STAT 5370 – Decision Theory

Homework 2

Matthew Aaron Looney

6/25/2018

Contents

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.

The general idea behind the Jeffrys prior is to find a prior that is non-informative. In other words, we as the researcher should not impose too much prior belief onto the model. We want the data to be allowed to speak for itself. Further, we may be in a situation where there is not much information available about the population under study and we need a way to develop a prior despite this lack of initial knowledge.

Sir Harold Jeffreys developed a method to obtain a non-informative prior based on a one-to-one transformation of the distributions parameters. This is sometimes refereed to as the Jeffreys' invariance principle because the parameters are invariant to affine transformations.

The Jeffreys prior is derived by the following steps:

- Based on the proposed DGP, identify a distribution for the data.
- Derive a likelihood function based on the data distribution
- Ensure the likelihood function is twice differentiable
- Calculate the Fisher Information Matrix
  - alternative forms of the Fisher information matrix may be employed depending on the nature of the independence of the data, ie. independent or iid. If the distribution is single parameter the calculation is vastly simplified.
- The Jeffreys prior is given by the following formula: $\pi(\theta) \propto \sqrt{det|\mathbf{I}(\theta)|}$
  - where $\mathbf{I}(\theta)$ is the Fisher information matrix, and $det|\cdot|$ is the determinant if the information matrix.

One of the benefits of using the Jeffreys prior is that it brings very little bias into the model. Using the Jeffreys prior reduces some of the criticism that the researcher is driving the results based on the prior distribution. As mentioned earlier, there may be little information available about the population model under study and having access to a non-informative prior can help. The Jeffreys prior can sometimes be very difficult or impossible to calculate depending on the number of parameters in the model and the nature of the independence of the data. In high dimensional space the Jeffreys prior can be challenging to obtain, and if obtainable can sometimes yield poor results.

(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:

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

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

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 α and β 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:

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

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 λ 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:

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

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

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

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:

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

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

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:

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

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

We recognize that equation 6 can be made to look like a Gamma distribution by subtracting one from the power on λ 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:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\int\limits_{A} p(y|\theta)p(\theta)d\theta} \propto \underbrace{p(y|\theta)}_{\text{L}(\theta)}\underbrace{p(\theta)}_{\text{Prior}} \tag{General}$$

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

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 α and β 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:

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

We can also calculate the Equal Tail Credible Intervals and the HPD Credible Intervals to determine the 95% probability that λ 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 β 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:

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

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

We recognize that equation 10 can be made to look like a Gamma distribution by the following $\beta^{0-1}e^{0}$ and we now have the Jeffrys prior which is $Gamma(0, 0)$. Kerman (2011) has argued that a more proper way to represent improper priors is by representing this form of Jeffreys prior as $Gamma(\varepsilon, \varepsilon)$ where $\varepsilon \approx 0$. The rationale being that as $\varepsilon \to 0$, the distribution appears flatter and flatter over the positive real axis. Either way the prior is still improper but is now defined. The posterior distribution is, however, proper and we can proceed.

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

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

$$\begin{aligned} p(\beta|y) &= \beta^n e^{-\beta\sum_{i=1}^{n} y_i} \cdot \beta^{0-1}e^{0} \\ &= \beta^{n-1}e^{-\beta\sum_{i=1}^{n} y_i} \end{aligned} \tag{11}$$

If we let $\sum_{i=1}^{n} y_i = n\bar{Y}$ then the posterior distribution is $Gamma(\overbrace{n}^{\alpha}, \overbrace{n\bar{Y}}^{\beta})$, where α and β 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:

$$\mathbb{E}[\beta|y] = \frac{\alpha}{\beta}$$
$$Var(\beta|y) = \frac{\alpha}{\beta^2} \tag{12}$$

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

Problem 2

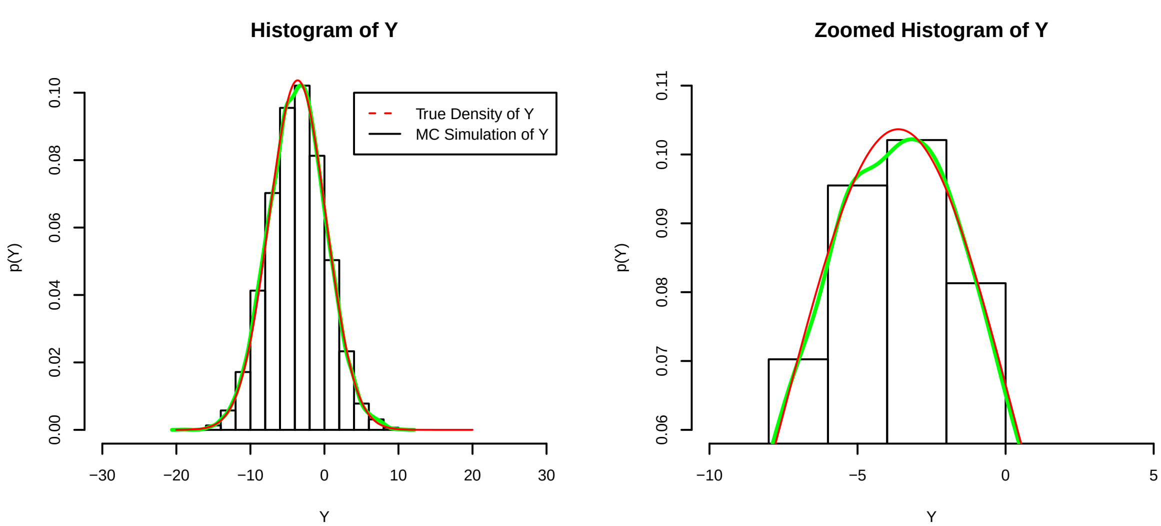
Let $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ such that X_1 and X_2 are independent. Define $Y = X_1 + X_2$. The goal of this problem is to determine the distribution of Y; i.e., the sum of two independent normal random variables.

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

The sum of two independent normally distributed random variables is normal.

$$Y = X_1 + X_2 = Y \sim \mathcal{N}(\mu_1 + \mu_2,\, \sigma_1^2 + \sigma_2^2)\,.$$

(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.



The above plots show the true density of Y (red) and the Monte Carlo simulation results (black). The histogram plot shows the probability distribution of Y. It is plain to see that the true density of Y and the Monte Carlo simulation results produce nearly identical results. The table below summarizes the results from this experiment. The last column of the Table 1 shows the difference between the

	Initial Values	True Values	MC Simulation (10,000 iter)	$\Delta_{(\mu,\sigma^2)}$
μ_1	-2.3449134	.	.	.
μ_2	-1.278761	.	.	.
σ_1^2	5.7285336	.	.	.
σ_1^2	9.0820779	.	.	.
$\mu = \mu_1 + \mu_2$	.	-3.6236744	.	.
$\sigma^2 = \sigma_1^2 + \sigma_2^2$	.	14.8106115	.	.
$mc.\mu$	.	.	-3.6515816	.
$mc.\sigma^2$	.	.	14.8546141	.
$\Delta_\mu = \mu - mc.\mu$	.	.	.	0.0279073
$\Delta_{\sigma^2} = \sigma^2 - mc.\sigma^2$	.	.	.	-0.0440025

Table 1: Summary of Results

true values of Y and the Monte Carlo simulation. This indicates that the Monte Carlo simulation technique works very well in this case.

(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.

I have leaned very heavily on several resources to learn the concept of convolutions as it relates to the sums of two continuous random variables. (Grinstead and Snell (1997), Robert and Casella (2010), G. Casella and Berger (2002))

Let X_1 and X_2 be two continuous random variables with density functions $f(x)$ and $g(x)$, respectively. Assume that both $f(x_1)$ and $g(x_2)$ are defined for all real numbers, x_1 and $x_2 \in \mathbb{R}$. Then the convolution $f * g$ of f and g is the function given by

$$\begin{aligned} f_Y(y) = (f * g)(y) &= \int_{-\infty}^{\infty} f(y - x_2) g(x_2) dx_2 \\ &= \int_{-\infty}^{\infty} g(y - x_1) f(x_1) dx_1 \end{aligned} \tag{13}$$

There are a few potential problems with equation 13. It is possible the integral does not have a closed form solution or the integral may be so intractable that an analytical solution is not practical. The good news is that we can use Monte Carlo simulation techniques to over come these problems.

A closer examination of the integral in equation 13 reveals that the form is nothing other than the formal definition of the expected value. If we apply the classical Monte Carlo integration techniques then we arrive at the following general equations:

$$E_f[h(X)] = \int_{\chi} h(x) f(x) dx$$

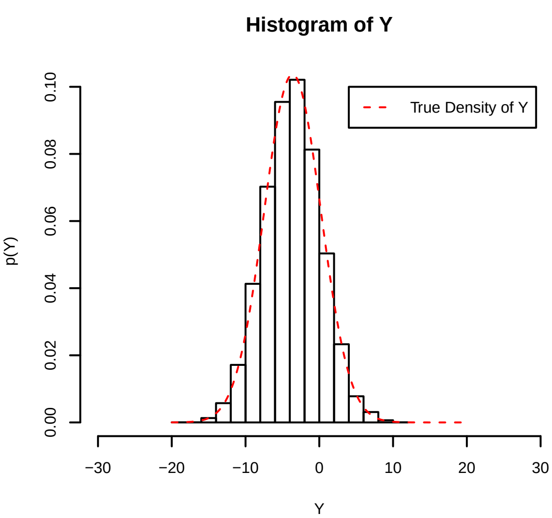
where, χ is the set where the random variable X takes its value, then χ is equal to the support of the density f.

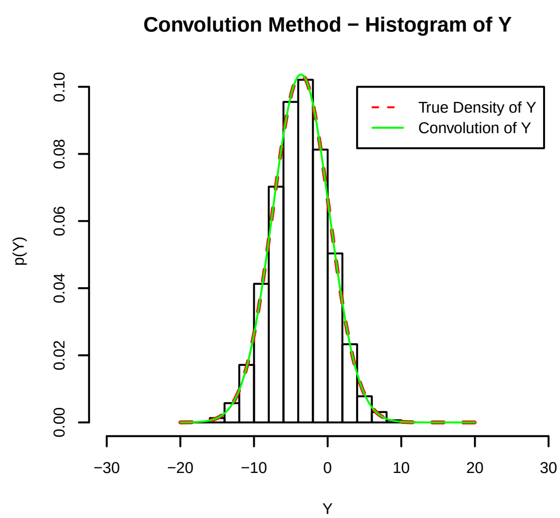
After we integrate out $f(x)$ we are left with the following equation which can be directly converted into a Monte Carlo integration algorithm:

$$\overline{h_n} = \frac{1}{n}\sum_{i=1}^{n} h(x_i) \tag{General}$$

$$f_Y(y) = \frac{1}{n}\sum_{i=1}^{n} f(y - x_2)dx_2 \tag{14}$$

The below plots show the results from the convolution method. The initial values for the distribution parameters are the same values listed in Table 1. We can see in the second plot that the true density (red) and the Monte Carlo simulation of the convolution (green) are all but identical. This indicates that the Monte Carlo method is highly successful at uncovering the true density.



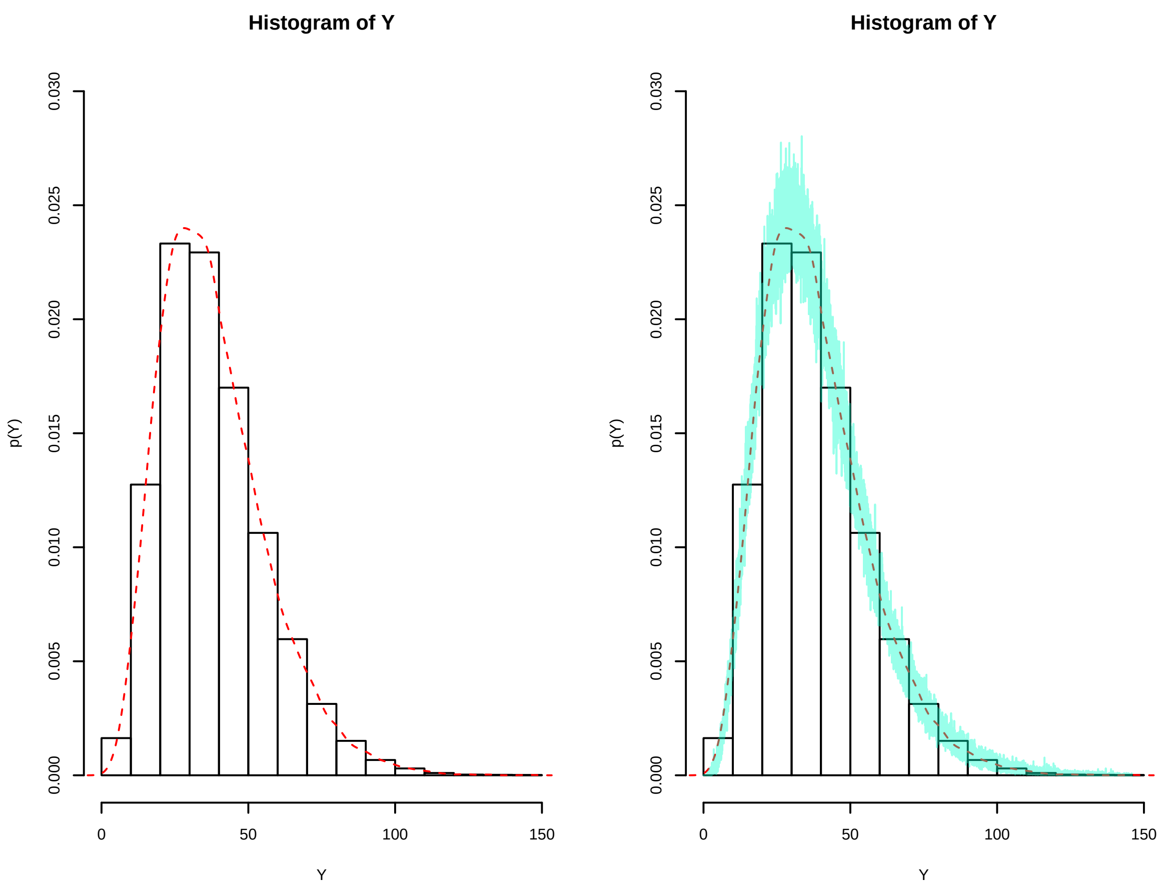


(d) Repeat the two approaches described in (b) and (c) above for $X_1 \sim Gamma(\alpha_1, \beta_1)$ and $X_2 \sim Gamma(\alpha_2, \beta_2)$. Note, unless $\beta_1 = \beta_2$, the resulting distribution of Y is not friendly. Also be cautious of the support.

The results from this experiment are similar to what we have seen in parts b and c of Problem 1. The method used in part b can be viewed in the first plot below. The convolution method of part c is the second plot. The initial values for the Gamma distribution can be seen in Table 2 below. I assigned the parameters for the Gamma distributions randomly using a uniform distribution which gave unusual values for the Gamma parameters. Even with these unusual parameter values the Monte Carlo simulation did a good job at uncovering the true distribution. The last column of Table 2 shows the difference in the mean values obtained by method b and method c. With more standard initial values of the Gamma distribution the difference would be smaller. This problem demonstrates how we can uncover a distribution that may not have a simple analytical solution (the sum of two Gamma distributions where $\beta_1 \neq \beta_2$).

	Initial Values	True Values	MC Simulation (10,000 iter)	$\Delta_{(\mu)}$
α_1	0.7639552	.	.	.
α_2	4.2768898	.	.	.
β_1	1.0914029	.	.	.
β_2	0.1140821	.	.	.
$\mu = \mu_1 + \mu_2$	.	38.0554245	.	.
$mc.\mu$	.	.	33.45	.
$\Delta_\mu = \mu - mc.\mu$	.	.	.	4.6054245

Table 2: Summary of Results



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.

The basic idea of the game of craps is summarized as follows:

- Roll two - six sided dice.
- If the sum of two rolls is in the set {7, 11}, then win.
- If the sum of two rolls is in the set {2, 3, 12}, then lose.
- Otherwise, the sum of the roll is the players "point". The player rolls the two dice repeatedly until either, the player rolls her "point" number or a 7. If the player rolls her "point" then she wins. If the player rolls her a 7 then she loses.

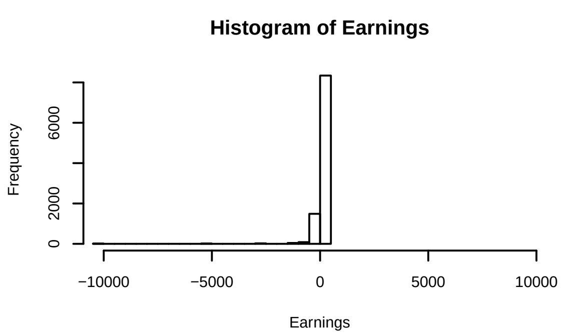
(b) Your initial bet will be 10 dollars. The return on any bet will be the amount that you bet; e.g., if you bet 10 dollars and win, then you win 10 dollars. You are going to play exactly 10 games of craps. If you lose on the previous game, your next bet requires you to "double down," that is, if on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

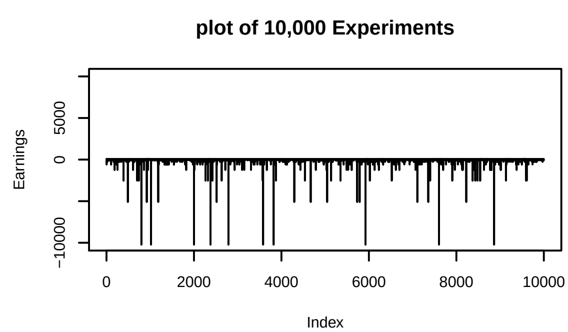
(c) Write a Monte Carol simulation to find your expected winnings after the 10 games. Also examine the distribution of your expected winnings. Does it appear that this is a good betting strategy

The dollar earnings from any single block of 10 games has a range from a total winnings of $100 to a total loss of $-10230. When we average over 10000 blocks the over all expected winnings is -5.134.

We can get a sense of how the distribution looks over the 10000 experiments by viewing the plots below. The histogram shows how the outcomes are distributed. The second plot shows the outcomes over the experiment. It is interesting to note that while the earnings from most blocks of 10 games hover around zero we obtain a fair amount of heavy losses which out weigh the wins. This can be seem more clearly in the frequency table below.

This is not a good betting strategy.





	Freq	% Valid	% Valid Cum.	% Total	% Total Cum.
-10230	10	0.10	0.10	0.10	0.10
-5100	12	0.12	0.22	0.12	0.22
-2540	10	0.10	0.32	0.10	0.32
-2530	14	0.14	0.46	0.14	0.46
-1260	10	0.10	0.56	0.10	0.56
-1250	22	0.22	0.78	0.22	0.78
-1240	11	0.11	0.89	0.11	0.89
-620	7	0.07	0.96	0.07	0.96
-610	30	0.30	1.26	0.30	1.26
-600	34	0.34	1.60	0.34	1.60
-590	13	0.13	1.73	0.13	1.73
-300	9	0.09	1.82	0.09	1.82
-290	43	0.43	2.25	0.43	2.25
-280	67	0.67	2.92	0.67	2.92
-270	39	0.39	3.31	0.39	3.31
-260	10	0.10	3.41	0.10	3.41
-140	14	0.14	3.55	0.14	3.55
-130	49	0.49	4.04	0.49	4.04
-120	92	0.92	4.96	0.92	4.96
-110	101	1.01	5.97	1.01	5.97
-100	58	0.58	6.55	0.58	6.55
-90	11	0.11	6.66	0.11	6.66
-60	8	0.08	6.74	0.08	6.74
-50	68	0.68	7.42	0.68	7.42
-40	162	1.62	9.04	1.62	9.04
-30	231	2.31	11.35	2.31	11.35
-20	170	1.70	13.05	1.70	13.05
-10	150	1.50	14.55	1.50	14.55
0	206	2.06	16.61	2.06	16.61
10	432	4.32	20.93	4.32	20.93
20	711	7.11	28.04	7.11	28.04
30	1118	11.18	39.22	11.18	39.22
40	1623	16.23	55.45	16.23	55.45
50	1775	17.75	73.20	17.75	73.20
60	1446	14.46	87.66	14.46	87.66
70	809	8.09	95.75	8.09	95.75
80	350	3.50	99.25	3.50	99.25
90	69	0.69	99.94	0.69	99.94
100	6	0.06	100.00	0.06	100.00
<NA>	0			0.00	100.00
Total	10000	100.00	100.00	100.00	100.00

References and Appendix

References:

Casella, G., & Berger, R. L. (2002). *Statistical Inference.* Thomson Learning.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., & Vehtari, A. (2014). *Bayesian Data Analysis* (Vol. 109).

Grinstead, C. M., & Snell, L. J. (1997). *Introduction to Probability.* (J. L. J. L. Snell & distributor Open Textbook Library, Eds.). Providence, RI : American Mathematical Society.

Kerman, J. (2011). Neutral noninformative and informative conjugate beta and gamma prior distributions, *5*, 1450–1470.

Ly, A., Marsman, M., Verhagen, J., Grasman, R., & Wagenmakers, E. J. (2017). A Tutorial on Fisher Information. *ArXiv e-prints*, *math.ST*.

Robert, C., & Casella, G. (2010). *Introducing Monte Carlo Methods with R* (Vol. 30). New York, NY: Springer New York, New York, NY.