Homework 2

1 Problem 1: 20pt

A researcher is trying to estimate the mean number of accidents per week within 100 feet of the Gervais Street/Assembly Street intersection in Columbia. She assumes a $Poisson(\lambda)$ model for the $number\ of\ accidents\ X\ per\ week$, so that the density function for X given λ is

$$p(x|\lambda) = \lambda^x \frac{e^{-\lambda}}{x!}, \qquad x = 0, 1, \dots, \qquad \lambda \geq 0$$

(a) She uses a standard exponential prior distribution for λ (i.e., an exponential with mean 1). Derive her posterior distribution for λ given a random sample $x_1, \dots, x_n$ from n weeks.

Answer : The prior distribution for λ is

$$p(\lambda) = e^{-\lambda}, \qquad \lambda \geq 0$$

Then the data-likelihood is computed by

$$L(\boldsymbol{x}; \lambda) = \prod_{i=1}^{n} p(x_i|\lambda) = e^{-n\lambda} \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!}$$

Therefore the posterior distribution of λ

$$p(\lambda|\boldsymbol{x}) \propto L(x;\lambda)p(\lambda) = e^{-(n+1)\lambda} * \frac{\lambda^{\sum x_i}}{\prod_{i=1}^{n} x_i!}$$

$$p(\lambda|\boldsymbol{x}) \propto e^{-(n+1)\lambda} * \lambda^{\sum x_i}$$

Since the kernel is Gamma kernel, we can derive that

$$p(\lambda|\boldsymbol{x}) \sim Gamma\left(\sum_{i=1}^{n} x_i + 1, n+1\right)$$

(b) Using your "expert knowledge" (i.e., any common-sense guess) of the accident rate around this busy intersection (and resisting the urge to use the data given in part (c)!),suggest a different /better prior for λ. Derive the corresponding posterior distribution for λ given a random sample $x_1, \dots, x_n$ from n weeks.

Answer : Since I have no expert information about λ, I prefer to use a non-informative prior. So Jeffrey's prior is a good choice.

$$p(\lambda) \propto I(\lambda)^{\frac{1}{2}}, \qquad I(\lambda) = -E_x\left[\frac{\partial^2 \log p(x|\lambda)}{\partial \lambda^2}\right]$$

$$\log p(x|\lambda) = -\lambda + x \log \lambda + \log x!$$

$$I(\lambda) = -E_x \left[-\frac{x}{\lambda^2} \right] = \frac{1}{\lambda}$$

Thus, the prior for λ is

$$p(\lambda) \propto \frac{1}{\sqrt{\lambda}}$$

Then the posterior distribution

$$p(\lambda|\boldsymbol{x}) \propto L(x;\lambda)p(\lambda) \propto e^{-n\lambda} \lambda^{\sum x_i} * \lambda^{-\frac{1}{2}}$$

Since the kernel is Gamma kernel, we can derive that

$$p(\lambda|\boldsymbol{x}) \sim Gamma\left(\sum_{i=1}^{n} x_i + \frac{1}{2}, n \right)$$

(c) If she gathers the following accident counts from 15 randomly selected weeks

1 0 4 1 4 2 5 3 0 3 1 2 2 4 1

find the posterior median and a 95% credible interval for λ using the standard exponential prior, along with these data.

Answer : Already known the posterior distribution for λ is

$$p(\lambda|\boldsymbol{x}) \sim Gamma\left(\sum_{i=1}^{n} x_i + 1, 16 \right)$$

which, using current data, is

$$p(\lambda|\boldsymbol{x}) \sim Gamma(34, 16)$$

Thus the posterior median

$$\text{median}(\lambda|\boldsymbol{x}) = qgamma(0.5, 34, 16) = 2.104$$

And 95% credible interval is

$$\lambda|\boldsymbol{x} \in [qgamma(0.025, 34,16), qgamma(0.975,34,16)]$$

$$\lambda|\boldsymbol{x} \in [\, 1.472, 2.897 \,]$$

(d) Give the posterior median and a 95% credible interval for λ using your own prior,along with the data in part (c).

Answer : Already known the posterior distribution for λ is

$$p(\lambda|\boldsymbol{x}) \sim Gamma\left(\sum_{i=1}^{n} x_i + \frac{1}{2}, n \right)$$

which, using current data, is

$$p(\lambda|\boldsymbol{x}) \sim Gamma(33.5, 15)$$

Thus the posterior median

$$\text{median}(\lambda|\boldsymbol{x}) = qgamma(0.5, 33.5, 15) = 2.2111$$

And 95% credible interval is

$$\lambda|\boldsymbol{x} \in [\,1.542\,, 3.051]$$

---

2 Problem 2: 45pt

The eBay selling prices for auctioned Palm M515 PDAs are assumed to follow $a\ normal\ distribution\ with\ \mu\ and\ \sigma^2\ unknown$. We wish to perform inference on the mean selling price μ.

(a) Suppose we assume an $IG(1100, 250000)$ prior for σ^2 and let the prior for $\mu|\sigma^2$ be

$$p(\mu|\sigma^2) \propto (\sigma^2)^{-\frac{1}{2}} \exp -\frac{1}{2\sigma^2/s_0}(\mu-\delta)^2$$

with $s_0 = 1$ and $\delta = 220$. If our sample data are:

$$(212, 249, 250, 240, 210, 234, 195, 199, 222, 213, 233, 251)$$

, then find a point estimate and 95% credible interval for μ. (Note, you can use either the conditional posterior or the marginal posterior of μ to obtain the interval.)

Answer : The prior distribution for σ^2 is

$$\sigma^2 \sim IG(\alpha,\beta), \qquad \alpha = 1100, \beta = 2.5 * 10^5$$

The prior for $\mu|\sigma^2$ is

$$p(\mu|\sigma^2) \propto (\sigma^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}(\mu-220)^2\right\}$$

The data-likelihood is

$$L(\boldsymbol{x};\mu,\sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x_i-\mu)^2\right\}$$

Bayesian point estimate is produced by summarizing the center of the posterior. Typically, we use the mean or mode of the posterior distribution, i.e.

$$\bar{\theta} = \int \theta * p(\theta|\boldsymbol{Data})\, d\theta$$

So we need to find the posterior distribution first.

The joint posterior for μ, σ^2 is

$$p(\mu,\sigma^2|\boldsymbol{x}) \propto L(\boldsymbol{x};\mu,\sigma^2) * p(\mu|\sigma^2) * p(\sigma^2)$$

$$\propto (\sigma^2)^{-\frac{n}{2}-\frac{3}{2}-\alpha} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 - \frac{1}{2\sigma^2}(\mu-\delta)^2 - \frac{\beta}{\sigma^2}\right\}$$

which is difficult to calculate by hand. So we need the MCMC process to help us.

First, we need to derive the full conditionals:

$$p(\mu|\boldsymbol{x},\sigma^2) \propto p(\boldsymbol{x}|\mu,\sigma^2) * p(\mu|\sigma^2)$$

$$\propto (\sigma^2)^{-\frac{n+1}{2}} \exp\left\{-\frac{1}{2\sigma^2}\left[(\mu-\delta)^2 + \sum_{i=1}^{n}(x_i-\mu)^2\right]\right\}$$

Thus

$$\mu|\boldsymbol{x},\sigma^2 \sim Normal\left(\frac{\delta+\sum x_i}{n+1}, \frac{\sigma^2}{n+1}\right), \qquad \delta = 220$$

Similarly,

$$p(\sigma^2|\boldsymbol{x},\mu) \propto p(\boldsymbol{x}|\mu,\sigma^2) * p(\mu|\sigma^2) * p(\sigma^2)$$

$$\propto (\sigma^2)^{-\frac{n}{2}-\frac{3}{2}-\alpha} \exp\left\{-\frac{1}{2\sigma^2}\left[2\beta + (\mu-\delta)^2 + \sum_{i=1}^{n}(x_i-\mu)^2\right]\right\}$$

Thus

$$\sigma^2|\boldsymbol{x},\mu \sim IG\left(\alpha + \frac{n+1}{2}, \beta + \frac{(\mu-\delta)^2}{2} + \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right)$$

where $\alpha = 1100,\ \beta = 2.5 * 10^5, \delta = 220$

Then we can run an gibbs sampling program to get the posterior distribution of μ, σ^2. Please see attached file "**question2_1.R**" for codes.

- Point Estimate for μ

$$\hat{\mu} = 225.2115$$

- The 95% credible interval

$$\mu|x,\sigma^2 \ \in [216.9408\,, 233.4332\,]$$

(b) Suppose now that we had assumned the independent improper priors

$$p(\mu) = 1, \qquad -\infty < \mu < \infty$$

$$p(\sigma) = \frac{1}{\sigma}, \qquad 0 < \sigma < \infty$$

Using the same data as in part (a), find a point estimate and 95% credible interval for μ.

Answer: The posterior distribution

$$p(\mu,\sigma^2|\boldsymbol{x}) \propto L(\boldsymbol{x};\mu,\sigma^2) * p(\mu) * p(\sigma^2)$$

$$p(\mu,\sigma^2|\boldsymbol{x}) \propto (\sigma^2)^{-\frac{n+1}{2}} * \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

is difficult to do calculus. So like question(a), we will also follow a MCMC process.

First we need to derive the full conditionals:

$$p(\mu|\boldsymbol{x},\sigma^2) \propto p(\boldsymbol{x}|\mu,\sigma^2) * p(\mu) \propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

Then we have

$$\mu|x,\sigma^2 \propto Normal\left(\frac{\sum x_i}{n}, \frac{\sigma^2}{n}\right)$$

Similarly, we can cmpute the full conditionals for σ^2

$$p(\sigma^2|\boldsymbol{x},\mu) \propto p(\boldsymbol{x}|\mu,\sigma^2) * p(\sigma^2)$$

where we need to do variable transformation

$$p(\sigma^2) = p(\sigma) * \left|\frac{\partial\sigma^2}{\partial\sigma}\right|^{-1} = \frac{1}{2\sigma^2}$$

So we can have

$$p(\sigma^2|\boldsymbol{x},\mu) \propto (\sigma^2)^{-\frac{n}{2}-1} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

Thus the full conditional is

$$\sigma^2|\boldsymbol{x},\mu \sim IG\left(\frac{n}{2}, \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right)$$

Then we can run an gibbs sampling program to get the posterior distribution of μ,σ^2. Please see attached file "**question2_2.R**" for codes.

- Point Estimate for μ

$$\hat{\mu} = 225.6525$$

- The 95% credible interval

$$\mu|x,\sigma^2 \in [212.9274, 238.2097]$$

(c) How do your substantive conclusions in parts (a) and (b) differ, and how is this related to the different choices of priors?

Answer: Compared with the outcome in (a), the point estimate is nearly the same in (b). However, the 95% credible interval differs. Part (b) get a wider credible interval. This is probably due to the non-informative priors used in (b) as a wider credible interval implicates more uncertainty.

(d) Now suppose (perhaps unrealistically) that we had known the true population variance was $\sigma^2 = 228$. Assuming a conjugate prior for μ with $\delta = 220$ and $\tau^2 = 25$, find a point estimate and 95% credible interval for the single unknown parameter μ.

Answer: The conjugate prior for μ is

$$p(\mu) = \frac{1}{\sqrt{2\pi}\,\tau}\exp\left\{-\frac{1}{2\tau^2}(\mu-\delta)^2\right\}$$

The data-likelihood is

$$L(\boldsymbol{x};\mu) = \prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2\sigma^2}(x_i-\mu)^2\right\}$$

so the posterior distribution is

$$p(\mu|\boldsymbol{x}) \propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 - \frac{1}{2\tau^2}(\mu-\delta)^2\right\}$$

$$\mu|\boldsymbol{x} \sim Normal\left(\frac{\tau^2\sum x_i + \delta\sigma^2}{n\tau^2+\sigma^2}, \frac{\tau^2\sigma^2}{n\tau^2+\sigma^2}\right)$$

Put in the specific data, we have

$$\mu|\boldsymbol{x} \sim Normal(223.220\,, 10.795)$$

So a point estimate is

$$\hat{\mu}|\boldsymbol{x} = E[\mu|\boldsymbol{x}] = 223.2197$$

The 95% credible interval is

$$\mu|\boldsymbol{x} \in [216.78\,, 229.659]$$

(e) How (if at all) does the inference in part (d) differ from the inferences in parts (a) and (b)? Explain your answer intuitively.

Answer: The point estimate in part (d) is smaller than point estimates in part (a) and (b), because the prior mean for μ is 220. And the 95% credible interval is narrower, since in part (d) we have a known σ^2, means that in part (d) we have less uncertainty compared with (a) and (b). So a narrower interval is reasonable.

---

3 Problem 3: 21 pt

Let $X_1, \ldots, X_n$ be i.i.d. data from a normal distribution with known mean 0 and unknown variance θ.

$$p(x|\theta) = \frac{1}{\sqrt{2\pi\theta}}\exp\left\{-\frac{1}{2\theta}x^2\right\}$$

(a) Write the likelihood L(θ|x).

Answer: The data-likelihood is

$$L(\theta;x) = \prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\theta}}\exp\left\{-\frac{1}{2\theta}x_i^2\right\} = \left(\frac{1}{\sqrt{2\pi\theta}}\right)^n \exp\left\{-\frac{1}{2\theta}\sum_{i=1}^{n}x_i^2\right\}$$

(b) Derive the Jeffreys prior for θ.

Answer: Definition of Jeffreys prior:

$$p(\theta) \propto I(\theta)^{\frac{1}{2}}, \qquad I(\theta) = -E_x\left[\frac{\partial^2 \log p(x|\theta)}{\partial\theta^2}\right]$$

where

$$\log p(x|\theta) = -\frac{1}{2\theta}x^2 - \frac{1}{2}\log\theta - \frac{1}{2}\log 2\pi$$

$$I(\theta) = -E_x\left[-\frac{1}{\theta^3}x^2 + \frac{1}{2\theta^2}\right] = \frac{1}{\theta^3}E_x[x^2] - \frac{1}{2\theta^2} = \frac{1}{2\theta^2}$$

So the jefferys prior is

$$p(\theta) \propto \frac{1}{\theta}$$

(c) Suppose we observe the 6 data values $x_1 = 2.75$, $x_2 = 1.78$, $x_3 = 0.36$, $x_4 = -1.64$, $x_5 = 0.17$, $x_6 = -2.03$. Write the posterior distribution, using your Jeffreys prior from part (b). Do you recognize the form of this posterior? Specify exactly what distribution it is, including the parameter values.

Answer: The posterior distribution is

$$p(\theta|\boldsymbol{x}) \propto L(\boldsymbol{x};\theta)p(\theta) \propto \theta^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\theta}\sum_{i=1}^{n} x_i^2\right\} * \theta^{-1}$$

Thus we can identify the posterior dist. of θ is

$$\theta|\boldsymbol{x} \sim IGamma\left(\frac{n}{2}, \frac{1}{2}\sum_{i=1}^{n} x_i^2\right)$$

Fit in the data, we have:

$$\theta|\boldsymbol{x} \sim IGamma(3,\ 8.850)$$

---

4 Problem 4: 14 pt

Two Bayesian statisticians, Barry and Brianna, are trying to estimate θ, the *mean survival time* for a population of terminally ill patients who have undergone a certain procedure meant to slow the spread of their disease. They consult with a medical expert, whose best guess of the most likely mean survival time is 400 days. The expert also believes there is a 2/3 chance that the mean survival time is between 315 and 485 days.

(a) Barry wishes to use *a normal prior* for θ. Based on the expert opinion, what parameters would be good choices for the parameters of his prior? Explain your reasoning clearly.

Answer: Suppose the prior for θ is

$$p(\theta) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2\sigma^2}(\theta-\mu)^2\right\}, \qquad Normal(\mu,\sigma^2)$$

From expert knowledge, we know the most likely mean survival time is 400 days, so

$$\mu = 400$$

Also, there is a 2/3 chance that the mean survival time is between 315 and 485 days, which means a (1-1/3) confidence interval for this data is [315, 485]. So

$$400 + z_{\frac{5}{6}} * \sigma = 485 \ \rightarrow \ \sigma = \frac{85}{0.96742} = 87.862$$

So $\mu = 400, \sigma = 87.862$ may be good choice since it fits the expert knowledge well.

(b) Brianna wishes to use a gamma prior for θ. Based on the expert opinion, what parameters would be good choices for the parameters of her prior? Explain your reasoning clearly.

Answer: Suppose the prior for θ is

$$\theta \sim Gamma(\alpha, \beta)$$

From expert knowledge, we know the most likely mean survival time is 400 days, so

$$E[\theta] = \frac{\alpha}{\beta} = 400$$

Also, there is a 2/3 chance that the mean survival time is between 315 and 485 days, which means a (1-1/3) confidence interval for this data is [315, 485]. So

$$qgamma\left(\frac{1}{6}, \alpha, \beta\right) \approx 315, \qquad qgamma\left(\frac{5}{6}, \alpha, \beta\right) \approx 485$$

Now we have three equations, implies there may me multiple choices. So a R program (grid search) is deployed to find those “good choice” parameters.

Finally we get

$$\alpha = 20.40, \qquad \beta = 0.051$$

with confidence interval

$$[314.62, 484.56]$$
