Midterm Examination Solutions

1. **Conjugate Priors.**

   (a) **(3 pts)** Define a *conjugate prior*.
   A class $\mathcal{P}$ of prior distributions for a parameter θ is called *conjugate* for a sampling model $p(y \mid \theta)$ if $p(\theta) \in \mathcal{P} \Rightarrow p(\theta \mid y) \in \mathcal{P}$.

   (b) **(4 pts)** What is the conjugate prior for the Exponential(λ) sampling distribution? Use the sampling distribution $f(x \mid \lambda) = \lambda \exp(-\lambda x)$. Both state the conjugate prior and show that it is conjugate.
   The conjugate prior for the exponential distribution is the gamma distribution. Given a sampling distribution $f(x \mid \lambda) = \lambda \exp(-\lambda x)$ and a prior distribution $\pi(\lambda \mid \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \exp(-\beta\lambda)\lambda^{\alpha-1}$, the posterior distribution is Gamma$(\alpha + n, \beta + \sum x_i)$.

2. **(5 pts) Jeffreys' Prior.** Derive the Jeffreys' prior for the normal mean assuming that the variance is known.
   Assume σ^2 is known. We have

   $$Y \sim \text{Normal}(\mu, \sigma^2)$$

   where

   $$f(y \mid \mu) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{1}{2\sigma^2}(y - \mu)^2) \ .$$

   Taking logs,

   $$\log(f(y \mid \mu)) = -\frac{1}{2\sigma^2}(y - \mu)^2 + \text{constant} \ .$$

   Taking partial derivatives,

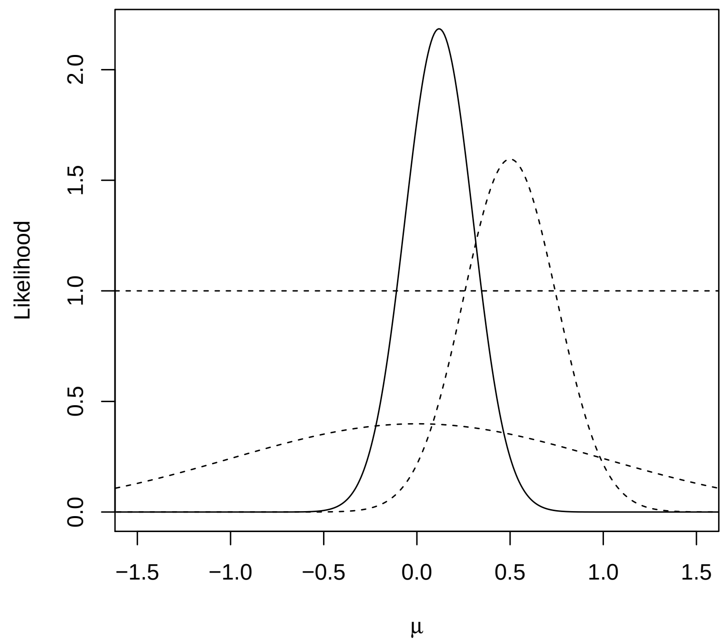
   $$\frac{\partial^2 \log(f(y \mid \mu))}{\partial \mu^2} = -\frac{1}{\sigma^2}$$

   This expression does not depend on y, so $I(\mu) = \frac{1}{\sigma^2} = \text{constant}$. This implies that our prior should be $\pi(\mu) \propto 1$ or uniform on the real line.

3. **Priors**. The plot is the likelihood function for the mean of a Normal$(\mu, 1)$ distribution evaluated using 30 observations.

   (a) **(3 pts)** Sketch (and label) a *diffuse* prior on the plot. Write a short definition of a diffuse prior.
   From Box and Tiao (1973), "a prior which is dominated by the likelihood is one which does not change *very much* over the region in which the likelihood is appreciable and does not assume large values outside that range."

   (b) **(3 pts)** Sketch (and label) an *informative* prior on the plot. Write a short definition of an informative prior.
   Colloquially, an informative prior is usually defined using its negative: any prior distribution trying to capture something besides "I have no idea."



(c) **(4 pts)** Sketch (and label) the Jeffreys' prior on the plot (see Question 2). Define an *improper* prior. Is the Jeffreys' prior proper or improper?
An *improper* prior is one that does not integrate to 1 (that does not have a finite integral). The Jeffreys' prior for the normal mean (variance known) is improper, as it is uniform over the real line.

4. **Predictive Distribution**.

(a) **(3 pts)** Let $X_1, \ldots, X_n \mid p \sim \text{Bernoulli}(p)$ and assume that we observe x successes and $n - x$ failures. Let the prior distribution for p be $\text{Beta}(\alpha, \beta)$. What is the posterior distribution for p?
The beta distribution is conjugate for the Bernoulli sampling distribution. The posterior distribution is $\text{Beta}(\alpha + x, \beta + n - x)$.

(b) **(3 pts)** Describe how to draw a sample from the posterior predictive distribution of X_{n+1} assuming that you have a sample $p^{(i)}, i = 1, \ldots, m$ from the posterior distribution of p.
For each sample $p^{(i)}$, draw $\tilde{y}^{(i)} \sim \text{Bernoulli}(p^{(i)})$. $\tilde{y}^{(i)}, i = 1, \ldots, m$ is a sample from the posterior predictive distribution.

(c) **(4 pts)** Using the posterior distribution from (4a), derive the predictive distribution for X_{n+1} assuming that it is drawn from the same Bernoulli(p) distribution.

$$
\begin{aligned}
f(x_{n+1} \mid x_1, \ldots, x_n) &= \int_0^1 f(x_{n+1} \mid p)\pi(p \mid x_1, \ldots, x_n)dp \\
&= \int_0^1 p^{x_{n+1}}(1-p)^{1-x_{n+1}} p^{\alpha+x-1}(1-p)^{\beta+n-x-1} \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+x)\Gamma(\beta+n-x)} dp \\
&= \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+x)\Gamma(\beta+n-x)} \int_0^1 p^{x_{n+1}+\alpha+x-1}(1-p)^{1-x_{n+1}+\beta+n-x-1} dp \\
&= \frac{\Gamma(\alpha+\beta+n)\Gamma(\alpha+x+x_{n+1})\Gamma(\beta+n-x+1-x_{n+1})}{\Gamma(\alpha+x)\Gamma(\beta+n-x)\Gamma(\alpha+\beta+n+1)}
\end{aligned}
$$

You get full credit if you got to this point.
Notice that x_{n+1} is a Bernoulli trial and can only take on values of 0 or 1.

$$
\begin{aligned}
P(X_{n+1} = 1) &= \frac{\Gamma(\alpha+\beta+n)\Gamma(\alpha+x+1)\Gamma(\beta+n-x)}{\Gamma(\alpha+x)\Gamma(\beta+n-x)\Gamma(\alpha+\beta+n+1)} \\
&= \frac{\Gamma(\alpha+\beta+n)\Gamma(\alpha+x+1)}{\Gamma(\alpha+x)\Gamma(\alpha+\beta+n+1)}
\end{aligned}
$$

Since $\Gamma(x+1) = x\Gamma(x)$, we can simplify this expression

$$
\begin{aligned}
P(X_{n+1} = 1) &= \frac{\Gamma(\alpha+\beta+n)\Gamma(\alpha+x+1)}{\Gamma(\alpha+x)\Gamma(\alpha+\beta+n+1)} \\
&= \frac{\alpha+x}{\alpha+\beta+n}
\end{aligned}
$$

5. **Sampling.**

   (a) **(5 pts)** Suppose that you have a posterior distribution $\pi(\theta \mid x_1, \ldots, x_n)$ with support on $(0, 1)$. Describe how you would use *brute force* sampling to draw a random sample from $\pi(\theta \mid x_1, \ldots, x_n)$.

   - Choose 10,000 values between 0 and 1.
   - Compute the value of the posterior distribution at each point of the 10,000 points.
   - Normalize by dividing each value by the sum of the 10,000 values.
   - Take a random sample with replacement from the 10,000 values using the normalized values as probabilities.

   (b) **(3 pts)** Suppose that we have a sample $\theta^{(i)}, i = 1, \ldots, m$ from $\pi(\theta \mid x_1, \ldots, x_n)$. How could we use this sample to evaluate the posterior probability that $\theta > 0.8$? Estimate the probability that $\theta > 0.8$ by the proportion of values from the sample that exceed 0.8.