STAT2651: Bayesian Inference	4/21/2021

HW5

Name: Kamaru-Deen Lawal
Email: kal246@pitt.edu

Chapter 6: Problem 2

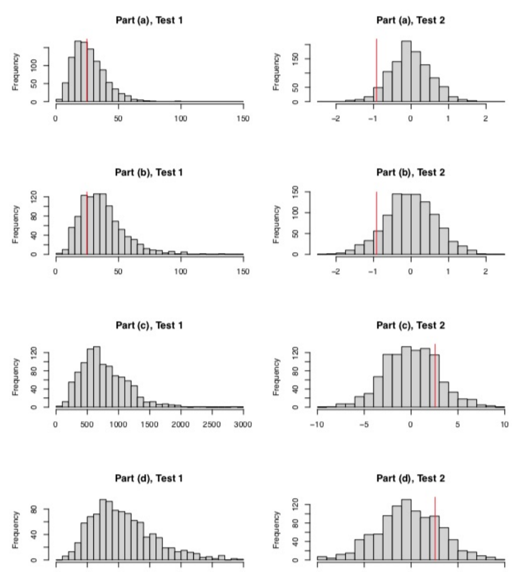
(a) T1 p-values: .60, .60 .17, .15
T2 p-values: .95, .91, .18, .23

Simulated y^{rep} as 10 draws from a Poisson Distribution with parameters given from 2.13. We test the independent poisson assumption with the LOOCV Mean Squared Error. The test statistic of interest is defined as $T_1(y) = \sum_{i=1}^{10}(y_i - \bar{y}_{-i})$. To test the no trend assumption, we fit a linear regression model such that $y_i = \beta_1 i + \beta_0$. From there we simply record β_1 as $T_2(y^{\text{rep}})$

(b) Below we see a table with the results of the simulations.

Part of Exercise 2.13	$T_1(\tilde{y})$	p-value for T_1	$T_2(\tilde{y})$	p-value for T_2
Part (a) Model	24.64	0.505	-0.92	0.946
Part (b) Model	24.64	0.742	-0.92	0.910
Part (c) Model	70788.14	≈ 0	2.55	0.180
Part (d) Model	70788.14	≈ 0	2.55	0.225

Below we see a histogram of the resultant distributions,



(c) The results of the posterior predictive checks agree with what was used in the predictive models. We also notice that the p-values for the first test for (a) and (b) are closer to .5 than for the first test for (c) and (d) which means they fit the data better. Both models show poor performance with respect to the second test statistic, which implies we can leave out the time component in our model.

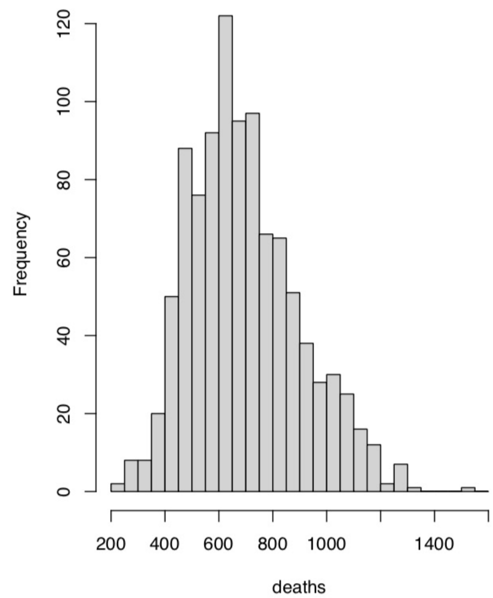
Chapter 6: Problem 3

(a) We will model airline fatalities using a compound Poisson model. Let A denote a random variable representing accidents, and D denote a random variable representing deaths. The model can be parameterized as follows (Note: The parameters for θ_1, θ_2 are drawn from models 2.13(a) and 2.13(c) respectively.)

$$A \sim \text{Poisson}(\theta_1),\ X_i \sim \text{Poission}(\theta_2),\ D = \sum_{i=1}^{A} X_i;\ \theta_1 \sim \text{Gamma}(\alpha_1, \beta_1),\ \theta_2 \sim \text{Gamma}(\alpha_2, \beta_2)$$

(b) Fitting the model resulted in the following values $\alpha_1 = 238, \beta_1 = 10, \alpha_2 = 28, \beta_2 = 1$

(c) Below we see the histogram of the prediction for 1986. The mean is 695 and the median is 668. The 95% confidence interval is $[367, 1144]$. The forecast is different because it does not assumes accidents and deaths are independent. Instead it assumes that the number of deaths is dependent on the number of accidents. In addition the posterior has a much larger variance than the fatality models from 2.13(c) and 2.13(d).



(d) Model results in a p-value of .119 for T1, and .468 for T2. The model having p-values closer to .5 than the airline fatality models in 2.13(c) and 2.13(d) tells us that the models that we developed fit the data better with respect to these tests.

Chapter 6: Problem 6

(a) Under the new protocol we have, $p(y, n|\theta) = \Pi_{i=1}^{n} \theta^{y_i}(1-\theta)^{1-y_i} \cdot 1_{y_n=0} \cdot 1_{(\sum_{i=1}^{n-1}(1-y_i)=12)}$. For the given data, this is just $\theta^7(1-\theta)^{13}$, which is the likelihood for the data under the "stop after 20 trials" rule. So if the prior distribution for θ is unchanged, which means the resultant posterior distribution is also unchanged.

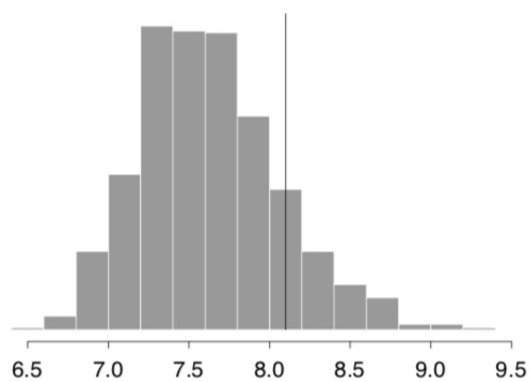
(b)
```
test <- NULL
for (i in 1:1000){
  theta <- rbeta (1,8,14)
  y.rep <- rbinom (1,1,theta)
  while (sum(y.rep==0) < 13)
    y.rep <- c(y.rep, rbinom(1,1,theta))
  n.rep <- length(y.rep)
  test <- c(test,
    sum (y.rep[2:n.rep] != y.rep[1:(n.rep-1)]))}
hist (test, xlab="T (y-rep)", yaxt="n",
  breaks=seq(-.5, max(test)+.5), cex=2)
```

The posterior predictive distribution differs in that it has a wider spread, which seems fairly intuitive since n is now a random variable instead of a constant. $T(y^{\text{rep}})$ is also much more likely to be even, specifically for low values of T, which is a a peculiarity of this particular sequential data collection rule.

Chapter 6: Problem 7

(a) For very large A, the posterior distribution for θ is approximately $N(5.5, \frac{1}{100})$. The posterior predictive distributino is $y^{\text{rep}}|y \sim N(5.5, \frac{1}{100}11^T)$, where 11^T is the $10x100$ square matrix with all elements equal to 1. We compute the distribution of $T(y^{\text{rep}})$ by simulation. The posterior The posterior predictive p-value is estimated as the proportion of simulations of $T(y^{\text{rep}})$ that exceed $T(y) = 8.1$; this is 14%.

```
test <- NULL
for (i in 1:1000){
  theta <- rnorm (1,5.1,1/sqrt(100))
  y.rep <- rnorm (100,theta,1)
  test <- c(test,max(abs(y.rep)))}
postscript ("fig6.6a.ps", horizontal=TRUE)
par (mar=c(5,5,4,1)+.1)
hist (test, xlab="T (y-rep)", yaxt="n",
  nclass=20, cex=2)
lines (rep(8.1,2), c(0,1000))
print (mean(test>8.1))
```



(b) For very large A, the prior predictive distribution for $T(y^{\text{rep}}) = \max_i |y_i^{\text{rep}}|$ will extend from 0 to the neighborhood of A. So, if A is large, the prior predictive p-value for $T(y) = 8.1$ will be close to 1.

(c) Since the prior distribution of θ is non-informative for the given data, we know that the data is primarily what determines the posterior distribution. If the data is fit to the model well, then it seems intuitive that the posterior predictive distribution is consistent with the data. On the other hand, we would not expect a diffuse prior to align with any observed data. More specifically, any finite dataset will be close to 0 compared to an improper distribution that will extend to ∞.

Chapter 6: Problem 8

(a) μ, τ are fixed, while θ varies. Let $\hat{\theta}$ and V denote the definitions for the posterior distribution and normal distribution with unknown mean and known variance, respectively. Let v^{rep} denote the sample variance of y.

$$
\begin{aligned}
p(y^{\text{rep}}) &= \int p(y^{\text{rep}}|\theta)p(\theta|y)d\theta && (1)\\
&= \int p(y^{\text{rep}}|\theta)p(\theta|\mu,\tau,y)d\theta && (2)\\
&= \int N(y^{\text{rep}}|\theta, v^{\text{rep}})N(\theta|\hat{\theta}, V)d\theta && (3)\\
&= N(r^{\text{rep}}|\mu, v^{\text{rep}} + V) && (4)
\end{aligned}
$$

An experimental replication that corresponds to this reference set could be applying a single coaching program to all of the schools. While the prior mean and variance are fixed there could be differences observed since the program is being applied to separate schools.

(b) Fixing μ and τ results in fewer random variables than in the model checks of Section 6.5. So it makes intuitive sense to expect a posterior distribution with significantly less variability, which should cause more extreme p-values for the maximum and minimum tests, and less extreme p-values for the standard deviation test.

(c) Results for each check are below. The p-values ended up being more extreme than the results from Section 6.5.