MATH50011 Statistical Modelling 1
Midterm Solutions

1. (a) Let X be a random variable with finite variance. Show that $Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$. (3 marks)

   Solution: We have

$$Var(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2]$$
$$= \mathbb{E}[X^2] - 2\mathbb{E}[X]^2 + \mathbb{E}[X]^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

   (b) Let γ be an unknown parameter belonging to a parameter space $\Gamma \subset \mathbb{R}$. Let S_n, $n \in \mathbb{N}$, be a sequence of estimators for γ. What does it mean for the sequence of estimators S_n, $n \in \mathbb{N}$, to be asymptotically unbiased for γ? (2 marks)

   Solution: It means that $\lim_{n\to\infty} \mathbb{E}[S_n] = \gamma$, *for every* $\gamma \in \Gamma$.

   (c) Let T and S be estimators of $\theta \in \mathbb{R}$. Let S be unbiased. Assume that $Var(T) < Var(S) - bias(T)$ for all $\theta \in \mathbb{R}$. Can we conclude that $MSE(T) < MSE(S)$ for all $\theta \in \mathbb{R}$? Motivate your answer (note that MSE stands for mean square error). (2 marks)

   Solution: No, we cannot conclude it because $MSE(T) = Var(T) + bias(T)^2$ *and* $MSE(S) = Var(S)$, *and so to conclude that* $MSE(T) < MSE(S)$ *we would need that* $Var(T) < Var(S) - bias(T)^2$ *but we only have* $Var(T) < Var(S) - bias(T)$.

   (d) Consider a sequence of iid random variables $X_1, X_2, ...$ with unknown parameter $\theta \in \mathbb{R}$. Assume that all the regularity conditions hold and assume that the Fisher information of X_1 is strictly positive. Let T_n, $n \in \mathbb{N}$, be sequence of unbiased estimators for $\theta \in \mathbb{R}$ such that $Var(T_n)$ attains the Cramer-Rao lower bound for every $n \in \mathbb{N}$. Is T_n always a consistent estimator of θ? Motivate your answer. (2 marks)

   Solution: Yes, it is always consistent. Indeed, $Var_\theta(T_n) = \frac{1}{nI_{f(1)}(\theta)}$ *which goes to 0 as* $n \to \infty$. *Since* T_n *is unbiased and* $Var_\theta(T_n) \to 0$, *by a result from lectures we conclude that* T_n *is consistent.*

   (e) Explain the likelihood ratio test, namely the hypothesis test based on the likelihood ratio statistics. In particular, your answer should include the definition of the relevant test statistics and the definition of the relevant critical value that determines the rejection region. (2 marks)

   Solution: See lecture notes.

2. Let X be a Bernoulli random variable with unknown parameter $\theta \in (0, 1)$, that is $P_\theta(X = 1) = \theta$ and $P_\theta(X = 0) = 1 - \theta$. Consider the random interval:

$$[L, U] = \begin{cases} [0, 1-\alpha], & \text{for } X = 0 \\ [\alpha, 1], & \text{for } X = 1. \end{cases}$$

   (a) Is $[L, U]$ a confidence interval (CI) for every $\alpha \in [0, 1]$? Motivate your answer. (2 marks)

   Solution: First, observe that.

$$P(\theta \in [L, U]) = P(\theta \in [L, U], X = 0) + P(\theta \in [L, U], X = 1),$$

Hence, we see that $[L, U]$ is indeed a $1-\alpha$ CI for $\alpha \leq 1/2$ because

$$P_\theta(\theta \in [L, U]) = \begin{cases} P_\theta(X=0) = 1-\theta \geq 1-\alpha & \text{for } \theta < \alpha, \\ 1 & \text{for } \alpha \leq \theta \leq 1-\alpha, \\ P_\theta(X=1) = \theta \geq 1-\alpha & \text{for } \theta > 1-\alpha. \end{cases}$$

However, for $\alpha > 1/2$ we have that $P_\theta(\theta \in [L, U]) = 0$ for $\theta \in (1-\alpha, \alpha)$ which shows that $[L, U]$ cannot be a CI for θ. Note that an acceptable solution is just to show that $[L, U]$ is not a CI for some $\alpha \in [0, 1]$.

(b) Assume that we observe $X = 1$ and consider the hypothesis $H_0 : 0.05 \leq \theta \leq 0.1$ vs $H_1 : \theta \in (0, 0.05) \cup (0.1, 1)$. Which are the values of $\alpha \in [0, 1/3]$ for which we reject H_0? Motivate your answer. (2 marks)

Solution: In this case $\Theta_0 = [0.05, 1]$. Since we observe $X = 1$ the observed CI is $[\alpha, 1]$, we reject H_0 when there is no intersection between Θ_0 and the observed CI, that is when $[0.05, 1] \cap [\alpha, 1] = \emptyset$. Thus, we reject H_0 for every $\alpha \in (0.1, 1/3]$ (it is fine to have written $\alpha > 0.1$).

3. Let $X_1, X_2, \ldots$ be a sequence of iid random variables, where X_1 has density

$$f_\lambda(x) = kx^{k-1}\lambda^{-k} \exp\left(-\left(\frac{x}{\lambda}\right)^k\right)$$

for $x > 0$, unknown parameter $\lambda > 0$ and known constant $k > 0$. All the regularity conditions are satisfied in this case.

(a) For fixed $n \in \mathbb{N}$, compute the maximum likelihood estimator (MLE) for λ based on the sample $X_1, \ldots, X_n$. Denote this estimator by $\hat{\lambda}_n$. (3 marks)

Solution: The log-likelihood is

$$\ell(\lambda) = \sum_{i=1}^{n} \log(k) + (k-1)\log(x_i) - k\log(\lambda) - (x_i/\lambda)^k$$

and its derivative is

$$\frac{\partial}{\partial\lambda}\ell(\lambda) = -nk/\lambda + \sum_{i=1}^{n} kx_i^k/\lambda^{k+1} = \frac{k}{\lambda}\Big(-n + \sum_{i=1}^{n} x_i^k/\lambda^k\Big)$$

by setting it equal to zero we obtain

$$\hat{\lambda}_n = \left(\frac{1}{n}\sum_{i=1}^{n} x_i^k\right)^{1/k}$$

This is indeed the MLE because the first derivative of the log-likelihood is positive for small values of λ and negative for large values of λ. Hence, the obtained stationary point is a maximum.

(b) Assume that $\hat{\lambda} = 4$, $k = 2$ and $n = 100$. Compute the 95% observed approximate confidence interval for λ. You might need to use that $P(Z > 1.96) = 0.025$ where $Z \sim N(0, 1)$. (2 marks)

Solution: The second derivative of the log-likelihood is

$$\frac{\partial^2}{\partial\lambda^2}\ell(\lambda) = nk/\lambda^2 - (k+1)k\sum_{i=1}^{n} x_i^k/\lambda^{k+2}$$

Moreover, using that the regularity conditions hold we obtain that $E_\lambda[\frac{\partial}{\partial\lambda}\ell(\lambda, X)] = 0$, *here* X *stands for* $(X_1, ..., X_n)$. *This is a result we have seen in the lectures and is obtained as follows*

$$E_\lambda[\frac{\partial}{\partial\lambda}\ell(\lambda, X)] = E[\frac{f_\lambda'(X)}{f_\lambda(X)}] = \int \frac{f_\lambda'(x)}{f_\lambda(x)} f_\lambda(x)dx = \int f_\lambda'(x)dx = \frac{\partial}{\partial\lambda}\int f_\lambda(x)dx = \frac{\partial}{\partial\lambda}1 = 0.$$

Hence, we have that

$$0 = E_\lambda[\frac{\partial}{\partial\lambda}\ell(\lambda, X)] = E_\lambda\left[\frac{k}{\lambda}\Big(-n + \sum_{i=1}^{n} X_i^k/\lambda^k\Big)\right]$$

$$\Rightarrow E_\lambda\left[-n + \sum_{i=1}^{n} X_i^k/\lambda^k\right] = 0 \Rightarrow E\left[\sum_{i=1}^{n} X_i^k\right] = n\lambda^k \Rightarrow E[X_1^k] = \lambda^k$$

We can now compute the Fisher information:

$$I(\lambda) = -E\left[\frac{\partial^2}{\partial\lambda^2}\ell(\lambda)\right] = -E\left[nk/\lambda^2 - (k+1)k\sum_{i=1}^{n} X_i^k/\lambda^{k+2}\right]$$

$$= -nk/\lambda^2 + \frac{(k+1)k}{\lambda^{k+2}} nE[X_1^k] = -nk/\lambda^2 + \frac{(k+1)k}{\lambda^{k+2}} n\lambda^k = \frac{nk^2}{\lambda^2}.$$

By the asymptotic normality of the MLE we know that $\sqrt{n}(\hat\lambda_n - \lambda) \xrightarrow{d} N(0, (I_{(1)}(\lambda))^{-1}) = N(0, \lambda^2/k^2)$. *Since the MLE is consistent, namely* $\hat\lambda_n \xrightarrow{p} \lambda$, *an estimate for the variance is* $\frac{\hat\lambda_n^2}{k^2}$. *In particular, we have that approximately*

$$\frac{\hat\lambda_n - \lambda}{\frac{\hat\lambda_n}{k\sqrt{n}}} \sim N(0, 1)$$

Hence, the approximate 95% CI for λ *is* $\hat\lambda_n \pm \dfrac{1.96\hat\lambda_n}{k\sqrt{n}}$ *which is* 4 ± 0.392 *using the given values.*

(Total 20 marks)