Answer sheet 5

Assignment 1. (a) $\widehat{\mu} = \overline{X}$.
(b) Using the CLT, it follows that $\sqrt{n}(\widehat{\mu} - \mu) \xrightarrow{d} N(0,1)$ as $n \to \infty$.
(c) Since $\widehat{\mu}$ is the MLE of μ, using a theorem done in the class, it follows that $\sqrt{n}(\widehat{\mu} - \mu) \xrightarrow{d} N(0, \mathcal{I}_1^{-1}(\mu))$, where $\mathcal{I}_1(\mu)$ is the Fisher information of μ from a single sample. Thus, it follows from part (b) that $\mathcal{I}_1(\mu) = 1$. So, $\mathcal{I}_n(\mu) = n\mathcal{I}_1(\mu) = n$. Hence, the Cramer-Rao lower bound for the variance of an unbiased estimator of μ is $\mathcal{I}_n^{-1}(\mu) = n^{-1}$.
(d) Since $\mathrm{Var}(\overline{X}) = n^{-1}$, it follows that $\widehat{\mu}$ satisfies the Cramer-Rao lower bound for all $n \geq 1$.
(e) $g(\mu) = \mathbb{P}[X_1 \leq 2] = \Phi(2 - \mu)$, where Φ is the cdf of the $N(0,1)$ distribution.
(f) Since g is a bijective function from $\mathbb{R}$ to $(0, \infty)$, it follows from the equivariance property of MLEs that the MLE of $g(\mu)$ is $g(\widehat{\mu}) = \Phi(2 - \overline{X})$.
(g) Note that $g'(\mu) = -\phi(2 - \mu)$, where ϕ is the density function of the $N(0,1)$ distribution. So, using the delta method, it follows that $\sqrt{n}\{g(\widehat{\mu}) - g(\mu)\} \xrightarrow{d} N(0, [g'(\mu)]^2) \equiv N(0, \phi^2(2-\mu))$ as $n \to \infty$.

Assignment 2. (a) This is a 1-parameter exponential family because the support $[\pi, \infty)$ does not depend on the parameter and

$$f(x;\alpha) = \exp\left(-\alpha \log x + \log \alpha + \alpha \log \pi - \log x\right) = \exp\left(\alpha T(x) - \gamma(\alpha) + S(x)\right).$$

Thus using the theorem from slide 100, we have

$$\begin{aligned} \mathbb{E} \log X &= -\mathbb{E}\, T(X) = -\gamma'(\alpha) = \frac{1}{\alpha} + \log \pi \\ \mathrm{Var} \log X &= \mathrm{Var} -T(X) = \mathrm{Var}\, T(X) = \gamma''(\alpha) = \frac{1}{\alpha^2}. \end{aligned}$$

Finally $\mathbb{E}[\log^2 X] = [\mathbb{E} \log X]^2 + \mathrm{Var} \log X = 2\alpha^{-2} + 2\frac{\log \pi}{\alpha} + (\log \pi)^2$.
(b) Note that the likelihood $L(\alpha)$ is zero outside of the set $\mathbf{1}(x_{(1)} \geq \pi)$, where $x_{(1)} = \min\{x_1, x_2, \ldots, x_n\}$. So, it is good enough to consider the maximization of $L(\alpha)$ when the sample points satisfy this condition. Then,

$$\begin{aligned} L(\alpha) &= \prod_{i=1}^{n} \left\{ \frac{\alpha \pi^{\alpha}}{x_i^{\alpha+1}} \right\} = \frac{\alpha^n \pi^{n\alpha}}{\left(\prod_{i=1}^{n} x_i\right)^{\alpha+1}} \\ \Rightarrow \quad \log L(\alpha) &= n \log \alpha + n\alpha \log \pi - (\alpha + 1) \sum_{i=1}^{n} \log x_i \\ \Rightarrow \quad \frac{\partial}{\partial \alpha} \log L(\alpha) &= \frac{n}{\alpha} + n \log \pi - \sum_{i=1}^{n} \log x_i. \end{aligned}$$

Setting $\Delta_\alpha \log L(\alpha) = 0$ yields the solution

$$\widehat{\alpha} = \frac{n}{\sum_{i=1}^{n} \log x_i - n \log \pi}.$$

Since $\partial^2 \log L(\alpha)/\partial \alpha^2 = -n/\alpha^2 < 0$, it follows that the $\widehat{\alpha}$ is the unique maximizer and hence the MLE of α.
(c) Observe that

$$\mathcal{I}_n(\alpha) = \mathbb{E}\left[-\frac{\partial^2 \log L(\alpha)}{\partial \alpha^2}\right] = \frac{n}{\alpha^2}.$$

So, $\mathcal{I}_1(\alpha) = \alpha^{-2}$. Thus, using a theorem done in the class, it follows that

$$\sqrt{n}(\widehat{\alpha} - \alpha) \xrightarrow{d} N(0, \alpha^2)$$

as $n \to \infty$.
(d) Note that for any $y > 0$, we have

$$\mathbb{P}[Y \le y] = \mathbb{P}[X \le \pi \exp(y)] = \int_\pi^{\pi \exp(y)} \frac{\alpha \pi^\alpha}{x^{\alpha+1}} dx = \pi^\alpha \left[\pi^{-\alpha} - (\pi \exp(y))^{-\alpha}\right] = 1 - \exp(-\alpha y).$$

So, the density of Y is given by $f_Y(y) = \alpha \exp(-\alpha y)$ if $y > 0$, and equals zero otherwise. Thus, $Y \sim Exp(\alpha)$.
(e) We know that the mean and the variance of the $Exp(\alpha)$ distribution are α^{-1} and α^{-2}, respectively. So, using the CLT, we have

$$\sqrt{n}\left(\frac{T(Y_1, Y_2, \ldots, Y_n)}{n} - \frac{1}{\alpha}\right) \xrightarrow{d} N\left(0, \frac{1}{\alpha^2}\right)$$

as $n \to \infty$.
(f) $\widehat{\alpha} = n/T(Y_1, Y_2, \ldots, Y_n)$.
Define $h(x) = x^{-1}$ on $(0, \infty)$. So, $h'(x) = -x^{-2}$. Using the delta method, it follows that

$$\begin{aligned}
&\sqrt{n}\left\{h\left(\frac{T(Y_1, Y_2, \ldots, Y_n)}{n}\right) - h\left(\frac{1}{\alpha}\right)\right\} \xrightarrow{d} N\left(0, \left[h'\left(\frac{1}{\alpha}\right)\right]^2 \frac{1}{\alpha^2}\right) \\
\Rightarrow \quad &\sqrt{n}(\widehat{\alpha} - \alpha) \xrightarrow{d} N\left(0, \alpha^2\right)
\end{aligned}$$

as $n \to \infty$. This is the same asymptotic distribution as that obtained in part (c).
(g) Here we have

$$\mathbb{E}\, X = \pi^\alpha \int_\pi^\infty \alpha x^{-\alpha} dx = \pi \frac{\alpha}{\alpha - 1}.$$

(If $\alpha \le 1$ the expectation is infinite.)
We obtain the equation

$$\overline{X}_n = m(\widetilde{\alpha}) = \pi \frac{\widetilde{\alpha}}{\widetilde{\alpha} - 1} = \pi + \frac{1}{\widetilde{\alpha} - \pi}$$

so that $\widetilde{\alpha} = 1 + \pi/(\overline{X}_n - \pi) = \overline{X}_n/(\overline{X}_n - \pi)$.
We will also need variance for the asymptotic distribution. We have $\mathbb{E}\, X^2 = \pi^2\alpha/(\alpha - 2)$ (infinite if $\alpha \le 2$) and $\operatorname{Var} X = \pi^2\alpha/[(\alpha - 2)(\alpha - 1)^2]$.
Thus by the central limit theorem

$$\sqrt{n}\left(\overline{X}_n - \pi\frac{\alpha}{\alpha - 1}\right) \to N\left(0, \pi^2 \frac{\alpha}{(\alpha - 2)(\alpha - 1)^2}\right), \qquad n \to \infty.$$

The function $m^{-1}(x) = 1 + \pi/(x - \pi)$ is differentiable at $\pi\alpha/(\alpha - 1) > 1$ with derivative $-(\alpha - 1)^2/\pi$ at that point. The delta method then gives

$$\sqrt{n}\,(\widetilde{\alpha} - \alpha) = \sqrt{n}\left(m^{-1}(\overline{X}_n) - m^{-1}\left(\pi\frac{\alpha}{\alpha - 1}\right)\right) \to \frac{(\alpha - 1)^2}{\pi} N(0, \operatorname{Var} X) = N\left(0, \frac{\alpha(\alpha - 1)^2}{\alpha - 2}\right).$$

(h)The asymptotic variance of the method of moments estimator $\widetilde{\alpha}$ is $\alpha(\alpha - 1)^2/[(\alpha - 2)n]$.

For the maximum likelihood estimator we have $\sqrt{n}(1/\widehat{\alpha}-1/\alpha) \to N(0,\alpha^{-2})$ and by the delta method the asymptotic variance of $\widehat{\alpha}$ is α^2/n. This is smaller than the asymptotic variance of the method of moments estimator because $\alpha^2 < \alpha(\alpha-1)^2/(\alpha-2)$; the difference is large when α is close to 2. Both asymptotic variances decay like $1/n$.

Assignment 3. The optimal value of a is $n-2=1$. The assignment can be carried out using the following code :

```
set.seed(18102017)
n <- 3
REP <- 1000
mu <- c(-1, 0, 1)
a <- n-2
MSE.mle <- MSE.stein <- numeric(REP)
for(i in 1:REP)
{
Y <- rnorm(n, mean = mu, sd = 1)
Y.norm <- sum(Y^2)
stein <- Y * (1 - a/Y.norm)
MSE.mle[i] <- sum((Y - mu)^2)
MSE.stein[i] <- sum((stein - mu)^2)
}
mean(MSE.mle)
mean(MSE.stein)
```

Assignment 4. (a) We have

$$\mathbb{E}\,\frac{1}{X} = \int_0^\infty \frac{\lambda^k x^{k-2} e^{-\lambda x}}{\Gamma(k)} dx = \frac{\lambda}{k-1}\int_0^\infty \frac{\lambda^{k-1} x^{k-2} e^{-\lambda x}}{\Gamma(k-1)} dx = \frac{\lambda}{k-1},$$

since $k>1$ and the last integrand is the density of a $Gamma(k-1,\lambda)$ distribution. If $k\le 1$, then $\mathbb{E}\,\frac{1}{X}=\infty$.
(b) Put $\lambda = 1/2$ and $k=n/2>1$ because $n>2$.
(c) Up to constants, the log likelihood is the negative of this sum of squares.
(d) The additive nature of the objective function allows for minimisation each μ_i separately. The first derivatives with respect to μ_i are

$$2\mu_i - 2y_i + 2\lambda\mu_i = 2[(1+\lambda)\mu_i - y_i]; \qquad \text{and} \qquad 2(1+\lambda)>0$$

so the unique minimum is attained at $\tilde{\mu}_i = y_i/(1+\lambda)$. In vector form, this can be written $\tilde{\mu}_\lambda = y/(1+\lambda)$.
(e) The mean squared error can be written as the expected value of

$$\sum_{i=1}^n (\tilde{\mu}_i-\mu_i)^2 = (1+\lambda)^{-2}\sum_{i=1}^n (y_i-\mu_i-\lambda\mu_i) = (1+\lambda)^{-2}\sum_{i=1}^n (y_i-\mu_i)^2 + \lambda^2\mu_i^2 - 2\lambda\mu_i(y_i-\mu_i).$$

Since $\mathbb{E}\,y_i=\mu_i$ the last term vanishes and since $\operatorname{Var} y_i = 1$ the first sum is n. Thus the mean squared error equals

$$\frac{1}{(1+\lambda)^2}\left(n+\lambda^2\sum_{i=1}^n \mu_i^2\right) = \frac{1}{(1+\lambda)^2}\left(n+\lambda^2\|\mu\|^2\right).$$

(f) The derivative of the mean squared error with respect to λ is

$$\frac{2\lambda\|\mu\|^2(1+\lambda)^2 - 2(1+\lambda)(n+\lambda^2\|\mu\|^2)}{(1+\lambda)^4} = \frac{2}{(1+\lambda)^3}\left(\lambda\|\mu\|^2 - n\right).$$

This is negative for small λ, so small but positive values of λ have a lower mean squared error than that of $\widehat{\lambda} = \tilde{\lambda}_0$.

(g) Since the derivative is negative for small λ and positive for for large λ, the unique minimum is attained when $\lambda = n/\|\mu\|^2$ (if $\mu \neq 0$). What this means is that the smaller $\|\mu\|$ is, the better it is to penalise it by choosing a high value of λ. In the extreme case where $\mu = 0$, the mean squared error is $n/(1+\lambda)^2$, which is strictly decreasing ; the more we penalise, the better.
The problem with this choice of λ is that it depends on the unknown value of μ. We will later see some ways of choosing λ in practice, most notably *cross-validation*. Note that this problem does not arise with the James–Stein estimator.