ANSWER BOOKLET

Feb 22, 2018 **Name:** ....

STAT 632 (Anirban Bhattacharya) **UIN:** ....

1	2	3	4	$\sum$

The exam is out of 80 points. For each problem a maximum of 20 points can be earned. You have 75 minutes to complete the exam.

- **Work your problems in the space provided.**
- **Show all work clearly and concisely.**
- **Justify your answers. State results you use.**
- **Box your $\boxed{\textbf{final answer}}$ whenever applicable.**
- $\mathbb{1}_A(x)$ denotes the indicator function of the set A, that is, $\mathbb{1}_A(x) = 1$ if $x \in A$ and is zero otherwise.

1. Suppose $x_1, \dots, x_n \mid \theta \sim U(0, \theta)$, where $U(a, b)$ denotes the uniform distribution on the interval $[a, b]$ with density $1/(b-a)$ on $[a, b]$. Our goal is to infer $\theta > 0$ from the data. Consider a Pareto prior on θ, $\theta \sim \mathcal{PA}(\alpha, \beta)$, which has density

$$\pi(\theta) = \frac{\alpha\beta^\alpha}{\theta^{\alpha+1}} \, \mathbb{1}_{(\beta,\infty)}(\theta),$$

where $\alpha, \beta > 0$ are respectively called the shape and scale parameters of the Pareto distribution.

(a) Write an expression the posterior distribution of θ. Can you identify what distribution it is? Find the posterior mean.

(b) Show that if $U \sim U(0, 1)$, then $\beta \, U^{-1/\alpha} \sim \mathcal{PA}(\alpha, \beta)$. Using this result, or otherwise, describe a procedure to obtain a 95% credible interval for $\log(\theta)$.

2. A random sample of 50 College Station residents are enquired about the number of traffic citations they received over the past year, and it was found that 35 individuals didn't receive any citations, 11 received 1 citation, 2 received 2 citations, and 2 individuals received 3 or more citations. Suppose we model the number of citations t received by an individual over the past year by

$$P(t = 0 \mid p) = p,\ P(t = 1 \mid p) = p(1 - p),\ P(t = 2 \mid p) = p(1 - p)^2,$$

where $p \in (0, 1)$. Also, assume that the number of citations received by individuals are independent of each other.

Assuming a Beta$(2, 1)$ prior on p, calculate its posterior distribution and posterior mean. Recall that a Beta(a, b) distribution has density

$$\frac{\Gamma(a + b)}{\Gamma(a)\,\Gamma(b)}\, x^{a-1}(1 - x)^{b-1},\ x \in (0, 1),$$

with mean $a/(a + b)$.

3. In psychometry studies, it is common to administer multiple tests on an individual to learn about certain unobserved cognitive traits. Assume a simplified setting where p tests are administered on an individual to learn about a single cognitive trait, which we shall call η. Let $y_1, \ldots, y_p$ denote the standardized test scores on the p different tests, which are related to the unobserved cognitive trait η through the model

$$y_i = \lambda_i \eta + \epsilon_i, \quad i = 1, \ldots, p.$$

Here $\lambda_i \in \mathbb{R}$ quantifies the difficulty level associated with the ith test, and $\epsilon_i \sim N(0,1)$ independently for $i = 1, \ldots, p$. For this problem, you may assume that the λ_is are known.

Assume a Gaussian prior $\eta \sim N(0, \kappa^{-1})$, for some fixed hyperparameter κ. Recall the $N(\mu, \sigma^2)$ density is $(2\pi\sigma^2)^{-1/2} e^{-(x-\mu)^2/(2\sigma^2)}$.

(a) Calculate the posterior distribution of $\eta \mid y_1, \ldots, y_p$.

(b) Calculate the marginal distribution of $(y_1, \ldots, y_p)$.

4. Let f and g be two distinct densities on $\mathbb{R}$. Let

$$h(x,y) = \frac{f(x)f(y) + g(x)g(y)}{2}, \quad x, y \in \mathbb{R}.$$

Show that h is a density on $\mathbb{R}^2$ and describe a procedure to sample from h. You may assume that we can sample from the densities f and g.