Foundations of Statistical Inference, BS2a
Exercises 1: brief solutions

Ex 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 $\widehat{\theta}$.

Sol Ex1
(a)

$$\begin{aligned} L(\lambda,\mathbf{x}) &= \prod_1^n e^{-\lambda m_i}(\lambda m_i)^{x_i}/x_i! \\ &= \exp\left\{(\log\lambda)\sum_1^n x_i - \lambda\sum_1^n m_i + \sum_1^n x_i\log m_i - \sum_1^n \log(x_i!)\right\} \end{aligned}$$

which is in canonical exponential form with $\theta = \log\lambda$, $B_1(x) = \sum_1^n x_i$
(b) The canonical (minimal) sufficient statistic is $\bar{X}$ ($n\bar{X}$ is fine as well). $\mathbb{E}[\bar{X}] = \lambda\bar{m}$. $\sum_1^n X_i$ is Poisson ($\lambda\sum_1^n m_i$) so $\mathrm{Var}(\bar{X}) = \lambda\bar{m}/n$.
(c) $\ell(\theta) = \text{const} + \theta\sum_1^n x_i - e^\theta\sum_1^n m_i$, $\partial\ell/\partial\theta = \sum_1^n x_i - e^\theta\sum_1^n m_i$, so $\widehat{\theta} = \log[\bar{x}/\bar{m}]$, provided $\bar{x} > 0$. If $\bar{x} = 0$ then $\widehat{\lambda} = 0$, $\widehat{\theta} = -\infty$. As $n\to\infty$ the probability that $\bar{X} = 0$ tends to zero if $\lambda > 0$.
(d) Fisher information $I_\theta = -\mathbb{E}[\partial^2 l/\partial\theta^2] = e^\theta\sum_1^n m_i$. The variance of $\widehat{\theta}$ is asymptotic to $I_\theta^{-1} = e^{-\theta}/\sum m_i$. There is a problem with the exact variance being infinite because $\bar{X} = 0$ has positive probability.

Ex 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?

Solution Ex 2

$$L(\theta;\mathbf{x}) = \alpha^{-n\beta}\exp\{-\alpha^{-\beta}\sum_1^n x_i^\beta\} \times \beta^n \prod_1^n x_i^{\beta-1}.$$

Assuming β is known constant, this is exponential form in canonical parameter $-\alpha^{-\beta}$. $n^{-1}\sum_1^n x_i^\beta$ is thus a (minimal) sufficient statistic for α if β is known. If α is known we cannot express $\sum_1^n x_i^\beta$ as a single function of $(x_1,\ldots,x_n)$ which does not depend on β times a function of β. The Weibull distribution cannot belong to a 2-parameter exponential family, since then there would exist sufficient statistics for α and β.

Ex 3. Let $X_1,\ldots,X_n$ 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 $X_1,\ldots,X_n$.
(b) Show that $\hat{\theta}$ is a sufficient for θ. Is the family a regular exponential family?
(c) Show that for all $\epsilon > 0$

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

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.

Solution Ex 3

$$L(\theta;\mathbf{x}) = e^{-\sum_1^n x_i + n\theta}\prod_{i=1}^n I[x_i > \theta]$$

Note that X_1 is just θ plus a mean 1 exponential r.v.
(a) $\widehat{\theta} = \min_i X_i$ maximizes $L(\theta,\mathbf{x})$.
(b)

$$L(\theta;\mathbf{x}) = e^{n\theta}\mathbb{1}_{\min(x_i)>\theta} \times e^{-n\bar{x}}$$

so it factorizes into $f_1(\min(x_i);\theta)h(x)$ and $\min(x_i)$ is sufficient. The family is not an exponential family
(c)

$$P(\widehat{\theta}-\theta > \epsilon) = P(\forall i X_i > \epsilon + \theta) = e^{-n\epsilon},\ \epsilon > 0$$

and

$$P(\widehat{\theta} - \theta < -\epsilon) = 0$$

Hence for all $\epsilon > 0$

$$\lim_n P_\theta[|\widehat{\theta} - \theta| > \epsilon] = 0$$

and it is consistent in probability Alternatively, since $\min(X_1, \ldots, X_n)$ is just $\theta + \min(Z_1, \ldots, Z_n) = \theta + \xi$ where the Z are iid exponential(1) variables and ξ is an exponential(n) rv, we have that

$$E[\hat{\theta}] = \theta + 1/n, \, V(\hat{\theta}) = n^{-2}.$$

We thus have

$$E[(\hat{\theta} - \theta)^2] = 2/n^2 = o(1)$$

it is not unbiased . An unbiased estimator is $\tilde{\theta} = \hat{\theta} - 1/n$.

Ex 4. Let $X_1, \ldots, X_n$ 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 $I_n(\theta)$.

Solutions to Ex 4

$$f(x;\theta) = \frac{1}{2}\theta^3 x^2 e^{-\theta x} = \exp\left\{-\theta x + \log(\frac{1}{2}x^2) + 3\log\theta\right\}$$

$$A(\theta) = -\theta, B(x) = x, C(x) = \log(\frac{1}{2}x^2), D(\theta) = 3\log\theta$$

or

$$A(\theta) = -\theta, B(x) = x, \omega(x) = \frac{1}{2}x^2, \psi(\theta) = \theta^3$$

(b)

$$L(\theta;x) \propto \theta^{3n} e^{-\theta\sum x_i} \times \prod x_i^2$$

By the factorization theorem $\bar{x}$ is a minimal sufficient statistic for θ ($K_1(\bar{x},\theta) = \theta^{3n}e^{-n\bar{x}}$ and $K_2(x) = \prod x_i^2$ and the mean is $3/\theta$.
(c) $l(\theta) = 3n\log\theta - \theta\sum x_i + \text{const}$, $l'(\theta) = 3n/\theta - \sum x_i$ so $\hat{\theta} = 3/\bar{x}$.

Recall: Gamma density with parameters (α,β) is $\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}$. If $X \sim \Gamma(a_1,\beta), Y \sim \Gamma(a_2,\beta)$ and independent then $X+Y \sim \Gamma(a_1+a_2,\beta)$. Mean of $\Gamma(\alpha,\beta)$ is α/β.

$\sum_1^n X_i$ has a Gamma distribution with density

$$\frac{\theta^{3n}}{\Gamma(3n)}x^{3n-1}e^{-\theta x} \quad x > 0$$

so

$$\begin{aligned}
\mathbb{E}[\widehat{\theta}] &= 3n\int_0^\infty \frac{\theta^{3n}}{\Gamma(3n)}x^{3n-2}e^{-\theta x}dx \\
&= 3n\cdot\frac{\theta^{3n}}{\Gamma(3n)}\cdot\frac{\Gamma(3n-1)}{\theta^{3n-1}} \\
&= \frac{3n\theta}{3n-1}
\end{aligned}$$

Thus $\widehat{\theta}$ is a biassed estimate of θ.
(d)

$$\begin{aligned}
\mathbb{E}[X_i^{-1}] &= \int_0^1 \frac{1}{2}\theta^3 x e^{-\theta x}dx \\
&= \frac{1}{2}\theta
\end{aligned}$$

so $\theta^* = (2/n)\sum_1^n X_i^{-1}$ is an unbiassed estimate of θ. Similarly, from the density, one can show that $\mathrm{Var}(\theta^*) = \theta^2/n$.
(e) Fisher's information is $I_\theta = -E[\frac{\partial^2}{\partial\theta^2}\ell(\theta)] = 3n/\theta^2$ $\mathrm{Var}(\theta^*) = \theta^2/n \geq \theta^2/(3n)$.

Ex 5.

Let $X_1,\dots,X_n$ 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.

Solution to Ex 5

(a)

$$\ell(\mu,\sigma^2) = \text{const} - \frac{n}{2}\log\sigma^2 - \frac{1}{2}\sum_{1}^{n}(x_i-\mu)^2/\sigma^2$$

so

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{1}^{n}(x_i-\mu)^2$$

and the MLE of μ is $\bar{x}$, uniformly in σ^2, so

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

Recap from Part A Statistics

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

$$\widehat{\sigma}^2 = \frac{n-1}{n}S^2 \sim \frac{\sigma^2}{n}\chi^2_{n-1}$$

χ^2_f has a density

$$\frac{1}{\Gamma(f/2)2^{f/2}}x^{f/2-1}e^{-x/2},\ x>0$$

which is a Gamma $(f/2, 1/2)$ density with mean $2\times f/2 = f$ and variance $4\times f/2 = 2f$.

(b) $\mathbb{E}(\widehat{\sigma}^2) = ((n-1)/n)\sigma^2$, $\mathrm{Bias}(\widehat{\sigma}^2) = -\sigma^2/n$, $\mathrm{Var}(\widehat{\sigma}^2) = (2(n-1)/n^2)\sigma^4$. Thus

$$\mathrm{MSE}(\widehat{\sigma}^2) = \mathrm{Var}(\widehat{\sigma}^2) + \mathrm{Bias}(\widehat{\sigma}^2)^2 = \frac{2n-1}{n^2}\sigma^4$$

Let

$$S^2 = (n-1)^{-1}\sum_{i=1}^{n}(X_i-\bar{X})^2 \sim \frac{\sigma^2}{n-1}\chi^2_{n-1},$$

then

$$\mathrm{MSE}(S^2) = \frac{2(n-1)}{(n-1)^2}\sigma^4 = \frac{2}{n-1}\sigma^4 > \mathrm{MSE}(\widehat{\sigma}^2) = \frac{2n-1}{n^2}\sigma^4.$$

(c) Let

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

A similar calculation to (b) shows that

$$\mathrm{MSE}(\sigma^{*2}) = \left(\frac{2}{n-1}b^2 + (b-1)^2\right)\sigma^4$$

where $b = (n-1)/(n+a)$. The MSE is minimal when

$$b = \frac{1}{\frac{2}{n-1}+1}, \text{ or } a = 1$$

That is the minimal MSE solution is

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

Exo 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

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

Solution to Ex 6

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

The mean of Z is

$$\mathbb{E}[Z] = \sum_{x\geq 2} x\frac{e^{-\lambda}}{1-e^{-\lambda}}\frac{\lambda^x}{x!} = \frac{e^{-\lambda}}{1-e^{-\lambda}}\sum_{x\geq 2}\frac{\lambda^x}{(x-1)!} = \frac{e^{-\lambda}}{1-e^{-\lambda}}\lambda(e^{\lambda}-1) = \lambda$$

To obtain the variance consider

$$\mathbb{E}[Z(Z-1)] = \sum_{x\geq 2} x(x-1)\frac{e^{-\lambda}}{1-e^{-\lambda}}\frac{\lambda^x}{x!} = \frac{e^{-\lambda}}{1-e^{-\lambda}}\sum_{x\geq 2}\frac{\lambda^x}{(x-2)!} = \frac{\lambda^2}{1-e^{-\lambda}}$$

Then

$$\mathrm{Var}(Z) = \frac{\lambda^2}{1-e^{-\lambda}} + \lambda - \lambda^2 = \lambda\left[1 + \frac{\lambda e^{-\lambda}}{1-e^{-\lambda}}\right]$$

The loglikelihood

$$l(\lambda) = -\lambda - \log(1-e^{-\lambda}) + x\log\lambda - \log x!$$

and

$$\begin{aligned}\frac{\partial l}{\partial\lambda} &= -\frac{1}{1-e^{-\lambda}} + \frac{x}{\lambda}\\ \frac{\partial^2 l}{\partial\lambda^2} &= \frac{e^{-\lambda}}{(1-e^{-\lambda})^2} - \frac{x}{\lambda^2}\end{aligned}$$

Fisher information for one observation is

$$\begin{aligned}i_\lambda &= -\mathbb{E}\left(\frac{\partial^2 l}{\partial\lambda^2}\right)\\ &= -\frac{e^{-\lambda}}{(1-e^{-\lambda})^2} + \frac{1}{\lambda^2}\left[\frac{\lambda e^{-\lambda}}{1-e^{-\lambda}} + \lambda\right]\\ &= \frac{1}{\lambda}\cdot\frac{1}{1-e^{-\lambda}}\left[1 - \frac{\lambda e^{-\lambda}}{1-e^{-\lambda}}\right]\end{aligned}$$

I have

$$\begin{aligned}i_\lambda &= -\mathbb{E}\left(\frac{\partial^2 l}{\partial\lambda^2}\right)\\ &= -\frac{e^{-\lambda}}{(1-e^{-\lambda})^2} + \frac{\lambda}{\lambda^2}\\ &= \frac{1}{\lambda}\cdot\frac{1}{1-e^{-\lambda}}\left[1 - \frac{\lambda e^{-\lambda}}{1-e^{-\lambda}}\right]\end{aligned}$$

$$\begin{aligned}\text{Efficiency} &= \left[I_\lambda \text{Var}(\bar{Z})\right]^{-1} \\ &= \frac{1-e^{-\lambda}}{1-\left(\frac{\lambda e^{-\lambda}}{1-e^{-\lambda}}\right)^2}.\end{aligned}$$

7. (a) Break the condition into two parts:

(*) $T(x) = T(y) = t$ implies $f(x;\theta)/f(y;\theta)$ is not a function of θ;

(**) $f(x;\theta)/f(y;\theta)$ not a function of θ implies $T(x) = T(y) = t$.

Let $f(x;\theta) = g(x|t(x),\theta)h(t|\theta)$ (with no assumption of sufficiency) and suppose $T(x) = T(y) = t$. If (*) holds then

$$\frac{f(x;\theta)}{f(y;\theta)} = \frac{g(x|t,\theta)}{g(y|t,\theta)} = c(x,y)$$

say, with c independent of θ (factors of h cancel). But then

$$\sum_{x:T(x)=t} g(x|t,\theta) = g(y|t,\theta) \sum_{x:T(x)=t} c(x,y)$$

so

$$g(y|t,\theta) = \left[\sum_{x:T(x)=t} c(x,y)\right]^{-1}$$

which is independent of θ, so T is sufficient for θ in f. If $f(x;\theta)/f(y;\theta)$ does depend on θ when $T(x) = T(y) = t$ then c depends on θ and the same reasoning shows T cannot be sufficient, so condition (*) is necessary for sufficiency. Let $U(x)$ be some sufficient statistic. We must show that T is a function of U, so T is minimal. It is enough to show that $U(x) = U(y)$ implies $T(x) = T(y)$. But $U(x) = U(y) = u$ implies $f(x;\theta)/f(y;\theta)$ is not a function of θ, and then $(**)$ implies $T(x) = T(y)$, so T is minimal sufficient.

(b) The intervals of a Poisson arrival process of rate λ are exponential so $X_i \sim \text{Exp}(\lambda)$ likelihood for $i = 1, 2, ..., N$. The probability that the final interval between time $Y = \sum_{i=1}^N X_i$ and S has no event is the probability that an $\text{Exp}(\lambda)$ random variable exceeds $S - Y$, that is, $\exp(-\lambda(S-Y))$.

The likelihood for λ given data $X = (x_1, ...x_n)$ is therefore

$$\begin{aligned} L(\theta; x) &= \left[\prod_{i=1}^{n} \lambda \exp(-\lambda x_i)\right] \exp(-\lambda(S - Y)) \\ &= \exp(-\lambda S)\lambda^n \end{aligned}$$

since $(S - Y) + x_n + ... + x_1 = S$ and so N is sufficient for λ by the factorization theorem ($L = K_1(x, \theta)K_2(x)$ with $K_1(x, \theta) = L$ and $K_2 = 1$). It is minimal sufficient by part (a) since, if $x = (x_1, ...x_n)$ and $y = y_1, ..., y_m$ then $L(x; \lambda)/L(y; \lambda)$ is independent of λ if and only if $n = m$.