1. Which of the following densities are within an exponential family? Explain your reasoning.

(a) Let $0 < \theta < 1$ and put
$$f(x, \theta) = (1 - \theta)^x; \quad x = 0, 1, 2, \ldots$$

(b)
$$f(x; \alpha, \lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} e^{-\lambda x} x^{\alpha - 1}, \quad x > 0,$$
where $\alpha > 0, \lambda > 0$;

(c)
$$f(x, \theta) = e^{-(x-\theta)}, \quad x \geq \theta.$$ 

**Solution:**

(a)
$$f(x, \theta) = (1 - \theta)^x; \quad x = 0, 1, 2, \ldots,$$
where $0 < \theta < 1$, can be written as
$$f(x, \theta) = \exp\{x \log \theta + \log(1 - \theta) + 0\}; \quad x = 0, 1, 2,$$
so $k = 1, c_1 = 1, h_1(x) = x, \phi_1(\theta) = \log \theta, c(\theta) = \log(1 - \theta), d(x) = 0$, and $\mathcal{X} = \{0, 1, 2, \ldots\}$ does not depend on $\theta$, so $f$ is in a 1-parameter exponential family.

(b)
$$f(x; \alpha, \lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} e^{-\lambda x} x^{\alpha - 1}, \quad x > 0,$$
where $\alpha > 0, \lambda > 0$, can be written as
$$f(x; \alpha, \lambda) = \exp\{-\lambda x + (\alpha - 1) \log(x) + \log(\lambda^\alpha / \Gamma(\alpha)) + 0\}$$
so with $\theta = (\alpha, \lambda)$, we can choose $k = 2, c_1 = c_2 = 1, h_1(x) = x, h_2(x) = \log(x), \phi_1(\theta) = \lambda, \phi_2(\theta) = \alpha - 1, c(\theta) = \log(\lambda^\alpha / \Gamma(\alpha)), d(x) = 0$, and $\mathcal{X} = (0, \infty)$ does not depend on $\theta$, so $f$ is in a 2-parameter exponential family.

(c)
$$f(x, \theta) = e^{-(x-\theta)}, \quad x \geq \theta,$$
can be written in the exponential family form with $k = 1, c_1 = 1, h_1(x) = 1, \phi_1(\theta) = \theta, d(x) = -x$, but $\mathcal{X} = [\theta, \infty)$ depends on $\theta$, so it is non-regular.
2. Suppose $X_1, X_2, \ldots, X_n$ is a random sample from the Pareto distribution $f(x, \lambda) = \frac{\lambda \alpha^\lambda}{x^{\lambda+1}}$ with $x > \alpha$, $\lambda > 0$, and $\alpha > 0$ known. Find a minimal sufficient statistic for $\lambda$.

**Solution:** The likelihood is 

$$L(\lambda, x) = \frac{(\lambda \alpha^\lambda)^n}{\prod_{j=1}^n x_j^{\lambda+1}} = \frac{\lambda^n \alpha^{n\lambda}}{(\prod_{j=1}^n x_j)^{\lambda+1}},$$

so $T = \prod_{j=1}^n X_j$ is sufficient.

For $x, y$ we have 

$$\frac{L(\lambda, x)}{L(\lambda, y)} = \left(\frac{\prod_{j=1}^n y_j}{\prod_{j=1}^n x_j}\right)^{\lambda+1}$$

is constant in $\lambda$ if and only if $t(x) = t(y)$, so $T$ is minimal sufficient.

3. Suppose $X_1, X_2, \ldots, X_n$ is a random sample from the log-normal distribution with density 

$$f(x, \mu, \phi) = \frac{1}{\sqrt{2\pi\phi}} \exp\left\{-\frac{1}{2\phi}(\log x - \mu)^2\right\}$$

with $\phi > 0$ (so that $\log X_j \sim N(\mu, \phi)$). Find a minimal sufficient statistic for the parameter $\theta = (\mu, \phi)$.

**Solution:**

$$L(\theta, x) = \frac{1}{(\prod x_i)(2\pi\phi)^{n/2}} \exp\left\{-\frac{1}{2\phi} \sum_{i=1}^n (\log x_i - \mu)^2\right\}$$

and the expression in the exponential can be written as 

$$-\frac{1}{2\phi} \sum (\log x_i)^2 + \frac{\mu}{\phi} \sum \log x_i - \frac{n\mu^2}{2\phi},$$

so $T = (\sum \log X_i, \sum (\log X_i)^2)$ is sufficient.

By considering the likelihood ratio we see that $T$ is indeed minimal sufficient.

4. Suppose $X_1, \ldots, X_n$ are independent and exponentially distributed, each with density function 

$$f(x; \theta) = \frac{1}{\theta} e^{-x/\theta}, \quad x \geq 0.$$ 

Let $\overline{X} = \frac{1}{n} \sum_{i=1}^n X_i$ and put 

$$T = \frac{\overline{X}}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X})^2}}.$$ 

Show that $T$ is an ancillary statistic. What does this say about $t$-tests on exponential data?
Solution: Note that $Y_i = \frac{1}{\theta} X_i$ has the exponential distribution with parameter 1, so its distribution does not depend on $\theta$. Next note that

$$T = \frac{\bar{Y}}{\sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \bar{Y})^2}}.$$

Hence the distribution of $T$ does not depend on $\theta$ neither; it is an ancillary statistic. Thus a $t$-test based on exponential data would not reveal any information about the parameter $\theta$.

5. Let $X_1, \ldots, X_n$ be i.i.d. uniform $U\left[\theta - \frac{1}{2}, \theta + \frac{1}{2}\right]$ random variables.
   a) Show that $(X_{(1)}, X_{(n)})$ is minimal sufficient for $\theta$.
   b) Show that $(S, A) = \left(\frac{1}{2}(X_{(1)} + X_{(n)}), X_{(n)} - X_{(1)}\right)$ is minimal sufficient for $\theta$, and that the distribution of $A$ is independent of $\theta$ (so $A$ is an ancillary statistic).
   c) Show that any value in the interval $[x_{(n)} - \frac{1}{2}, x_{(1)} + \frac{1}{2}]$ is a maximum-likelihood-estimator for $\theta$.

Solution:
   a) The likelihood is

   $$L(\theta, x) = 1 \left( \theta - \frac{1}{2} \leq x_1, \ldots, x_n \leq \theta + \frac{1}{2} \right)$$

   $$= 1 \left( \theta - \frac{1}{2} \leq x_{(1)}, x_{(n)} \leq \theta + \frac{1}{2} \right)$$

   which is a function of $(\theta, x_{(1)}, x_{(n)})$, so $(X_{(1)}, X_{(n)})$ is sufficient. For minimal sufficiency, note that $L(\theta, x) = L(\theta, y)$ if and only if $(x_{(1)}, x_{(n)}) = (y_{(1)}, y_{(n)})$, proving the first assertion.
   b) We have that

   $$X_{(1)} = S - \frac{1}{2} A$$

   $$X_{(n)} = S + \frac{1}{2} A$$

   so $L(\theta, x)$ is a function of $(\theta, s, a)$, hence $(S, A)$ is sufficient, and $(S, A)$ is a function of a minimal sufficient statistic, so must be minimal sufficient itself. To see that the distribution of $A$ is independent of $\theta$, write $Y_i = X_i - \theta$, then $Y_i \sim U(-1/2, 1/2)$, and $Y_{(i)} = X_{(i)} - \theta$. Hence $A = X_{(n)} - X_{(1)} = Y_{(n)} - Y_{(1)}$, the difference of order statistics from a distribution which does not involve $\theta$. So the distribution of $A$ does not depend on $\theta$.
   c) Follows directly from the likelihood - the likelihood

   $$L(\theta, x) = 1 \left( x_{(n)} - \frac{1}{2} \leq \theta \leq x_{(1)} + \frac{1}{2} \right)$$

   is constant equal to 1 on the interval $[x_{(n)} - \frac{1}{2}, x_{(1)} + \frac{1}{2}]$, and 0 otherwise.

6. The random variables $X_1, \ldots, X_n$ are independent with geometric distribution $P(X_i = x) = p(1 - p)^{x-1}$ for $x = 1, 2, \ldots$. Let $\theta = p^{-1}$.  

3
(i) Find the maximum likelihood estimator for $p$. Considering $n = 1$, is it unbiased?
(ii) Show that $\hat{\theta} = \overline{X}$ is the maximum likelihood estimator for $\theta$. Is it unbiased?
(iii) Compute the expected Fisher information for $\theta$.
(iv) Does $\hat{\theta}$ attain the Cramer-Rao lower bound?

Solution:

a) We have that

$$L(p, x) = p^n (1 - p)^{\sum_{i=1}^n x_i - n}$$

and so

$$\ell(p) = n \log p + \left( n \sum_{i=1}^n x_i - n \right) \log(1 - p);$$

differentiate:

$$\ell'(p) = \frac{n}{p} - \frac{n \sum_{i=1}^n x_i - n}{1 - p};$$

the only solution for $\ell'(p) = 0$ is $p = \frac{1}{\overline{X}}$. The second derivative is

$$\ell''(\theta) = -\frac{n}{p^2} - \frac{n \sum_{i=1}^n x_i - n}{(1 - p)^2} < 0,$$

so that $\hat{p} = \frac{1}{\overline{X}}$ is the mle.

For $n = 1$,

$$E \left( \frac{1}{\overline{X}} \right) = \sum_{x=1}^{\infty} \frac{1}{x} p(1 - p)x^{-1}$$

$$= p + \sum_{x=2}^{\infty} \frac{1}{x} p(1 - p)x^{-1}$$

$$> p,$$

so $\hat{p}$ is not unbiased.

b) We have that

$$L(\theta, x) = \theta^{-n}(1 - \theta^{-1})^{\sum_{i=1}^n x_i - n}$$

and so

$$\ell(\theta) = -n \log \theta + \sum_{i=1}^n \left( x_i - 1 \right) \log(1 - \theta^{-1});$$
differentiate:

\[ \ell'(\theta) = \frac{n}{\theta(\theta - 1)}(\bar{x} - \theta); \]

the only solution for \( \ell'(\theta) = 0 \) is \( \theta = \bar{x} \). The second derivative is

\[ \ell''(\theta) = \frac{n}{(\theta(\theta - 1))^2} \left\{ -\theta(\theta - 1) - (\bar{x} - \theta)(2\theta - 1) \right\}; \]

so that

\[ \ell''(\hat{\theta}) = \frac{n}{(\hat{\theta}(\hat{\theta} - 1))^2}(-\hat{\theta}(\hat{\theta} - 1)) < 0 \]

if \( \hat{\theta} = \bar{x} > 1 \), so if \( \hat{\theta} > 1 \), the mle is \( \hat{\theta} = \bar{x} \).

If \( \bar{x} = 1 \), then

\[ L(\theta, x) = \theta^{-n}, \]

and we know that \( \theta \geq 1 \), so the likelihood is maximized for \( \theta = 1 \).

c) Calculate that \( E_\theta \bar{X} = \theta, \operatorname{Var}_\theta(\bar{X}) = \frac{\theta(\theta - 1)}{n} \), so, from \( \ell'(\theta) \),

\[ I(\theta) = \frac{n^2}{\theta^2(\theta - 1)^2} \operatorname{Var}_\theta(\bar{X}) \]

\[ = \frac{n}{\theta(\theta - 1)} \]

which equals the inverse of the variance of \( \bar{X} \). So it attains the Cramer-Rao lower bound.

Note: As \( E_\theta \bar{X} = \theta \), the estimator is unbiased, and we have seen that it attains the Cramer-Rao lower bound, hence it has minimum variance among all unbiased estimators.

7. Let \( X_1, \ldots, X_n \) be i.i.d. \( U[0, \theta] \), having density

\[ f(x; \theta) = \frac{1}{\theta}, \quad 0 \leq x \leq \theta \]

with \( \theta > 0 \).

(i) Estimate \( \theta \) using both the method of moments and maximum likelihood.

(ii) Calculate the means and variances of the two estimators.

(iii) Which one should be preferred and why?

Solution: a) We have \( E_\theta(X) = \frac{\theta}{2} \), so the m.o.m. is \( \tilde{\theta} = 2\bar{X} \).

The likelihood is

\[ L(\theta, x) = \theta^{-n} 1(0 \leq x_1, \ldots, x_n \leq \theta) \]

\[ = \theta^{-n} 1(x_{(n)} \leq \theta) \]

5
and this is maximized for $\hat{\theta} = x(n)$.

b) We have by construction

$$E_{\theta}(\hat{\theta}) = \theta,$$

and

$$Var_{\theta}(\hat{\theta}) = \frac{4\theta^2}{12n} = \frac{\theta^2}{3n}.$$  

For $\hat{\theta}$, calculate (see second year probability, or the book by Rice)

$$E_{\theta}(\hat{\theta}) = \frac{n}{n+1}\theta, \quad Var_{\theta}(\hat{\theta}) = \frac{n}{(n+1)^2(n+2)}\theta^2.$$  

c) The m.l.e. is not unbiased but asymptotically unbiased, and has much smaller variance than the m.o.m., so I would prefer it. You may decide otherwise if you put more emphasis on unbiasedness.

8. Suppose $X_1, \ldots, X_n$ are a random sample with mean $\mu$ and finite variance $\sigma^2$. Use the delta method to show that, in distribution, 

$$\sqrt{n}(\bar{X}^2_n - \mu^2) \to \mathcal{N}(0, 4\mu^2\sigma^2).$$

What would you suggest if $\mu = 0$?

*Solution:* From the central limit theorem we know that $\sqrt{n}(\bar{X}_n - \mu) \to \mathcal{N}(0, \sigma^2)$. We use the function $g(s) = s^2$, so that $g'(s) = 2s$; now the delta method gives that

$$\bar{X}^2_n \approx \mathcal{N}(\mu^2, 4\mu^2\sigma^2 / n).$$

Standardizing gives the result.

If $\mu = 0$ then the result just gives convergence to point mass at 0, which is not informative. Instead we could use that $\sqrt{n}\frac{\bar{X}^2_n}{\sigma^2} \to \mathcal{N}(0,1)$, so that $n\frac{\bar{X}^2_n}{\sigma^2} \approx \chi^2_1$, and, in distribution,

$$n\frac{\bar{X}^2_n}{\sigma^2} \to \sigma^2\chi^2_1.$$