STAT 4224/5224 **Midterm** **Name: Key**
Spring 2023 **Solution** **Uni:**

1. True or False:

   a. To generate S independent random draws from some posterior predictive distribution $X_{new}|x_1, \ldots, x_n$, we first simulate a single draw from the posterior, $\theta_0 \sim f(\theta|\, x_1, \ldots, x_n)$, and then use it to generate iid $X_{new}^{(1)}, \ldots, X_{new}^{(S)} \sim f(x|\theta_0)$. F

   b. For any two evets $P(A \cup B) \leq \max\{P(A), P(B)\}$. F

   c. When analyzing a posterior distribution by computing a grid approximation of the posterior density, it is important that the grid region includes almost all of the support of the posterior probability. T

   d. In the multivariate Normal model with multivariate normal prior on $\boldsymbol{\theta}$ and known covariance matrix $\boldsymbol{\Sigma}$, the posterior predictive distribution of $\boldsymbol{x}_{new}|\boldsymbol{x}$ is also multivariate normal. T

   e. In the Binomial model with Beta(a, b) prior, the posterior variance is always smaller than the prior variance, regardless of the values of n, x, a and b. F

   f. The variance of the prior predictive distribution for the Binomial model with Beta(a, b) prior is given by $\frac{(a+x)(b+n-x)}{(a+b+n)^2}$. F

   g. If $\boldsymbol{X} \sim \mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $X_i \sim N(\mu_i, \sigma_{ii}),\ \ i = 1, \ldots, p$. T

   h. If $X_1, \ldots, X_n|\theta \sim$ iid Poisson(θ), then $\Gamma(a, b)$ is a conjugate prior. T

   i. For sampling from $f(x_1, x_2, x_3)$ the Gibbs sampler will iterate between $f(x_1|x_2)$, $f(x_2|x_3)$ and $f(x_3|x_1)$. F

   j. A finite state space Markov chain $X_0, X_2, \ldots$ satisfies the following equation: $P(X_n = s_n|X_{n-1} = s_{n-1},\ \ldots,\ X_0 = s_0) = P(X_n = s_n|X_{n-1} = s_{n-1}), \forall n$ T

   k. *Bonus*: Let X, Y, Z be random variables with joint density $f(x, y, z)$. If there exist functions g and h such that $f(x, y, z) \propto g(x, y)h(x, z)$, then Y and Z are conditionally independent given X. T

2. Let $X_1, X_2 \mid \theta, \sigma^2$ be iid $N(\theta, \sigma^2)$, where $\theta \mid \sigma^2 \sim N(0, 1)$ and $\sigma^2 \sim IG(1, ½)$.
Note that if $Y \sim IG(a, b)$, then $f(y) = \frac{b^a}{\Gamma(a)} y^{-a-1} e^{-\frac{b}{y}},\ y > 0$.

a. Derive (up to a normalizing constant) the posterior density $f(\theta|\sigma^2, x_1, x_2)$ of $\theta \mid \sigma^2, x_1, x_2$

$$f(x_1, x_2, \theta, \sigma^2) \propto (\sigma^2)^{-1} e^{-\frac{1}{2}\left[\frac{(x_1-\theta)^2}{\sigma^2} + \frac{(x_2-\theta)^2}{\sigma^2}\right]} e^{-\frac{1}{2}\theta^2} (\sigma^2)^{-2} e^{-\frac{1}{2\sigma^2}}$$
$$= (\sigma^2)^{-3} e^{-\frac{1}{2}\left[\frac{2}{\sigma^2}\theta^2 + \theta^2 - \frac{2(x_1+x_2)}{\sigma^2}\theta + \frac{x_1^2+x_2^2}{\sigma^2}\right]}$$
$$= (\sigma^2)^{-3} e^{-\frac{1}{2\sigma^2}(x_1^2+x_2^2)} e^{\frac{1}{2}\left(\frac{2+\sigma^2}{\sigma^2}\right)\left(\frac{x_1+x_2}{2+\sigma^2}\right)^2} e^{-\frac{1}{2}\left[\frac{2+\sigma^2}{\sigma^2}\left(\theta - \frac{x_1+x_2}{2+\sigma^2}\right)^2\right]}$$
$$\Rightarrow f(\theta|\sigma^2, x_1, x_2) \propto e^{-\frac{1}{2}\left[\frac{2+\sigma^2}{\sigma^2}\left(\theta - \frac{x_1+x_2}{2+\sigma^2}\right)^2\right]}$$
$$\Rightarrow \theta|\sigma^2, x_1, x_2 \sim N\left(\frac{x_1 + x_2}{2 + \sigma^2}, \frac{\sigma^2}{2 + \sigma^2}\right)$$

b. Derive (up to a normalizing constant) the posterior density $f(\sigma^2|x_1, x_2)$ of $\sigma^2 \mid x_1, x_2$

$$f(\sigma^2|x_1, x_2) \propto (\sigma^2)^{-3} e^{-\frac{1}{2\sigma^2}(x_1^2+x_2^2)} e^{\frac{1}{2}\left(\frac{2+\sigma^2}{\sigma^2}\right)\left(\frac{x_1+x_2}{2+\sigma^2}\right)^2} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left[\frac{2+\sigma^2}{\sigma^2}\left(\theta - \frac{x_1+x_2}{2+\sigma^2}\right)^2\right]} d\theta$$
$$= (\sigma^2)^{-3} e^{-\frac{1}{2\sigma^2}(x_1^2+x_2^2)} e^{\frac{1}{2\sigma^2}\frac{(x_1+x_2)^2}{2+\sigma^2}} \sqrt{2\pi \frac{\sigma^2}{2 + \sigma^2}}$$
$$\propto \frac{(\sigma^2)^{-2}}{\sqrt{2 + \sigma^2}} e^{-\frac{1}{2\sigma^2}\left[x_1^2 + x_2^2 - \frac{(x_1+x_2)^2}{2+\sigma^2}\right]}$$

Note this is not a standard distribution.

c. If you could simulate both densities in a) and b), explain briefly how you would use them to estimate $E(\theta|x_1, x_2)$.

1. Generate σ^2 from $f(\sigma^2|x_1, x_2)$ in part b.
2. Use the generated σ^2 from step 1 to generate θ from $f(\theta|\sigma^2, x_1, x_2)$ in part a.
3. Repeat steps 1 and 2 S times to obtain a sequence $(\theta, \sigma^2)^{(1)}, \ldots, (\theta, \sigma^2)^{(S)}$
4. Calculate $\frac{1}{S}\sum_{i=1}^{S} \theta^{(i)} \approx E(\theta|x_1, x_2)$.

3. In a sample of 100 adults, 21 of them have a travel credit card. Let θ represent the proportion of the entire population that has a travel credit card and assume Beta(a, b) prior on it. Use without proof the fact that the posterior distribution is:

$$\theta \mid x \sim \text{Beta}(a + x, b + n - x), \text{ where } x = \sum_{i=1}^{n} x_i.$$

a) If we strongly believe that $\theta < 0.5$, which of the following priors is the most appropriate to use?

   i. Beta(1, 1)
   ii. Beta(5, 3)
   iii. Beta(3, 5)
   iv. Beta(½, ½)

b) Regardless of your answer to part a), suppose we want to use Beta(1, 1) prior. Find the posterior:

   i. Mean

$$E(\theta \mid x) = \frac{a+x}{a+b+n} = \frac{1+21}{1+1+100} = 0.22$$

   ii. Mode

$$\frac{a+x-1}{a+b+n-2} = \frac{1+21-1}{1+1+100-2} = 0.21$$

   iii. Standard deviation

$$Var(\theta \mid x) = \frac{(a+x)(b+n-x)}{(a+b+n)^2(a+b+n+1)} = \frac{(1+21)(1+100-21)}{(1+1+100)^2(1+1+100+1)}$$

$$= 0.000016$$

$$\sigma(\theta \mid x) = 0.004$$

c) Suppose you want to calculate a 95% posterior quantile-based confidence interval using statistical software. Write down the exact command you would use.

~~quantile~~ qbeta(22, 80, c(0.025, 0.975))

d) Explain briefly how you can approximate the posterior median.

median(rbeta(10000, 22, 80))

4. A hospital receives 70% of its flu vaccine from Company A, 20% from Company B and the remainder from Company C. Each shipment contains a large vial of vaccine. It is known that from Company A, 10% of the vials are ineffective; from Company B, 86% are ineffective, and from Company C 47% are ineffective. The hospital received one shipment and randomly selected without replacement to test for effectiveness 10 of the vials.

a) Find the probability that exactly 2 of the selected vials are ineffective.

$$P(x=2) = P(A)\cdot P(x=2|A) + P(B)\,P(x=2|B) + P(C)\,P(x=2|C)$$

$$= 0.7\left[\binom{10}{2}(0.1)^2(0.9)^8\right] + 0.2\left[\binom{10}{2}(0.86)^2(0.14)^8\right]$$

$$+ 0.1\left[\binom{10}{2}(0.47)^2(0.53)^8\right]$$

$$= 0.7(0.194) + 0.2(0) + 0.1(0.062)$$

$$= 0.142$$

b) Given that exactly 2 of the selected vials are ineffective, compute each of the probabilities that it came from Company A, B or C.

$$P(A|x=2) = \frac{P(A)\,P(x=2|A)}{P(x=2)} = \frac{0.7(0.194)}{0.142}$$

$$= 0.956$$

$$P(B|x=2) \approx 0$$

$$P(C|x=2) = \frac{0.1(0.062)}{0.142} = 0.044$$