ECMA 31000: Solutions to Problem Set 4

Joe Hardwick

Due Nov 4 by 11:59PM

Question 1 (Partial Identification) Suppose the entire population of 22-25 year olds is surveyed on their preference for rock music. For each individual i, let $Y_i = 1$ if i likes rock music and $Y_i = 0$ if not. Not all individuals respond. For these individuals, Y_i is unobserved. Let X_i be the age of individual i, in years. Suppose we have the following data about the joint distribution of taste for rock music and age:

	Age in Years			
Likes Rock?	22	23	24	25
$Y = 1$	0.05	0.10	0.05	0.15
$Y = 0$	0.10	0.10	0.10	0
Y unobserved	0.05	0.10	0.15	0.05

We wish to learn $\mathrm{P}(Y = 1)$.

a) Using the data alone, what can you conclude about $\mathrm{P}(Y = 1)$?

ANS: There is a range of possible values for $\mathrm{P}(Y = 1)$, since we do not know the preferences of those who did not respond. The total proportion of individuals that did not respond is 0.35. The total proportion of individuals who report liking rock is 0.35. Therefore, $\mathrm{P}(Y = 1) \in [0.35, 0.7]$.

b) You make the assumption that preferences for rock are independent of age. Is this assumption consistent with the data? If it is, what can you conclude about $\mathrm{P}(Y = 1)$ now?

ANS: This assumption is not consistent with the data. The data reveal that

$$\frac{1}{4} \leq \mathrm{P}(Y = 1|X = 22) \leq \frac{1}{2}, \qquad \frac{3}{4} \leq \mathrm{P}(Y = 1|X = 25) \leq 1$$

However, If preference for rock is independent of age, then

$$\mathrm{P}(Y = 1|X = 22) = \mathrm{P}(Y = 1|X = 25)$$

must hold. This is clearly impossible since $\mathrm{P}(Y = 1|X = 22) \leq \frac{1}{2}$, while $\frac{3}{4} \leq \mathrm{P}(Y = 1|X = 25)$.

c) You make the assumption that older individuals are more likely to have a taste for rock music. That is, you assume $\mathrm{P}(Y = 1|X = x)$ is non-decreasing in x. Is this assumption consistent with the data? If so, what can you conclude about $\mathrm{P}(Y = 1)$ now?

ANS: This time the assumption is consistent with the data. Reasoning as in part b) we obtain:

$$\begin{aligned}
\mathrm{P}(Y=1|X=22) &\in \left[\frac{1}{4},\frac{1}{2}\right], \\
\mathrm{P}(Y=1|X=23) &\in \left[\frac{1}{3},\frac{2}{3}\right], \\
\mathrm{P}(Y=1|X=24) &\in \left[\frac{1}{6},\frac{2}{3}\right], \\
\mathrm{P}(Y=1|X=25) &\in \left[\frac{3}{4},1\right].
\end{aligned}$$

One vector of probabilities that would be consistent with both the data and assumption is

$$\begin{aligned}
\mathrm{P}(Y=1|X=22) &= \frac{1}{4}, \\
\mathrm{P}(Y=1|X=23) &= \frac{1}{3}, \\
\mathrm{P}(Y=1|X=24) &= \frac{1}{3}, \\
\mathrm{P}(Y=1|X=25) &= \frac{3}{4}.
\end{aligned}$$

Next, we use the assumption that older individuals are more likely to have a preference for rock to tighten the bounds derived above. The assumption implies

$$\begin{aligned}
\mathrm{P}(Y=1|X=22) &\in \left[\frac{1}{4},\frac{1}{2}\right], \\
\mathrm{P}(Y=1|X=23) &\in \left[\frac{1}{3},\frac{2}{3}\right], \\
\mathrm{P}(Y=1|X=24) &\in \left[\frac{1}{3},\frac{2}{3}\right], \\
\mathrm{P}(Y=1|X=25) &\in \left[\frac{3}{4},1\right].
\end{aligned}$$

Now we use the law of total probability to work out the logical bounds on $\mathrm{P}(Y=1)$. We have

$$\mathrm{P}(Y=1) = \sum_{j=22}^{25} \mathrm{P}(Y=1|X=j)\,\mathrm{P}(X=j).$$

Using our revised bounds, we obtain

$$\frac{1}{4}\cdot\frac{1}{5}+\frac{1}{3}\cdot\frac{3}{10}+\frac{1}{3}\cdot\frac{3}{10}+\frac{3}{4}\cdot\frac{1}{5} \le \mathrm{P}(Y=1) \le \frac{1}{2}\cdot\frac{1}{5}+\frac{2}{3}\cdot\frac{3}{10}+\frac{2}{3}\cdot\frac{3}{10}+1\cdot\frac{1}{5},$$

or

$$\mathrm{P}(Y=1) \in [0.4, 0.7].$$

d) A third of the 24 year olds who did not respond initially now respond. They report that they do not like rock. How does your answer to parts a) and c) change?

ANS: For part a), we simply reduce the upper bound by 0.05, so now $\mathrm{P}(Y=1) \in [0.35, 0.65]$. For part c), note that based on previous work and the new data,

$$\begin{aligned}
\mathrm{P}(Y=1|X=22) &\in \left[\frac{1}{4}, \frac{1}{2}\right], \\
\mathrm{P}(Y=1|X=23) &\in \left[\frac{1}{3}, \frac{2}{3}\right], \\
\mathrm{P}(Y=1|X=24) &\in \left[\frac{1}{3}, \frac{1}{2}\right], \\
\mathrm{P}(Y=1|X=25) &\in \left[\frac{3}{4}, 1\right].
\end{aligned}$$

Our assumption implies that we now must revise the upper bound for $\mathrm{P}(Y=1|X=23)$ down to $\frac{1}{2}$, since no more than $\frac{1}{2}$ of 24 year olds like rock. It follows that

$$\begin{aligned}
\mathrm{P}(Y=1|X=22) &\in \left[\frac{1}{4}, \frac{1}{2}\right], \\
\mathrm{P}(Y=1|X=23) &\in \left[\frac{1}{3}, \frac{1}{2}\right], \\
\mathrm{P}(Y=1|X=24) &\in \left[\frac{1}{3}, \frac{1}{2}\right], \\
\mathrm{P}(Y=1|X=25) &\in \left[\frac{3}{4}, 1\right].
\end{aligned}$$

which yields

$$\mathrm{P}(Y=1) \in [0.4, 0.6].$$

e) Now suppose <u>all</u> 24 year olds who did not respond initially now report they don't like rock. How does your answer to parts a) and c) change?

ANS: For a), we reduce the upper bound further, so $\mathrm{P}(Y=1) \in [0.35, 0.55]$. For part c), we note that the assumption that $\mathrm{P}(Y=1|X=x)$ is non-decreasing in x is no longer consistent with the data, since $\mathrm{P}(Y=1|X=23) \geq \frac{1}{3}$ but $\mathrm{P}(Y=1|X=24) = \frac{1}{6}$.

Question 2 Suppose we have an iid sample $\{X_i\}_{i=1}^n$ drawn from a distribution represented by the pdf

$$f_\theta(x) = \begin{cases} \theta x^{\theta-1} & 0 < x < 1, \\ 0 & \text{otherwise,} \end{cases}$$

for some $\theta > 0$.

a) Find a Method of Moments estimator and the Maximum Likelihood estimator of θ.

ANS: We use the first moment of X to derive a method of moments estimator. Note that

$$\mathrm{E}(X) = \int_0^1 x \cdot \theta x^{\theta-1} \mathrm{d}x = \frac{\theta}{\theta+1}.$$

Using the sample analogue principle, we get

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \frac{\hat{\theta}_{MM}}{\hat{\theta}_{MM}+1},$$

which yields

$$\hat{\theta}_{MM} = \frac{\bar{X}_n}{1-\bar{X}_n}.$$

To derive the MLE of θ, we construct the likelihood function using the fact that the X_i are iid:

$$\begin{aligned} L(\theta) &= \Pi_{i=1}^{n} f_\theta(X_i) \\ &= \theta^n \left(\Pi_{i=1}^{n} X_i\right)^{\theta-1}. \end{aligned}$$

The log-likelihood is maximized if and only if the likelihood is maximized, so we work with this instead. We have

$$l(\theta) = \ln L(\theta) = n\ln(\theta) + (\theta-1)\sum_{i=1}^{n}\ln(X_i).$$

Note that this is clearly a concave function of θ, so the FOC will suffice to find the maximizer. We have

$$l'(\theta) = \frac{n}{\theta} + \sum_{i=1}^{n}\ln(X_i).$$

Setting $l'(\theta)=0$ gives the MLE:

$$\hat{\theta}_{MLE} = -\frac{1}{\frac{1}{n}\sum_{i=1}^{n}\ln(X_i)}.$$

b) Show that both estimators are consistent.

ANS: For the method of moments estimator, note that $\bar{X}_n \xrightarrow{a.s.} \mathrm{E}(X) = \frac{\theta}{\theta+1}$ by the SLLN. Applying the continuous mapping theorem with $g(x) = \frac{x}{1-x}$ gives

$$\hat{\theta}_{MM} = \frac{\bar{X}_n}{1-\bar{X}_n} \xrightarrow{a.s.} \frac{\frac{\theta}{\theta+1}}{1-\frac{\theta}{\theta+1}} = \theta.$$

For the MLE, note that since $\{\ln(X_i)\}_{i\geq 1}$ is an iid sequence and

$$\exp\left[\mathrm{E}(\ln(X_i))\right] \leq \mathrm{E}(\exp(\ln(X_i))) = \mathrm{E}(X_i) = \frac{\theta}{1+\theta}$$

by Jensen's inequality, it follows that $\mathrm{E}(\ln(X_i))$ exists, so we can apply the SLLN to the sequence $\{\ln(X_i)\}_{i\geq 1}$ and conclude that

$$\frac{1}{n}\sum_{i=1}^{n}\ln(X_i) \xrightarrow{a.s.} \mathrm{E}(\ln(X_i)) = \int_0^1 \ln(x)\,\theta x^{\theta-1}\mathrm{d}x = -\frac{1}{\theta}.$$

where the last equality follows using integration by parts:

$$\int_0^1 \ln(x)\,\theta x^{\theta-1}\mathrm{d}x = \left[\ln(x)\,x^\theta\right]_0^1 - \int_0^1 x^{\theta-1}\mathrm{d}x = -\frac{1}{\theta}.$$

By the continuous mapping theorem with $\mathrm{g}(x) = -\frac{1}{x}$, we find that

$$\hat{\theta}_{ML} \overset{a.s.}{\to} -\frac{1}{-\frac{1}{\theta}} = \theta,$$

since $\theta > 0$.

c) Derive the asymptotic distribution of your Method of Moments estimator.

ANS: Note that since $\bar{X}_n \overset{a.s.}{\to} \mathrm{E}(X) = \frac{\theta}{\theta+1}$, and

$$\begin{aligned}
Var(X) &= \mathrm{E}\left(X^2\right) - \mathrm{E}(X)^2 \\
&= \int_0^1 x^2 \cdot \theta x^{\theta-1}\mathrm{d}x - \left(\frac{\theta}{\theta+1}\right)^2 \\
&= \frac{\theta}{\theta+2} - \left(\frac{\theta}{\theta+1}\right)^2 \\
&= \frac{\theta}{(\theta+1)^2(\theta+2)} < \infty,
\end{aligned}$$

the CLT implies

$$\sqrt{n}\left(\bar{X}_n - \frac{\theta}{\theta+1}\right) \overset{d}{\to} \mathcal{N}\left(0, \frac{\theta}{(\theta+1)^2(\theta+2)}\right).$$

Now let $g(x) = \frac{x}{1-x}$ and note that $g'(x) = \frac{1}{(1-x)^2}$. Use the delta method to conclude that

$$\sqrt{n}\left(\hat{\theta}_{MM} - \theta\right) \overset{d}{\to} \mathcal{N}\left(0, g'\left(\frac{\theta}{\theta+1}\right)^2 \cdot \frac{\theta}{(\theta+1)^2(\theta+2)}\right).$$

Noting that $g'\left(\frac{\theta}{\theta+1}\right)^2 = (1+\theta)^4$gives the result

$$\sqrt{n}\left(\hat{\theta}_{MM} - \theta\right) \overset{d}{\to} \mathcal{N}\left(0, \frac{\theta(\theta+1)^2}{\theta+2}\right).$$

d) Derive the asymptotic distribution of the maximum likelihood estimator.

ANS: First we compute $\mathrm{E}\left(\ln(X_i)^2\right)$ using integration by parts twice. Note that:

$$\begin{aligned}
\mathrm{E}\left(\ln(X_i)^2\right) &= \int_0^1 \ln(x)^2\,\theta x^{\theta-1}\mathrm{d}x \\
&= \left[\ln(x)^2\,x^\theta\right]_0^1 - 2\int_0^1 x^{\theta-1}\ln(x)\,\mathrm{d}x
\end{aligned}$$

$$
\begin{aligned}
&= 0 - 2\left(\left[\frac{x^{\theta}\ln(x)}{\theta}\right]_0^1 - \int_0^1 \frac{x^{\theta-1}}{\theta}\mathrm{d}x\right)\\
&= 0 - 0 + \frac{2}{\theta^2}.
\end{aligned}
$$

Next, since $\frac{1}{n}\sum_{i=1}^{n}\ln(X_i) \xrightarrow{a.s.} \mathrm{E}(\ln(X_i)) = -\frac{1}{\theta}$, and

$$
Var(\ln(X_i)) = \mathrm{E}\left(\ln(X_i)^2\right) - \mathrm{E}(\ln(X_i))^2 = \frac{1}{\theta^2},
$$

it follows by the CLT that

$$
\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\ln(X_i) - \left(-\frac{1}{\theta}\right)\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{\theta^2}\right).
$$

Finally, we employ the delta method with $g(x) = -\frac{1}{x}$ to conclude that

$$
\sqrt{n}\left(\hat{\theta}_{ML} - \theta\right) \xrightarrow{d} \mathcal{N}\left(0, g'\left(-\frac{1}{\theta}\right)^2 \cdot \frac{1}{\theta^2}\right).
$$

Noting that

$$
g'\left(-\frac{1}{\theta}\right)^2 = \frac{1}{\left(-\frac{1}{\theta}\right)^2} = \theta^4
$$

gives the result

$$
\sqrt{n}\left(\hat{\theta}_{ML} - \theta\right) \xrightarrow{d} \mathcal{N}\left(0, \theta^2\right).
$$

e) Compute the information matrix $I(\theta)$ defined in Lecture 9. Compare your answer to the solution in d). Compare the asymptotic variances of the MLE and MoM estimator.

ANS: The information matrix is given by

$$
I(\theta) = -\mathrm{E}\left(\frac{\partial^2}{\partial\theta^2}\ln f_\theta(X)\right) = -\mathrm{E}\left(-\frac{1}{\theta^2}\right) = \frac{1}{\theta^2}.
$$

Therefore $I(\theta)^{-1} = \theta^2$, which is the asymptotic variance of the Maximum Likelihood estimator derived in part d). The MLE has lower asymptotic variance than the MoM estimator, which is not as efficient asymptotically.

Question 3 a) Let $\{X_i\}_{i\geq 1}$ be an iid sequence with $\mathrm{E}(X_i) = \mu$ and $Var(X_i) = \sigma^2 > 0$. Calculate MSE of the (biased) estimate of σ^2 given by

$$
S_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2.
$$

Decompose your result into $Var\left(S_n^2\right) + Bias\left(S_n^2\right)^2$. Use your result to deduce the MSE of $\tilde{S}_n^2 =$

$\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}_n\right)^2$ stated in class.

ANS: From Lecture 6, note that

$$\begin{aligned}\left(S_n^2-\sigma^2\right)^2 &= \left(\frac{1}{n}\sum_{i=1}^{n}\left[(X_i-\mu)^2-\sigma^2\right]-\left(\bar{X}_n-\mu\right)^2\right)^2\\ &= \left(\frac{1}{n}\sum_{i=1}^{n}\left[(X_i-\mu)^2-\sigma^2\right]\right)^2\\ &\quad -2\left(\bar{X}_n-\mu\right)^2\frac{1}{n}\sum_{i=1}^{n}\left[(X_i-\mu)^2-\sigma^2\right]\\ &\quad +\left(\bar{X}_n-\mu\right)^4.\end{aligned}$$

Note that the first term is a sum of mean zero iid variables with variance $\mathrm{E}\left(X_i-\mu\right)^4-\sigma^4$, so

$$\mathrm{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n}\left[(X_i-\mu)^2-\sigma^2\right]\right)^2\right]=\frac{1}{n^2}\sum_{i=1}^{n}Var\left((X_i-\mu)^2-\sigma^2\right)=\frac{\mathrm{E}\left(X_i-\mu\right)^4-\sigma^4}{n}.$$

For the second term, expanding gives

$$\left(\bar{X}_n-\mu\right)^2\frac{1}{n}\sum_{i=1}^{n}\left[(X_i-\mu)^2-\sigma^2\right]=\frac{1}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\left(X_j-\mu\right)\left(X_k-\mu\right)\left[(X_i-\mu)^2-\sigma^2\right].$$

The only nonzero terms in the expectation occur when $i=j=k$ (n times), yielding

$$-2\mathrm{E}\left(\left(\bar{X}_n-\mu\right)^2\frac{1}{n}\sum_{i=1}^{n}\left[(X_i-\mu)^2-\sigma^2\right]\right)=-2\cdot\frac{\mathrm{E}\left(X_i-\mu\right)^4-\sigma^4}{n^2}.$$

For the final term, expanding gives

$$\left(\bar{X}_n-\mu\right)^4=\frac{1}{n^4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\left(X_i-\mu\right)\left(X_j-\mu\right)\left(X_k-\mu\right)\left(X_l-\mu\right).$$

The only nonzero terms in the expectation occur when $i=j=k=l$ (n times), or $i=j\neq k=l; i=k\neq j=l; i=l\neq j=k$ ($3n\left(n-1\right)$ times), giving

$$\mathrm{E}\left(\left(\bar{X}_n-\mu\right)^4\right)=\frac{\mathrm{E}\left(X_i-\mu\right)^4}{n^3}+\frac{3\left(n-1\right)}{n^3}\sigma^4.$$

Adding terms gives

$$\begin{aligned}MSE\left(S_n^2\right)&=\mathrm{E}\left(X_i-\mu\right)^4\left(\frac{1}{n}-\frac{2}{n^2}+\frac{1}{n^3}\right)+\sigma^4\left(-\frac{1}{n}+\frac{2}{n^2}+\frac{3\left(n-1\right)}{n^3}\right)\\ &=\mathrm{E}\left(X_i-\mu\right)^4\frac{\left(n-1\right)^2}{n^3}-\sigma^4\frac{n^2-5n+3}{n^3}\end{aligned}$$

$$
\begin{aligned}
&= \mathrm{E}\left(X_i-\mu\right)^4 \frac{(n-1)^2}{n^3} - \sigma^4 \frac{n^2-4n+3}{n^3} + \frac{\sigma^4}{n} \\
&= \underbrace{\mathrm{E}\left(X_i-\mu\right)^4 \frac{(n-1)^2}{n^3} - \sigma^4 \frac{(n-1)(n-3)}{n^3}}_{Var(S_n^2)} + \underbrace{\frac{\sigma^4}{n^2}}_{Bias(S_n^2)^2}
\end{aligned}
$$

where the last equality was derived by recalling that

$$
Bias\left(S_n^2\right) = \frac{\sigma^2}{n}.
$$

Since $\tilde{S}_n^2 = \frac{n}{n-1} S_n^2$ and $\tilde{S}_n^2$ is unbiased, we deduce that

$$
\begin{aligned}
MSE\left(\tilde{S}_n^2\right) &= Var\left(\tilde{S}_n^2\right) \\
&= \frac{n^2}{(n-1)^2} Var\left(S_n^2\right) \\
&= \frac{\mathrm{E}\left(X_i-\mu\right)^4}{n} - \frac{\sigma^4 (n-3)}{n(n-1)}.
\end{aligned}
$$

b) Consider the family of estimators $\left\{k\tilde{S}_n^2 : k \in \mathbb{R}\right\}$ of σ^2. Calculate the value k^* which minimizes the MSE of $k\tilde{S}_n^2$, and show that for any distribution F with $0 < Var(F) < \infty$,

$$
k^* \leq \frac{n^2-n}{n^2-n+2}.
$$

Show, in particular, that for the normal distribution,

$$
k^* = \frac{n-1}{n+1}.
$$

Hint: Use the optimal value of k derived in Lecture 8.

ANS: From lecture 8, the optimal value of k is given by

$$
k^* = \frac{\sigma^4}{\frac{\mathrm{E}(X_i-\mu)^4}{n} - \frac{\sigma^4(n-3)}{n(n-1)} + \sigma^4} \leq \frac{n\sigma^4}{\sigma^4\left(1 - \frac{n-3}{n-1} + n\right)} = \frac{n^2-n}{n^2-n+2},
$$

where the first inequality follows by Jensen's inequality, since

$$
\mathrm{E}\left(X_i-\mu\right)^4 \geq \sigma^4.
$$

For the normal distribution,

$$
k^* = \frac{\sigma^4}{\frac{\mathrm{E}(X_i-\mu)^4}{n} - \frac{\sigma^4(n-3)}{n(n-1)} + \sigma^4} = \frac{\sigma^4}{\frac{3\sigma^4}{n} - \frac{\sigma^4(n-3)}{n(n-1)} + \sigma^4} = \frac{n-1}{n+1}.
$$

Question 4 Suppose you observe an iid sample $\{X_i\}_{i=1}^n$, where $X_i \sim U[0,\theta]$.

a) Compute $\mathrm{E}\left(X^k\right)$, and derive the method of moments estimator ($\hat{\theta}_k$) of θ based on this calculation.

ANS: We have

$$\mathrm{E}\left(X^k\right) = \int_0^\theta \frac{x^k}{\theta}\mathrm{d}x = \left[\frac{x^{k+1}}{(k+1)\,\theta}\right]_0^\theta = \frac{\theta^k}{k+1}.$$

Replacing the expectation with the sample analogue and the true parameter with the estimate gives

$$\frac{1}{n}\sum_{i=1}^n X_i^k = \frac{\hat{\theta}_k^k}{k+1} \implies \hat{\theta}_k = \left(\frac{k+1}{n}\sum_{i=1}^n X_i^k\right)^{\frac{1}{k}}.$$

b) Show that $\hat{\theta}_{MM}^k \overset{a.s.}{\to} \theta$. Is $\hat{\theta}_{MM}^k$ an unbiased estimator of θ?

ANS: By the strong law of large numbers and the continuous mapping theorem with $g(x) = ((k+1)\,x)^{1/k}$, we obtain

$$\left(\frac{k+1}{n}\sum_{i=1}^n X_i^k\right)^{\frac{1}{k}} \overset{a.s.}{\to} \left((k+1)\,\frac{\theta^k}{k+1}\right)^{\frac{1}{k}} = \theta.$$

However, $\hat{\theta}_{MM}^k$ is biased downward, since by Jensen's inequality:

$$\mathrm{E}\left[\left(\frac{k+1}{n}\sum_{i=1}^n X_i^k\right)^{\frac{1}{k}}\right] \le \left[\mathrm{E}\left(\frac{k+1}{n}\sum_{i=1}^n X_i^k\right)\right]^{\frac{1}{k}} = \theta.$$

c) Find the asymptotic distribution of $\hat{\theta}_{MM}^k$. How does it vary with k?

ANS: Our goal is to apply the CLT to a sample average and then get at the k-th root via the delta method. Note that

$$\begin{aligned}\sqrt{n}\left(\hat{\theta}_k^k - \theta^k\right) &= (k+1)\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n X_i^k - \frac{\theta^k}{k+1}\right)\\ &\overset{d}{\to} \mathcal{N}\left(0,(k+1)^2\,Var\left(X_i^k\right)\right).\end{aligned}$$

by the CLT and continuous mapping theorem with $g(x) = (k+1)\,x$. Note that

$$\begin{aligned}Var\left(X_i^k\right) &= \mathrm{E}\left(X_i^{2k}\right) - \mathrm{E}\left(X_i^k\right)^2\\ &= \frac{\theta^{2k}}{2k+1} - \frac{\theta^{2k}}{(k+1)^2}\\ &= \frac{k^2\theta^{2k}}{(2k+1)(k+1)^2},\end{aligned}$$

so

$$\sqrt{n}\left(\hat{\theta}_k^k - \theta^k\right) \overset{d}{\to} \mathcal{N}\left(0,\frac{k^2\theta^{2k}}{2k+1}\right).$$

Now we apply the delta method, with $h(x) = x^{1/k}$. Since

$$h'(x) = \frac{x^{(1-k)/k}}{k};$$
$$\left(h'\left(\theta^k\right)\right)^2 = \frac{\theta^{2-2k}}{k^2};$$

we conclude that

$$\begin{aligned}\sqrt{n}\left(h\left(\hat{\theta}_k^k\right) - h\left(\theta^k\right)\right) &= \sqrt{n}\left(\hat{\theta}_k - \theta\right) \\ &\xrightarrow{d} \mathcal{N}\left(0, \left(h'\left(\theta^k\right)\right)^2 \frac{k^2\theta^{2k}}{2k+1}\right) \\ &\stackrel{d}{=} \mathcal{N}\left(0, \frac{\theta^2}{2k+1}\right).\end{aligned}$$

The asymptotic variance is decreasing in k.

d) Compute the Maximum Likelihood Estimator $\left(\hat{\theta}_{ML}\right)$ of θ. How does it relate to the method of moments estimator based on computing $\mathrm{E}\left(X^k\right)$? (Think about large k).

ANS: Note that the likelihood given an iid sample $X^n = \{X_i\}_{i=1}^n$ is:

$$\begin{aligned}p_\theta\left(X^n\right) &= \Pi_{i=1}^n\left[\frac{1}{\theta}\mathbf{1}\left(X_i \leq \theta\right)\right] \\ &= \frac{1}{\theta^n}\mathbf{1}\left(X_i \leq \theta \text{ for all } i\right).\end{aligned}$$

This function is decreasing in θ, but is 0 if any X_i is larger than θ, so the best that can be achieved is

$$\hat{\theta}_{ML} = \max_{i \leq n} X_i.$$

Next, note that

$$\hat{\theta}_k = \left(\frac{k+1}{n}\sum_{i=1}^n X_i^k\right)^{\frac{1}{k}} = \left(\frac{k+1}{n}\right)^{\frac{1}{k}}\left(\sum_{i=1}^n X_i^k\right)^{\frac{1}{k}}.$$

It can be shown that for any $n \geq 1$:

$$\lim_{k\to\infty}\left(\frac{k+1}{n}\right)^{\frac{1}{k}} = 1$$

and

$$\lim_{k\to\infty}\left(\sum_{i=1}^n X_i^k\right)^{\frac{1}{k}} = \max_{i \leq n} X_i,$$

so

$$\lim_{k\to\infty}\hat{\theta}_k = \hat{\theta}_{ML}.$$

Based on this result, we may already suspect that $\hat{\theta}_{ML}$ will converge in probability to θ at a faster rate than $\hat{\theta}_k$, which converges at rate $\sqrt{n}$, because the asymptotic distribution of $\hat{\theta}_k$ has a very small variance when k is large.

e) Show that $\hat{\theta}_{ML}$ is a consistent estimator of θ and compute its asymptotic distribution.

ANS: This was derived in problem set 2. We have

$$\begin{aligned}
\mathrm{P}\left(\max_{i\leq n} X_i \leq x\right) &= \mathrm{P}\left(\cap_{i=1}^{n}\left\{X_i \leq x\right\}\right) \\
&= \Pi_{i=1}^{n}\mathrm{P}\left(X_i \leq x\right) \\
&= \begin{cases} 0 & \text{if } x < 0, \\ \left(\frac{x}{\theta}\right)^n & \text{if } 0 \leq x \leq \theta, \\ 1 & \text{if } x > \theta. \end{cases}
\end{aligned}$$

It follows that

$$F_{X_n}(x) \to \begin{cases} 0 & x < \theta \\ 1 & x \geq 1 \end{cases},$$

which is the distribution of the constant θ. Since convergence in distribution to a constant implies convergence in probability, the result follows.

Now we derive the asymptotic distribution. Since we cannot use the CLT here, we analyse the distribution of $n\left(\theta - \max_{i\leq n} X_i\right)$ directly.

Note that for $x \geq 0$:

$$\begin{aligned}
\mathrm{P}\left(n\left(\theta - \max_{i\leq n} X_i\right) \leq x\right) &= \mathrm{P}\left(\theta - \frac{x}{n} \leq \max_{i\leq n} X_i\right) \\
&= 1 - \mathrm{P}\left(\max_{i\leq n} X_i \leq \theta - \frac{x}{n}\right) \\
&= 1 - \left(1 - \frac{x/\theta}{n}\right)^n \\
&\to 1 - \exp\left(-\frac{x}{\theta}\right).
\end{aligned}$$

Indeed, we see that $\hat{\theta}_{ML}$ converges in probability to θ at rate n.

f) Is $\hat{\theta}_{ML}$ unbiased? If not, construct an unbiased estimate of θ based on $\hat{\theta}_{ML}$. Call it $\tilde{\theta}$.

ANS: $\hat{\theta}_{ML}$ is biased. To see this, note that since

$$F_{\hat{\theta}_{ML}}(x) = \begin{cases} 0 & \text{if } x < 0, \\ \left(\frac{x}{\theta}\right)^n & \text{if } 0 \leq x \leq \theta, \\ 1 & \text{if } x > \theta. \end{cases}$$

the pdf of $\hat{\theta}_{ML}$ on $[0,\theta]$ is given by

$$f_{\hat{\theta}_{ML}}(x) = F'_{\hat{\theta}_{ML}}(x) = \frac{nx^{n-1}}{\theta^n}.$$

It follows that

$$\begin{aligned}
\mathrm{E}\left(\hat{\theta}_{ML}\right) &= \int_0^\theta x \cdot \frac{nx^{n-1}}{\theta^n}\mathrm{d}x \\
&= \frac{n}{n+1}\left[\frac{x^{n+1}}{\theta^n}\right]_0^\theta \\
&= \frac{n}{n+1}\cdot\theta.
\end{aligned}$$

An unbiased estimator is therefore given by

$$\frac{n+1}{n}\hat{\theta}_{ML}.$$

g) Compute the MSE of $\hat{\theta}_{ML}$ and $\tilde{\theta}$. Which is lower?

ANS: From part f), the bias of $\hat{\theta}_{ML}$ is given by

$$\mathrm{E}\left(\hat{\theta}_{ML} - \theta\right) = -\frac{\theta}{n+1}.$$

The variance of $\hat{\theta}_{ML}$ is given by

$$\begin{aligned}
Var\left(\hat{\theta}_{ML}\right) &= \mathrm{E}\left(\hat{\theta}_{ML}^2\right) - \mathrm{E}\left(\hat{\theta}_{ML}\right)^2 \\
&= \int_0^\theta x^2 \cdot \frac{nx^{n-1}}{\theta^n}\mathrm{d}x - \left(\frac{n\theta}{n+1}\right)^2 \\
&= \theta^2\left(\frac{n}{n+2} - \frac{n^2}{(n+1)^2}\right) \\
&= \theta^2\left(\frac{n(n+1)^2 - n^2(n+2)}{(n+2)(n+1)^2}\right) \\
&= \frac{\theta^2 n}{(n+2)(n+1)^2}.
\end{aligned}$$

The MSE is therefore given by

$$MSE\left(\hat{\theta}_{ML}\right) = \frac{\theta^2 n}{(n+2)(n+1)^2} + \frac{\theta^2}{(n+1)^2} = \frac{2\theta^2}{(n+1)(n+2)}.$$

Based on the variance calculation, observe that

$$MSE\left(\tilde{\theta}\right) = Var\left(\tilde{\theta}\right)$$

$$= \left(\frac{n+1}{n}\right)^2 Var\left(\hat{\theta}_{ML}\right)$$
$$= \frac{\theta^2}{n\left(n+2\right)}.$$

We note that the MSE of $\tilde{\theta}$ is strictly lower than that of $\hat{\theta}_{ML}$ when $n > 1$.

h) Use your answer to part g) to conclude that $\hat{\theta}_{ML} \overset{a.s.}{\to} \theta$.

ANS: Let $\hat{\theta}_{ML}^n = \max_{i \leq n} X_i$. We need to show that

$$\lim_{n\to\infty} \mathrm{P}\left(\cup_{k\geq n}\left\{\left|\hat{\theta}_{ML}^n - \theta\right| > \epsilon\right\}\right) = 0.$$

Note that

$$\begin{aligned}
\mathrm{P}\left(\cup_{k\geq n}\left\{\left|\hat{\theta}_{ML}^k - \theta\right| > \epsilon\right\}\right) &\leq \sum_{k=n}^{\infty} \mathrm{P}\left(\left|\hat{\theta}_{ML}^k - \theta\right| > \epsilon\right) \\
&\leq \sum_{k=n}^{\infty} \frac{\mathrm{E}\left(\left|\hat{\theta}_{ML}^k - \theta\right|^2\right)}{\epsilon^2} \\
&= \sum_{k=n}^{\infty} \frac{2\theta^2}{\left(k+1\right)\left(k+2\right)\epsilon^2} \\
&\leq \frac{2\theta^2}{\epsilon^2}\sum_{k=n}^{\infty}\frac{1}{k^2} \to 0
\end{aligned}$$

as $n \to \infty$.

Question 5 (Not quite a sample average) Let $\{X_i\}_{i=1}^n$ be an iid sample, with $X_i \sim U\left[0,\theta\right]$ for some $\theta > 0$. Consider the following estimator of θ:

$$\hat{\theta}_n = \left(\Pi_{i=1}^n X_i\right)^{\frac{1}{n}}.$$

a) Is $\hat{\theta}_n$ a consistent estimator of θ? If not, find a function f such that $\tilde{\theta}_n = f\left(\hat{\theta}_n\right)$ is a consistent estimator of θ.

ANS: No. To analyse the probability limit of $\hat{\theta}_n$, we first take logs:

$$\ln\left(\hat{\theta}_n\right) = \frac{1}{n}\sum_{i=1}^n \ln\left(X_i\right).$$

By the SLLN, it follows that

$$\begin{aligned}
\ln\left(\hat{\theta}_n\right) &\overset{a.s.}{\to} \mathrm{E}\left(\ln\left(X_i\right)\right) \\
&= \int_0^\theta \frac{\ln\left(x\right)}{\theta}\mathrm{d}x.
\end{aligned}$$

$$
\begin{aligned}
&= \left[\frac{x\ln(x)}{\theta}\right]_0^\theta - \int_0^\theta \frac{1}{\theta}\mathrm{d}x \\
&= \ln(\theta) - 1.
\end{aligned}
$$

By the continuous mapping theorem,

$$\hat{\theta}_n = \exp\left(\ln\left(\hat{\theta}_n\right)\right) \overset{a.s.}{\to} \exp\left(\ln(\theta) - 1\right) = \frac{\theta}{e}.$$

Therefore, we set

$$\tilde{\theta} = f\left(\hat{\theta}_n\right) = e\hat{\theta}_n \overset{a.s.}{\to} \theta.$$

b) Find the asymptotic distribution of $\tilde{\theta}_n$.

ANS: When establishing consistency, we studied a sample average

$$\frac{1}{n}\sum_{i=1}^n \ln(X_i).$$

We now apply the CLT to this sample average:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \left(\ln(X_i) - [\ln(\theta) - 1]\right) \overset{d}{\to} \mathcal{N}\left(0, Var\left(\ln(X_i)\right)\right).$$

We compute

$$
\begin{aligned}
\mathrm{E}\left(\ln(X_i)^2\right) &= \int_0^\theta \frac{\ln(x)^2}{\theta}\mathrm{d}x \\
&= \left[\frac{x\ln(x)^2}{\theta}\right]_0^\theta - \int_0^\theta x \cdot \frac{2\ln(x)}{x\theta}\mathrm{d}x \\
&= \ln(\theta)^2 - 2\int_0^\theta \frac{\ln(x)}{\theta}\mathrm{d}x \\
&= \ln(\theta)^2 - 2\ln(\theta) + 2,
\end{aligned}
$$

from which it follows that

$$Var\left(\ln(X_i)\right) = 1.$$

Therefore,

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n \left(\ln(X_i) - [\ln(\theta) - 1]\right)\right) \overset{d}{\to} \mathcal{N}(0, 1).$$

Finally, we apply the delta method with $g(x) = \exp(x+1)$ to obtain

$$\sqrt{n}\left(\tilde{\theta}_n - \theta\right) \overset{d}{\to} \mathcal{N}\left(0, g'\left(\ln(\theta) - 1\right)^2\right) \overset{d}{=} \mathcal{N}\left(0, \theta^2\right).$$

Question 6 Suppose you observe a random sample of $\{Y_i, X_i\}_{i=1}^n$ where

$$Y_i = \begin{cases} 1 & \text{if individual } i \text{ is employed,} \\ 0 & \text{if individual } i \text{ is unemployed.} \end{cases}$$

is a scalar outcome indicating employment status and X_i is a $(K \times 1)$ vector of observable characteristics. You model the probability of employment conditional on X as follows:

$$\mathrm{P}\left(Y = 1 | X = x\right) = F\left(x'\beta\right),$$

for some known distribution function F and unknown parameter β. Write down the conditional log likelihood of the observed sample, as a function of β.

ANS: The pmf of the conditional distribution of Y given X is

$$f_\beta\left(y|x\right) = F\left(x'\beta\right)^y \left(1 - F\left(x'\beta\right)\right)^{1-y}.$$

Given an iid sample $\{Y_i, X_i\}_{i=1}^n$, the conditional likelihood is given by the joint pmf of the observations Y, conditional on X, evaluated at the sample points:

$$l_n\left(\theta\right) = \Pi_{i=1}^n F\left(X_i'\beta\right)^{Y_i} \left(1 - F\left(X_i'\beta\right)\right)^{1-Y_i}.$$

The conditional log-likelihood is given by

$$L_n\left(\theta\right) := \frac{1}{n} \ln\left(l_n\left(\theta\right)\right) = \frac{1}{n} \sum_{i=1}^n Y_i \ln\left(F\left(X_i'\beta\right)\right) + \left(1 - Y_i\right) \ln\left(1 - F\left(X_i'\beta\right)\right).$$

If F is the CDF of a standard normal variable, then the model is called a probit model. If F is the logistic CDF $F\left(s\right) = \frac{1}{1+\exp(-s)}$,then it is a logit model.