

Homework 2

AMS 207

Wyara Vanesa Moura e Silva

Spring 2018

May 8

Question 8

Write the Bayes factor, BIC, DIC and Gelfand and Ghost criterion to compare a model where n observations are assumed to be sampled with a poisson distribution with a gamma prior, to a model where the observations are sampled from a binomial distribution, with a fixed, large, number of trials and beta prior for the probability of success.

8.1

Consider the data on deaths by horsekicks in the Prussian army. Fit the data using the two different models.

Solution:

Model 1:

Using a model that assuming that the n observations came from a poisson distribution, and with prior given by a gamma distribution, we have as functions for the model:

Likelihood:

$$f(\mathbf{y}|\lambda) = \prod_{i=1}^n \frac{1}{y_i!} \exp\{-\lambda\} \lambda^{y_i} \mathbb{1}(y_i \in \mathbb{N}) \mathbb{1}(\lambda \in \mathbb{R}^+)$$

Prior:

$$\pi(\lambda) \propto \exp\{-b\lambda\} \lambda^{a-1} \mathbb{1}(\lambda \in \mathbb{R}^+)$$

Posterior:

$$\begin{aligned} \pi(\lambda|\mathbf{y}) &\propto f(\mathbf{y}|\lambda)\pi(\lambda) \\ &\propto \exp\{-(b+n)\lambda\} \lambda^{a+\sum_{i=1}^n y_i-1} \mathbb{1}(\lambda \in \mathbb{R}^+) \end{aligned}$$

Therefore, the posterior distribution has a kernel of the distribution Gamma ($a + \sum_{i=1}^n y_i, b + n$). Now we can fit the data to the model. The parameters chose for the prior are $a = 10$ and $b = 1$, in addition, the sufficient statistics $\sum_{i=1}^n y_i = 196$, and size of sample $n = 20$.

In Figure 1 on the left side we can see the histogram for the given data. The sample mean is 9.8 and the sample variance is 19.32, slightly away from the value of the average. On the right side of Figure 1 it is displayed the likelihood, prior and posterior distribution for the parameter λ , we can notice that the posterior mean is around 9.8 approximately.

Some summaries of the posterior distribution are given by: mean = 9.8095, mode = 9.7619 and credible interval = (8.5156;11.1936). These values were obtained using the closed form of the mean and mode and for the credible interval the function `qgamma` in R.

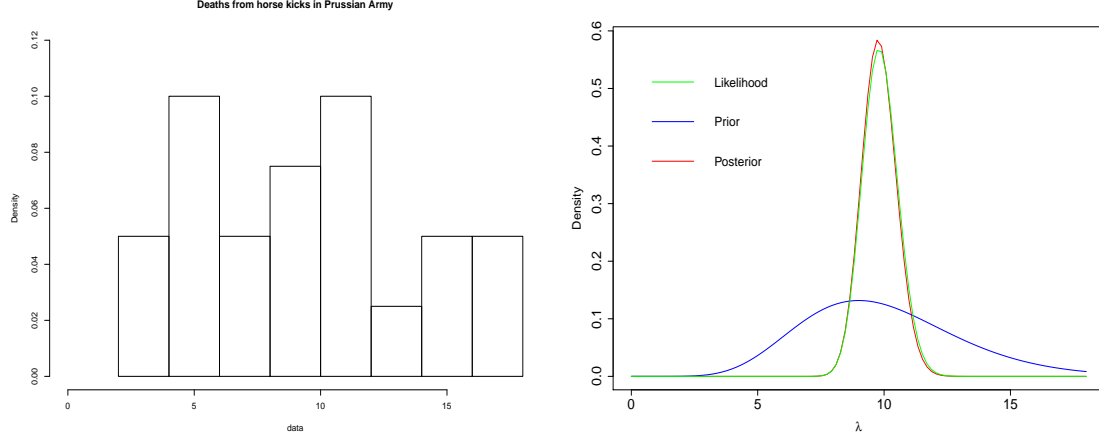


Figure 1: Histogram for the data of deaths by horsekicks and the densities for the posterior, prior and likelihood for the parameter λ .

Model 2:

Using a fixed large number of $n = 10^4$ for the trials in the binomial model. The law of rare events assures that a poisson distribution can be approximated to a binomial model with a large number for the trials and probability of success small. In addition, we are going to use a conjugate beta prior distribution, then, the functions used in this model are given by:

Likelihood:

$$f(\mathbf{y}|\theta) = \prod_{i=1}^n \binom{N}{y_i} \theta^{y_i} (1 - \theta)^{n - y_i} \mathbf{1}(y_i \in \mathbb{N}) \mathbf{1}(\theta \in (0, 1))$$

Prior:

$$\pi(\theta) \propto \theta^{\alpha-1} (1 - \theta)^{\beta-1} \mathbf{1}(\theta \in (0, 1))$$

Posterior:

$$\pi(\theta|\mathbf{y}) \propto \theta^{\sum_{i=1}^n y_i + \alpha - 1} (1 - \theta)^{n * N - \sum_{i=1}^n y_i + \beta - 1} \mathbf{1}(\theta \in (0, 1))$$

Thus, the posterior distribution has a kernel of the distribution Beta $(\sum_{i=1}^n y_i + \alpha, n * N - \sum_{i=1}^n y_i + \beta)$. Now we can fit the data to the model.

The parameters chose for the prior were $\alpha = 1$ and $\beta = 1$, in addition, the sufficient statistics $\sum_{i=1}^n y_i = 196$, and size of sample $n = 20$. Figure 2 on the left side, we can see the distribution for the likelihood, prior and posterior, however, it is not a good graph to look. On the right side, we have an approximation for the data using the poisson mass function in the lines, with the binomial mass function in the dots, fitted with the posterior mean of the parameters.

Some summaries of the posterior distribution are given by: mean = 0.000984, mode = 0.000980 and credible interval = (0.000852; 0.001127). These values were obtained using the closed form of the mean and mode and for the credible interval the function `qbeta` in R.

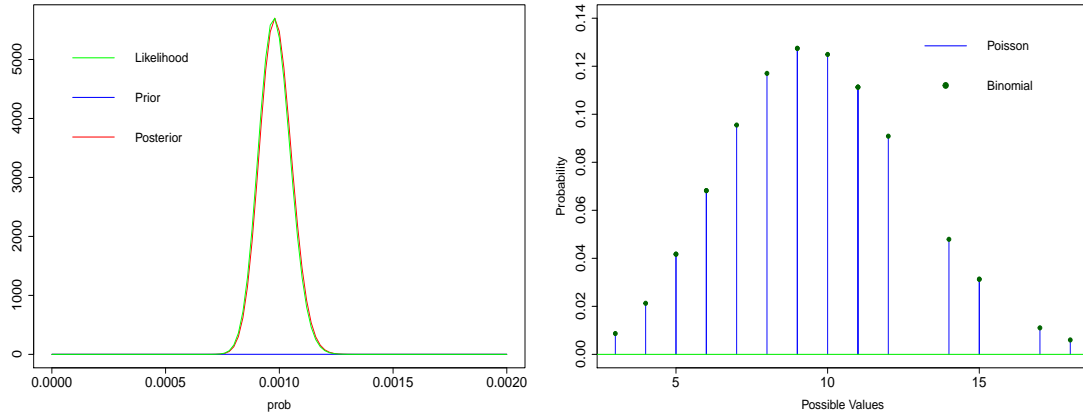


Figure 2: densities for the posterior, prior and likelihood for the parameter θ and mass function of Poisson and Binomial with the posterior mean as parameters.

8.2

Perform a prior sensitivity analysis.

Solution:

Due to the relatively small number of observations, the posterior can be affected by both the prior and the likelihood. In order to conduct sensitivity analysis, we will try these approach with various different priors in both models to get an idea of how different priors affect the resulting posterior.

Model 1:

For the model with gamma density as posterior distribution with parameters of shape and rate, $(a + \sum_{i=1}^n y_i, b + n)$, respectively. In Figure 3 we can see on the right side the graphs for the likelihood and log-likelihood. Maximizing the log-likelihood we get as maximum likelihood $\hat{\lambda}_{MLE} = 9.8000$.

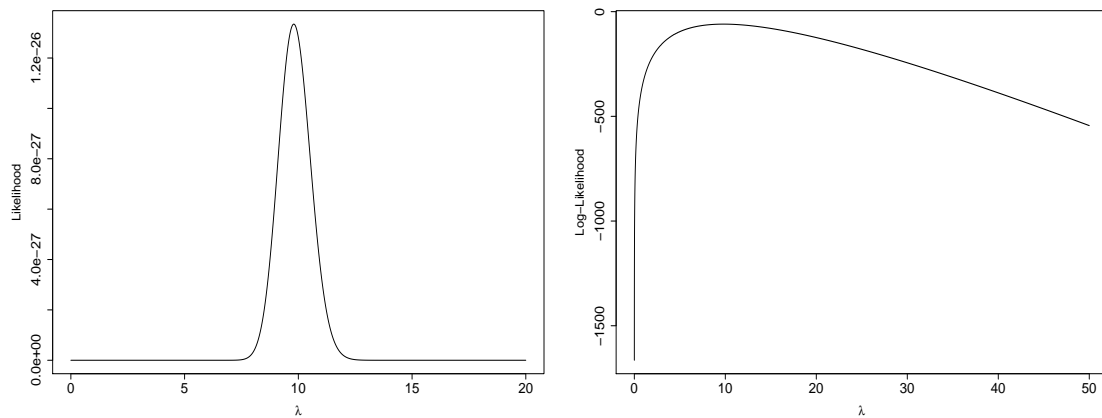


Figure 3: Likelihood and Log-Likelihood of the Poisson distribution given the deaths by horse-kicks.

In Figure 4 was generate the density for different gamma prior on the left side, and the right side a graph for the posterior distribution with those different priors.

We can see how the posterior density change when the parameters for the prior distribution changes in some cases. Moreover, we can notice that when the prior get more centered on the mean of the sample which is the same as the maximum likelihood estimator, the posterior gets better.

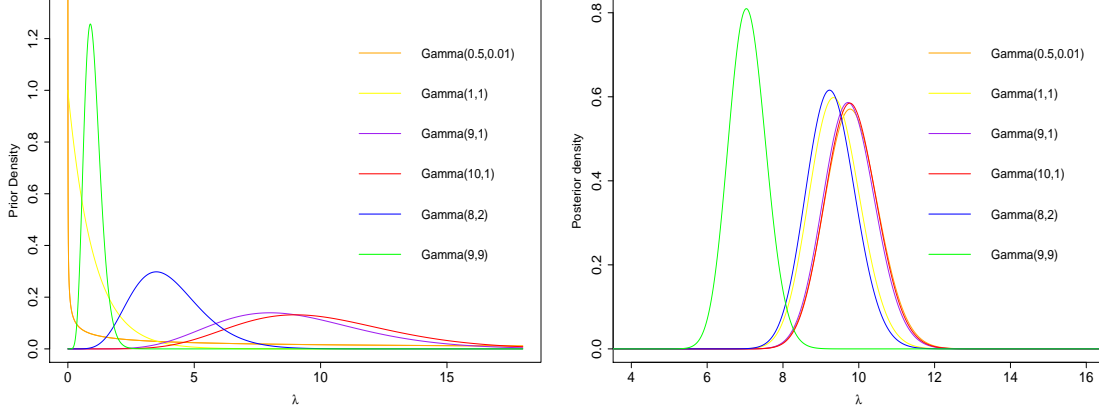


Figure 4: Prior and Posterior distribution with those different priors .

In Table 1 was summarize some posterior mean and mode for these different prior, and it can be noticed how the results change the mean posterior for the parameters when the shape is smaller or greater than the rate parameter.

The prior distribution of Gamma (0.5, 0.01) it is an approximation to the Jeffreys' prior, in order to make proper, given that the Jeffreys' prior would be a Gamma (0.5, 0), but this it is not necessarily a distribution. For this prior, we have close results with the posterior mean to the Gamma (10, 1). For the non-informative Gamma (1, 1) the value for the posterior mean is a little away from the others priors values. The graphs for the posterior in Figure 6 displays the same behavior. While the choices of the others parameters in the priors have a relative effect on the change of the posterior mean.

Table 1: Summaries of mean, mode and credibility intervals of posterior distribution using different priors.

Prior	Posterior Mean	Posterior Mode	Credibility Interval
Gamma (0.5,0.01)	9.82009	9.77011	(8.494982;11.239831)
Gamma (1,1)	9.38095	9.33333	(8.116650;10.73543)
Gamma (9,1)	9.76190	9.71428	(8.471257;11.142725)
Gamma (10,1)	9.80952	9.76190	(8.515620;11.19360)
Gamma (8,2)	9.27272	9.22727	(8.043862;10.587667)
Gamma (9,9)	7.06896	7.03448	(6.134359;8.068870)

Model 2:

For the model with beta density as posterior distribution we have as the two parameters of shapes, $(\sum_{i=1}^n y_i + \alpha, n * N - \sum_{i=1}^n y_i + \beta)$, respectively. In Figure 5 we can see the graphs for the likelihood and log-likelihood. Maximizing the log-likelihood we get as maximum likelihood $\hat{\theta}_{MLE} = 0.000979$.

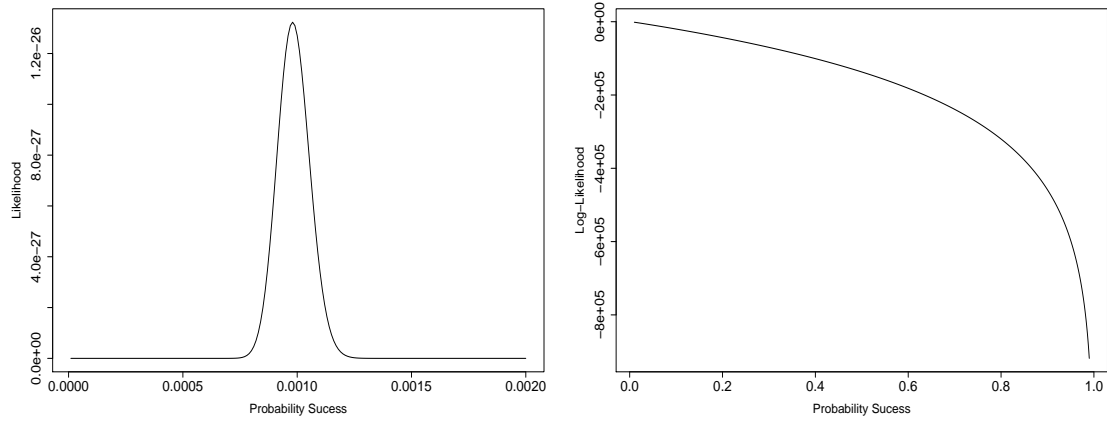


Figure 5: Likelihood and Log-Likelihood of the binomial distribution given the deaths by horse-kicks.

Figure 6 shows the density for different beta prior on the left side and on the right side a graph for the posterior distribution with those different priors. It can be noticed how the posterior density change when the parameters for the prior distribution change. Then, We can notice that when the prior put more weight on the extremes to the left or equal weights support, the posterior gets better.

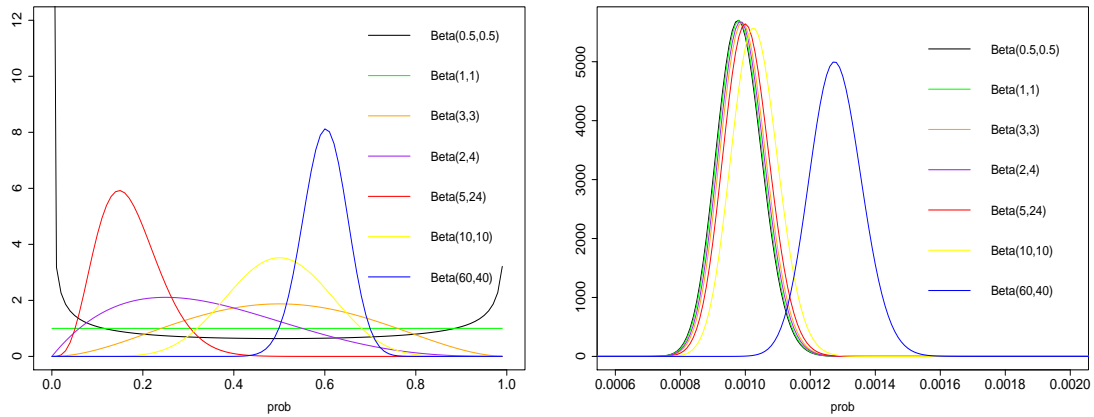


Figure 6: Prior and Posterior distribution with those different priors .

The priors distributions with parameters $\text{Beta}(0.5, 0.5)$ and $\text{Beta}(1, 1)$ are special cases of non-informative priors, being the Jeffreys and uniform prior, respectively.

In Table 2 we can notice that the posterior mean did not change much between the uniform and Jeffreys priors, and neither the priors $\text{Beta}(3, 3)$ and $\text{Beta}(2, 4)$. However, the other priors distribution have a certain influence on the posterior. The graphs for the posterior in Figure 6 displays this behavior.

Table 2: Summaries of mean, mode and credibility intervals of posterior distribution using different priors.

Prior	Posterior Mean	Posterior Mode	Credibility Interval
Beta (0.5,0.5)	0.0009825	0.0009775	(0.0008499;0.0011244)
Beta (1,1)	0.0009849	0.0009800	(0.0008522;0.0011271)
Beta (3,3)	0.0009949	0.0009899	(0.0008615;0.0011378)
Beta (2,4)	0.0009899	0.0009849	(0.0008569;0.0011324)
Beta (5,24)	0.0010048	0.0009998	(0.0008707;0.0011483)
Beta (10,10)	0.0010298	0.0010249	(0.0008941;0.0011751)
Beta (60,40)	0.0012793	0.0012743	(0.0011275;0.0014406)

8.3

Present a model comparison analysis using the criteria mentioned above.

Solution:

First, for the criterion method that gives evidence to a model without reference to prior distributions, only the likelihood, which is the BIC (Bayesian Information Criterion) favor the Model 1 (see Table 3). The maximum likelihood estimator were found using the function `optimize` in R. And it was just implemented the function below in R.

$$BIC = -2 * \log(f(\mathbf{y}|\hat{\theta}_{mle})) + k \log(n).$$

Second, the criterion that uses the information from the posterior instead of only the likelihood, which is the DIC (Deviance Information Criteria) favor the Model 1. This criterion was implemented in R by using the functions below.

$$\begin{aligned} D(\theta) &= -2 * \log(f(\mathbf{y}|\theta)) + 2 \log(h(\mathbf{y})). \\ DIC &= \bar{D} + p_D \\ p_{DIC} &= \bar{D} - D(\bar{\theta}) \end{aligned}$$

where the $D(\theta)$ is the deviance statistics, $D(\bar{\theta})$ means posterior expectation deviance, and p_{DIC} is the penalty.

The samples for the posteriors were generated by direct sampling using the functions `rgamma` and `rbeta` in R (see the R code). In addition, the values for the goodness fit and the effective size favor Model 1, too. Although, we can notice that the DIC difference between the two models is small, as well as the effective size have close values (see Table 3). The parameters for the priors distributions were the same as stated in Part 8.1.

Moreover, the Gelfand and Gosh criterion, which is known as posterior predictive loss approach, then it is not based on the likelihood. This criterion put more beliefs in favor of Model 1, as can be seen in Table 3. For the goodness of fit, the Model 1 is a little smaller than Model 2 and also, the penalty for Model 1 is smaller than Model 2 (see Table 3). This criterion was implemented in R by using the function below.

$$\begin{aligned} D &= G + P \\ G &= \sum_{l=1}^n (y_l - \mu_l)^2 \quad \text{and} \quad P = \sum_{l=1}^n \sigma_l^2 \end{aligned}$$

which $\mu_l = E(z_l|\mathbf{y})$ and $\sigma_l^2 = \text{Var}(z_l|\mathbf{y})$, in which z_l are the replicas of the posterior values for each data point (see the R code).

Regarding the Bayes Factor, first it was tried to integrate the parameters out in order to find the marginal, however, this method was unsuccessful, given that with the parameters chosen to the prior makes the integral diverge (see the R code). Then, it was done using the ratio of the exact marginal distributions of the two models, which are given by:

The Bayes Factor is given by the ratio:

$$BF_{12} = \frac{\int f(\mathbf{y}|\lambda)\pi(\lambda)d\lambda}{\int f(\mathbf{y}|\theta)\pi(\theta)d\theta} = \frac{f(\mathbf{y}^{(M1)})}{f(\mathbf{y}^{(M2)})}$$

Then, for the marginal we get,

Model 1:

$$\begin{aligned} f(\mathbf{y}^{(M1)}) &= \int f(\mathbf{y}|\lambda)\pi(\lambda)d\lambda \\ &= \int \frac{\beta^\alpha}{\prod_{i=1}^n y_i! \Gamma(\alpha)} \exp\{-(b+n)\lambda\} \lambda^{\alpha+\sum_{i=1}^n y_i-1} d\lambda \\ f(\mathbf{y}^{(M1)}) &= \frac{\Gamma(\alpha + \sum_{i=1}^n y_i)}{\Gamma(\alpha) \prod_{i=1}^n y_i!} \frac{\beta^\alpha}{(n+\beta)^{\alpha+\sum_{i=1}^n y_i}} \end{aligned}$$

Model 2:

$$\begin{aligned} f(\mathbf{y}^{(M2)}) &= \int f(\mathbf{y}|\theta)\pi(\theta)d\theta \\ &= \int \prod_{i=1}^n \binom{N}{y_i} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{\sum_{i=1}^n y_i + \alpha - 1} (1-\theta)^{nm - \sum_{i=1}^n y_i + \beta - 1} d\theta \\ f(\mathbf{y}^{(M2)}) &= \prod_{i=1}^n \binom{N}{y_i} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a + \sum_{i=1}^n y_i) \Gamma(n * N - \sum_{i=1}^n y_i + b)}{\Gamma(a+b+n * N)} \end{aligned}$$

It was used the ratio for the log of the marginals and then it was taken the exponential of the subtraction of the marginals (see the R code).

Then as shown in Table 3, the value for the Bayes Factor favor Model 1 as the best one. However, this analysis is using as parameters for the prior: Gamma(10,1) and Beta(1,1). If we change the prior to make them non-informative as choosing Gamma(1,1) and Beta(1,1), we get a value that favor Model 2, but not such a big number value for the ratio as you can see in the Table 3. Also, using non-informative priors makes the results smaller for the bayes factor, given that we are not adding much information with the prior.

However, If we keep the non-informative priors and change the value for the size of trials of the binomial to $N = 10^6$ instead of $N = 10^4$ (value used in the previous analysis), then the Bayes Factor favor the Model 1 as the best one. This confirms that the bayes factor it is very affected by the choice of the prior, and also in this model by the number of the binomial size. Given that it changes the scale of the parameters that makes the distributions to be more close or not to each other.

In addition, we are going to use an approximation, as another method to approximate the Bayes Factor. Thus, we will sample parameters (the size of the sample was $n = 500000$) from the prior densities for both models, and then we will compute the likelihood given these parameters (see the R code).

In addition, we take the average of all those likelihoods for both models and compute their ratio, then we get an approximation of the Bayes Factor (Christense et al., 2011). This method is also known as the naive Monte Carlo estimator of marginal likelihood (Gronau et al. 2017).

In Table 3, the values shown regarding the Bayes Factor for the approximation follows the conclusions reached from the ratio of the marginals.

Reference

Christensen, R., Johnson, W., Branscum, A., & Hanson, T. E. (2011). Bayesian ideas and data analysis: an introduction for scientists and statisticians. CRC Press.

Gronau, Q. F., Sarafoglou, A., Matzke, D., Ly, A., Boehm, U., Marsman, M., Leslie, D.S., Forster, J.J., Wagenmakers, E.J. & Steingroever, H. (2017). A tutorial on bridge sampling. Journal of mathematical psychology, 81, 80-97.

Table 3: Summaries of expected value and mode of posterior distribution using different priors.

Criterion		Model 1	Model 2
BIC		122.1529	122.1701
DIC		121.0584	121.1927
Gelfand and Gosh	Goodness Fit	119.1578	119.1784
	Effective Size	1.9005	2.0142
		572.0319	577.3273
	Goodness Fit	367.4867	369.1573
Bayes Factor (Beliefs in)	Penalty	204.5451	208.1700
	Marginals ⁽¹⁾	1254.15	0.000797
	Sampling ⁽¹⁾	1312.22	0.000762
	Marginals ⁽²⁾	0.67553	1.480313
	Sampling ⁽²⁾	0.68250	1.465198
	Marginals ⁽³⁾	66.9495	0.014936
Sampling ⁽³⁾		42.3343	0.023621

Note: ⁽¹⁾ N=10⁴ and Prior Gamma(10,1) and Beta(1,1). ⁽²⁾ N=10⁴ and Prior Gamma(1,1) and Beta(1,1). ⁽³⁾ N=10⁶ and Prior Gamma(1,1) and Beta(1,1).

R Code

See below all the R Code used to perform the analysis.

```
rm(list=ls(all=TRUE))
set.seed(7)

# Question 8

# number of deaths, number of occurrences
# Annual deaths from horse kicks in the Prussian army (1875-1894)

# data
n = 20
data = c(3, 5, 7, 9, 10, 18, 6, 14, 11, 9, 5, 11, 15, 6, 11, 17, 12, 15, 8, 4)

mean(data)
var(data)

# plot for data
plot(data, type="l", main="Deaths from horse kicks in Prussian Army")
plot(density(data))
hist(data, freq=F, ylim=c(0,0.12), xlim=c(0,18),
main="Deaths from horse kicks in Prussian Army")
lines(density(data, bw = "SJ"))

##### Model 1 - Poisson #####

# mle estimator
loglike_pois <- function(par, data){
  x = data
  lambda = par[1]
  sum(dpois(x, lambda, log = TRUE))
}

mle_pois = optim(par=c(0.3), loglike_pois, data=data, method="BFGS",
hessian = TRUE, control = list(fnscale = -1))
mle_pois$par

#plot logLike:
p.seq <- seq(0, 50, 0.001)
plot(p.seq, sapply(p.seq, loglike_pois, data=data), type="l", cex.axis=1.6,
cex.lab=1.4, ylab="Log-Likelihood", xlab=expression(paste(lambda)))

# like
like_pois = function(data,lambda){prod(dpois(data, lambda, log = FALSE))}
# plot Like
p.seq <- seq(0, 20, 0.001)
plot(p.seq, sapply(p.seq, like_pois,data=data), type="l", cex.axis=1.6,
cex.lab=1.4, ylab="Likelihood", xlab=expression(paste(lambda)))
```

```

# plot for the mass function
plot(data, dpois(data, mle_pois$par), ylim=c(0,0.14), type="h", lwd=1,
col="blue", cex.axis=1.6, cex.lab=1.4,
      main="", ylab="Probability", xlab="Number of occurrences" )
abline(h=0, col="green2")
N=10^4
points(data, dbinom(data, N, 9.8/N), pch=19, col="darkgreen")
legend(13,0.145,lty = c(1,NA), pch=c(NA,19), cex=1.4, bty = "n",
lwd = 1, col=c("blue","darkgreen"), legend=c("Poisson","Binomial") )

# plot posterior, prior likelihood
colors = c("orange","yellow", "purple","red","blue","green")
a = c(0.5,1,9,10,8,9)
b = c(0.01,1,1,1,2,9)
curve(dgamma(x,shape=a[4] + sum(data),rate=b[4] + n),cex.axis=1.6,
cex.lab=1.4,col="red",xlab=expression(paste(lambda)),ylab="Density",
lwd=1,from=0,to=18)
curve(dgamma(x,shape=a[4],rate=b[4]),col="blue",lwd=1,add=TRUE)
curve(dgamma(x, shape=sum(data)+1,rate=n), add=TRUE,type="l",
col="green", xlab="Number of occurrences",ylab="Probability")
legend(0,0.58,lty = c(1), cex=1.4, bty = "n", lwd = 1,
col=c("green","blue","red"), legend=c("Likelihood","Prior","Posterior") )

# prior plots
colors = c("orange","yellow", "purple","red","blue","green")
a = c(0.5,1,9,10,8,9)
b = c(0.01,1,1,1,2,9)
range_pois = seq(0,18,0.001)
plot(range_pois, dgamma(range_pois,a[1] , b[1]), ylim=c(0,1.3),
col=colors[1],type="l", ylab="Prior Density", xlab = expression(paste(lambda)),
cex.axis=1.6, cex.lab=1.4, main="")
for (i in 1:length(a))
{
  lines(range_pois, dgamma(range_pois, a[i] , b[i]),
        type="l", ylab="", col=colors[i], lty=1)
}

legend(10.7,1.3,lty = c(1,rep(1,length(a))), cex=1.4, bty = "n", lwd = 1,
      legend=c("Gamma(0.5,0.01)","Gamma(1,5)","Gamma(9,1)","Gamma(10,1)",
"Gamma(8,2)","Gamma(9,9)"),
      , col=colors)

# posterior plots
dev.off()
colors = c("orange","yellow", "purple","red","blue","green")
a = c(0.5,1,9,10,8,9)
b = c(0.01,1,1,1,2,9)
range_pois = seq(0,18,0.001)

```

```

plot(range_pois, dgamma(range_pois,a[1] + sum(data), b[1] + n),
     col=colors[1], ylim=c(0,0.8), type="l", ylab="Posterior density",
     xlab = expression(paste(lambda)),cex.axis=1.6, cex.lab=1.4, main="")
for (i in 2:length(a))
{
  lines(range_pois, dgamma(range_pois, a[i] + sum(data), b[i] + n),
        type="l", ylab="", col=colors[i], lty=1)
}
legend(11.8,0.8,lty = c(1,rep(1,length(a))), cex=1.4, bty = "n", lwd = 1,
      legend=c("Gamma(0.5,0.01)","Gamma(1,5)","Gamma(9,1)","Gamma(10,1)",
               "Gamma(8,2)","Gamma(9,9)"),
      col=colors)
#abline(v=mle_pois$par)

```

```

# posterior mean
a = c(0.5,1,9,10,8,9)
b = c(0.01,1,1,1,2,9)
mean_pois=rep(0,length(a))
mode_pois=rep(0,length(a))
var_pois=rep(0,length(a))
quant_pois=matrix(0,length(a),3)

```

```

par(mfrow=c(2,2))
for (i in 1:length(a))
{
  mean_pois[i] = (a[i] + sum(data))/ (b[i] + n)
  mode_pois[i] = (a[i] + sum(data) - 1)/ (b[i] + n)
  var_pois[i] = (a[i] + sum(data))/ (b[i] + n)^2
  quant_pois[i,] = qgamma(c(0.025, .5, 0.975), a[i] + sum(data), b[i] + n)
}

```

```

mean_pois
mode_pois
quant_pois

```

Model 2 - Binomial

```

#large number of trials
N = 10^4

```

```

# mle estimator
loglike_bino <- function(par, data){
  x = data
  prob = par[1]
  sum(dbinom(x, N, prob, log = TRUE))
}

```

```

}

mle_bino = optim(par=c(0.01), loglike_bino, data=data, method="Brent",
lower = 0, upper = 1, control = list(fnscale = -1))

mle_bino$par
loglike_bino(c(mle_bino$par),data=data)

dev.off()
# plot for the mass function
plot(data, dbinom(data, N, mle_bino$par), type="h", lwd=1, col="blue",
ylim=c(0,0.14),cex.axis=1.6, cex.lab=1.4,
      main="", ylab="Probability", xlab="Possible Values")
abline(h=0, col="green2")
points(data, dbinom(data, N, 9.8/N), pch=19, col="darkgreen")

legend(13,0.145,lty = c(1,NA), pch=c(NA,19), cex=1.4, bty = "n", lwd = 1,
col=c("blue","darkgreen"), legend=c("Poisson","Binomial") )

#plot logLike:
p.seq = seq(0.01, 0.99, 0.001)
plot(p.seq, sapply(p.seq, loglike_bino, data=data), type="l", cex.axis=1.6,
cex.lab=1.4, ylab="Log-Likelihood",xlab="Probability Sucess")

# likelihood posterior prior
colors = c("black","green","orange","purple","red","yellow","blue")
alpha=c(0.5,1,3,2,5,10,60)
beta=c(0.5,1,3,4,24,10,40)
curve(dbeta(x, alpha[2] + sum(data), N*n - sum(data) + beta[2]),cex.axis=1.6,

cex.lab=1.4,col="red",xlab="prob",ylab="Density",lwd=1,from=0,to=0.002)
curve(dbeta(x, alpha[2], beta[2]),col="blue",lwd=1,add=TRUE)
curve(dbeta(x, sum(data), N*n - sum(data)), add=TRUE,type="l", col="green",

xlab="Number of occurrences",ylab="Probability")
legend(0,5700,lty = c(1), cex=1.4, bty = "n", lwd = 1,
col=c("green","blue","red"), legend=c("Likelihood","Prior","Posterior") )

# prior plots
colors = c("black","green","orange","purple","red","yellow","blue")
alpha=c(0.5,1,3,2,5,10,60)
beta=c(0.5,1,3,4,24,10,40)
range_prior = seq(0.0001,1,10^-2)

plot(range_prior, dbeta(range_prior, alpha[1], beta[1]), col=colors[1],
ylim=c(0,12), xlim=c(0,1),
      type="l", ylab="", xlab = "prob", cex.axis=1.6, cex.lab=1.4, main="")
for (i in 2:length(alpha))

```

```

{
  lines(range_prior, dbeta(range_prior, alpha[i], beta[i]),
        type="l", ylab="", col=colors[i], lty=1)
}

legend(0.65,12.9,lty = c(1,rep(1,length(a))), cex=1.4, bty = "n", lwd = 1,
      legend=c("Beta(0.5,0.5)","Beta(1,1)","Beta(3,3)","Beta(2,4)","Beta(5,24)",
        "Beta(10,10)","Beta(60,40)"),
      , col=colors)

# posterior plots
dev.off()
colors = c("black","green","orange","purple","red","yellow","blue" )
alpha=c(0.5,1,3,2,5,10,60)
beta=c(0.5,1,3,4,24,10,40)
range_post = seq(0.00001,0.003,10^-5)
plot(range_post, dbeta(range_post, alpha[1] + sum(data), N*n - sum(data) +
beta[1]), col=colors[1],
      type="l", ylab="", xlab = "prob", cex.axis=1.6, cex.lab=1.4,
      main="",xlim=c(0.0001,0.0030))
for (i in 2:length(alpha))
{
  lines(range_post, dbeta(range_post, alpha[i] + sum(data), N*n - sum(data) +
beta[i]),
        type="l", ylab="", col=colors[i], lty=1)
}
legend(0.0017,5900,lty = c(1,rep(1,length(a))), cex=1.4, bty = "n", lwd = 1,
      legend=c("Beta(0.5,0.5)","Beta(1,1)","Beta(3,3)","Beta(2,4)","Beta(5,24)",
        "Beta(10,10)","Beta(60,40)"),
      col=colors)
#abline(v=mle_bino$par)

# posterior mean
alpha=c(0.5,1,3,2,5,10,60)
beta=c(0.5,1,3,4,24,10,40)
mean_bino=rep(0,length(alpha))
mode_bino=rep(0,length(alpha))
quant_bino=matrix(0,length(alpha),3)

par(mfrow=c(2,2))
for (i in 1:length(alpha))
{
  mean_bino[i] = (alpha[i] + sum(data))/ (alpha[i] + beta[i] + N*n)
  mode_bino[i] = (alpha[i] + sum(data) - 1)/ ( alpha[i] + beta[i] + N*n - 2)
  quant_bino[i,] = qbeta(c(0.025, .5, 0.975), alpha[i] + sum(data), N*n -
sum(data) + beta[i])
}

```

```
mean_bino
mode_bino
quant_bino
```

```
##### Part 3 #####
```

```
# Information Criteria
```

```
a = c(0.5,1,9,10,8,9)
```

```
b = c(0.01,1,1,1,2,9)
```

```
alpha=c(0.1,1,3,2,5,10,40)
```

```
beta=c(0.1,1,2,4,14,20,60)
```

```
i=4;j=2
```

```
# data
```

```
N = 10^4
```

```
n = 20
```

```
data = c(3, 5, 7, 9, 10, 18, 6, 14, 11, 9, 5, 11, 15, 6, 11, 17, 12, 15, 8, 4)
```

```
ns = 10000
```

```
fm1 = function(lambda){sum(dpois(data,lambda,log=TRUE))}
```

```
fm2 = function(prob){sum(dbinom(data,N,prob,log=TRUE))}
```

```
# maximum likelihood estimation
```

```
lambda.mle = optimize(fm1,lower=0, upper=15, maximum=TRUE)$m
```

```
prob.mle = optimize(fm2,lower=0, upper=1, maximum=TRUE)$m
```

```
# AIC and BIC
```

```
k = 1
```

```
AIC1 = -2*fm1(lambda.mle)+k
```

```
AIC2 = -2*fm2(prob.mle)+k
```

```
BIC1 = -2*fm1(lambda.mle)+k*log(n)
```

```
BIC2 = -2*fm2(prob.mle)+k*log(n)
```

```
AICs = c(AIC1,AIC2)
```

```
BICs = c(BIC1,BIC2)
```

```
# posterior distribution
```

```
lambda.post = rgamma(ns, a[i] + sum(data), b[i] + n)
```

```
prob.post = rbeta(ns, alpha[j] + sum(data), N*n - sum(data) + beta[j])
```

```
# DIC (deviance information criteria)
```

```
means1 = mean(lambda.post)
```

```
means2 = mean(prob.post)
```

```
mu11 = fm1(means1)
```

```

mu22    = fm2(means2)

pred1 = NULL; pred2 = NULL
for(t in 1:ns){
  pred1[t] = fm1(lambda.post[t])
  pred2[t] = fm2(prob.post[t])
}

pD1 = 2*(mu11 - mean(pred1))
pD2 = 2*(mu22 - mean(pred2))

DIC1 = -2*mu11 + 2*pD1
DIC2 = -2*mu22 + 2*pD2

pDs = c(pD1,pD2)
DICs = c(DIC1,DIC2)

#Gelfand and Ghosh
a = c(0.5,1,9,10,8,9)
b = c(0.01,1,1,1,2,9)

alpha=c(0.1,1,3,2,5,10,40)
beta=c(0.1,1,2,4,14,20,60)

i=4;j=2
nsp = 10000

# posterior distribution
lambda.post = rgamma(nsp,a[i] + sum(data), b[i] + n)
prob.post = rbeta(nsp, alpha[i] + sum(data), N*n - sum(data) + beta[i])

# predicted values
pred_values_1 = matrix(0,nsp,n)
pred_values_2 = matrix(0,nsp,n)

for(i in 1:length(lambda.post))
{
  #obtain one prediction for death
  pred_values_1[i,]=rpois(rep(1,n),rep(lambda.post[i],n))
  pred_values_2[i,]=rbinom(rep(1,n),N,rep(prob.post[i],n))
}

#G term of gelfand and ghosh
G1 = sum((apply(pred_values_1, 2,mean)-data)^2)
P1 = sum(apply(pred_values_1,2,var))
D1G = G1 + P1

G2 = sum((apply(pred_values_2, 2,mean)-data)^2)
P2 = sum(apply(pred_values_2,2,var))

```

```

D2G = G2 + P2

GGs = c(D1G,D2G)

## ALL 4 criterion
cbind(AICs, BICs, DICs,GGs)

# Bayes Factor
require(rmutil)

a = c(0.5,1,9,10,8,9)
b = c(0.01,1,1,1,2,9)

alpha=c(0.1,1,3,2,5,10,40)
beta=c(0.1,1,2,4,14,20,60)

N=10^4
i=4;j=2
ns = 50000

# generate ns parameters from the prior
BFprior1 = rgamma(ns, a[i], b[i])
BFprior2 = rbeta(ns, alpha[j], beta[j])

# density of the parameters
f1 = function(lambda){sum(dpois(data,lambda,log=TRUE))}
f2 = function(prob){sum(dbinom(data,N,prob,log=TRUE))}

# density of the parameters
BFden1 = matrix(0, ns, length(data))
BFden2 = matrix(0, ns, length(data))
for (i in 1:length(data)){
  BFden1[,i] = dpois(data[i],BFprior1,log=TRUE)
  BFden2[,i] = dbinom(data[i],N,BFprior2,log=TRUE)
}

BF1 = apply(BFden1, 1, sum)
BF2 = apply(BFden2, 1, sum)

BF12 = mean(exp(BF1)) / mean(exp(BF2))
BF12
BF21 = mean(exp(BF2)) / mean(exp(BF1))
BF21

cbind(BF12,BF21)

# using marginals
poissongammaaden = function(data,n,a,b) {
  a*log(b) - ((a+sum(data))*(log(n+b))) - sum(log(factorial(data))) +

```



```

(lgamma(a+sum(data))) - (lgamma(a))}

binomialbetaden = function(data,n,N,alpha,beta) {
  sum(log(choose(N, data))) + lgamma(alpha+beta) + lgamma(alpha+sum(data)) +
  lgamma(n*N - sum(data) +beta) -
  lgamma(alpha) - lgamma(beta) - lgamma(n*N + alpha +beta)}

PG = poissongammaden(data,n, a[4], b[4])
BB = binomialbetaden(data,n,N,alpha[2],beta[2])

exp(PG-BB)
exp(BB-PG)

# using the integration of the marginal
poisson_gamma_pdf = function(theta, par)
  # theta: parameter to be integrate
  # par: a list consists of data, alpha and beta
{
  #lambda = par$lambda
  a = par$a
  b = par$b
  data = par$data
  sum(dpois(data, lambda = theta, log=TRUE)) * dgamma(theta, shape = a,
  rate = b, log=TRUE)
}

bino_beta_pdf = function(theta, par)
  # theta: parameter to be integrate
  # par: a list consists of data, N, alpha and beta
{
  N = par$N
  #p = par$p
  data = par$data
  alpha = par$alpha
  beta = par$beta
  sum(dbinom(data, size=N, prob=theta, log=TRUE)) *
  dbeta(theta, shape1=alpha, shape2=beta, log=T)
}

bayes.factor =function(a, b, alpha, beta, N, data)
{
  m1=integrate(poisson_gamma_pdf, 0, 100, par=list(a=a, b=b, data=data))
  m2=integrate(bino_beta_pdf, 0, 1, par=list(N=N, data = data, alpha=alpha,
  beta=beta))
  m1$value / m2$value
}

bayes.factor(10, 1, 1, 1, N, data)

```