**LaTeX Code for HW 1**

\documentclass[10pt,a4paper]{article}

\usepackage[utf8]{inputenc}

\usepackage{amsmath}

\usepackage{amsfonts}

\usepackage{amsthm}

\usepackage{amssymb}

\usepackage{listings}

\usepackage{color}

\definecolor{light-gray}{gray}{0.92}

\usepackage{graphicx}

\usepackage[left=2cm,right=2cm,top=2cm,bottom=2cm]{geometry}

\usepackage{relsize}

\usepackage[english]{babel}

\usepackage[utf8]{inputenc}

\usepackage{fancyhdr}

\linespread{1.2}

\pagestyle{fancy}

\fancyhf{}

\rhead{\textit{Joseph High \ \ Hopkins ID: 9E1FDC}}

\lhead{\textit{550.792 HW 1}}

\begin{document}

\title{EN.553.732 Homework 1}

\author{Joseph High \ Hopkins ID: 9E1FDC}

\date{\today}

\maketitle

**\section{Problem 1}**

Let $p(y\_1, ..., y\_6\vert\theta)$ denote the sampling model and $\pi(\theta)$ denote the prior distribution. It is given that

\\\

$y\_1, ..., y\_6\vert\theta \sim U[\theta-\frac{1}{2}, \theta+\frac{1}{2}]$ and that $\theta \sim U[10, 20]$. We then have that

\\\

$p(y\_1, ..., y\_6\vert\theta)=\displaystyle{\frac{1}{(\theta+\frac{1}{2})-(\theta-\frac{1}{2})}} = 1$, \ $y\_i \in (\theta-\frac{1}{2}, \theta+\frac{1}{2}), \ i =1,..., 6$

\\\

\\\

and \ $\pi(\theta) = \frac{1}{10}$, \ $\theta \in [10,20]$

\\\

\\\

Note that $\theta-\frac{1}{2} < y\_i < \theta+\frac{1}{2} \ \implies \ \vert y-\theta \vert < \frac{1}{2} \ \implies \ \vert \theta-y\_i \vert < \frac{1}{2} \ \implies \ y\_i-\frac{1}{2} < \theta < y\_i+\frac{1}{2}, \ i = 1,..., 6$

\\\

\\\

Using the available data $y\_i \in\{11.0, 11.5, 11.7, 11.1, 11.4, 10.9\}$, we can find updated bounds for $\theta$.

\\\

\\\

Indeed, since the above inequality holds for all $i=1,..., 6$, then $\underset{i}{\max}\{y\_i\}-\frac{1}{2}<\theta<\underset{i}{\min}\{y\_i\}+\frac{1}{2}$

\\\

\\\

$\implies 11.7 - \frac{1}{2} < \theta < 10.9+\frac{1}{2} \ \implies \ 11.2<\theta<11.4$

\\\

The posterior distribution of $\theta$ is $p(\theta\vert y\_1,..., y\_6) \propto p(y\_1, ..., y\_6\vert\theta)\pi(\theta)$, implying that the posterior distribution is also uniformly distributed.

More precisely, by Bayes' rule $$p(\theta\vert y\_1,..., y\_6) = \frac{p(y\_1, ..., y\_6\vert\theta)\pi(\theta)}{p(y\_1,...,y\_6)} = \frac{(1/10)}{p(y\_1,...,y\_6)}, \ \ 11.2<\theta<11.4$$

\\\

Now,

$1 =\mathlarger{\int\_{11.2}^{11.4}p(\theta\vert y\_1,..., y\_6)d\theta} = \mathlarger{\int\_{11.2}^{11.4}\frac{(1/10)}{p(y\_1,...,y\_6)}d\theta} = \mathlarger{\frac{1}{p(y\_1,...,y\_6)}\int\_{11.2}^{11.4}\frac{1}{10}d\theta} = \mathlarger{\frac{0.2/10}{p(y\_1,...,y\_6)}}$

\\\

\\\

\\\

$\implies \ \mathlarger{\frac{1}{p(y\_1,...,y\_6)}} = 50$

\\\

\\\

Thus, $$p(\theta\vert y\_1,..., y\_6) = \frac{p(y\_1, ..., y\_6\vert\theta)\pi(\theta)}{p(y\_1,...,y\_6)} =(\frac{1}{10})(50) = 5$$

\\\

Therefore, the posterior distribution of $\theta$ is $p(\theta\vert y\_1,..., y\_6) = 5, \ 11.2<\theta<11.4$.

**\section{Problem 2}**

Let $p(y\_1, ..., y\_{20} \vert \theta)$ denote the sampling model and $\pi(\theta)$ denote the prior distribution. It is given that

\\\

$y\_1, ..., y\_{20} \vert \theta \sim$ Exponential($\theta$) and that $\theta \sim$ Gamma($\alpha$, $\beta$) where E[$\theta$] = 0.2 and $\sigma[\theta]$ = 1.

\\\

The liklihood function is then $L(\theta$) = $\mathlarger{\theta^{20}e^{-(\sum\limits\_{i=1}^{20}y\_i)\theta}}$ and prior distribution is of the form $\pi(\theta)\propto \theta^{\alpha-1}e^{-\beta\theta}, \ \theta>0$

\\\

\\\

The posterior of $\theta$ is of then of the form $$p(\theta\vert y\_1,..., y\_{20}) \propto L(\theta)\pi(\theta) \propto \theta^{\alpha+19}e^{-(\sum\limits\_{i=1}^{20}y\_i + \beta)\theta}, \ \ \theta>0$$

which is proportional to the pdf of a gamma distribution with parameters $\alpha+20$ and $\beta+\sum\limits\_{i=1}^{20}y\_i$

\\\

\\\

To determine the paramaters $\alpha$ and $\beta$ we note that since $\theta \sim$ Gamma($\alpha$, $\beta$), \ E[$\theta$] = $\mathlarger{\frac{\alpha}{\beta}}$

\\\

and $\mathlarger{\sigma[\theta] = \sqrt{\frac{\alpha}{\beta^2}}}$

\\\

Then from the given information, we have that

$\mathlarger{\frac{\alpha}{\beta}}$ = 0.2 and $\mathlarger{\sqrt{\frac{\alpha}{\beta^2}}}$ = 1 $\implies \alpha = 0.2\beta \implies \frac{0.2}{\beta} = 1$

\\\

$\implies \beta = 0.2 \implies \alpha = 0.04$.

\\\

\\\

Moreover, we are given that the average time to serve a customer from a sample of 20 customers is 3.8

\\\

minutes. That is, $\bar{y}$ = $\frac{1}{20}\sum\limits\_{i=1}^{20}y\_i = 3.8 \implies \sum\limits\_{i=1}^{20}y\_i = 76$.

\\\

\\\

Thus, we have that the parameters of the posterior distribution are $\alpha+20 = 20.04$ and $\beta+\sum\limits\_{i=1}^{20}y\_i = 76.2$

\\\

Therefore, $\theta \vert y\_1, ..., y\_{20} \sim$ Gamma(20.04, 76.2).

**\section{Problem 3}**

\begin{proof}

It is given that the sampling model, $f(y\_1, ..., y\_n \vert \ p)$, has the negative binomial distribution with unknown parameter $p$ (and known r) and the prior distribution, $\pi(p)$, is the beta distribution with parameters $\alpha$ and $\beta$.

\\\

That is,

\\\

\\\

$L(p) = \mathlarger{\prod\_{i=1}^{n}f(y\_i \ \vert p)} = \mathlarger{\prod\_{i=1}^{n}\binom{y\_i + r -1}{y\_i}(1-p)^rp^{y\_i} \implies L(p) \propto (1-p)^{nr}p^{\sum\_{i=1}^{n}y\_i}}$

\\\

\\\

\\\

$\pi(p) = \mathlarger{\frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}p^{\alpha-1}(1-p)^{\beta-1} \implies \pi(p) \propto p^{\alpha-1}(1-p)^{\beta-1}}$

\\\

\\\

Therefore, the posterior distribution is of the form $$f(p \ \vert y\_1, ..., y\_n) \propto \ L(p)\pi(p) \propto \ p^{\alpha +\sum\_{i=1}^{n}y\_i - 1}(1-p)^{\beta+nr-1} \longrightarrow \textrm{Beta}(\alpha+\sum\_{i=1}^{n}y\_i, \ \beta+nr)$$

\\\

Thus, the posterior distribution, again, follows a beta distribution with parameters $\alpha+\sum\_{i=1}^{n}y\_i$ and $\beta+nr$

\\\

Therefore, the family of beta distributions is a conjugate family of prior distributions for samples from a negative binomial distribution with one unknown parameter.

\end{proof}

**\section{Problem 4}**

\textbf{Part 1:} For each i $\in[1, 100]$ \ $\mathrm{Pr}(Y\_i = y\_i \vert \ \theta) = \theta^{y\_i}(1-\theta)^{1-y\_i}$ \ since each $Y\_i$ is a binary random variable.

\\\

Then, with the assumption of conditional independence of the $Y\_i$ on $\theta$, we have $$\mathrm{Pr}(Y\_1 = y\_1, ..., Y\_{100} = y\_{100} \vert \ \theta) = \mathlarger{\prod\_{i=1}^{100}\mathrm{Pr}(Y\_i = y\_i \vert \ \theta) = \prod\_{i=1}^{100}\theta^{y\_i}(1-\theta)^{1-y\_i} = \theta^{\sum\_{i=1}^{100}y\_i}(1-\theta)^{100 - \sum\_{i=1}^{100}y\_i}}$$

\\\

Then $\mathrm{Pr}(\sum\_{i=1}^{100}Y\_i = y\_i \vert \ \theta)$ is the probability that the sum of the binary random variables is equal to $y$, where the sum of $y$ from 100 binary random variables can be achieved in $\binom{100}{y}$ distinct ways. Thus, $$\mathrm{Pr}(\sum\_{i=1}^{100}Y\_i = y\_i \vert \ \theta) = \binom{100}{y}\theta^y(1-\theta)^{100-y}$$

\\\

\textbf{Part 2:} The R code attached computes $\mathrm{Pr}(\sum\_{i=1}^{100}Y\_i = 57 \vert \ \theta)$ for each $\theta$. Refer to R code for computation.

\\\

\\\

From part 1 we know that $\mathlarger{\mathrm{Pr}(\sum\_{i=1}^{100}Y\_i = 57 \vert \ \theta) = \binom{100}{57}\theta^{57}(1-\theta)^{43}}$.

\\\

The corresponding outputs for each $\theta \in \{0.0, 0.1, ... , 0.9, 1.0\}$ are below.

\begin{table}[h]

\centering

\begin{tabular}{c c c c c c c c c c c c}

\hline\hline

$\theta$ & \ \ 0.000 & 0.1 & 0.2 & 0.3 & 0.4 & 0.5 \\

\hline

$\mathlarger{\mathrm{Pr}(\sum\_{i=1}^{100}Y\_i = 57 \vert \ \theta)}$ & \ \ 0.000 & \ $4.107 \times 10^{-31}$ & \ $3.738 \times 10^{-16}$ & \ $1.307 \times 10^{-8}$ & \ $2.286 \times 10^{-4}$ & \ $3.007 \times 10^{-2}$ \\

\hline

\hline

\end{tabular}

**\label{tab:hresult}**

\end{table}

\begin{table}[h]

\centering

\begin{tabular}{c c c c c c c c c c c c}

\hline\hline

$\theta$ & \ \ 0.6 & 0.7 & 0.8 & 0.9 & 1.0 \\

\hline

$\mathlarger{\mathrm{Pr}(\sum\_{i=1}^{100}Y\_i = 57 \vert \ \theta)}$ & \ \ $6.673 \times 10^{-2}$ & \ $1.853 \times 10^{-3}$ & \ $1.004 \times 10^{-7}$ & \ $9.396 \times 10^{-18}$ & \ 0.000 \\

\hline

\hline

\end{tabular}

**\label{tab:hresult}**

\end{table}

\\\

\textbf{R code and Plot for Part 2}

\lstset{backgroundcolor=\color{light-gray}, frame=single, basicstyle = \ttfamily\small}

\begin{lstlisting}{language=R}

> #Part 2

> theta<-c(0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0)

> Y<-rep(1,11)

> for(i in 1:11)

+ Y[i]<-choose(100,57)\*(theta[i]^57)\*(1-theta[i])^43

> print(Y)

[1] 0.000000e+00 4.107157e-31 3.738459e-16 1.306895e-08 2.285792e-04 3.006864e-02

[7] 6.672895e-02 1.853172e-03 1.003535e-07 9.395858e-18 0.000000e+00

>

> plot(theta, Y, type = "h", main = "Problem 4, Part 2",

+ xlab = expression(paste(theta)), ylab="Pr(Y=57 | theta)")

\end{lstlisting}

**\includegraphics**[scale= 0.65]{P4P2.pdf}

\\\

\\\

\textbf{Part 3:} Since $p(\theta = 0.0) = p(\theta = 0.1) = ... = p(\theta = 1.0)$ then since $\theta$ discrete, we have

\\\

\\\

$p(\theta = 0.0) = p(\theta = 0.1) = ... = p(\theta = 1.0) = \mathlarger{\frac{1}{11}}$

Then, by Bayes' rule, we have for each $\theta \in \{0.0, 0.1,..., 1.0\}$

$$p(\theta \ \vert \ \sum\_{i=1}^{100}Y\_i = 57) = \frac{\mathlarger{p(\sum\_{i=1}^{100}Y\_i = 57 \vert \ \theta)p(\theta)}}{\mathlarger{p(\sum\_{i=1}^{100}Y\_i = 57)}} = \frac{\mathlarger{\binom{100}{57}\theta^{57}(1-\theta)^{43}(\frac{1}{11})}}{p(\mathlarger{\sum\_{i=1}^{100}Y\_i = 57)}}$$

where\ $\mathlarger{\frac{1}{p(\sum\_{i=1}^{100}Y\_i = 57)}}$ \ is the normalization constant. Then,

$$\mathlarger{\mathrm{Pr}(\sum\_{i=1}^{100}Y\_i = 57) = \sum\_{\theta}\binom{100}{57}\theta^{57}(1-\theta)^{43}(\frac{1}{11})}$$

From the R code (below), the values of $\mathlarger{p(\theta \ \vert \ \sum\_{i=1}^{100}Y\_i = 57)}$ were found to be:

\begin{table}[h]

\centering

\begin{tabular}{c c c c c c c c c c c c}

\hline\hline

$\theta$ & \ \ 0.000 & 0.1 & 0.2 & 0.3 & 0.4 & 0.5 \\

\hline

$\mathlarger{p(\theta \ \vert \ \sum\_{i=1}^{100}Y\_i = 57)}$ & \ \ 0.000 & \ $4.154 \times 10^{-30}$ & \ $3.781 \times 10^{-15}$ & \ $1.322 \times 10^{-7}$ & \ $2.312 \times 10^{-3}$ & \ $3.041 \times 10^{-1}$ \\

\hline

\hline

\end{tabular}

**\label{tab:hresult}**

\end{table}

\begin{table}[h]

\centering

\begin{tabular}{c c c c c c c c c c c c}

\hline\hline

$\theta$ & \ \ 0.6 & 0.7 & 0.8 & 0.9 & 1.0 \\

\hline

$\mathlarger{p(\theta \ \vert \ \sum\_{i=1}^{100}Y\_i = 57)}$ & \ \ $6.749 \times 10^{-1}$ & \ $1.874 \times 10^{-2}$ & \ $1.015 \times 10^{-6}$ & \ $9.502 \times 10^{-17}$ & \ 0.000 \\

\hline

\hline

\end{tabular}

**\label{tab:hresult}**

\end{table}

\\\

\\\

\\\

\textbf{R code and Plot for Part 3}

\begin{lstlisting}{language=R}

> #Part 3

> x<-rep(1,11)

> x1<-rep(1,11)

> for(i in 1:11)

+ x[i]<-(choose(100,57)\*(theta[i]^57)\*(1-theta[i])^43)\*(1/11)

> NormConstant<-1/(sum(x))

> x1<-x\*NormConstant

> print(x1)

[1] 0.000000e+00 4.153701e-30 3.780824e-15 1.321705e-07 2.311695e-03 3.040939e-01

[7] 6.748515e-01 1.874172e-02 1.014907e-06 9.502335e-17 0.000000e+00

>

> plot(theta, x1, type = "h", main = "Problem 4 Part 3",

+ xlab = expression(paste(theta)), ylab = "Posterior")

\end{lstlisting}

**\includegraphics**[scale= 0.65]{P4P3.pdf}

\\\

\\\

\textbf{Part 4:} Using the uniform density on [0,1] for the prior, the posterior density $p(\theta) \times$Pr$(\sum\_{i=1}^{100}Y\_i = 57 \ \vert \ \theta)$ is such that

$$p(\theta \ \vert \ \sum\_{i=1}^{100}Y\_i = 57) = \frac{\mathlarger{\binom{100}{57}\theta^{57}(1-\theta)^{43}\times(1)}}{\mathlarger{\mathrm{Pr}(\sum\_{i=1}^{100}Y\_i = 57)}} \ \ , \ \ \mathrm{for} \ \ 0 \le \theta \le 1$$

This plot for this posterior density is below.

\\\

\textbf{R code and Plot for Part 4}

\begin{lstlisting}{language=R}

> #Part 4

> f<-curve(choose(100,57)\*x^57\*((1-x)^43), from = 0, to = 1,

+ main = "Problem 4, Part 4", xlab = expression(paste(theta)),

+ ylab = "posterior")

\end{lstlisting}

**\includegraphics**[scale= 0.65]{P4P4.pdf}

\\\

\textbf{Part 5:} From the above we see that the posterior distribution $\sim$ Beta(58, 44). However, we are not using the normalizing constant here. Rather, what is being shown is that the posterior maintains the same shape and proportions as the sampling model's joint distribution.

\\\

Since $p(\theta \ \vert \ \sum\_{i=1}^{100}Y\_i = 57) \propto \theta^{57}(1-\theta)^{43}$ \ then \ $\theta \ \vert \ \sum\_{i=1}^{100}Y\_i = 57 \sim \mathrm{Beta}(58, 44)$

\\\

Below is the plot of the posterior.

\\\

\\\

\textbf{R code and Plot for Part 5}

\begin{lstlisting}{language=R}

> #Part 5

> v<-seq(0, 1, length = 200)

> z<-dbeta(v, 58, 44)

> plot(v, z, type = "l", main = "Problem 4, Part 5 Beta(58,44) Distribution",

+ xlab = expression(paste(theta)), ylab = "Posterior")

\end{lstlisting}

**\includegraphics**[scale= 0.65]{P4P5.pdf}

\\\

\textit{Relationship Between Plots:} While the plots in parts 2 and 3 are discrete and continuous in parts 3 and 4, each set of plots is unimodal with a mode around 0.6 and has a similar shape in distribution. However, this shouldn't be too surprising since each set of plots is based on the same data/results. In parts 4 and 5 we notice that the distributions have the same shape, but the distribution in part 4 is normalized and includes the binomial coefficient. However, what is important to recognize is that the shape of the posterior distribution can be derived from the liklihood $\times$ prior, that is, without any constants or functions that are not dependent on $\theta$. In other words, the statement poster $\propto$ liklihood $\times$ prior (or sampling model $\times$ prior) is meaningful. We are able extract sufficient information about the posterior from the sampling model and the prior.

**\section{Problem 5}**

\textbf{R Code and Plot}

\begin{lstlisting}{language=R}

> #Problem 5

> theta\_0<-c(0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9)

> n\_0<-c(1,2,8,16,32)

>

> a<-matrix(0L, nrow =length(theta\_0), ncol =length(n\_0))

> b<-matrix(0L, nrow =length(theta\_0), ncol =length(n\_0))

> for (i in 1:length(theta\_0))

+ {for (j in 1:length(n\_0))

+ {a[i,j]=theta\_0[i]\*n\_0[j]

+ b[i,j]=(1-theta\_0[i])\*n\_0[j]

+ }

+ }

> Pr<-matrix(0L, nrow =length(theta\_0), ncol =length(n\_0))

> for (i in 1:length(theta\_0))

+ {for (j in 1:length(n\_0))

+ {

+ f <- function(x)

+ {choose(100,57)\*(x^57)\*((1- x)^43)\*(gamma(a[i,j]+b[i,j])/

+ (gamma(a[i,j])\*gamma(b[i,j])))\*(x^(a[ i,j]-1))\*(1-x)^(b[i,j]-1)}

+ bot<-integrate(f,0, 1, rel.tol=1e-10)$value

+ top<-integrate(f,0.5, 1, rel.tol=1e-10)$value

+ Pr[i,j]<-top/bot

+

+ }

+ }

> contour(theta\_0, n\_0, Pr,main = "Problem 5 Countour Plot",

+ xlab=expression(paste(theta)), ylab='n\_0')

\end{lstlisting}

**\includegraphics**[scale= 0.65]{P5.pdf}

\\\

The contour plot suggests that "lower values of $n\_0$ are generally 90\% or more certain that" $\theta > 0.5$

\\\

(as cited in Hoff, pgs. 6-7). In other words, according to the countour plot, we can be at least 90\% certain that $\theta > 0.5$, which is quite significant. It is reasonable to believe, then, that $\theta > 0.5$.

\\\

\\\

\\\

\textit{Reference for Response/Answer to Problem 5:}

\\\

Hoff, Peter D. (2009). \textit{A First Course in Bayesian Statistical Methods}. New York, NY: Springer.

\end{document}