# Stat 543 Exam 1
# Spring 2016

**I have neither given nor received unauthorized assistance on this exam.**

KEY

**Name Signed** **Date**

**Name Printed**

This Exam consists of ~~9~~ 11 questions that will be scored at 10 points apiece (making ~~90~~ 110 points possible).

There is also on the last page of the Exam an "Extra Credit" question that will be scored out of 10 points. Any Extra Credit obtained will be recorded and used at the end of the course at Vardeman's discretion in deciding borderline grades. DO NOT spend time on this question until you are done with the entirety of the regular exam.

**1.** Consider two distributions for a bivariate random vector $\mathbf{X} = (X_1, X_2)$ taking values in the unit square $[0,1]^2$ with (joint) pdfs

$$f(\mathbf{x}\,|\,0) = I\left[\mathbf{x} \in [0,1]^2\right] \text{ and } f(\mathbf{x}\,|\,1) = (x_1 + x_2) I\left[\mathbf{x} \in [0,1]^2\right]$$

**a)** Identify a minimal sufficient statistic for the two-class model $\mathcal{P} = \{P_0, P_1\}$ corresponding to these two pmfs. The likelihood ratio is minimal sufficient in a 2-class model. This is

$$\ell(\underset{\sim}{x}) = \frac{f(\underset{\sim}{x}\,|\,1)}{f(\underset{\sim}{x}\,|\,0)} = x_1 + x_2$$

**b)** Set up completely (including limits of integration), but do not try to evaluate, a double integral giving the K-L information regarding $f(\mathbf{x}\,|\,1)$ relative to $f(\mathbf{x}\,|\,0)$.

We want $E_{f_1} \ln\left(\frac{f(X\,|\,1)}{f(X\,|\,0)}\right)$. Here this is

$$\int_0^1 \int_0^1 \Big(\ln(x_1 + x_2)\Big)(x_1 + x_2)\, dx_1\, dx_2$$

**c)** For a uniform prior distribution (on the two-point parameter space $\Theta = \{0,1\}$), i.e. one with $g(0) = g(1) = \frac{1}{2}$, what is the posterior probability of $\theta = 1$ given the observation $\mathbf{X} = (x_1, x_2)$? (Give a function $g(1|\mathbf{x}) : \Re^2 \to [0,1]$.)

$$g(1|\underset{\sim}{x}) = \frac{g(1)f(\underset{\sim}{x}|1)}{g(1)f(\underset{\sim}{x}|1) + g(0)f(\underset{\sim}{x}|0)}$$

$$= \frac{\frac{1}{2}(x_1+x_2)}{\frac{1}{2}(x_1+x_2) + \frac{1}{2}(1)} = \frac{x_1+x_2}{x_1+x_2+1}$$

(The likelihood ratio is always minimal sufficient in a 2-class problem)

**d)** For a decision problem with $\mathcal{A} = \Theta = \{0,1\}$ and 0-1 loss ($L(\theta, a) = I[a \neq \theta]$) and the uniform prior of part **c)** what is a Bayes optimal decision function, $d : [0,1]^2 \to \{0,1\}$? (Identify the set of $\mathbf{x} \in [0,1]^2$ for which one should take action $a = 1$.)

The minimum risk decision function is of the form

$$I\left[g(1|\underset{\sim}{x}) > \tfrac{1}{2}\right]$$

This is $I[x_1 + x_2 > 1] = d(\underset{\sim}{x})$

**e)** For your decision function, $d$, from part **d)**, evaluate the two values of the risk function $R(\theta,d)$, namely $R(0,d)$ and $R(1,d)$. (If you could not do part **d)** you may use the incorrect decision function $d(\mathbf{x}) = I[x_1 < x_2]$.)

$$R(0,d) = P_0[X_1+X_2>1] = \iint_{\nabla} 1 \, dx_1 dx_2 = \frac{1}{2}$$

$$R(1,d) = P_1[X_1+X_2<1] = \iint_{\Delta} (x_1+x_2) \, dx_1 dx_2$$

$$= \int_0^1 \int_0^{1-x_2} x_1+x_2 \, dx_1 dx_2$$

$$= \int_0^1 \frac{(1-x_2)^2}{2} + x_2(1-x_2) \, dx_2$$

$$= -\frac{(1-x_2)^3}{6}\Big|_0^1 + \left(\frac{x_2^2}{2} - \frac{x_2^3}{3}\right)\Big|_0^1 = \frac{1}{6}+\frac{1}{2}-\frac{1}{3} = \frac{1}{3}$$

2. Consider a class of discrete distributions on the small sample space $\mathcal{X} = \{0,1,2\}$ with pmfs

$$f(x \mid \eta) \propto \exp(\eta x^2)$$

for parameter space $\Gamma = \Re$. Suppose that $X_1, X_2, \ldots, X_n$ are iid according to pmf $f(x \mid \eta)$.

**a)** Identify a minimal sufficient statistic $T(\mathbf{X})$ in this context and argue very carefully that it is indeed minimal sufficient (applying whatever results from class or B&D are helpful).

This is an exponential family with natural parameter $\eta$ in a parameter space that has an interior. So the statistic

$$T(X) = \sum_{i=1}^{n} X_i^2$$

is minimal sufficient.

**b)** Consider method of moments estimation of $\eta$ based on $\widehat{\mu_{1}}_n = \bar{x}$. Find the estimating equation that a MOM estimator $\hat{\eta}_n^{\text{MOM}}$ must solve here.

Note that $\exp(\eta \cdot 0) + \exp(\eta(1)^2) + \exp(\eta(2)^2) = \frac{1}{C(\eta)}$

and then that $E_\eta X = \dfrac{1\exp(\eta) + 2\exp(4\eta)}{1 + \exp(\eta) + \exp(4\eta)}$

Then, the estimating equation is

$$\bar{X}_n = \frac{\exp(\eta) + 2\exp(4\eta)}{1 + \exp(\eta) + \exp(4\eta)}$$

**c)** Consider maximum likelihood estimation of $\eta$. Find an estimating equation that the Maximum Likelihood estimator $\hat{\eta}_n^{\text{MLE}}$ must solve here.

$$\frac{1}{n}\sum_{i=1}^{n} T(X_i) = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

$$E_\eta T(X) = E_\eta X^2 = \frac{\exp(\eta) + 4\exp(4\eta)}{1 + \exp(\eta) + \exp(4\eta)}$$

and the estimating equation is

$$\frac{1}{n}\sum X_i^2 = \frac{\exp(\eta) + 4\exp(4\eta)}{1 + \exp(\eta) + \exp(4\eta)}$$

(In an exponential family the MLE must satisfy

$$E_\eta T(X) = \frac{1}{n}\sum_{i=1}^{n} T(x_i) \quad )$$

**d)** What basis is there for Vardeman's expectation that the MLE from **c)** will have better theoretical properties than the MOM estimator from **b)**?

The MLE is based on a minimal sufficient statistic while the MOM estimator is not. One might expect the MLE to be better than the MOM estimator, as the latter is built on a statistic that is not sufficient.

**3.** Consider a situation in which $X \sim \text{Poisson}(\lambda)$ for $\lambda \in \Re$ but one observes only the (Bernoulli) variable $Y = I[X > 0]$.

**a)** Find the Fisher Information in $Y$ about $\lambda$ at the parameter value $\lambda_0$.

$Y \sim \text{Ber}(1-e^{-\lambda})$ So $f(y|\lambda) = (1-e^{-\lambda})^y (e^{-\lambda})^{(1-y)}$

$$\log f(y|\lambda) = y\log(1-e^{-\lambda}) + (1-y)\log(e^{-\lambda})$$

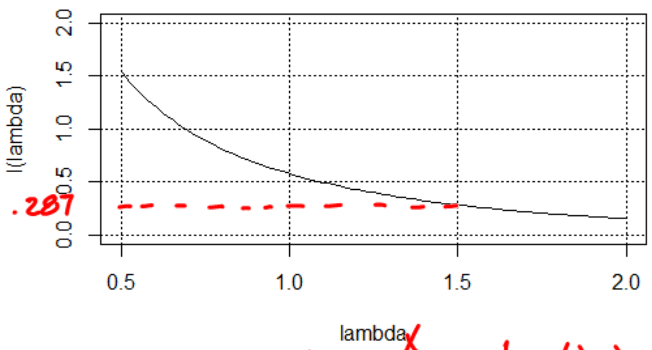
$$\frac{d}{d\lambda}\log f(y|\lambda) = y\frac{e^{-\lambda}}{1-e^{-\lambda}} + y - 1$$

and $$I_Y(\lambda_0) = \text{Var}_{\lambda_0}\left(Y\frac{1}{1-e^{-\lambda_0}}\right)^2 = \left(\text{Var}_{\lambda_0} Y^2\right)\left(\frac{1}{1-e^{-\lambda_0}}\right)^2$$

$$= \frac{e^{-\lambda_0}}{1-e^{-\lambda_0}}$$

$(\text{Var}\, Y^2 = \text{Var}\, Y = (1-e^{-\lambda})(e^{-\lambda}))$

**b)** Below is a plot of the correct answer to **a)**. Suppose that in fact $\lambda = 1.5$. How large must be $n$ in order for a sample $Y_1, Y_2, \ldots, Y_n$ (derived from independent Poisson$(\lambda)$ variables $X_1, \ldots, X_n$ as indicated above) to be as informative about $\lambda$ as a single $X \sim$ Poisson$(\lambda)$?




$$I_Y(1.5) = \frac{e^{-1.5}}{1-e^{-1.5}} = .287$$

$$f(x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$$

$$\log f(x|\lambda) = -\lambda + x\log\lambda - \log x!$$

$$\frac{d}{d\lambda} f(x|\lambda) = -1 + \frac{x}{\lambda}$$

$$I_X(\lambda_0) = \mathrm{Var}_{\lambda_0}\frac{X}{\lambda_0} = \frac{1}{\lambda_0^2}(\lambda_0) = \frac{1}{\lambda_0} \quad \text{So } I_X = \frac{1}{1.5} = .67.$$

Since FI adds for independent variables, $n$ iid $Y$'s have FI $n(.287)$. In order to have $n(.287) \geq .67$ we need $n \geq 3$.

4. Extra Credit – (Weighted Squared Error Loss) Suppose that in a Bayes statistical model some real-valued parametric function $\gamma(\boldsymbol{\theta})$ is of interest and will be estimated under weighted squared error loss

$$L(\boldsymbol{\theta}, a) = w(\boldsymbol{\theta})(\gamma(\boldsymbol{\theta}) - a)^2$$

for a known function $w(\boldsymbol{\theta}) \geq 0$. By considering the posterior distribution for observed data $X = x$, describe an optimal $d(x)$ in terms of expectations derived from this conditional distribution.

$$E[L(\theta,a) \mid X=x] = E[w(\theta)\gamma^2(\theta) - 2w(\theta)\gamma(\theta)a + w(\theta)a^2 \mid X=x]$$

$$= E[w(\theta)\gamma^2(\theta) \mid X=x] - 2a\,E[w(\theta)\gamma(\theta) \mid X=x] + a^2\,E[w(\theta) \mid X=x]$$

This quadratic in $a$ has a minimum where

$$\frac{d}{da}(\text{above}) = -2\,E[w(\theta)\gamma(\theta) \mid X=x] + 2a\,E[w(\theta) \mid X=x] = 0$$

i.e. where $a = \dfrac{E[w(\theta)\gamma(\theta) \mid X=x]}{E[w(\theta) \mid X=x]}$