p124-125 AoS book. Consider $X_1, \ldots, X_n \overset{iid}{\sim} U(0, \theta)$ i.e. the uniform DGP where the upper bound of the support is unknown (θ). We already derived the "silly" MM estimator $\hat{\theta}^{MM} = 2\bar{X}$ but can the MLE do "better"? Let's derive the MLE:

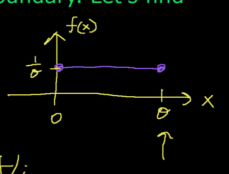
$1/\theta \, \mathbb{1}_{x \in [0, \theta]}$

$$0 \overset{set}{=} \sum_{i=1}^{n} \frac{\partial}{\partial \theta}\left[\ln(f(x_i; \theta))\right] = \sum \frac{\partial}{\partial \theta}\left[\underbrace{\ln\left(\frac{1}{\theta}\right)}_{-\ln(\theta)}\right] = -\sum \frac{\partial}{\partial \theta}\left[\ln(\theta)\right]$$

$$= -\sum \frac{1}{\theta} = -\frac{n}{\theta} = 0 \Rightarrow$$ no solution for θ!

The MLE is not at a critical point. It's at a boundary. Let's find the likelihood function explicitly.

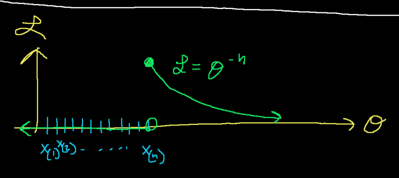
$$f(x; \theta) = \begin{cases} \frac{1}{\theta} & \text{if } 0 \le x \le \theta \\ 0 & o/t \end{cases}$$



$$\underbrace{f(x_1, \ldots, x_n; \theta)}_{} = \prod_{i=1}^{n} f(x_i; \theta) = \begin{cases} \frac{1}{\theta^n} & \text{if } 0 \le x_i \le \theta \;\; \forall i \\ 0 & o/t \end{cases}$$

$$= \mathcal{L}(\theta; x_1, \ldots, x_n) = \begin{cases} \frac{1}{\theta^n} & \text{if } \theta \ge x_i \;\forall i \Leftrightarrow \theta \ge x_{(n)} \\ 0 \end{cases}$$

Definition of "order statistics". Given $x_1, \ldots, x_n$, order their values from minimum to maximum and denote them as $x_{(1)}, \ldots, x_{(n)}$ where the minimum is $x_{(1)}$ and the max is $x_{(n)}$.



$\hat{\theta}^{MLE} := \text{argmax}(\mathcal{L}) = X_{(n)}$, the maximum of the data

$\Rightarrow \hat{\theta}^{MLE} \sim$ GeneralBeta(n, 1, θ), by MATH 368

$$Var[\hat{\theta}^{MLE}] = \theta^2 \frac{n}{(n+1)^2(n+2)} \quad \text{by MATH 368}$$

$$Var[\hat{\theta}^{MM}] = 4Var[\bar{X}] = 4\frac{Var[X]}{n} \overset{(b-a)^2/12 \text{ from 241}}{=} 4\frac{\theta^2}{12n} = \theta^2 \frac{1}{3n} \quad \text{(demo next class)}$$

Which estimator is better? Which has lower variance? (Ignore bias for now). The ratio is called "relative efficiency" (RE)

$$RE = \frac{Var[\hat{\theta}^{MM}]}{Var[\hat{\theta}^{MLE}]} = \frac{\cancel{\theta^2}\frac{1}{3n}}{\cancel{\theta^2}\frac{n}{(n+1)^2(n+2)}} = \frac{(n+1)^2(n+2)}{3n^2} > 1 \quad \forall n$$

$\Rightarrow$ MLE better!

You may want to compare the ratio of MSE's since the MLE is biased but the bias is so small it doesn't change the story. The MLE is hands-down better.

Questions:
1. Is there a theoretical minumum MSE when estimating θ for each DGP?
2. If (1), for any DGP, is there a procedure to locate that "best" estimator?

The answer to both is "no" because the class of "all estimators" is too large and the choice of θ matters. Here's why:

$X_1, \ldots, X_n \overset{iid}{\sim} Bern(\theta)$, $\hat{\theta} = \bar{X}$, $\hat{\theta}_{BAD} = \frac{1}{2}$.

At $\theta = \frac{1}{2}$, $MSE[\hat{\theta}] = \frac{1}{4n}$, $MSE[\hat{\theta}_{BAD}] = 0 \Rightarrow \hat{\theta}_{BAD}$ wins!

To make question 1's answer a "yes" we limit the class of all allowable estimators to only those that are unbiased. So...
Is there a theoretical minimum MSE for all unbiased estimators given a DGP? YES. It is called the "uniformly minimum variance unbiased estimator" (UMVUE). It's also the minimum MSE (since MSE = variance when unbiased). Denote it $\hat{\theta}^*$. Then,

$$Var[\hat{\theta}^*] \le Var[\hat{\theta}] \quad \forall \text{ unbiased } \hat{\theta}$$

Is there a closed form expression for Var[$\hat{\theta}^*$]? Yes. It is called the Cramer-Rao Lower bound (CRLB) proven by Cramer and Rao in 1945-1946.

If you can show that the variance of your unbiased estimator = CRLB, then your estimator is the UMVUE!!

Let's first prove the closed form of the CRLB: consider a DGP $X_1, \ldots, X_n \overset{iid}{\sim} f(x; \theta)$ then for an unbiased estimator $\hat{\theta}$:

$$Var[\hat{\theta}] \ge \frac{I(\theta)^{-1}}{n} \quad \text{where} \quad I(\theta) := E_X\left[\ell'(\theta; x)^2\right]$$

I(θ) is called the "Fisher Information" defined by Fisher in 1922. It is the derivative wrt θ of the log-likelihood function squared.

From MATH 368, we prove that for any two rv's Q and S,

$Cov[Q, S]^2 \le Var[Q] \, Var[S]$

$$\Rightarrow Var[Q] \ge \frac{Cov[Q, S]^2}{Var[S]} = \frac{(E[QS] - E[Q]E[S])^2}{E[S^2] - E[S]^2}$$

let $Q = \hat{\theta}$ since it is unbiased, $E[\hat{\theta}] = \theta$

$$\Rightarrow Var[\hat{\theta}] \ge \frac{(E[\hat{\theta} S] - \theta E[S])^2}{E[S^2] - E[S]^2}$$

Let S be the "Score Function", $S := \frac{\partial}{\partial \theta}\left[\ln(f(X_1, \ldots, X_n; \theta))\right]$ def 1

$$\overset{iid}{:=} \frac{\partial}{\partial \theta}\left[\ln\left(\prod_{i=1}^{n} f(X_i; \theta)\right)\right] \quad \text{def 3}$$

$$:= \frac{\partial}{\partial \theta}\left[\sum_{i=1}^{n} \ln(f(X_i; \theta))\right] \quad \text{def 4}$$

$$:= \frac{\frac{\partial}{\partial \theta}\left[f(X_1, \ldots, X_n; \theta)\right]}{f(X_1, \ldots, X_n; \theta)} \quad \text{def 2}$$

These are multiple definitions of the score function or alternate notations. They are useful in different contexts.

$$:= \sum_{i=1}^{n} \frac{\partial}{\partial \theta}\left[\ln(f(X_i; \theta))\right] \quad \text{def 5}$$

$$:= \frac{\partial}{\partial \theta}\left[\ell(\theta; X_1, \ldots, X_n)\right] \quad \text{def 6}$$

$$:= \ell'(\theta; X_1, \ldots, X_n) \quad \text{def 7}$$

$$\overset{iid, \text{ def 4}}{:=} \sum_{i=1}^{n} \ell'(\theta; X_i) \quad \text{def 8}$$

$$E[S] \overset{\text{def 2}}{=} E_X\left[\frac{\frac{\partial}{\partial \theta}\left[f(X_1, \ldots, X_n; \theta)\right]}{f(X_1, \ldots, X_n; \theta)}\right]$$

$$= \int_{Supp[X_1]} \cdots \int_{Supp[X_n]} \frac{\frac{\partial}{\partial \theta}\left[f(x_1, \ldots, x_n; \theta)\right]}{\cancel{f(x_1, \ldots, x_n; \theta)}} \cancel{f(x_1, \ldots, x_n; \theta)} \, dx_1 \ldots dx_n$$

Assume we can interchange differentiation and integration (for details of when this is legal, study real analysis)

$$= \frac{\partial}{\partial \theta}\left[\int_{Support} \cdots \int f(x_1, \ldots, x_n; \theta) \, dx_1 \ldots dx_n\right] = \frac{\partial}{\partial \theta}[1] = 0$$

$\Rightarrow E[S] = 0$ (Fact 1a)

$$0 = E[S] \overset{\text{def 8}}{=} E\left[\sum \ell'(\theta; X_i)\right] \overset{\forall i}{=} n \, E\left[\ell'(\theta; X_i)\right] \Rightarrow E\left[\ell'(\theta; X_i)\right] = 0 \quad \text{(Fact 1b)} \;\; \forall i$$

$$Var[S] = E[S^2] - E[S]^2 = E[S^2] \overset{\text{def 8}}{=} E\left[\left(\sum_{i=1}^{n} \ell'(\theta; X_i)\right)^2\right]$$

$$(q_1 + q_2 + \ldots + q_n)^2 = \sum_{i=1}^{n} q_i^2 + \sum_{i \ne j} q_i q_j$$

$$= E\left[\sum_{i=1}^{n} \ell'(\theta; X_i)^2 + \sum_{i \ne j} \ell'(\theta; X_i) \, \ell'(\theta; X_j)\right]$$

If U, V are independent rv's then E[UV] = E[U] E[V]

$$= \sum_{i=1}^{n} \underbrace{E\left[\ell'(\theta; X_i)^2\right]}_{I(\theta)} + \sum_{i \ne j} \underbrace{E\left[\ell'(\theta; X_i)\right]}_{0} \underbrace{E\left[\ell'(\theta; X_j)\right]}_{0}$$

Fact 1b

$$= \underbrace{n \, E\left[\ell'(\theta; X)^2\right]}_{I_n(\theta)}$$

The "Fisher Information"