MAST20005/MAST90058: Assignment 1

Due date: 11am, Friday 27 August 2021

Instructions: See the LMS for the full instructions, including the submission policy and how to submit your assignment. Remember to submit early and often: multiple submission are allowed, we will only mark your final one. Late submissions will receive **zero** marks.

Problems:

1. **(R)** In week 1 we did a 'fun quiz' to predict how long the border between NSW and Victoria will be closed. The (anonymised) responses to this are in the file `quiz.txt`. Please answer the following questions using the data in that file.

   (a) Give basic descriptive statistics for these data and produce a box plot. Briefly comment on the center, spread and shape of the distribution.

   (b) Assuming a gamma distribution for these data, compute maximum likelihood estimates for the parameters.

   (c) Draw a density histogram and superimpose a pdf for a gamma distribution using the estimated parameters.

   (d) Draw a QQ plot to compare the data against the fitted gamma distribution. Include a reference line. Comment on the fit of the model to the data.

2. A discrete random variable X has the following pmf:

x	1	2	3
$p(x)$	θ^2	$2\theta(1-\theta)$	$(1-\theta)^2$

   A random sample of size $n = 20$ produced the following observations:

   1, 1, 2, 3, 1, 2, 1, 3, 2, 2, 2, 1, 3, 1, 3, 1, 1, 2, 1, 2.

   For each of the following quantities, derive a general formula and, where applicable, calculate it using the given data.

   (a) i. Find $\mathbb{E}(X)$ and $\mathrm{var}(X)$.
   ii. Find the method of moments estimator and estimate of θ.
   iii. Find the standard error of this estimate.

   (b) Let F_1, F_2 and F_3 denote the sample frequencies of 1, 2 and 3, respectively.
   i. Find the likelihood function in terms of F_1, F_2 and F_3.
   ii. Find that the maximum likelihood estimator and estimate of θ.
   iii. Find the variance of this estimator.
   (*Hint:* write the estimator in terms of the sample mean.)

3. Let $X \sim \mathrm{Unif}(0, \theta)$, a continuous uniform distribution with an unknown endpoint θ.

   (a) Suppose we have a single observation on X.
   i. Find the method of moments estimator (MME) for θ and derive its mean and variance.
   ii. Find the maximum likelihood estimator (MLE) for θ and derive its mean and variance.

(b) The *mean square error* (MSE) of an estimator is defined as $\text{MSE}(\hat{\theta}) = \mathbb{E}\left[\left(\hat{\theta} - \theta\right)^2\right]$.

i. Let $\text{bias}(\hat{\theta}) = \mathbb{E}(\hat{\theta}) - \theta$. Show that,

$$\text{MSE}(\hat{\theta}) = \text{var}(\hat{\theta}) + \text{bias}(\hat{\theta})^2.$$

ii. Compare the MME and MLE from above in terms of their mean square errors.

iii. Find an estimator with smaller MSE than either of the above estimators.

(c) Suppose we have a random sample of size n from X.

i. Find the MME and derive its mean, variance and MSE.

ii. Find the MLE and derive its mean, variance and MSE.

iii. Consider the estimator $a\hat{\theta}$ where $\hat{\theta}$ is the MLE. Find a that minimises the MSE.

Some information that might be useful:

$$\mathbb{E}\left(X_{(1)}\right) = \frac{\theta}{n+1}, \quad \mathbb{E}\left(X_{(1)}^2\right) = \frac{2\theta^2}{(n+1)(n+2)}, \quad \mathbb{E}\left(X_{(n)}\right) = \frac{n\theta}{n+1}, \quad \mathbb{E}\left(X_{(n)}^2\right) = \frac{n\theta^2}{n+2}$$

4. **(R)** We have a random sample of size 10 from a normal distribution. We wish to estimate the population mean. Damjan suggests taking the average of the sample minimum and sample maximum. Julia proposes using the sample median. Martina thinks we should use the sample mean instead. Do a simulation in R to compare these three estimators in terms of their bias and variance. Include a side-by-side boxplot that compares their sampling distributions.

5. Let X_1, X_2, X_3, X_4 be iid rvs with $\mathbb{E}(X_i) = \mu$ and $\text{var}(X_i) = \sigma^2 > 0$, for $i = 1, 2, 3, 4$. Consider the following four estimators of μ:

$$T_1 = \frac{1}{3}(X_1 + X_2) + \frac{1}{6}(X_3 + X_4) \qquad T_2 = \frac{1}{6}(X_1 + 2X_2 + 3X_3 + 4X_4)$$
$$T_3 = \frac{1}{4}(X_1 + X_2 + X_3 + X_4) \qquad T_4 = \frac{1}{3}(X_1 + X_2 + X_3) + \frac{1}{4}X_4^2$$

(a) Which of these estimates are unbiased? Show your working.

(b) Among the unbiased estimators, which one has the smallest variance?