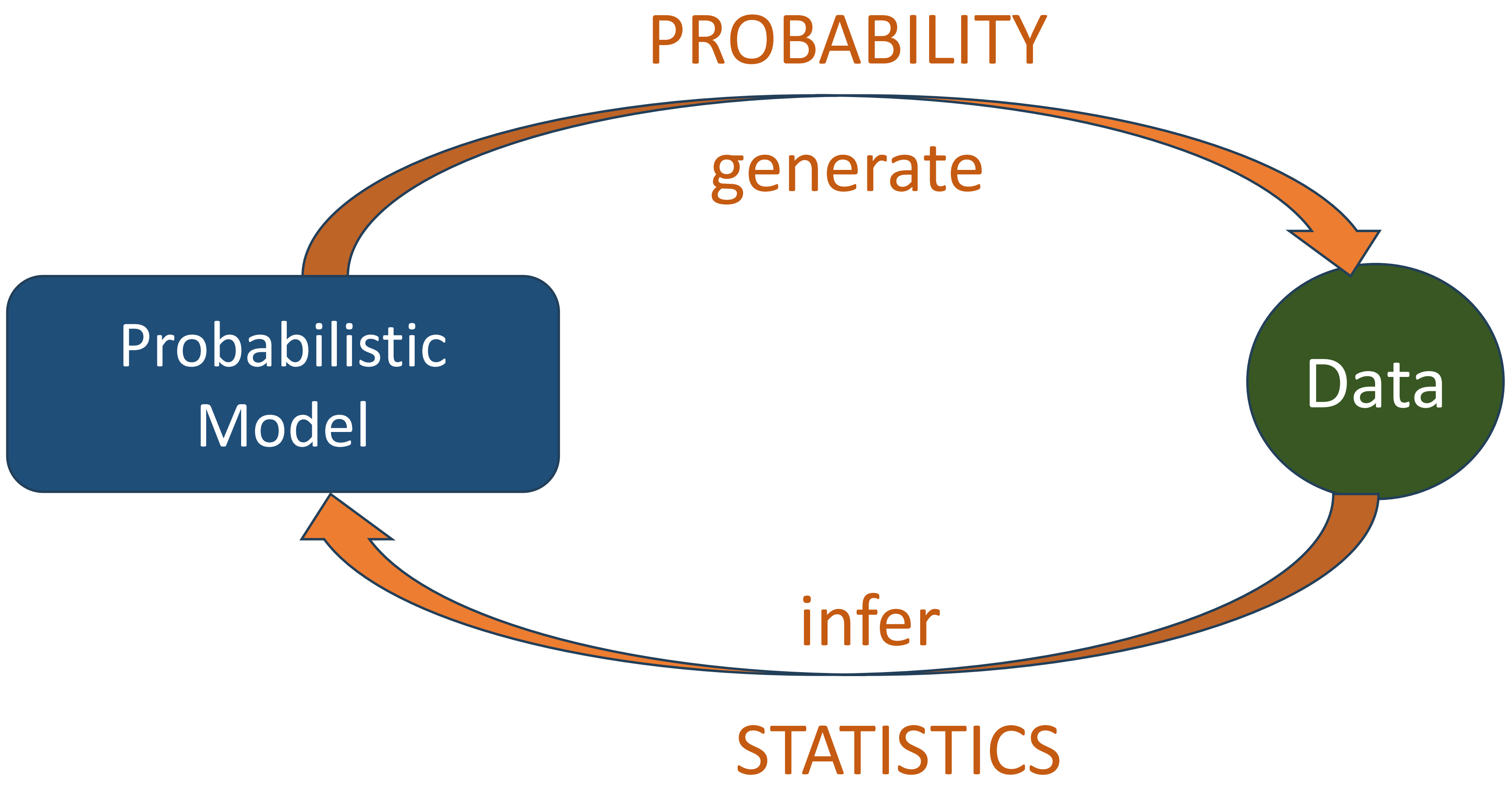
Chapter 15
Estimators for Mean & Variance

Probability vs. Statistics



<u>This chapter</u>: Given some data, infer the mean and variance of the underlying distribution.

Motivation for Estimation

Call this unknown θ

Goal: Estimate the mean height of women

Obviously can't measure heights of all women, so sample heights of n women, say $n = 5$.

Use **sample average** as estimation of θ.

Definition of Estimator

e.g. $\theta = \boldsymbol{E}[X]$, where X is unknown

θ: quantity we're trying to estimate

think of these as i.i.d. instances of X

$X_1, X_2, \dots, X_n$: sampled data

This is a r.v. because it's a function of r.v.s

Sometimes just write $\hat{\theta}$

$\hat{\theta}(X_1, X_2, \dots, X_n)$: estimator of the unknown θ

This is a constant

$\hat{\theta}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$: estimation of θ based on specific instantiation of the data

Sample Mean Estimator

Defn 15.2: Let $X_1, X_2, \ldots, X_n$ be i.i.d. samples of r.v. X with unknown mean $\theta = \boldsymbol{E}[X]$.

The **sample mean** is a point estimator of θ:

$$\hat{\theta}(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \mathrm{M_n} = \bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Alternate name for sample mean

Alternate name for sample mean

Choices of estimators

Let $X_1, X_2, \dots, X_n$ be i.i.d. samples of r.v. X with unknown finite mean $\theta = \boldsymbol{E}[X]$ and unknown finite variance $\sigma^2 = \boldsymbol{Var}(X)$.

Two possible estimators for θ:

$$\hat{\theta}_A = \bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

$$\hat{\theta}_B = X_2$$

Q: How do we choose between estimators?

Desirable properties of a point estimator

<u>Defn 15.3</u>: Let $\hat{\theta}(X_1, X_2, \ldots, X_n)$ be a point estimator for θ.
The **bias** of $\hat{\theta}$ is:

$$\boldsymbol{B}(\hat{\theta}) = \boldsymbol{E}[\hat{\theta}] - \theta$$

If $\boldsymbol{B}(\hat{\theta}) = 0$, we say that $\hat{\theta}$ is and **unbiased estimator** of θ.

unbiased estimator is desirable property #1

Which of $\hat{\theta}_A = \bar{X}$ and $\hat{\theta}_B = X_2$ are unbiased estimators?

Both!

So why do we prefer $\hat{\theta}_A$?

Desirable properties of a point estimator

Defn 15.4: The **mean squared error (MSE)** of estimator $\hat{\theta}(X_1, X_2, \ldots, X_n)$ is:

$$\boldsymbol{MSE}(\hat{\theta}) = \boldsymbol{E}\left[(\hat{\theta} - \theta)^2\right]$$

Low MSE is desirable property #2

What metric does this remind you of?

Desirable properties of a point estimator

Defn 15.4: The **mean squared error (MSE)** of estimator $\hat{\theta}(X_1, X_2, \ldots, X_n)$ is:

$$\boldsymbol{MSE}(\hat{\theta}) = \boldsymbol{E}\left[(\hat{\theta} - \theta)^2\right]$$

Low MSE is desirable property #2

Lemma 15.5: If $\hat{\theta}(X_1, X_2, \ldots, X_n)$ is an unbiased estimator, then

$$\boldsymbol{MSE}(\hat{\theta}) = \boldsymbol{Var}(\hat{\theta})$$

Proof: $\boldsymbol{MSE}(\hat{\theta}) = \boldsymbol{E}\left[(\hat{\theta} - \theta)^2\right] = \boldsymbol{E}\left[(\hat{\theta} - \boldsymbol{E}[\hat{\theta}])^2\right] = \boldsymbol{Var}(\hat{\theta})$

Desirable properties of a point estimator

$$MSE(\hat{\theta}) = E\left[(\hat{\theta} - \theta)^2\right] = Var(\hat{\theta})$$

if $\hat{\theta}(X_1, X_2, \ldots, X_n)$ is unbiased

Q: Which of $\hat{\theta}_A = \bar{X}$ and $\hat{\theta}_B = X_2$ have lower MSE?

Both $\hat{\theta}_A = \bar{X}$ and $\hat{\theta}_B = X_2$ are unbiased estimators.

$$\Rightarrow MSE(\hat{\theta}_A) = Var(\hat{\theta}_A) = \frac{1}{n^2} \cdot n\, Var(X) = \frac{Var(X)}{n}$$

$$\Rightarrow MSE(\hat{\theta}_B) = Var(\hat{\theta}_B) = Var(X_2) = Var(X)$$

$\hat{\theta}_A$ is way better!

Desirable properties of a point estimator

We also want our estimator to become more accurate as the sample size increases.

Defn 15.6: Let $\hat{\theta}_1(X_1), \hat{\theta}_2(X_1, X_2), \hat{\theta}_3(X_1, X_2, X_3), \ldots$ be a sequence of point estimators of θ, where $\hat{\theta}_n(X_1, X_2, \ldots, X_n)$ is a function of n i.i.d. samples.

We say that r.v. $\hat{\theta}_n$ is a **consistent estimator** of θ if, $\forall \epsilon > 0$,

$$\lim_{n\to\infty} \boldsymbol{P}\left\{\left|\hat{\theta}_n - \theta\right| \geq \epsilon\right\} = 0$$

consistency is desirable property #3

Seems hard to prove that our estimator is consistent ….

Fortunately, there's a helpful lemma!

Desirable properties of a point estimator

Lemma 15.7: If $\lim_{n\to\infty} \boldsymbol{MSE}(\hat{\theta}_n) = 0,$ then $\hat{\theta}_n$ is a consistent estimator.

Proof:

$$\boldsymbol{P}\{|\hat{\theta}_n - \theta| \geq \epsilon\} = \boldsymbol{P}\left\{|\hat{\theta}_n - \theta|^2 \geq \epsilon^2\right\} \leq \frac{\boldsymbol{E}\left[|\hat{\theta}_n - \theta|^2\right]}{\epsilon^2}$$

$$\leq \frac{\boldsymbol{E}\left[(\hat{\theta}_n - \theta)^2\right]}{\epsilon^2} = \frac{\boldsymbol{MSE}(\hat{\theta}_n)}{\epsilon^2}$$

$$\Rightarrow \lim_{n\to\infty} \boldsymbol{P}\{|\hat{\theta}_n - \theta| \geq \epsilon\} = \lim_{n\to\infty} \frac{\boldsymbol{MSE}(\hat{\theta}_n)}{\epsilon^2} = 0$$

Desirable properties of a point estimator

Lemma 15.7: If $\lim_{n\to\infty} \boldsymbol{MSE}(\hat{\theta}_n) = 0$, then $\hat{\theta}_n$ is a consistent estimator.

Q: Is $\hat{\theta}_A = \bar{X} = M_n$ a consistent estimator of $\boldsymbol{E}[X]$?

Yes!

$$\boldsymbol{MSE}(\hat{\theta}_A) = \boldsymbol{Var}(\hat{\theta}_A) = \frac{\boldsymbol{Var}(X)}{n}$$

Assuming $Var(X)$ finite

$$\Rightarrow \lim_{n\to\infty} \boldsymbol{MSE}(\hat{\theta}_A) = \lim_{n\to\infty} \boldsymbol{Var}(\hat{\theta}_A) = \lim_{n\to\infty} \frac{\boldsymbol{Var}(X)}{n} = 0$$

$\Rightarrow$ $\hat{\theta}_A$ is a consistent estimator

An estimator for Variance

Call this unknown θ

Goal: Estimate the variance of height of women

Obviously can't measure heights of all women, so sample heights of n women, say $n = 5$.

Q: Can we use sample variance as an estimation of θ?

Q: And how do we define sample variance ?

An estimator for Variance

Two cases:

1. The case where *know* $\boldsymbol{E}[X]$ and want to estimate $\boldsymbol{Var}(X)$

2. The case where *don't know* $\boldsymbol{E}[X]$ and want to estimate $\boldsymbol{Var}(X)$

Estimating Variance when Mean is Known

Assume $\mu = \boldsymbol{E}[X]$ is known.

Let $\hat{\theta}$ be an estimator of $\theta = \boldsymbol{Var}(X)$:

$$\hat{\theta}(X_1, X_2, \ldots, X_n) = \overline{S^2} \equiv \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

mean squared distance from mean

sample variance

Estimating Variance when Mean is Known

Assume $\mu = \boldsymbol{E}[X]$ is known.

Let $\hat{\theta}$ be an estimator of $\theta = \boldsymbol{Var}(X)$:

$$\hat{\theta}(X_1, X_2, \ldots, X_n) = \overline{S^2} \equiv \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

Is $\overline{S^2}$ an unbiased estimator?

$\boldsymbol{Yes}$!

$$\boldsymbol{E}\left[\overline{S^2}\right] = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{E}[(X_i - \mu)^2] = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{Var}(X_i)$$

$$= \boldsymbol{Var}(X)$$

Estimating Variance when Mean is UNknown

How about we replace μ by $\bar{X}$?

Assume $\mu = \boldsymbol{E}[X]$ is **not** known.

Let $\hat{\theta}$ be an estimator of $\theta = \boldsymbol{Var}(X)$:

$$\hat{\theta}(X_1, X_2, \ldots, X_n) = \overline{S^2} \equiv \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

$$\hat{\theta}(X_1, X_2, \ldots, X_n) = \overline{S^2} \equiv \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Estimating Variance when Mean is UNknown

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Assume $\mu = \boldsymbol{E}[X]$ is **not** known.

Let $\hat{\theta}$ be an estimator of $\theta = \boldsymbol{Var}(X)$:

$$\hat{\theta}(X_1, X_2, \ldots, X_n) = \overline{S^2} \equiv \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Q: Is $\overline{S^2}$ an unbiased estimator?

A: Not quite. In Exercise 15.4, you will prove that:

$$\boldsymbol{E}\left[\overline{S^2}\right] = \frac{n-1}{n} \cdot \boldsymbol{Var}(X)$$

So need to multiply $\overline{S^2}$ by $\frac{n}{n-1}$ to get unbiased estimator

Estimating Variance when Mean is UNknown

Defn 15.8: Let $X_1, X_2, X_3, \ldots$ be i.i.d. samples of r.v X with **unknown** mean and variance. The **sample variance** is a point estimator of $\theta = \boldsymbol{Var}(X)$. It is denoted by S^2 and defined by:

$$\hat{\theta}(X_1, X_2, \ldots, X_n) = \frac{n}{n-1}\overline{S^2} = \underbrace{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}_{\text{sample variance}} \equiv S^2$$

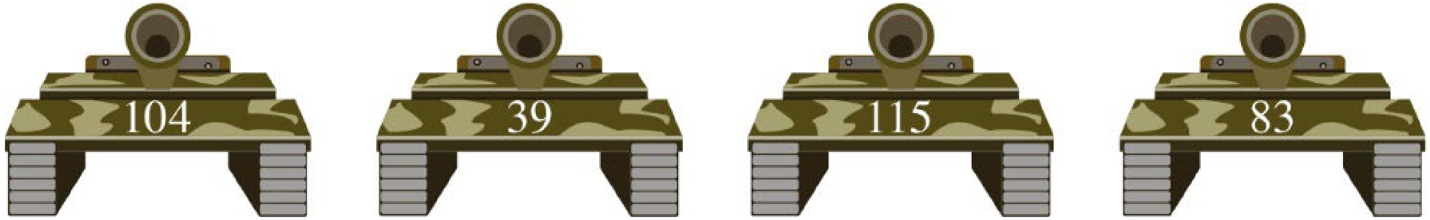
sample variance

More complex estimators

The sample mean estimator is often a component of more complex estimators. Consider the following example:

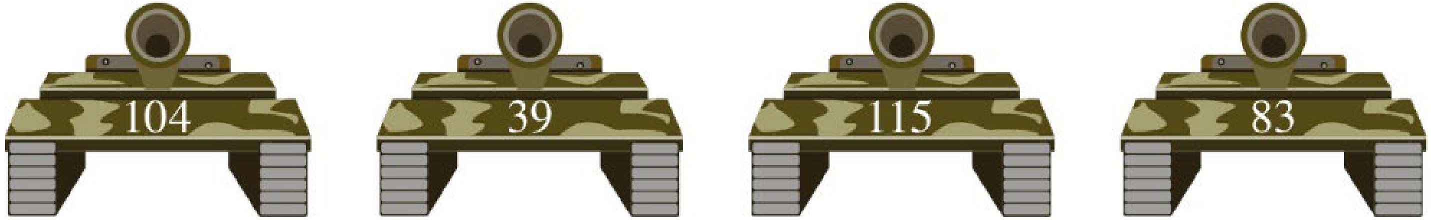
Goal: Estimate # German tanks in WW II.

Each tank has serial number. Below are the tanks we've seen so far …



Mathematically, we're trying to estimate a maximum, call it θ, based on seeing n samples, $X_1, X_2, \ldots, X_n$, each randomly picked without replacement, from the integers $1, 2, 3, \ldots, \theta$.

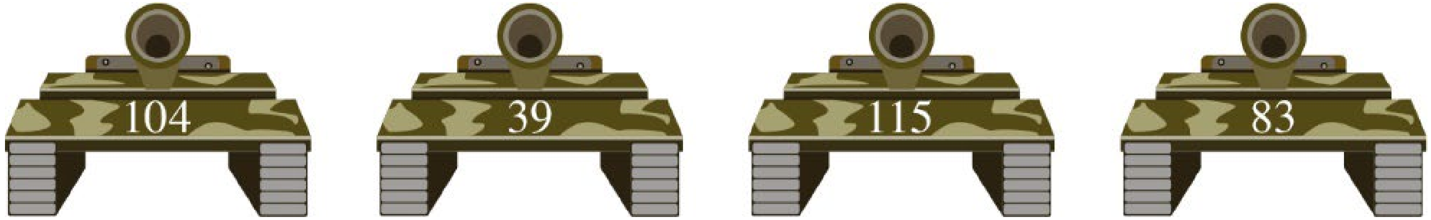
Estimating the max



Mathematically, we're trying to estimate a maximum, call it θ, based on seeing n samples, $X_1, X_2, \ldots, X_n$, each randomly picked without replacement, from the integers $1, 2, 3, \ldots, \theta$.

We know how to estimate the mean. Can we write the max in terms of the sample mean?

Estimating the max



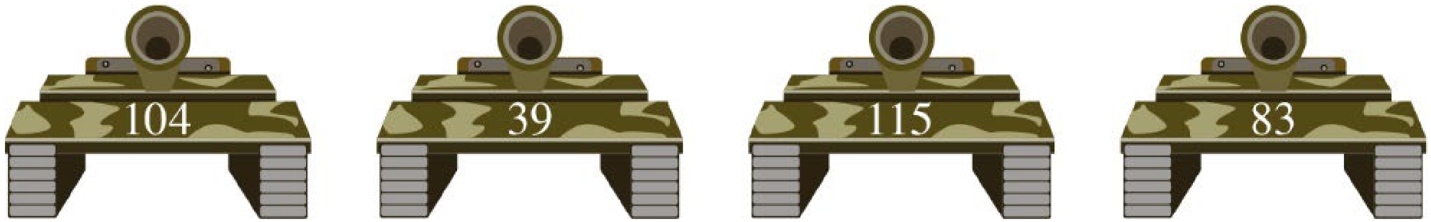
Mathematically, we're trying to estimate a maximum, call it θ, based on seeing n samples, $X_1, X_2, \ldots, X_n$, each randomly picked without replacement, from the integers $1, 2, 3, \ldots, \theta$.

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

$$\boldsymbol{E}[\bar{X}] = \frac{1}{n}(\boldsymbol{E}[X_1] + \boldsymbol{E}[X_2] + \cdots + \boldsymbol{E}[X_n]) = \frac{\theta + 1}{2}$$

$$\boldsymbol{E}[X_i] = 1 \cdot \frac{1}{\theta} + 2 \cdot \frac{1}{\theta} + 3 \cdot \frac{1}{\theta} + \cdots + \theta \cdot \frac{1}{\theta} = \frac{\theta + 1}{2}$$

Estimating the max



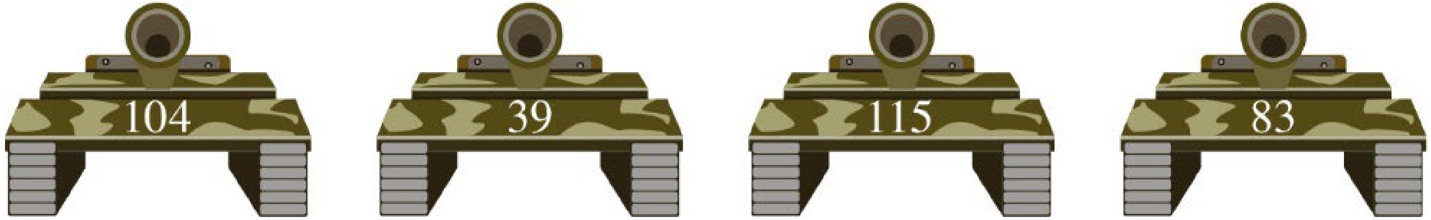
Mathematically, we're trying to estimate a maximum, call it θ, based on seeing n samples, $X_1, X_2, \dots, X_n$, each randomly picked without replacement, from the integers $1, 2, 3, \dots, \theta$.

$$\boldsymbol{E}[\bar{X}] = \frac{1}{n}(\boldsymbol{E}[X_1] + \boldsymbol{E}[X_2] + \cdots + \boldsymbol{E}[X_n]) = \frac{\theta + 1}{2}$$

Hence a reasonable estimator for θ is:

$$\hat{\theta}(X_1, X_2, \dots, X_n) = 2\bar{X} - 1$$

Estimating the max



Mathematically, we're trying to estimate a maximum, call it θ, based on seeing n samples, $X_1, X_2, \ldots, X_n$, each randomly picked without replacement, from the integers $1, 2, 3, \ldots, \theta$.

$$\boldsymbol{E}[\bar{X}] = \frac{1}{n}\left(\boldsymbol{E}[X_1] + \boldsymbol{E}[X_2] + \cdots + \boldsymbol{E}[X_n]\right) = \frac{\theta + 1}{2}$$

Is $\hat{\theta}$ unbiased?

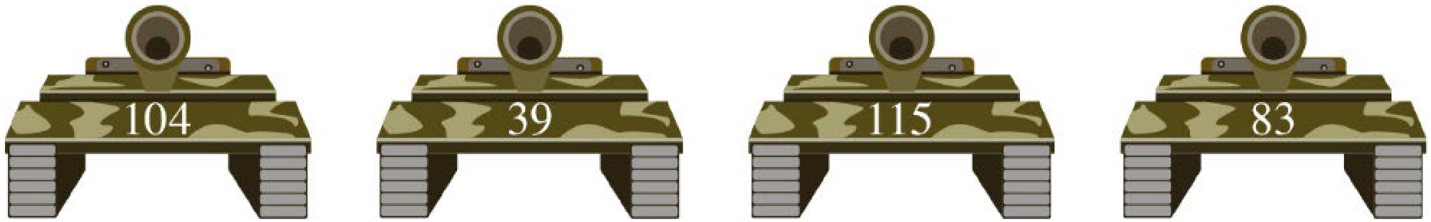
Hence a reasonable estimator for θ is:

$$\hat{\theta}(X_1, X_2, \ldots, X_n) = 2\bar{X} - 1$$

Yes!

$$\boldsymbol{E}[\hat{\theta}] = 2\boldsymbol{E}[\bar{X}] - 1$$
$$= 0$$

Estimating the max



Mathematically, we're trying to estimate a maximum, call it θ, based on seeing n samples, $X_1, X_2, \dots, X_n$, each randomly picked without replacement, from the integers $1, 2, 3, \dots, \theta$.

$$\boldsymbol{E}[\bar{X}] = \frac{1}{n}(\boldsymbol{E}[X_1] + \boldsymbol{E}[X_2] + \cdots + \boldsymbol{E}[X_n]) = \frac{\theta + 1}{2}$$

Is $\hat{\theta}$ good?

Hence a reasonable estimator for θ is:

$$\hat{\theta}(X_1, X_2, \dots, X_n) = 2\bar{X} - 1$$

Not necessarily.
$\hat{\theta} = 2\bar{X} - 1$
could be smaller
than some samples