1 Exercise - Bayesian modelling of discrete data - Solution

(a) Let y_i denote the number of fatal accidents in year i, for $i = 1, ..., 10$, and let θ bet he expected number of accidents in a year. The model for the data is $y_i|\theta \sim \text{Poisson}(\theta)$.

Let's use the conjugate family of distributions for convenience. That is, let $\theta \sim \text{Gamma}(\alpha, \beta)$ and so $\theta|\mathbf{y} \sim \text{Gamma}(\alpha + \sum_{i=1}^{n} y_i, \beta + n)$. We need to set the prior parameters (α, β). Let's assume a noninformative prior distribution $(\alpha, \beta) = (1, 0)$ (corresponds to a uniform flat prior). This is an improper prior but should be okay since we have enough information here: $n = 10$. Then the posterior distribution is $\theta|\mathbf{y} \sim \text{Gamma}(239, 10)$.

Let $\tilde{y}$ be the number of fatal accidents in 1986. Given θ, the predicitive distribution for $\tilde{y}$ is Poisson(θ).

Simulation: Draw θ from $p(\theta|\mathbf{y})$ and then draw $\tilde{y}$ from $p(\tilde{y}|\theta)$.

```
> alpha<-1
> beta<-0
> accid<-c(24,25,31,31,22,21,26,20,16,11)
> theta.post<-rgamma(1000,alpha+sum(accid),beta+length(accid))
> y.1986<-rpois(1000,theta.post)
> quantile(y.1986,c(0.025,0.975))
 2.5% 97.5%
   13    34
```

The computed 95% predictive interval is [13,34].

(b) The estimated number of passenger miles in each year is 'passenger deaths/death rate'. For example, in 1976 the estimated passenger miles is $734/0.19 = 3863 \times 10^{11}$.

Let x_i be the number of passenger miles flown in year i and let θ be the expected accident rate per passenger mile. The model for the data is $p(y_i|x_i, \theta) \sim \text{Poisson}(x_i\theta)$. Again assume the noninformative prior Gamma(1,0) for $p(\theta)$.
Then the posterior distribution is $\theta|\mathbf{y}, \mathbf{x} \sim \text{Gamma}(\alpha + \sum_{i=1}^{n} y_i, \beta + \sum_{i=1}^{n} x_i) = \text{Gamma}(238, 57.159 \times 10^{11})$.

```
> deaths<-c(734,516,754,877,814,362,764,809,223,1066)
> d.rate<-c(0.19,0.12,0.15,0.16,0.14,0.06,0.13,0.13,0.03,0.15)
> miles<-deaths/d.rate
> theta.post<-rgamma(1000,alpha+sum(accid),beta+sum(miles))
> y.1986<-rpois(1000,theta.post*8000)
> quantile(y.1986,c(0.025,0.975))
  2.5%  97.5%
20.975 45.000
```

The computed 95% predictive interval is [21,45].

(c)

```
> theta.post<-rgamma(1000,alpha+sum(deaths),beta+length(deaths))
> y.1986<-rpois(1000,theta.post)
> quantile(y.1986,c(0.025,0.975))
 2.5% 97.5%
  637   750
```

The computed 95% predictive interval is [637,750].

(d)

```
> theta.post<-rgamma(1000,alpha+sum(deaths),beta+sum(miles))
> y.1986<-rpois(1000,theta.post*8000)
> quantile(y.1986,c(0.025,0.975))
    2.5%    97.5%
 905.975 1029.025
```

The computed 95% predictive interval is [906,1029].

(e) Conditional on exposure, it is reasonable to consider the number of fatal accidents by year as independent observations. Case (d) is not appropriate because deaths from the same accident are not independent. Hence, we consider case (b) to be more or less reasonable.