MAST20005/MAST90058: Assignment 1

Due date: 11am, Friday 30 August 2019

Instructions: Questions labelled with '**(R)**' require use of R. Please provide appropriate R commands and their output, along with sufficient explanation and interpretation of the output to demonstrate your understanding. **Such R output should be presented in an integrated form together with your explanations; do not attach them as separate sheets.** All other questions should be completed without reference to any R commands or output, except for looking up quantiles of distributions where necessary. Make sure you give enough explanation so your tutor can follow your reasoning if you happen to make a mistake. Please also try to be as succinct as possible. Each assignment will include marks for good presentation and for attempting all problems.

Problems:

1. **(R)** Let X be a random variable representing distance travelled (in kilometers) until a tire is worn out. The following are 16 observations of X:

```
41300  40300  43200  41100  39300  42100  42700  41300
38900  41200  44600  42300  40700  43500  39800  40400
```

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

   (b) Assuming a normal distribution, compute maximum likelihood estimates for the parameters.

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

   (d) Draw a QQ plot to compare the data against the fitted normal 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.

   (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 $X \sim \text{Unif}(0, \theta)$, a 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. Let $X_1, \ldots, X_n$ be a random sample from the lognormal distribution, $\text{Lognormal}(\mu, \lambda)$, whose pdf is:
$$f(x \mid \mu, \lambda) = \frac{1}{x\sqrt{2\pi\lambda}} \exp\left\{-\frac{(\ln x - \mu)^2}{2\lambda}\right\}, \quad x > 0.$$

   (a) Show that the MLE of μ and λ are $\hat{\mu} = \frac{1}{n}\sum_{i=1}^n \ln X_i$ and $\hat{\lambda} = \frac{1}{n}\sum_{i=1}^n (\ln X_i - \hat{\mu})^2$.

   (b) It is known that $\ln X_i \sim \text{N}(\mu, \lambda)$. Derive a $100 \cdot (1 - \alpha)\%$ CI for λ.

   (c) **(R)** Consider the following dataset:

   ```
   0.27, 3.30, 4.58,  2.61, 0.38, 3.77, 1.11, 1.15, 4.11, 2.10,
   0.07, 1.74, 2.11, 12.79, 1.85, 0.30, 0.34, 1.31, 0.14, 0.74
   ```

      i. Assuming a lognormal distribution is an appropriate model for these data, compute the maximum likelihood estimate of λ and give a 95% CI.

      ii. Draw a QQ plot to compare these data to the fitted lognormal distribution, $\text{Lognormal}(\hat{\mu}, \hat{\lambda})$. Is this model appropriate for these data?

   Hint: Quantiles of the lognormal distribution can be computed using the `qlnorm()` function.

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?