MATH 8820 (Fall 2018)
Homework 2

Shirong Zhao

1. (a) **Solution.**

From the Wikipedia(https://en.wikipedia.org/wiki/Jeffreys_prior)

In Bayesian probability, the Jeffreys prior, named after Sir Harold Jeffreys, is a non-informative (objective) prior distribution for a parameter space; it is proportional to the square root of the determinant of the Fisher information matrix:

$$p(\vec{\theta}) \propto \sqrt{det I(\vec{\theta})}$$

Where $\vec{\theta}$ is the parameter vector.

Jeffreys prior has the key feature that its functional dependence on the likelihood L is invariant under reparameterization of the parameter vector $\vec{\theta}$.

The principle behind the construction Jeffreys prior is invariant to smooth, monotone transformation of the parameter. Here we briefly comment why it is non-informative. It turns out that the Jeffreys prior is indeed the uniform prior over the parameter space Θ, but not under the Euclidean geometry (pdfs depend on the geometry, as they give limits of probability of a set over the volume of the set, and volume calculation depends on geometry).

If the model only has one parameter, the Fisher Information will be just a scalar. However, when we have more than 1 parameter, it will be complicated to get the Fisher Information matrix, which is the expectation of the negative Hessian matrix of likelihood function. □

(b) **Solution.**

Assuming $Y_i \overset{iid}{\sim} Bernoulli(p)$, then

$$f(Y_i|p) = p^{Y_i}(1-p)^{1-Y_i}$$

Therefore

$$\begin{aligned} LF(p|\boldsymbol{Y}) &= \prod_{i=1}^{n} p^{Y_i}(1-p)^{1-Y_i} \\ &= p^{n\bar{Y}}(1-p)^{n-n\bar{Y}} \end{aligned}$$

Where $\bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n}$, $E(\bar{Y}) = p$, Hence

$$LLF(p|\boldsymbol{Y}) = n\bar{Y}\log p + (n - n\bar{Y})\log(1-p)$$

Therefore,

$$\begin{aligned} I(p) &= E[-\frac{\partial^2 LLF}{\partial p^2}] \\ &= E[n\bar{Y}p^{-2} + (n - n\bar{Y})(1-p)^{-2})] \\ &= \frac{n}{p(1-p)} \end{aligned}$$

Therefore, the Jeffrey's prior is

$$\begin{aligned} P(p) &\propto \sqrt{\frac{n}{p(1-p)}} \\ &\propto \sqrt{\frac{1}{p(1-p)}} \\ &\propto p^{\frac{1}{2}-1}(1-p)^{\frac{1}{2}-1} \end{aligned}$$

Hence, $p \sim \text{Beta}(\frac{1}{2}, \frac{1}{2})$, Therefore the posterior distribution is

$$\begin{aligned} P(p|\boldsymbol{Y}) &\propto p^{\frac{1}{2}-1}(1-p)^{\frac{1}{2}-1}p^{n\bar{Y}}(1-p)^{n-n\bar{Y}} \\ &\propto p^{n\bar{Y}+\frac{1}{2}-1}(1-p)^{n-n\bar{Y}+\frac{1}{2}-1} \end{aligned}$$

Hence, $p|\boldsymbol{Y} \sim \text{Beta}(n\bar{Y} + \frac{1}{2}, n - n\bar{Y} + \frac{1}{2})$,

To construct posterior inference, we could report the posterior mean or mode of p, and construct the credible interval use the posterior $(1-\alpha)\%$ quantiles, or use the HPD interval. □

(c) **Solution.**

Assuming $Y_i \overset{iid}{\sim} Poisson(\lambda)$, then

$$f(Y_i|\lambda) = \frac{\lambda^{Y_i}\exp(-\lambda)}{Y_i!}$$

Therefore

$$\begin{aligned} LF(\lambda|\boldsymbol{Y}) &= \prod_{i=1}^{n} \frac{\lambda^{Y_i}\exp(-\lambda)}{Y_i!} \\ &= \frac{\lambda^{n\bar{Y}}\exp(-n\lambda)}{\prod_{i=1}^{n} Y_i!} \end{aligned}$$

Where $\bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n}$, $E(\bar{Y}) = \lambda$, Hence

$$LLF(\lambda|\boldsymbol{Y}) = n\bar{Y}\log\lambda - n\lambda + Constant$$

Therefore,

$$\begin{aligned}
I(\lambda) &= E[-\frac{\partial^2 LLF}{\partial\lambda^2}] \\
&= E[n\bar{Y}\lambda^{-2}] \\
&= n\lambda^{-1}
\end{aligned}$$

Therefore, the Jeffrey's prior is

$$\begin{aligned}
P(\lambda) &\propto \sqrt{n\lambda^{-1}} \\
&\propto \sqrt{\frac{1}{\lambda}} \\
&\propto \lambda^{\frac{-1}{2}}
\end{aligned}$$

Hence, it's unknown distribution Therefore the posterior distribution is

$$\begin{aligned}
P(\lambda|\boldsymbol{Y}) &\propto \lambda^{\frac{-1}{2}} \frac{\lambda^{n\bar{Y}}\exp(-n\lambda)}{\prod_{i=1}^{n} Y_i!} \\
&\propto \lambda^{n\bar{Y}+\frac{1}{2}-1}\exp(-n\lambda)
\end{aligned}$$

Hence, $\lambda|\boldsymbol{Y} \sim \text{Gamma}(n\bar{Y}+\frac{1}{2}, n)$,

To construct posterior inference, we could report the posterior mean or mode of λ, and construct the credible interval use the posterior $(1-\alpha)\%$ quantiles, or use the HPD interval. □

(d) **Solution.**

Assuming $Y_i \overset{iid}{\sim} \exp(\beta)$, then

$$f(Y_i|\beta) = \beta\exp(-\beta)$$

Therefore

$$\begin{aligned}
LF(\beta|\boldsymbol{Y}) &= \prod_{i=1}^{n} \beta\exp(-\beta) \\
&= \beta^n \exp(-\beta n\bar{Y})
\end{aligned}$$

Where $\bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n}$, $E(\bar{Y}) = \frac{1}{\beta}$, Hence

$$LLF(\beta|\boldsymbol{Y}) = n\log\beta - n\bar{Y}\beta$$

Therefore,

$$\begin{aligned}
I(\beta) &= E[-\frac{\partial^2 LLF}{\partial\beta^2}] \\
&= E[n\beta^{-2}] \\
&= n\beta^{-2}
\end{aligned}$$

Therefore, the Jeffrey's prior is

$$\begin{aligned}
P(\beta) &\propto \sqrt{n\beta^{-2}} \\
&\propto \sqrt{\beta^{-2}} \\
&\propto \beta^{-1}
\end{aligned}$$

Hence, it's unknown distribution Therefore the posterior distribution is

$$\begin{aligned}
P(\beta|\boldsymbol{Y}) &\propto \beta^{-1}\beta^{n}\exp(-\beta n\bar{Y}) \\
&\propto \beta^{n-1}\exp(-\beta n\bar{Y})
\end{aligned}$$

Hence, $\beta|\boldsymbol{Y} \sim \text{Gamma}(n, n\bar{Y})$,

To construct posterior inference, we could report the posterior mean or mode of β, and construct the credible interval use the posterior $(1-\alpha)\%$ quantiles, or use the HPD interval. □

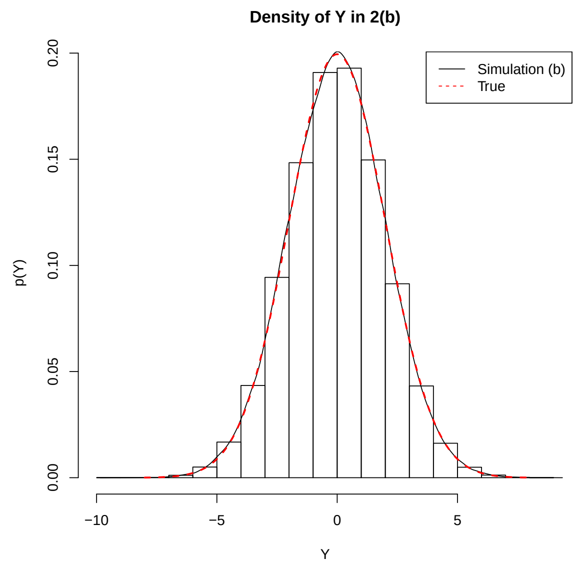
2. (a) **Solution.**

$Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

□

(b) **Solution.**

Assume $X_1 \sim N(-1, 1)$ and $X_2 \sim N(1, 3)$. Then we sample X_1, $M = 100000$ times from $N(-1, 1)$ and X_2, $M = 100000$ times from $N(1, 3)$, and then sum of X_1 and X_2 to get $M = 100000$ samples of Y. Then we can plot the histograms and density of Y.



□

(c) **Solution.**

By convolution theorem, we have the following density for $Y = X_1 + X_2$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X_1}(y-x) f_{X_2}(x) dx$$

This integral is not easy to deal with, however, notice that $f_Y(y) = E_{X_2} f_{X_1}(y - x_2)$, and hence by the law of large numbers,

$$f_Y(y) \approx \frac{1}{N} \sum_{i=1}^{n} f_{X_1}(y - x_i)$$

where $x_i, i = 1, ..., n$ is a random sample from f_{x_2}. Hence the algorithm is

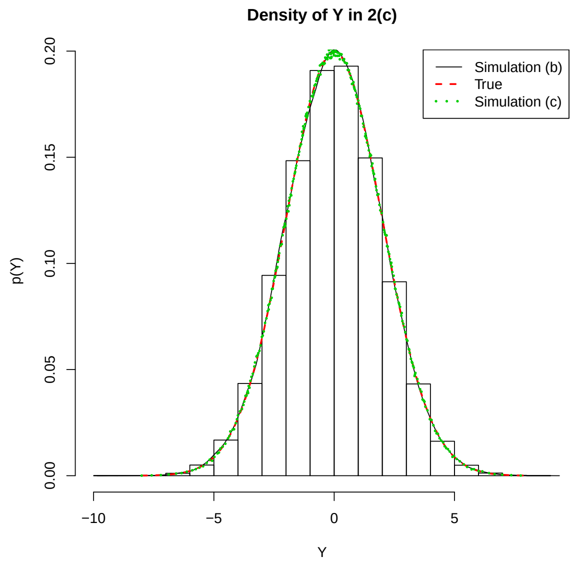
Step 1: let $y = seq(-6, -6, by = 0.01)$

Step 2: for j=0, take $yj = y[j]$

Step 3: dram N=1000, random sample observations $\{x_i\}_{i=1}^N$ from $N(\mu_2, \sigma_2^2)$, and then calculate $\hat{f}_Y(yj) = \frac{1}{N} \sum_{i=1}^n f_{X_1}(yj - x_i)$

Step 4: Repeat Step 2 until Step 3 for $j = 1, 2, 3, ...$ until all the values in $y = seq(-6, -6, by = 0.01)$ are exploited.

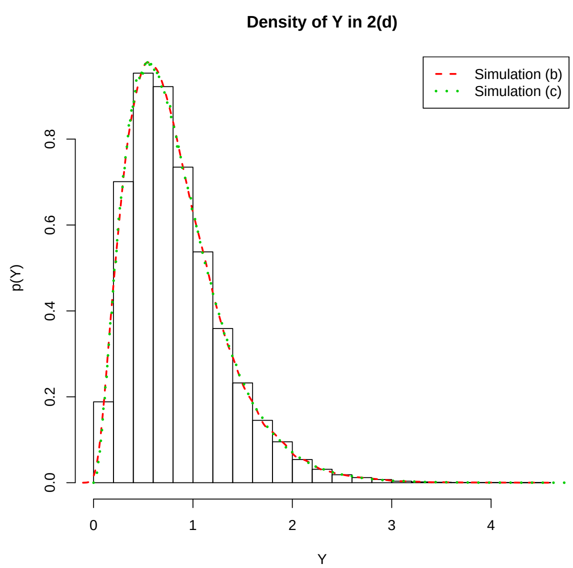
Step 5: then the density of $Y = X_1 + X_2$ can be approximated by $\hat{f}_Y(yj), j = 0, 1, 2, ...$



Simulation (b) means using methods from (b), and Simulation (c) means using methods from (c). From the graph, we can see that the two simulations did a very good job. □

(d) **Solution.**

Using the same logic as that in (b) and (c), we can generate the following graph for Y.



Simulation (b) means using methods from (b), and Simulation (c) means using methods from (c). From the graph, we can see that the two simulations did a very good job. □

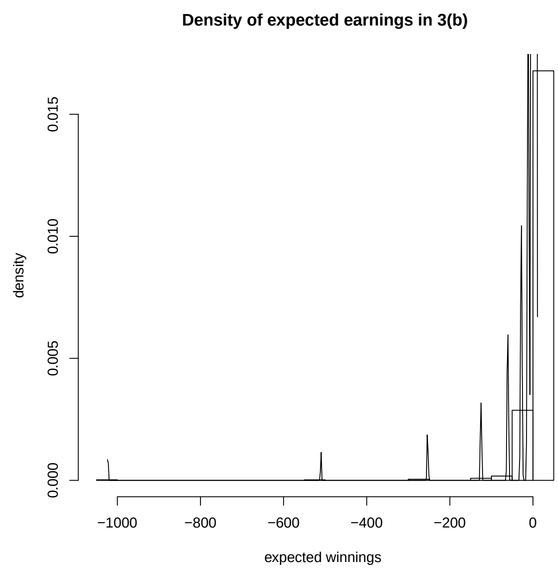
3. (a) **Solution.**

A game played with two dice. If the total is 7 or 11 (a "natural"), the thrower wins and retains the dice for another throw. If the total is 2, 3, or 12 ("craps"), the thrower loses but retains the dice. If the total is any other number (called the thrower's "point"), the thrower must continue throwing and roll the "point" value again before throwing a 7. If he succeeds, he wins and retains the dice, but if a 7 appears first, the player loses and passes the dice. Source http://mathworld.wolfram.com/Craps.html

□

(b) **Solution.**

I simulate 100000 times, the density of the expected earning is shown below. We can see that this density is highly left skewed, it means we have very high possibility to draw a very negative number, which is not good. Also, we see that the mean of expected earnings is -0.5, and the minimum is -1023, the maximum is 10. Hence double down is not a good strategy.



□

4. (a) **Solution.**

the parameter space is $(\mathbf{p}_A, \mathbf{p}_B, \mathbf{p}_C)$, Where

$$\mathbf{p}_A = (\frac{1}{3}, \frac{1}{3}, \frac{1}{6}, \frac{1}{6})$$

$$\mathbf{p}_B = (\frac{1}{6}, \frac{1}{3}, \frac{1}{3}, \frac{1}{6})$$

$$\mathbf{p}_C = (\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{2})$$

□

(b) **Solution.**

let $C = \binom{29}{5}\binom{29}{11}\binom{29}{6}\binom{29}{7}$

If it is Die A, then $LF = C * (\frac{1}{3})^5(\frac{1}{3})^{11}(\frac{1}{6})^6(\frac{1}{6})^7 = C * 3^{-16}6^{-13} = C * 1.778663 * 10^{-18}$

If it is Die B, then $LF = C * (\frac{1}{6})^5(\frac{1}{3})^{11}(\frac{1}{3})^6(\frac{1}{6})^7 = C * 3^{-17}6^{-12} = C * 3.557326 * 10^{-18}$

If it is Die C, then $LF = C * (\frac{1}{6})^5(\frac{1}{6})^{11}(\frac{1}{6})^6(\frac{1}{2})^7 = C * 2^{-7}6^{-22} = C * 5.935571 * 10^{-20}$

Hence the maximum likelihood estimate of $\mathbf{p}(p_1, p_2, p_3, p_4)$ is $\mathbf{p}_B = (\frac{1}{6}, \frac{1}{3}, \frac{1}{3}, \frac{1}{6})$. □

(c) **Solution.** Here we consider likelihood ratio test

$$LR = \frac{LF(x|\mathbf{p}_C)}{\max_{i=A,B,C}(LF(x|\mathbf{p}_i))}$$

where x is a randomly sample drawn from Die C for 29 times. We first use Monte Carlo simulation method to get the distribution of LR, and then for our sample, we calculate

$$LR0 = \frac{LF(x|\mathbf{p}_C)}{\max_{i=A,B,C}(LF(x|\mathbf{p}_i))} = 0.01668549$$

then we calculate p-value as the percentage of $LR < LR0$, the p-value is $0.00245 < 0.01 = \alpha$, Hence we reject the null hypothesis.

□