# Tank Estimators

Jo Hardin

Fall 2020

How can a random sample of integers between 1 and $N$ (with $N$ unknown to the researcher) be used to estimate $N$? This problem is known as the German tank problem and is derived directly from a situation where the Allies used maximum likelihood to determine how many tanks the Axes had produced. See `https://en.wikipedia.org/wiki/German_tank_problem`.

1. The tanks are numbered from 1 to $N$. Working with your group, randomly select five tanks, without replacement, from the bowl. The tanks are numbered:

2. Think about how you would use your data to estimate $N$. (Come up with at least 3 estimators.) Come to a consensus within the group as to how this should be done. One person from your group will report out after the warm-up is over. Ideally, the person to report out will be someone who has not yet spoken in class this semester. Step-up if you haven't yet spoken. Step back if you speak regularly.

   Our estimates of $N$ are:

   Our rules or formulas for the estimators of $N$ based on a sample of n (in your case 5) integers are:

Assuming the random variables are distributed according to a discrete uniform. (Tbh, this model is with replacement, but the answers you get aren't much different than without replacement if $n << N$.)

$$X_i \sim P(X = x|N) = \frac{1}{N} \qquad x = 1, 2, \ldots, N \quad i = 1, 2, \ldots, n$$

3. What is the method of moments estimator of N?

4. What is the maximum likelihood estimator of N?

## Theoretical Mean Squared Error

Most of our estimators are made up of four basic functions of the data: the mean, the median, the min, and the max. Fortunately, we know something about the moments of these functions:

| g($\underline{X}$) | E[ g($\underline{X}$) ] | Var( g($\underline{X}$) ) |
|---|---|---|
| $\overline{X}$ | $\frac{N+1}{2}$ | $\frac{(N+1)(N-1)}{12n}$ |
| median($\underline{X}$) = M | $\frac{N+1}{2}$ | $\frac{(N-1)^2}{4n}$ |
| min($\underline{X}$) | $\frac{(N-1)}{n}+1$ | $\left(\frac{N-1}{n}\right)^2$ |
| max($\underline{X}$) | $N-\frac{(N-1)}{n}$ | $\left(\frac{N-1}{n}\right)^2$ |

Using this information, we can calculate the MSE for 4 of the estimators that we have derived. (Remember that MSE = Variance + Bias$^2$.)

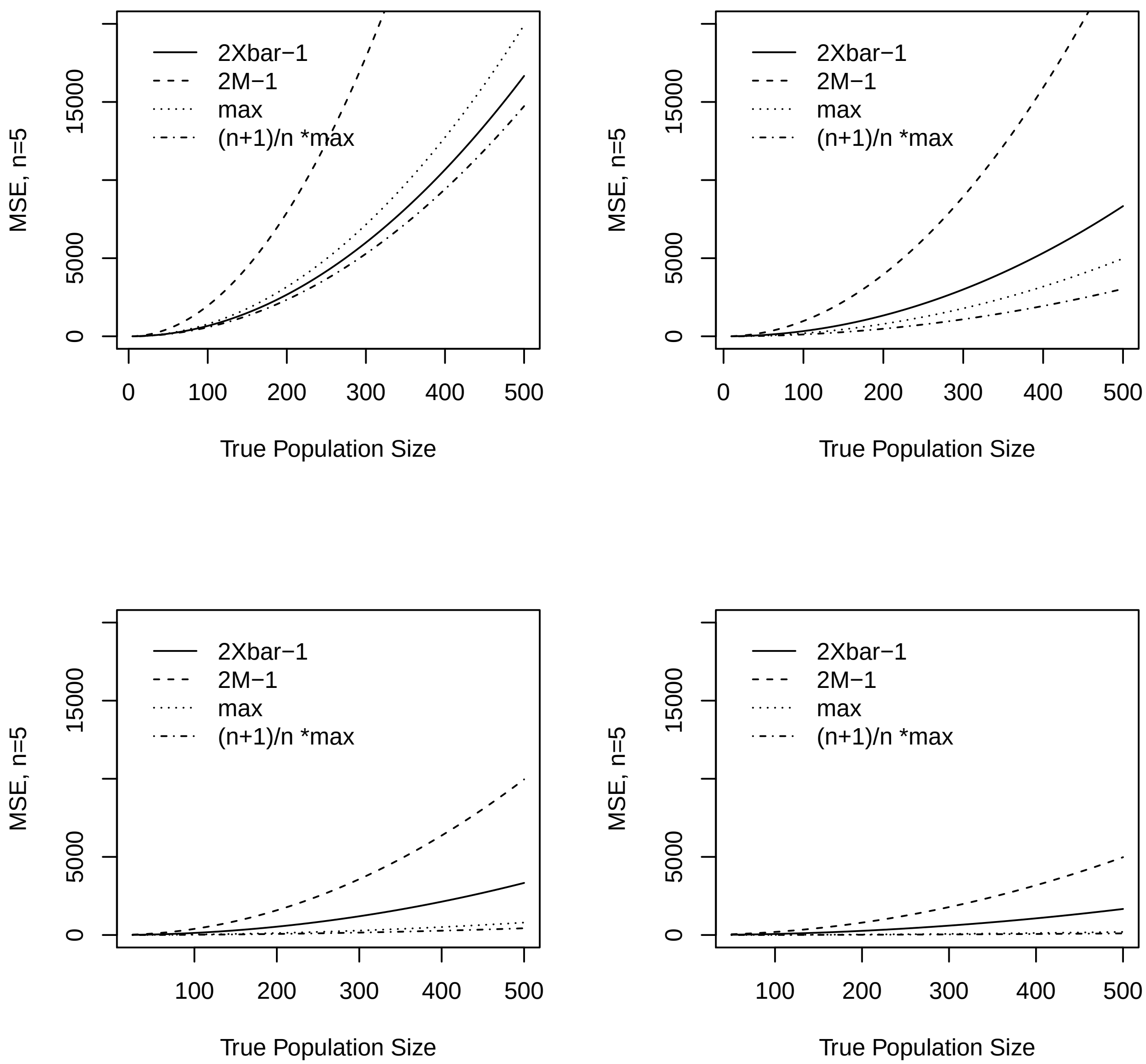
$$\begin{aligned}
\text{MSE }(2\cdot\overline{X}-1) &= \frac{4(N+1)(N-1)}{12n}+\left(2\left(\frac{N+1}{2}\right)-1-N\right)^2 \\
&= \frac{4(N+1)(N-1)}{12n} \qquad (1)
\end{aligned}$$

$$\begin{aligned}
\text{MSE }(2\cdot M-1) &= \frac{4(N-1)^2}{4n}+\left(2\left(\frac{N+1}{2}\right)-1-N\right)^2 \\
&= \frac{4(N-1)^2}{4n} \qquad (2)
\end{aligned}$$

$$\begin{aligned}
\text{MSE }(\max(\underline{X})) &= \left(\frac{N-1}{n}\right)^2+\left(N-\frac{(N-1)}{n}-N\right)^2 \\
&= \left(\frac{N-1}{n}\right)^2+\left(\frac{N-1}{n}\right)^2 = 2*\left(\frac{N-1}{n}\right)^2 \qquad (3)
\end{aligned}$$

$$\text{MSE }\left(\left(\frac{n+1}{n}\right)\max(\underline{X})\right) = \left(\frac{n+1}{n}\right)^2\left(\frac{N-1}{n}\right)^2+\left(\left(\frac{n+1}{n}\right)\left(N-\frac{N-1}{n}\right)-N\right)^2 \qquad (4)$$

MSE for different estimates of the Population Size



## Empirical Mean Squared Error

Compare the following six estimators:

$$
\begin{aligned}
\hat{N}_1 &= 2 \cdot \overline{X} - 1 \\
\hat{N}_2 &= 2 \cdot median(\underline{X}) - 1 \\
\hat{N}_3 &= max(\underline{X}) \\
\hat{N}_4 &= \frac{n+1}{n} max(\underline{X}) \\
\hat{N}_5 &= max(\underline{X}) + min(\underline{X}) \\
\hat{N}_6 &= \frac{n+1}{n-1}[max(\underline{X}) - min(\underline{X})]
\end{aligned}
$$

When $n = 5$, which means the sample size is 5 (whereas the population number is 447), the measurements are shown below. We want empirical MSE to be lowest and bias close to 0. The modified maximum has the lowest MSE and thus this estimator is the best when $n = 5$. The histograms estimating the sampling distributions are illustrated below.

```
## # A tibble: 6 x 6
##   method         mean median     bias    var    mse
##   <chr>         <dbl>  <dbl>    <dbl>  <dbl>  <dbl>
## 1 2 medians      448.   451    1.16   28343. 28344.
## 2 2 xbars        447.   449.   0.0882 13176. 13176.
## 3 diff max min   448.   462    1.01   13805. 13806.
## 4 modified max   448.   467.   0.748   5604.  5605.
## 5 sample max     373.   389  -73.9     3892.  9349.
## 6 sum min max    448.   448    0.574   9504.  9504.
```