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 $I(\theta) = -\mathbb{E}\left[\ddot{\ell}(\theta)\right]$,

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

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

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

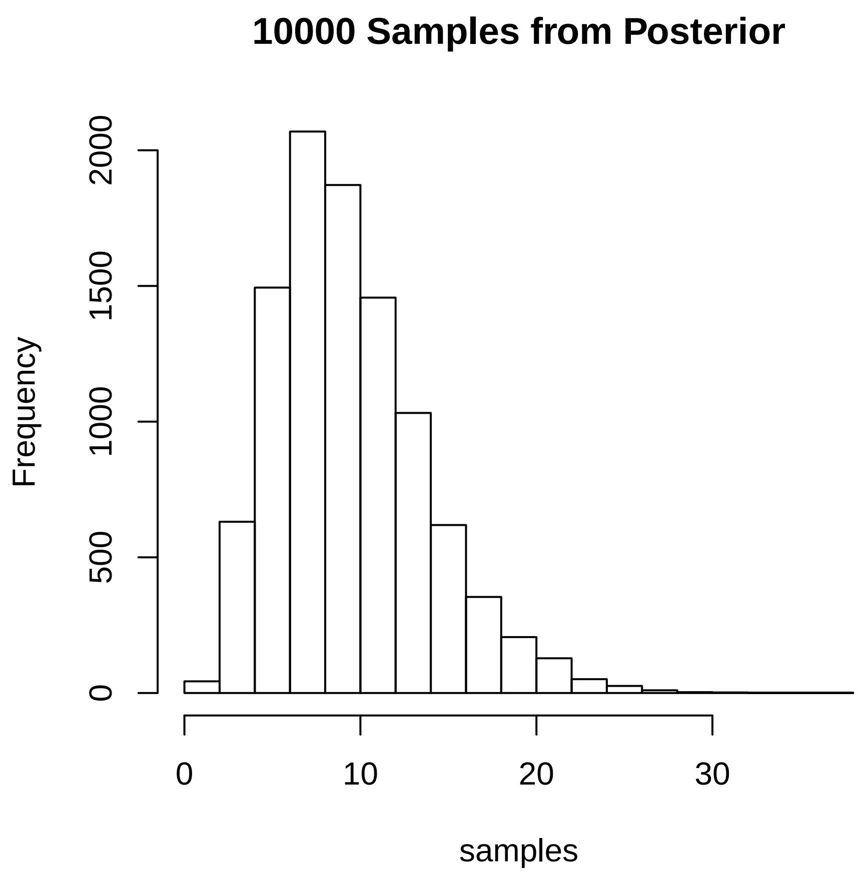
Since the MLE is asymptotically normal, we can approximate the 95% confidence interval with a normal distribution,

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

(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).