STAT 576 SOLUTION FOR HOMEWORK 1

1 (Single Parameter Bayesian Inference)**.** Suppose $X_1, X_2, \ldots,$ be an i.i.d. sequence of Bernoulli(θ) trials with unknown parameter $\theta \in (0,1)$. Let Y_k be the number of trials before k-th failures for some fixed $k \geq 1$. That is

$$Y_k = \min \left\{ n : \sum_{i=1}^{n} (1 - X_i) \geq k \right\}$$

with the convention that $\min \varnothing = +\infty$.

(a) Show that $\mathbb{P}[Y_k < \infty] = 1$.
Hint: First show that $Y_k = Z_1 + Z_2 + \cdots + Z_k$ with $Z_1, \ldots, Z_k \sim$ Geometric$(1-\theta)$. Then for any constant c, show that $\mathbb{P}[Y_k \geq c] \leq c^{-1}\mathbb{E}[Y_k]$, which converges to 0 when $c \to \infty$.
(b) Compute $p(y_k \mid \theta)$ and relate it to the negative-binomial distribution.
(c) Find the conjugate prior on θ for the distribution in (b). Identify the hyperparameters.
(d) Use the prior in part (c) to get the posterior distribution.
(e) What are the MAP estimator and the posterior mean estimator for θ?
(f) Compute the Jeffreys prior and figure out whether it belongs to the conjugate prior family.

Solution.

(a) Let Z_i be the number of trials between the $(i-1)$-th failure (exclusive) and the i-th failure (inclusive). Since all trails are independent, the Z_i's are independent and each Z_i follows a geometric distirbution with parameter $1-\theta$. Notice that $Y_k = Z_1 + Z_2 + \cdots + Z_k$. Therefore $\mathbb{E}[Y_k] = \mathbb{E}[Z_1] + \mathbb{E}[Z_2] + \cdots + \mathbb{E}[Z_k] = k/(1-\theta) < \infty$. By Markov inequality, we have $\mathbb{P}[Y_k \geq c] \leq \mathbb{E}[Y_k]/c = k/[c(1-\theta)]$. Let $c \to \infty$. We have $\mathbb{P}[Y_k < \infty] = 1$.
(b) We have

$$\begin{aligned} p(y_k \mid \theta) &= \mathbb{P}[k-1 \text{ failures in the first } y_k - 1 \text{ trials}] \times \mathbb{P}[y_k\text{-th trial is a failure}] \\ &= \binom{y_k - 1}{k-1} \theta^{y_k - k} (1-\theta)^k \end{aligned}$$

By comparing this probability density(mass) function with that of negative binomial distributions, we find that $y_k - k$ follows a negative binomial distribution with parameters $(k, 1-\theta)$.
(c) The conjugate prior should have a density $p(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$, which is the Beta$(\alpha, \beta)$ distribution. (α, β) is the hyperparameter.
(d) We choose $p(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$. The posterior is

$$p(theta \mid y_k) \propto p(y_k \mid \theta) p(\theta) \propto \theta^{\alpha + y_k - k - 1}(1-\theta)^{\beta + k - 1}$$

Hence, the posterior is Beta$(\alpha + y_k - k, \beta_k)$.
(e) Use the properties of the Beta distribution. The MAP estimator is

$$\frac{\alpha + y_k - k - 1}{\alpha + \beta + y_k - 2}.$$

The posterior mean is

$$\frac{\alpha + y_k - k}{\alpha + \beta + y_k}$$

(f) The log-likelihood function is

$$\ell(\theta) = (y_k - k)\log\theta + k\log(1-\theta) + C$$

for some constant C. The score function is

$$\ell'(\theta) = \frac{y_k - k}{\theta} - \frac{k}{1-\theta}.$$

The second-order derivative is

$$\ell''(\theta) = -\frac{y_k - k}{\theta^2} - \frac{k}{(1-\theta)^2}.$$

Then we get the Fisher's information:

$$\mathcal{I}(\theta) = -\mathbb{E}[\ell''(\theta)] = \frac{k}{\theta(1-\theta)^2}$$

The correspondingly Jeffreys prior is

$$p(\theta) \propto \sqrt{\mathcal{I}(\theta)} = \frac{1}{(1-\theta)\sqrt{\theta}},$$

which corresponds to a Beta$(1/2, 0)$ distribution.

2 (Textbook Problems)**.** Finish Problems 5 and 12 in Chapter 2 of the textbook.

Solution.
Problem 5:

(a) The prior predictive distribution is

$$\begin{aligned}\mathbb{P}(y = k) &= \int_0^1 \binom{n}{k} \theta^k (1-\theta)^{n-k} d\theta \\ &= \binom{n}{k} \frac{\Gamma(k+1)\Gamma(n-k+1)}{\Gamma(n+2)} \\ &= \frac{n!}{k!(n-k)!} \frac{k!(n-k)!}{(n+1)!} \\ &= \frac{1}{n+1}\end{aligned}$$

(b) The posterior of θ is Beta$(\alpha+y, \beta+n-y)$. Therefore the posterior mean of θ is $(\alpha+y)/(\alpha+\beta+n)$, which is always between $\alpha/(\alpha+\beta)$ and y/n.

(c) The variance for the uniform prio is $1/12$. For the postetior, the variance is

$$\frac{y(n-y)}{n^2(n+1)} \leq \frac{n^2/4}{n^2(n+1)} = \frac{1}{4(n+1)} \leq \frac{1}{12}.$$

The first inequality becomes equality when $n = 2y$ and the second inequality becomes equaity when $n = 1$. Therefore, we will not have two equalities simultaneously. Hence the posterior variance is strictly less than the prior's.

(d) It happens when the observations and the prior are relatively contraversal. For example, we take $\alpha = 1, \beta = 5$ and $n = y = 5$.

Problem 12:
The probability mass function is

$$p(y \mid \theta) = \frac{\theta^y e^{-\theta}}{y!}$$

Then take the second-order derivative of the log-likelihood:

$$\ell''(\theta) = -\frac{y}{\theta^2}$$

Then the Fisher's information is

$$\mathcal{I}(\theta) = -\mathbb{E}[\ell''] = \frac{1}{\theta}.$$

The corresponding Jeffreys prior is

$$p(\theta) \propto \sqrt{\mathcal{I}(\theta)} = \theta^{-1/2},$$

which is the Gamma distribution with parameters $\alpha = 1/2$ and $\beta = 0$.

3 (Normal with Precision Parameter)**.**

(a) Consider the **Normal-Gamma** distribution with density

$$p(x, \tau \mid \mu, \lambda, \alpha, \beta) = \frac{\beta^\alpha \sqrt{\lambda}}{\Gamma(\alpha)\sqrt{2\pi}} \tau^{\alpha-1/2} e^{-\beta\tau} e^{-\lambda\tau(x-\mu)^2/2}$$

Find the conditional distritions $p(x \mid \tau)$ and $p(\tau \mid x)$, and the marginal distributions $p(x)$ and $p(\tau)$.

(b) Consider a normal distribution with precision parametrization, that is,

$$p(y \mid \mu, \tau) = \frac{\sqrt{\tau}}{\sqrt{2\pi}} e^{-\frac{\tau(y-\mu)^2}{2}}$$

Use the Normal-Gamma distribution as a conjugate prior and find the posterior distribution.

(c) Find the marginal posterior distributions for μ and τ correspondingly.

Solution.

(a) From the joint distribution, we figure out

$$\begin{aligned} &p(x \mid \tau) \propto e^{-\lambda\tau(x-\mu)^2/2} \\ &p(\tau \mid x) \tau^{\alpha-1/2} e^{-\beta\tau} e^{-\lambda\tau(x-\mu)^2/2}. \end{aligned}$$

Therefore,

$$\begin{aligned} &x \mid \tau \sim \mathcal{N}(\mu, 1/(\lambda\tau)) \\ &\tau \mid x \sim \text{Gamma}\left(\alpha + 1/2, \beta + \lambda(x-\mu)^2/2\right) \end{aligned}$$

We can write down the conditional densities (up to constants):

$$\begin{aligned} &p(x \mid \tau) \propto \sqrt{\tau} e^{-\lambda\tau(x-\mu)^2/2} \\ &p(\tau \mid x) \propto \left(\beta + \frac{\lambda(x-\mu)^2}{2}\right)^{\alpha+1/2} \tau^{\alpha-1/2} e^{-\beta\tau} e^{-\lambda\tau(x-\mu)^2/2} \end{aligned}$$

Then we can get the marginal densities:

$$\begin{aligned} p(\tau) &= \frac{p(x,\tau)}{p(x \mid \tau)} \propto \tau^{\alpha-1} e^{-\beta\tau} \\ p(x) &= \frac{p(x,\tau)}{p(\tau \mid x)} \propto \left(1 + \frac{\lambda(x-\mu)^2}{2\beta}\right)^{-\alpha-1/2} \end{aligned}$$

The corresponding distributions are

$$\begin{aligned} &\tau \sim \text{Gamma}(\alpha, \beta) \\ &x \sim t_{2\alpha}(\mu, \beta/(\lambda\alpha)) \end{aligned}$$

(b) We assume the prior for (μ, τ) is Normal-Gamma with parameters $(\mu_0, \lambda, \alpha, \beta)$ The posterior distribution is

$$\begin{aligned} p(\mu, \tau \mid y) &\propto p(y \mid \mu, \tau) p(\mu, \tau) \\ &\propto \tau^\alpha e^{-\beta\tau} e^{-\lambda\tau(\mu-\mu_0)^2/2} e^{-\tau(y-x)^2/2} \\ &\propto \tau^\alpha e^{-\beta_1\tau} e^{-\lambda_1\tau(\mu-\mu_1)^2/2}, \end{aligned}$$

where

$$\beta_1 = \beta + \frac{\lambda}{2(\lambda+1)}(y-\mu_0)^2, \quad \lambda_1 = \lambda + 1, \quad \mu_1 = \frac{\lambda\mu_0 + y}{\lambda+1}$$

This is a Normal-Gamma distribution with parameters

$$\left(\frac{\lambda\mu_0 + y}{\lambda+1}, \lambda+1, \alpha + \frac{1}{2}, \beta + \frac{\lambda}{2(\lambda+1)}(y-\mu_0)^2\right)$$

(c) Using results from part (a), we have

$$\mu \mid y \sim t_{2\alpha+1}\left(\frac{\lambda\mu_0 + y}{\lambda+1}, \frac{\beta + \frac{\lambda}{2(\lambda+1)}(y-\mu_0)^2}{(\lambda+1)(\alpha+1/2)}\right)$$
$$\tau \mid y \sim \text{Gamma}\left(\alpha + \frac{1}{2}, \beta + \frac{\lambda}{2(\lambda+1)}(y-\mu_0)^2\right)$$

4 (Textbook Problems)**.** Finish Problems 10 in Chapter 3 of the textbook.

Solution.
The likelihood function for group j is

$$p(y_{j1}, \ldots, y_{jn_j} \mid \mu_j, \sigma_j^2) \propto \frac{1}{\sigma_j^2}\prod_{i=1}^{n_j} e^{-\frac{(y_{ji}-\mu_j)^2}{2\sigma_j^2}} \propto (\sigma_j^2)^{-n_j/2} e^{-\frac{n}{2\sigma_j^2}(\bar{y}_j-\mu_j)^2 - \frac{(n_j-1)s_j^2}{2\sigma_j^2}}$$

Then the posterior is given by

$$p(\mu_j, \sigma_j^2 \mid y_{j1}, \ldots, y_{jn_j}) \propto (\sigma_j^2)^{-n_j/2-1} e^{-\frac{n}{2\sigma_j^2}(\bar{y}_j-\mu_j)^2 - \frac{(n_j-1)s_j^2}{2\sigma_j^2}}$$

The marginal posterior distribution for σ_j^2 is

$$\sigma_j^2 \mid y_{j1}, \ldots, y_{jn_j} \sim \text{Inv} - \text{Gamma}((n_j-1)/2, (n_j-1)s_j^2/2)$$

Then $\sigma_j^{-2} \mid y_{j1}, \ldots, y_{jn_j} \sim \text{Gamma}((n_j-1)/2, (n_j-1)s_j^2/2)$ and $(n_j-1)s_j^2/\sigma_j^2 \mid y_{j1}, \ldots, y_{jn_j} \sim \text{Gamma}((n_j-1)/2, 1/2) \sim \chi^2_{n_j-1}$. Therefore, we have $(n_1-1)s_1^2/\sigma_1^2 \sim \chi^2_{n_1-1}$ and $(n_2-1)s_2^2/\sigma_2^2 \sim \chi^2_{n_2-1}$ independently. The result is now immediate.