Bayesian Analysis of Movie Body Count

Group E: Fengmei Liu, Nate Kotila, Jun Wang, Nazia Khan

1 Problem Statement

This project is to analyze kill rate per hour in Quentin Tarantino's movie , with the ultimate goal to predict body count in his next movie. We did the counts in one hour prediction with kill rate per hour analysis, and the movie length prediction separately, then combine these two independent variables together for the next movie body count prediction.

2 Sampling Model

For this body count problem, every death outcome should be an integer, and our interest is the kill rate per hour, which is a positive real number. Thus, let y_i denote the number of deaths in movie i, t_i denote the length of movie i and let λ denote the kill rate per hour. We can define our sample space: $\{y \in \mathbb{N}\}$, the parameter space: $\{\lambda \in \mathbb{R}^+\}$.

Consider that y_is are exchangeable because the order-release year of each movie doesn't matter and we assume the movie population is way larger than the sample size, y_i are conditionally i.i.d given the underlying distribution parameter.
We firstly think about Poisson model. However, for Poisson model, we know that $E(Y) = Var(Y)$. This makes the Poisson model very restrictive on the applications. Due to the heterogeneity in the movie, deaths in different intervals could be different, similarly, for different movies, the deaths per hour will be different. As suggested by Gelman(BDA3), in practice, the counts problem are often overdispersed, with variance greater than mean, so he suggested that the negative binomial distribution is a robust alternative to Poisson.
We justify the model would meet the following assumptions.
A1: 0 deaths occurred at the start of the movie.
A2: For all deaths, and for any two time intervals, I1 and I2, of equal length, Pr(deaths in I1) not necessarily the same as Pr(deaths in I2)
A3: Deaths that occur in non-overlapping time intervals are not mutually independent.
A4: More then one deaths can't happen at the same time in the movie, even if people were killed at the same scenery, there's still tiny time difference of the deaths.
A5: Probability of no death in a movie $\in (0,1)$ i.e. $0 < Pr\{y(t) = 0\} < 1$ for all $t > 0$.
As we argued above, A2 and A3 of Poisson were replaced by heterogeneity, by adding dispersion parameter. We get Negative Binomial Distribution

$$p(y_i|\alpha, \beta, t_i) = \frac{\Gamma(\alpha + y)}{\Gamma(\alpha) y!} \left(\frac{\beta}{\beta + t_i}\right)^{\alpha} \left(\frac{t_i}{\beta + t_i}\right)^{y_i}, \quad \alpha > 0, \beta > 0, t_i > 0, y_i = 0, 1, 2, ... \tag{1}$$

Here, Negative Binomial distribution $NB(\alpha, \beta)$ is a mixture of $Possion(\lambda)$, with λ follow the Gamma distribution, $Gamma(\alpha, \beta)$.
For the reason of including kill rate as our interest parameter, we reparameterize the model to be $NB(\lambda, \phi)$, of which λ is the mean, and ϕ^{-1} is the overdispersion level(ODL).
y_i denotes the number of death in i^{th} movie in one hour. so $t_i = 1$ here.
The model is re-written as:

$$p(y_i|\lambda, \phi) = \frac{\Gamma(\phi^{-1} + y)}{\Gamma(\phi^{-1}) y!} \left(\frac{1}{\lambda\phi + 1}\right)^{\phi^{-1}} \left(\frac{\lambda}{\lambda + \phi^{-1}}\right)^{y_i}, \quad \lambda > 0, \phi > 0, y_i = 0, 1, 2, ... \tag{2}$$

With this model, we have two parameters λ and ϕ, This ϕ in the variance expression allows us to construct a more accurate model for the counts. we can easily tell the conjugate prior of λ, but for ϕ it's very complex to get the conjugate prior, We reviewed some literature in which polynomial approximation was used for the prior,but it ends up chunky and less feasible to utilize. As suggested by (Zhou and Carin, 2013), in their paper, maximum likelihood estimation could be utilized for the purpose of getting ϕ. Also, in our problem, ϕ is a nuisance parameter, so we use an likelihood estimation to approximate it, then find the conjugate prior of key parameter λ.

3 Data

The "historical data": after searching similar movies to Tarantino, we use the information from website ScoopWhoop "If you love Quentin-Tarantino, also check out these 15 movies". Which should provide pretty good knowledge of the prior about Tarantino's movies, as they emulate his style very close. We gathered the information about 10 such movies.
The sampling data: we used the information about 8 movies from the inforgraphic. We kept "The hateful eight" for validating the model. This movie is not included in the sampling data.

Name	Death	Movie Length	Kill Per Hour
Battle Royale	54	2.03	27
The Boondock	34	1.83	19
City on Fire	10	1.75	6
True Romance	20	2.02	10
Audition	2	1.92	1
Killing Zoe	20	1.63	12
Rubber	27	1.42	19
Django	180	2.75	65
Lucky Number Slevin	19	1.83	10
The Way of The Gun	20	1.98	10

Table 1: Historical data

Name	Death	Movie Length	Kill Per Hour
Reservoir Dogs	11	1.67	7
Pulp Fiction	7	2.97	2
Jackie Brown	4	2.57	2
Kill Bill vol.1	62	1.87	33
Kill Bill vol.2	13	2.3	6
Death Proof	6	2.12	3
Inglourious Basterds	396	2.55	155
DjanGo Unchained	64	2.75	23

Table 2: Sample data

4 Prior

4.1 Maximum Likelihood Estimation of ϕ

(Piegorsch, 1990) gave the way to get the MLE estimation of ϕ, compared to the Pseudo-Likelihood in Robinson&Simith's paper, we get reasonable estimation. The MLE estimator of ϕ is derived by :

$$\nabla_{\phi} l = \frac{1}{n}\sum_{i=1}^{n}\sum_{\nu=0}^{y_i-1}\frac{\nu}{1+\phi\nu} + \phi^{-2}log\{1+\phi\lambda\} - \frac{\lambda(\bar{y}+\phi^{-1})}{1+\phi\lambda} \tag{3}$$

solve by set $\nabla_{\phi} l = 0$ and $\hat{\lambda} = \bar{y}$, plug in our data, we get $\hat{\phi} = 0.67$

4.2 Conjugate prior for λ

First, we can prove that NB(λ, ϕ) is an exponential family, see attached file.
The conjugate prior

$$\begin{aligned} p(\lambda|\alpha_0, \beta_0) &\propto \left(\frac{1}{\lambda\phi+1}\right)^{\alpha_0-1}\left(\frac{\lambda}{\lambda+\phi^{-1}}\right)^{\beta_0-1} \\ p(\lambda|\alpha_1, \beta_1) &\propto (\lambda\phi)^{\alpha_1-1}(1+\lambda\phi)^{-\alpha_1-\beta_1} \end{aligned} \tag{4}$$

The closed form would be: $p(\lambda|\alpha_1,\beta_1) = \frac{\Gamma(\alpha_1+\beta_1)}{\Gamma(\alpha_1)\Gamma(\beta_1)}(\lambda\phi)^{\alpha_1-1}(1+\lambda\phi)^{-\alpha_1-\beta_1}$
The prior $p(\lambda|\alpha_1,\beta_1) \sim$ genearlized beta prime distribution, To estimate α_1, β_1, we use historical data to get the moment matching.

$$\begin{aligned} E(\lambda) &= \frac{\alpha_1}{(\beta_1-1)\phi} = \text{expected count per hour} = 22.8 \\ Var(\lambda) &= \frac{\alpha(\alpha_1+\beta_1-1)}{(\beta_1-1)^2(\beta_1-2)\phi^2} = \text{variance of counts per hour} = 561.36 \end{aligned} \tag{5}$$

plug in $\hat{\phi} = 0.67$, solve to get $\alpha_1 = 30.6$, $\beta_1 = 3$. So we have the prior $0.67\lambda \sim \beta'(30.6, 3, 1, 1)$, $\lambda \sim \beta'(30.6, 3, 1, \frac{1}{0.67})$

5 Posterior

5.1 Posterior of λ

Using the prior and likelihood, the posterior of λ is

$$\begin{aligned} p(\lambda|y,\hat{\phi}) &\propto \left(\lambda\hat{\phi}\right)^{\alpha_1+\sum_{i=1}^{8} y_i-1}\left(1+\lambda\hat{\phi}\right)^{-\alpha_1-\beta_1-\sum_{i=1}^{8} y_i-8\hat{\phi}^{-1}} \\ &\propto \left(\lambda\hat{\phi}\right)^{261.6-1}\left(1+\lambda\hat{\phi}\right)^{-261.6-15} \end{aligned} \tag{6}$$

So we have, $0.67\lambda|y \sim \beta'(261.6, 15, 1, 1)$, $\quad \lambda|y \sim \beta'(261.6, 15, 1, \frac{1}{0.67})$

5.2 Highest Posterior Density Region of λ

The 95% HDI region is : (15.7, 44.7). We are mostly sure the kill rate per hour is between 15.7-44.7

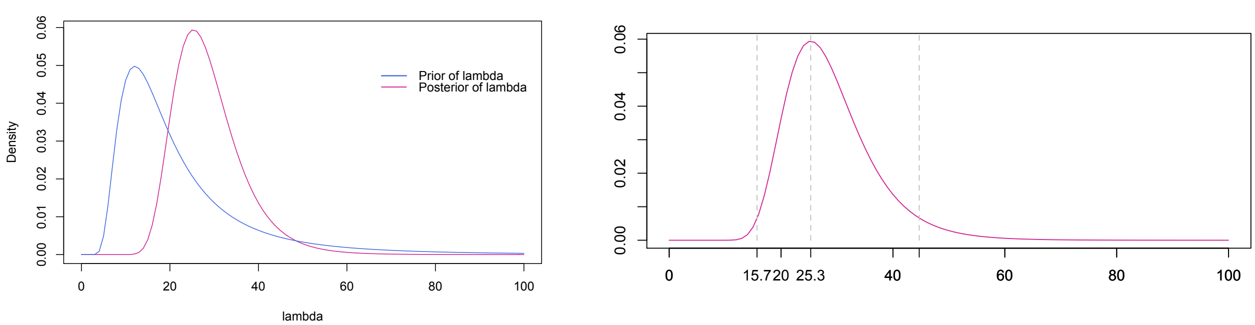


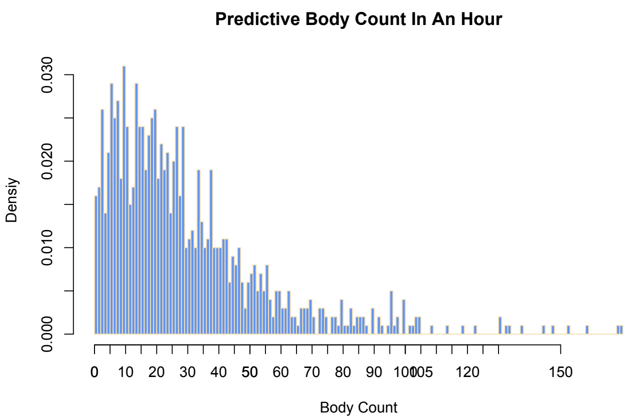
Figure 1: Posterior Distribution of $p(\lambda|y)$; Right: Predictive Body Count An Hour

From the prior-posterior graph, posterior kill rate per hour shifted to the right. The prior similar movies give us some idea about Tarantino's style, but the true data of his movies affected the result, making our belief about the kill rate more concentrated and shift right, generally we update the belief of the most possible kill rate to be larger, around 25.3 as shown in right plot.

5.3 Posterior predictive distribution

$$p(\tilde{y}|y) = \int_0^\infty p(\tilde{y}|\lambda)p(\lambda|y)d\lambda \tag{7}$$

Using simulation method, for λ do 1000 draws from beta prime with $rpearsonVI(1000, 247, 14, 1, 1/0.67)$, then for y draw from negative binomial $rnbinom(1000, size = 1/0.67, mu = \lambda)$. The most possible value for death in an hour is 13. The 95% confidence interval of the deaths are between 0-81.



5.4 Predictive Posterior Checking

The data has one extreme large number, test the minimum, maximum $T_1 = Min(Y), T_2 = Max(Y)$. Test the mean and variance $T_3 = mean(Y), T_4 = sd(Y)$.As we made the assumption of the variance bigger than mean, so first we test $T_5 = var(Y) - mean(Y)$, Also we believe the model should have long right tail, so test $T_6 = |Max(Y) - Mean(Y)| - |Min(Y) - Mean(Y)|$. The test shows our model can reflects minimum and mean well, it describes some discrepancy between our model and the data in the maximum, s.d, because the data has one extreme large point. Variance-mean test and right long tail test also show the model can fit the data, although not very well.Considering the extreme points in our data, we think all the tests show our model can fit the data well.

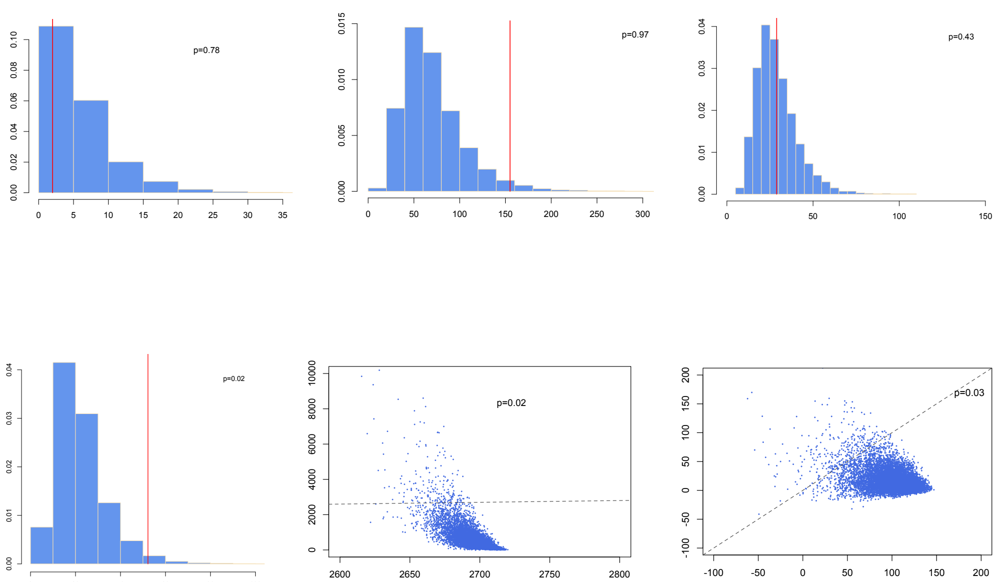


Figure 2: Predictive Posterior Checking Test- Min, Max, Mean, SD, Var-Mean, Long right tail

6 Movie Length Distribution

Sampling: $l \sim Normal(2, \sigma^2)$ We compared the quantiles of sample length to the quantiles of Normal distribution. We see that the sample length seems to be from the Normal distribution family.

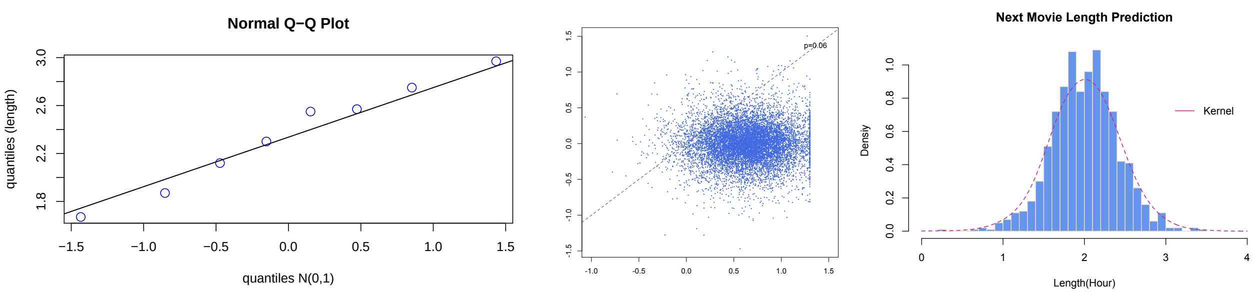
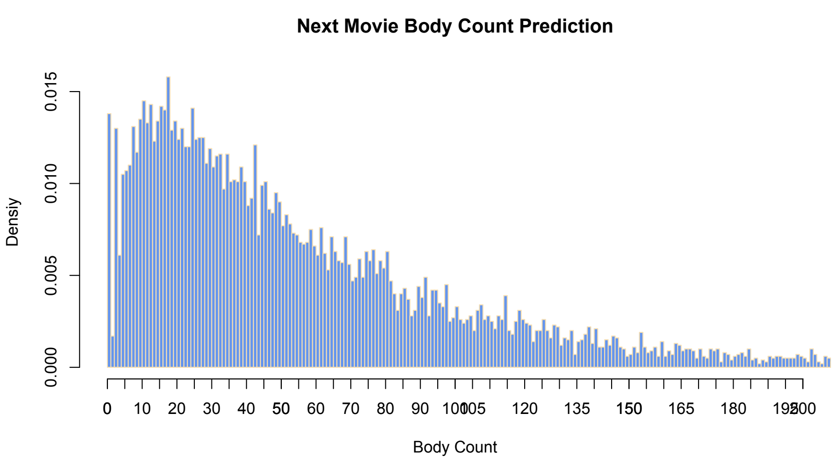


Figure 3: Movie Length QQ plot and Posterior Predictive Simulation

Suppose the movie duration l(hour) to be normal distributed with mean=2, and unknown.
Prior: $\sigma^2 \sim Inv - chi(\nu_0, \sigma_0^2)$ As shown in textbook (Gelman,BDA3) $\sigma^2 \sim Inv - Chi(\nu_0, \sigma_0^2)$
Posterior: $\sigma^2|l \sim Inv - Chi(\nu_0 + n, \frac{\nu_0\sigma_0^2 + n\nu}{\nu_0 + n})$ ν_0 is the prior degree of freedom , 10-1=9, σ_0^2 is the scale $\sigma^2|l \sim Inv - Chi(17, 0.1391)$
The posterior predictive checking has been down on minimum, maximum and symmetric test, only the plot of symmetric test is shown in Fig.4.We can say the normal model can be used for this data. Posterior predictive distribution.$p(\tilde{l}|l) = \int_0^\infty p(\tilde{l}|\sigma^2)p(\sigma^2|l)d\sigma^2$, Draw 1000 σ^2 from the Inv-Chi distribution and then draw normal distribution with $\mu = 2, \sigma^2$. The simulated result are shown in Fig.4.

7 Prediction of body count in next movie

The movie length and movie body counts are independent, the deaths are not correlated with the movie length. $p(\tilde{n}|y, l) = p(\tilde{y}\tilde{l}|y, l) = p(\tilde{y}|y) * p(\tilde{l}|l)$.
Using "The hateful eight", which has 19 body count as the test data.From this simulation of the prediction, We have the expected value 57, variance 2895. 80% confidence interval (0,163). And the most likely body count in the next movie is 17. "The hateful eight" fits in this simulation prediction well.



References

[1] Gary King, Variance Specification in Event Count Models: From Restrictive Assumptions to a Generalized Estimator

[2] Zhou Mingyuan, Lawrence Carin, 2013, Negative Binomial Process Count and Mixture Modeling

[3] Mark D Robinson, Gordon k Smith, 2007, Small-sample estimation of negative binomial dispersion, with applications to SAGE data.

[4] C.I.Bliss, R.A. Fisher, 1953, Fitting the negative binomial distribution to biological data

[5] Andrew Gelman, BDA3

[6] Walter Piegorsch, 1990, Maximum Likelihood Estimation for Negative Binomial Dispertion Parameter

8 Appendix

8.1 MLE calculation of ϕ in R

```
temp <- 0
v <- movie$KillRate

phi <- function(x){
   for (i in 1:length(v)){
      for(j in 0:(v[i]-1)){
         temp <- temp + j/(1+x*j)
      }
   }
   1/length(v)*temp + x^(-2)*log(1+x*mean(v)) -
       mean(v)*(mean(v)+x^(-1))/(1+x*mean(v))
}

xstart <- c(0.1,0.99)
nleqslv(xstart, phi, control=list(trace=1,btol=.01,delta="cauchy"))
```

8.2 Negative Binomial Exponential Family Proof

$$p(y|\lambda,\phi) = \frac{\Gamma(\phi^{-1}+y)}{\Gamma(\phi^{-1})y!}\left(\frac{1}{\lambda\phi+1}\right)^{\phi^{-1}}\left(\frac{\lambda}{\lambda+\phi^{-1}}\right)^{y}, \quad \lambda > 0, \phi > 0, y_i = 0,1,2,...$$

$$p(y|\lambda,\phi) = \frac{\Gamma(\phi^{-1}+y)}{\Gamma(\phi^{-1})y!}\left(\frac{1}{\lambda\phi+1}\right)^{\phi^{-1}} \exp\left(y\log\left(\frac{\lambda}{\lambda+\phi^{-1}}\right)\right)$$

Since, we know $\phi = 0.67$

$$p(y|\lambda,\phi) = f(y).g(\lambda).\exp\left[t(y).h(\lambda)\right]$$

where $f(y) = \frac{\Gamma(\phi^{-1}+y)}{\Gamma(\phi^{-1})y!}$; $\quad g(\lambda) = \left(\frac{1}{\lambda\phi+1}\right)^{\phi^{-1}}$; $\quad t(y) = y$; $\quad h(\lambda) = \log\left(\frac{\lambda}{\lambda+\phi^{-1}}\right)$

This is a one-parameter exponential family.