Spring 2002, Aaron Lanterman

ECE 7251: Signal Detection and Estimation
In-Class Quiz I - Solutions

1. This is a problem on exponential families and sufficient statistics.

   (a) For each of the following parametric densities, state whether the density is in an exponential family. A simple "yes" or "no" will suffice; no explanation is needed. Don't spend too much time on these; go with your gut feeling.

   i. (2 points)

$$p(y;\theta) = \begin{cases} \frac{1}{\theta}\exp\left[-\frac{y}{\theta}\right] & \text{for } y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

   Solution: Yes.

   ii. (2 points)

$$p(y;\theta) = \begin{cases} \frac{1}{1-e^{-\theta}}\exp[-y] & \text{for } 0 \leq y \leq \theta \\ 0 & \text{otherwise} \end{cases}$$

   Solution: No. If the support of the density (i.e. set over which it is nonzero) depends on θ, then you automatically know it can't be an exponential family.

   iii. (2 points)

$$p(y;\theta) = \begin{cases} \exp[-(y-\theta)] & \text{for } y \geq \theta \\ 0 & \text{otherwise} \end{cases}$$

   Solution: No, same reason as (b).

   iv. (2 points)

$$p(y;\theta) = \frac{1}{2\theta}\exp\left[-\frac{|y|}{\theta}\right]$$

   Solution: Yes.

   v. (2 points)

$$p(y;\theta) = \frac{1}{2}\exp[-|y-\theta|]$$

   Solution: No.

   (b) (5 points) Suppose that y_i are independent and identically distributed samples from the density function

$$p(y;\theta) = \begin{cases} (\theta-1)y^{-\theta} & \text{for } y \geq 1 \\ 0 & \text{otherwise} \end{cases}$$

   Suppose n independent observations are made. Find a sufficient statistic for θ.

Solution:

$$p(y;\theta) = \prod_{i=1}^{n}(\theta-1)y_i^{-\theta} = (\theta-1)^n \prod_{i=1}^{n} y_i^{-\theta} = (\theta-1)^n \exp[-\theta \sum_{i=1}^{n} \ln y_i]$$

This is the decomposition we need. $\sum_{i=1}^{n} \ln y_i$ is a sufficient statistic, which you could also write as $\ln \prod_{i=n} y_i$; hence $\prod_{i=1}^{n} y_i$ is also a sufficient statistic. Either answer will do.

2. Dilbert's evil pointy-haired boss asks him to consider the case of estimating the mean θ of a Poisson random variable from n independent observations y_i, $i = 1, \ldots, n$. After some work (like what we did in class one day), Dilbert learns that the maximum-likelihood estimator is $\hat{\theta}_{ML}(y) = \sum_{i=1}^{n} y_i/n$ and that the Fisher information is $F(\theta) = N/\theta$.

   A few days later, Dilbert's evil pointy-haired boss walks by and says, "Oh, I'm sorry. What we *really* need is the ML estimate and the Fisher information for the parameter $\gamma = \ln(\theta)$. You have to start all over." Since you're in the cubicle next to Dilbert's cubicle, you overhear the conversation. After the evil pointy-haired boss leaves, you offer to help.

   (a) (5 points) What is the maximum-likelihood estimate of γ?

   Solution:

$$\hat{\gamma}(y) = \ln(\hat{\theta}(y)) = \ln(\sum_{i=1}^{n} y_i/n)$$

   (b) (5 points) What is the Fisher information for estimating γ? (Be sure to express your answer in terms of γ, i.e. don't leave any leftover θ's in your answer.)

   Solution: Define the forward and inverse mappings $g(x) = \ln(x)$ and $g^{-1}(x) = e^x$; There are two ways to do this.

   Approach 1:

$$\frac{dg(\theta)}{d\theta} = \frac{d\ln(\theta)}{d\theta} = \frac{1}{\theta}$$

$$F_\gamma(\theta) = (\frac{dg(\theta)}{d\theta})^{-2} F(\theta) = \theta^2 \frac{N}{\theta} = \theta N$$

   Evaluating that at $\theta = g^{-1}(\gamma) = e^\gamma$ yields $F_\gamma(\gamma) = e^\gamma N$.

   Approach 2:

$$\frac{dg^{-1}(\gamma)}{d\gamma} = \frac{de^\gamma}{d\gamma} = e^\gamma$$

$$F_\gamma(\gamma) = (\frac{dg^{-1}(\gamma)}{d\gamma})^2 F(g^{-1}(\gamma)) = e^{2\gamma}\frac{N}{e^\gamma} = e^\gamma N$$

(c) (5 points) Later that day, you meet Ratbert and Catbert in the cafeteria. Ratbert tells you about an estimator he developed for a certain inference problem (not related to Dilbert's problem above) which has a variance of

$$\text{var}_\theta(\hat{\theta}_{RB}) = \frac{n-1}{n}\frac{1}{F(\theta)},$$

where n is the number of independent data samples collected, $F(\theta)$ is the Fisher information for Ratbert's problem, and "RB" stands for Ratbert.

Catbert looks at Ratbert's work and snorts: "No, no, no. This can't be right. Your estimator violates the Cramer-Rao bound! Your calculations must be screwed up."

You check over Ratbert's calculations and find they are correct. What is wrong with Catbert's reasoning? What does Ratbert's result tell us about his estimator?

Solution: The CR bound which Catbert is thinking of only applies to unbiased estimators. This tells us that Ratbert's estimator is biased. Both Poor's and Hero's texts give examples of this; to reduce total MSE, it sometimes pays to accept some bias in return for a great reduction in variance.

3. Let Y be uniformly distributed on the interval $[0, \pi]$. (Yes, I meant π and not 2π). Suppose that $\Theta = \cos(Y)$ is estimated by $aY + b$.

You may find the following integrals useful in solving this problem:

$$\begin{aligned} \int x\cos(x)dx &= \cos(x) + x\sin(x) \\ \int \cos^2(x)dx &= \frac{1}{2}[\cos(x)\sin(x) + x] \end{aligned}$$

Also, recall that $Var[Y] = \pi^2/12$.

(a) (8 pts) What numerical values for a and b minimize the mean square error?

Solution: Use

$$E[\theta|Y] = E[\theta] + \frac{\text{cov}[\theta, Y]}{\text{var}[Y]}[Y - E[Y]]$$

$Var[Y]$ is given in the problem statement. $E[Y]$ is just the mean of a uniform random variable, so clearly $E[Y] = \pi/2$. Other things we need include:

$$E[\theta] = E[\cos Y] = \frac{1}{\pi}\int_0^\pi \cos y dy = 0$$

(You don't even really need to work out that integral explicitly; it's obvious when you think about what cosine looks like.)

$$E[\theta Y] = E[(\cos Y)Y] = \frac{1}{\pi}\int y\cos y dy = \frac{1}{\pi}[\cos(y) + y\sin(y)]\Big|_0^\pi = \frac{1}{\pi}(-1-1) = -\frac{2}{\pi}$$

Since $E[\theta] = 0$, $Var[Y] = E[\theta Y] = -2/\pi$.
So

$$E[\theta|Y] = \frac{-2/\pi}{\pi^2/12}[Y - \frac{\pi}{2}] = \frac{-24}{\pi^3}[Y - \frac{\pi}{2}]$$

We see $a = -24/\pi^3$ and $b = 12/\pi^2$.

(b) (8 pts) Compute $E[(\Theta - (aY + b))^2]$.

Solution: Use

$$Cov[error] = E[\theta^2] - \frac{\text{cov}^2[\theta, Y]}{\text{var}[Y]}$$

The only new thing we need is

$$E[\theta^2] = \frac{1}{\pi}\int_0^\pi \cos^2 y dy = \frac{1}{\pi}\frac{1}{2}[\cos y \sin y + y]\Big|_0^\pi = \frac{1}{2\pi}\pi = \frac{1}{2}$$

So

$$Cov[error] = \frac{1}{2} - \frac{(-2/\pi)^2}{\pi^2/12} = \frac{1}{2} - \frac{48}{\pi^4}$$

(c) (2 pts) In the last two parts, you computed the linear MMSE estimate and its mean square error. What is the actual (possibly nonlinear) MMSE estimator of Θ given Y?

Solution: Clearly, $\hat{\theta} = \cos(y)$.

(d) (2 pts) What is the mean square error of the MMSE estimator you gave in part (c)?

Solution: Clearly, MSE $= 0$.

4. Let the observation Y have the conditionally uniform density

$$p(y|\theta) = \begin{cases} 1/\theta & \text{for } 0 < y \leq \theta \\ 0 & \text{otherwise} \end{cases}$$

where Θ is a random variable with the density

$$p(\theta) = \begin{cases} \theta \exp(-\theta) & \text{for } \theta \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

You may find the following integral useful:

$$\int_v^\infty ue^{-u} du = (v+1)e^{-v} \quad \text{for } v \geq 0$$

(a) (7 pts) Find the MAP (maximum *a posteriori*) estimator of θ.
Solution:

$$p(y|\theta)p(\theta) = \begin{cases} \frac{1}{\theta}\theta e^{-\theta} = e^{-\theta} & \text{for } \theta \geq y \\ 0 & \text{otherwise} \end{cases}$$

Hence $\hat{\theta}_{MAP}(y) = y$.

(b) (7 pts) Find the MMSE (minimum mean squared error) estimator of θ.
Solution:

$$\hat{\theta}_{MMSE}(Y) = E[\Theta|Y=y] = \frac{\int \theta p(y,\theta)d\theta}{\int p(y,\theta)d\theta} = \frac{\int_y^\infty \theta e^{-\theta}d\theta}{\int_y^\infty e^{-\theta}d\theta} = \frac{(y+1)e^{-y}}{e^{-y}} = y+1$$

(c) (7 pts) Find the MAE (minimum absolute error) estimator of θ.
Solution:

$$\frac{\int_{-\infty}^{\hat{\theta}} p(y,\theta)d\theta}{\int p(y,\theta)d\theta} = \frac{1}{2}$$

$$\frac{\int_y^{\hat{\theta}} e^{\theta}d\theta}{\int_y^\infty e^{-\theta}d\theta} = \frac{e^{-y} - e^{-\hat{\theta}}}{e^{-y}} = \frac{1}{2}$$

$$\begin{aligned} e^{-y} - e^{-\hat{\theta}} &= \frac{e^{-y}}{2} \\ e^{\hat{\theta}} &= \frac{e^{-y}}{2} \\ -\hat{\theta} &= -y - \ln 2 \end{aligned}$$

So $\hat{\theta}_{MAE}(y) = y + \ln 2$.

(d) (2 pts) Suppose we want to estimate $\phi = 4\theta + 3$. Is the MAP estimate of ϕ given by

$$\hat{\phi}_{MAP}(y) = 4\theta_{MAP}(y) + 3?$$

A simple yes or no will suffice.
Solution: Yes, since the transformation is linear.

(e) (2 pts) Suppose we want to estimate $\beta = 4\theta^2 + 3$. Is the MAP estimate of β given by

$$\hat{\beta}_{MAP}(y) = 4(\theta_{MAP}(y))^2 + 3?$$

A simple yes or no will suffice.
Solution: No, since the transformation is nonlinear.

5. (15 points) In the examples we've done in class, we've been fortunate that the loglikelihood was bounded above, so that maximum-likelihood parameter estimates existed. There are some parametric models which are much more problematic. One of these is the three-parameter Weibull model. The probability density function of a three-parameter Weibull random variable is given by

$$p(y; \alpha, \gamma, \lambda) = \frac{\alpha}{\lambda} \left(\frac{y - \gamma}{\lambda} \right)^{\alpha - 1} \exp \left[- \left(\frac{y - \gamma}{\lambda} \right)^{\alpha} \right]$$

for $y \geq \gamma$.

Suppose we collect n i.i.d. samples $y_1, \cdots, y_n$. Show that the maximum likelihood estimate of the parameter triple $(\alpha, \gamma, \lambda)$ does not exist. (Hint: start by writing down the loglikelihood, and think about what would make it blow up to positive infinity.)

Solution: Consider the

$$(\alpha - 1) \sum_{i=1}^{n} \ln \left(\frac{y_i - \gamma}{\lambda} \right)$$

term in the loglikelihood. Let γ equal one of the data points. The logarithm blows "down" to negative infinity; choosing $\alpha < 1$ makes the loglikelihood blow "up" to positive infinity.