The exam is open book and open notes. No laptops are permitted but you may use a calculator. Giving or receiving assistance from other students is not allowed. Show work to receive full credit! Partial credit will be given, but only for work written on the exam. If you can’t solve a problem completely, give the steps you would take to solve the problem to receive partial credit.

For questions 1a, 2b, and 3 give the parametric distribution, e.g., Gamma($n, \alpha$), to receive full credit.

1. (25 points) Assume the population follows a normal distribution with mean $\mu$ and variance 1. Before conducting a large study, we do a literature review and find a pilot study which collected data from the same population. Let $n_1$ and $\bar{Y}_1$ be the sample size and sample mean from this pilot study. Based on this, we form a prior $\mu \sim N(\bar{Y}_1, 1)$. Now we collect a large sample with $n_2$ observations. Our estimate of $\mu$ will be its posterior mean given these $n_2$ observations.

(a) Compute the posterior distribution of $\mu$.

In class we saw that $\mu \mid y \sim N\left(\frac{\frac{n_2}{\bar{Y}_2}}{\sigma^2} + \frac{\frac{1}{\sigma^2}}{\sigma^2}, \frac{1}{\sigma^2 + \frac{1}{\sigma^2}}\right)$

Plugging in $\sigma^2 = \bar{Y}_1$, $\sigma^2 = 1$ gives:

$\mu \mid y \sim N\left(\frac{\frac{n_2}{\bar{Y}_2} \bar{Y}_1}{\sigma^2 + 1}, \frac{1}{\sigma^2 + 1}\right)$

(b) Compute the bias, variance and mean squared error of the posterior mean (assuming both the pilot and large studies are random).

$E(\mu \mid y) = \hat{\mu} = \frac{n_2 \bar{Y}_2 + \bar{Y}_1}{n_2 + 1}$

$E(\bar{Y}_1) = \frac{n_2 E(\bar{Y}_2) + E(\bar{Y}_1)}{n_2 + 1} = \frac{n_2 \mu + \mu}{n_2 + 1} = \mu$

so $\hat{\mu}$ is unbiased.

$\text{MSE}(\hat{\mu}) = \text{Var}(\hat{\mu}) = \frac{n_2^2}{(n_2 + 1)^2} \text{Var}(\bar{Y}_2) = \frac{1}{(n_2 + 1)^2} \text{Var}(\bar{Y}_1) = \frac{1}{(n_2 + 1)^2}$

(c) How does this compare with a standard MLE analysis?

The MLE is $\bar{Y}_2$ which is unbiased with variance $\text{MSE} = \frac{1}{n_2}$.

Both $\hat{\mu}$ and the MLE are unbiased, and

$\text{MSE}(\hat{\mu}) = \frac{n_1 n_2 + 1}{n_1 (n_2 + 1)^2} \leq \frac{n_1 n_2 + n_1}{n_1 (n_2 + 1)^2} = \frac{1}{n_2 + 1} < \frac{1}{n_2} = \text{MSE}(\text{MLE})$. 

1
2. (25 points) Assume $Y$ follows a geometric distribution with $P(Y = y) = \theta(1 - \theta)^y$ and $E(Y) = (1 - \theta)/\theta$ for $\theta \in (0, 1)$.

(a) Find the Jeffreys' prior for $\theta$.

$$l = \log(f(y)) = \log(\theta) + y \log(1 - \theta)$$

$$E(y) = \frac{1}{\theta}$$

$$E(\theta) = \frac{1}{\theta^2 (1 - \theta)}$$

So the Jeffreys' prior is

$$p(\theta) \propto \sqrt{-E(\theta^{1/2})}$$

$$= \sqrt{\frac{1}{\theta^2} + \frac{1}{(1-\theta)^2}}$$

$$= \frac{1}{1 - \theta}$$

(b) Compute the posterior distribution of $\theta$ under a Uniform$(0, 1)$ prior.

$$p(\theta | y) \propto \theta^{y+1} (1-\theta)^{y+1}$$

$$\Rightarrow \theta | y \sim \text{Beta}(y+1, y+1)$$

(c) Assuming the prior in (b), use the Bayesian central limit theorem to find a normal approximation to the posterior. Would you recommend this approximation in this case?

From (a), the posterior mode is

$$\hat{\theta} = \frac{1}{1 - \theta} = 0$$

From (b), $-E' = I = (y+1)^2 \frac{y}{(1 - \frac{1}{y+1})^2}$

$$\hat{\theta} = \frac{1}{y+1}$$

$$= \frac{1}{y+1}$$

$$= \frac{1}{y+1}$$

$$\Rightarrow \theta | y \sim N\left(\frac{1}{y+1}, \frac{1}{(y+1)^2 + (y+1)^2} \right)$$

2
3. (25 points) Consider the time series model $Y_1 \sim N(0, 10^2)$ and $Y_t | Y_{t-1}, ..., Y_1 \sim N(\rho Y_{t-1}, \sigma^2)$ for $t = 2, ..., n$. Assume priors $\rho \sim N(0, \tau^2)$ and $\sigma^2 \sim \text{InvGamma}(a, b)$.

(a) Compute the posterior distribution of $\rho | \sigma^2, Y_1, ..., Y_n$.

$$f(\rho | y, \sigma^2) \propto \frac{1}{\sigma^2} \exp \left( -\frac{1}{2\sigma^2} \left( \sum_{t=2}^{n} (Y_t - \rho Y_{t-1})^2 \right) \right) \exp \left( -\frac{1}{2\tau^2} \left( \sum_{t=2}^{n} Y_{t-1}^2 \right) \right)$$

$$\propto -\frac{1}{2} \left[ \sum_{t=2}^{n} \left( \frac{1}{\sigma^2} Y_{t-1}^2 - 2\rho Y_{t-1} + \rho^2 Y_{t-1}^2 + \frac{\rho^2}{\tau^2} \right) \right]$$

$$\propto -\frac{1}{2} \left[ -2 \left( \frac{1}{\sigma^2} \sum_{t=2}^{n} Y_{t-1} \right) \rho + \left( \frac{1}{\sigma^2} \sum_{t=2}^{n} Y_{t-1}^2 + \frac{1}{\tau^2} \right) \rho^2 \right]$$

So $\rho | \text{rest} \sim N \left( \frac{\sum_{t=2}^{n} Y_{t-1} Y_t}{\sum_{t=2}^{n} Y_{t-1}^2 + \frac{1}{\tau^2}}, \frac{1}{\sigma^2} \sum_{t=2}^{n} Y_{t-1}^2 + \frac{1}{\tau^2} \right)$

(b) Compute the posterior distribution of $\sigma^2 | \rho, Y_1, ..., Y_n$.

$$f(\sigma^2 | \rho, y) \propto \frac{1}{\sigma^{n/2}} \exp \left( -\frac{1}{2\sigma} \left( \sum_{t=2}^{n} (Y_t - \rho Y_{t-1})^2 \right) \right) \exp \left( -\frac{n}{2\sigma^2} \right)$$

$$\propto \frac{1}{\sigma^{n/2}} \exp \left( -\frac{1}{2\sigma} \sum_{t=2}^{n} (Y_t - \rho Y_{t-1})^2 \right) \exp \left( -\frac{n}{2\sigma^2} \right)$$

So $\sigma^2 | \text{rest} \sim \text{InvGamma}(\frac{n}{2} + a, \frac{1}{2} \text{SSE} + b)$
4. (25 points) A drug company has developed $n$ new drugs. Based on initial studies, the posterior probability that drug $j$ is effective is $\theta_j \in (0,1)$, independent over $j$. The company must now decide whether the drugs are promising enough to conduct a large-scale trial to conclusively demonstrate their effectiveness. The company has unlimited money to initiate these trials. If drug $j$ is really effective but they decide not to pursue a trial, it costs the company $L_j$ dollars in lost profits. If the company pursues a trial for drug $j$ and it turns out not to be effective, it costs the company $K_j$ dollars. What is the Bayes decision rule in this case? That is, a vector $d = (d_1, \ldots, d_n)$ where $d_j = 1$ if a trial is initiated for drug $j$ and $d_j = 0$ otherwise.

Define $T_j = I(\text{drug } j \text{ is effective})$

The loss is

$$l(d,T) = \sum_{j=1}^n I(d_j = 0, T_j = 1) L_j + I(d_j = 1, T_j = 0) K_j$$

The expected loss (risk) is

$$E(l(d,T)) = \sum_{j=1}^n E(I(d_j = 0, T_j = 1) L_j) + E(I(d_j = 1, T_j = 0) K_j)$$

$$= \sum_{j=1}^n I(d_j = 0) \theta_j L_j + I(d_j = 1) (1-\theta_j) K_j$$

Since the risk is additive in the decisions, the Bayes rule simply minimizes each term separately.

So

$$d_j = \begin{cases} 1 & (1-\theta_j) K_j < \theta_j L_j \\ 0 & \text{otherwise} \end{cases}$$