

# pstat115hw1

Hedi Xia, Kevin Goode

## 3

(a)

```
p_1 = 1 - data.matrix(sentiment_frequencies_mat)[17, 2]
p_2 = 1 - data.matrix(sentiment_frequencies_mat)[8, 2]
n1 = data.matrix(sentiment_frequencies_mat)[17, 3]
n2 = data.matrix(sentiment_frequencies_mat)[8, 3]
sd_diff = sqrt(p_1 * (1 - p_1) / n1 + p_2 * (1 - p_2) / n2)
mean_diff = p_1 - p_2
sprintf("The lower limit is %s and the upper limit is %s.",
        mean_diff - 1.96 * sd_diff, mean_diff + 1.96 * sd_diff)
```

```
## [1] "The lower limit is 0.157734830363307 and the upper limit is 0.295204912436079."
```

The lower limit is 0.157734830363307 and the upper limit is 0.295204912436079. 0 is not contained in the interval. We are 95 percent confident that the sentiment changes between these two chapters.

(b)

95 percent is the probability that the confidence interval we calculate contains the actual difference of frequencies.

(c)

It makes sense since C17 is much negatively sentimented than C8 in our estimates, and in actual circumstance “murder” is more negative than “wedding”.

## 4

(a)

Data Generating Process: generate  $\lambda_i$  with distribution  $\Gamma(\alpha, \beta)$  and then generate  $Y_i$  with  $Poisson(\lambda_i)$ . Assuming  $\alpha = 10$  and  $\beta = 1$ ,

```
lambdas = rgamma(1000, 10, 1)
ys = rep(lambdas)
for (i in 1:length(ys)){ys[i] = rpois(1, ys[i]);}
sprintf("The empirical mean is %s, the empirical variance is %s.", mean(ys), var(ys))
```

```
## [1] "The empirical mean is 9.985, the empirical variance is 22.3471221221221."
```

Theoretically the mean and variance of poisson distribution is the same. However, in our case it is not, while the variance is around two times the mean. The mean is around 10, which is the mean of distribution of  $\lambda_i$ , but the variance is around twice the variance of  $\lambda_i$ .

(b)

$$\begin{bmatrix} & Known & Unknown \\ Variables & \lambda_i & Y_i \\ Constants & y_i, n, \alpha, \beta & \end{bmatrix}$$

(c)

Notice that

$$P(\lambda_i|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda_i^{\alpha-1} e^{-\beta\lambda_i}$$

And

$$P(Y_i = x|\lambda_i) = \frac{\lambda_i^x e^{-\lambda_i}}{x!}$$

If we multiply them together,

$$p(Y_i, \lambda_i|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda_i^{x+\alpha-1} e^{-(\beta+1)\lambda_i}$$

Notice that

$$\int_{d\lambda_i} \frac{(\beta+1)^{x+\alpha}}{\Gamma(x+\alpha)} \lambda_i^{x+\alpha-1} e^{-(\beta+1)\lambda_i} = 1$$

Therefore,

$$p(Y_i = x|\alpha, \beta) = \int_{d\lambda_i} p(Y_i, \lambda_i|\alpha, \beta) = \frac{\beta^\alpha}{(\beta+1)^{x+\alpha}} \binom{x+\alpha-1}{\alpha-1}$$

Or we can write it as

$$p(Y_i|\alpha, \beta) = \frac{\beta^\alpha}{(\beta+1)^{Y_i+\alpha}} \binom{Y_i+\alpha-1}{\alpha-1}$$

(d)

It is a Negative binomial distribution  $NB(\alpha, 1/(\beta+1))$