Assignment

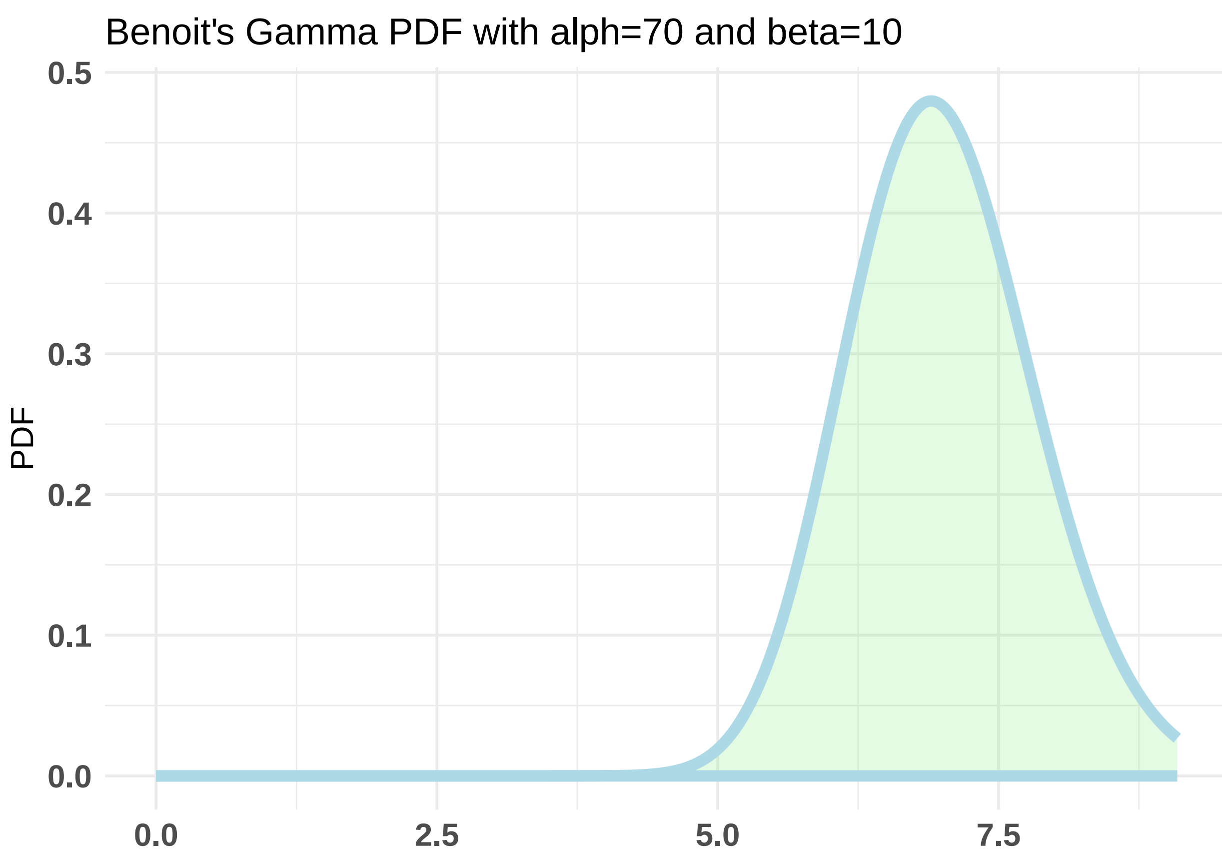
2023-03-17

Question 1

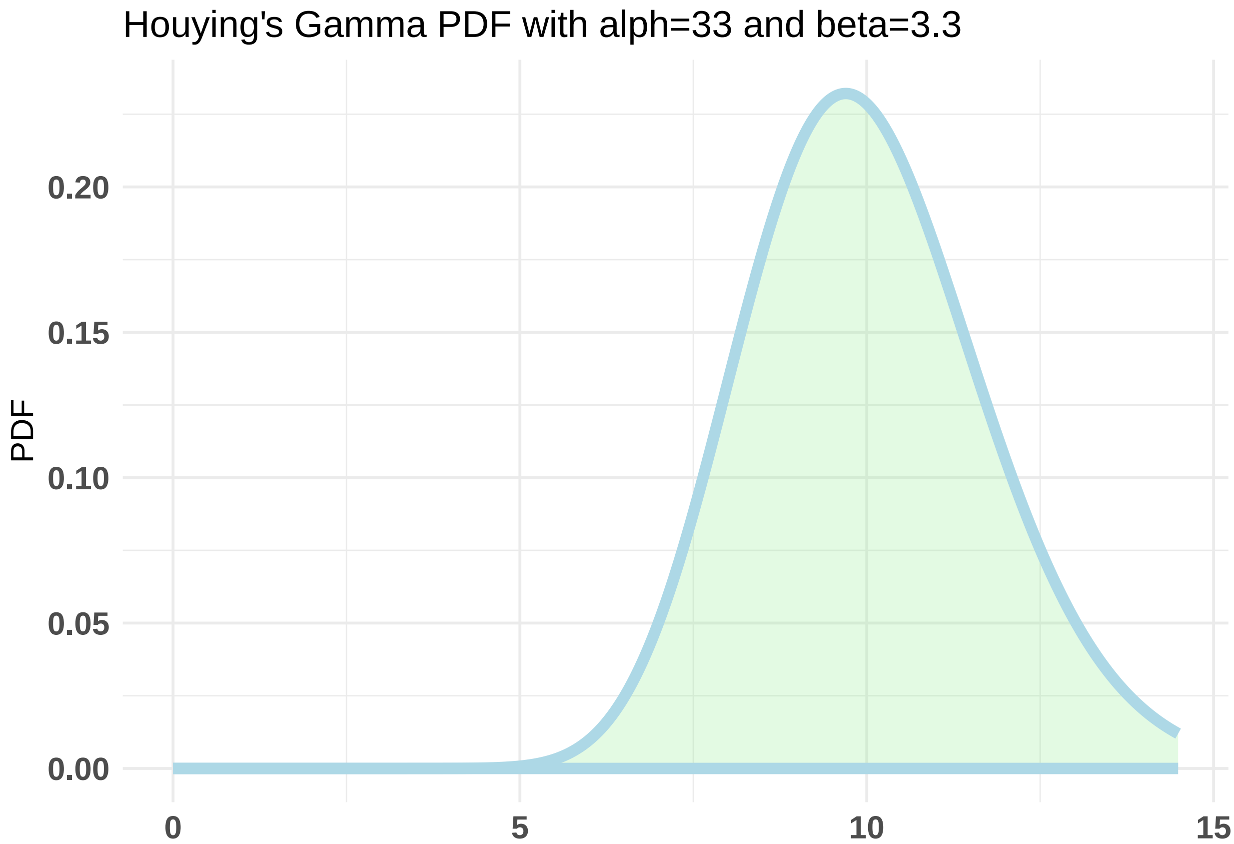
PART 1

(a) Provide a plot representing two Gamma priors.

```
library(bayesAB)
library(ggplot2)
plotGamma(70,10)+ggtitle("Benoit's Gamma PDF with alph=70 and beta=10")
```




```
plotGamma(33,3.3)+ggtitle("Houying's Gamma PDF with alph=33 and beta=3.3")
```



(b) Compare the priors of Benoit and Houying with respect to the averagevalue and spread. Which person believes that there will be more ER visits, on average? Which person is more confident of his/her best guess" at the average numberof ER visits? Provide your reasoning.

Consider the gamma pdf: $G(\lambda, \alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma\alpha}\lambda^{\alpha-1}e^{-}\beta\lambda$ Recall that: $E(\lambda) = \frac{\alpha}{\beta}$ and $Var(\lambda) = \frac{\alpha}{\beta^2}$
We have average and spread as follows:
Benoit $= \begin{array}{cc} Average & Variance \\ \hline 7.0 & 0.7 \end{array}$ and Houying $= \begin{array}{cc} Average & Variance \\ \hline 10.0 & 3.03 \end{array}$

Conclusion:

Houying believes that there will be more ER visits, on average, as her mean is greater than that of Benoit. Benoit is more confident of his best guess on the average number of ER visits t. This is because the variance of 0.7 is less than that of Houying, 3.03, and signifies small deviations of the observations from the mean. Also, Benoit's variance is very samll.

(c) Construct 90% interval estimates for λ using Benoit's prior and Houying's prior.

Recall that the Confidence Interval formula is as follows:

$$CI = \bar{X} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

.

$$\begin{aligned} &\text{Where:} \\ \bar{X} &= E(\lambda, \\ \alpha &= 0.1, \\ Z_{\frac{\alpha}{2}} &= 1.645, \\ \sigma &= \sqrt{v}ar(\lambda), \\ n &= \text{sample size} \end{aligned}$$

Benoits 90% CI

$$= 7.0 \pm 1.645 \frac{0.837}{\sqrt{n}}$$

and
Houying's 90% CI

$$= 10.0 \pm 1.645 \frac{1.74}{\sqrt{n}}$$

(d). After some thought, Benoit believes that his best prior guess at λ is correct, but he is less confident in this guess. Explain how Benoit can adjust the parameters of his Gamma prior to reflect this new prior belief.

By collection of real world data whose probability distribution (PDF) is known, Benoit will be able to estimate his posterior distribution that will, in turn, be used to construct the confidence interval for the parameter λ and ascertain that the true value of the parameter lies within the interval.

(e). Houying also revisits her prior. Her best guess at the average number of ER visits is now 3 larger than her previous best guess, but the degree of confidence in this guess hasn't changed. Explain how Houying can adjust the parameters of her Gamma prior to reflect this new prior belief.

With a posterior distribution in place, Houying woul repeat the estimation experiment severally with varying datasets of larger sample sizes, each time recording the value of the confidence interval of ER visits. Thereafter, an average of the different values of the confidence intervals would give a smaller interval for easier estimation and reduce the variance of the value of ER visits

PART 2.

A hospital collects the number of patients in the emergency room admitted between 10 pm and 11 pm for each day of a week. The following table records the day and the number of ER visits for the given day.

Day	$Number of ER visits$
$Sunday$	8
$Monday$	6
$Tuesday$	6
$Wednesday$	9
$Thursday$	8
$Friday$	9
$Saturday$	7

Suppose one assumes Poisson sampling for the counts and a conjugate Gamma prior with parameters $\alpha = 70$ and $\beta = 10$ for the Poisson rate parameter λ.

(f) .Given the sample shown in the Table, obtain the posterior distribution for λ through the Gamma-Poisson conjugacy. Obtain a 95% posterior credible interval for λ.

Let the prior distribution $p(\lambda) \sim \frac{\beta^\alpha}{\Gamma\alpha}\lambda^{\alpha-1}e^{-\beta\lambda}$
We need to find posterior distribution $f(\lambda|x)$ such that:

$$f(\lambda|x) = \frac{f(x|\lambda).p(\lambda)}{f(x)}$$

Where:
$f(x|\lambda)$ is the likelihood function of data X given λ and
$f(x)$ is the marginal density of X, data summed or integrated over λ

The Likelihood:

$$f(x|\lambda) = \prod_{i=1}^{n} \frac{\lambda^x e^{-\lambda}}{x!} = \frac{\lambda^{\sum_{i=1}^{n} xi} e^{-n\lambda}}{\prod_{i=1}^{n} xi!}$$

The Joint pdf is:

$$f(x|\lambda) \cdot p(\lambda) = \frac{\lambda^{\sum_{i=1}^{n} xi} e^-\lambda}{\prod_{i=1}^{n} xi!} \cdot \frac{\beta^\alpha}{\Gamma\alpha}\lambda^{\alpha-1}e^-\beta\lambda$$

$$= \frac{\beta^\alpha}{\prod_{i=1}^{n} xi!\Gamma\alpha}\lambda^{\alpha+\sum_{i=1}^{n} xi-1}e^{-(\beta+n)\lambda}$$

The marginal density is:

$$f(x) = \int_0^\infty f(x|\lambda).p(\lambda)\, d\lambda$$

$$= \int_0^\infty \frac{\beta^\alpha}{\prod_{i=1}^{n} xi!\Gamma\alpha}\lambda^{\alpha+\sum_{i=1}^{n} xi-1}e^{-(\beta+n)\lambda}\, d\lambda = 1$$

$$= \frac{\beta^\alpha}{\prod_{i=1}^{n} xi!\Gamma\alpha}\int_0^\infty \lambda^{\alpha+\sum_{i=1}^{n} xi-1}e^{-(\beta+n)\lambda}\, d\lambda = 1$$

Recall that:

$$\int_0^\infty G(\lambda, \alpha, \beta) = \int_0^\infty \frac{\beta^\alpha}{\Gamma\alpha}\lambda^{\alpha-1}e^{-\beta\lambda} = 1$$

$$<=> \int_0^\infty \lambda^{\alpha-1} e^- \beta\lambda = \frac{\Gamma\alpha}{\beta^\alpha}$$

Let $\alpha = \alpha + \sum_{i=1}^n xi$ and $\beta = \beta + n$, then

$$<=> \int_0^\infty \lambda^{\alpha+\sum_{i=1}^n xi-1} e^-(\beta+n)\lambda \, d\lambda = \frac{\Gamma(\alpha+\sum_{i=1}^n xi)}{(\beta+n)^{(\alpha+\sum_{i=1}^n xi)}}$$

$$f(x) = \int_0^\infty \frac{\beta^\alpha}{\prod_{i=1}^n xi!\Gamma\alpha} \lambda^{\alpha+\sum_{i=1}^n xi-1} e^-(\beta+n)\lambda \, d\lambda = \frac{\beta^\alpha}{\prod_{i=1}^n xi!\Gamma\alpha} \cdot \frac{\Gamma(\alpha+\sum_{i=1}^n xi)}{(\beta+n)^{(\alpha+\sum_{i=1}^n xi)}}$$

Upon substituting, we get

$$f(\lambda|x) = \frac{f(x|\lambda).p(\lambda)}{f(x)} = \frac{\frac{\beta^\alpha}{\prod_{i=1}^n xi!\Gamma\alpha} \lambda^{\alpha+\sum_{i=1}^n xi-1} e^{-(\beta+n)\lambda}}{\frac{\beta^\alpha}{\prod_{i=1}^n xi!\Gamma\alpha} \cdot \frac{\Gamma(\alpha+\sum_{i=1}^n xi)}{(\beta+n)^{(\alpha+\sum_{i=1}^n xi)}}}$$

Like terms cancels, so we get:

$$f(\lambda|x) = \frac{\lambda^{\alpha+\sum_{i=1}^n xi-1} e^{-(\beta+n)\lambda}}{G(\alpha+\sum_{i=1}^n xi, \beta+n)}$$

$$OR$$

$$f(\lambda|x) = G(\alpha + \sum_{i=1}^n xi, \beta+n) = \frac{(\beta+n)^(\alpha+\sum_{i=1}^n xi)}{\Gamma(\alpha+\sum_{i=1}^n xi)} \cdot \lambda^{\alpha+\sum_{i=1}^n xi-1} e^{-(\beta+n)\lambda}$$

Substituting values of the $\alpha, \beta, \sum_{i=1}^n xi$, the posterior distribution is simplified as:

$$f(\lambda|x) = \frac{(17)^{123}}{\Gamma 123} \lambda^{122} e^{-17\lambda} = \frac{(17)^{123}}{122!} \lambda^{122} e^{-17\lambda}$$

The 95% credible interval for λ is:

```
qgamma(c(0.025, 0.975), shape = 123, rate = 17)
```

```
## [1] 6.013237 8.568716
```

g).Suppose a hospital administrator states that the average number of ER visits during any evening hour does not exceed 6. By computing a posterior probability, evaluate the validity of the administrator's statement.

```
pgamma(6,shape = 123,rate = 17)
```

```
## [1] 0.02368644
```

A probability of 0.02368644 indicates that the administrator's statement is valid by just 2.4%

(h).The hospital is interested in predicting the number of ER visits between 10pm and 11pm for another week. Use simulations to generate posterior predictions of the number of ER visits for another week (seven days).

Using the case when $mean = \frac{\alpha}{\beta}$ then our simulations for anther week is:

```r
Sim<-rgamma(n=7,shape = 123,rate = 17)
Sim
```

```
## [1] 9.072572 7.421906 7.626768 7.348346 8.324085 7.610421 7.997249
```

```r
wk2<-data.frame(Day= c("Sunday","Monday","Tuesday","Wednesday","Thursday","Friday","Saturday"),
                round(Sim,2))
colnames(wk2)=c("Day","Simulated No of ER visits")
wk2
```

```
##         Day Simulated No of ER visits
## 1    Sunday                      9.07
## 2    Monday                      7.42
## 3   Tuesday                      7.63
## 4 Wednesday                      7.35
## 5  Thursday                      8.32
## 6    Friday                      7.61
## 7  Saturday                      8.00
```

Question 2.

PART 1. The Exponential distribution is often used as a model to describe the time between events, such as traffic accidents. A random variable Y has an Exponential distribution if its pdf is as follows.

$$f(y|\lambda) = \{ \ \lambda e^{-\lambda y}, if y \geqslant 0, if y \leqslant 0\}$$

(a).Use the prior distribution $\lambda \sim$ Gamma(a, b), and find its posterior distribution π (λ | y1, · · · , yn), where yi i.i.d. $\sim$ Exponential(λ) for i = 1, · · · , n. Do you recognise a known distribution? Is the Gamma prior a conjugate prior for this model?

Note that:

$$f(y|\lambda) \sim \lambda e^{-\lambda y}$$

,

$$p(\lambda) \sim \frac{b^a}{\Gamma a}\lambda^{a-1}e^{-b\lambda}$$

and

$$\pi(\lambda|y) = \frac{L(y|\lambda)\cdot p(\lambda)}{f(x)}$$

$$L(y|\lambda) = \prod_{i=1}^{n} f(y|\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda y}$$

$$= \lambda^n e^{-\lambda \sum_{i=1}^{n} yi}$$

$$\therefore L(y|\lambda)\cdot p(\lambda) = \lambda^n e^{-\lambda \sum_{i=1}^{n} yi} \cdot \frac{b^a}{\Gamma a}\lambda^{a-1}e^{-b\lambda}$$

$$= \frac{b^a}{\Gamma a}\lambda^{a+n-1}e^{-(b+\sum_{i=1}^{n} yi)\lambda}$$

Now,

$$f(y) = \int_0^\infty \frac{b^a}{\Gamma a}\lambda^{a+n-1}e^{-(b+\sum_{i=1}^{n} yi)\lambda}\, d\lambda = \frac{b^a}{\Gamma a}\int_0^\infty \lambda^{a+n-1}e^{-(b+\sum_{i=1}^{n} yi)\lambda}\, d\lambda$$

Recall that $G(a,b) = \frac{b^a}{\Gamma a}\lambda^{a-1}e^{-b\lambda}$ is a PDF and integrates to 1, that is

$$\int_0^\infty \frac{b^a}{\Gamma a}\lambda^{a-1}e^{-b\lambda}\ d\lambda = 1$$

$$<=> \int_0^\infty \lambda^{a-1}e^{-b\lambda} = \frac{\Gamma a}{b^a}$$

Let $a = a+n, b = (b+\sum_{i=1}^n yi)$.

Thus,

$$\int_0^\infty \lambda^{a+n-1}e^{-(b+\sum_{i=1}^n yi)\lambda}\ d\lambda = \frac{\Gamma(a+n)}{(b+\sum_{i=1}^n yi)^{(a+n)}}$$

$$\therefore f(y) = \frac{b^a}{\Gamma a} \cdot \frac{\Gamma(a+n)}{(b+\sum_{i=1}^n yi)^{(a+n)}}$$

The posterior distribution of is given as:

$$f(\lambda|y) = \frac{\frac{b^a}{\Gamma a}\lambda^{a+n-1}e^{-(b+\sum_{i=1}^n yi)\lambda}}{\frac{b^a}{\Gamma a} \cdot \frac{\Gamma(a+n)}{(b+\sum_{i=1}^n yi)^{(a+n)}}} = \frac{\lambda^{a+n-1}e^{-(b+\sum_{i=1}^n yi)\lambda}}{\frac{\Gamma(a+n)}{(b+\sum_{i=1}^n yi)^{(a+n)}}}$$

$$= \frac{\lambda^{a+n-1}e^{-(b+\sum_{i=1}^n yi)\lambda}}{G(a+n, b+\sum_{i=1}^n yi)}$$

OR

$$= \frac{(b+\sum_{i=1}^n yi)^{(a+n)}}{\Gamma(a+n)} \cdot \lambda^{a+n-1}e^{-(b+\sum_{i=1}^n yi)\lambda}$$

Conclusion:

Yes, I recognize a known posterior distribution, which is Gamma

$$G(shape = (a+n), rate = (b+\sum_{i=1}^n yi)$$

and Gamma prior is indeed a conjugate prior since both the prior distribution of λ and the posterior distribution $f(\lambda|y)$ falls in the same family of distributions.