Statistics GIDP
Ph.D. Qualifying Exam
Theory

May 27, 2014, 9:00am-1:00pm

Instructions: Provide answers on the supplied pads of paper; write on only one side of each sheet. Complete exactly 2 of the first 3 problems, and 2 of the last 3 problems. Turn in only those sheets you wish to have graded. Stay calm and do your best; good luck.

1. An urn initially contains b black balls and r red balls. A ball is drawn at random and then replaced along with another ball of the same color. This procedure is then repeated one more time. Let X_n equal 0 if the ball at the nth draw is black and equal 1 if it is red. Let
$$Y = X_1 + X_2.$$
Calculate the mean and variance of Y.

2. (In this question, you may assume without proof that all moments required exist, are finite, and are non-zero.) For a random variable Y with mean μ_Y and variance σ_Y^2,"heavy-tailedness may be measured by kurtosis
$$\tau_Y = \frac{\mathrm{E}[(Y-\mu_Y)^4]}{\sigma_Y^4}.$$
If $Y_1, Y_2, \cdots, Y_n$ are i.i.d. from a distribution with mean $\mu_Y = 0$, kurtosis τ_Y and variance σ_Y^2, show that $T = \sum_{i=1}^{n} Y_i$ has kurtosis
$$\tau_T = \tau_Y + 3(n-1)/n.$$

3. Let $(X_i, Y_i), i = 1, 2, \cdots$ be a sequence of i.i.d. random vectors, where $X_i \sim \exp(\theta)$ for each i, with the pdf
$$f_\theta(x) = \frac{1}{\theta}\exp\left(-\frac{x}{\theta}\right), \quad x > 0,$$
with $\theta > 0$ and mean $\mathrm{E}(X_i) = \theta$. And $Y_i \sim \exp(\frac{1}{\theta}), i = 1, \cdots, n$, are independent and identically distributed. Write
$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad \bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i.$$
   (a) Consider a linear estimator
   $$\hat{\theta}_a = a\bar{X}_n$$
   where a is a constant. Find the value of a that minimizes $\mathrm{E}(\hat{\theta}_a - \theta)^2$.

(b) Show that $\bar{X}_n$ is a consistent and asymptotically normal estimator of θ and find its asymptotic variance.

(c) Show that $1/\bar{Y}_n$ is a consistent and asymptotically normal estimator of θ and find its asymptotic variance.

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

$$f(x|\theta_1, \theta_2) = \begin{cases} (\theta_1 + \theta_2)^{-1} \exp(-x/\theta_1), & x > 0, \\ (\theta_1 + \theta_2)^{-1} \exp(x/\theta_2), & x < 0. \end{cases}$$

where $\theta_1 > 0$ and $\theta_2 > 0$ are unknown parameters.

(a) Find $\Pr(X_1 > 0)$.

(b) Is this an exponential family? Justify your answer.

(c) Obtain a sufficient statistic for θ_1 and θ_2.

(d) Suppose that you observed the following data set of size $n = 3$:

$$x_1 = 2.0, \quad x_2 = -0.6, \quad x_3 = -0.4.$$

Use these data to find the maximum likelihood estimates for θ_1 and θ_2.

5. Let $(Y_{i1}, Y_{i2})^T, i = 1, \cdots, n$ be n independent random vectors which are distributed as bivariate Normal $N(\begin{pmatrix} \mu_i \\ \mu_i \end{pmatrix}, \Sigma)$ with $\mu_i = \alpha + \beta x_i$ for $i = 1, \cdots, n$ and covariance matrix $\Sigma = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$. The parameters $\alpha, \beta, \sigma^2, \rho$ are unknown. Here $x_1, \cdots, x_n$ are known constants. Define

$$\bar{Y}_{..} = \frac{1}{n} \sum_{i=1}^{n} \bar{Y}_{i\cdot}, \quad \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

and

$$\bar{Y}_{i\cdot} = \frac{Y_{i1} + Y_{i2}}{2}, \quad i = 1, \cdots, n.$$

(a) Obtain a minimal sufficient statistic for $(\alpha, \beta, \sigma^2, \rho)$.

(b) It can be shown that the MLE of β is

$$\hat{\beta}_{MLE} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(\bar{Y}_{i\cdot} - \bar{Y}_{..})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Obtain $\mathrm{E}\left(\hat{\beta}_{MLE}\right)$. Establish that $\hat{\beta}_{MLE}$ is the UMVUE of β.

(c) Establish the distribution of

$$S_1 = \sum_{i=1}^{n} (Y_{i1} - Y_{i2})^2 = 2 \sum_{i=1}^{n} \sum_{j=1}^{2} (Y_{ij} - \bar{Y}_{i\cdot})^2.$$

(d) Use S_1 to obtain the UMVUE for the parameter $\theta = \sigma^2(1-\rho)$.

6. Assume that a chemical experiment can produce one of five possible consequences, with the following probability distribution

Consequence (k)	1	2	3	4	5
Probability	p^4	$4p^3q$	$6p^2q^2$	$4pq^3$	q^4

where $0 < p < 1$ is unknown and $q = 1 - p$. Now we repeat the chemical experiment n times. Let X_k denote the total counts of runs which have Consequence k in n runs, for $k = 1, ..., 5$. Assume the consequences of individual experiments are independent.

(a) Show that the distribution of the counts $(X_1, X_2, X_3, X_4, X_5)$ belongs to the exponential family.

(b) Show that $T = 4X_1 + 3X_2 + 2X_3 + X_4$ is a sufficient statistic for p.

(c) Is T a complete statistic? Justify your answer.

(d) Show that $T \sim \text{Bin}(4n, p)$.

(e) Derive the uniformly most powerful (UMP) level α test of the hypotheses

$$H_0 : p \leq \frac{1}{2} \quad \text{vs} \quad H_1 : p > \frac{1}{2}.$$

For each part, you are allowed to use results or conclusions drawn in the earlier parts.

Solutions:

1. X_1 is obviously a Bernoulli r.v. with $p = \frac{r}{b+r}$, X_2 is a Bernoulli r.v. with

$$p = \frac{r}{r+b} \cdot \frac{r+1}{r+b+1} + \frac{b}{r+b} \cdot \frac{r}{r+b+1} = \frac{r}{r+b}.$$

Therefore, $\mathrm{E}(Y) = \mathrm{E}(X_1) + \mathrm{E}(X_2) = 2\frac{r}{r+b}$.

$$\begin{aligned}\mathrm{Var}(Y) &= \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + 2(\mathrm{E}(X_1X_2) - \mathrm{E}(X_1)\mathrm{E}(X_2)) \\ &= 2\frac{r}{r+b} \cdot \frac{b}{r+b} + 2(\frac{r}{r+b} \cdot \frac{r+1}{r+b+1} - \frac{r}{b+r} \cdot \frac{r}{b+r}) \\ &= 2\frac{br}{(b+r)^2} \cdot \frac{b+r+2}{b+r+1}.\end{aligned}$$

2. (a) Without loss of generality, assume $\mu_Y = 0$. First, it is easy to find $\sigma_T^2 = n\sigma_Y^2$. When to find the relationship between $\mathrm{E}T^4$ and $\mathrm{E}Y^4$, we only need to consider items with all even powers, because with odd powers, at least one $\mathrm{E}Y_i$ must be involved. The total number of $\mathrm{E}Y_i^2Y_j^2$ is $\binom{n}{2}\binom{4}{2}$, therefore,

$$\tau_T = \frac{\mathrm{E}(\sum_{i=1}^n Y_i)^4}{\sigma_T^4} = \frac{n\mathrm{E}Y^4 + 6n(n-1)/2 \cdot (\mathrm{E}Y^2)^2}{n^2\sigma_Y^4} = \frac{1}{n}\tau_Y + 3(n-1)/n.$$

(b) Using the result from part (a), we first get the kurtosis of a Bernoulli r.v.

$$\tau_Y = \frac{\frac{1}{2}(1-p)^4 + \frac{1}{2}(0-p)^4}{[p(1-p)]^2} = \frac{(1-p)^4 + p^4}{2p^2(1-p)^2}.$$

Hence, the kurtosis for $Bin(n,p)$ is $\tau_T = \tau_Y/n + 3(n-1)/n$.

3. (a) Let $h(a)$ denote the MSE at a and decompose as,

$$\begin{aligned}h(a) &= \mathrm{E}(\hat{\theta}_a - \theta)^2 = [\mathrm{Bias}(\hat{\theta}_a)]^2 + \mathrm{Var}(\hat{\theta}_a) \\ &= (a-1)^2\theta^2 + a^2\frac{\theta^2}{n}\end{aligned}$$

Minimizing $h(a)$, we get $a = \frac{n}{n+1}$.

(b) Consistency can be showed by LLN. By CLT,

$$\sqrt{n}(\bar{X}_n - \theta) \xrightarrow{D} N(0, \theta^2).$$

(c) By CLT,

$$\sqrt{n}(\bar{Y}_n - 1/\theta) \xrightarrow{D} N(0, 1/\theta^2).$$

Apply the delta method and obtain

$$\sqrt{n}(1/\bar{Y}_n - \theta) \xrightarrow{D} N(0, \theta^2).$$

4. (a) $P(X_1 > 0) = \int_0^\infty (\theta_1+\theta_2)^{-1} \exp\{-x/\theta_1\} dx = \frac{\theta_1}{\theta_1+\theta_2}$.

(b) The density of X can be expressed as

$$f(x|\theta_1,\theta_2) = \frac{1}{\theta_1+\theta_2} \exp[\frac{-xI(x>0)}{\theta_1} + \frac{xI(x<0)}{\theta_2}].$$

It is a two-dimensional exponential family with $T_1(X) = -XI(X > 0)$ and $T_2(X) = XI(X < 0)$, and $w_1(\theta) = 1/\theta_1$ and $w_2(\theta) = 1/\theta_2$.

(c) The pair $(-\sum_{i=1}^n [x_i I(x_i > 0)], \sum_{i=1}^n [x_i I(x_i < 0)])$ is sufficient for (θ_1, θ_2).

(d) The log likelihood is

$$\log L(\theta_1,\theta_2|x_1,\cdots,x_n) = -n\log(\theta_1+\theta_2) + \frac{\sum_{i=1}^n [-x_i I(x_i>0)]}{\theta_1} + \frac{\sum_{i=1}^n [x_i I(x_i<0)]}{\theta_2}.$$

Define $x_+ = \sum_{i=1}^n x_i I(x_i > 0)$ and $x_- = \sum_{i=1}^n x_i(x_i < 0)$. Taking the partial first-order derivatives with respect to θ_1 and θ_2, we get the equation system

$$\begin{aligned} \frac{x_+}{\theta_1^2} &= \frac{n}{\theta_1+\theta_2} \\ \frac{x_-}{\theta_2^2} &= -\frac{n}{\theta_1+\theta_2}. \end{aligned}$$

Plug in $x_+ = \sum_{i=1}^n x_i I(x_i > 0) = 2$ and $x_- = \sum_{i=1}^n x_i I(x_i < 0) = -1$, and solve $\hat\theta_1 = \frac{2+\sqrt{2}}{3}$, $\hat\theta_2 = \frac{1+\sqrt{2}}{3}$. Check the second-order derivative too.

5. (a) Note that the inverse $\Sigma^{-1} = \begin{pmatrix} a & b \\ b & a \end{pmatrix}$ where $a = \frac{1}{1-\rho^2}$ and $b = -\frac{\rho}{1-\rho^2}$. There is a one-to-one correspondence between the pair (ρ, σ^2) and (a, b). The determinant $|\Sigma| = \sigma^2(1-\rho^2)$. The joint pdf of $(Y_{i1}, Y_{i2})^T, i = 1, \cdots, n$ is

$$\begin{aligned} & \prod_{i=1}^n f(y_{i1}, y_{i2}|\alpha,\beta,a,b) = \prod_{i=1}^n (2\pi)^{-1}(a^2-b^2)^{1/2} \cdot \\ & \exp\left\{-\frac{1}{2}\left[\begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} - \begin{pmatrix} \alpha+\beta x_i \\ \alpha+\beta x_i \end{pmatrix}\right]^T \Sigma^{-1} \left[\begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} - \begin{pmatrix} \alpha+\beta x_i \\ \alpha+\beta x_i \end{pmatrix}\right]\right\} \\ = & \ (2\pi)^{-n}(a^2-b^2)^{n/2} \exp\{-\frac{1}{2}T\} \end{aligned}$$

where

$$\begin{aligned} T &= \sum_{i=1}^n \begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix}^T \begin{pmatrix} a & b \\ b & a \end{pmatrix} \begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} - 2\begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix}^T \begin{pmatrix} a & b \\ b & a \end{pmatrix} \begin{pmatrix} \alpha+\beta x_i \\ \alpha+\beta x_i \end{pmatrix} + \begin{pmatrix} \alpha+\beta x_i \\ \alpha+\beta x_i \end{pmatrix}^T \begin{pmatrix} a & b \\ b & a \end{pmatrix} \begin{pmatrix} \alpha+\beta x_i \\ \alpha+\beta x_i \end{pmatrix} \\ &= \sum_{i=1}^n \left[aY_{i1}^2 + 2bY_{i1}Y_{i2} + aY_{i2}^2 - 2\alpha(a+b)(Y_{i1}+Y_{i2}) - 2\beta(a+b)(Y_{i1}+Y_{i2})x_i\right] + \text{constant} \\ &= a\sum_{i=1}^n \sum_{j=1}^2 Y_{ij}^2 + 2b(\sum_{i=1}^n Y_{i1}Y_{i2}) - 4n\alpha(a+b)\bar{Y}_{\cdot\cdot} - 4\beta(a+b)(\sum_{i=1}^n x_i \bar{Y}_{i\cdot}) + \text{constant}, \end{aligned}$$

Therefore, $\left(\sum_{i,j} Y_{ij}^2, \sum_i Y_{i1}Y_{i2}, \bar{Y}_{\cdot\cdot}, \sum_i x_i \bar{Y}_{i\cdot}\right)$ is complete and sufficient for (a, b, α, β), also for $(\sigma^2, \rho, \alpha, \beta)$.

(b) $\hat{\beta}_{MLE} = \frac{\sum_{i=1}^n x_i \bar{Y}_{i\cdot} - n\bar{x}\bar{Y}_{\cdot\cdot}}{\sum_{i=1}^n (x_i - \bar{x})^2}$. Note $E(\bar{Y}_{i\cdot}) = \alpha + \beta x_i$, $E(\bar{Y}_{\cdot\cdot}) = \alpha + \beta \bar{x}$. As a function of complete and sufficient statistics, $\hat{\beta}_{MLE}$ is the UMVUE of β.

(c) Note $Y_{i1} - Y_{i2} \sim N(0, 2\sigma^2(1-\rho))$, and $Y_{i1} - Y_{i2}$'s are all indepdendent. Therefore, $\frac{S_1}{2\theta} = \frac{\sum_{i=1}^n (Y_{i1} - Y_{i2})^2}{2\sigma^2(1-\rho)} \sim \chi_n^2$. And $E(S_1) = 2n\theta$. Note

$$S_1 = \sum_{i=1}^n (Y_{i1} - Y_{i2})^2 = \sum_{i=1}^n (Y_{i1}^2 + Y_{i2}^2 - 2Y_{i1}Y_{i2}) = \sum_{i,j} Y_{ij}^2 - 2\sum_{i=1}^n Y_{i1}Y_{i2}$$

is a function of complete and sufficient statistics. The UMVUE of θ is $\frac{S_1}{2n}$.

6. (a) $(X_1, X_2, X_3, X_4, X_5)$ follows a multinomial distribution with the joint pdf

$$\begin{aligned} f_p(x) = P(X_1 = x_1, ..., X_5 = x_5) &= \frac{n!}{x_1! \cdots x_5!} p_1^{x_1} \cdots p_5^{x_5} \\ &= \frac{n!}{x_1! \cdots x_5!} (p^4)^{x_1} (4p^3 q)^{x_2} (6p^2 q^2)^{x_3} (4pq^3)^{x_4} (q^4)^{x_5} \\ &\propto p^{4x_1 + 3x_2 + 2x_3 + x_4} q^{x_2 + 2x_3 + 3x_4 + 4x_5} \\ &\propto p^{T(x)} q^{4n - T(x)} \\ &\propto q^{4n} \exp[T(x) \log(\frac{p}{q})] \end{aligned}$$

where $\eta = \log(\frac{p}{q})$ is the natural parameter. This belong an exponential family.

(b) $T(X) = 4X_1 + 3X_2 + 2X_3 + X_4$ is sufficient for p.

(c) The parameter space $\{\log(\frac{p}{1-p}) : 0 < p < 1\}$ is an open set in R, so T is complete.

(d) Note T essentially counts the total number of heads in tossing $4n$ iid coins. Alternatively, use the MGF argument.

(e) Assume $0 < p_0 < p_1 < 1$, then the likelihood ratio

$$l(x) = f_{p_1}(x)/f_{p_0}(x) = [\frac{q_1}{q_0}]^{4n} \exp[T(x) \log(\frac{p_1}{p_0}\frac{q_0}{q_1})]$$

is monotonically increasing in T. Using Karlin-Rubin, the UMP test is to reject H_0 iff $T > c$, with c satisfying $P_{p=\frac{1}{2}}(T > c) = \alpha$.