STA 114: Midterm II

Total time: 1hr 10min

The **six** questions below carry a total of 43 points. Your exam will be graded out of 40 points – your score is either the points you secure or 40, whichever is less. Answer each question to the best of your ability and show work to guarantee partial/full marks. Make use of the tables and basic probability facts attached at the end. You'd be provided with white papers to write your answers. Please write your name on each sheet of paper and remember to staple them before turning in.

1. Complete the statements below each with a correct option and justify. [$3\times 3 = 9$ points]

   (a) For a model $X \sim f(x|\theta)$, $\theta \in [0, 4]$ is assigned a $\xi(\theta) = \mathsf{Uniform}(0, 4)$ prior pdf. An observation x gives $L_x(1)/L_x(2) = 3$. Then $\xi(1|x)/\xi(2|x)$ must equal
   A. 3   B. 1/3   C. 1   D. Cannot be determined.
   3. Because $\xi(\theta|x) = \text{const} \times L_x(\theta)\xi(\theta)$ and so $\xi(1|x)/\xi(2/|x) = L_x(1)/L_x(2) \times \xi(1)/\xi(2) = 3$ as $\xi(1)/\xi(2) = 1$.

   (b) Area measurements X_1, X_2 (in mm^2) of two microchips produced by a machine are modeled as independent observations from the pdf $g(x_i|\theta) = 1 - |x_i - \theta|$, if $x_i \in [\theta - 1, \theta + 1]$ and $g(x_i|\theta) = 0$ otherwise. The unknown "target area" $\theta \in [1, 5]$ is assigned a $\mathsf{Uniform}(1, 5)$ prior. Based on observation $X_1 = 3$, the posterior predictive probability of $X_2 < 1$ is
   A. 1   B. $\frac{1}{2}$ C. $\frac{1}{4}$ D. 0
   0. Because
   $$P(X_2 < 1|X_1 = 3) = \int P(X_2 < 1|\theta)\xi(\theta|3)d\theta = \int_2^4 P(X_2 < |\theta)\xi(\theta|3)d\theta,$$
   since given $X_1 = 3$, the posterior pdf $\xi(\theta|3) = \text{const} \times g(3|\theta)\xi(\theta)$ is zero for θ outside $[2, 4]$. But $P(X_2 < 1|\theta) = 0$ for $\theta \in [2, 4]$.

   (c) Annual hurricane counts $X_1, \cdots, X_n$ from n successive years are modeled as $X_i \overset{\text{IID}}{\sim} \mathsf{Poisson}(\mu)$, $\mu \in (0, \infty)$. A researcher used some prior distribution on μ and based on the observed counts $x_1, \cdots, x_{n-1}$ from the first $n - 1$ years (that's all she had access to), reported the posterior pdf of μ as $\mathsf{Gamma}(150, 10)$. If she could also use the n-th year's count $X_n = 15$, she would have reported the posterior pdf
   A. $\mathsf{Gamma}(15, 1)$ B. $\mathsf{Gamma}(150, 10)$ C. $\mathsf{Gamma}(165, 11)$ D. Cannot be determined.
   $\mathsf{Gamma}(165, 15)$. Because, to include the additional observation we could take the posterior $\mathsf{Gamma}(150, 10)$ after first $n - 1$ observations as the new prior, and $X_n \sim \mathsf{Poisson}(\mu)$ as the model. By conjugacy, the resulting posterior is $\mathsf{Gamma}(150 + 15, 10 + 1) = \mathsf{Gamma}(165, 11)$.

2. Let $X_1, \cdots, X_n$ denote the first-serve success rates of a tennis player from n matches. Consider the model $X_i \overset{\text{IID}}{\sim} g(x_i|\theta)$, $\theta \in (0, \infty)$, where the pdf $g(x_i|\theta) = \theta x_i^{\theta-1}$ for $0 < x_i < 1$ and is zero elsewhere. [$5 + 3 + 2 = 10$ points]

   (a) Identify the Jeffreys' prior pdf for θ and show that the corresponding posterior pdf of θ given observations $x_1, \cdots, x_n$, is $\mathsf{Gamma}(n, -\sum_{i=1}^n \log x_i)$. [Hint: $x_i^{\theta-1} = x_i^{-1}e^{\theta \log x_i}$]

The log-likelihood function is $\ell_x(\theta) = \text{const} + n\log\theta + \theta\sum_{i=1}^n \log x_i$. Hence $\ddot{\ell}_x(\theta) = -1/\theta^2$ and so $I^F(\theta) = 1/\theta^2$. Therefore, $\xi^J(\theta) = \sqrt{I^F(\theta)} = 1/\theta$. With this prior the posterior equals: $\xi^J(\theta|x) = \text{const.} \times \theta^n e^{\theta\sum_{i=1}^n \log x_i} \times (1/\theta) = \text{const} \times \theta^{n-1}e^{-(-\sum_{i=1}^n \log x_i)\theta}$ which matches the pdf of $\mathsf{Gamma}(n, -\sum_{i=1}^n \log x_i)$.

(b) For any $a \geq 30$ and any $b > 0$, the quantiles of a $\mathsf{Gamma}(a, b)$ distribution are well approximated by those of a $\mathsf{Normal}(c, d^2)$ distribution with $c = a/b$ and $d^2 = a/b^2$. Use this approximation to give a central 95% posterior credible interval for θ under the Jeffreys prior, based on observations with $n = 40$ and $\sum_{i=1}^n \log x_i = -8.2$.

The central 95% posterior credible interval for θ under the Jeffreys' prior is $[\theta_{.025}(x), \theta_{.975}(x)]$ where $\theta_u(x)$ denotes the u-th quantile of $\mathsf{Gamma}(40, 8.2)$ distribution. By the stated result, this interval can be approximated by $c \mp 1.96d$, where $c = 40/8.2 = 4.88$ and $d = \sqrt{40/8.2^2} = 0.77$. So the interval equals $[3.37, 6.39]$.

(c) Under the model, the average success rate of the player is given by the quantity $\eta = \mathrm{E}_{[X_i|\theta]}X_i = \frac{\theta}{\theta+1}$. We could rewrite the model as $X_i \overset{\text{IID}}{\sim} h(x_i|\eta)$, $\eta \in (0,1)$ where $h(x_i|\eta) = \frac{\eta}{1-\eta}x_i^{\eta/(1-\eta)-1}$ for $0 < x_i < 1$ and analyze it with a Jeffreys prior for η. What would be the central 95% posterior credible interval for η under this analysis based on the observation in part (c)? Explain.

Because η is a monotone transform of θ and we are using Jeffreys' priors for both formulations, the central 95% posterior credible interval of η must be that of θ transformed in a similar manner, i.e., $[\frac{3.37}{3.37+1}, \frac{6.39}{6.39+1}] = [0.77, 0.86]$.

3. A count data X is modeled as $X \sim f(x|\theta)$, $\theta \in \{\frac{1}{4}, \frac{1}{2}, \frac{3}{4}\}$, where the pmfs $f(x|\theta)$, defined over $x \in \{0, 1, 2, 3\}$ (with $f(x|\theta) = 0$ for any other x) are as in the table below. Assume that θ is assigned a discrete uniform prior pmf: $\xi(\frac{1}{4}) = \xi(\frac{1}{2}) = \xi(\frac{3}{4}) = 1/3$.

θ	$f(0\|\theta)$	$f(1\|\theta)$	$f(2\|\theta)$	$f(3\|\theta)$
$\frac{1}{4}$	$\frac{27}{64}$	$\frac{27}{64}$	$\frac{9}{64}$	$\frac{1}{64}$
$\frac{1}{2}$	$\frac{8}{64}$	$\frac{24}{64}$	$\frac{24}{64}$	$\frac{8}{64}$
$\frac{3}{4}$	$\frac{1}{64}$	$\frac{9}{64}$	$\frac{27}{64}$	$\frac{27}{64}$

A future variable X^* depends on the same θ through pmfs $f^*(x^*|\theta)$ given by $f^*(0|\theta) = 1-\theta$, $f^*(1|\theta) = \theta$ and $f^*(x^*|\theta) = 0$ for all other x^*. Assume X and X^* are conditionally independent given θ. Calculate $P(X^* = 1|X = x)$ for each $x = 0, 1, 2, 3$. [9 points]

The posterior pmfs $\xi(\theta|x)$ can be found by normalizing each column of the above table (because of uniform prior):

θ	$\xi(\theta\|0)$	$\xi(\theta\|1)$	$\xi(\theta\|2)$	$\xi(\theta\|3)$
$\frac{1}{4}$	$\frac{27}{36}$	$\frac{27}{60}$	$\frac{9}{60}$	$\frac{1}{36}$
$\frac{1}{2}$	$\frac{8}{36}$	$\frac{24}{60}$	$\frac{24}{60}$	$\frac{8}{36}$
$\frac{3}{4}$	$\frac{1}{36}$	$\frac{9}{60}$	$\frac{27}{60}$	$\frac{27}{36}$

and so,

$$P(X^* = 1|X = 0) = \frac{1}{4} \times \frac{27}{36} + \frac{1}{2} \times \frac{8}{36} + \frac{3}{4} \times \frac{1}{36} = 0.19 + 0.11 + 0.02 = 0.32$$
$$P(X^* = 1|X = 1) = \frac{1}{4} \times \frac{27}{60} + \frac{1}{2} \times \frac{24}{60} + \frac{3}{4} \times \frac{9}{60} = 0.11 + 0.2 + 0.11 = 0.42$$
$$P(X^* = 1|X = 2) = \frac{1}{4} \times \frac{9}{60} + \frac{1}{2} \times \frac{24}{60} + \frac{3}{4} \times \frac{27}{60} = 0.04 + 0.2 + 0.34 = 0.58$$
$$P(X^* = 1|X = 3) = \frac{1}{4} \times \frac{1}{36} + \frac{1}{2} \times \frac{8}{36} + \frac{3}{4} \times \frac{27}{36} = 0.01 + 0.11 + 0.56 = 0.68$$

4. Smile durations (in seconds) of a baby are modeled as independent observations from a $\mathsf{Uniform}(0, \theta)$ with $\theta \in (0, \infty)$ assigned a $\mathsf{Pareto}(2, 15)$ prior. [3 + 2 = 5 points]

   (a) What's the posterior predictive probability that the next smile would last longer than 20 seconds given past data: 10, 20, 13, 15, 1, 1, 6, 7, 9, 9?
   The posterior pdf of θ is $\mathsf{Pareto}(2 + n, \max(15, x_{\max}) = \mathsf{Pareto}(12, 20)$. Given θ, the probability of the next smile lasting loner than 20 seconds is 0 if $\theta \le 20$ and $1-20/\theta$ if $\theta > 20$. So, the posterior predictive probability of the next smile lasting longer than 20 seconds is:

   $$1 - \int_{20}^{\infty} \frac{20}{\theta} \frac{12 \times 20^{12}}{\theta^{13}} d\theta = 1 - 12 \cdot 20^{13} \int_{20}^{\infty} \theta^{-14} d\theta = 1 - 12 \cdot 20^{13} \left(-\frac{\theta^{-13}}{13} \right) \Bigg|_{20}^{\infty} = \frac{1}{13}$$

   (b) What's the answer to the above question based on an MLE plug-in approach?
   The MLE of θ is 20, therefore the plug-in answer is 0.

5. Data $X = (X_1, \cdots, X_n)$, $Y = (Y_1, \cdots, Y_m)$ are modeled as $X_i \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu_1, \sigma^2)$, $Y_j \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu_2, \sigma^2)$, X_i's and Y_j's are independent with $\xi(\mu_1, \mu_2, \sigma^2) = \text{const}/\sigma^2$, for which the central $100(1-\alpha)\%$ posterior credible intervals of $\eta = \mu_1 - \mu_2$ are given by

   $$(\bar{x} - \bar{y}) \mp z_{n+m-2}(\alpha) \sqrt{\left(\frac{1}{n} + \frac{1}{m} \right) \frac{(n-1)s_x^2 + (m-1)s_y^2}{n + m - 2}}.$$

   What is the posterior probability that μ_1 exceeds μ_2 based on observations x, y with $n = 12, m = 7$, $\bar{x} = 119$, $\bar{y} = 111.5$, $s_x = 21.4$, $s_y = 20.6$? [4 points]

   For observed data, the above intervals are $7.5 \mp z_{17}(\alpha) \sqrt{(.08 + .14) \frac{5037+2546}{17}} = 7.5 \mp z_{17}(\alpha)\sqrt{98.1} = 7.5 \mp 9.9 z_{17}(\alpha)$. Now, the posterior probability of $\mu_1 > \mu_2$ is $1 - \alpha/2$ where α is the largest positive fraction for which the left end of the above interval just touches the point 0, i.e., α satisfies $z_{17}(\alpha) = 7.5/9.9 = 0.76$, i.e., $1 - \alpha/2 = \Phi_{17}(0.76)$.

   You could also arrive at this answer in a more direct way. We have $\xi(\mu_1, \mu_2, \sigma^2 | x, y) = \mathsf{NN}\chi^{-2}(119, 12, 111.5, 7, 17, 446)$ and so the posterior pdf of $(\mu_1 - \mu_2, \sigma^2)$ is $\mathsf{N}\chi^{-2}(7.5, (1/12 + 1/7)^{-1}, 17, 446)$ and so the posterior pdf of $T = \frac{\mu_1 - \mu_2 - 7.5}{9.9}$ is $t(17)$ giving $P(\mu_1 > \mu_2 | x, y) = P(T > -0.76 | x, y) = \Phi_{17}(0.76)$.

6. Let $C_0, C_1, \cdots$ denote the US population counts (in millions) in the census years since 1910 (so, C_0 is the count of 1910, C_1 is the count of 1920 and so on). Suppose these counts are modeled by the growth equation $C_t = C_{t-1}e^{X_t}$, $t = 1, 2, \cdots$ and the log-growth rates $X_1, X_2, \cdots$ are modeled as $X_i \overset{\text{IID}}{\sim} \mathsf{Normal}(\mu, \sigma^2)$ where σ is fixed at 0.03 and $\mu \in (-\infty, \infty)$ is assigned the Jeffreys' prior $\xi(\mu) = \text{const}$. Give a central 95% posterior predictive interval for the population count at the next census (in 2020) based on the following observations. [6 points]

Year	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000	2010
Population	92	106	123	132	151	179	203	226	248	281	308

We have data $X_1, \cdots, X_{10}$ and want to predict $308e^{X_{11}}$. The recorded data on $X_1, \cdots, X_{10}$ are (differences of the log populations values): 0.14, 0.15, 0.07, 0.13, 0.17, 0.13, 0.11, 0.09, 0.12, 0.09, with $\bar{x} = 0.12$. Therefore th 95% central posterior predictive interval for X_{11} is: $0.12 \mp 1.96 \times 0.03\sqrt{1 + 1/10} = 0.12 \times 0.061 = [0.059, 0.181]$. Therefore the central 95% posterior credible interval for next census population is $[308 \cdot e^{0.059}, 308 \cdot e^{0.181}] = [308 \cdot 1.06, 308 \cdot 1.20] = [326, 369]$.