

Stat 516

Homework 5

Alden Timme and Marco Ribeiro

Due Date: Thursday, November 13

1

- (a) For the model $Y|\theta \sim \text{Poisson}(E \times \theta)$, where E is the “expected” number of cases, Y is the count of disease cases, and $\theta > 0$ is the relative risk, we have the likelihood function

$$L(\theta) = p_\theta(Y = y) = \frac{(E\theta)^y}{y!} e^{-E\theta}$$

which yields the log likelihood

$$\ell(\theta) = \log L(\theta) = y \log(E\theta) - \log(y!) - E\theta$$

To find Fisher’s (expected) information $I(\theta)$, we use $I(\theta) = -\mathbb{E} [\ddot{\ell}(\theta)]$,

$$\begin{aligned}\ell(\theta) &= y \log E + y \log \theta - \log(y!) - E\theta \\ \dot{\ell}(\theta) &= \frac{y}{\theta} - E \\ \ddot{\ell}(\theta) &= -\frac{y}{\theta^2} \\ I(\theta) &= -\mathbb{E} [\ell''(\theta)] = -\mathbb{E} \left[-\frac{Y}{\theta^2} \right] = \frac{1}{\theta^2} \mathbb{E}[Y] = \frac{E}{\theta}\end{aligned}$$

Maximizing $\ell(\theta)$ to find the MLE, we have

$$\hat{\theta} = \frac{y}{E}$$

The variance of the MLE $\hat{\theta}$ is then given by

$$\text{Var}(\hat{\theta}) = I(\theta)^{-1} = \frac{\theta}{E}$$

- (b) If we assume a prior of $\theta \sim \text{Gamma}(a, b)$, we have

$$\begin{aligned}p(\theta|y) &\propto p(y|\theta)p(\theta) \\ &\propto \theta^y e^{-E\theta} \theta^{a-1} e^{-b\theta} \\ &= \theta^{y+a-1} e^{-(E+b)\theta}\end{aligned}$$

so we see $\theta|y \sim \text{Gamma}(y + a, E + b)$.

- (c) When we see $y = 4$ cases of leukemia with an expected number $E = 0.25$, the MLE is

$$\hat{\theta} = \frac{y}{E} = 16$$

with variance

$$\text{Var}(\hat{\theta}) = \frac{\theta}{E} = 64$$

Since the MLE is asymptotically normal, we can approximate the 95% confidence interval with a normal distribution,

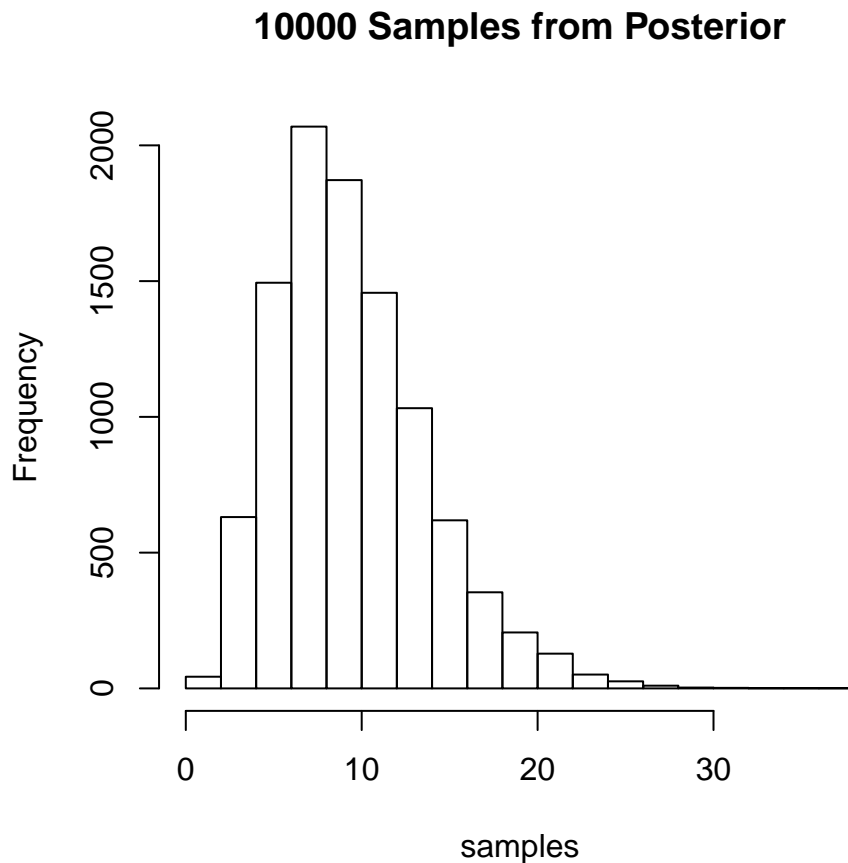
$$\hat{\theta} \pm 1.96 \times \sqrt{\text{Var}(\hat{\theta})} = 16 \pm 1.96 \times 8 \approx (0.32, 31.68)$$

- (d) To find the a and b which give a gamma prior with 90% interval $[0.1, 10]$, we use the R function `optim`,

```
priorch <- function(x, q1, q2, p1, p2) {  
  (p1 - pgamma(q1, x[1], x[2]))^2 + (p2 - pgamma(q2, x[1], x[2]))^2  
}  
opt <- optim(par=c(1,1), fn=priorch, q1=0.1, q2=10, p1=0.05, p2=0.95)  
a <- opt$par[1]  
b <- opt$par[2]
```

yielding $a = 0.8405$ and $b = 0.2679$.

Using this prior and the data from part (c), we arrive at a posterior of $\theta|y \sim \text{Gamma}(4.8405, 0.5179)$. Sampling from this distribution 1000 times, we get the following histogram,



A 95% credible interval using the 0.025- and 0.975-quantiles is (2.9642, 19.3296).

- (e) Using the 95% confidence interval from the MLE and its variance, we would say there is *not* evidence of excess risk for these data, because the value $\theta = 1$ corresponding to “null” risk is contained in the asymptotic 95% confidence interval. From the Bayesian analysis, we would say there *is* evidence of excess risk for these data, because the 95% credible interval does not contain $\theta = 1$ (“null” risk).

2

- (a) MLEs of p_{12} p_{21} are given by

$$\hat{p}_{12} = \frac{n_{12}}{n_{1+}} = \frac{223}{588} \approx 0.3793$$

$$\hat{p}_{21} = \frac{n_{21}}{n_{2+}} = \frac{250}{456} \approx 0.5482$$

The asymptotic variances of \hat{p}_{12} and \hat{p}_{21} are then

$$\widehat{\text{Var}}(\hat{p}_{12}) \approx \frac{\hat{p}_{12}(1 - \hat{p}_{12})}{n_{1+}} \approx 4.0037 \times 10^{-4}$$

$$\widehat{\text{Var}}(\hat{p}_{21}) \approx \frac{\hat{p}_{21}(1 - \hat{p}_{21})}{n_{2+}} \approx 5.4314 \times 10^{-4}$$

Using these estimates, we can form 95% asymptotic confidence intervals using the asymptotic normality of the MLEs,

$$\text{CI}_{12} = \hat{p}_{12} \pm 1.96 \times \sqrt{\widehat{\text{Var}}(\hat{p}_{12})} \approx (0.34, 0.4185)$$

$$\text{CI}_{21} = \hat{p}_{21} \pm 1.96 \times \sqrt{\widehat{\text{Var}}(\hat{p}_{21})} \approx (0.5026, 0.5939)$$

- (b) With independent uniform priors, we have $p_{12} \sim \text{Beta}(1, 1)$ and $p_{21} \sim \text{Beta}(1, 1)$. The posterior distributions are then

$$p_{12}|n_{1\cdot} \sim \text{Beta}(n_{12} + 1, n_{11} + 1)$$

$$p_{21}|n_{2\cdot} \sim \text{Beta}(n_{21} + 1, n_{22} + 1)$$

The posterior medians are then

$$\tilde{p}_{12} \approx \frac{n_{12} + 1 - 1/3}{n_{1+} - 2/3} \approx 0.3795$$

$$\tilde{p}_{21} \approx \frac{n_{21} + 1 - 1/3}{n_{2+} - 2/3} \approx 0.5481$$

A 95% credible interval for each parameter can then be obtained from the 0.025- and 0.975-quantiles of each parameter’s respective posterior,

$$\text{CI}_{12} \approx (0.3409, 0.4192)$$

$$\text{CI}_{21} \approx (0.5023, 0.5933)$$

- (c) Using a likelihood ratio test, we want to test the null hypothesis H_0 that the weather on each day is independent versus H_1 that the weather on one day depends on the weather the previous day. Under H_0 , $\hat{p}_j = n_j/n$, whereas under H_1 , $\hat{p}_{ij} = n_{ij}/n_{i+}$. Using the likelihood ratio test statistic, we have

$$T = 2(\hat{\ell}_1 - \hat{\ell}_0)$$

$$= 2 \sum_{i=1}^2 \sum_{j=1}^2 n_{ij} \log \left(\frac{\hat{p}_{ij}}{\hat{p}_j} \right)$$

$$= 2 \sum_{i=1}^2 \sum_{j=1}^2 n_{ij} \log \left(\frac{n_{ij}/n_{i+}}{n_j/n} \right)$$

$$\approx 25.0195$$

Under the null, $T \sim \chi_1^2$, and we have $P(T > 25.0195) \approx 5.6753 \times 10^{-7}$, so the likelihood ratio test would reject the null hypothesis that the weather on each day is independent.
Under the Bayesian paradigm, ...