# Introduction to Probability

Lecture 11: Estimators (Part II)

Mateja Jamnik, Thomas Sauerwald

University of Cambridge, Department of Computer Science and Technology
email: {mateja.jamnik,thomas.sauerwald}@cl.cam.ac.uk

Easter 2024

# Outline

Recap

Estimating Population Size (First Version)

Mean Squared Error

Estimating Population Size (Second Version)

# Recap: Unbiased Estimators and Bias

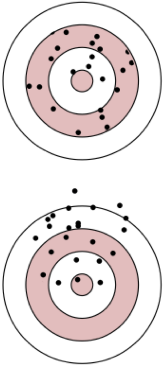
**Definition**

An estimator $T$ is called an unbiased estimator for a parameter $\theta$ if

$$\mathbf{E}[\,T\,] = \theta,$$

irrespective of the value $\theta$. The bias is defined as

$$\mathbf{E}[\,T\,] - \theta = \mathbf{E}[\,T - \theta\,].$$



Source: Edwin Leuven (Point Estimation)

???

- How can we measure the accuracy of an estimator? $\rightsquigarrow$ bias and mean-squared error
- If there are several unbiased estimators, which one to choose? $\rightsquigarrow$ mean-squared error (or variance)

# An Unbiased Estimator may not always exist

**Example 6**

Suppose that we have one sample $X \sim Bin(n, p)$, where $0 < p < 1$ is unknown but $n$ is known. Prove there is no unbiased estimator for $1/p$.

Answer

- First a simpler proof which exploits that $p$ might be arbitrarily small
- Intuition: By making $p$ smaller and smaller, we force $\max_{0 \le k \le n} T(k)$, $k \in \{0, 1, \ldots, n\}$ to become bigger and bigger
- Formal Argument:
  - Fix any estimator $T(X)$
  - Define $M := \max_{0 \le k \le n} T(k)$. Then,

$$\begin{aligned} \mathbf{E}[\,T(X)\,] &= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} \cdot T(k) \\ &\le M \cdot \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = M. \end{aligned}$$

  - Hence this estimator does not work for $p < \frac{1}{M}$, since then $\mathbf{E}[\,T(X)\,] \le M < \frac{1}{p}$ (negative bias!)
- The next proof will work even if $p \in [a, b]$ for $0 < a < b \le 1$.

# An Unbiased Estimator may not always exist (cntd. - non-examinable)

**Example 6 (cntd.)**

Suppose that we have one sample $X \sim Bin(n, p)$, where $0 < p < 1$ is unknown but $n$ is known. Prove there is no unbiased estimator for $1/p$.

Answer

- Suppose there exists an unbiased estimator with $\mathbf{E}[\,T(X)\,] = 1/p$.
- Then

$$\begin{aligned} 1 &= p \cdot \mathbf{E}[\,T(X)\,] \\ &= p \cdot \sum_{k=0}^{n} \mathbf{P}[\,X = k\,] \cdot T(k) \\ &= p \cdot \sum_{k=0}^{n} \binom{n}{k} p^k \cdot (1-p)^{n-k} \cdot T(k) \end{aligned}$$

- Last term is a polynomial of degree $n+1$ with constant term zero

  $\Rightarrow p \cdot \mathbf{E}[\,T(X)\,] - 1$ is a (non-zero) polynomial of degree $\leq n+1$

  $\Rightarrow$ this polynomial has at most $n+1$ roots

  $\Rightarrow \mathbf{E}[\,T(X)\,]$ can be equal to $1/p$ for at most $n+1$ values of $p$, and thus cannot be an unbiased.

# Outline

Recap

Estimating Population Size (First Version)

Mean Squared Error

Estimating Population Size (Second Version)

# Estimating Population Size (First Version)

- Suppose we have a sample of a few serial numbers (IDs) of some product
- We assume IDs are running from 1 to an unknown parameter $N$ (so $N = \theta$)
- Each of the IDs is drawn without replacement from the discrete uniform distribution over $\{1, 2, \ldots, N\}$
- This is also known as Tank Estimation Problem or (Discrete) Taxi Problem

$$7, 3, 10, 46, 14$$

**Warning**

- As before, we denote the samples $X_1, X_2, \ldots, X_n$
- Since sampling is without replacement:
  - they are not independent! (but identically distributed)
  - their number must satisfy $n \leq N$

# First Estimator Based on Sample Mean

**Example 1**

Construct an unbiased estimator using the sample mean.

Answer

- The sample mean is

$$\overline{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}.$$

- Linearity of expectation applies (even for dependent random var.!):

$$\begin{aligned} \mathbf{E}\left[\,\overline{X}_n\,\right] &= \frac{n \cdot \mathbf{E}\left[\,X_1\,\right]}{n} = \mathbf{E}\left[\,X_1\,\right] \\ &= \sum_{i=1}^{N} i \cdot \frac{1}{N} = \frac{N+1}{2}. \end{aligned}$$

- Thus we obtain an unbiased estimator by

$$T_1 := 2 \cdot \overline{X}_n - 1.$$

# Example: Odd Behaviour of $T_1$

- Suppose $n = 5$
- Let the sample be

$$7, 3, 10, 46, 14$$

- The estimator returns:

$$T_1 = 2 \cdot \overline{X}_n - 1 = 2 \cdot \frac{80}{5} - 1 = 31 \quad ☹$$

This estimator will often unnecessarily underestimate the true value $N$.

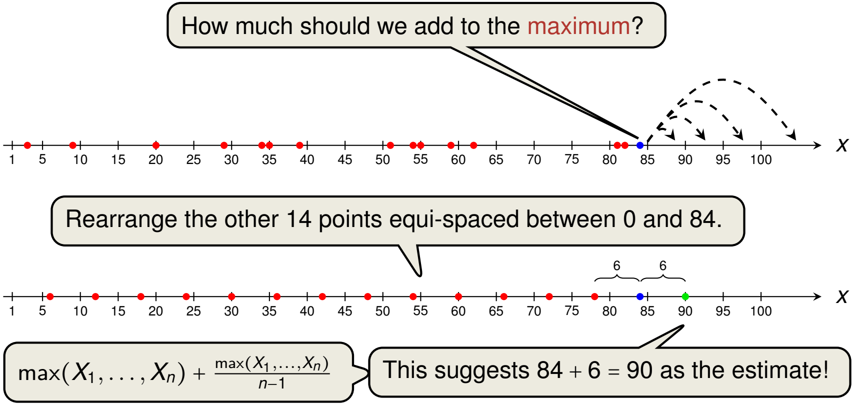
Challenging exercise: Find a lower bound on $\mathbf{P}\left[\, T_1 < \max(X_1, X_2, \ldots, X_n) \,\right]$

- Achieving unbiasedness alone is not a good strategy
- Improvement: find an estimator which always returns a value at least $\max(X_1, X_2, \ldots, X_n)$

# Intuition: Constructing an Estimator based on Maximum

- Suppose $n = 15$
- Our samples are:

$$9, 82, 39, 35, 20, 51, 54, 62, 81, 29, \textcolor{blue}{84}, 59, 3, 34, 55$$



# Deriving the Estimator Based on Maximum

**Example 2**

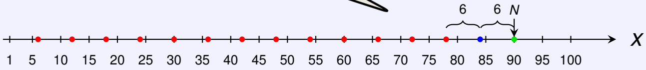
Construct an unbiased estimator using $\max(X_1, \ldots, X_n)$

Answer

- Calculate expectation of the maximum (for details see Dekking et al.)

$$\mathbf{E}\left[\max(X_1, \ldots, X_n)\right] = \ldots = \frac{n}{n+1} \cdot N + \frac{n}{n+1} = \frac{n}{n+1} \cdot (N+1).$$

Equi-spaced configuration would suggest $\max(X_1, \ldots, X_n) \approx \frac{n-1}{n} \cdot N$

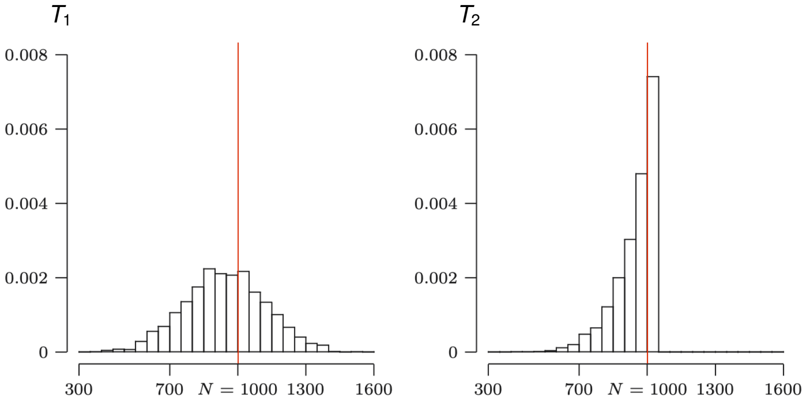



- Hence we obtain an unbiased estimator by

$$T_2 := \frac{n+1}{n} \cdot \max(X_1, \ldots, X_n) - 1.$$

- For our samples before, we get $t_2 = \frac{16}{15} \cdot 84 - 1 = 88.6$.

# Empirical Analysis of the two Estimators




Source: Modern Introduction to Statistics

Figure: Histogram of 2000 values for $T_1$ and $T_2$, when $N = 1000$ and $n = 10$.

Can we find a quantity that captures the superiority of $T_2$ over $T_1$?

# Outline

# Mean Squared Error

**Mean Squared Error Definition**

Let $T$ be an estimator for a parameter $\theta$. The mean squared error of $T$ is

$$\mathbf{MSE}[\,T\,] = \mathbf{E}\left[(T-\theta)^2\right].$$

- According to this, estimator $T_1$ better than $T_2$ if $\mathbf{MSE}[\,T_1\,] < \mathbf{MSE}[\,T_2\,]$.

**Bias-Variance Decomposition**

The mean squared error can be decomposed into:

$$\mathbf{MSE}[\,T\,] = \underbrace{(\mathbf{E}[\,T\,]-\theta)^2}_{=\,\text{Bias}^2} + \underbrace{\mathbf{V}[\,T\,]}_{=\,\text{Variance}}$$

- If $T_1$ and $T_2$ are both unbiased, $T_1$ is better than $T_2$ iff $\mathbf{V}[\,T_1\,] < \mathbf{V}[\,T_2\,]$.

## Bias-Variance Decomposition: Derivation

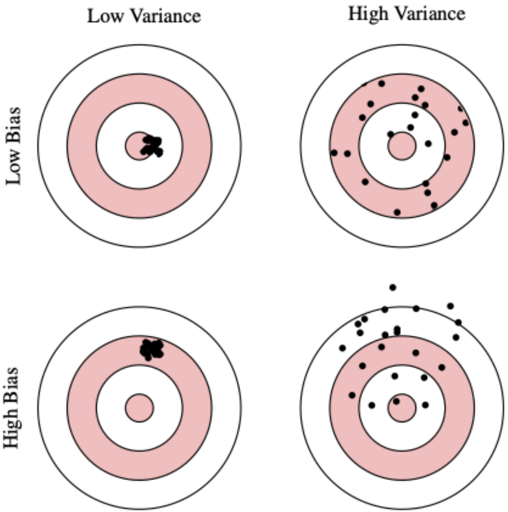
**Example 3**

We need to prove: $\mathbf{MSE}[\,T\,] = (\mathbf{E}[\,T\,] - \theta)^2 + \mathbf{V}[\,T\,]$.

Answer

$$\begin{aligned}
\mathbf{MSE}[\,T\,] &= \mathbf{E}\left[(T-\theta)^2\right] \\
&= \mathbf{E}\left[T^2 - 2T\theta + \theta^2\right] \\
&= \mathbf{E}[\,T\,]^2 - 2\cdot\mathbf{E}[\,T\,]\cdot\theta + \theta^2 + \mathbf{E}\left[T^2\right] - \mathbf{E}[\,T\,]^2 \\
&= (\mathbf{E}[\,T\,] - \theta)^2 + \mathbf{V}[\,T\,].
\end{aligned}$$

# Bias-Variance Decomposition: Illustration




Source: Edwin Leuven (Point Estimation)

Example 4

It holds that $\mathbf{MSE}[T_1] = \Theta\left(\frac{N^2}{n}\right)$, where $T_1 = 2 \cdot \overline{X}_n - 1$.

Answer

- Since $T_1$ is unbiased, $\mathbf{MSE}[T_1] = (\mathbf{E}[T_1] - \theta)^2 + \mathbf{V}[T_1] = \mathbf{V}[T_1]$, and

$$\mathbf{V}[T_1] = \mathbf{V}\left[2 \cdot \overline{X}_n - 1\right] = 4 \cdot \mathbf{V}\left[\overline{X}_n\right] = \frac{4}{n^2} \cdot \mathbf{V}[X_1 + \cdots + X_n]$$

- Note: The $X_i$'s are not independent!
- Use generalisation of $\mathbf{V}[X_1 + X_2] = \mathbf{V}[X_1] + \mathbf{V}[X_2] + 2 \cdot \mathbf{Cov}[X_1, X_2]$ (Exercise Sheet) to $n$ r.v.'s, and then that the $X_i$'s are identically distributed, and also the $(X_i, X_j)$, $i \neq j$:

$$\mathbf{V}[X_1 + \cdots + X_n] = \sum_{i=1}^{n} \mathbf{V}[X_i] + 2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} \mathbf{Cov}[X_i, X_j]$$
$$= n \cdot \mathbf{V}[X_1] + 2\binom{n}{2} \cdot \mathbf{Cov}[X_1, X_2].$$

- By definition of the discrete uniform distribution, $\mathbf{V}[X_1] = \frac{(N+1)(N-1)}{12}$
- Intuitively, $X_1$ and $X_2$ are negatively correlated, which would be sufficient to complete the proof. For a more rigorous and precise derivation (see Dekking et al.):

$$\mathbf{Cov}[X_1, X_2] = -\frac{1}{12}(N+1).$$

- Rearranging and simplifying gives

$$\mathbf{V}[T_1] = \frac{(N+1)(N-n)}{3n}.$$

# Analysis of the MSE for $T_2$ (non-examinable)

Example 5

It holds that $\mathbf{MSE}[T_2] = \Theta\left(\frac{N^2}{n^2}\right)$, where $T_2 = \frac{n+1}{n} \cdot \max(X_1, \ldots, X_n) - 1$.

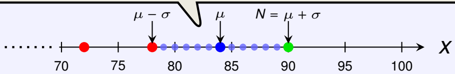
Answer

- $T_2$ is unbiased $\Rightarrow$ need $\mathbf{V}[T_2]$ which reduces to $\mathbf{V}[\max(X_1, \ldots, X_n)]$
- One can prove:

For details see Dekking et al.

$$\mathbf{V}[\max(X_1, \ldots, X_n)] = \cdots = \frac{n(N+1)(N-n)}{(n+2)(n+1)^2} = \Theta\left(\frac{N^2}{n^2}\right)$$

Equi-spaced (idealised) configuration suggests a standard deviation of $\sigma \approx \frac{N}{n}$




Maximum could have equally likely taken any value between 79 and 90

- $\mathbf{MSE}[T_2]$ is much lower than $\mathbf{MSE}[T_1] = \Theta\left(\frac{N^2}{n}\right)$, i.e., $\frac{\mathbf{MSE}[T_1]}{\mathbf{MSE}[T_2]} = \frac{n+2}{3}$
- $\Rightarrow$ confirms simulations suggesting that $T_2$ is better than $T_1$!
- can be shown $T_2$ is the best unbiased estimator, i.e., it minimises MSE.

# Outline

Recap

Estimating Population Size (First Version)

Mean Squared Error

Estimating Population Size (Second Version)

# A New Estimation Problem

Previous Model

- Population/ID space $S = \{1, 2, \ldots, N\}$
- We take uniform samples from $S$ without replacement
- **Goal:** Find estimator for $N$

Similar idea applies to situations where elements are not labelled before we see them first time (Mark & Recapture Method)

New Model

- Population/ID space of size $|S| = N$
- We take uniform samples from $S$ with replacement
- **Goal:** Find estimator for $N$

- Suppose $n = 6$, $N = 11$, $S = \{3, 4, 7, 8, 10, 15.83356, 20, 21, 56, 81, 10000\}$
- Let the sample be

$$10, 81, 20, 3, 81, 10000$$

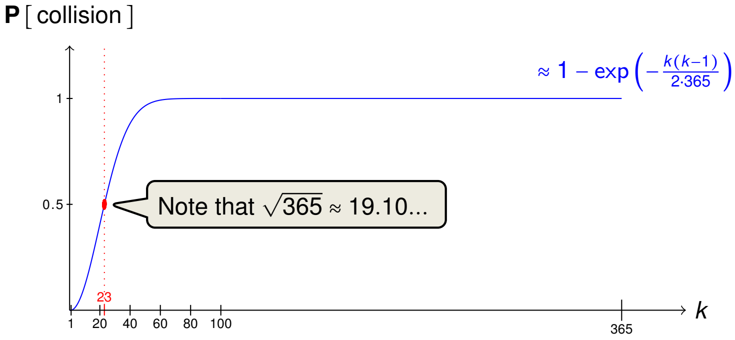
Let us call this a **collision**

As we do not know $S$, our only clue are elements that **were sampled twice.**

# Birthday Problem

**Birthday Problem:** Given a set of $k$ people

- What is the probability of having two with the same birthday (i.e., having at least one collision)?
- What is the expected number of people one needs to ask until the first collision occurs?



# Estimation via Collision: The Algorithm

Recall: As we do not know $S$, our only information are **collisions**.

```
FIND-FIRST-COLLISION(S)
C = ∅
For i = 1, 2, ...
    Take next i.i.d. sample X_i from S
    If X_i ∉ C then C ← C ∪ {X_i}
    else return T(i)
End For
```

$T(i)$ will be the value of the estimator if algo returns after $i$ rounds. (We want $T$ unbiased)

- Running Time: The expected time until the algorithm stops is:
  = the expected number of samples until a collision...

Same as the birthday problem, but now with $|S| = N$ days... ☺

**Expected Running Time (Knuth, Ramanujan)**

$$\sqrt{\frac{\pi N}{2}} - \frac{1}{3} + O\left(\frac{1}{\sqrt{N}}\right).$$

**Exercise:** Prove a bound of $\leq 2 \cdot \sqrt{N}$

# Estimation via Collision: Getting the Estimator Unbiased

Example 6

One can define $T(i)$, $i \in \mathbb{N}$, such that $\mathbf{E}[\,T\,] = |S|$ for any finite, non-empty set $S$.

Answer

- We outline a construction by induction.
- Case $|S| = 1$: Algo always stops after $i = 2$ rounds and returns $T(2)$. We want

$$1 = \mathbf{E}[\,T\,] = T(2) \qquad \Rightarrow \qquad T(2) = 1.$$

- Case $|S| = 2$: Algo stops after 2 or 3 rounds (w.p. 1/2 each). We want

$$2 = \mathbf{E}[\,T\,] = \frac{1}{2} \cdot T(2) + \frac{1}{2} \cdot T(3) \qquad \Rightarrow \qquad T(3) = 3.$$

- Case $|S| = 3$: gives $3 = \mathbf{E}[\,T\,] = \frac{1}{3} \cdot T(2) + \frac{4}{9} \cdot T(3) + \frac{2}{9} \cdot T(4)$
$\Rightarrow T(4) = 6$, similarly, $T(5) = 10$ etc.
- can continue to define $T(i)$ inductively in this way (note $T$ is unique) (a proof that $T(i) = \binom{i}{2}$ is harder)