We are going to examine several Poisson models for count data.

- Data: series of coal mine disasters over 112-year history (1851-1962) in the U.K.. The data are characterized as relatively high disaster counts in the early era and low disaster counts in the later era.

- Question of interest: did improvement in technology and safety practices have an actual effect of the rate of serious accidents? When did the change actually occur?

- We will do three things: fit a simple poisson model using Monte Carlo methods, fit a change point poisson model using Gibbs sampler (fun programming!), use a R package (MCMCpack to solve this problem.

**Poisson model** Let $Y_1, Y_2, \ldots, Y_n$ represent a series of count data of mine disasters. We can assume $Y_i$s follow a Poisson distribution with parameter $\lambda$.

$$Y_i \sim \mathcal{P}(\lambda), \ i = 1, \ldots, n$$

$\lambda$ is both the mean and variance of this Poisson distribution.

$$E(Y) = \lambda, \ V(Y) = \lambda$$

Given data, we are interested in estimating $\lambda$.

We can attach $\lambda$ has a prior distribution of Gamma

$$\lambda \sim \mathcal{G}(a_0, b_0), \ a_0 > 0, b_0 > 0$$

The mean of the prior distribution of $\lambda$ is $a_0/b_0$, and the variance is $a_0/b_0^2$.

Gamma is a conjugate prior distribution for the parameter of Poisson distribution. Hence, we can write down the posterior distribution of $\lambda$ as
\[ \lambda | Y_1, \ldots, Y_n \sim G(a, b) \]

where \( a = a_0 + \sum_{i=1}^{n} Y_i \), \( b = b_0 + n \). So the posterior mean of \( \lambda \) (the mean number of mine disasters) is \( a/b \).

There are couple of things we can do to improve the analysis:

- To conduct a sensitivity analysis, change prior values, assess the results.
- Use hyperprior. Instead assigning a value for the prior scale parameter \( b_0 \), we assign a hyperprior for \( b_0 \)
  \[ b_0 \sim IG(u_0, v_0) \]

Now we can find the posterior distribution by iteratively updating prior given hyperprior, posterior given data and prior.

\[ b_0 \sim IG(a_0 + u_0, \lambda + 1/v_0) \]
\[ \lambda | Y_1, \ldots, Y_n \sim G(a_0 + \sum_{i=1}^{n} Y_i, b_0 + n) \]

The following simple R code implement the above algorithms.

```r
# prior mean of lambda
a0 <- 2

# prior variance of lambda
b0 <- 1

### now the prior mean of gamma is the same as the sample mean
```
cat("prior mean of lambda = ", a0/b0, 
")
cat("prior variance of lambda = ", a0/b0^2, 
")
plot(density(rgamma(1000, a0, b0), main="prior distribution of
lambda")
abline(v=a0/b0)

# posterior mean of lambda
a <- a0 + n* mu.Y

# posterior variance of lambda
b <- b0 + n
cat("posterior mean of lambda = ", a/b, 
")
cat("posterior variance of lambda = ", a/b^2, 
")
cat("frequentist estimate of lambda = ", mu.Y, 
")
plot(density(rgamma(1000, a, b), main="posterior distribution of
lambda")

## option
### use hyperior to reduce prior dependency
u0<-0
v0<-1
nchain<-1000
lambda.post<-rep(0,nchain)
b0.post<-rep(0,nchain)

for (i in 1:nchain) {
    lambda.post[i]<-lambda<-rgamma(1, a0+sum(Y), b0+n)
    b0.post[i]<-b0<-1/rgamma(1, a0+u0, lambda+v0)
}
Poisson with two change points (Carlin, Gelfand and Smith, 1992) It looks that there might be a change point. Now let’s assume the change occurs at year $k$. Before year $k$, the count of mine disasters follow a Poisson distribution with mean $\lambda_1$

$$Y_i \sim P(\lambda_1), \quad i = 1, \ldots, k$$

After year $k$, the count of mine disasters follow a second Poisson distribution with mean $\lambda_2$

$$Y_j \sim P(\lambda_2), \quad j = k + 1, \ldots, n$$

Now we have three parameters of interest $\lambda_1$, $\lambda_2$, and $k$ (at the $k$th year change occurs).

For each parameter, we specify a prior distribution

$$\lambda_1 \sim \mathcal{G}(a_0, b_0)$$
$$\lambda_2 \sim \mathcal{G}(c_0, d_0)$$
$$k \sim \mathcal{U}[1, 2, \ldots, n]$$

Now we can write down the posterior distribution (up to a normalization factor)

$$P(\lambda_1, \lambda_2, k|Y) = \frac{f(Y|\lambda_1, \lambda_2, k)P(\lambda_1)P(\lambda_2)P(k)}{\int f(Y|\lambda_1, \lambda_2, k)P(\lambda_1)P(\lambda_2)P(k)\,dk}$$

$$\propto f(Y|\lambda_1, \lambda_2, k)P(\lambda_1)P(\lambda_2)P(k)$$

$$\propto \left( \prod_{i=1}^{k} \frac{e^{-\lambda_1}Y_i}{y_i!} \right) \left( \prod_{i=k+1}^{n} \frac{e^{-\lambda_2}Y_i}{y_i!} \right) \left( \frac{b_0^{a_0}}{\Gamma(a_0)} \lambda_1^{a_0-1}e^{-b_0\lambda_1} \right) \left( \frac{d_0^{c_0}}{\Gamma(c_0)} \lambda_2^{c_0-1}e^{-d_0\lambda_2} \right)$$

Trick: in conditional posterior distribution, the parameters that are conditioned on are treated as constants.

Now the conditional posterior distribution of $\lambda_1$ given data $Y$ and $k$, $\lambda_2$ is

$$P(\lambda_1|Y_1, \ldots, Y_n, \lambda_2, k) =$$

$$P(\lambda_1|Y_1, \ldots, Y_k) \sim \mathcal{G}(a_0 + \sum_{i=1}^{K} Y_i, b_0 + k)$$

Similarly, the conditional posterior distribution of $\lambda_2$ given data $Y$ and $k$, $\lambda_1$ is
The conditional posterior distribution of $k$ can be simplified as

\[ P(k|Y, \lambda_1, \lambda_2) \propto e^{-k(\lambda_1 - \lambda_2)}(\lambda_1/\lambda_2) \sum_{i=1}^{k} Y_i \]

Note the right hand side of the above equation is not a probability distribution but it is proportional to the actual distribution. We call it the "kernel" of the distribution, since it is only different by a constant. To draw from this distribution (it is discrete, how convenient!), we can calculate RHS for $k = 1, \ldots, n$, and rescale these quantities by their sum, then we get the actual conditional posterior distribution. If the above distribution is continuous, we need to use rejection sampling techniques.

The following R code implement the above algorithms.

```r
##############################################
a0<-2
b0<-1
c0<-2
d0<-1

nchain<-100
lambda1.post<-rep(0, nchain)
lambda2.post<-rep(0, nchain)
k.post<-rep(0, nchain)

lambda1<-rgamma(1, a0, b0)
lambda2<-rgamma(1, c0, d0)
k<-sample(1:n, 1)

for (s in 1:nchain) {
    lambda1.post[s]<-lambda1<-rgamma(1, a0+sum(Y[1:k]), b0+k)
    lambda2.post[s]<-lambda2<-rgamma(1, c0+sum(Y[(k+1):n]), d0+n-k)
    print(lambda1-lambda2)
}
```
###conditional posterior distribution of k

```r
k.post.dist <- function(t1, t2, X) {
    n <- length(X)
    post.dist <- rep(0, n)
    for (i in 1:n)
        post.dist[i] <- exp(-(t1-t2)*i)*(t1/t2)^sum(X[1:i])
    post.dist <- post.dist/sum(post.dist)
    return(post.dist)
}
```

Using **MCMCpack** In R package **MCMCpack**, Martin and Quinn implemented a different method to model changepoint in Poisson process (Chib, 1998). Chib’s method allows us to model more than 2 change points.

```r
library(MCMCpack)
model1 <- MCMCpoissonChangepoint(Y, m=1, c0=6.85, d0=1, verbose = 10000,
                                 marginal.likelihood="Chib95")
model2 <- MCMCpoissonChangepoint(Y, m=2, c0=6.85, d0=1, verbose = 10000,
                                 marginal.likelihood="Chib95")
model3 <- MCMCpoissonChangepoint(Y, m=3, c0=6.85, d0=1, verbose = 10000,
                                 marginal.likelihood="Chib95")

print(BayesFactor(model1, model2, model3))
## Draw plots using the "right" model
plotState(model1)
```
Bayesian model selection  How to select among different models? We can Bayes Factor.

$$\kappa = \frac{P(Y|\text{model 1})}{P(Y|\text{model 2})}$$

$P(Y|\text{model 1})$ is called the marginal likelihood of model 1. It is calculated by averaging over likelihood given different values of parameters weighted by their posterior distribution.

$$P(Y|\text{model 1}) = \int P(\theta|Y, \text{model 1})P(Y|\theta, \text{model 1})d\theta$$

<table>
<thead>
<tr>
<th>Ratio</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>less than 1:1</td>
<td>favor model 2</td>
</tr>
<tr>
<td>3:1</td>
<td>Barely worth mentioning</td>
</tr>
<tr>
<td>10:1</td>
<td>Substantial evidence favor model 1</td>
</tr>
<tr>
<td>30:1</td>
<td>Strong</td>
</tr>
<tr>
<td>100:1</td>
<td>Very strong</td>
</tr>
<tr>
<td>more than 100:1</td>
<td>Decisive</td>
</tr>
</tbody>
</table>

Table 1: Use Bayes factor to select models

Reference
