UNIFORM ESTIMATION

K.N. RUMSEY

1. The Problem

In this short paper, we will consider the rather simple problem of estimating θ based on a random sample from the $U(0, \theta)$ distribution.

$$X_1, X_2, \cdots X_n \overset{iid}{\sim} U(0, \theta) \tag{1}$$

$$f(x_i|\theta) = \frac{\mathcal{I}(0 \le x_i \le \theta)}{\theta} \tag{2}$$

We will begin by briefly presenting the Method of Moments and Maximum Likelihood estimators, and then moving onto some possible improvements. We will present the well known Uniform Minimum Variance Unbiased Estimator (UMVUE) correction, an estimator which minimizes the MSE, a jack-knife resampling estimator to remove bias and finally a probability of error estimator which does not depend on n.

2. Classical Estimators

2.1. **Method of Moments.** In a single parameter model, the Method of Moments estimator simply sets the sample mean $\bar{X}$ equal to the first moment μ and solves algebraically for θ. In this case, we have

$$\mu = E(X_1) = \frac{\theta}{2} \tag{3}$$

which implies

$$\bar{X} = \frac{\theta}{2} \quad \Rightarrow \quad \hat{\theta}_{mom} = 2\bar{X} \tag{4}$$

The bias and variance calculations for this estimator are quite easy leading to a tractable formula for MSE which we will show later in this section.

2.2. **Maximum Likelihood.** The Maximum Likelihood Estimator for this problem is also relatively simple using a trick with the indicator function. We begin by writing the Likelihood function for this model where $x_{(1)}$ and $x_{(n)}$ are the maximum and minimum order statistics of the sample respectively.

$$L(\theta|X_1, \cdots X_n) = \prod_{i=1}^n f(x_i|\theta) = \frac{\prod_{i=1}^n \mathcal{I}(0 \le x_i \le \theta)}{\theta^n} = \frac{\mathcal{I}(0 \le x_{(1)})\mathcal{I}(x_{(n)} \le \theta)}{\theta^n} \tag{5}$$

As a function of θ we see that this function (ignoring indicator functions) is monotonically decreasing. Hence we need to make θ as small as possible. However the likelihood becomes 0 when $\theta < x_{(n)}$, hence to maximize the likelihood we should choose

$$\hat{\theta}_{mle} = X_{(n)} \tag{6}$$

Using the four-parameter beta distribution, we will be able to determine the MSE of this estimator as well.

2.3. **Mean Squared Error - Method of Moments.** The bias of an estimator is given by $B(\hat{\theta}) = E(\hat{\theta}) - \theta$. It is easy to show that the Method of Moments estimator is unbiased.

$$B(\hat{\theta}_{mom}) = E(2\bar{X}) - \theta = 2E(X_1) - \theta = 2\frac{\theta}{2} - \theta = 0 \tag{7}$$

The variance is also straightforward here.

$$Var(\hat{\theta}_{mom}) = Var\left(\frac{2\sum_{i=1}^{n} x_i}{n}\right) = \frac{4}{n^2}\left(\sum_{i=1}^{n} Var(X_i)\right) = \frac{4}{n^2}\frac{n\theta^2}{12} = \frac{\theta^2}{3n} \tag{8}$$

Putting these together, we get the MSE.

$$MSE(\hat{\theta}_{mom}) = \left[B(\hat{\theta}_{mom})\right]^2 + Var(\hat{\theta}_{mom}) = \frac{\theta^2}{3n} \tag{9}$$

2.4. **Mean Squared Error - Maximum Likelihood.** We will need to use the following well known fact.

Proposition*:*
If $X_1, X_2, \cdots X_n$ is a random sample from the $U(0, \theta)$ distribution, then the maximum order statistic $X_{(n)}$ follows the four parameter beta distribution $Be_{[0,\theta]}(n, 1)$.

A simple way of leveraging this is to notice that $\hat{\theta}_{mle} = X_{(n)} \approx \theta Y$ where $\approx$ means "equal in distribution" and $Y \sim Be(n, 1)$. Thus to calculate the bias and variance we can use the moments of the two-parameter beta distribution.

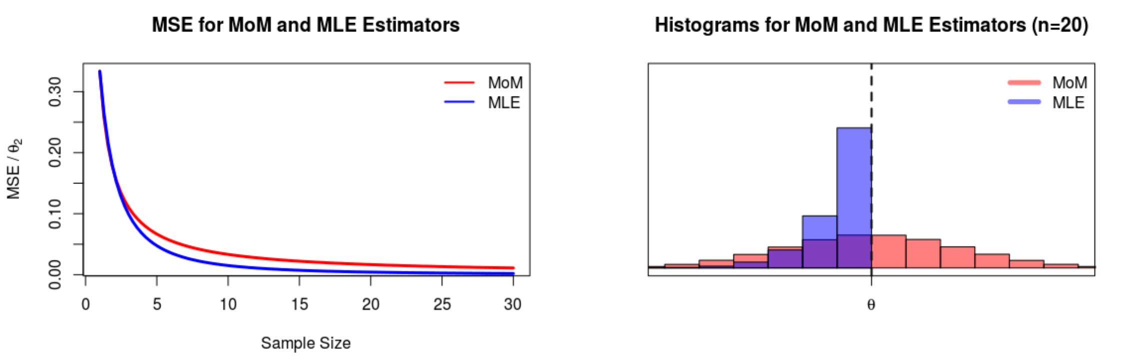
$$B(\hat{\theta}_{mle}) = \theta E(Y) - \theta = \theta\left(\frac{n}{n+1} - 1\right) = \frac{-\theta}{n+1} \tag{10}$$

$$Var(\hat{\theta}_{mle}) = \theta^2 Var(Y) = \frac{\theta^2 n}{(n+1)^2(n+2)} \tag{11}$$

$$MSE(\hat{\theta}_{mle}) = \left(\frac{-\theta}{n+1}\right)^2 + \frac{\theta^2 n}{(n+1)^2(n+2)} = \theta^2\frac{(n+2)+n}{(n+1)^2(n+2)} = \frac{2\theta^2}{(n+1)(n+2)} \tag{12}$$

We conclude this section with plots of the MSE as a function of n and histograms based on simulations with $n = 20$.

Figure 1



3. Bias Elimination

Although the Method of Moments produces an unbiased estimate, the variance (and thus the MSE) tends to be large compared to the Maximum Likelihood estimate. In this section, we consider two methods of bias elimination. First we give the UMVUE, which is well known and by definition the least variance unbiased estimate. Secondly, we consider a re-sampling technique called Jack-knifing and demonstrate it's unbiasedness numerically.

3.1. **The UMVUE.** Without going into details, we give another proposition which ties the UMVUE to completeness and sufficiency of a statistic.

Proposition*: If T is a complete and sufficient statistic for a parameter θ and $\phi(T)$ is an estimator based on the data only through T, then $\phi(T)$ is the (unique) minimum variance estimator of its expected value*

Without justification, we note that $X_{(n)}$ is complete and sufficient for θ. In the previous section, we found that $E(X_{(n)}) = \frac{\theta n}{n+1}$. Through a simple bias correction, we obtain the estimator

$$\hat{\theta}umvue = \frac{n+1}{n} X_{(n)} \tag{13}$$

which is clearly unbiased for θ. Since $\hat{\theta}umvue$ depends on the data only through $X_{(n)}$ we use the proposition to conclude that it is in fact the UMVUE. The MSE calculation is simple given the work done in the previous section.

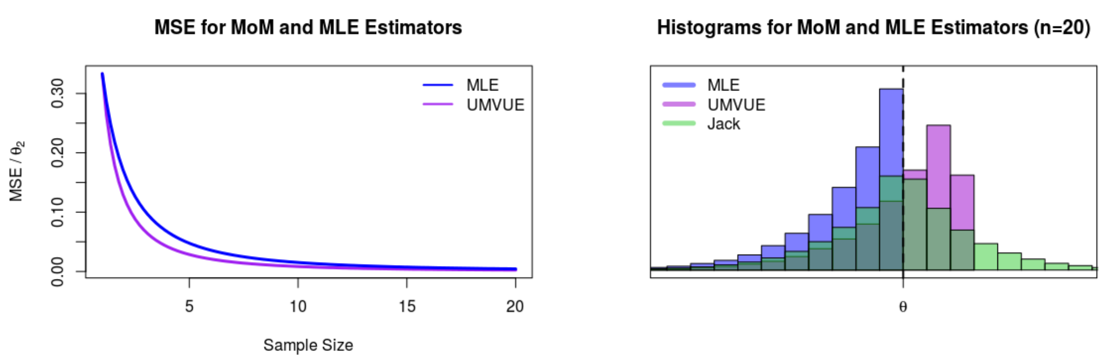
$$MSE(\hat{\theta}umvue) = 0^2 + \theta^2 \left(\frac{n+1}{n}\right) Var(X_{(n)}) = \theta^2 \frac{(n+1)^2}{n^2} \frac{n}{(n+1)^2(n+2)} = \frac{\theta^2}{n(n+2)} \tag{14}$$

3.2. **Jackknife Estimator.** Jackknifing is a resampling method which provides a way to estimate the bias. In this case, Jackknifing really isn't necessary because the exact bias can be calculated. It's inclusion in this paper is for illustrative purposes, since the Jack-Knife can often be a useful tool when the bias can't be calculated. Given an estimator $\hat{\theta}$ (we will use the mle), let $\hat{\theta}_{-i}$ to be the same estimator calculated with the i^{th} observation eliminated from the sample.

$$\hat{\theta}_{jack} = n\hat{\theta}_{mle} - \frac{n-1}{n} \sum_{i=1}^{n} \hat{\theta}_{-i} \tag{15}$$

Calculating the MSE requires the joint distribution of 2 order statistics so we exclude it for simplicity.

Figure 2



4. Further Extensions

In this section, we will use optimization to construct two more estimators. The first is similar to the UMVUE with the objective being to minimize the MSE. The second minimizes the probability of estimation error, which leads to an interesting estimator which (upon numerical analysis) does not depend on the sample size n.

4.1. Minimizing MSE.

In the UMVUE approach we saw earlier, the idea was straightforward. The MLE provides a good starting place, but we know that $X_{(n)}$ must always be smaller than θ. Thus we multiply by a constant (which is greater than 1) to eliminate the bias. In doing so, we add additional variance but nonetheless we see an improvement in MSE.

We propose considering estimators of the form

$$\hat{\theta}_c = cX_{(n)} \tag{16}$$

for $c \geq 1$. Immediately, we see two special cases: $\hat{\theta}_{mle} = \hat{\theta}_1$ and $\hat{\theta}_{umvue} = \hat{\theta}_{\frac{n+1}{n}}$. Our goal is to find c which minimizes $MSE(\hat{\theta}_c)$. Recall that $cX_{(n)} \approx c\theta Y$ where $\approx$ means equal in distribution and $Y \sim Be(n, 1)$.

$$B(\hat{\theta}_c) = \theta\left(\frac{cn}{n+1} - 1\right) \tag{17}$$

$$Var(\hat{\theta}_c) = \theta^2\left(\frac{c^2 n}{(n+1)^2(n+2)}\right) \tag{18}$$

$$MSE(\hat{\theta}_c) = \theta^2\left(\left(\frac{cn}{n+1} - 1\right)^2 + \frac{c^2 n}{(n+1)^2(n+2)}\right) \tag{19}$$

This becomes a straightforward one-dimensional optimization problem which can be done using Calculus. It is easy to derive that the optimal c is given by

$$c_{opt} = \frac{(n+1)(n+2)}{n(n+2)+1} \tag{20}$$

A simple algebraic argument shows that $\hat{\theta}_{mle} < \hat{\theta}_{c_{opt}} < \hat{\theta}_{umvue}$. The improvement in MSE comes from the bias variance trade off. The UMVUE eliminates the bias, where as $\hat{\theta}_{c_{opt}}$ only reduces it. We are allowing the estimator to be slightly biased so long as the reduction in variance is worth it the MSE sense. In conclusion, we have found that of all estimators of the form given in (16), the one with the minimal MSE is given by:

$$\hat{\theta}_{c_{opt}} = \frac{(n+1)(n+2)}{n(n+2)+1} X_{(n)} \tag{21}$$

4.2. Minimzing Probability of Error.

For motivation, imagine an application where we are okay with a certain amount of error, say ϵ. Supposing that any estimation error which is greater than ϵ is catastrophic, it makes sense to find the estimator which minimizes the probability of catastrophe.

First, notice that ϵ should be fairly small. Otherwise the probability of catastrophe will converge to 0 for all reasonable estimators. Secondly, we will see shortly that this formulation in terms of ϵ will cause some issues, but we address that momentarily. Let us again look at estimators of the form given in (16), but this time the goal is to minimize:

$$f(c) = P(|\hat{\theta}_c - \theta| > \epsilon) \tag{22}$$

For the last time, we will leverage the distributional equivalence of $cX_{(n)}$ and $c\theta Y$ for $Y \sim Be(n,1)$.

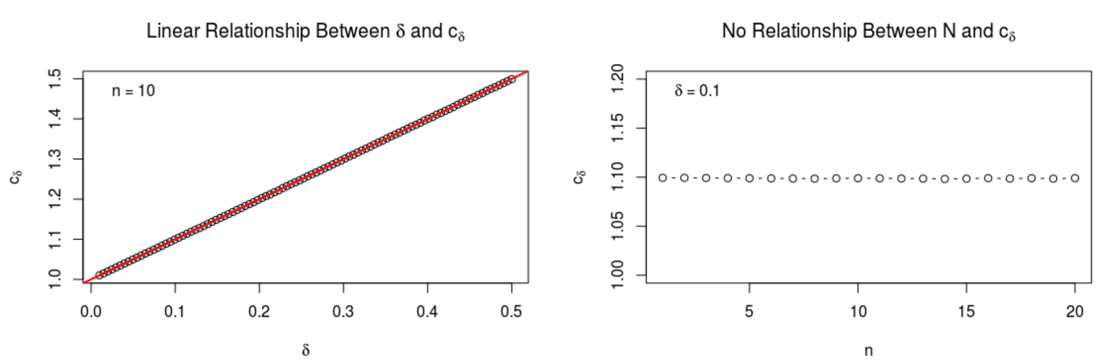
$$\begin{aligned} f(c) &= P(|\hat{\theta}_c - \theta| > \epsilon) \\ &= P(\hat{\theta}_c < \theta - \epsilon) + P(\hat{\theta}_c > \theta + \epsilon) \\ &= P\left(Y < \frac{\theta - \epsilon}{c\theta}\right) + P\left(Y > \frac{\theta + \epsilon}{c\theta}\right) \\ &= P\left(Y < \frac{1}{c}\left(1 - \epsilon/\theta\right)\right) + 1 - P\left(Y > \frac{1}{c}\left(1 + \epsilon/\theta\right)\right) \\ &= 1 + F_Y(L) - F_Y(R) \\ &\text{where } L = \frac{1}{c}\left(1 - \epsilon/\theta\right) \text{ and } R = \frac{1}{c}\left(1 + \epsilon/\theta\right) \end{aligned}$$

If there is a way to minimize this function directly, it is beyond the scope of this simple exploratory paper. A simple numerical optimization at first produces troublesome results. The optimal c depends on θ which is strictly violates the properties of an estimator. Therefore, at this point we substitute $\epsilon = \delta\theta$ for $\delta < 1$. This yeilds:

$$\begin{aligned} f(c) &= 1 + F_Y(L') - F_Y(R') \\ &\text{where } L' = (1-\delta)/c \text{ and } R' = (1-\delta)/c \end{aligned}$$

Keeping in mind that δ should be small (otherwise there is no unique solution up to machine epsilon), we optimize over a grid of n and δ values. Interestingly, we see that the optimal multiplier denoted c_δ does not depend on n and the relationship between c_δ and δ is linear.

Figure 3



The red line in the left image gives the optimal relationship between c_δ and δ.

$$c_\delta = 1 + \delta \tag{23}$$

$$\hat{\theta}_\delta = (1+\delta)X_{(n)} \tag{24}$$

If the sample size is unknown, we are unable to use any of the previously discussed bias reduction techniques. This estimator provides us with a justifiable way to increase the MLE in absence of the sample size. If n is unknown however, there is no way to compute the MSE and hence no way of knowing how to choose δ in terms of the MSE.

5. Concluding Remarks

To summarize, we have proposed 6 estimators of θ. We began with 2 classical estimation techniques, and then discussed methods of bias reduction/elimination for the MLE approach. We concluded with estimators which minimized the MSE and probability of error. In the following table, we give the equation for each estimator, as well it's Bias, Variance and MSE (if analytically computable). Finally, we give the estimated MSE for each estimator using $100,000$ simulations setting $\theta = 1$ and $n = 20$.

Table 1. Summary of Results

Method	Equation	Bias	Variance	$1000 \times \hat{MSE}$
Method of Moments	$\hat{\theta} = 2\bar{X}$	0	$\frac{\theta^2}{3n}$	16.64
Maximum Likelihood	$\hat{\theta} = X_{(n)}$	$\frac{-\theta}{n+1}$	$\frac{n\theta^2}{(n+1)^2(n+2)}$	4.32
UMVUE	$\hat{\theta} = \frac{n+1}{n} X_{(n)}$	0	$\frac{\theta^2}{n(n+2)}$	2.27
Jackknife	$\hat{\theta} = nX_{(n)} - \frac{n-1}{n} \sum_{i=1}^{n} \hat{\theta}_{-i}$	0	$-$	4.15
Min MSE	$\hat{\theta} = \frac{(n+1)(n+2)}{n(n+2)+1} X_{(n)}$	$\frac{-\theta}{n(n+2)+1}$	$\frac{n(n+2)\theta^2}{(n(n+2)+1)^2}$	2.26
Min Probability of Catastrophe	$\hat{\theta} = (1+\delta)X_{(n)}$	$\frac{\delta n - 1}{n+1}\theta$	$\frac{(1+\delta)^2 n}{(n+1)^2(n+2)}\theta^2$	4.76

Figure 4