BASKIN SCHOOL OF ENGINEERING

Department of Applied Mathematics and Statistics

2015 First Year Exam: June 5, 2015

INSTRUCTIONS

If you are on the Applied Mathematics track, you are required to complete problems 1(AMS 203), 2(AMS 211), 3(AMS 212A), 4(AMS 212B), 5(AMS 213), and 6(AMS 214).

If you are on the Statistics track, you are required to complete problems 1(AMS 203), 2(AMS 211), 7(AMS 205B), 8(AMS 206B), 9(AMS 207), and 10(AMS256).

Please complete all required problems on the **supplied exam papers**. Write your exam ID number and problem number on each page. Use only the **front side** of each page.

Problem 1 (AMS 203):

Nic is vacationing in Monte Carlo. The amount X (in dollars) he takes to the casino each evening is a random variable with probability density function

$$f_X(x) = \begin{cases} ax, & 0 \le x \le 40, \\ 0, & \text{otherwise.} \end{cases}$$

At the end of each night, the amount Y that he has on leaving the casino is uniformly distributed between zero and twice the amount he took in.

(a) (10%) Find the constant a.

(b) (25%) Determine the joint probability density function $f_{X,Y}(x, y)$. Be sure to indicate what the sample space is.

(c) (25%) What is the probability that on any given night Nic makes a positive profit at the casino? Justify your reasoning.

(d) (40%) Find the probability density function of Nic's profit on any particular night, $Z = Y - X$. What is $E(Z)$?

Problem 2 (AMS 211):

Problem 3 (AMS 212A):

Problem 4 (AMS 212B):

Problem 5 (AMS 213):

Problem 6 (AMS 214):

Problem 7 (AMS 205B):

Suppose $X_1, ..., X_n \stackrel{iid}{\sim} N(\mu_1, \sigma_1^2)$ and $Y_1, ..., Y_m \stackrel{iid}{\sim} N(\mu_2, \sigma_2^2)$, with $\mu_1, \mu_2 \in \mathbb{R}$ and $\sigma_1, \sigma_2 > 0$. Moreover, assume that $\{X_1, ..., X_n\}$ and $\{Y_1, ..., Y_m\}$ are mutually independent.

(a) (30%) Derive a set of minimal sufficient statistics for $(\mu_1, \mu_2, \sigma_1, \sigma_2)$. Is this collection of statistics complete under the additional assumption that $\mu_1 = \mu_2$? Justify your answer.

(b) (40%) Find the UMVUE for $\theta = (\mu_1\mu_2)/(\sigma_1^r\sigma_2^s)$, where $1 < r < n-1$ and $1 < s < m-1$. Show your work.
You may use without proof the following fact: if $Z \sim IG(\alpha, \beta)$ for $\alpha > \eta \geq 1$ and $\beta > 0$, in the parameterization in which $E(Z) = \beta/(\alpha - 1)$, then

$$E\left(Z^\eta\right) = \frac{\beta^\eta\,\Gamma(\alpha-\eta)}{\Gamma(\alpha)}\,.$$

(c) (30%) Now assume that σ_1^2 and σ_2^2 are known, $\sigma_1^2 = \sigma_2^2 \equiv \sigma^2$, and $\mu_1 = \mu_2 \equiv \mu$. Explain why a UMP level α test for testing H_0: $\mu \leq \mu_0$ versus H_1: $\mu > \mu_0$ exists, and specify this test explicitly.

Problem 8 (AMS 206B):

Let X_i denote the number of fires in a town for week i, $i = 1, \ldots, n$. Suppose the observations, $\boldsymbol{x} = (x_1, \ldots, x_n)$, form an i.i.d. sample from a Poisson distribution with mean θ.

(1) (10%) Find the Jeffreys prior for θ, $\pi_J(\theta)$.

(2) (20%) Use the prior in part (1) to find the resulting posterior distribution $\pi_J(\theta \mid \boldsymbol{x})$.

(3) (30%) Under the posterior distribution in part (2), find the Bayes estimate of θ with the following loss function: $L(\theta, a) = \theta(\theta - a)^2$.

(4) (40%) Now assume instead that the prior on θ is given by an exponential distribution, with density $\pi(\theta \mid \lambda) = \lambda \exp(-\lambda\,\theta)$ for known $\lambda > 0$. Suppose that, under the Poisson sampling model above and with the same sample $\boldsymbol{x}$ of data, it is desired to test H_0: $\theta \leq c$ versus H_1: $\theta > c$, for some constant $c > 0$. Let $F(t \mid \alpha, \beta)$ be the CDF of the gamma(a, b) distribution, in the parameterization in which the mean is α/β.

   (a) (35%) Using $F(t \mid \alpha, \beta)$ and having observed $\boldsymbol{x}$, give an explicit expression for the resulting Bayes factor in favor of H_0.

   (b) (5%) If this Bayes factor worked out numerically to be 6.25, what would your interpretation be about the weight of data evidence for or against H_0?

Problem 9 (AMS 207):

Count data with an overabundance of zeros can be modeled through the zero-inflated Poisson model (ZIP). The p.m.f. of the zero-inflated Poisson distribution, ZIP(λ, π), is given by

$$f(y|\lambda, \pi) = \pi I(y = 0) + (1 - \pi) f_0(y|\lambda), \quad y = 0, 1, 2, \ldots,$$

where $0 \leq \pi \leq 1$, $I(\cdot)$ is the indicator function, and $f_0(y|\lambda) = \lambda^y e^{-\lambda}/y!$, with $\lambda > 0$.

1. (10%) Let $Y \sim$ ZIP(λ, π). Find $\Pr(Y = 0|\lambda, \pi)$, and $\Pr(Y = y|\lambda, \pi)$ for $y \neq 0$.

2. (90%) Given λ and π, let $Y_1, \ldots, Y_n$ be i.i.d. ZIP(λ, π), with priors $\lambda \sim$ gamma(a, b) and $\pi \sim$ Beta(c, d). Assume this model is used to describe the number of cigarettes smoked by the respondents of a health survey. More specifically, suppose that out of 10 respondents, the number of cigarettes smoked by each respondent are:

$$y_1 = 0,\ y_2 = 5,\ y_3 = 0,\ y_4 = 0,\ y_5 = 10,\ y_6 = 0,\ y_7 = 0,\ y_8 = 0,\ y_9 = 3,\ y_{10} = 0.$$

   (a) (15%) Write down $p(\lambda, \pi|y_1, \ldots, y_{10})$ up to a proportionality constant.
   (b) (15%) Find $p(\lambda|\pi, y_1, \ldots, y_{10})$ and $p(\pi|\lambda, y_1, \ldots, y_{10})$ (up to a proportionality constant in each case). Are these distributions available in closed form?
   (c) (60%) Introduce auxiliary variables Z_i such that: $Y_i = 0$ if $Z_i = 1$, and $Y_i \sim$ Poisson(λ) if $Z_i = 0$, with $Z_i \sim$ Bernoulli(π). Derive an MCMC algorithm to sample from $p(\lambda, \pi, Z_1, \ldots, Z_{10}|y_1, \ldots, y_{10})$. Provide the details of the steps in your algorithm, i.e., for Gibbs steps, provide the full conditional distributions, for Metropolis-Hastings steps, provide the proposal distributions and the acceptance probabilities.

Some useful formulas:

- The p.m.f. of a random variable X with a Poisson distribution with mean λ is
$$f(x) = \frac{1}{x!}\lambda^x e^{-\lambda}, \quad x = 0, 1, 2, \ldots, \quad \lambda > 0.$$
- The p.d.f. of a random variable X with a gamma(a, b) distribution is
$$f(x) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}, \quad x > 0, \quad a > 0,\ b > 0.$$
- The p.d.f. of a random variable X with a Beta(c, d) distribution is
$$f(x) = \frac{\Gamma(c+d)}{\Gamma(c)\Gamma(d)} x^{c-1}(1-x)^{d-1}, \quad x \in [0, 1], \quad c > 0,\ d > 0.$$

Problem 10 (AMS 256):

Consider a two-way ANOVA model

$$y_{i,j} = \alpha_i + \tau_j + \epsilon_{i,j}, \quad i = 1,2;\;\; j = 1,2 \tag{1}$$

where the parameters α_i and τ_j are unknown. For the errors, assume that $\epsilon_{i,j} \overset{iid}{\sim} \mathrm{N}(0, \sigma^2)$.

1. (10%) Express these observations into a general linear model form, $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{\beta} = [\alpha_1, \alpha_2, \tau_1, \tau_2]^T$. That is, specify $\boldsymbol{y}$, $\mathbf{X}$ and $\boldsymbol{\epsilon}$. Identify the rank of $\mathbf{X}$.

2. (15%) Show that for this design, if a function of $\boldsymbol{\beta}$, $c_1\alpha_1 + c_2\alpha_2 + d_1\tau_1 + d_2\tau_2$ is estimable, then $c_1 + c_2 = d_1 + d_2$.

3. (40%) Consider the function of $\boldsymbol{\beta}$ defined by $\alpha_1 - \alpha_2$. A g-inverse of the matrix $\mathbf{X}^T\mathbf{X}$ for the model in (1) is given by

$$(\mathbf{X}^T\mathbf{X})^- = \frac{1}{4}\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 4 & -2 & -2 \\ 0 & -2 & 3 & 1 \\ 0 & -2 & 1 & 3 \end{bmatrix}$$

   Find the best linear unbiased estimator of the function. Also, find its probability distribution (assume that σ^2 is known for this part).

4. (15%) Is y_{11} the best linear unbiased estimator of $\alpha_1 + \tau_1$? Explain why or why not.

5. (20%) Assume that σ^2 is unknown. Derive a test for $H_0 : \alpha_1 = \alpha_2$ at significance level γ. That is, describe a test statistic, its distribution under the null and the alternative, and a rejection region. Expressing SSE in the form of vectors is enough for an answer.