STA 131B HW2

7.5

8. The likelihood function is

$$f_n(\mathbf{x}|\theta) = \begin{cases} \exp(n\theta - \sum_{i=1}^n x_i) & \text{for } \min(x_1, \dots, x_n) > \theta \\ 0 & \text{otherwise.} \end{cases}$$

(a) For each value of $\mathbf{x}$, $f_n(\mathbf{x}|\theta)$ will be a maximum when θ is made as large as possible subject to the strict inequality $\theta < \min(x_1, \dots, x_n)$. Therefore, the value $\theta = \min(x_1, \dots, x_n)$ cannot be used and there is no MLE.

(b) Change the pdf of X_i to the equivalent

$$f(x|\theta) \begin{cases} e^{\theta - x} & \text{for } x \geq \theta \\ 0 & \text{for } x < \theta \end{cases}$$

(only the inequality sign is changed), then the likelihood function is

$$f_n(\mathbf{x}|\theta) = \begin{cases} \exp(n\theta - \sum_{i=1}^n x_i) & \text{for } \min(x_1, \dots, x_n) \geq \theta \\ 0 & \text{otherwise.} \end{cases}$$

The likelihood function $f_n(\mathbf{x}|\theta)$ will be nonzero for $\theta \leq \min(x_1, \dots, x_n)$ and the MLE will be $\hat{\theta} = \min(x_1, \dots, x_n)$.

10. The likelihood function is

$$f_n(\mathbf{x}|\theta) = \frac{1}{2^n} \exp\left\{ -\sum_{i=1}^n |x_i - \theta| \right\}.$$

Therefore, the MLE of θ will be the value that minimizes $\sum_{i=1}^n |x_i - \theta|$. $\hat{\theta}$ = the sample median of x_i is one of the solutions (*non-unique*) to this minimization problem. To see this:

$$\begin{aligned} \frac{\partial |x_i - \theta|}{\partial \theta} &= \begin{cases} -1, & \text{if } \theta < x_i \\ 1, & \text{if } \theta > x_i \\ \text{does not exist}, & \text{if } \theta = x_i \end{cases} \\ \frac{\partial l(\theta)}{\partial \theta} &= \sum_{i=1}^n \frac{\partial |x_i - \theta|}{\partial \theta} \\ &= \sum_{x_i > \theta} 1 + \sum_{x_i < \theta} -1 \\ &= -\#\{x_i : x_i < \theta\} + \#\{x_i : x_i > \theta\}, \quad \text{if } \theta \neq x_i \text{ for all } i, \end{aligned}$$

where $\#\{x_i : x_i < \theta\}$ means the number of x_is that are smaller than θ. $\frac{\partial l(\theta)}{\partial \theta} > 0$ ($l(\theta)$ increasing) when there are more x_i lying left to θ, and is positive ($l(\theta)$ decreasing) when more x_i are lying right to θ, so the maximum of $l(\theta)$ is taken when there are equal number of x_i lying on either sides of θ. ($l(\theta)$ is continuous, so finitely number of non-differentiable points do not affect the results.) We have $\hat{\theta}$ is the middle value among $x_i, \ldots, x_n$ if n is odd; or is any point between the two middle values among $x_1, \ldots, x_n$ if n is even (again, *non-unique*).

Another way to see $\hat{\theta}$ is a minimizer of $\sum_{i=1}^n |x_i - \theta|$ is as follows (similar to book Theorem 4.5.3): for any $\theta \geq \hat{\theta}$, we have

$$\begin{aligned}
\sum_{i=1}^n |x_i - \theta| - \sum_{i=1}^n |x_i - \hat{\theta}| &= \sum_{i=1}^n (|x_i - \theta| - |x_i - \hat{\theta}|) \\
&= \sum_{i:x_i<\hat{\theta}} (|x_i - \theta| - |x_i - \hat{\theta}|) + \sum_{i:\hat{\theta}\leq x_i \leq \theta} (|x_i - \theta| - |x_i - \hat{\theta}|) + \sum_{i:\theta < x_i} (|x_i - \theta| - |x_i - \hat{\theta}|) \\
&= \sum_{i:x_i<\hat{\theta}} (\theta - \hat{\theta}) + \sum_{i:\hat{\theta}\leq x_i \leq \theta} (\theta + \hat{\theta} - 2x_i) + \sum_{i:\theta < x_i} (-\theta + \hat{\theta}) \\
&\geq \sum_{i:x_i<\hat{\theta}} (\theta - \hat{\theta}) + \sum_{i:\hat{\theta}\leq x_i \leq \theta} (\theta - \hat{\theta}) + \sum_{i:\theta < x_i} (-\theta + \hat{\theta}) \\
&= (\theta - \hat{\theta})(\#\{x_i : x_i \leq \theta\} - \#\{x_i : x_i > \theta\})
\end{aligned}$$

We know $\#\{x_i : x_i \leq \theta\} - \#\{x_i : x_i > \theta\}$ is non-negative because $\theta \geq \hat{\theta}$ (the median), so $\sum_{i=1}^n |x_i - \theta| - \sum_{i=1}^n |x_i - \hat{\theta}| \geq 0$ for all $\theta > \hat{\theta}$. The case $\theta < \hat{\theta}$ can be proven similarly and we then have $\sum_{i=1}^n |x_i - \theta| - \sum_{i=1}^n |x_i - \hat{\theta}| \geq 0$ for all θ. So $\hat{\theta}$ is a minimizer.

7.6

3. The median of an exponential distribution with parameter β is the number m such that

$$\int_0^m \beta \exp(-\beta x)\, \mathrm{d}x = \frac{1}{2}.$$

By calculus $m = \log(2)/\beta$, and it follows from the invariant property of MLE that the MLE $\hat{m} = \log(2)/\hat{\beta}$, where $\hat{\beta}$ is the MLE of β. The joint pdf of $X_1, \ldots, X_n$ is

$$f(\mathbf{x}|\beta) = \begin{cases} \beta^n \exp(-\beta \sum_{i=1}^n x_i) & x_1, \ldots, x_n > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Set $\frac{\partial f(\mathbf{x}|\beta)}{\partial \beta} = 0$ we find $\hat{\beta} = 1/\bar{X}_n$. So $\hat{m} = \log(2)\bar{X}_n$.

8. By the invariant property of MLE, the MLE for $\theta = \Gamma'(\alpha)/\Gamma(\alpha)$ is $\hat{\theta} = \Gamma'(\hat{\alpha})/\Gamma(\hat{\alpha})$. By equation (7.6.5) we know the MLE $\Gamma'(\hat{\alpha})/\Gamma(\hat{\alpha}) = \frac{1}{n}\sum_{i=1}^n \log(x_i)$.

23.

(a) The means of X_i and X_i^2 are respectively $\alpha/(\alpha+\beta)$ and $\alpha(\alpha+1)/[(\alpha+\beta)(\alpha+\beta+1)]$. We set these equal to the first two sample moments $m_1 = \frac{1}{n}\sum_{i=1}^n x_i$ and $m_2 = \frac{1}{n}\sum_{i=1}^n x_i^2$ and solve for α and β. Solve

$$\begin{cases} m_1 = \frac{\alpha}{\alpha+\beta} \\ \frac{m_2}{m_1} = \frac{\alpha+1}{\alpha+\beta+1}, \end{cases}$$

which is a linear system

$$\begin{cases} (m_1-1)\alpha + m_1\beta = 0 \\ (m_2-m_1)\alpha + m_2\beta + m_2 - m_1 = 0 \end{cases}$$

we have

$$\hat{\alpha} = \frac{m_1(m_1-m_2)}{m_2-m_1^2}$$
$$\hat{\beta} = \frac{(1-m_1)(m_1-m_2)}{m_2-m_1^2}.$$

(b) The MLE for a beta distribution has no closed form. One can verify the method of moments estimator is not the MLE by a simulation example: obtain the MLE by numerical optimization and then compare with the method of moments estimator. (Alternatively, MLE has no closed form but MOM estimator has closed form, so they cannot be the same.)

7.10

10. $\theta = (1/3)(1+\beta)$ and $0 \le \beta \le 1$ implies θ must lie in the interval $[1/3, 2/3]$. The log-likelihood function is

$$l(\theta) = \log(f(\mathbf{x}|\theta)) = n\bar{x}\log(\theta) + (n-n\bar{x})\log(1-\theta),$$
$$\frac{\partial l(\theta)}{\partial\theta} = \frac{n\bar{x}-n\theta}{\theta(1-\theta)}.$$

The unconstrained MLE is obtained by setting $\frac{\partial l(\theta)}{\partial \theta} = 0 \implies \hat{\theta} = \bar{x}$. The MLE for θ is then

$$\hat{\theta} = \begin{cases} m_1 & \text{if } 1/3 \le \bar{x} \le 2/3 \\ 1/3 & \text{if } \bar{x} < 1/3, \text{because } \frac{\partial l(\theta)}{\partial\theta} > 0 \text{ when } \theta < \bar{x}, \\ 2/3 & \text{if } \bar{x} > 2/3, \text{because } \frac{\partial l(\theta)}{\partial\theta} < 0 \text{ when } \theta > \bar{x}. \end{cases}$$

Then by the invariant property of MLE, $\hat{\beta} = 3\hat{\theta} - 1$.

Additional problems

1.

(a) Since $E(X_i) = \lambda^{-1}$, we set the sample mean $m_1 = \lambda^{-1}$ and obtain the method of moment estimator $\hat{\lambda}_1 = 1/m_1$.

(b) Since $E(X_i^2) = 2/\lambda^2$, we set $m_2 = 2/\lambda^2$ and obtain the method of moment estimator $\hat{\lambda}_2 = \sqrt{\frac{2}{m_2}}$.

(c) Since $\mathrm{var}(X_i) = E(X_i^2) - E(X_i)^2 = 1/\lambda^2$, we set the sample variance $m_2 - m_1^2 = 1/\lambda^2$ and obtain $\hat{\lambda}_3 = 1/\sqrt{m_2 - m_1^2}$.

(d) $P(X_1 \geq 1) = e^{-\lambda}$. By handout 3, the method of moment estimator using the first moment is $e^{-\hat{\lambda}_1} = \exp\{-1/m_1\}$ (you can as well plug in a MOM $\hat{\theta}$ based on a different moment like the ones obtained in (b) and (c)).

2.

(a) (Supplementary Exercise #4) The joint pdf is

$$f(\mathbf{x}|\theta) = \begin{cases} \frac{1}{\theta^n} & \text{if } \theta \leq x_i \leq 2\theta \text{ for } i = 1, \ldots, n \\ 0 & \text{otherwise.} \end{cases}$$

$\theta \leq x_i \leq 2\theta$ for $i = 1, \ldots, n$ if and only if $\frac{1}{2}\max(x_1, \ldots, x_n) \leq \theta \leq \min(x_1, \ldots, x_n)$. Since $f(\mathbf{x}|\theta)$ is decreasing on its support, the MLE $\hat{\theta} = \frac{1}{2}\max(x_1, \ldots, x_n)$ is the smallest possible value of θ.

(b) Since $E(X_i) = 1.5\theta$, set $m_1 = 1.5\theta$ we have $\hat{\theta}_1 = 2m_1/3$.

(c) Since $E(X_i^2) = \frac{7\theta^2}{3}$, set $m_2 = \frac{7\theta^2}{3}$ we have $\hat{\theta}_2 = \sqrt{\frac{3m_2}{7}}$.

(d) The MLE cannot be the same as any of the method of moment estimators in general, because method of moment estimators are functions of only m_1, m_2, m_3, $\ldots$, but the MLE involves $\max(x_1, \ldots, x_n)$.