THE UNIVERSITY OF SYDNEY
SCHOOL OF MATHEMATICS AND STATISTICS

Tutorial Week 5 Solution

STAT3023: Statistical Inference — Semester 2, 2023

1. (a) For each X_i, we have the pmf

$$p_{X_i}(x_i) = P(X_i = x_i) = \theta^{x_i}(1-\theta)^{1-x_i}, \quad x_i \in \{0,1\}.$$

The likelihood of θ is therefore,

$$L(\theta) = \prod_{i=1}^n p_{X_i}(X_i) = \prod_{i=1}^n \theta^{X_i}(1-\theta)^{1-X_i} = \theta^{\sum_{i=1}^n X_i}(1-\theta)^{n-\sum_{i=1}^n X_i}$$
$$= \left(\frac{\theta}{1-\theta}\right)^{\sum_{i=1}^n X_i}(1-\theta)^n.$$

This likelihood can be written in the canonical form of the exponential family as

$$L(\theta) = I_{\mathcal{A}}(\mathbf{x}) \exp\left\{\log\left(\frac{\theta}{1-\theta}\right)\sum_{i=1}^n X_i + n\log(1-\theta)\right\},$$

where $\mathcal{A} = \{(X_1,\ldots,X_n) : X_i \in \{0,1\}\}$. It is in the exponential family form with $\eta = \log\left(\frac{\theta}{1-\theta}\right)$, $T(\mathbf{x}) = \sum_{i=1}^n X_i$ being the natural parameter and a sufficient statistic, respectively. To identify $A^*(\eta)$, note that from $\eta = \log\left(\frac{\theta}{1-\theta}\right)$, we have

$$e^\eta = \frac{\theta}{1-\theta}, \quad e^\eta - \theta e^\eta = \theta, \quad e^\eta = \theta(e^\eta+1),$$

so $\theta = \frac{e^\eta}{1+e^\eta}$. Therefore,

$$A^*(\eta) = -n\log(1-\theta) = -n\log\left(1-\frac{e^\eta}{1+e^\eta}\right) = -n\log\left(\frac{1}{1+e^\eta}\right) = n\log(1+e^\eta).$$

From this exponential family form, we have

$$E\{T(\mathbf{x}\} = \frac{dA^*}{d\eta} = n\frac{e^\eta}{1+e^\eta} = n\theta,$$

$$\mathrm{Var}\{T(\mathbf{x})\} = \frac{d^2A^*}{d\eta^2} = n\frac{e^\eta(1+e^\eta)-e^\eta e^\eta}{(1+e^\eta)^2} = \frac{ne^\eta}{(1+e^\eta)^2} = \frac{ne^\eta}{1+e^\eta}\frac{1}{1+e^\eta} = n\theta(1-\theta).$$

(b) By a similar argument, the likelihood of θ is

$$L(\theta) = \prod_{i=1}^n f_{X_i}(X_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}(X_i-\theta)^2\right)$$
$$= (2\pi)^{-n/2}\exp\left(-\frac{1}{2}\sum_{i=1}^n X_i^2 + \theta\sum_{i=1}^n X_i - \frac{n}{2}\theta^2\right)$$
$$= (2\pi)^{-n/2}\exp\left(-\frac{1}{2}\sum_{i=1}^n X_i^2\right)\exp\left(\theta\sum_{i=1}^n X_i - \frac{n}{2}\theta^2\right).$$

This likelihood function is already in the canonical form of the exponential family, with $h(\mathbf{x}) = (2\pi)^{-n/2} \exp\left(-(1/2)\sum_{i=1}^{n} X_i^2\right)$ (we don't write out the support function explicitly since there is no constraint on X_i). The natural parameter is $\eta = \theta$, and $T(\mathbf{x}) = \sum_{i=1}^{n} X_i$ is a sufficient statistic. The function $A^*(\eta) = (n/2)\theta^2 = (n/2)\eta^2$. From this exponential family form, we have

$$
\begin{aligned}
E\{T(\mathbf{x}\} &= \frac{dA^*}{d\eta} = n\eta = n\theta, \\
\mathrm{Var}\{T(\mathbf{x})\} &= \frac{d^2A^*}{d\eta^2} = n.
\end{aligned}
$$

(c) The likelihood of θ is

$$
\begin{aligned}
L(\theta) = \prod_{i=1}^{n} f_{X_i}(X_i) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\theta}} \exp\left(-\frac{1}{2\theta}X_i^2\right) \\
&= (2\pi)^{-n/2} \exp\left(-\frac{1}{2\theta}\sum_{i=1}^{n} X_i^2 - \frac{n}{2}\log(\theta)\right)
\end{aligned}
$$

This likelihood function is already in the canonical form of the exponential family, with $h(\mathbf{x}) = (2\pi)^{-n/2}$ (we don't write out the support function explicitly since there is no constraint on X_i). The natural parameter is $\eta = -1/(2\theta)$, and $T(\mathbf{x}) = \sum_{i=1}^{n} X_i^2$ is a sufficient statistic.

2. (a) We note that $E(X) = \alpha\beta$. Hence, for the method-of-moment estimator, we set the first sample moment $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ equals to $E(X)$, i.e.

$$\bar{X} = \alpha\beta.$$

Hence, a method-of-moment estimate for β is $\hat{\beta}_{\mathrm{mm}} = \dfrac{\bar{X}}{\alpha}$.

For the maximum likelihood estimator, the likelihood function of β is

$$
\begin{aligned}
L(\beta) = \prod_{i=1}^{n} f_{X_i}(X_i) &= \prod_{i=1}^{n} \frac{1}{\Gamma(\alpha)\beta^\alpha}(X_i)^{\alpha-1}e^{-X_i/\beta} \\
&= \{\Gamma(\alpha)\beta^\alpha\}^{-n}\left(\prod_{i=1}^{n} X_i\right)^{\alpha-1} e^{-\sum_{i=1}^{n} X_i/\beta}.
\end{aligned}
$$

The corresponding log likelihood is

$$\ell(\beta) = \log L(\beta) = -n\log\Gamma(\alpha) - n\alpha\log(\beta) + (\alpha-1)\log\left(\prod_{i=1}^{n} X_i\right) - \frac{\sum_{i=1}^{n} X_i}{\beta}. \tag{1}$$

To find the maximum likelihood estimator, we differentiate the log likelihood with respect to β and set it equal to zero. In this context,

$$\frac{\partial\ell}{\partial\beta} = \frac{-n\alpha}{\beta} + \frac{\sum_{i=1}^{n} X_i}{\beta^2} = \frac{-n\alpha\beta + \sum_{i=1}^{n} X_i}{\beta^2} = 0.$$

Hence,

$$-n\alpha\beta + \sum_{i=1}^{n} X_i = 0, \quad \hat{\beta}_{mle} = \frac{\sum_{i=1}^{n} X_i}{n\alpha} = \frac{\bar{X}}{\alpha}.$$

(b) If both α and β are unknown, then we need to maximize the log likelihood 1 with respect to both α and β. Hence, we need to differentiate the log likelihood with both α and β and set them equals zero. Specifically,

$$\frac{\partial \ell}{\partial \alpha} = -n\frac{\Gamma'(\alpha)}{\Gamma(\alpha)} - n\log(\beta) + \log\left(\prod_{i=1}^{n} X_i\right) = 0; \tag{2}$$

$$\frac{\partial \ell}{\partial \beta} = \frac{-n\alpha}{\beta} + \frac{\sum_{i=1}^{n} X_i}{\beta^2} = \frac{-n\alpha\beta + \sum_{i=1}^{n} X_i}{\beta^2} = 0. \tag{3}$$

From (3), we have $\beta = \sum_{i=1}^{n} X_i/(n\alpha)$, or $\beta = \bar{X}/\alpha$. Substituting in into (2), we then have the equation to solve for the maximum likelihood of α is

$$-n\frac{\Gamma'(\alpha)}{\Gamma(\alpha)} - n\log\left(\frac{\bar{X}}{\alpha}\right) + \log\left(\prod_{i=1}^{n} X_i\right) = 0.$$

3. (a) Note that for each X_i, the pmf is

$$p_{X_i}(x) = P(X_i = x) = e^{-\lambda}\frac{\lambda^{x_i}}{x_i!}.$$

The likelihood function of λ is therefore given by

$$L(\lambda) = \prod_{i=1}^{n} p_{X_i}(X_i) = \prod_{i=1}^{n} e^{-\lambda}\frac{\lambda^{X_i}}{X_i!} = e^{-n\lambda}\frac{\lambda^{\sum_{i=1}^{n} X_i}}{\prod_{i=1}^{n} X_i!}.$$

Therefore, the corresponding log likelihood is

$$\ell(\lambda) = \log L(\lambda) = -n\lambda + \left(\sum_{i=1}^{n} X_i\right)\log(\lambda) - \log\left(\prod_{i=1}^{n} X_i!\right).$$

To find the maximum likelihood estimator, we differentiate the log likelihood function with respect to λ and set it equal to zero. In this context,

$$\frac{\partial \ell}{\partial \lambda} = -n + \frac{\sum_{i=1}^{n} X_i}{\lambda} = 0,$$

so the maximum likelihood estimator of λ is $\hat{\lambda}_{mle} = \dfrac{\sum_{i=1}^{n} X_i}{n} = \bar{X}$. Hence, the corresponding maximum likelihood estimator for $\theta = e^{-\lambda}$ is $\hat{\theta}_1 = e^{-\bar{X}}$.

(b) To find the bias and variance of $\hat{\theta}_1$, we need to compute $E(\hat{\theta}_1) = E(e^{-\bar{X}})$ and $E(\hat{\theta}_1^2) = E(e^{-2\bar{X}})$. We can use the moment-generating function to do it. The moment generating function of each X_i is

$$M_{X_i}(t) = e^{\lambda(e^t - 1)}.$$

Since X_i are iid, then we have

$$M_{\sum_{i=1}^{n} X_i}(t) = \prod_{i=1}^{n} M_{X_i}(t) = e^{n\lambda(e^t - 1)}.$$

Therefore,

$$M_{\bar{X}}(t) = M_{\sum_{i=1}^n X_i}(t/n) = \prod_{i=1}^{n} M_{X_i}(t) = e^{n\lambda(e^{t/n}-1)} = \theta^{-n(e^{t/n}-1)}.$$

As a result,

$$\begin{aligned} E(\hat{\theta}_1) &= E(e^{-\bar{X}}) = M_{\bar{X}}(-1) = \theta^{-n(e^{-1/n}-1)}, \\ E(\hat{\theta}_1^2) &= E(e^{-2\bar{X}}) = M_{\bar{X}}(-2) = \theta^{-n(e^{-2/n}-1)}. \end{aligned}$$

Finally the corresponding bias and variance of $\hat{\theta}_1$ are given by

$$\begin{aligned} \text{Bias}(\hat{\theta}_1) &= E(\hat{\theta}_1) - \theta = \theta^{-n(e^{-1/n}-1)} - \theta; \\ \text{Var}(\hat{\theta}_1) &= E(\hat{\theta}_1^2) - \Big\{E(\hat{\theta}_1)\Big\}^2 = \theta^{-n(e^{-2/n}-1)} - \theta^{-2n(e^{-1/n}-1)}. \end{aligned}$$

(c) Y just counts the number of zeros out of n trials, so $Y \sim \text{Binomial}(n, \theta)$.

(d) An unbiased estimator for θ is $\hat{\theta}_2 = \dfrac{Y}{n}$. The corresponding variance of $\hat{\theta}_2$ is

$$\text{Var}(\hat{\theta}_2) = \text{Var}\left(\frac{Y}{n}\right) = \frac{1}{n^2}\text{Var}(Y) = \frac{1}{n^2} n\theta(1-\theta) = \frac{\theta(1-\theta)}{n}.$$

(e) Since $E(X) = \text{Var}X = \lambda$. Hence, by the central limit theorem, we have

$$\sqrt{n}(\bar{X} - \lambda) \xrightarrow{d} N(0, \lambda).$$

Consider $g(\lambda) = e^{-\lambda} = \theta$, so $\lambda = -\log(\theta)$ then $g'(\lambda) = -e^{-\lambda}$. An application of the Delta method gives

$$\sqrt{n}(e^{-\bar{X}} - e^{-\lambda}) \xrightarrow{d} N(0, \lambda e^{-2\lambda}),$$

or in other words,

$$\sqrt{n}(\hat{\theta}_1 - \theta) \xrightarrow{d} N(0, -\theta^2 \log \theta).$$

Therefore, as $n \to \infty$, the relative efficiency of $\hat{\theta}_1$ and $\hat{\theta}_2$ is

$$\frac{-\theta^2 \log(\theta)}{\theta(1-\theta)} = \frac{-\theta \log(\theta)}{1-\theta}.$$