Stat 516, Homework 5

Due date: Thursday, November 16.

Names: Andy Magee Neal Marquez.

Note: Do this homework in *pairs*; two students turning in a single joint solution. No two Statistics, Biostatistics, QERM, EE, Econ, AMath, . . . students may work together. (Exceptions as needed based on the make-up of the class.)

1. Consider the model $Y|\theta \sim \text{Poisson}(E \times \theta)$ where E is a known "expected number" of cases, Y is the count of disease cases and $\theta > 0$ is the relative risk with $\theta = 1$ corresponding to "null" risk.

   (a) Find expressions for the likelihood function $L(\theta)$, the log likelihood function $l(\theta)$, the score function $S(\theta)$ and Fisher's (expected) information $I(\theta)$. Find the MLE and its variance.

$$\mathcal{L}(Y|\theta, E) = \frac{(E\theta)^Y e^{-E\theta}}{Y!}$$

$$\mathcal{L}(Y|\theta, E) = \prod_{i=1}^{n} \frac{(E\theta)^{y_i} e^{-E\theta}}{y_i!}$$

$$\mathcal{L}(Y|\theta, E) = \frac{(E\theta)^{\sum_{i=1}^{n} y_i} e^{-nE\theta}}{\prod_{i=1}^{n} y_i!}$$

$$\ell(Y|\theta, E) = \sum_{i=1}^{n} y_i log(E\theta) - nE\theta + c$$

$$S(\theta) = \frac{d\ell}{d\theta} = -n + \frac{1}{E\theta} \sum_{i=1}^{n} y_i$$

$$I(\theta) = -\mathbf{E}\left[\frac{d^2\ell}{d\theta^2}\right] = \frac{1}{E\theta^2} \sum_{i=1}^{n} y_i = \frac{n}{\theta}$$

$$MLE(\theta) = \hat{\theta} = \frac{1}{nE} \sum_{i=1}^{n} y_i$$

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

   (b) Suppose we assume a prior of $\theta \sim \text{Gamma}(a, b)$ so that

$$p(\theta) = \frac{b^a}{\Gamma(a)} \theta^{a-1} \exp(-b\theta),$$

   with $a, b > 0$. Show that the posterior $\theta|y$ is also gamma and find its parameters.

$$\text{Posterior} \propto \text{Likelihood} \times \text{Prior}$$

$$= \frac{b^a \theta^{a-1} e^{-b\theta}}{\Gamma(a)} \frac{(E\theta)^Y e^{-E\theta}}{Y!}$$

$$= \frac{b^a}{\Gamma(a)} \frac{E^Y}{Y!} \theta^{a-1} e^{-b\theta} \theta^Y e^{-E\theta}$$

$$\propto \theta^{a-1} e^{-b\theta} \theta^Y e^{-E\theta}$$

$$= \theta^{Y+a-1} e^{-\theta(b+E)}$$

$$\propto \frac{(b+E)^{Y+a} \theta^{Y+a-1} e^{-\theta(b+E)}}{\Gamma(Y+a)}$$

$$\sim \text{Gamma}(Y+a, b+E)$$

(c) Close to a nuclear reprocessing plant $y = 4$ cases of leukemia were observed with an expected number of $E = 0.25$. Give the MLE, its variance and a 95% confidence interval based on a normal approximation.

$$\hat{\theta} = \frac{4}{.25} = 16$$

$$Var(\hat{\theta}) = \frac{16}{1} = 16$$

$$CI_{.95} = \left(\hat{\theta} - 1.96 * \sqrt{Var(\hat{\theta})}, \hat{\theta} + 1.96 * \sqrt{Var(\hat{\theta})} \right)$$

$$= (8.16, 23.84)$$

(d) Find the a and b which give a gamma prior that assigns 0.9 probability to the interval [0.1,10]. Find the posterior corresponding to this prior and generate samples from the posterior. Give a histogram of the posterior and state a 95% credible interval.
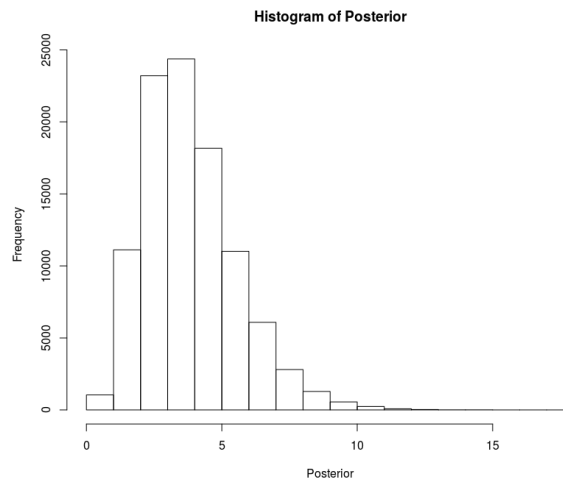
Minimizing the following function in R gives us a distribution Gamma(.84,.27) which satisfies the above criteria.

```
fnOpt <- function(par) {
  a <- par[1]
  b <- par[2]
  return( ( ((0.1 - qgamma(0.05,a,b))/0.1)^2 + ((10 - qgamma(0.95,a,b))/10)^2 ) )
}
```

The 95% credible interval obtained from this process yields values $(2.94, 19.33)$. A histogram of samples from the posterior follows.

Histogram of Posterior

(e) Is there evidence of excess risk for these data? Discuss the differences between the likelihood-based and Bayesian analyses.

Based on the standard criteria for assesing evidence of risk both the MLE and the Bayesian approach provide 95% intervals, confidence and credible respectively, of the parameter estimate $\hat{\theta}$ which do not cover the null risk value of 1. Because of this our data shows evidence for higher risk than expected at a strong statistical level.

2. Consider the Snoqualmie falls data available on the class website. We will analyze the data for June across the 36 years; compare how the data was processed in examples in the lecture notes.

(a) Using the multiple years of data estimate the transition probability matrix for the first-order Markov chain model, with p_{12} the probability of wet given dry and p_{21} the probability of dry given wet. Obtain MLEs and 95% asymptotic confidence intervals for p_{12} and p_{21}.

```
y <- scan("snoqualmie.txt")
nodays <- rep(c(365,365,365,366),9) # account for leap years
june <- matrix(0, nrow=36, ncol=30) # build empty data mtarix
daysum <- 31*3 + 30 + 29 # our start point is this number of days in the future
for (i in 1:36){
    if (i>1) daysum <- daysum + nodays[i-1] # if past the first year add 365|366
    june[i,] <- (y[daysum+1:30] > 0) * 1 # turn into indicators
}

n11 <- sum(june[,1:29] == 0 & june[,2:30] == 0)
n21 <- sum(june[,1:29] == 1 & june[,2:30] == 0)
n12 <- sum(june[,1:29] == 0 & june[,2:30] == 1)
n22 <- sum(june[,1:29] == 1 & june[,2:30] == 1)

# sanity check
```

```
sum(n11 + n12 + n21 + n22) == 36*29


p_hat_11 <- n11 / (n11 + n12)
p_hat_12 <- n12 / (n11 + n12)
p_hat_21 <- n21 / (n21 + n22)
p_hat_22 <- n22 / (n21 + n22)


P_hat <- rbind(c(p_hat_11, p_hat_12), c(p_hat_21, p_hat_22))
colnames(P_hat) <- c("dry", "wet")
row.names(P_hat) <- c("dry", "wet")


N_obs <- rbind(c(n11, n12), c(n21, n11))
colnames(N_obs) <- c("dry", "wet")
row.names(N_obs) <- c("dry", "wet")
P_stderr <- sqrt(P_hat * (1 - P_hat) / rowSums(N_obs))
P_hat
P_hat - 1.96 * P_stderr
P_hat + 1.96 * P_stderr
```

$$\hat{p}_{12} = .253$$

$$\text{CI}_{\hat{p}_{12}} = (.219, .288)$$

$$\hat{p}_{21} = .349$$

$$\text{CI}_{\hat{p}_{21}} = (.305, .393)$$

(b) Under the assumption of independent uniform priors, give the form of the posterior distributions for the parameters p_{12}, p_{21}. Obtain the posterior and report posterior medians and 95% credible intervals for each of the two parameters.

A Uniform(0,1) prior is a Beta(1,1) prior, making the posterior distribution beta as well (since the Beta is the 2D Dirichlet distribution, and the Dirichlet is the conjugate prior for unrestricted transition matrices).

Specifically, with a Beta(1,1) prior, the posterior $p_{ij} \mid n_{ij}, n_{ii} \sim Beta(1 + n_{ij}, 1 + n_{ii})$

This means we can get our posterior medians and confidence intervals easily, as:

```
posterior_median_p_12 <- qbeta(0.5,1+n12,1+n11)
posterior_median_p_21 <- qbeta(0.5,1+n21,1+n22)
posterior_ci_p_12 <- qbeta(c(0.025,0.975),1+n12,1+n11)
posterior_ci_p_21 <- qbeta(c(0.025,0.975),1+n21,1+n22)
```

Giving us very similar results.

$$\hat{p}_{12} = .254$$

$$\text{CI}_{\hat{p}_{12}} = (.220, .290)$$

$$\hat{p}_{21} = .350$$

$$\text{CI}_{\hat{p}_{21}} = (.306, .395)$$

(c) Test the null of independence versus the Markov model, using a likelihood ratio test and a Bayes factor.

The independence model is one with the following transition matrix:

$$\begin{bmatrix} p_{11} & p_{12} \\ p_{11} & p_{12} \end{bmatrix}$$

That is, both rows are identical, and there is only one free parameter in the model. We note that in the indendence model, p_{11} is the probability of transitioning to a dry day (given a wet or a dry day) and p_{12} is the probability of transitioning to a wet day (given a dry or wet day).

It is then clear that the MLE of,p_{11}, $\hat{p}^{*}_{11} = \frac{n_{11}+n_{21}}{n}$. Similarly the MLE of,$p_{12}$, $\hat{p}^{*}_{12} = \frac{n_{12}+n_{22}}{n}$.

The likelihood ratio test here is distriuted $\chi^2(df = 1)$, as the independent model is found by restricting the alternative model to have $p_{21} = p_{11}$ and $p_{22} = p_{12}$

```
p_hat_star_11 <- (n11 + n21)/n
p_hat_star_12 <- (n21 + n22)/n
lnL_independence <- (n11+n21)*log(p_hat_star_11) + (n21+n22)*log(p_hat_star_12)
lnL_Markov <- n11*log(p_hat_11) + n12*log(p_hat_12) + n21*log(p_hat_21) +
    n22*log(p_hat_22)


LR_vs_independence <- 2 * (lnL_Markov - lnL_independence)


LR_vs_independence


pchisq(LR_vs_independence,1,lower.tail=FALSE)
```

Giving us a chi squared value of 167 with probability 2.47×10^{-38}.

To get the Bayes Factor, we note that the marginal likelihood of the independence model, $Pr(\mathbf{y} \mid H_o)$ is given by

$$\int_0^1 Pr(\mathbf{y} \mid p_{11},\ H_0)\, p(p_{11} \mid H_0) = \frac{\Gamma(1+1)}{\Gamma(1)\Gamma(1)} \int_0^1 (p_{11})^{n_{11}+n_{21}+1-1}(1-p_{11})^{n_{12}+n_{22}+1-1}dp_{11}$$

$$= \frac{\Gamma(1+1)}{\Gamma(1)\Gamma(1)} \frac{\Gamma(n_{11}+n_{21}+1)\Gamma(n_{12}+n_{22}+1)}{\Gamma(n+1+1)}$$

$$= \frac{\Gamma(n_{11}+n_{21}+1)\Gamma(n_{12}+n_{22}+1)}{\Gamma(n+1+1)}$$

The marginal likelihood for the Markov model, $Pr(\mathbf{y} \mid H_1)$ is given by

$$\int_0^1 Pr(\mathbf{y} \mid p, \ H_1) \ p(p_{11}, p_{21} \mid H_1) = \frac{\Gamma(1+1)}{\Gamma(1)\Gamma(1)} \frac{\Gamma(1+1)}{\Gamma(1)\Gamma(1)}$$

$$\times \int_0^1 \int_0^1 (p_{11})^{n_{11}+1-1}(1-p_{11})^{n_{12}+1-1}(p_{21})^{n_{21}+1-1}(1-p_{21})^{n_{22}+1-1} dp_{11} dp_{21}$$

$$= \frac{\Gamma(n_{11}+1)\Gamma(n_{12}+1)}{\Gamma(n_{11}+n_{12}+1+1)} \frac{\Gamma(n_{21}+1)\Gamma(n_{22}+1)}{\Gamma(n_{21}+n_{22}+1+1)}$$

```
marginal_indpendence <- exp(lgamma(n11+n21+1) + lgamma(n12+n22+1) - lgamma(n+1))

marginal_Markov <- exp(lgamma(n11+1) + lgamma(n12+1) + lgamma(n21+1) +
                       lgamma(n22+1) - lgamma(n11+n12+2) - lgamma(n21+n22+2))


BF_vs_independence <- marginal_Markov/marginal_indpendence
```

Giving us a probability 2.39×10^{-32}

(d) Now consider the null of common transition probabilities across years versus the alternative that each year possesses its own pair of probabilities. First, estimate the probabilities in each year and provide a figure of the logits of the probabilities, i.e. $\log\left(\frac{p_{12}}{1-p_{12}}\right)$ and $\log\left(\frac{p_{21}}{1-p_{21}}\right)$, plotted versus year. Second, carry out a likelihood ratio test to formally test the hypothesis. Third, evaluate the Bayes factor to examine the evidence for each of the two hypotheses. What do you conclude?
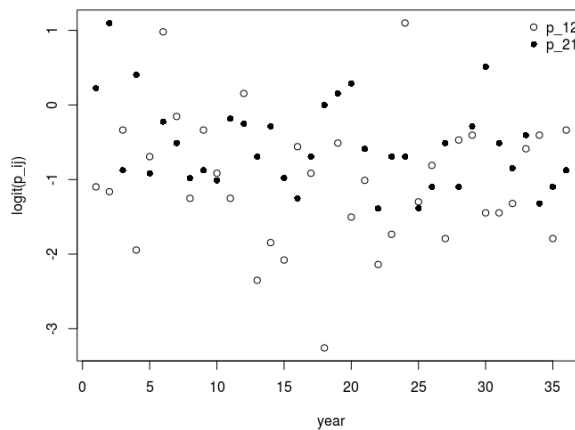
First, we loop over all the years to record the year-specific TPMs for June, so we can plot the $logit(p_{ij})$. Then we loop over the years again to calculate the likelihood ratio.

To conduct the LRT, we note that the test statistic is distributed $\chi^2(df = 70)$, since there are $36 * 2 = 72$ free parameters in the model where each year has its own matrix, and 2 in the simpler model.

Lastly, we loop over all the years one last time so we can get the marginal likelihood of the year-specific model. Since all the years are independent, our 72-dimensional integral is easy to decompose into 36 that look a lot like the ones we calculated above.

The R code is as follows.

```r
#### Likelihood Ratio Test
yearly_P <- apply(june,1,function(june_year){
  n11 <- sum(june_year[1:29] == 0 & june_year[2:30] == 0)
  n12 <- sum(june_year[1:29] == 0 & june_year[2:30] == 1)
  n21 <- sum(june_year[1:29] == 1 & june_year[2:30] == 0)
  n22 <- sum(june_year[1:29] == 1 & june_year[2:30] == 1)


  n <- n11 + n12 + n21 + n22


  p_hat_11 <- n11 / (n11 + n12)
  p_hat_12 <- n12 / (n11 + n12)
  p_hat_21 <- n21 / (n21 + n22)
  p_hat_22 <- n22 / (n21 + n22)


  return(c(p_hat_11,p_hat_12,p_hat_21,p_hat_22))


})


# Calculate and plot the logit pij as requested
logit_p12 <- log(yearly_P[2,]/(1-yearly_P[2,]))
logit_p21 <- log(yearly_P[3,]/(1-yearly_P[3,]))


plot(logit_p12,xlab="year",ylab="logit(p_ij)",pch=1,ylim=c(min(c(logit_p12,logit_p21)),max(c(logit_p12,
points(logit_p21,pch=16)
legend("topright",legend=c("p_12","p_21"),pch=c(1,16),border=NA,bty="n")


# Carry out the LRT
yearly_lnL_vector <- apply(june,1,function(june_year){
  n11 <- sum(june_year[1:29] == 0 & june_year[2:30] == 0)
  n12 <- sum(june_year[1:29] == 0 & june_year[2:30] == 1)
  n21 <- sum(june_year[1:29] == 1 & june_year[2:30] == 0)
  n22 <- sum(june_year[1:29] == 1 & june_year[2:30] == 1)


  n <- n11 + n12 + n21 + n22


  p_hat_11 <- n11 / (n11 + n12)
  p_hat_12 <- n12 / (n11 + n12)
  p_hat_21 <- n21 / (n21 + n22)
  p_hat_22 <- n22 / (n21 + n22)


  lnL <- n11*log(p_hat_11) + n12*log(p_hat_12) + n21*log(p_hat_21) + n22*log(p_hat_22)


  return(lnL)
```

```r
})


# The log-likelihood for the "every year to itself" model is the sum of the
    log-likelihoods of each year
lnL_yearly <- sum(yearly_lnL_vector)


LR_vs_yearly <- 2 * (lnL_yearly - lnL_Markov)


pchisq(LR_vs_yearly,36*2-2,lower.tail=FALSE)


### Bayes Factor
yearly_lnL_marginal_likelihood <- apply(june,1,function(june_year){
  n11 <- sum(june_year[1:29] == 0 & june_year[2:30] == 0)
  n12 <- sum(june_year[1:29] == 0 & june_year[2:30] == 1)
  n21 <- sum(june_year[1:29] == 1 & june_year[2:30] == 0)
  n22 <- sum(june_year[1:29] == 1 & june_year[2:30] == 1)

  n <- n11 + n12 + n21 + n22
  n <- n11 + n12 + n21 + n22

  p_hat_11 <- n11 / (n11 + n12)
  p_hat_12 <- n12 / (n11 + n12)
  p_hat_21 <- n21 / (n21 + n22) n11 <- sum(june_year[1:29] == 0 & june_year[2:30] == 0)
  n12 <- sum(june_year[1:29] == 0 & june_year[2:30] == 1)
  n21 <- sum(june_year[1:29] == 1 & june_year[2:30] == 0)
  n22 <- sum(june_year[1:29] == 1 & june_year[2:30] == 1)

  n <- n11 + n12 + n21 + n22
  n <- n11 + n12 + n21 + n22

  p_hat_11 <- n11 / (n11 + n12)
  p_hat_22 <- n22 / (n21 + n22)

  this_year_log_marginal <- lgamma(n11+1) + lgamma(n12+1) + lgamma(n21+1) +
      lgamma(n22+1) - lgamma(n11+n12+2) - lgamma(n21+n22+2)

  return(this_year_log_marginal)

})


marginal_yearly <- exp(sum(yearly_lnL_marginal_likelihood))


BF_vs_yearly <- marginal_yearly/marginal_Markov
  p_hat_11 <- n11 / (n11 + n12)
  p_hat_12 <- n12 / (n11 + n12)
```

```
    p_hat_21 <- n21 / (n21 + n22)
    p_hat_22 <- n22 / (n21 + n22)


    this_year_log_marginal <- lgamma(n11+1) + lgamma(n12+1) + lgamma(n21+1) +
        lgamma(n22+1) - lgamma(n11+n12+2) - lgamma(n21+n22+2)


    return(this_year_log_marginal)


  })


  marginal_yearly <- exp(sum(yearly_lnL_marginal_likelihood))


  BF_vs_yearly <- marginal_yearly/marginal_Markov
```

On the basis of the likelihood test, we would reject the simpler model where there is a single TPM shared between all Junes in all years in favor of each June getting its own TPM.

On the basis of the Bayes Factor, we would not reject the single TPM for all Junes. In fact, since the simpler model has a marginal likelihood on the order of 10^{16} times greater than the TPM-per-yaer model, we would be inclined to view this as strong evidence in favor of the simpler model.

On the one hand, our plot shows a great deal of year-to-year heterogeneity in the transition matrix parameters. The LRT sees this as very important. On the other hand, each of those is based on 29 transitions, from which 2 free parameters are being estimated, so it feels like we might be overfitting a bit, and the Bayes Factor punishes us more heavily here for the high dimensionality of our more complicated model.

Our LRT returns a p-value of ≈ 0.028, reasonable evidence against the simple model. But our Bayes Factors says that the more complicated model is $\approx 4.39 \times 10^{-17}$ as probable as our simple model, which is a stronger preference in the opposite direction. It does not seem unreasonable to conclude that a single TPM is the better choice.

3. For ternary observations (i.e., $s = 3$ states), consider a Markov chain model in which the 3×3 transition probability matrix $\mathbf{P} = (p_{ij})$ is unrestricted. With the row sums being one, we may think of there being 6 parameters, say p_{ij} with $1 \leq i \leq 3$ and $1 \leq j \leq 2$.

  (a) Specialize the result on likelihood inference for parametric transition probabilities to this case and write out the 6×6 joint asymptotic covariance matrix for the MLE

$$(\hat{p}_{11}, \hat{p}_{12}, \hat{p}_{21}, \hat{p}_{22}, \hat{p}_{31}, \hat{p}_{32}),$$

giving formulas in terms of $\mathbf{P}$ and the stationary distribution π (which is assumed to exist).

The log likelihood for the above function may be written as

$$\ell(\mathbf{p}) = N_{11}\log(p_{11}) + N_{12}\log(p_{12}) + N_{13}\log(1 - p_{11}p_{12}) +$$
$$N_{21}\log(p_{21}) + N_{22}\log(p_{22}) + N_{23}\log(1 - p_{21}p_{22}) +$$
$$N_{31}\log(p_{31}) + N_{32}\log(p_{32}) + N_{33}\log(1 - p_{31}p_{32})$$

Where...

$$E[N_{ij}] = n\pi_i p_{ij}$$

Which when taking second ordervpartial derivatives yields the follwing asymptotic covariance matrix Σ.

$$\Sigma = \begin{bmatrix} \frac{p_{11}(1-p_{11})}{n\pi_1} & \frac{-p_{11}p_{12}}{n\pi_1} & 0 & 0 & 0 & 0 \\ \frac{-p_{11}p_{12}}{n\pi_1} & \frac{p_{12}(1-p_{12})}{n\pi_1} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{p_{21}(1-p_{21})}{n\pi_2} & \frac{-p_{21}p_{22}}{n\pi_2} & 0 & 0 \\ 0 & 0 & \frac{-p_{21}p_{22}}{n\pi_2} & \frac{p_{22}(1-p_{22})}{n\pi_2} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{p_{31}(1-p_{31})}{n(1-\pi_1-\pi_2)} & \frac{-p_{31}p_{32}}{n(1-\pi_1-\pi_2)} \\ 0 & 0 & 0 & 0 & \frac{-p_{31}p_{32}}{n(1-\pi_1-\pi_2)} & \frac{p_{32}(1-p_{32})}{n(1-\pi_1-\pi_2)} \end{bmatrix}$$

or

for $1 \leq i \leq 3$ and $1 \leq j \leq 2$ and $1 \leq k \leq 3$ and $1 \leq l \leq 2$

$$\Sigma_{\hat{p}_{ij},\hat{p}_{kl}} = \begin{cases} 0 & \text{if } i \neq k \\ \frac{p_{ij}(1-p_{ij})}{n\pi_i} & \text{if } i = k \ \& \ j = l \\ \frac{-p_{ij}p_{il}}{n\pi_i} & \text{if } i = k \ \& \ j \neq l \end{cases}$$

(b) Write out the 9×9 joint asymptotic covariance matrix for the MLE

$$(\hat{p}_{11}, \hat{p}_{12}, \hat{p}_{13}, \ \hat{p}_{21}, \hat{p}_{22}, \hat{p}_{23}, \ \hat{p}_{31}, \hat{p}_{32}, \hat{p}_{33}).$$

for $i, j, k, l \in \{1, 2, 3\}$

$$\Sigma_{\hat{p}_{ij},\hat{p}_{kl}} = \begin{cases} 0 & \text{if } i \neq k \\ \frac{p_{ij}(1-p_{ij})}{n\pi_i} & \text{if } i = k \ \& \ j = l \neq 3 \\ \frac{-p_{ij}p_{il}}{n\pi_i} & \text{if } i = k \ \& \ j \neq l \ \& \ j \neq 3 \ \& \ l \neq 3 \\ \frac{(1-p_{i1}-p_{i2})(p_{i1}+p_{i2})}{n\pi_i} & \text{if } i = k \ \& \ j = l = 3 \\ \frac{-p_{ij}(1-p_{i1}-p_{i2})}{n\pi_i} & \text{if } i = k \ \& \ j \neq l = 3 \\ \frac{-(1-p_{i1}-p_{i2})p_{il}}{n\pi_i} & \text{if } i = k \ \& \ l \neq j = 3 \end{cases}$$

(c) Discuss the asymptotic dependences you find.

Each state i can be used as a starting point and all points in the asymptotic covariance matrix for the MLE may be seen as a trinomial likelihood where the elements of the covariance matrix Σ may be populated by the inverse of the fisher information matrix $I(\theta)$. For each starting point the

sub matrix has asymptotic dependencies such that any two values p_{i1}, p_{i2}, and p_{i3} are negatively correlated which, because of their sum to one constraint, makes sense with our intuition for the model.