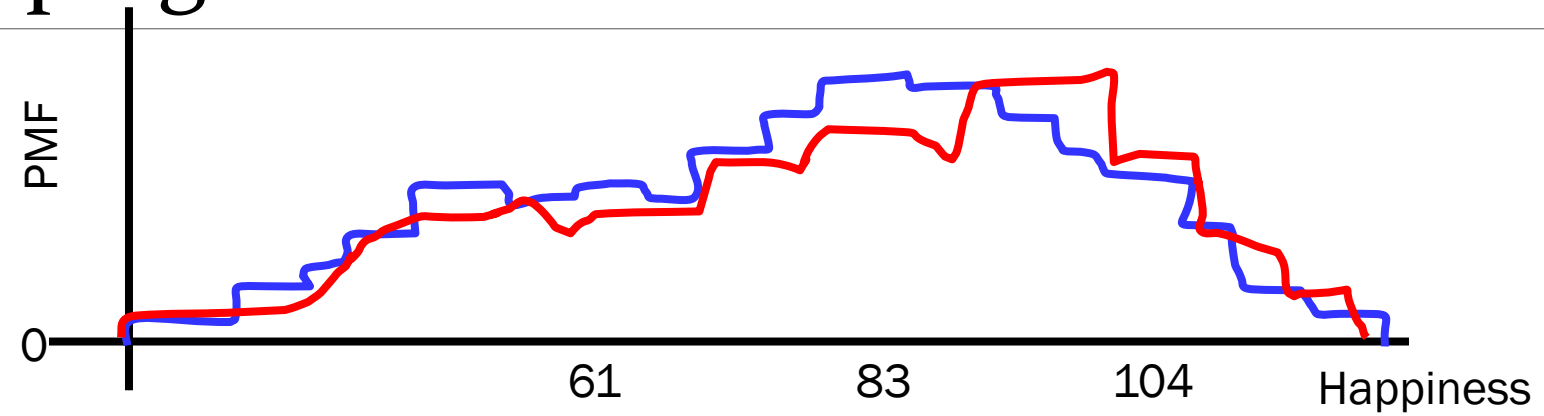


Bootstrap Algorithm (sample):

1. Estimate the **PMF** using the sample
2. Repeat **10,000** times:
   a. Draw **sample.size()** new samples from PMF
   b. **Recalculate the var on the resample**
3. You now have a **distribution of your vars**

Vars = [472.7]

Bootstrapping of Variance

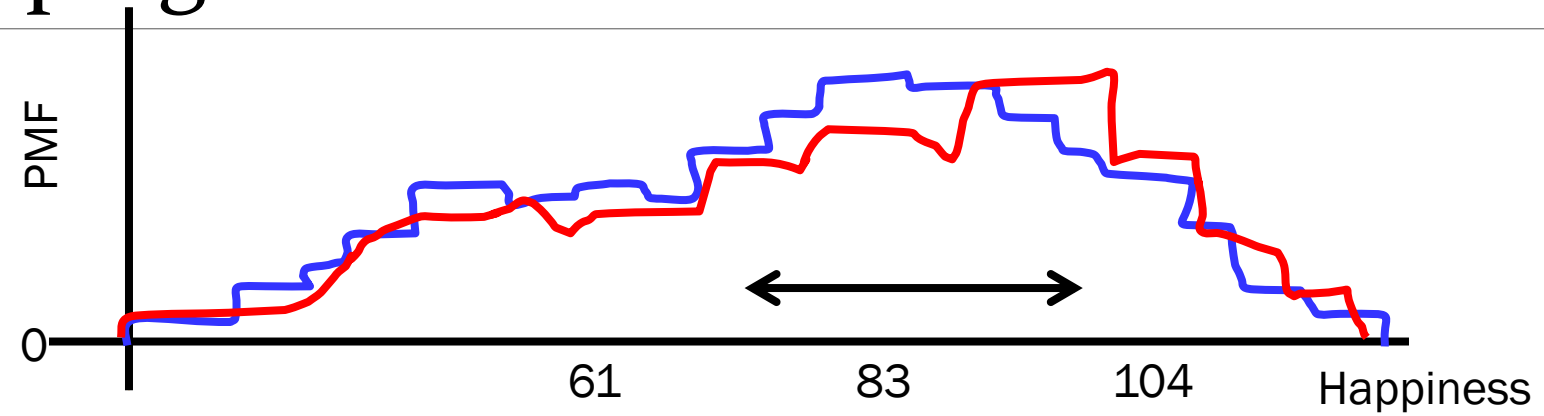


Bootstrap Algorithm (sample):

1. Estimate the **PMF** using the sample
2. Repeat **10,000** times:
   a. Draw **sample.size()** new samples from PMF
   b. **Recalculate the var on the resample**
3. You now have a **distribution of your vars**

Vars = [472.7]

Bootstrapping of Variance

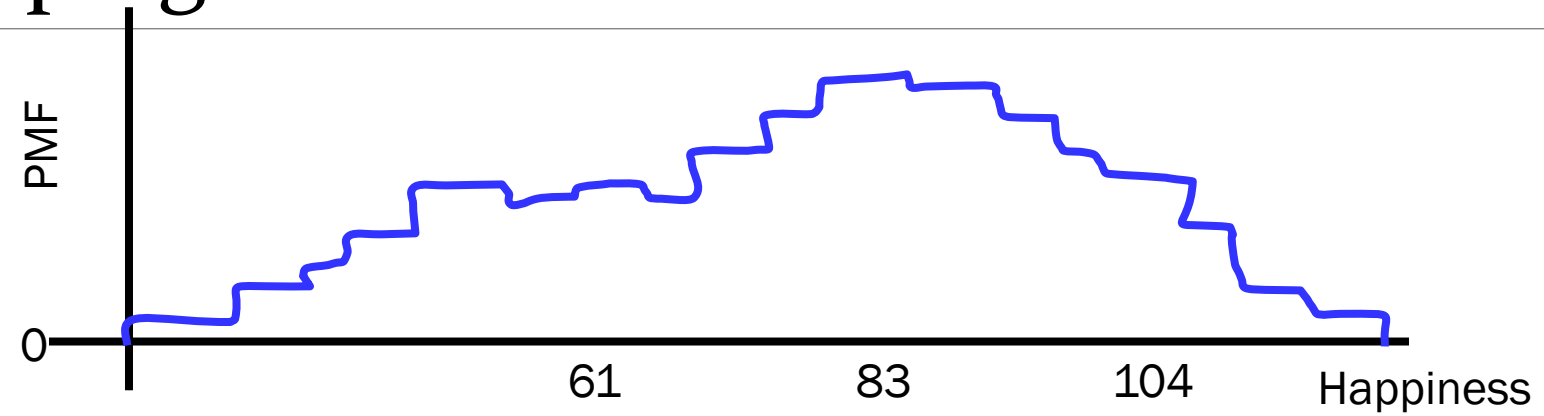


Bootstrap Algorithm (sample):

1. Estimate the **PMF** using the sample
2. Repeat **10,000** times:
   a. Draw **sample.size()** new samples from PMF
   b. **Recalculate the var on the resample**
3. You now have a **distribution of your vars**

Vars = [472.7, 478.4]

Bootstrapping of Variance

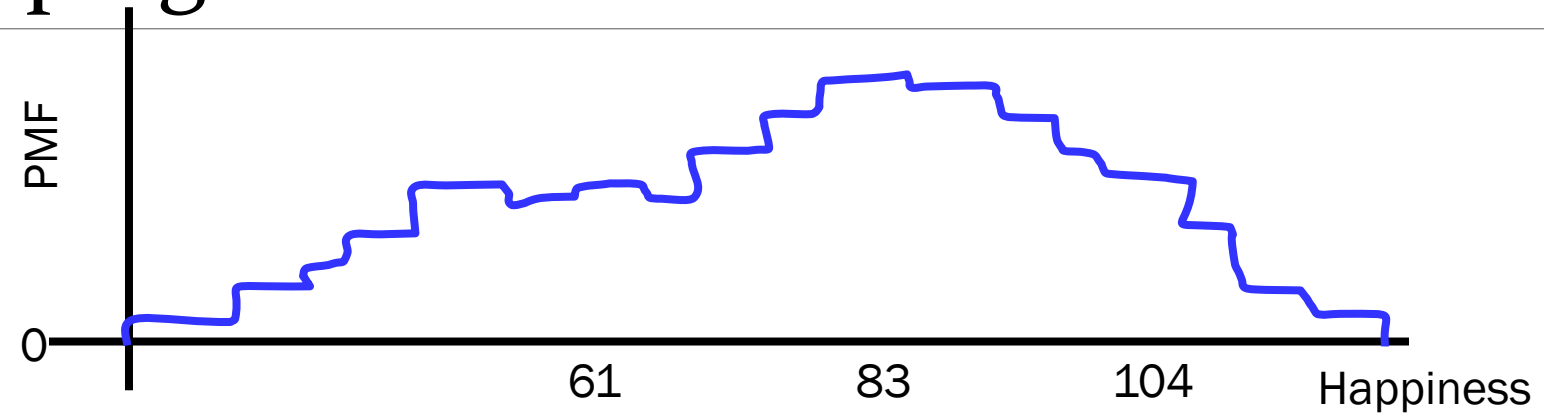


Bootstrap Algorithm (sample):

1. Estimate the **PMF** using the sample
2. Repeat **10,000** times:
   a. Draw **sample.size()** new samples from PMF
   b. **Recalculate the var on the resample**
3. You now have a **distribution of your vars**

Vars = [472.7, 478.4]

Bootstrapping of Variance



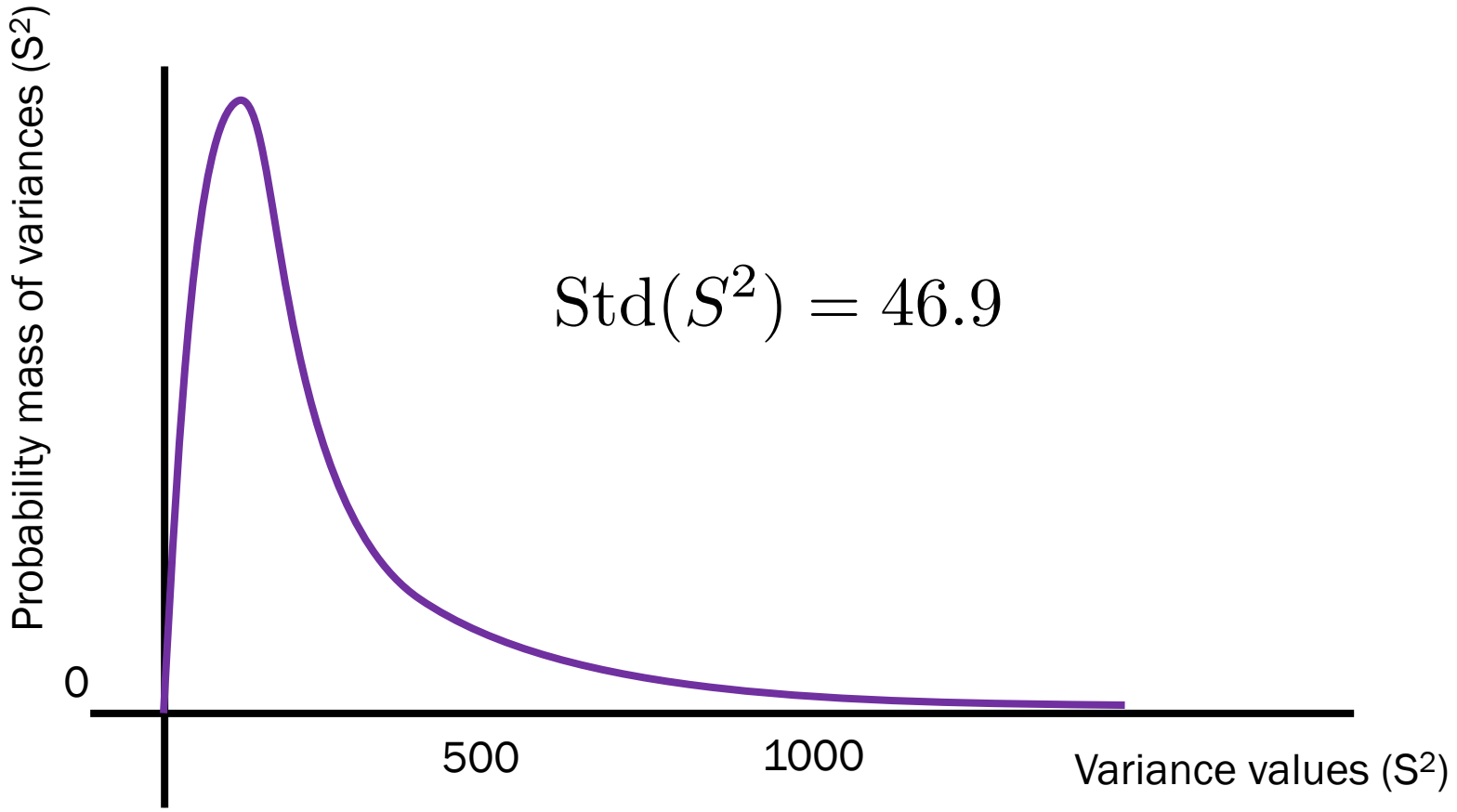
Bootstrap Algorithm (sample):

1. Estimate the **PMF** using the sample
2. Repeat **10,000** times:
   a. Draw **sample.size()** new samples from PMF
   b. **Recalculate the var on the resample**
3. You now have a **distribution of your vars**

Vars = [472.7, 478.4, 469.2, ..., 476.2]

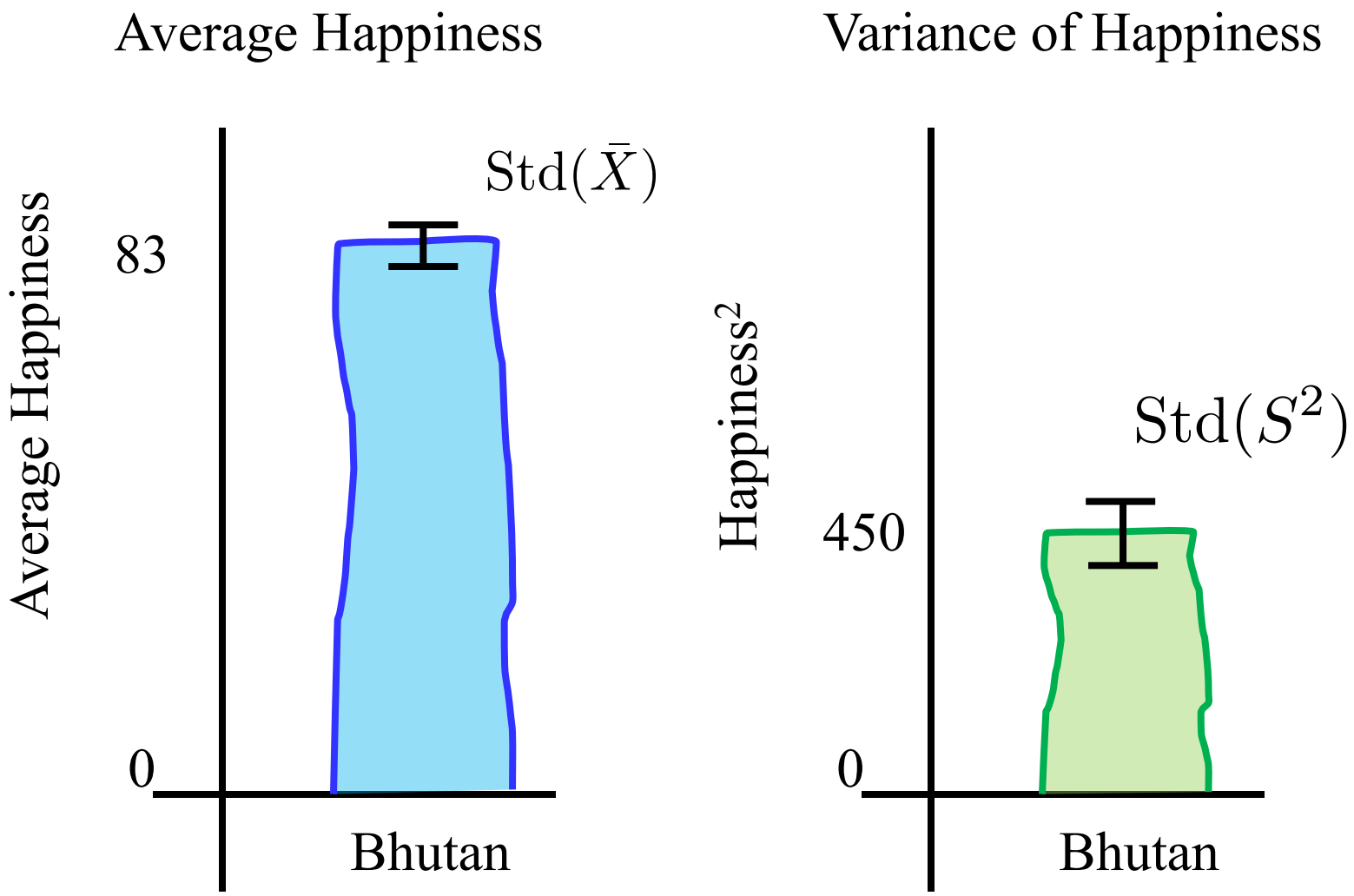
Bootstrapping of Variance

Sample Vars = [472.7, 478.4, 469.2, ..., 476.2]



$$\text{Std}(S^2) = 46.9$$

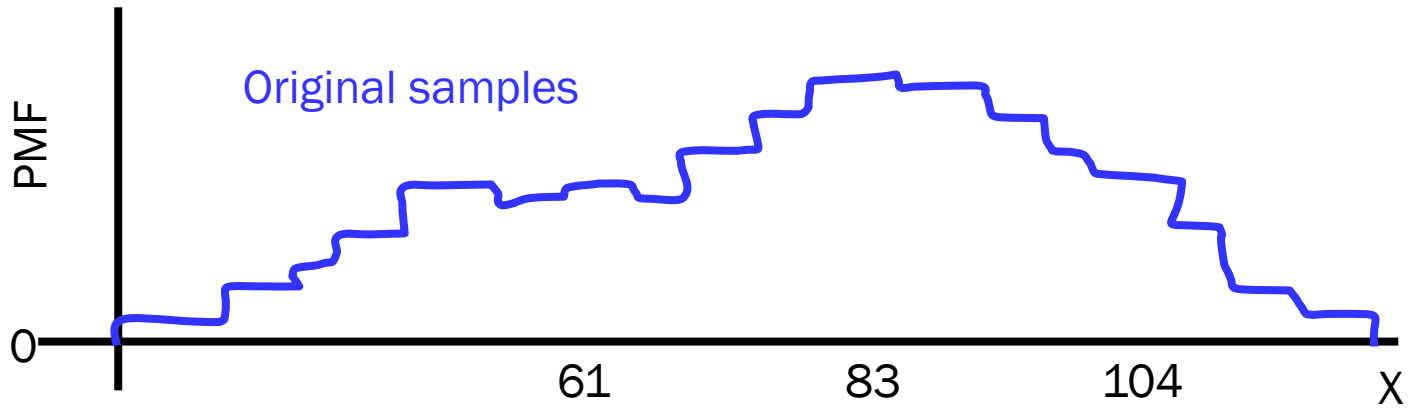
Bootstrapping of Variance



Claim: The average happiness of Bhutan is 83 ± 2

Bootstrapping in Practice

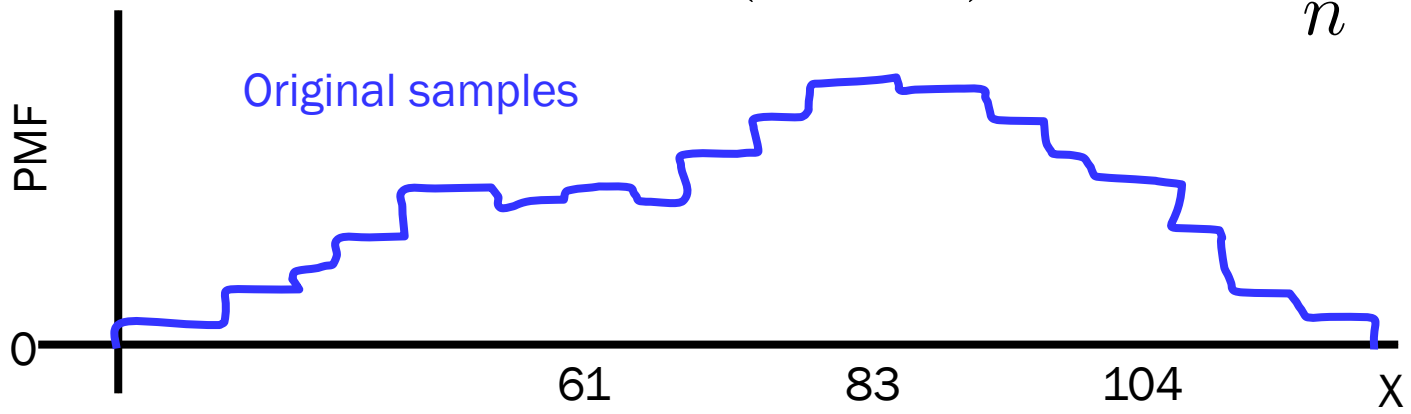
```
def resample(samples):
    # Estimate the PMF using the samples
    # Draw K new samples from the PMF
```



Bootstrapping in Practice

```
def resample(samples):
    # Estimate the PMF using the samples
    # Draw K new samples from the PMF
    return np.random.choice(samples, K,
                            replace = True)
```

$$P(X = k) = \frac{\text{count}(X = k)}{n}$$



OG Bootstrapping

Bootstrap Algorithm (sample):

1. Estimate the **PMF** using the sample
2. Repeat **10,000** times:
   a. Resample **sample.size()** from PMF
   b. Recalculate the stat on the resample
3. You now have a **distribution of your stat**

Bootstrapping in Practice

Bootstrap Algorithm (sample):

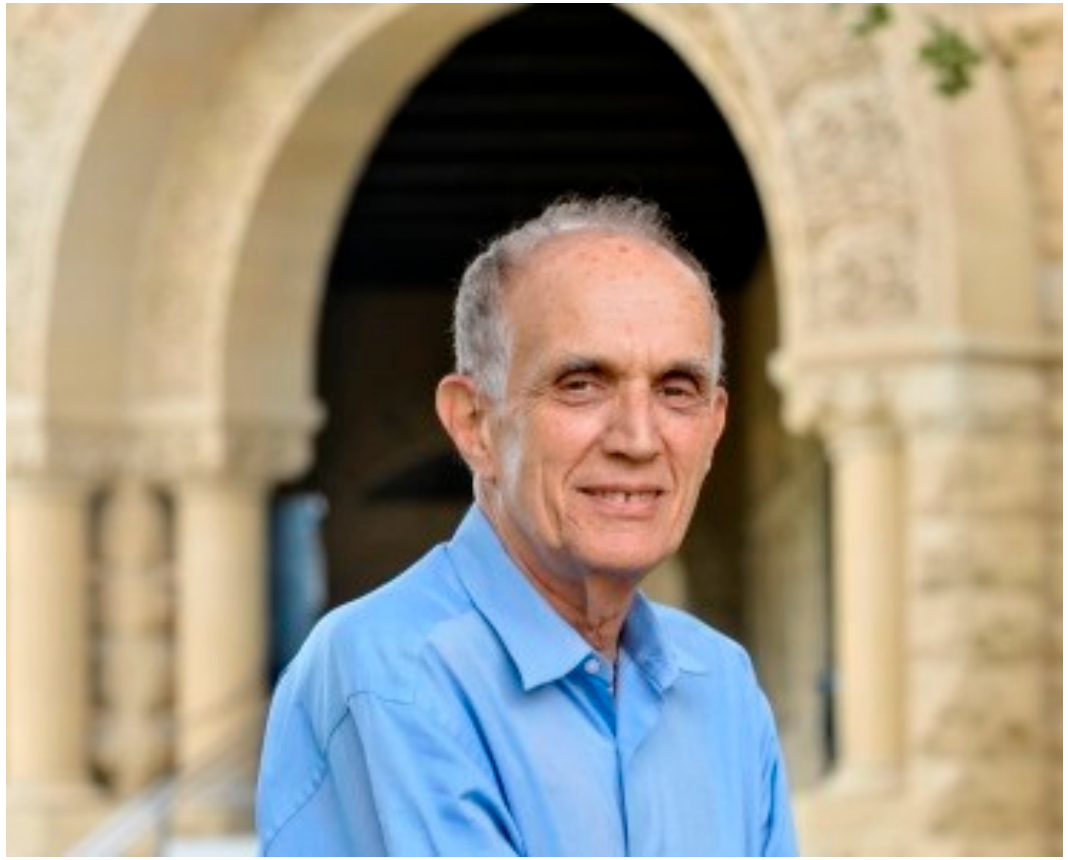
1. Repeat **10,000** times:
   a. **Choose sample.size elems from sample, with replacement**
   b. Recalculate the stat on the resample
2. You now have a **distribution of your stat**

To the code!

Bootstrap provides a way to calculate probabilities of statistics using code.


Bootstrap

Bradley Efron



Invented bootstrapping in 1979

Still a professor at Stanford

Won a National Science Medal

Works for any statistic*

*as long as your samples are IID and the underlying distribution doesn’t have a long tail

The Classic Science Test

Group 1	Group 2
4.44	2.15
3.36	3.01
5.87	2.02
2.31	1.43
...	...
3.70	1.83
$\mu_1 = 3.1$	$\mu_2 = 2.4$

Claim: Group 1 and Group 2 are samples from different distributions with a 0.7 difference of means.

How confident are you in this claim?

A real difference?

Learning in Context A	Learning in Context B
4.44	2.15
3.36	3.01
5.87	2.02
2.31	1.43
...	...
3.70	1.83

18 students

23 students

$\mu_1 = 3.1$

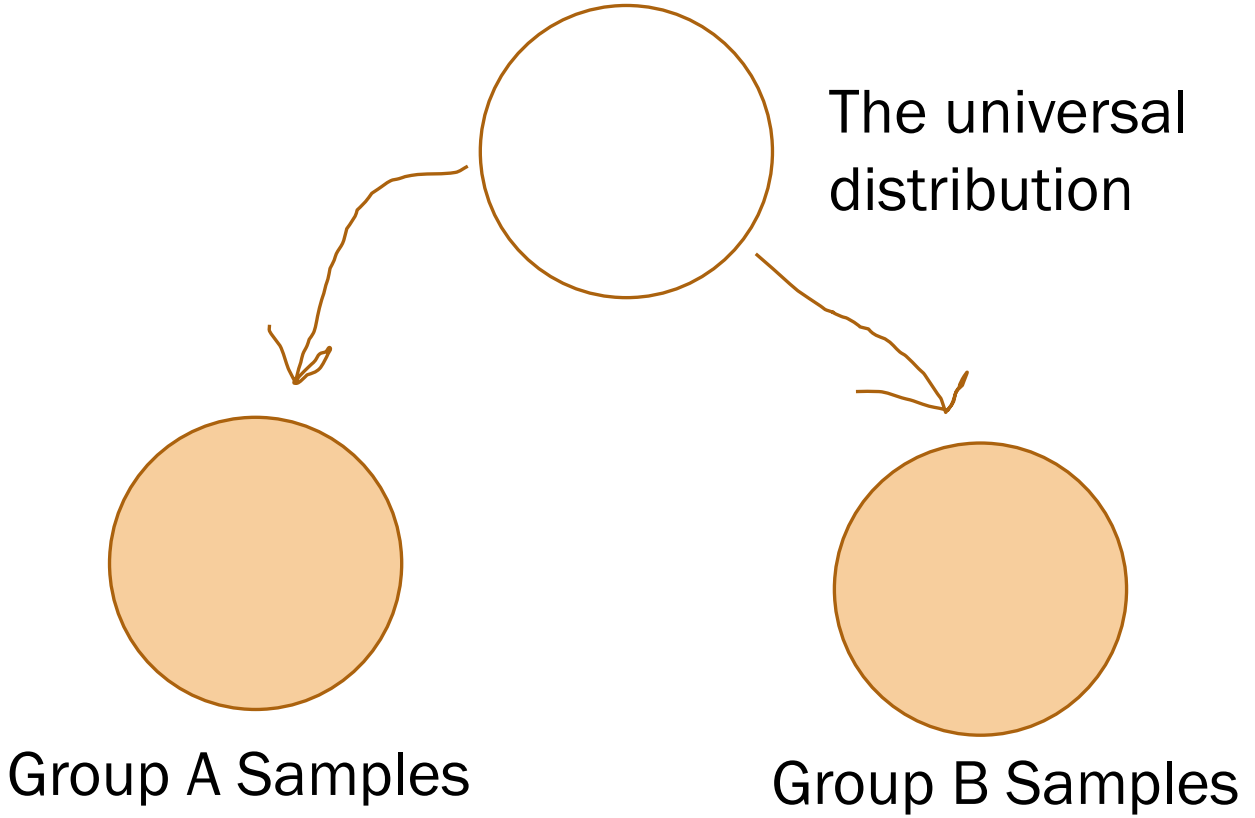
$\mu_2 = 2.4$

Claim: Group 1 and Group 2 are samples from different distributions with a 0.7 difference of means.

How confident are you in this claim?

The Null Hypothesis

There is no difference between the two groups, so everyone is drawn from the same distribution. Any difference you observe is due to sampling error.



To the code!