# STAT 5370 – Decision Theory

Homework 2

*Matthew Aaron Looney*

*6/22/2018*

# Contents

(b) Your initial bet will be 10 dollars. The return on any bet will be the amount that you bet; e.g., if you bet 10 dollars and win, then you win 10 dollars. You are going to play exactly 10 games of craps. If you lose on the previous game, your next bet requires you to "double down," that is, if on one game you bet X dollars and lose, on the next game you have to bet 2X dollars. If you win on the previous game, your next bet will be 10 dollars. . . . . . . . . . . . . . . . 8

(c) Write a Monte Carol simulation to find your expected winnings after the 10 games. Also examine the distribution of your expected winnings. Does it appear that this is a good betting strategy? . . . . . . . . . . . . . . . . . . . . . . . . . . . 8

# Problem 1

In this problem we will explore an alternate class of priors.

**(a) Find a resource and write a basic description of what a Jeffrey's prior is. Be thorough, your exposition here needs to be such that someone who knows nothing of what a Jeffrey's prior is could learn the salient features. Identify why people like Jeffery's priors and discuss drawbacks.**

**(b) Derive the Jeffery's prior for the Bernoulli model (i.e., assuming $Y_i \overset{iid}{=} Bernoulli(p)$), and find the posterior distribution of $p$ under this prior. Is it a known distribution? Discuss how you would proceed to conduct posterior inference.**

Assuming $Y_i \overset{iid}{=} Bernoulli(p)$, then we can say that $Y|p \sim Binomial(n,p)$ and we can obtain the likelihood function as:

$$L(p|Y) = \prod_{i=1}^{n} p^{y_i}(1-p)^{1-y_i} = p^{\sum_{i=1}^{n} y_i}(1-p)^{\sum_{i=1}^{n}(1-y_i)} = p^{\sum_{i=1}^{n} y_i}(1-p)^{n-\sum_{i=1}^{n} y_i}$$

Letting $k = \sum_{i=1}^{n} y_i$ we obtain the log likelihood as:

$$\log L(p|Y) = k\log(p) + (n-k)\log(1-p)$$

The next step in the derivation of the Jeffreys Prior is to calculate the Fisher information matrix. We are told the data are iid and we can see that the log likelihood function is twice differentiable which allows us to use an alternative form of the Fisher information matrix. Further, calculation is vastly simplified since our likelihood function is low dimensional (single parameter). The Fisher information is given by the following:

$$\mathbf{I_{j,k}}(\boldsymbol{\theta}) = -\mathbb{E}_{\mathbf{y}|\theta}\left[\frac{\partial^2 \log L(\mathbf{y}|\boldsymbol{\theta})}{\partial \theta_j \partial \theta_k}\right] \tag{General}$$

$$\mathbf{I}(p) = -\left[\frac{-\mathbb{E}[k]}{p^2} - \frac{(n-\mathbb{E}[k])}{(1-p)^2}\right] \tag{1}$$

Recalling that $Y|p \sim Binomial(n,p)$ and we defined $k$ as $\sum_{i=1}^{n} y_i$, then $\mathbb{E}[k] = np$ and we obtain the following results:

$$\mathbf{I}(p) = \frac{n}{p(1-p)}$$

The Jefferys prior is defined as:

$$\pi(\theta) \propto \sqrt{det|\mathbf{I}(\theta)|} \tag{General}$$

$$\pi(p) \propto \sqrt{det|\mathbf{I}(\mathbf{p})|} \propto p^{-0.5}(1-p)^{-0.5} \tag{2}$$

We recognize that equation 2 can be made to look like a Beta distribution by subtracting one from each of the powers to obtain $\pi(p) \propto p^{0.5-1}(1-p)^{0.5-1}$ and we now have the Jeffrys prior which is $Beta(\frac{1}{2}, \frac{1}{2})$.

With the non-informative Jefferys prior in hand we can proceed to build the posterior distribution in the usual way which gives:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\int_A p(y|\theta)p(\theta)d\theta} \propto \underbrace{p(y|\theta)}_{\text{L}(\theta)}\underbrace{p(\theta)}_{\text{Prior}} \qquad \text{(General)}$$

$$\begin{aligned}
\pi(p|y) &= p^{\sum_{i=1}^{n} y_i}(1-p)^{n-\sum_{i=1}^{n} y_i} \cdot p^{0.5-1}(1-p)^{0.5-1} \\
&= p^{\sum_{i=1}^{n} y_i+0.5-1}(1-p)^{n-\sum_{i=1}^{n} y_i+0.5-1}
\end{aligned} \qquad (3)$$

If we let $\sum_{i=1}^{n} y_i = n\bar{Y}$ then the posterior distribution is $Beta(\overbrace{n\bar{Y}+0.5}^{\alpha}, \overbrace{n-n\bar{Y}+0.5}^{\beta})$, where $\alpha$ and $\beta$ are the parameters of the Beta distribution and $\alpha > 0$ (first shape parameter), $\beta > 0$ (second shape parameter).

The Beta distribution is well known with all of the most relevant moments documented. We can obtain the posterior mean and variance by applying the following formulas and making the appropriate substitutions of the posterior distributions parameters:

$$\begin{aligned}
\mathbb{E}[p|y] &= \frac{\alpha}{\alpha + \beta} = \hat{\mu} \\
Var(p|y) &= \frac{\hat{\mu}(1 - \hat{\mu})}{\alpha + \beta + 1}
\end{aligned} \qquad (4)$$

We can also calculate the Equal Tail Credible Intervals and the HPD Credible Intervals to determine the 95% probability that $p$ falls within some upper and lower bound.

**(c) Derive the Jeffery's prior for the Poisson model (i.e., assuming $Y_i \overset{iid}{=} Poisson(\lambda)$), and find the posterior distribution of $\lambda$ under this prior. Is it a known distribution? Discuss how you would proceed to conduct posterior inference.**

Assuming $Y_i \overset{iid}{=} Poisson(\lambda)$, then we can say that $Y|\lambda \sim Poisson(\lambda)$ and we can obtain the likelihood function as:

$$L(\lambda|y) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^y}{y!} = \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} y_i}}{\prod_{i=1}^{n} y!} \propto e^{-n\lambda}\lambda^{\sum_{i=1}^{n} y_i}$$

Letting $k = \sum_{i=1}^{n} y_i$ we obtain the log likelihood as:

$$\log L(\lambda|Y) = -n\lambda + k\log(\lambda)$$

4

The next step in the derivation of the Jeffreys Prior is to calculate the Fisher information matrix. We are told the data are iid and we can see that the log likelihood function is twice differentiable which allows us to use an alternative form of the Fisher information matrix. Further, calculation is vastly simplified since our likelihood function is low dimensional (single parameter). The Fisher information is given by the following:

$$\mathbf{I_{j,k}}(\boldsymbol{\theta}) = -\mathbb{E}_{\mathbf{y}|\theta}\left[\frac{\partial^2 \log L(\mathbf{y}|\boldsymbol{\theta})}{\partial\theta_j\partial\theta_k}\right] \qquad \text{(General)}$$

$$\mathbf{I}(\lambda) = -\frac{\mathbb{E}[\mathbf{k}]}{\lambda^{\mathbf{2}}} \qquad (5)$$

Recalling that $Y|\lambda \sim Poisson(\lambda)$ and we defined $k$ as $\sum_{i=1}^{n} y_i = n\bar{Y}$, then $n\mathbb{E}[\bar{Y}] = n\lambda$ and we obtain the following results:

$$\mathbf{I}(\lambda) = \frac{\mathbf{n}}{\lambda}$$

The Jefferys prior is defined as:

$$\pi(\theta) \propto \sqrt{det|\mathbf{I}(\theta)|} \qquad \text{(General)}$$

$$\pi(\lambda) \propto \sqrt{det|\mathbf{I}(\lambda)|} \propto \lambda^{-0.5} \qquad (6)$$

We recognize that equation 6 can be made to look like a Gamma distribution by subtracting one from the power on $\lambda$ to obtain $\pi(\lambda) \propto \lambda^{0.5-1}e^0$ and we now have the Jeffrys prior which is $Gamma(\frac{1}{2}, 0)$.

With the non-informative Jefferys prior in hand we can proceed to build the posterior distribution in the usual way which gives:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\int_A p(y|\theta)p(\theta)d\theta} \propto \underbrace{p(y|\theta)}_{\text{L}(\theta)}\underbrace{p(\theta)}_{\text{Prior}} \qquad \text{(General)}$$

$$\begin{aligned}\pi(\lambda|y) &= e^{-n\lambda}\lambda^{\sum_{i=1}^{n} y_i} \cdot e^0\lambda^{0.5-1} \\ &= \lambda^{\sum_{i=1}^{n} y_i+0.5-1}e^{-n\lambda}\end{aligned} \qquad (7)$$

If we let $\sum_{i=1}^{n} y_i = n\bar{Y}$ then the posterior distribution is $Gamma(\overbrace{n\bar{Y}+0.5}^{\alpha}, \overbrace{n}^{\beta})$, where $\alpha$ and $\beta$ are the parameters of the Gamma distribution and $\alpha > 0$ (shape parameter), $\beta > 0$ (rate parameter).

The Gamma distribution is well known with all of the most relevant moments documented. We can obtain the posterior mean and variance by applying the following formulas and making the appropriate substitutions of the posterior distributions parameters:

$$\begin{aligned}\mathbb{E}[\lambda|y] &= \frac{\alpha}{\beta} \\ Var(\lambda|y) &= \frac{\alpha}{\beta^2}\end{aligned} \qquad (8)$$

We can also calculate the Equal Tail Credible Intervals and the HPD Credible Intervals to determine the 95% probability that $\lambda$ falls within some upper and lower bound.

**(d) Derive the Jeffery's prior for the exponential model (i.e., assuming $Y_i \overset{iid}{=} Exponential(\beta)$), and find the posterior distribution of $\beta$ under this prior. Is it a known distribution? Discuss how you would proceed to conduct posterior inference.**

Assuming $Y_i \overset{iid}{=} Exponential(\beta)$, then we can say that $Y|\beta \sim Exponential(\beta)$ and we can obtain the likelihood function as:

$$L(\beta|y) = \prod_{i=1}^{n} \beta e^{-\beta y_i} = \beta^n e^{-\beta \sum_{i=1}^{n} y_i}$$

Letting $k = \sum_{i=1}^{n} y_i$ we obtain the log likelihood as:

$$\log L(\beta|Y) = n \log(\beta) - \beta k$$

The next step in the derivation of the Jeffreys Prior is to calculate the Fisher information matrix. We are told the data are iid and we can see that the log likelihood function is twice differentiable which allows us to use an alternative form of the Fisher information matrix. Further, calculation is vastly simplified since our likelihood function is low dimensional (single parameter). The Fisher information is given by the following:

$$\mathbf{I_{j,k}}(\boldsymbol{\theta}) = -\mathbb{E}_{\mathbf{y}|\theta} \left[ \frac{\partial^2 \log L(\mathbf{y}|\boldsymbol{\theta})}{\partial \theta_j \partial \theta_k} \right] \tag{General}$$

$$\mathbf{I}(\beta) = \frac{n}{\beta^2} \tag{9}$$

The Jefferys prior is defined as:

$$\pi(\boldsymbol{\theta}) \propto \sqrt{det|\mathbf{I}(\theta)|} \tag{General}$$

$$\pi(\beta) \propto \sqrt{det|\mathbf{I}(\beta)|} \propto \beta^{-1} \tag{10}$$

# Problem 2

Let $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ such that $X_1$ and $X_2$ are independent. Define $Y = X_1 + X_2$. The goal of this problem is to determine the distribution of $Y$; i.e., the sum of two independent normal random variables.

**(a) Obviously, this is trivial, so simply state the distribution of $Y$.**

**(b) Approach 1: Consider using Monte Carlo sampling to obtain a histogram and kernel density estimate (see the code that I have provided) of the pdf of $Y$ by directly sampling both $X_1$ and $X_2$. Over plot the true density of $Y$ and comment. Note, you should make use of large enough Monte Carlo sample that your results are reasonable.**

**(c) Approach 2: Note, the distribution of $Y$ can also be obtained through the convolution of the probability distributions of $X_1$ and $X_2$. Sketch out theoretically how this would be done. Based on this idea, create a Monte Carlo sampling technique which can be used to approximate the pdf of $Y$ evaluated at any point in the support. Use this function and add the approximation based on this technique to the Figure described in the part (b) above.**

**(d) Repeat the two approaches described in (b) and (c) above for $X_1 \sim Gamma(\alpha_1, \beta_1)$ and $X_2 \sim Gamma(\alpha_2, \beta_2)$. Note, unless $\beta_1 = \beta_2$, the resulting distribution of $Y$ is not friendly. Also be cautious of the support.**

# Problem 3

Game of chance, the game of craps and a betting strategy are going to be explored in this Monte Carlo simulation experiment.

**(a) Do a search and outline the rules of craps.**

**(b) Your initial bet will be 10 dollars. The return on any bet will be the amount that you bet; e.g., if you bet 10 dollars and win, then you win 10 dollars. You are going to play exactly 10 games of craps. If you lose on the previous game, your next bet requires you to "double down," that is, if on one game you bet X dollars and lose, on the next game you have to bet 2X dollars. If you win on the previous game, your next bet will be 10 dollars.**

**(c) Write a Monte Carol simulation to find your expected winnings after the 10 games. Also examine the distribution of your expected winnings. Does it appear that this is a good betting strategy?**