Foundations of Statistical Inference, BS2a, Exercises 1

1. Let $X_1, \dots, X_n$ be independent Poisson random variables with means $\mathbb{E}(X_i) = \lambda m_i$, $i = 1, \dots, n$ where $\lambda > 0$ is unknown and $m_1, \dots, m_n$ are known constants.
(a) Show that the model defines a canonical exponential family with canonical parameter $\theta = \log \lambda$.
(b) What is the canonical sufficient statistic? Find its mean and variance.
(c) Find the MLE $\widehat{\theta}$ of θ.
(d) Find the Fisher information for θ. What statements can you make about the variance of $\widehat{\theta}$.

2. A random sample $X_1, \dots, X_n$ is taken from the Weibull distribution

$$\frac{\beta}{\alpha^\beta} x^{\beta-1} \exp\left\{ - \left(\frac{x}{\alpha} \right)^\beta \right\}, \ x > 0, \alpha > 0, \beta > 0.$$

Assuming that β is known, find a single function of $X_1, \dots, X_n$ which is sufficient for α. Show that however if α is known there is no single function of $X_1, \dots, X_n$ which is sufficient for β. Does the Weibull distribution belong to a 2-parameter exponential family?

3. Let $X_1, \dots, X_n$ be a random sample from the density

$$f(x; \theta) = e^{-(x-\theta)}, \ x > \theta$$

(a) Show that the MLE $\widehat{\theta}$ of θ is the minimum of $X_1, \dots, X_n$.
(b) Show that $\hat{\theta}$ is a sufficient for θ. Is the family a regular exponential family?
(c) Show that for all $\epsilon > 0$

$$P_\theta[|\widehat{\theta} - \theta| > \epsilon] \leq e^{-n\epsilon},$$

deduce that $\widehat{\theta}$ is consistent in probability and in quadratic mean, but is biased estimator of θ with $\mathbb{E}[\widehat{\theta}] = \theta + 1/n$. Suggest an unbiased and consistent estimator and find its variance.

4. Let $X_1, \dots, X_n$ be a random sample from a distribution with density

$$f(x; \theta) = \frac{1}{2} \theta^3 x^2 e^{-\theta x}, \ x > 0.$$

(a) Rewrite the density in standard exponential form, giving $A(\theta)$, $B(X)$, $C(X)$, $D(\theta)$ explicitly.

(b) Find a minimal sufficient statistic for θ, $T(X)$. Find the expected value of the statistic.
(c) Find the maximum likelihood estimator for θ. Is it unbiassed for θ?
(d) Show that $\theta^* = (2/n)\sum_{i=1}^n X_i^{-1}$ is an unbiased estimator of θ and find its variance.
(e) Compute the Fisher information matrix $I_n(\theta)$ of the model and compare the variance of θ^* with $I_n(\theta)$.

5. Let $X_1, \ldots, X_n$ be a sample from $N(\mu, \sigma^2)$.
(a) Show that the MLE of σ^2 is

$$\widehat{\sigma}^2 = n^{-1}\sum_{i=1}^n (X_i - \bar{X})^2.$$

(b) Show that $\widehat{\sigma}^2$ has a smaller mean square error than

$$(n-1)^{-1}\sum_{i=1}^n (X_i - \bar{X})^2.$$

(c) For which value of a is the MSE of

$$(n+a)^{-1}\sum_{i=1}^n (X_i - \bar{X})^2$$

the smallest. Hint: For (b) and (c) you will need to find $\mathrm{Var}(\chi^2_{n-1})$ which is a special case of the variance of a gamma distribution.

6. a) Let $Y_1, \cdots, Y_n$ be a random sample from a Posson distribution with parameter $\lambda > 0$. One observes only $W_i = \mathbf{1}_{Y_i > 0}$. Compute the likelihood associated with the sample $(W_1, \cdots, W_n)$ and the MLE in λ. Show that it is consistent in probability.
b) Let $X_1, \ldots, X_n$ be a random sample from a truncated Poisson distribution with distribution

$$f(x; \lambda) = \frac{e^{-\lambda}}{1 - e^{-\lambda}} \cdot \frac{\lambda^x}{x!},\ x = 1, 2, \ldots.$$

For $i = 1, \ldots, n$ a random variable Z_i is defined by

$$Z_i = X_i \text{ if } X_i \geq 2 \text{ or } Z_i = 0 \text{ if } X_i = 1$$

Show that $\bar{Z}$ is an unbiased estimator of λ with efficiency

$$\frac{1 - e^{-\lambda}}{1 - \left(\frac{\lambda e^{-\lambda}}{1 - e^{-\lambda}}\right)^2}.$$

7. (a) (optional bookwork) Let X be a discrete random variable with pmf $f(x;\theta)$ with parameter $\theta \in \Theta$ and sample space $X \in \chi$. Let $T(x)$ be a function of x. Suppose $f(x;\theta)/f(y;\theta)$ is not a function of θ if and only if $T(x) = T(y)$. Show that $T(x)$ is minimal sufficient for θ.
(b) Let $N = N(0, S]$ be the number of events in a Poisson arrival process of rate λ acting over time s in the interval $0 < s \leq S$. Suppose we observe arrivals in the process at times $X_1, X_2, ..., X_N$, and wish to use these data to estimate λ. Show that N is minimal sufficient for λ (assume the result in (a) holds for any sufficiently regular family of probability distributions).