STAT 4224/5224 **Midterm** **Name:**
Spring 2023 **Practice - Solutions** **Uni:**

Important:
You are allowed to bring one letter-sized page written on one side of your own notes.

1. True or False:

   a. There exist r.v.'s X and Y such that Var(Y | X) > Var(Y) with positive probability. T

   b. The Monte Carlo method in Bayesian inference is useful for approximating posterior expectations E[$g(\theta)|\boldsymbol{x}$] but cannot be used to compute approximate Bayesian confidence intervals. F

   c. If $X_1, \ldots, X_n | \theta$ are iid Poisson(θ) and $\theta \sim \Gamma(a, b)$, then the posterior predictive distribution of X_{new} is Poisson distributed as well. F

   d. In the Binomial model with uniform prior on θ, the mean of the posterior predictive distribution can be written as weighted average of the population mean θ and the sample mean $\bar{x}$. F

   e. Let X, Y, Z be random variables with joint density $p(x, y, z)$ such that there exist functions f, g, and h with the property that $p(x, y, z) \propto f(x)g(y)h(z)$. Then X, Y and Z are mutually independent. T

   f. If $X_1, X_2, \ldots, X_n$ are exchangeable then they are also iid. F (only the reverse implication is true)

   g. If $\theta^{(1)}, \ldots, \theta^{(S)}$ are iid from the posterior $f(\theta|x_1, \ldots, x_n)$, then a Monte Carlo approximation of $E(g(\theta)|x_1, \ldots, x_n)$ is given by $\frac{1}{S}\sum_{i=1}^{S} g\left(\theta^{(i)}\right)$. T

   h. A probability density function that satisfies $f(x) \propto e^{ax^2 + bx}$ for any real x, and some $a < 0$, is necessarily that of a normal distribution. T

   i. The R command `pnorm(-4.5, -3.4, sqrt(10))` gives the probability P(X > -4.5), where $X \sim N(\mu = -3.4, \sigma^2 = 10)$. F because `pnorm` gives P($X$ < -4.5)

   j. Suppose we have one observation X such that:
   $$X \mid \theta \sim N(\theta, \sigma^2)$$
   $$\theta \sim N(\mu, \tau^2)$$
   Then the marginal distribution of X is normal. T

   k. Bonus: A 95% Bayesian confidence interval for θ is a random interval [$l(\boldsymbol{X})$, $u(\boldsymbol{X})$] such that, for any value of $\theta \in \Theta$, P[$l(\boldsymbol{X}) < \theta < u(\boldsymbol{X})|\theta$] = 0.95. F

2. Let $X_1 \mid \theta \sim N(\theta, 1)$, where $\theta \sim N(0, 1)$.

   a. Derive the posterior distribution of $\theta \mid x_1$

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

   b. Derive the posterior predictive distribution of $X_{\text{new}} \mid x_1$.

$$f(x_{new}|x_1) = \int_{-\infty}^{\infty} f(x_{new}|\theta, x_1) f(\theta|x_1) d\theta = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x_{new}-\theta)^2} \frac{1}{\sqrt{\pi}} e^{-\frac{1}{2}2\left(\theta-\frac{1}{2}x_1\right)^2} d\theta$$
$$\Rightarrow f(x_{new}|x_1) \propto \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\theta^2-2\theta x_{new}+x_{new}^2+2\theta^2-2\theta x_1+\frac{1}{4}x_1^2\right)} d\theta$$
$$\propto e^{-\frac{1}{2}x_{new}^2} \int_{-\infty}^{\infty} e^{-\frac{1}{2}[3\theta^2-2\theta(x_{new}+x_1)]} d\theta$$
$$\propto e^{-\frac{1}{2}x_{new}^2} e^{\frac{1}{2}\left[\frac{1}{3}(x_{new}+x_1)^2\right]} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left[3\left(\theta-\frac{1}{3}(x_{new}+x_1)\right)^2\right]} d\theta$$
$$\propto e^{-\frac{1}{2}\left(x_{new}^2-\frac{1}{3}x_{new}^2-\frac{2}{3}x_{new}x_1\right)} \sqrt{2\pi \times \frac{1}{3}}$$
$$\propto e^{-\frac{1}{2}\left(\frac{2}{3}x_{new}^2-\frac{2}{3}x_{new}x_1\right)} \propto e^{-\frac{1}{2}\left[\frac{2}{3}\left(x_{new}-\frac{1}{2}x_1\right)^2\right]}$$
$$\Rightarrow X_{new}|x_1 \sim N\left(\frac{1}{2}x_1, 1.5\right)$$

Note: for alternative proof see lecture notes.

3. Suppose we have independent $X_1, \ldots, X_n \sim$ Bernoulli(θ) and $\theta \sim$ Beta(a, b). Use without proof the fact that the posterior distribution is: $\theta \mid x \sim$ Beta($a + x$, $b + n - x$), where $x = \sum_{i=1}^{n} x_i$.

a) Suppose $a = b = 1$. For which case will the 95% posterior interval be shorter: ($n = 10, x = 4$) or ($n = 20, y = 8$)? Explain why, but no need to prove it rigorously.

$$Var(\theta|x) = \frac{(a+x)(b+n-x)}{(a+b+n)^2(a+b+n+1)} = \frac{35}{1872} = 0.01869658$$
$$Var(\theta|y) = \frac{(a+y)(b+n-y)}{(a+b+n)^2(a+b+n+1)} = \frac{117}{11132} = 0.01051024$$

Therefore, the interval will be shorter for ($n = 20$, $y = 8$).

b) Suppose the observed data are $n = 20$ and $x = 8$. For which prior distribution will the 95% posterior interval be shorter, Beta(2, 3) or Beta(4, 6)? Again, just explain why.

The interval will be shorter for the Beta(4, 6) prior.

c) Derive the general formula for the posterior predictive distribution of X_{new}.

$$X_{new}|x_1, \ldots x_n \sim Bernoulli\left(\frac{a + \sum_{i=1}^{n} x_i}{a+b+n}\right)$$

Done in class/homework.

d) If $n = 2$, $x = 0$, $a = 4$, and $b = 4$, which one is smaller, the posterior predictive variance or the prior predictive variance?

The prior predictive distribution is:

$$X_{new}|x_1, \ldots x_n \sim Bernoulli\left(\frac{a}{a+b}\right)$$
$$\Rightarrow Var_{PriorPred}(X_{new}) = \left(\frac{a}{a+b}\right)\left(1 - \frac{a}{a+b}\right) = \frac{ab}{(a+b)^2} = \frac{16}{64} = \frac{1}{4} = .25$$
$$Var_{PostPred}(X_{new}) = \left(\frac{a+x}{a+b+n}\right)\left(1 - \frac{a+x}{a+b+n}\right) = \frac{(a+x)(b+n-x)}{(a+b+n)^2} = \frac{24}{100} = .24 < .25$$

4. Based on Paul's symptoms, he believes it is equally likely that he has either Covid or the flu. Thus, each disease has a prior probability of 0.5; assume there is zero probability that he has both, and zero probability he has neither.
   It is known that:
   - A rapid Covid test is 70% sensitive and 97% specific. Meaning, if Paul has Covid, the probability of a negative test is 0.3; if he has the flu, the probability of a negative test is 0.97.
   - An influenza test is 60% sensitive and 80% specific. So, if Paul has Covid, the probability of a negative flu test is 0.8; if Paul has the flu, the probability of a negative test is 0.4.

a) Suppose Paul takes a rapid Covid test and gets a negative result. What is his updated belief about the probability he has Covid versus the flu?

Define θ_C = "Paul has Covid", θ_F = "Paul has flu", x_1 = "Covid test is negative" and x_2 = "Flu test is negative".
Then:

$$P(\theta_C|x_1) = \frac{P(\theta_C)P(x_1|\theta_C)}{P(x_1)} = \frac{0.5(0.3)}{0.5(0.3) + 0.5(0.97)} = 0.236$$

$$P(\theta_F|x_1) = 1 - 0.236 = 0.764$$

b) Suppose that in addition to the negative Covid test, Paul takes an independent flu test and gets a negative result there as well. What is his updated belief about the probability he has Covid versus the flu?

$$P(\theta_C|x_1, x_2) = \frac{P(\theta_C|x_1)P(x_2|\theta_C, x_1)}{P(x_1, x_2)} = \frac{0.236(0.8)}{0.236(0.8) + 0.764(0.4)} = 0.382$$

$$P(\theta_F|x_1, x_2) = 1 - 0.382 = 0.618$$

c) Which result changed Paul's belief more dramatically, the negative Covid test, the negative flu test, or did they both have an equal effect? Does it make intuitive sense?

The negative Covid test was more informative, which makes sense since it is the test with higher sensitivity and specificity.