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

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

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

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

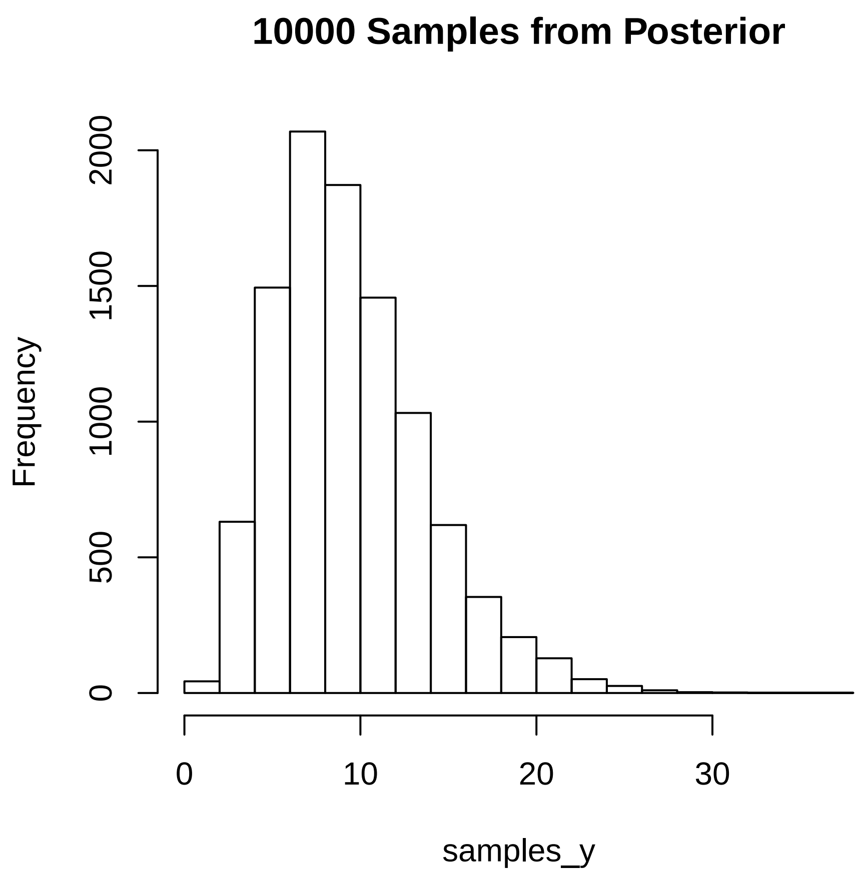
Since the MLE is asymptotically normal, we can approximate the 95% confidence interval with a normal distribution,

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

(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 $a = 0.8405201$ and $b = 0.267862$.
Using this prior and the data from part (c), we arrive at a posterior of $\theta|y \sim \mathrm{Gamma}(4.8405201, 0.517862)$.
Sampling from this distribution 1000 times, we get the following histogram,



A 95% credible interval using the 0.025- and 0.975-quantiles is $(2.9641727, 19.3295729)$.

(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).

Since we only have one observation, the confidence interval from the MLE is very wide. The prior in the Bayesian analysis acts as if we had "previous observations", thus tightening the confidence interval.

2

(a) The frequencies are given by:

$$\begin{bmatrix} 448 & 152 \\ 155 & 289 \end{bmatrix}$$

MLEs of p_{12} p_{21} are given by

$$\hat{p}_{12} = \frac{n_{12}}{n_{1+}} = \frac{152}{600} \approx 0.2533333$$
$$\hat{p}_{21} = \frac{n_{21}}{n_{2+}} = \frac{155}{444} \approx 0.3490991$$

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

$$\widehat{\mathrm{Var}(\hat{p}_{12})} \approx \frac{\hat{p}_{12}(1-\hat{p}_{12})}{n_{1+}} \approx 3.1525926 \times 10^{-4}$$
$$\widehat{\mathrm{Var}(\hat{p}_{21})} \approx \frac{\hat{p}_{21}(1-\hat{p}_{21})}{n_{2+}} \approx 5.1177684 \times 10^{-4}$$

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

$$\mathrm{CI}_{12} = \hat{p}_{12} \pm 1.96 \times \sqrt{\widehat{\mathrm{Var}(\hat{p}_{12})}} \approx (0.2185325, 0.2881342)$$
$$\mathrm{CI}_{21} = \hat{p}_{21} \pm 1.96 \times \sqrt{\widehat{\mathrm{Var}(\hat{p}_{21})}} \approx (0.304759, 0.3934392)$$

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

$$p_{12}|n_{1\cdot} \sim \mathrm{Beta}(n_{12}+1, n_{11}+1)$$
$$p_{21}|n_{2\cdot} \sim \mathrm{Beta}(n_{21}+1, n_{22}+1)$$

The posterior medians are then

$$\tilde{p}_{12} \approx \frac{n_{12}+1-1/3}{n_{1+}-2/3} \approx 0.2538804$$
$$\tilde{p}_{21} \approx \frac{n_{21}+1-1/3}{n_{2+}-2/3} \approx 0.349551$$

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,

$$\mathrm{CI}_{12} \approx (0.2201944, 0.2896583)$$
$$\mathrm{CI}_{21} \approx (0.3062282, 0.3945995)$$

(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

$$\begin{aligned} T &= 2(\hat{\ell}_1 - \hat{\ell}_0) \\ &= 2\sum_{i=1}^{2}\sum_{j=1}^{2} n_{ij} \log\left(\frac{\hat{p}_{ij}}{\hat{p}_j}\right) \\ &= 2\sum_{i=1}^{2}\sum_{j=1}^{2} n_{ij} \log\left(\frac{n_{ij}/n_{i+}}{n_j/n}\right) \\ &\approx 170.3160316 \end{aligned}$$

Under the null, $T \sim \chi^2_1$, and we have $\Pr(T > 170.3160316) \approx 0$, so the likelihood ratio test would reject the null hypothesis that the weather on each day is independent.
Under the Bayesian paradigm, we compare the independence assumption with prior $p_1 \sim \text{Beta}(1,1)$ versus the Markov assumption with priors $p_{12} \sim \text{Beta}(1,1)$ and $p_{21} \sim \text{Beta}(1,1)$. Under the independence assumption, the posterior for p_1 is

$$p_1|n \sim \text{Beta}(n_1+1, n_2+1)$$

while under the Markov assumption we have, as before,

$$\begin{aligned} p_{12}|n_{1\cdot} &\sim \text{Beta}(n_{12}+1, n_{11}+1) \\ p_{21}|n_{2\cdot} &\sim \text{Beta}(n_{21}+1, n_{22}+1) \end{aligned}$$

To find the Bayes factor, we must compute $\Pr(y|H_0)$ and $\Pr(y|H_1)$. For H_0, we have

$$\begin{aligned} \Pr(y|H_0) &= \binom{n}{n_1}\frac{\Gamma(2)}{\Gamma(1)\Gamma(1)}\frac{\Gamma(n_1+1)\Gamma(n_2+1)}{\Gamma(n+2)} \\ &= \binom{n}{n_1}\frac{\Gamma(n_1+1)\Gamma(n_2+1)}{\Gamma(n+2)} \\ &\approx 9.2506938 \times 10^{-4} \end{aligned}$$

while under H_1, we have

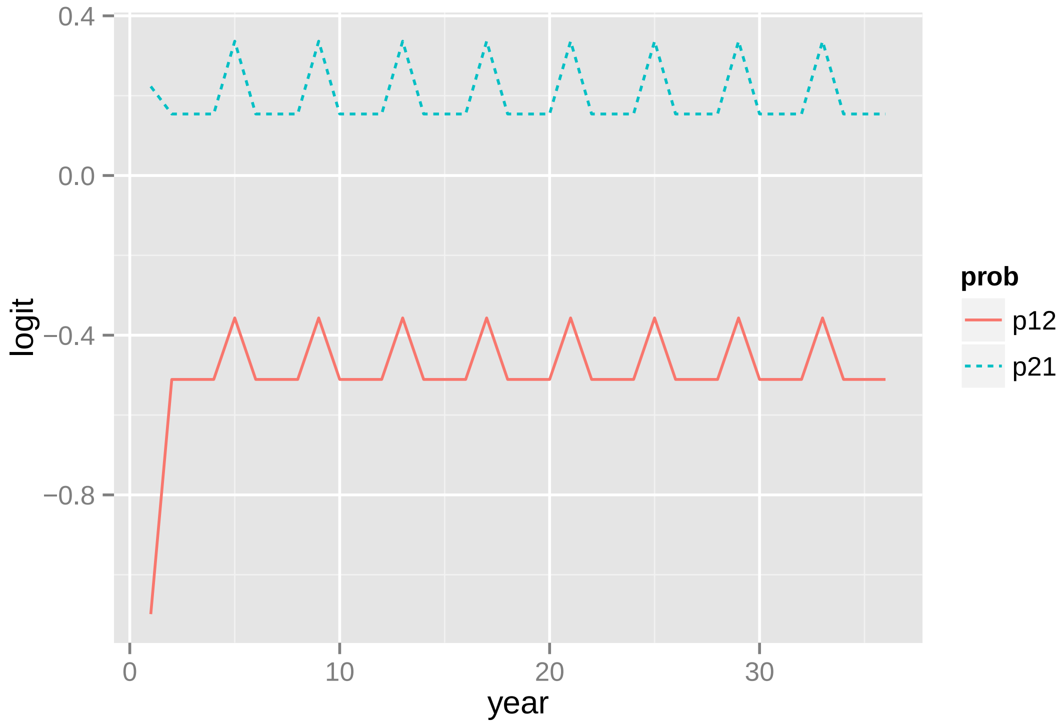
$$\begin{aligned} \Pr(y|H_1) &= \int_0^1\int_0^1 p_{12}^{n_{12}}(1-p_{12})^{n_{11}} p_{21}^{n_{21}}(1-p_{21})^{n_{22}} \left(\frac{\Gamma(2)}{\Gamma(1)\Gamma(1)}\right)^2 dp_{12}dp_{21}\binom{n_{1+}}{n_{12}}\binom{n_{2+}}{n_{21}} \\ &= \binom{n_{1+}}{n_{12}}\binom{n_{2+}}{n_{21}}\int_0^1\int_0^1 p_{12}^{n_{12}}(1-p_{12})^{n_{11}} p_{21}^{n_{21}}(1-p_{21})^{n_{22}} dp_{12}dp_{21} \\ &= \binom{n_{1+}}{n_{12}}\binom{n_{2+}}{n_{21}}\int_0^1 p_{12}^{n_{12}}(1-p_{12})^{n_{11}} dp_{12} \times \int_0^1 p_{21}^{n_{21}}(1-p_{21})^{n_{22}} dp_{21} \\ &= \binom{n_{1+}}{n_{12}}\binom{n_{2+}}{n_{21}}\frac{\Gamma(n_{12}+1)\Gamma(n_{11}+1)}{\Gamma(n_{1+}+2)} \cdot \frac{\Gamma(n_{21}+1)\Gamma(n_{22}+1)}{\Gamma(n_{2+}+2)} \\ &\approx 3.7390865 \times 10^{-6} \end{aligned}$$

The Bayes factor is then

$$\text{BF} = \frac{\Pr(y|H_0)}{\Pr(y|H_1)} \approx 247.4051804$$

Using the Kass and Raftery suggestions for intervals of Bayes Factors, we note that $1/\text{BF} \approx 0.004042 < 1$, so we would say there is *not* sufficient evidence against the null hypothesis of independence.

(d) Estimating the probabilites p_{12} and p_{21} in each year by the MLE, we have below a plot of the logits of the probabilities,



To test the null hypothesis that the transition probabilities are the same every year versus the alternative that they are different every year, we need to find $\hat{\ell}_0$ and $\hat{\ell}_1$. Letting $n_{ij}^{(k)}$ be the number of transitions from state i to state j in year k and $p_{ij}^{(k)}$ the transition probabilities for year k, we have

$$\hat{\ell}_1 = \sum_{k=1}^{36} \sum_{i=1}^{2} \sum_{j=1}^{2} n_{ij}^{(k)} \log \hat{p}_{ij}^{(k)} \approx -702.8216408$$

whereas for $\hat{\ell}_0$, we have

$$\hat{\ell}_0 = \sum_{i=1}^{2} \sum_{j=1}^{2} n_{ij} \log \hat{p}_{ij} \approx -704.2026072$$

giving us a likelihood ratio test statistic of

$$T = 2(\hat{\ell}_1 - \hat{\ell}_0) = 2.7619328$$

which follows a $\chi^2_{lrt.df}$, where the degrees of freedom arise from the fact that under H_1 we have $2 \times 36 = 72$ parameters to estiamte and under H_0 we have only 2 parameters. With $T \sim \chi^2_{lrt.df}$, we have $\Pr(T > 2.7619328) \approx 1$, so we would not reject the null hypothesis that there are common transition probabilities across years.
Under the Bayesian paradigm, ...