STAT 510 — Mathematical Statistics

Exam: Midterm Exam **Date:** October 20th 2022, 9:30am-10:50am

Problem 1.
1. Write $X_i = Y_i + \theta$, where $Y_i \sim f(y|\beta,\theta) = \frac{1}{\beta}e^{-\frac{y}{\beta}}\mathbb{1}_{\{y\geq 0\}}$

Since

$$\mathbb{E}Y_i = \int_0^\infty \frac{y}{\beta}e^{-\frac{y}{\beta}}dy = \beta\int_0^\infty ze^{z}dz = -\beta\int_0^\infty zde^{-z} = -\beta\left[ze^{-z}|_0^\infty - \int_0^\infty e^{-z}dz\right] = \beta$$

$$\mathbb{E}Y_i^2 = \int_0^\infty \frac{y^2}{\beta}e^{-\frac{y}{\beta}}dy = \beta^2\int_0^\infty z^2e^{-z}dz = -\beta^2\int_0^\infty z^2de^{-z} = 2\beta^2\int_0^\infty e^{-z}dz = 2\beta^2$$

then $Var(X_i) = Var(Y_i) = \beta^2,\ \mathbb{E}X_i^2 = Var(X_i) + (\mathbb{E}X_i)^2 = \beta^2 + (\beta+\theta)^2$

Let $m_k = \dfrac{1}{n}\sum\limits_{i=1}^n X_i^k$ be the k-th moment. Method of moment estimator $(\tilde{\beta},\tilde{\theta})$ satisfies:

$$\begin{cases} m_1 = \mathbb{E}X_1 = \beta+\theta \\ m_2 = \mathbb{E}X_1^2 = \beta^2 + (\beta+\theta)^2 \end{cases}$$

$$\implies m_2 - m_1^2 = \beta^2 \implies \tilde{\beta} = \sqrt{m_2 - m_1^2}, \tilde{\theta} = m_1 - \tilde{\beta} = m_1 - \sqrt{m_2 - m_1^2}$$

Thus, $(\tilde{\beta},\tilde{\theta}) = (\sqrt{m_2 - m_1^2}, m_1 - \sqrt{m_2 - m_1^2})$

2. joint likelihood of $\tilde{X} = (X_1,\cdots,X_n)$:

$$\begin{aligned} f(\tilde{x}|\beta,\theta) &= \prod_{i=1}^n\left[\frac{1}{\beta}e^{-\frac{x_i-\theta}{\beta}}\mathbb{1}_{(x_i\geq\theta)}\right] \\ &= \beta^{-n}e^{-\frac{\sum_{i=1}^n(x_i-\theta)}{\beta}}\mathbb{1}_{(\min\limits_{1\leq i\leq n} x_i\geq\theta)} \\ &= \beta^{-n}e^{-\frac{n\bar{x}_n}{\beta}+\frac{n\theta}{\beta}}\mathbb{1}_{(\min\limits_{1\leq i\leq n} x_i\geq\theta)} \end{aligned}$$

where $\bar{x}_n = \frac{1}{n}\sum\limits_{i=1}^n x_i$.

Note that $f(\tilde{x}|\beta,\theta)$ is increasing w.r.t. θ, which means MLE $\hat{\theta} = \min_{1\le i\le n} X_i$

Then $\max_{\beta>0} \; log f(\tilde{x}|\beta,\hat{\theta}) = -n \; log\beta - \frac{n\bar{x}_n}{\beta} + \frac{n\hat{\theta}}{\beta}$

$$[\beta]: \qquad -\frac{n}{\beta} + \frac{n(\bar{x}_n - \hat{\theta})}{\beta^2} = 0$$

$\Longrightarrow \; \hat{\beta} = \bar{X}_n - \hat{\theta} = \bar{X}_n - \min_{1\le i\le n} X_i$

3. Want to compute

$$\mathbb{E}\hat{\theta} = \mathbb{E}\min_{1\le i\le n}(Y_i + \theta) = \mathbb{E}\min_{1\le i\le n} Y_i + \theta$$

where Y_i are iid exponential r.v.'s with scale parameter β.
<u>Claim</u>: $\mathbb{E}\min_{1\le i\le n} Y_i > 0$.
<u>Check</u>:

$$\begin{aligned} P(\min_{1\le i\le n} Y_i > t) &= P(Y_i \ge t, \forall i \in [n]) \\ &= \prod_{i=1}^{n} P(Y_i \ge t) \\ &= \left[\int_t^\infty \frac{1}{\beta} e^{-\frac{y}{\beta}} dy\right]^n \\ &= \left(\int_{\frac{t}{\beta}}^\infty e^{-z} dz\right)^n \\ &= e^{-\frac{nt}{\beta}} \end{aligned}$$

$\Longrightarrow \; \frac{d}{dt} P(\min_{1\le i\le n} Y_i \le t) = \frac{n}{\beta} e^{-\frac{nt}{\beta}}$ is the pdf of $\min_{1\le i\le n} Y_i$

$\Longrightarrow \; \mathbb{E}\min_{1\le i\le n} Y_i = \int_0^\infty t\frac{n}{\beta} e^{-\frac{nt}{\beta}} dt = \frac{\beta}{n}\int_0^\infty z e^{-z} dz = \frac{\beta}{n} > 0$
(in the second equality: changing variables $z = \frac{nt}{\beta}$)

$\Longrightarrow \; \mathbb{E}\hat{\theta} = \theta + \mathbb{E}\min_{1\le i\le n} Y_i > \theta$ and $\mathbb{E}\hat{\beta} = \mathbb{E}\bar{X}_n - \mathbb{E}\min_{1\le i\le n} X_i = \mathbb{E}\bar{Y}_n - \mathbb{E}\min_{1\le i\le n} Y_i = \beta - \frac{\beta}{n} < \beta$, which means both $\hat{\beta}$ and $\hat{\theta}$ are biased.

4. Recall the joint likelihood function is

$$f(\tilde{x}|\beta,\theta) = \beta^{-n} e^{\frac{n\theta}{\beta}} e^{-\frac{n\bar{x}_n}{\beta}} \mathbb{1}_{(\min_{1\le i\le n} x_i \ge \theta)}$$

Use Neyman factorization theorem with $h(\tilde{x}) = 1$, we conclude that $(\bar{X}_n, \min_{1\le i\le n} X_i)$ is a sufficient statistics for (β, θ). ∎

Problem 2.

1. joint pmf of $\tilde{X} = (X_1, \cdots, X_n)$ is given by

$$f(\tilde{x}|p) = \prod_{i=1}^{n} \left[(1-p)^{x_i - 1} p\right] = (1-p)^{\sum_{i=1}^{n}(x_i - 1)} p^n$$

$$\Longrightarrow \ l(p|\tilde{x}) := log\ f(\tilde{x}|p) = n(\bar{x}_n - 1)log(1-p) + nlogp$$

$$[p]: \ \frac{n(\bar{x}_n - 1)}{1-p} = \frac{n}{p} \Longrightarrow p\bar{x}_n - p = 1 - p \Longrightarrow \hat{p}_{MLE} = \frac{1}{\bar{X}_n}$$

2. Using Bayes' rule:

$$\begin{aligned}
\pi(p|\tilde{x}) &\propto \pi(p) f(\tilde{x}|p) \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}(1-p)^{n(\bar{x}_n - 1)} p^n \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha+n-1}(1-p)^{\beta+n(\bar{x}_n-1)-1}
\end{aligned}$$

where the last term is the kernal (i.e., unnormalized pdf) of the distribution $Beta(\alpha+n, \beta+n(\bar{x}_n - 1))$
For any $\alpha, \beta > 0$, if $p \sim Beta(\alpha, \beta)$.

$$\begin{aligned}
\mathbb{E}p &= \int_0^1 p \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1} dp \\
&= \left[\int_0^1 \frac{\Gamma(\alpha+\beta+1)}{\Gamma(\alpha+1)\Gamma(\beta)} p^{(\alpha+1)-1}(1-p)^{\beta-1} dp\right] \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)} \frac{\Gamma(\alpha+1)}{\Gamma(\alpha+\beta+1)}
\end{aligned}$$

The above integral = 1 because the integrand is the pdf of $Beta(\alpha+1, \beta)$
Using recursive relation,

$$\begin{cases} \Gamma(\alpha+\beta+1) = (\alpha+\beta)\Gamma(\alpha+\beta) \\ \Gamma(\alpha+1) = \alpha\Gamma(\alpha) \end{cases}$$

$$\Longrightarrow \ \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)} \frac{\Gamma(\alpha+1)}{\Gamma(\alpha+\beta+1)} = \frac{\alpha}{\alpha+\beta}$$

So, when $p|\tilde{X} \sim Beta(\alpha+n, \beta+n(\bar{x}_n - 1))$, the posterior mean

$$\mathbb{E}[p|\tilde{x}] = \frac{\alpha+n}{\alpha+n+\beta+n(\bar{x}_n - 1)} = \frac{\alpha+n}{\alpha+\beta+n\bar{x}_n}$$

$\blacksquare$

Problem 3.

1. joint pdf of $\tilde{X} = (X_1, \cdots, X_4)$ is given by

$$
\begin{aligned}
f(\tilde{x}|\theta) &= \prod_{i=1}^{4} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i-\theta)^2}{2}} \\
&= (2\pi)^{-2} e^{-\frac{\sum_{i=1}^{4}(x_i^2 - 2x_i\theta + \theta^2)}{2}} \\
&= (2\pi)^{-2} e^{-\frac{1}{2}\sum_{i=1}^{4} x_i^2} e^{(\sum_{i=1}^{4} x_i)\theta - 2\theta^2}
\end{aligned}
$$

Let

$$h(\tilde{x}) = (2\pi)^{-2} e^{-\frac{1}{2}\sum_{i=1}^{4} x_i^2}$$

and

$$g(T_1(\tilde{x}), T_2(\tilde{x})|\theta) = e^{(\sum_{i=1}^{4} x_i)\theta - 2\theta^2}$$

where $g(t_1, t_2) = e^{(t_1+t_2)\theta - 2\theta^2}$, $T_1(\tilde{x}) = x_1 + x_2$, $T_2(\tilde{x}) = x_3 + x_4$.

By Neyman factorization theorem, (T_1, T_2) is a suficient statistics for θ.

2. Since $T_3 = X_1 + 2X_2 \sim N(3\theta, 5)$, $T_4 = X_3 + X_4 \sim N(2\theta, 2)$ and $T_3 \perp\!\!\!\perp T_4$, the joint pdf of (T_3, T_4) is given by

$$q(t_3, t_4|\theta) = \frac{1}{\sqrt{2\pi \cdot 5}} e^{-\frac{(t_3-3\theta)^2}{10}} \times \frac{1}{\sqrt{2\pi \cdot 2}} e^{-\frac{(t_4-2\theta)^2}{4}}$$

Compute

$$
\begin{aligned}
\frac{f(\tilde{x}|\theta)}{q(T_3(\tilde{x}), T_4(\tilde{x})|\theta)} &= \frac{(2\pi)^{-2} e^{-\frac{\sum_{i=1}^{4}(x_i-\theta)^2}{2}}}{(2\pi)^{-1}\frac{1}{\sqrt{10}} e^{-\frac{(x_1+2x_2-3\theta)^2}{10}} e^{-\frac{(x_3+x_4-2\theta)^2}{4}}} \\
&= \frac{\sqrt{10}}{2\pi} exp\left[h(\tilde{x}) + \theta\sum_{i=1}^{4} x_i - \frac{6}{10}\theta(x_1 + 2x_2) - \theta(x_3 + x_4) - 2\theta^2 + \frac{9}{10}\theta^2 + \theta^2\right]
\end{aligned}
$$

where $h(\tilde{x})$ is a function depends only on $\tilde{x}$.

Clearly, the above ratio depends on θ. By definition of sufficient statistic, (T_3, T_4) is not a sufficient statistics for θ.

$\blacksquare$