Qualify Exam 2015

Mingwei Fei

December 12, 2022

1 MVN

Suppose that $Y \sim N(\mu, \Sigma)$ where Σ is symmetric and full rank, Let A be a symmetric matrix.

1.1 Quadratic Form $Y^T AY$ and Chi-square distribution

Show that the quadratic form $Y^T AY$ can be represented as

$$Y^T AY = \sum_{i=1}^{k} \lambda_i W_i$$

where the W_i's are independently distributed as noncentral chi-squared variables with d_i degrees of freedom and noncentrality parameter δ_i, that is, $W_i \sim \chi^2_{d_i}(\delta_i), i = 1, 2, ..., k$. Indicate what λ_i, d_i, δ_i are equal to.

1.1.1 Question

Suppose $Y_{n\times n}$, and Σ is full rank. so Σ is $n \times n$ dimension matrix,

(i) *Normal distribution vs. Chi-square:*

We can transform Y_i into $N(\mu, 1)$ distribution, so that the quadratic form will be a non-central chi-square distribution.

If $Z_1, ..., Z_k$ are independent, standard normal random variables, then the sum of their squares is chi-square distribution,

$$Q = Z_i^2 \sim \chi^2(k)$$
$$p(k) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} exp(-\frac{x}{2})$$

(ii) *Non-central Chi-square:*

Here the k is unknown, we need to show that the sum of non-central chi-square distribution is also a non-central chi-square distribution with distribution transformation. The distribution transformation generally use Moment Generating Function.

Lemma:

Let $Q_i \sim \chi^2_{k_i}(\lambda_i)$ for $i = 1, \dots, n$, be independent. Then, $Q = \sum_{i=1}^n Q_i$ is a noncentral $\chi^2_k(\lambda)$, where $k = \sum_{i=1}^n k_i$ and $\lambda = \sum_{i=1}^n \lambda_i$.

Chi-square distribution and non-central chi-square distribution are totally different. I need to understand the components δ_i and d_i in the non-central chi-square distribution.

The non-central chi-square distribution: Let $(X_1, X_2, \dots, X_i, \dots, X_k)$ be k independent, normally distributed random variables with means μ_i and unit variances. Then the random variable

$$Q = \sum_{i=1}^k X_i^2 \sim \chi^2(k, \lambda), \qquad \lambda = \sum_{i=1}^k \mu_i^2$$

(iii) Here A matrix is not necessarily inverse of Σ, it could be any symmetric matrix. So this is a general case of linear combination of non-central chi-square.

$$\begin{aligned}
\Sigma &= QQ^T \\
Y^T A Y &= (Q^T Y)^T diag\{\lambda_1, ..., \lambda_k\}(Q^T Y) \\
Q^T Y &= \Sigma^{-1/2} Y \sim N(\mu, 1) \\
A &= \Sigma^{-1/2} diag\{\lambda_1, ..., \lambda_k\}\Sigma^{-1/2} \\
A^T &= A, \qquad \text{A is symmetric}
\end{aligned}$$

1.1.2 Proof

$$Y^T A Y = \sum_{i=1}^k \lambda_i W_i$$

where W_i are independently distributed as noncentral chi-squared variables with d_i degrees of freedom and noncentrality parameter δ_i, that is, $W_i \sim \chi^2_{d_i}(\delta_i), i = 1, 2, ..., k$. Indicate what λ_i, d_i, δ_i are equal to.

$$\begin{aligned}
\Sigma &= QQ^T, \qquad \text{by semi-definite matrix} \\
Q^{-1}Y &= (Z_i), \qquad Z_i \sim N(\mu_i, I) \\
A &= Q\Lambda Q^T, \qquad \Lambda = diag\{\lambda_1, ...\lambda_k\} \\
Y^T AY &= Y^T Q\Lambda Q^T Y = (Q^T Y)^T diag\{\lambda_1, ...\lambda_k\}(Q^T Y) \\
&= \sum_{i=1}^{k} \lambda_i Z_i^2 =\sim \sum_{i=1}^{k} \lambda_i \chi^2(d_i, \delta_i), \qquad \delta_i = \mu_i^2 \\
Y^T AY &= \sum_{i=1}^{k} \lambda_i W_i, \qquad W_i \sim \chi^2_{d_i}(\delta_i)
\end{aligned}$$

λ_i is the eigenvalue of matrix A , d_i is the number of same eigenvalue λ_i.

W_i is the non-central chi-square distribution with noncentrality parameter $\delta_i = \sum_{i=1}^{d_i} \mu_i^2$.

1.2 MGF of $Y^T AY$

Use part (a) to derive the moment generating function of $Y^T AY$. Let m(t) denote the moment generating function. Show that m(t) exits in a small neighborhood of t = 0, say, $|t| < t_0$ for some positive constant t_0. Find the maximal value of t_0 i.i.d chi-square distribution sum MGF.

1.2.1 Proof

$$\begin{aligned}
M(t) &= \prod_{i=1}^{k} M_i(t) \\
p(x_i) &= Q^{-1}Y = \frac{1}{\sqrt{2\pi}} exp(-\frac{(X-\mu_i)^2}{2}) \\
M_i(t) &= E[x_i^2 t] = \frac{1}{\sqrt{2\pi}} \int exp[-\frac{(1-2t)x^2 - 2\mu x + \mu_i^2}{2}]dx \\
&= \frac{1}{\sqrt{2\pi}} \int exp[-\frac{x^2 - 2\mu_i/(1-2t)x + \mu_i^2/(1-2t)^2 - \mu_i^2/(1-2t)^2 + \mu_i^2/(1-2t)}{2((1-2t)^{-1})}]dx \\
&= exp[\frac{\mu_i^2/(1-2t)^2 - \mu_i^2/(1-2t)}{2((1-2t)^{-1})}](1-2t)^{-1/2} \int \frac{1}{\sqrt{2\pi(1-2t)^{-1}}} exp[-\frac{[x-\mu_i/(1-2t)]^2}{2((1-2t)^{-1})}]dx \\
&= (1-2t)^{-1/2} exp[\frac{\mu_i^2 t}{(1-2t)}] \\
M(t) &= \prod_{i=1}^{k} (1-2t)^{-1/2} \frac{\mu_i^2 t}{(1-2t)} = (1-2t)^{-k/2} exp[\frac{\sum_i^k \mu_i^2 t}{(1-2t)}]
\end{aligned}$$

In which, $(1-2t) > 0, \qquad t < 1/2$. We can see that the product of non-centrality chi-square distributions is also a non-central chi-square distribution.

Another method is to let $Z \sim N(0,1)$, then $(Z+\mu)^2$ has a noncentral chi-square distribution with one degree of freedom, the MGF of $(Z+\mu)^2$

$$
\begin{aligned}
E[exp(t(Z+\mu)^2)] &= \frac{1}{\sqrt{2\pi}}\int exp(t(Z+\mu)^2)exp(-\frac{Z^2}{2}) \\
&= \frac{1}{\sqrt{2\pi}}\int exp[-\frac{(1-2t)Z^2-2\mu Z+\mu_i^2}{2}]dZ \\
&= (1-2t)^{-1/2}exp[\frac{\mu^2 t}{(1-2t)}]
\end{aligned}
$$

By definition, a non-central chi-square random variable $\chi^2_{n,\lambda}$ with n df and parameters $\lambda = \sum_i^n \mu_i^2$ is the sum of n independent normal variables $X_i = Z_i + \mu_i, i = 1, 2, ..n$. **Remember multivariate normal distribution, μ_i are different**.

$$
\begin{aligned}
\chi^2_{n,\lambda} &= \sum_i^n X_i^2 = \sum_i^n (Z_i+\mu_i)^2 \\
M(t) &= \prod_{i=1}^k M_i(t) = \prod_i^k (1-2t)^{-1/2}exp[\frac{\mu^2 t}{(1-2t)}] \\
&= (1-2t)^{-k/2}exp[\frac{\sum_i^n \mu_i^2 t}{(1-2t)}] = (1-2t)^{-n/2}exp[\frac{\lambda t}{(1-2t)}]
\end{aligned}
$$

1.3 $A = \Sigma^{-1}$

Use part (a) to show that $tr[(A\Sigma)^2] = tr(A\Sigma) = r$, where r is the rank of A, then Y^TAY has a chi-squared distribution. Determine its degrees of freedom and noncentrality parameter.

1.3.1 Question

1.3.2 Proof

From part (a) that

$$
\begin{aligned}
A = Q\Lambda Q^T, &\qquad \Sigma = QQ^T \\
(A\Sigma)^2 = A\Sigma A\Sigma &= [Q\Lambda Q^TQQ^T][Q\Lambda Q^TQQ^T] = Q\Lambda^2 Q^T \\
tr((A\Sigma)^2) = tr(A\Sigma), &\qquad \lambda_i^2 = \lambda_i, \qquad \lambda_i = 1, 0
\end{aligned}
$$

As $r = \sum_{i=1}^k \lambda_i$ is the rank of A, then we have

$$A\Sigma = diagBlk\{I_{r\times r}, \quad 0_{n-r\times n-r}\}$$

Then Y^TAY is the sum of r chi-square $\chi^2(1, \delta_i)$

$$Y^T AY = (Q^T Y)^T I_{r\times r}(Q^T Y) = \chi^(r, \delta)$$
$$\delta = \sum_{i=1}^{r} \mu_i^2$$

The degrees of freedom is r, the non-centrality parameter is $\delta = \sum_{i=1}^{r} \mu_i^2$.

1.4 $YTAY$ Distribution

Show that $Y^T AY$ has a noncentral chi-squared distribution if and only if $A\Sigma$ is idempotent.

1.4.1 Questions

Need to link the piece of information together. In order to have noncentral chi-square,

$$A = PP^T, \qquad \text{Symmetric matrix}$$
$$Y^T AY = (P^T Y)^T (P^T Y) \sim \chi^2(r, \delta), \qquad P^T Y \sim N(\mu, I)$$
$$P^T Y = \Sigma^{-1/2} Y \sim N(\mu, I)$$

Idempotent is $A^2 = A, A^T = A$. To prove if and only if, we need to demonstrate both way. And often times we need to show contradiction.

From part(c), we already show that when A is idempotent, the $Y^T AY$ has a non-central chi-squared distribution.

$A\Sigma = (A\Sigma)^2$, $A\Sigma$ is idempotent.

1.4.2 Proof

we have the MGF of linear combination of non-central chi-square distribution Y

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(w_i t)$$
$$= \prod_{i=1}^{n} (1 - 2w_j t)^{-1/2} exp\left(\frac{\lambda w_j t}{1 - 2w_j t}\right)$$

Then we can see that the shape parameter is $\frac{1}{2w_i}$. If we want to have a non-central chi-square distribution for Y, then all w_j need to be the same.

$$M_Y(t) = \prod_{i=1}^{n} (1 - 2w_i t)^{-1/2} exp\left(\frac{\lambda_i w_i t}{1 - 2w_i t}\right)$$
$$= (1 - 2wt)^{-n/2} exp\left(\frac{\sum_{i=1}^{n} \lambda_i wt}{1 - 2wt}\right)$$

And for chi-square distribution, the shape parameter has to be 1/2, so the $w_i = 1$. So we prove that if Y is a non-central chi-square distribution, A has to have the eigenvalues either 1 or 0.

The other way is also proved from part (c).

2 Likelihood function in regression model/ different link functions

3 Likelihood function for random effect, two level distribution

3.1 Bayesian Statistics

The Bayesian statistics could be used to construct likelihood function, we introduce hidden variables that could be integrate out to get the marginal distribution.

The below question is about "matched pair"

Consider independent observations $(X_1, Y_1), ..., (X_n; Y_n)$ where Y_i takes values 0 and 1. Suppose that $X_i|(Y_i = m) \sim N(\mu_m, \sigma^2)$ and $P(Y_i = m) = \pi_m$ for $m = 0, 1$, where $\pi_0 + \pi_1 = 1$, and $0 < \pi_0 < 1$. Show that $P(Y_i = m|X_i), m = 0, 1$, satisfies the logistic model, that is

$$logit\,(P(Y_i = 1|X_i, \alpha)) = \alpha_0 + \alpha_1 X_i$$

We have distribution of $P(Y_i = m|X_i), m = 0, 1$

$$\begin{aligned}
P(Y_i = m|X_i, \alpha) &= \frac{P(Y_i, X_i)}{P(X_i)} = \frac{P(X_i|Y_i)P(Y_i)}{P(X_i)} \\
P(Y_i = 1|X_i, \alpha) &= \frac{P(X_i|Y_i = 1)P(Y_i = 1)}{P(X_i)} \\
&= \frac{exp(-1/2\sigma^2(x_i - \mu_1)^2)\pi_1}{exp(-1/2\sigma^2(x_i - \mu_1)^2)\pi_1 + exp(-1/2\sigma^2(x_i - \mu_0)^2)\pi_0} \\
P(Y_i = 0|X_i, \alpha) &= \frac{P(X_i|Y_i = 0)P(Y_i = 0)}{P(X_i)} \\
&= \frac{exp(-1/2\sigma^2(x_i - \mu_0)^2)\pi_0}{exp(-1/2\sigma^2(x_i - \mu_1)^2)\pi_1 + exp(-1/2\sigma^2(x_i - \mu_0)^2)\pi_0} \\
logit\,(P(Y_i = 1|X_i, \alpha)) &= log\frac{P(Y_i = 1|X_i, \alpha)}{P(Y_i = 0|X_i, \alpha)} \\
&= log(\pi_1/\pi_0) - \frac{(x_i - \mu_1)^2}{2\sigma^2} + \frac{(x_i - \mu_0)^2}{2\sigma^2} \\
&= log(\pi_1/\pi_0) + \frac{\mu_0^2 - \mu_1^2}{2\sigma^2} + \frac{(\mu_1 - \mu_0)}{\sigma^2}x_i \\
\text{In which,}\alpha = (\alpha_0, \alpha_1) &= \left(log(\pi_1/\pi_0) + \frac{\mu_0^2 - \mu_1^2}{2\sigma^2}, \frac{(\mu_1 - \mu_0)}{\sigma^2}\right)^T
\end{aligned}$$

3.1.1 Likelihood function for conditional distribution

item[(b)] Based on the logistic model in part (a), give the explicit form of the Newton-Raphson algorithm for calculating the maximum likelihood estimate of α, denoted by $\hat{\alpha} = (\hat{\alpha_0}, \hat{\alpha_1})$, and derive the asymptotic covariance matrix of α.

This question will need to give the likelihood function first.

$Y_i|X_i$ follows a binomial distribution

$$\begin{aligned}
p(Y_i|\alpha) &= P(Y_i = 1|X_i, \alpha)^{I(y_i=1)} P(Y_i = 0|X_i, \alpha)^{I(y_i=0)} \\
logp(Y_i|\alpha) &= I(y_i = 1)logP(Y_i = 1|X_i, \alpha) + I(y_i = 0)logP(Y_i = 0|X_i, \alpha) \\
ln(Y_i|\alpha) &= \sum_{i=1}^{n} I(y_i = 1)logP(Y_i = 1|X_i, \alpha) + I(y_i = 0)logP(Y_i = 0|X_i, \alpha) \\
&= \sum_{i=1}^{n} I(y_i = 1)logP(Y_i = 1) + (1 - I(y_i = 1))log(1 - P(Y_i = 1)) \\
&= \sum_{i=1}^{n} I(y_i = 1)logP(Y_i = 1)/(1 - P(Y_i = 1)) + log(1 - P(Y_i = 1))
\end{aligned}$$

Let $\theta = logP(Y_i = 1)/(1 - P(Y_i = 1))$

$$\begin{aligned}
ln(Y_i|\theta) &= \sum_{i=1}^{n} I(y_i = 1)\theta - log(1 + exp(\theta)) \\
ln(Y_i|\alpha) &= \sum_{i=1}^{n} y_i(\alpha_0 + \alpha_1 x_i) - log\,(1 + exp(\alpha_0 + \alpha_1 x_i))
\end{aligned}$$

Find MLE for α

$$\begin{aligned}
\frac{\partial ln(Y_i|\alpha)}{\partial \alpha_0} &= \sum_{i=1}^{n} y_i - (1 + exp(\alpha_0 + \alpha_1 x_i))^{-1} exp(\alpha_0 + \alpha_1 x_i) \\
\frac{\partial ln(Y_i|\alpha)}{\partial \alpha_1} &= \sum_{i=1}^{n} y_i x_i - (1 + exp(\alpha_0 + \alpha_1 x_i))^{-1} exp(\alpha_0 + \alpha_1 x_i) x_i \\
\frac{\partial ln^2(Y_i|\alpha)}{\partial \alpha_0^2} &= -\sum_{i=1}^{n} \frac{exp(\alpha_0 + \alpha_1 x_i)}{[1 + exp(\alpha_0 + \alpha_1 x_i)]^2}, \qquad E[-\frac{\partial ln^2(Y_i|\alpha)}{\partial \alpha_0^2}] = n\pi_1(1 - \pi_1) \\
\frac{\partial ln^2(Y_i|\alpha)}{\partial \alpha_1^2} &= -\sum_{i=1}^{n} \frac{exp(\alpha_0 + \alpha_1 x_i)}{[1 + exp(\alpha_0 + \alpha_1 x_i)]^2} x_i x_i^T \\
\frac{\partial ln^2(Y_i|\alpha)}{\partial \alpha_0 \alpha_1} &= -\sum_{i=1}^{n} \frac{exp(\alpha_0 + \alpha_1 x_i)}{[1 + exp(\alpha_0 + \alpha_1 x_i)]^2} x_i \\
I_n(\alpha) &= -E[\frac{\partial ln^2(Y_i|\alpha)}{\partial \alpha^2}] \\
&= \begin{bmatrix} n\pi_1(1 - \pi_1) & \sum_{i=1}^{n} \pi_1(1 - \pi_1) x_i \\ \sum_{i=1}^{n} \pi_1(1 - \pi_1) x_i & \sum_{i=1}^{n} \pi_1(1 - \pi_1) x_i x_i^T \end{bmatrix}
\end{aligned}$$

So the N-R algorithm is

$$\alpha_{k+1} = \alpha_k - I_n(\alpha_k)^{-1}\frac{\partial ln(Y_i|\alpha_k)}{\partial \alpha_k}$$

The asymptotic distribution of α by CLT and covariance matrix

$$\sqrt{n}(\hat{\alpha} - \alpha) \xrightarrow{d} N\left(0, \Sigma\right)$$
$$\Sigma = \{\frac{1}{n}I_n(\alpha)\}^{-1}$$

3.1.2 Likelihood function for joint distribution

Write down the joint distribution of $\{(X_iY_i) : i = 1, 2..n\}$ and calculate the maximum likelihood estimate of θ, denoted by θ_F , and its asymptotic covariance matrix.
The joint distribution of $\{(X_iY_i) : i = 1, 2..n\}$

$$\begin{aligned}
p(X_i, Y_i) &= P(X_i|Y_i)P(Y_i) \\
p(Y_i = 1, X_i) &= \frac{1}{\sqrt{2\pi\sigma}}exp(-1/2\sigma^2(x_i - \mu_1)^2)\pi_1 \\
p(Y_i = 0, X_i) &= \frac{1}{\sqrt{2\pi\sigma}}exp(-1/2\sigma^2(x_i - \mu_0)^2)\pi_0 \\
p(X_i, Y_i) &= P(Y_i = 1, X_i)^{I(y_i=1)}P(Y_i = 0, X_i)^{I(y_i=0)} \\
&= \{\frac{1}{\sqrt{2\pi\sigma}}exp(-1/2\sigma^2(x_i - \mu_1)^2)\pi_1\}^{y_i}\{\frac{1}{\sqrt{2\pi\sigma}}exp(-1/2\sigma^2(x_i - \mu_0)^2)\pi_0\}^{1-y_i} \\
logp(X_i, Y_i) &= log\frac{1}{\sqrt{2\pi\sigma}} + y_ilog\pi_1 + (1 - y_i)log(1 - \pi_1) - \frac{(x_i - \mu_i)^2}{2\sigma^2}y_i - \frac{(x_i - \mu_0)^2}{2\sigma^2}(1 - y_i)
\end{aligned}$$

The log-likelihood function of $\{(X_i, Y_i) : i = 1, 2..n\}$

$$ln(X, Y) = nlog\frac{1}{\sqrt{2\pi\sigma}} + \sum_{i=1}^{n} y_ilog\pi_1 + (1 - y_i)log(1 - \pi_1) - \frac{(x_i - \mu_1)^2}{2\sigma^2}y_i - \frac{(x_i - \mu_0)^2}{2\sigma^2}(1 - y_i)$$

The MLE of θ could get by taking derivatives to log-likelihood function

$$\frac{\partial lnp(X,Y|\theta)}{\partial \pi_1} = \sum_{i=1}^{n} y_i/\pi_1 - (1-y_i)/(1-\pi_1) = 0$$

$$\frac{\partial lnp(X,Y|\theta)}{\partial \mu_1} = \sum_{i=1}^{n} \frac{y_i(x_i-\mu_1)}{\sigma^2} = 0$$

$$\frac{\partial lnp(X,Y|\theta)}{\partial \mu_0} = \sum_{i=1}^{n} \frac{(1-y_i)(x_i-\mu_0)}{\sigma^2} = 0$$

$$\frac{\partial lnp(X,Y|\theta)}{\partial \sigma^2} = -\frac{n}{2}1/\sigma^2 + \sum_{i=1}^{n} \frac{(x_i-\mu_1)^2 y_i}{2\sigma^4} + \sum_{i=1}^{n} \frac{(x_i-\mu_0)^2(1-y_i)}{2\sigma^4} = 0$$

$$\hat{\sigma^2} = \frac{\sum_{i=1}^{n}[(x_i-\mu_1)^2 y_i + (x_i-\mu_0)^2(1-y_i)]}{n}$$

$$\hat{\pi}_1 = \frac{\sum_{i=1}^{n} y_i}{n}, \qquad \hat{\mu_1} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} y_i}, \qquad \hat{\mu_0} = \frac{\sum_{i=1}^{n} x_i(1-y_i)}{\sum_{i=1}^{n}(1-y_i)}$$

The Fisher information matrix

$$\frac{\partial ln^2(X,Y|\theta)}{\partial \pi_1^2} = \sum_{i=1}^{n} -\frac{y_i}{\pi_1^2} - \frac{(1-y_i)}{(1-\pi_1)^2}, \qquad E[-\frac{\partial ln^2(X,Y|\theta)}{\partial \pi_1^2}] = \frac{1}{\pi_1(1-\pi_1)}$$
$$\frac{\partial ln^2(X,Y|\theta)}{\partial \mu_1^2} = \sum_{i=1}^{n} -\frac{y_i}{\sigma^2}, \qquad E[-\frac{\partial ln^2(X,Y|\theta)}{\partial \mu_1^2}] = \frac{\pi_1}{\sigma^2}$$
$$\frac{\partial ln^2(X,Y|\theta)}{\partial \mu_0^2} = \sum_{i=1}^{n} -\frac{(1-y_i)}{\sigma^2}, \qquad E[-\frac{\partial ln^2 p(X,Y|\theta)}{\partial \mu_0^2}] = \frac{1-\pi_1}{\sigma^2}$$
$$\frac{\partial ln^2(X,Y|\theta)}{\partial (\sigma^2)^2} = \frac{n}{2(\sigma^2)^2} - \sum_{i=1}^{n} \frac{(x_i-\mu_1)^2 y_i}{(\sigma^2)^3} - \sum_{i=1}^{n} \frac{(x_i-\mu_0)^2(1-y_i)}{(\sigma^2)^3}$$
$$E[-\frac{\partial ln^2(X,Y|\theta)}{\partial (\sigma^2)^2}] = \frac{1}{2\sigma^4}$$
$$\frac{\partial ln^2(X,Y|\theta)}{\partial \pi_1 \mu_1} = 0$$
$$\frac{\partial ln^2(X,Y|\theta)}{\partial \pi_1 \mu_0} = 0$$
$$\frac{\partial ln^2(X,Y|\theta)}{\partial \pi_1 \sigma} = 0$$
$$\frac{\partial ln^2(X,Y|\theta)}{\partial \mu_1 \mu_0} = 0$$
$$\frac{\partial ln^2(X,Y|\theta)}{\partial \mu_1 \sigma} = \sum_{i=1}^{n} -\frac{y_i(x_i-\mu_1)}{(\sigma^2)^2}, \qquad E[-\frac{\partial ln^2(X,Y|\theta)}{\partial \mu_1 \sigma}] = 0$$
$$\frac{\partial ln^2(X,Y|\theta)}{\partial \mu_0 \sigma} = \sum_{i=1}^{n} -\frac{(1-y_i)(x_i-\mu_0)}{(\sigma^2)^2}, \qquad E[-\frac{\partial ln^2(X,Y|\theta)}{\partial \mu_0 \sigma}] = 0$$

So we have covariance matrix, by CLT

$$I(\theta) = \frac{1}{n} I_n(\theta) = \frac{1}{n} E[-\frac{\partial ln^2 p(X,Y|\theta)}{\partial \theta^2}] = \begin{bmatrix} \frac{1}{\pi_1(1-\pi_1)} & 0 & 0 & 0 \\ 0 & \frac{\pi_1}{\sigma^2} & 0 & 0 \\ 0 & 0 & \frac{1-\pi_1}{\sigma^2} & 0 \\ 0 & 0 & 0 & \frac{1}{2\sigma^4} \end{bmatrix}$$
$$\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{d} N\left(0, \Sigma\right), \qquad \Sigma(\theta) = I(\theta)^{-1} = \begin{bmatrix} \pi_1(1-\pi_1) & 0 & 0 & 0 \\ 0 & \frac{\sigma^2}{\pi_1} & 0 & 0 \\ 0 & 0 & \frac{\sigma^2}{1-\pi_1} & 0 \\ 0 & 0 & 0 & 2\sigma^4 \end{bmatrix}$$

3.1.3 Fisher Information, Delta Method, Asymptotic Covariance

Calculate the asymptotic covariance matrix of $h(\hat{\theta}^F)$.

$$
\begin{aligned}
h(\theta^F) &= (\alpha_0, \alpha_1) = \left(log(\frac{\pi_1}{1-\pi_1}) + \frac{\mu_0^2 - \mu_1^2}{2\sigma^2}, \frac{(\mu_1 - \mu_0)}{\sigma^2} \right)^T \\
\frac{\partial h(\theta^F)}{\partial \pi_1} &= (\frac{1}{\pi_1} + \frac{1}{1-\pi_1}, 0)^T \\
\frac{\partial h(\theta^F)}{\partial \mu_1} &= (-\frac{\mu_1}{\sigma^2}, \frac{1}{\sigma^2})^T \\
\frac{\partial h(\theta^F)}{\partial \mu_0} &= (\frac{\mu_0}{\sigma^2}, -\frac{1}{\sigma^2})^T \\
\frac{\partial h(\theta^F)}{\partial \sigma^2} &= \left(-\frac{(\mu_0^2 - \mu_1^2)}{2\sigma^4}, -\frac{(\mu_1 - \mu_0)}{\sigma^4} \right)^T \\
\sqrt{n}(h(\hat{\theta}^F) - h(\theta^F)) &\xrightarrow{d} N\left(0, \Sigma_h\right)
\end{aligned}
$$

By delta method,

$$
\begin{aligned}
\Sigma^h &= h(\theta^F)'\Sigma(\theta)(\theta^F)'^T \\
&= \begin{bmatrix} \frac{1}{\pi_1} + \frac{1}{1-\pi_1} & -\frac{\mu_1}{\sigma^2} & \frac{\mu_0}{\sigma^2} & -\frac{(\mu_0^2-\mu_1^2)}{2\sigma^4} \\ 0 & \frac{1}{\sigma^2} & -\frac{1}{\sigma^2} & -\frac{(\mu_1-\mu_0)}{\sigma^4} \end{bmatrix} \begin{bmatrix} \pi_1(1-\pi_1) & 0 & 0 & 0 \\ 0 & \frac{\sigma^2}{\pi_1} & 0 & 0 \\ 0 & 0 & \frac{\sigma^2}{1-\pi_1} & 0 \\ 0 & 0 & 0 & 2\sigma^4 \end{bmatrix} \begin{bmatrix} \frac{1}{\pi_1} + \frac{1}{1-\pi_1} & 0 \\ -\frac{\mu_1}{\sigma^2} & \frac{1}{\sigma^2} \\ \frac{\mu_0}{\sigma^2} & -\frac{1}{\sigma^2} \\ -\frac{(\mu_0^2-\mu_1^2)}{2\sigma^4} & -\frac{(\mu_1-\mu_0)}{\sigma^4} \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{\pi_1(1-\pi_1)} + \frac{\mu_0}{(1-\pi_1)\sigma^2} + \frac{\mu_1}{\pi_1\sigma^2} + \frac{(\mu_0^2-\mu_1^2)^2}{2\sigma^4} & -\frac{1}{\sigma^2}(\frac{\mu_0}{1-\pi_1} + \frac{\mu_1}{\pi_1}) + \frac{(\mu_1-\mu_0)(\mu_0^2-\mu_1^2)}{\sigma^4} \\ -\frac{1}{\sigma^2}(\frac{\mu_0}{1-\pi_1} + \frac{\mu_1}{\pi_1}) + \frac{(\mu_1-\mu_0)(\mu_0^2-\mu_1^2)}{\sigma^4} & \frac{1}{\sigma^2\pi_1(1-\pi_1)} + \frac{2(\mu_1-\mu_0)^2}{\sigma^4} \end{bmatrix}
\end{aligned}
$$

3.1.4 Invariance of MLE estimator

MLE estimators does not depend on the log-likelihood function. And it does not change with the form.

In this part, suppose that $\mu_0 = \mu_1$. Show that $Cov(\hat{\alpha})^{-1}Cov(h(\hat{\theta}^F))$ converges to a matrix which does not depend on θ. Interpret this result.

When the parameters change, the likelihood function will change, and so does the Fisher Information. So we need to recalculate all the covariance matrix.

If $\mu_0 = \mu_1$, then $\alpha = (\alpha_0, \alpha_1)^T = (log(\pi_1/\pi_0), 0)^T$ The covariance matrix of α

$$\begin{aligned}
\alpha_0 &= log(\pi_1/\pi_0) \\
ln(Y_i|\alpha) &= \sum_{i=1}^{n} y_i(\alpha_0) - log\left(1 + exp(\alpha_0)\right) \\
\frac{\partial ln(Y_i|\alpha)}{\partial \alpha_0} &= \sum_{i=1}^{n} y_i - \frac{exp\alpha_0}{1 + exp\alpha_0} \\
\frac{\partial ln^2(Y_i|\alpha)}{\partial \alpha_0^2} &= \sum_{i=1}^{n} -\frac{exp\alpha_0}{(1 + exp\alpha_0)^2} \\
I_n(\alpha) &= E[-\frac{\partial ln^2(Y_i|\alpha)}{\partial \alpha_0^2}] = \sum_{i=1}^{n} \frac{exp\alpha_0}{(1 + exp\alpha_0)^2}
\end{aligned}$$

Need to pay attention that, we used the logistic model for $Y|X$ to get the Fisher information for α.

$$\begin{aligned}
ln(\theta) &= nlog\frac{1}{\sqrt{2\pi}\sigma} + \sum_{i=1}^{n} y_i log\pi_1 + (1 - y_i)log(1 - \pi_1) - \frac{(x_i - \mu)^2}{2\sigma^2} \\
\frac{\partial ln(\theta)}{\partial \pi_1} &= \sum_{i=1}^{n} \frac{y_i}{\pi_1} - \frac{1 - y_i}{1 - \pi_1} \\
\frac{\partial ln^2(\theta)}{\partial \pi_1^2} &= \sum_{i=1}^{n} -\frac{y_i}{\pi_1^2} - \frac{1 - y_i}{(1 - \pi_1)^2}, \qquad E[-\frac{\partial ln^2 p(\theta)}{\partial \pi_1^2}] = n\frac{\pi_1}{(1 - \pi_1))} \\
\frac{\partial ln(\theta)}{\partial \mu} &= \sum_{i=1}^{n} \frac{x_i - \mu}{\sigma^2} \\
\frac{\partial ln^2(\theta)}{\partial \mu^2} &= \sum_{i=1}^{n} -\frac{1}{\sigma^2} \\
\frac{\partial ln(\theta)}{\partial \sigma^2} &= -\frac{n}{2} 1/\sigma^2 + \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma^4} \\
\frac{\partial ln^2(\theta)}{\partial (\sigma^2)^2} &= \frac{n}{2(\sigma^2)^2} - \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{\sigma^6}, \qquad E[-\frac{\partial ln^2 p(\theta)}{\partial (\sigma^2)^2}] = \frac{n}{2\sigma^4} \\
\frac{\partial ln^2(\theta)}{\partial \mu \sigma^2} &= \sum_{i=1}^{n} -\frac{x_i - \mu}{\sigma^4}, \qquad E[-\frac{\partial ln^2(\theta)}{\partial \mu \sigma^2}] = 0
\end{aligned}$$

Then we have Fisher information $I_n(\theta)$

$$
\begin{aligned}
I_n(\theta) &= E[-\frac{\partial ln^2(\theta)}{\partial \theta^2}] \\
&= \begin{bmatrix} n\frac{\pi_1}{(1-\pi_1))} & 0 & 0 \\ 0 & \frac{n}{\sigma^2} & 0 \\ 0 & 0 & \frac{n}{2\sigma^4} \end{bmatrix} \\
Cov(\hat{\alpha})^{-1} &= I_n(\alpha) = n\pi_1(1-\pi_1) \\
\frac{\partial h}{\partial \theta} &= (\frac{1}{\pi_1(1-\pi_1)}, 0, 0)
\end{aligned}
$$

Then we have

$$
\begin{aligned}
Cov(\hat{\alpha})^{-1}\Sigma^h &= I_n(\alpha)\frac{\partial h}{\partial \theta} I_n(\theta)^{-1} \frac{\partial h}{\partial \theta}^T \\
&= n\pi_1(1-\pi_1)(\frac{1}{\pi_1(1-\pi_1)}, 0, 0) \begin{bmatrix} \pi_1(1-\pi_1)/n & 0 & 0 \\ 0 & \sigma^2/n & 0 \\ 0 & 0 & 2\sigma^4/n \end{bmatrix} (\frac{1}{\pi_1(1-\pi_1)}, 0, 0)^T \\
&= 1
\end{aligned}
$$

So we have $Cov(\hat{\alpha})^{-1}Cov(h(\hat{\theta}^F))$ converges to a matrix which does not depend on θ.

3.1.5 Different Scenarios

Now suppose that π_1 is known. Will the results of (b) - (e) be changed? Please explain. If so, then derive the corresponding results and compare with those obtained above.

If π_1 is known,

(i) For (b), does not change as the parameters are $\alpha = (\alpha_0, \alpha_1)^T$ which does not involve π_1.

$$
\begin{aligned}
I_n(\alpha) &= -E[\frac{\partial ln^2 p(Y_i|\alpha)}{\partial \alpha^2}] \\
&= \begin{bmatrix} n\pi_1(1-\pi_1) & \sum_{i=1}^n \pi_1(1-\pi_1)x_i \\ \sum_{i=1}^n \pi_1(1-\pi_1)x_i & \sum_{i=1}^n \pi_1(1-\pi_1)x_i x_i^T \end{bmatrix} \\
Cov(\alpha) = I_n(\alpha)^{-1} &= \frac{1}{[\sum_{i=1}^n nx_i^2 - (\sum_{i=1}^n x_i)^2]\pi_1(1-\pi_1)} \begin{bmatrix} \sum_{i=1}^n nx_i^2 & -\sum_{i=1}^n x_i \\ -\sum_{i=1}^n x_i \sum_{i=1}^n \pi_1(1-\pi_1)x_i & n \end{bmatrix}
\end{aligned}
$$

(ii) For (c), it involves π_1, so the result will change. We have covariance matrix for

$\theta = (\mu_1, \mu_0, \sigma^2)^T$,

$$I(\theta) = E[-\frac{1}{n}\frac{\partial ln^2 p(X,Y|\theta)}{\partial \theta^2}], \qquad = \begin{bmatrix} \frac{\pi_1}{\sigma^2} & 0 & 0 \\ 0 & \frac{1-\pi_1}{\sigma^2} & 0 \\ 0 & 0 & \frac{1}{2\sigma^4} \end{bmatrix}$$

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \Sigma), \qquad \Sigma = I(\theta)^{-1} = \begin{bmatrix} \frac{\sigma^2}{\pi_1} & 0 & 0 \\ 0 & \frac{\sigma^2}{1-\pi_1} & 0 \\ 0 & 0 & 2\sigma^4 \end{bmatrix}$$

(iii) For (d), the $h(\theta)$ does not involve π_1, but the Jacobian matrix and $I(\theta)$ will change when π_1 is known. We have covariance matrix for $h(\theta) = c(\mu, \sigma^2)$.

$$h(\theta^F) = (\alpha_0, \alpha_1) = \left(log(\frac{\pi_1}{1-\pi_1}) + \frac{\mu_0^2 - \mu_1^2}{2\sigma^2}, \frac{(\mu_1 - \mu_0)}{\sigma^2} \right)^T$$

$$\sqrt{n}(h(\hat{\theta}^F) - h(\theta^F)) \xrightarrow{d} N(0, \Sigma_h)$$

$$h(\theta^F)' = \begin{bmatrix} -\frac{\mu_1}{\sigma^2} & \frac{\mu_0}{\sigma^2} & -\frac{(\mu_0^2-\mu_1^2)}{2\sigma^4} \\ \frac{1}{\sigma^2} & -\frac{1}{\sigma^2} & -\frac{(\mu_1-\mu_0)}{\sigma^4} \end{bmatrix}$$

$$\Sigma(\theta) = \begin{bmatrix} \frac{\sigma^2}{\pi_1} & 0 & 0 \\ 0 & \frac{\sigma^2}{1-\pi_1} & 0 \\ 0 & 0 & 2\sigma^4 \end{bmatrix}$$

$$\Sigma^h = h(\theta^F)'\Sigma(\theta)(\theta^F)'^T$$

$$= \begin{bmatrix} \frac{\mu_0}{(1-\pi_1)\sigma^2} + \frac{\mu_1}{\pi_1\sigma^2} + \frac{(\mu_0^2-\mu_1^2)^2}{2\sigma^4} & -\frac{1}{\sigma^2}(\frac{\mu_0}{1-\pi_1} + \frac{\mu_1}{\pi_1}) + \frac{(\mu_1-\mu_0)(\mu_0^2-\mu_1^2)}{\sigma^4} \\ -\frac{1}{\sigma^2}(\frac{\mu_0}{1-\pi_1} + \frac{\mu_1}{\pi_1}) + \frac{(\mu_1-\mu_0)(\mu_0^2-\mu_1^2)}{\sigma^4} & \frac{2(\mu_1-\mu_0)^2}{\sigma^4} \end{bmatrix}$$

(iv) For (e), the only parameter that need to estimate is $\alpha_0 = log(\pi_1/(1-\pi_1))$, which is now known. The question is meaningless.

4 Likelihood for one random variable

To calculate the covariance matrix, we will use the MGF and take derivatives. Or use the cumulant function KGF to get the covariance.
Use one random variable for the two way contingency table. While the Fisher information is the inverse of the covariance matrix, however we don't use Fisher information to calculate covariance matrix due to the math computation.

For one random variable Y:

$$\begin{aligned}
p(\theta) &= \prod_{i=1}^{n}\prod_{j=1}^{J}\pi_j^{I(Y_i=j)}, \qquad \theta = (\pi_1, \pi_2, ...\pi_J)' \\
lnp(\theta) &= \sum_{i=1}^{n}\sum_{j=1}^{J} I(Y_i = j)log(\pi_j) = \sum_{j=1}^{J} n_j log(\pi_j) \\
M_X(t) &= E[exp(t^T X)] = E[exp(t^T(Y_1 + Y_2 + ...Y_n))] = E[exp(t^T Y_1 + t^T Y_2 + ...t^T Y_n)] \\
&= E[\prod_{i=1}^{n} exp(t^T Y_i)] \\
&= \prod_{i=1}^{n} E[exp(t^T Y_i)] \qquad \text{(by independence)} \\
&= \prod_{i=1}^{n} M_{Y_i}(t) = \prod_{i=1}^{n} P(Y_i = 1)e^{ty_i} \qquad \text{by MGF of discrete variable } Y_i \\
&= \left(\sum_{j=1}^{J} \pi_j exp(t_j)\right)^n \qquad \text{by MGF of multinoulli}
\end{aligned}$$

The MGF for bernoulli distribution

$$M_X(t) = 1 - p + pexp(t), \qquad K_X(t) = log(1 - p + pexp(t))$$

For multinomial distribution

$$\begin{aligned}
M_X(t) &= (1 - p + pexp(t))^n, \qquad K_X(t) = nlog(1 - p + pexp(t)) \\
E[n_j] &= n\pi_j, \qquad Var[n_j] = n\pi_j(1 - \pi_j), \qquad Cov(n_j, n_k) = -n\pi_j\pi_k, (j \neq k)
\end{aligned}$$

Thus to compute covariance matrix

$$\begin{aligned}
E(X_1X_2) &= \frac{\partial^2 M_X(t)}{\partial t_i \partial t_j}|_{t_i=t_j=0} \\
&= \frac{\partial \left(n(\pi_i e^{t_i})(\sum_{k=1}^K \pi_k e^{t_k})^{n-1}\right)'}{\partial t_j} \\
&= n(n-1)(\sum_{k=1}^K \pi_k e^{t_k})^{n-2}\pi_i\pi_j|_{t_i=t_j=0} = n(n-1)\pi_i\pi_j \\
E(X_i) &= n\pi_i \\
Cov(X_i, X_j) &= E(X_iX_2) - E(X_i)E(X_j) = n(n-1)\pi_i\pi_j - n^2\pi_i\pi_j = -n\pi_i\pi_j \\
Var(X_i) &= E(X_i^2) - E(X_i)^2 \\
E(X_i^2) &= \frac{\partial^2 M(t)}{\partial t\, \partial t} = \frac{\partial \left(n(\pi_i e^{t_i})(\sum_{k=1}^K \pi_k e^{t_k})^{n-1}\right)'}{\partial t_i} \\
&= n(\sum_{k=1}^K \pi_k e^{t_k})^{n-1}\pi_i e^{t_i} + n(n-1)(\sum_{k=1}^K \pi_k e^{t_k})^{n-2}\pi_i\pi_i e^{2t_i}|_{t_i=0} \\
&= n\pi_i + n(n-1)\pi_i^2 = n\pi_i(1-\pi) \\
Var(X_i/n) &= \frac{1}{n^2}Var(X_i) = \frac{1}{n}\pi_i(1-\pi_i)
\end{aligned}$$

Thus the covariance matrix is

$$\begin{aligned}
\Sigma &= \begin{bmatrix} \pi_1(1-\pi_1) & -\pi_1\pi_2 & & -\pi_i\pi_j \\ -\pi_j\pi_i & \pi_i(1-\pi_i) & & \\ .. & .. & .. & .. \end{bmatrix} \\
&= diag(\pi_j) - \theta\theta^T
\end{aligned}$$

Here is the question, why do we think the covariance matrix of X is the covariance matrix of π?

$$\begin{aligned}
n^{-1}(n_1, n_2, ..n_I) &= n^{-1}\sum_{i=1}^n [1(X_i = 1), 1(X_i = 2), ..1(X_i = I)] \\
&= E[1(X_i = 1), 1(X_i = 2), ..1(X_i = I)] = [\pi_1, \pi_2, ..\pi_I]
\end{aligned}$$

4.1 Likelihood for multinomial sampling variable in contingency table

$$
\begin{aligned}
p(\pi_{ij}) &= \prod_{i=1}^{I}\prod_{j=1}^{J}\pi_{ij}^{n_{ij}}, \qquad \pi_{ij} > 0, \quad \sum_i\sum_j \pi_{ij} = 1 \\
\theta &= c(\pi_{11}, \pi_{12}, \pi_{21}) \\
ln(\theta) &= \sum_i\sum_j n_{ij} log\pi_{ij} = n_{11}log\pi_{11} + n_{12}log\pi_{12} + n_{21}log\pi_{21} + n_{22}log\pi_{22} \\
&= n_{11}log\pi_{11} + n_{12}log\pi_{12} + n_{21}log\pi_{21} + n_{22}log(1-\pi_{11}-\pi_{12}-\pi_{21})
\end{aligned}
$$

We can calculate the MLE estimate of π_{ij}

$$
\begin{aligned}
\frac{\partial ln(\theta)}{\partial \pi} &= \frac{n_{11}}{\pi_{11}} - \frac{n_{22}}{(1-\pi_{11}-\pi_{12}-\pi_{21})} = 0, \\
\pi_{11} &= \frac{n_{11}}{n_{22}}\pi_{22}, \qquad \pi_{12} = \frac{n_{12}}{n_{22}}\pi_{22}, \qquad \pi_{21} = \frac{n_{21}}{n_{22}}\pi_{22}, \qquad \pi_{22} = \frac{n_{22}}{n} \\
\pi_{ij} &= \frac{n_{ij}}{n}
\end{aligned}
$$

Similarly as above, we need to find the $Cov(\theta)$, start from finding $Var(\pi_{11}, \pi_{12}), Cov(\pi_{11}, \pi_{12})$.

4.2 Pearson Statistics

Question: why the Pearson Statistics use the square of difference between sample mean and expected mean, then divided by the expected mean?

We need to know what is the distribution of the Pearson Statistics. First, we start from the asymptotic distribution of the sample percentage $\hat{\pi} = \frac{n_i}{n}$.

$$
\begin{aligned}
\sqrt{n}(\frac{n_1}{n} - \pi_1, \frac{n_2}{n} - \pi_2, ..\frac{n_I}{n} - \pi_I) &\xrightarrow{L} N(0, \Sigma^*) \\
\Sigma^* &= diag\{\pi\} - \pi\pi^T
\end{aligned}
$$

We need to pay attention that, the $\pi_1, \pi_2, ..\pi_I$ are joint distributed. The Pearson statistics comes from a function of $(\frac{n_1}{n} - \pi_1, \frac{n_2}{n} - \pi_2, ..\frac{n_I}{n} - \pi_I)$, which could use delta method. The normal distribution is always associated with chi-square distribution.

$$
\begin{aligned}
\Gamma &= diag\{\pi_1, \pi_2, ...\pi_I\} \\
\sqrt{n}\Gamma^{-1/2}\left(\frac{n_1}{n} - \pi_1, \frac{n_2}{n} - \pi_2, ..\frac{n_I}{n} - \pi_I\right) &\xrightarrow{L} N(0, \Gamma^{-1/2}\Sigma^*\Gamma^{-1/2})
\end{aligned}
$$

Because Γ is a diagonal matrix, so it could be multiplied directly to the left or right of a matrix, and it only works on the diagonal element.

$$
\begin{aligned}
\Gamma^{-1/2}\Sigma^*\Gamma^{-1/2} &= \Gamma^{-1/2}\Gamma^{1/2}(I-\sqrt{\pi}^{\otimes 2})\left(\Gamma^{-1/2}\Gamma^{1/2}\right)^T \\
tr(I-\sqrt{\pi}^{\otimes 2}) &= I-1 \\
tr(\Gamma^{-1/2}\Sigma^*\Gamma^{-1/2}) &= tr(\Sigma^*\Gamma^{-1/2}\Gamma^{-1/2}) = tr(\Sigma^*\Gamma^{-1}) \\
&= tr([\Gamma-\pi\pi^T]\Gamma^{-1}) = tr(\Gamma\Gamma^{-1}) - tr(\pi\pi^T\Gamma^{-1}) = I-1
\end{aligned}
$$

The Pearson Chi-square statistic is defined as

$$
\chi^2 = n\sum_{j=1}^{I}(\frac{n_j}{n}-\pi_j)^2/\pi_j = \left[\sqrt{n}\Gamma^{-1/2}\left(\frac{n_1}{n}-\pi_1, \frac{n_2}{n}-\pi_2, ..\frac{n_I}{n}-\pi_I\right)\right]^{\otimes 2}
$$

which converge to $\chi^2(I-1)$ as $n \to \infty$.

4.3 Odds ratio

The covariance of odds ratio by delta method. We simplify 2×2 table as $\pi_{11}=\pi_1, \pi_{12}=\pi_2, \pi_{21}=\pi_3, \pi_{22}=\pi_4$.

$$
\begin{aligned}
g(\pi) &= \frac{\pi_{22}\pi_{11}}{\pi_{12}\pi_{21}} \qquad \pi = (\pi_{11}, \pi_{12}, \pi_{21}, \pi_{22}) \\
\sqrt{n}\left(g(\hat{\pi})-g(\pi)\right) &\xrightarrow{d} N\left(0, \left(\frac{\partial g(\pi)}{\partial \pi}\right)\Sigma\left(\frac{\partial g(\pi)}{\partial \pi}\right)^T\right) \\
\frac{\partial g(\pi)}{\partial \pi} &= \left(\frac{\partial g}{\partial \pi_{11}}, \frac{\partial g}{\pi_{12}}, \frac{\partial g}{\partial \pi_{21}}, \frac{\partial g}{\partial \pi_{22}}\right)^T \\
&= \left(\frac{\pi_{22}}{\pi_{21}\pi_{12}}, \frac{-\pi_{11}\pi_{22}}{\pi_{21}\pi_{12}^2}, \frac{-\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}^2}, \frac{\pi_{11}}{\pi_{21}\pi_{12}}\right)^T \\
\Sigma^* &= g(\pi)^2(\frac{1}{\pi_{11}}+\frac{1}{\pi_{12}}+\frac{1}{\pi_{21}}+\frac{1}{\pi_{22}})
\end{aligned}
$$

So that,

$$
Var(\hat{R}) = \frac{1}{n}\Sigma^*
$$

We consider $log\hat{R}$ instead of $\hat{R}$, because $log\hat{R}$ converges rapidly to a normal distribution compared to $\hat{R}$.

$$\begin{aligned}
log(\hat{R}) &= log\pi_1 + \log\pi_2 - \log\pi_3 \log\pi_4 \\
\frac{\partial g(\pi)}{\partial\pi} &= \left(\frac{1}{\pi_{11}}, -\frac{1}{\pi_{12}}, -\frac{1}{\pi_{21}}, \frac{1}{\pi_{22}}\right)^T \\
Var(log(\hat{R})) &= \frac{1}{n}\tilde{\Sigma} \\
\tilde{\Sigma} &= \left(\frac{\partial g(\pi)}{\partial\pi}\right)^T \Sigma \left(\frac{\partial g(\pi)}{\partial\pi}\right) \\
log(\hat{R}) &= \frac{1}{n}\left(\frac{1}{\hat{\pi}_{11}} + \frac{1}{\hat{\pi}_{12}} + \frac{1}{\hat{\pi}_{21}} + \frac{1}{\hat{\pi}_{22}}\right) \\
s.e.log(\hat{R}) &= \frac{1}{\sqrt{n}}\sqrt{\frac{1}{\hat{\pi}_{11}} + \frac{1}{\hat{\pi}_{12}} + \frac{1}{\hat{\pi}_{21}} + \frac{1}{\hat{\pi}_{22}}}
\end{aligned}$$

4.4 Retrospective vs. Prospective vs. Cross Sectional Study

4.4.1 Retrospective

For retrospective study, the Y is fixed

$$\begin{aligned}
\theta = p(X=1|Y=1) &= \frac{\pi_{11}}{\pi_{11}+\pi_{21}} \\
1-\theta = p(X=0|Y=1) &= \frac{\pi_{21}}{\pi_{11}+\pi_{21}} \\
\gamma = p(X=1|Y=0) &= \frac{\pi_{12}}{\pi_{12}+\pi_{22}} \\
1-\gamma = p(X=0|Y=0) &= \frac{\pi_{22}}{\pi_{12}+\pi_{22}}
\end{aligned}$$

$X|Y$ are binomial distribution, which is different from above multinomial distribution. And the $X|Y=0, X|Y=1$ are independent.

$$\begin{aligned}
p(\theta,\gamma) &= \theta^{n_{11}}(1-\theta)^{n_{21}}\gamma^{n_{12}}(1-\gamma)^{n_{22}} \\
lnp(\theta,\gamma) &= n_{11}log\theta + n_{21}(1-\theta) + n_{12}log\gamma + n_{22}log(1-\gamma) \\
\frac{\partial ln}{\partial\theta} &= \frac{n_{11}}{\theta} - \frac{n_{21}}{1-\theta} = 0 \\
\hat{\theta} &= \frac{n_{11}}{n_{11}+n_{21}} \\
\frac{\partial ln}{\partial\gamma} &= \frac{n_{12}}{\gamma} - \frac{n_{22}}{1-\gamma} = 0 \\
\hat{\gamma} &= \frac{n_{12}}{n_{12}+n_{22}}
\end{aligned}$$

Then get covariance matrix by delta method, binomial distribution variance is $np(1-p)$

$$
\begin{aligned}
g(\theta) &= \frac{n_{11}n_{22}}{n_{21}n_{12}} = \frac{\theta/(1-\theta)}{\gamma/(1-\gamma)} \\
\sqrt{n}\left(\theta - \hat{\theta}\right) &\xrightarrow{d} N(0, \Sigma) \\
\Sigma &= \begin{bmatrix} \theta(1-\theta) & 0 \\ 0 & \gamma(1-\gamma) \end{bmatrix} \\
\sqrt{n}\left(g(\hat{\theta}) - g(\theta)\right) &\xrightarrow{d} N(0, g(\theta)'\Sigma^{New} g(\theta)'^{T}) \\
g(\theta)' &= \left(\frac{(1-\gamma)/\gamma}{1/(1-\theta)^2}, \frac{\theta/(1-\theta)}{-1/\gamma^2}\right)
\end{aligned}
$$

The standard error for odds ratio in retrospective study

$$
\begin{aligned}
se(\hat{R}) &= \hat{R}\sqrt{\frac{1}{n_{.1}\hat{\pi}_{X=2|Y=1}\hat{\pi}_{X=1|Y=1}} + \frac{1}{n_{.2}\hat{\pi}_{X=2|Y=2}\hat{\pi}_{X=1|Y=2}}} \\
\hat{\pi}_{X=2|Y=1} &= \frac{n_{21}}{n_{11}+n_{21}} \\
\hat{\pi}_{X=1|Y=1} &= \frac{n_{11}}{n_{11}+n_{21}} \\
\hat{\pi}_{X=2|Y=2} &= \frac{n_{12}}{n_{12}+n_{22}} \\
\hat{\pi}_{X=1|Y=2} &= \frac{n_{12}}{n_{12}+n_{22}}
\end{aligned}
$$

$n_{.1} = n_{11} + n_{21}, \quad n_{.2} = n_{12} + n_{22}$

$$
\begin{aligned}
se(\hat{R}) &= \frac{n_{22}n_{11}}{(n_{21}n_{12})}\sqrt{\frac{n_{11}+n_{21}}{n_{11}n_{21}} + \frac{n_{12}+n_{22}}{n_{12}n_{22}}} \\
&= \frac{n_{22}n_{11}}{(n_{21}n_{12})}\sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}}
\end{aligned}
$$

4.4.2 Prospective

The standard error for odds ratio in prospective study

$$
\begin{aligned}
se(\hat{R}) &= \hat{R}\sqrt{\frac{1}{n_{1.}\hat{\pi}_{Y=2|X=1}\hat{\pi}_{Y=1|X=1}} + \frac{1}{n_{2.}\hat{\pi}_{Y=2|X=2}\hat{\pi}_{Y=1|X=2}}} \\
\hat{\pi}_{Y=2|X=1} &= \frac{n_{12}}{n_{11}+n_{12}} \\
\hat{\pi}_{Y=1|X=1} &= \frac{n_{11}}{n_{11}+n_{12}} \\
\hat{\pi}_{Y=2|X=2} &= \frac{n_{22}}{n_{21}+n_{22}} \\
\hat{\pi}_{Y=1|X=2} &= \frac{n_{21}}{n_{21}+n_{22}}
\end{aligned}
$$

$n_{1.} = n_{11} + n_{12}, \quad n_{2.} = n_{21} + n_{22}$

$$
\begin{aligned}
se(\hat{R}) &= \frac{n_{22}n_{11}}{(n_{21}n_{12})}\sqrt{\frac{n_{11}+n_{12}}{n_{11}n_{12}} + \frac{n_{21}+n_{22}}{n_{21}n_{22}}} \\
&= \frac{n_{22}n_{11}}{(n_{21}n_{12})}\sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}}
\end{aligned}
$$

4.4.3 Cross-Sectional

For cross-sectional study, we only have the total n fixed. That is the difference for each scenario.
To calculate the covariance matrix, we will use the MGF and take derivatives. Or use the cumulant function KGF to get the covariance.
Use one random variable for the two way contingency table. While the Fisher information is the inverse of the covariance matrix, however we don't use Fisher information to calculate covariance matrix due to the math computation.

Show that the sample odds ratio $\hat{R} = n_{22}n_{11}/(n_{21}n_{12})$ has the same standard error for cross-sectional, prospective and retrospective studies.

The standard error for odds ratio in cross sectional study

$$
\begin{aligned}
se(\hat{R}) &= \frac{\hat{R}}{\sqrt{n}}\sqrt{\frac{1}{\hat{\pi_{11}}} + \frac{1}{\hat{\pi_{12}}} + \frac{1}{\hat{\pi_{21}}} + \frac{1}{\hat{\pi_{22}}}} \\
&= \frac{n_{22}n_{11}}{(n_{21}n_{12})}\sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}}
\end{aligned}
$$

By comparing the above standard errors in three types of studies, we see that they have same standard errors. Odds ratio is invariant in terms of sampling method. Similarly the coefficient of a particular covariate is associated with the odds ratio of the covariate, which is invariant with prospective and retrospective studies. Check out p747.

4.5 Hypergeometric distribution

Dervie the hypergeometric distribution

$$\begin{aligned}
p(n_{11}|n_{1.}, n_{.1}, n, \Xi) &= \frac{p(n_{11}, n_{1.}, n_{.1}, |n)}{p(n_{1.}, n_{.1}, |n)} \\
&= \frac{n!}{n_{11}!n_{12}!n_{21}!n_{22}!}\Xi^{n_{11}}\frac{n!}{n_{11}!n_{12}!n_{21}!n_{22}!} \\
&= \frac{n!n_{1.}!(n-n_{1.})!}{n_{1.}!(n-n_{1.})!n_{11}!n_{12}!n_{21}!n_{22}!} \\
&= \binom{n}{n_{1.}}\binom{n_{1.}}{n_{11}}\binom{n-n_{1.}}{n_{.1}-n_{11}}
\end{aligned}$$

4.6 Contingency Table- Relationship between Poisson and Multinomial distribution

Consider a $I \times J$ contingency table of cell counts, where each cell count is denoted by $n_{ij}, i = 1, ..I, j = 1, ..J$, and thus n_{ij} denotes the cell count of ith row and jth column, and $n_{ij} \sim Poisson(\mu_{ij})$ and independent. Further, let $n = \sum_{j=1}^{J}\sum_{i=1}^{I} n_{ij}$ denote the grand total.

(a) Derive the joint distribution of $(n_{11}, n_{12}, ...n_{ij})$ conditional on grand total n. By poisson distribution of each cell counts

$$n = \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij} \sim \frac{exp(-\mu)\mu^n}{n!}, \qquad \mu = \sum_{i=1}^{I}\sum_{j=1}^{J}\mu_{ij}$$

$$\begin{aligned}
p(n_{11}, ..n_{ij}|n) &= \frac{\prod_{i=1}^{I}\prod_{j=1}^{J}\frac{exp(-\mu_{ij})\mu_{ij}{}^{n_{ij}}}{n_{ij}!}}{\frac{exp(-\mu)\mu^n}{n!}} \\
&= \binom{n}{n_{11}n_{12}...n_{ij}}\frac{\prod_{i=1}^{I}\prod_{j=1}^{J}\mu_{ij}{}^{n_{ij}}}{\mu^n} \\
&= \binom{n}{n_{11}n_{12}...n_{ij}}\prod_{i=1}^{I}\prod_{j=1}^{J}\left(\frac{\mu_{ij}}{\mu}\right)^{n_{ij}}
\end{aligned}$$

The joint distribution is Multinomial $(n; \pi_{11}, \pi_{12}, ..\pi_{IJ})$, where $\pi_{ij} = \frac{\mu_{ij}}{\sum_{i=1}^{I}\sum_{j=1}^{J}\mu_{ij}}$

(b) Suppose all of the rows margins are assumed fixed. Derive the joint distribution of $(n_{11}, n_{12}, ...n_{ij})$.

$$n_{i+} = \sum_{j=1}^{J} n_{ij}$$

$$n_{i+} \sim Poisson(\sum_{j=1}^{J} \mu_{ij})$$

$$\begin{aligned} p(n_{11}, ..n_{ij}|n_{i+}) &= \prod_{i=1}^{I}\prod_{j=1}^{J} \frac{exp(-\mu_{ij}){\mu_{ij}}^{n_{ij}}}{n_{ij}!} \Big/ \prod_{i=1}^{I} \frac{exp(-\mu_i)\mu_i^{n_{i+}}}{n_{i+}!} \\ &= \prod_{i=1}^{I} \binom{n_{i+}}{n_{ij}} \prod_{i=1}^{I}\prod_{j=1}^{J} \left(\frac{\mu_{ij}}{\sum_{j=1}^{J}\mu_{ij}}\right)^{n_{ij}} \end{aligned}$$

(c) Suppose all of the columns margins are assumed fixed. Derive the joint distribution of $(n_{11}, n_{12}, ...n_{ij})$.

$$n_{+j} = \sum_{i=1}^{I} n_{ij}$$

$$n_{+j} \sim Poisson(\sum_{i=1}^{I} \mu_{ij})$$

$$\begin{aligned} p(n_{11}, ..n_{ij}|n_{+j}) &= \prod_{i=1}^{I}\prod_{j=1}^{J} \frac{exp(-\mu_{ij}){\mu_{ij}}^{n_{ij}}}{n_{ij}!} \Big/ \prod_{j=1}^{J} \frac{exp(-\mu_i)\mu_i^{n_{+j}}}{n_{+j}!} \\ &= \prod_{j=1}^{J} \binom{n_{+j}}{n_{ij}} \prod_{i=1}^{I}\prod_{j=1}^{J} \left(\frac{\mu_{ij}}{\sum_{i=1}^{I}\mu_{ij}}\right)^{n_{ij}} \end{aligned}$$

(d) Suppose that $I = 2$ and $J = 2$, and both the rows margins and column margins are fixed. Derive the joint distribution of $(n_{11}|n_{1+}, n_{+1}n)$, where $n_{1+} = n_{11} + n_{12}, n_{+1} = n_{11} + n_{21}$.

$$p(n_{11}|n_{1+}, n_{+1}n) = \frac{p(n_{11}, n_{1+}, n_{+1}n)}{p(n_{1+}, n_{+1}n)}$$

$$\begin{aligned} p(n_{ij}) &= \prod_{i=1}^{2}\prod_{j=1}^{2} \frac{exp(-\mu_{ij})\mu_{ij}^{n_{ij}}}{n_{ij}!} \\ &= \frac{exp(-\mu_{11})\mu_{11}^{n_{11}}}{n_{11}!} \frac{exp(-\mu_{12})\mu_{12}^{n_{12}}}{n_{12}!} \frac{exp(-\mu_{21})\mu_{21}^{n_{21}}}{n_{21}!} \frac{exp(-\mu_{22})\mu_{22}^{n_{22}}}{n_{22}!} \\ n_{12} &= n_{1+} - n_{11}, \qquad n_{21} = n_{+1} - n_{11}, \\ n_{22} &= n - n_{12} - n_{21} - n_{11} = n - n_{1+} - n_{+1} + n_{11} \end{aligned}$$

$$p(n_{11}, n_{1+}, n_{+1}n) = \frac{exp(-\mu_{11})\mu_{11}^{n_{11}}}{n_{11}!} \frac{exp(-\mu_{12})\mu_{12}^{n_{1+}-n_{11}}}{(n_{1+} - n_{11})!} \frac{exp(-\mu_{21})\mu_{21}^{n_{+1}-n_{11}}}{(n_{+1} - n_{11})!} \frac{exp(-\mu_{22})\mu_{22}^{n-n_{1+}-n_{+1}+n_{11}}}{(n - n_{1+} - n_{+1} + n_{11})!}$$

The Jacobian transformation matrix

$$J = \begin{pmatrix} \frac{\partial n_{11}}{\partial n_{11}} & \frac{\partial n_{11}}{\partial n_{1+}} & \frac{\partial n_{11}}{\partial n_{+1}} & \frac{\partial n_{11}}{\partial n} \\ \frac{\partial n_{12}}{\partial n_{11}} & \frac{\partial n_{12}}{\partial n_{1+}} & \frac{\partial n_{21}}{\partial n_{+1}} & \frac{\partial n_{22}}{\partial n} \\ \frac{\partial n_{21}}{\partial n_{11}} & \frac{\partial n_{21}}{\partial n_{1+}} & \frac{\partial n_{21}}{\partial n_{+1}} & \frac{\partial n_{22}}{\partial n} \\ \frac{\partial n_{22}}{\partial n_{11}} & \frac{\partial n_{22}}{\partial n_{1+}} & \frac{\partial n_{22}}{\partial n_{+1}} & \frac{\partial n_{22}}{\partial n} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 1 & -1 & -1 & 1 \end{pmatrix}$$

$$\|J\| = 1$$

Then we can get the $p(n_{1+}, n_{+1}, n)$ by summing over n_{11}. We have $n_{11} <= n_{1+}, n_{11} <= n_{+1}$, and $n_{11} >= -n + n_{1+} + n_{+1}$.

$$\begin{aligned} p(n_{11}, n_{1+}, n_{+1}n) &= \frac{exp(-\mu_{11})\mu_{11}^{n_{11}}}{n_{11}!} \frac{exp(-\mu_{12})\mu_{12}^{n_{1+}-n_{11}}}{(n_{1+}-n_{11})!} \frac{exp(-\mu_{21})\mu_{21}^{n_{+1}-n_{11}}}{(n_{+1}-n_{11})!} \frac{exp(-\mu_{22})\mu_{22}^{n-n_{1+}-n_{+1}+n_{11}}}{(n-n_{1+}-n_{+1}+n_{11})!} \\ &= \frac{exp(-\sum_{i=1}^{2}\sum_{j=1}^{2}\mu_{ij}) \left(\frac{\mu_{11}\mu_{22}}{\mu_{12}\mu_{21}}\right)^{n_{11}} \left(\frac{\mu_{12}}{\mu_{22}}\right)^{n_{1+}} \left(\frac{\mu_{21}}{\mu_{22}}\right)^{n_{+1}} \mu_{22}^{n}}{n_{11}!(n_{1+}-n_{11})!(n_{+1}-n_{11})!(n-n_{1+}-n_{+1}+n_{11})!} \\ p(n_{1+}, n_{+1}n) &= \sum_{\max(0,-n+n_{1+}+n_{+1})}^{\min(n_{1+},n_{+1})} \frac{exp(-\sum_{i=1}^{2}\sum_{j=1}^{2}\mu_{ij}) \left(\frac{\mu_{11}\mu_{22}}{\mu_{12}\mu_{21}}\right)^{n_{11}} \left(\frac{\mu_{12}}{\mu_{22}}\right)^{n_{1+}} \left(\frac{\mu_{21}}{\mu_{22}}\right)^{n_{+1}} \mu_{22}^{n}}{n_{11}!(n_{1+}-n_{11})!(n_{+1}-n_{11})!(n-n_{1+}-n_{+1}+n_{11})!} \end{aligned}$$

So we can have

$$\begin{aligned} p(n_{11}|n_{1+}, n_{+1}n) &= \frac{p(n_{11}, n_{1+}, n_{+1}n)}{p(n_{1+}, n_{+1}n)} \\ &= \frac{exp(-\sum_{i=1}^{2}\sum_{j=1}^{2}\mu_{ij}) \left(\frac{\mu_{11}\mu_{22}}{\mu_{12}\mu_{21}}\right)^{n_{11}} \left(\frac{\mu_{12}}{\mu_{22}}\right)^{n_{1+}} \left(\frac{\mu_{21}}{\mu_{22}}\right)^{n_{+1}} \mu_{22}^{n}}{n_{11}!(n_{1+}-n_{11})!(n_{+1}-n_{11})!(n-n_{1+}-n_{+1}+n_{11})!} \\ &\Bigg/ \sum_{\max(0,-n+n_{1+}+n_{+1})}^{\min(n_{1+},n_{+1})} \frac{exp(-\sum_{i=1}^{2}\sum_{j=1}^{2}\mu_{ij}) \left(\frac{\mu_{11}\mu_{22}}{\mu_{12}\mu_{21}}\right)^{n_{11}} \left(\frac{\mu_{12}}{\mu_{22}}\right)^{n_{1+}} \left(\frac{\mu_{21}}{\mu_{22}}\right)^{n_{+1}} \mu_{22}^{n}}{n_{11}!(n_{1+}-n_{11})!(n_{+1}-n_{11})!(n-n_{1+}-n_{+1}+n_{11})!} \end{aligned}$$

Which we can rewrite

$$\begin{aligned} p(n_{11}|n_{1+}, n_{+1}n) = &\binom{n_{1+}}{n_{11}}\binom{n-n_{1+}}{n_{+1}-n_{11}}\left(\frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}}\right)^{n_{11}} \\ &\Bigg/ \sum_{x\in\max(0,-n+n_{1+}+n_{+1})}^{\min(n_{1+},n_{+1})} \binom{n_{1+}}{x}\binom{n-n_{1+}}{n_{+1}-x}\left(\frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}}\right)^{x} \end{aligned}$$

(e) Let π_{ij} denote the cell probability and assume n is fixed. Consider testing $H_0 : \pi_{ij} = \pi_{i+}\pi_{+j}, i = 1, ..I, j = 1, ..J$. Derive the MLE of π_{ij} under H_0.

The H_0 could be written as

$$H_0 : \pi_{ij} = \pi_{i+}\pi_{+j}$$

The multinomial distribution of π_{ij}

$$p(\pi_{ij}) = \binom{n}{n_{11}n_{12}n_{21}n_{22}}\pi_{ij}^{n_{ij}}, \sum_{i=1}^{I}\sum_{j=1}^{J}\pi_{ij} = 1$$

The log-likelihood function

$$logp(\pi_{ij}) = log\binom{n}{n_{11}n_{12}n_{21}n_{22}} + n_{ij}log\pi_{ij}, \sum_{i=1}^{I}\sum_{j=1}^{J}\pi_{ij} = 1$$

Under H_0, the log-likelihood

$$logp(\pi_{ij}) = log\binom{n}{n_{11}n_{12}n_{21}n_{22}} + n_{ij}log\pi_{i+}\pi_{+j}, \sum_{i=1}^{I}\pi_{i+} = 1, \sum_{j=1}^{J}\pi_{+j} = 1$$

By Lagrangian multiplier theorem,

$$\begin{aligned} ln(\pi_{ij}) &= nlog\binom{n}{n_{11}n_{12}n_{21}n_{22}} + \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}log\pi_{i+}\pi_{+j} + \lambda(\sum_{i=1}^{I}\sum_{j=1}^{J}\pi_{ij} - 1), \\ &= nlog\binom{n}{n_{11}n_{12}n_{21}n_{22}} + \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij}log\pi_{i+} + \sum_{j=1}^{J}\sum_{i=1}^{I} n_{ij}log\pi_{+j} - \lambda(\sum_{i=1}^{I}\pi_{i+} - 1) \end{aligned}$$

Take first derivative of log-likelihood

$$\begin{aligned} \frac{\partial ln}{\partial \pi_{i+}} &= \frac{\sum_{j=1}^{J} n_{ij}}{\pi_{i+}} + \lambda = 0 \\ \hat{\pi}_{i+} &= \frac{\sum_{j=1}^{J} n_{ij}}{\lambda} \\ \sum_{i=1}^{I}\pi_{i+} = 1, &\qquad \lambda = \sum_{j=1}^{J}\sum_{i=1}^{I} n_{ij} \\ \hat{\pi}_{i+} &= \frac{n_{i+}}{n} \end{aligned}$$

Similarly, we have $\hat{\pi}_{+j} = \frac{n_{+j}}{n}$, the MLE of π_{ij} under H_0 is

$$\hat{\pi}_{ij} = \hat{\pi}_{i+}\hat{\pi}_{+j} = \frac{n_{i+}n_{+j}}{n^2}$$

(f) Derive the likelihood ratio test for the hypothesis in part (e) and derive its asymptotic distribution under H_0. From part (e), we have the parameter estimates under H_0. While under alternative hypothesis, we have $\mu_{ij} = n_{ij}$.

$$
\begin{aligned}
LRT_n = 2(LR(\pi_{H_1}) - LR(\pi_{H_0})) &= 2\left(\sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij} log \pi_{ij} - \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij} log \pi_{i+}\pi_{+j}\right) \\
&= 2\left(\sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij} log \frac{\pi_{ij}}{\pi_{i+}\pi_{+j}}\right) \\
&= 2\left(\sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij} log \frac{n_{ij} n}{n_{i+}n_{+j}}\right) \sim \chi^2_{(I-1)(J-1)}
\end{aligned}
$$

Note that the full model has $(IJ - 1)$ parameters, and the null hypothesis has $(I-1) + (J-1)$ parameters.

$$
\begin{aligned}
df &= I \times J - 1 - (I-1) - (J-1) \\
&= (I-1)(J-1)
\end{aligned}
$$

(g) Suppose that π_{11}, π_{12} are parameters of interest and the rest of the parameters are treated as nuisance. Derive the conditional likelihood of (π_{11}, π_{12}) and the conditional MLE's of (π_{11}, π_{12}). If not specified, we treat as general contingency table that total n is fixed. If only π_{11}, π_{12} are parameters of interest and the rest of the parameters are treated as nuisance, then we will set the rest of the parameters as one parameter, and get its distribution, which is to find the sufficient statistics for rest of the parameters. Write the Multinomial distribution in exponential family distribution.
We can find marginal distribution by summing over along all possible values of (n_{11}, n_{12}). Note that $n_{11} \leq \min n_{1+} - n_{12}, n_{+1}$ for a given value of n_{12}. Similarly, $n_{12} \leq \min n_{1+} - n_{11}, n_{+1}$ for a given value of n_{11}.
Additionally,

$$
\begin{aligned}
n &\geq n_{1+} + n_{+1} + n_{+2} - n_{11} - n_{12} \\
n_{11} + n_{12} &\geq \max 0, n_{+1} + n_{1+} + n_{+2}
\end{aligned}
$$

Let

$$
\begin{aligned}
S(n_{11}, n_{12}) = \{(n_{11}, n_{12}) : n_{11} + n_{12} \geq \max 0, n_{+1} + n_{1+} + n_{+2}, \\
n_{11} \leq \min{(n_{1+} - n_{12}, n_{+1})}, n_{12} \leq \min{(n_{1+} - n_{11}, n_{+1})}\}
\end{aligned}
$$

The conditional distribution

$$
\begin{aligned}
p(n_{11}, n_{12} | n_{13}, ...n_{IJ}, n) &= \frac{p(n_{ij}}{p(S_n)} \\
&= \frac{\frac{1}{n_{11}! n_{12}!}\pi_{11}^{n_{11}}\pi_{12}^{n_{12}}}{\sum_{(x,y \in S_n)} \frac{1}{x! y!}\pi_{11}^{x}\pi_{12}^{y}}
\end{aligned}
$$

And $\hat{\pi}_{11}, \hat{\pi}_{12}$ are the CMLE that maximize $p(n_{11}, n_{12}|n_{13}, ...n_{IJ}, n)$.

5 Practice

5.1 Contingency table parameters

(a) Get MLE of π and prove CLT.
The multinomial distribution based on total n.

$$p(\theta) = n! \prod_{i=0}^{1} \prod_{j=0}^{1} \frac{\pi_{ij}^{n_{ij}}}{n_{ij}!}, \qquad \theta = (\pi_{00}, \pi_{01}, \pi_{10}, \pi_{11})^T$$

$$\begin{aligned} lnp(\theta) &= logn! + \sum_{i=0}^{1} \sum_{j=0}^{1} n_{ij} log(\pi_{ij}) - logn_{ij}! \\ &= logn! + n_{00} log\pi_{00} + n_{01} log\pi_{01} + n_{10} log\pi_{10} + n_{11} log(1 - \pi_{00} - \pi_{01} - \pi_{10}) \end{aligned}$$

The MLE of the θ by taking derivative to the log-likelihood

$$\begin{aligned} \frac{\partial ln(\theta)}{\partial \pi_{00}} &= \frac{n_{00}}{\pi_{00}} - \frac{n_{11}}{1 - \pi_{00} - \pi_{01} - \pi_{10}} = 0 \\ \frac{\partial ln(\theta)}{\partial \pi_{01}} &= \frac{n_{01}}{\pi_{01}} - \frac{n_{11}}{1 - \pi_{00} - \pi_{01} - \pi_{10}} = 0 \\ \frac{\partial ln(\theta)}{\partial \pi_{10}} &= \frac{n_{10}}{\pi_{10}} - \frac{n_{11}}{1 - \pi_{00} - \pi_{01} - \pi_{10}} = 0 \\ \hat{\pi_{00}} &= \frac{n_{00}}{n} \\ \hat{\pi_{01}} &= \frac{n_{01}}{n} \\ \hat{\pi_{10}} &= \frac{n_{10}}{n} \\ \hat{\pi_{11}} &= \frac{n_{11}}{n}, \qquad n = n_{00} + n_{01} + n_{10} + n_{11} \end{aligned}$$

Let $Z_i = I(X = x, Y = y) \sim$ multi $(1, \pi_{00}, \pi_{01}, \pi_{10}, \pi_{11})$.

$$
\begin{aligned}
Z_1 &= I[(X, Y) = (0, 0)] \\
Z_2 &= I[(X, Y) = (0, 1)] \\
Z_3 &= I[(X, Y) = (1, 0)] \\
Z_4 &= I[(X, Y) = (1, 1)] \\
p(\theta) &= \prod_k \pi_k^{I(Z_k=1)}
\end{aligned}
$$

$$
\begin{aligned}
M_Z(t) &= E[exp(t^T Z)] = E[exp(t^T(Z_1 + Z_2 + ...Z_n))] = E[exp(t^T Z_1 + t^T Z_2 + ...t^T Z_n)] \\
&= E[\prod_{i=1}^n exp(t^T Z_i)] \\
&= \prod_{i=1}^n E[exp(t^T Z_i)] \qquad \text{(by independence)} \\
&= \prod_{i=1}^n M_{Z_i}(t) = \prod_{i=1}^n P(Z_i = 1)e^{tz_i} \qquad \text{by MGF of discrete variable } Z_i \\
&= \left(\sum_{j=1}^J \pi_j exp(t_j) \right)^n \qquad \text{by MGF of multinoulli}
\end{aligned}
$$

Then the covariance matrix of θ could be calculated by MGF.

$$
\begin{aligned}
E(Z_1 Z_2) &= \frac{\partial^2 M_Z(t)}{\partial Z_i \partial Z_j}|_{t_i=t_j=0} \\
&= \frac{\partial \left(n(\pi_i e^{t_i})(\sum_{k=1}^K \pi_k e^{t_k})^{n-1} \right)'}{\partial t_j} \\
&= n(n-1)(\sum_{k=1}^K \pi_k e^{t_k})^{n-2} \pi_i \pi_j |_{t_i=t_j=0} = n(n-1)\pi_i \pi_j \\
E(X_i) &= n\pi_i \\
Cov(Z_i, Z_j) &= E(Z_i Z_2) - E(Z_1)E(Z_j) = n(n-1)\pi_i\pi_j - n^2 \pi_i \pi_j = -n\pi_i\pi_j \\
Var(Z_i) &= E(Z_i^2) - E(Z_i)^2 \\
E(Z_i^2) &= \frac{\partial \left(n(\pi_i e^{t_i})(\sum_{k=1}^K \pi_k e^{t_k})^{n-1} \right)'}{\partial t_i} \\
&= n(\sum_{k=1}^K \pi_k e^{t_k})^{n-1} \pi_i e^{t_i} + n(n-1)(\sum_{k=1}^K \pi_k e^{t_k})^{n-2} \pi_i \pi_i e^{2t_i}|_{t_i=0} \\
&= n\pi_i + n(n-1)\pi_i^2 = n\pi_i(1-\pi) \\
Var(Z_i/n) &= \frac{1}{n^2} Var(Z_i) = \frac{1}{n}\pi_i(1-\pi_i)
\end{aligned}
$$

Thus the covariance matrix is

$$\Sigma = \begin{bmatrix} \pi_{00}(1-\pi_{00}) & -\pi_{00}\pi_{01} & -\pi_{00}\pi_{10} & -\pi_{00}\pi_{11} \\ -\pi_{01}\pi_{00} & \pi_{01}(1-\pi_{01}) & -\pi_{01}\pi_{10} & -\pi_{01}\pi_{11} \\ -\pi_{10}\pi_{00} & -\pi_{10}\pi_{01} & \pi_{10}(1-\pi_{10}) & -\pi_{10}\pi_{11} \\ -\pi_{11}\pi_{00} & -\pi_{11}\pi_{01} & -\pi_{11}\pi_{10} & \pi_{11}(1-\pi_{11}) \end{bmatrix} = diag(\pi_{ij}) - \theta\theta^T$$

By Central limit theroem,

$$\sqrt{n}(\hat{\pi_{00}} - \pi_{00}, \hat{\pi_{01}} - \pi_{01}, \hat{\pi_{10}} - \pi_{10}, \hat{\pi_{11}} - \pi_{11})^T \xrightarrow{d} N(0, \Sigma)$$

(b) Let R denote the odds ratio. Find the maximum likelihood estimate of log(R) and derive its asymptotic distribution.
By invariance of MLE:

$$\begin{aligned} R &= \frac{\pi_{00}\pi_{11}}{\pi_{01}\pi_{10}} \\ g(R) = logR &= log\pi_{00} + log\pi_{11} - log\pi_{01} - log\pi_{10} \\ log\hat{R} &= log\hat{\pi_{00}} + log\hat{\pi_{11}} - log\hat{\pi_{01}} - log\hat{\pi_{10}} \\ &= log\frac{n_{00}n_{11}}{n_{01}n_{10}} \end{aligned}$$

By Central limit theorem, we have

$$\sqrt{n}\left(g(\hat{R}) - g(R)\right) \xrightarrow{d} N\left(0, \frac{\partial g(R)}{\partial\theta}\Sigma\frac{\partial g(R)}{\partial\theta}^T\right)$$

By delta method,

$$\begin{aligned} \frac{\partial g(R)}{\partial\theta} &= \left(\frac{1}{R}\frac{\partial R}{\partial\pi_{00}}, \frac{1}{R}\frac{\partial R}{\partial\pi_{01}}, \frac{1}{R}\frac{\partial R}{\partial\pi_{10}}, \frac{1}{R}\frac{\partial R}{\partial\pi_{11}}\right) \\ &= \left(\frac{1}{\pi_{00}}, -\frac{1}{\pi_{01}}, -\frac{1}{\pi_{10}}, \frac{1}{\pi_{11}}\right) \\ \Sigma^R &= \frac{\partial g(R)}{\partial\theta}\Sigma\frac{\partial g(R)}{\partial\theta}' \\ &= \left(\frac{1}{\pi_{00}}, -\frac{1}{\pi_{01}}, -\frac{1}{\pi_{10}}, \frac{1}{\pi_{11}}\right)\begin{bmatrix} \pi_{00}(1-\pi_{00}) & -\pi_{00}\pi_{01} & -\pi_{00}\pi_{10} & -\pi_{00}\pi_{11} \\ -\pi_{01}\pi_{00} & \pi_{01}(1-\pi_{01}) & -\pi_{01}\pi_{10} & -\pi_{01}\pi_{11} \\ -\pi_{10}\pi_{00} & -\pi_{10}\pi_{01} & \pi_{10}(1-\pi_{10}) & -\pi_{10}\pi_{11} \\ -\pi_{11}\pi_{00} & -\pi_{11}\pi_{01} & -\pi_{11}\pi_{10} & \pi_{11}(1-\pi_{11}) \end{bmatrix}\begin{bmatrix} \frac{1}{\pi_{00}} \\ -\frac{1}{\pi_{01}} \\ -\frac{1}{\pi_{10}} \\ \frac{1}{\pi_{11}} \end{bmatrix} \\ &= (\frac{1}{\pi_{00}} + \frac{1}{\pi_{01}} + \frac{1}{\pi_{10}} + \frac{1}{\pi_{11}}) \end{aligned}$$

We have the asymptotic distribution of $log(R)$

$$\sqrt{n}(log\hat{R} - logR) \xrightarrow{d} N\left(0, (\frac{1}{\pi_{11}} + \frac{1}{\pi_{12}} + \frac{1}{\pi_{21}} + \frac{1}{\pi_{22}})\right)$$

(c) Construct an approximate 95% confidence interval for the odds ratio R.
From part (b), we have the asymptotic normal distribution of $logR$. We have the asymptotic distribution of R.

$$f = exp(g) = R, \qquad f(g)' = R$$

$$\sqrt{n}(f(\hat{g}) - f(g)) \xrightarrow{d} N\left(0, f(g)'(\frac{1}{\pi_{11}} + \frac{1}{\pi_{12}} + \frac{1}{\pi_{21}} + \frac{1}{\pi_{22}})f(g)'^{T}\right)$$

$$\sqrt{n}(\hat{R} - R) \xrightarrow{d} N\left(0, R^2(\frac{1}{\pi_{11}} + \frac{1}{\pi_{12}} + \frac{1}{\pi_{21}} + \frac{1}{\pi_{22}})\right)$$

$$(\hat{R} - R) \xrightarrow{d} N\left(0, \frac{1}{n}R^2(\frac{1}{\pi_{11}} + \frac{1}{\pi_{12}} + \frac{1}{\pi_{21}} + \frac{1}{\pi_{22}})\right)$$

The 95% confidence interval for the odds ratio R

$$\{R : \hat{R} - 1.96\hat{R}\sqrt{\frac{1}{\pi_{11}} + \frac{1}{\pi_{12}} + \frac{1}{\pi_{21}} + \frac{1}{\pi_{22}}} \le R \le \hat{R} + 1.96\hat{R}\sqrt{\frac{1}{\pi_{11}} + \frac{1}{\pi_{12}} + \frac{1}{\pi_{21}} + \frac{1}{\pi_{22}}}\}$$

(d) Under the assumptions of part (a), further assume that$\pi_{1+} = \pi_{11} + \pi_{10} = \frac{exp(\alpha)}{1+\exp(\alpha)}$ and $\pi_{+1} = \pi_{11} + \pi_{01} = \frac{exp(\alpha+\beta)}{1+\exp(\alpha+\beta)}$. Derive the maximum likelihood estimates of (α, β), denoted by $(\hat{\alpha}; \hat{\beta})$.

$$\pi_{01} + \pi_{11} = \frac{exp(\alpha)}{1 + \exp(\alpha)}$$

$$exp(\alpha) = \frac{\pi_{10} + \pi_{11}}{\pi_{01} + \pi_{00}}, \qquad \alpha = log\left(\frac{\pi_{10} + \pi_{11}}{\pi_{01} + \pi_{00}}\right)$$

$$\pi_{10} + \pi_{11} = \frac{exp(\alpha + \beta)}{1 + \exp(\alpha + \beta)}$$

$$\alpha + \beta = log\left(\frac{\pi_{01} + \pi_{11}}{\pi_{10} + \pi_{00}}\right)$$

$$\beta = log\left(\frac{\pi_{01} + \pi_{11}}{\pi_{10} + \pi_{00}}\right) - log\frac{\pi_{10} + \pi_{11}}{\pi_{01} + \pi_{00}}, \qquad \beta \ = log\left(\frac{(\pi_{01} + \pi_{11})(\pi_{01} + \pi_{00})}{(\pi_{10} + \pi_{00})(\pi_{10} + \pi_{11})}\right)$$

By invariance of MLE,

$$\hat{\alpha} = log\left(\frac{\hat{\pi_{10}} + \hat{\pi_{11}}}{\hat{\pi_{01}} + \hat{\pi_{00}}}\right) = log\left(\frac{n_{10} + n_{11}}{n_{01} + n_{00}}\right)$$

$$\hat{\beta} = log\left(\frac{(\hat{\pi}_{01} + \hat{\pi}_{11})(\hat{\pi}_{01} + \hat{\pi}_{00})}{(\hat{\pi}_{10} + \hat{\pi}_{00})(\hat{\pi}_{10} + \hat{\pi}_{11})}\right) = log\left(\frac{(n_{01} + n_{11})(n_{01} + n_{00})}{(n_{10} + n_{00})(n_{10} + n_{11})}\right)$$

(e) Using the assumptions of part (d), derive the asymptotic distribution of (α, β) (properly normalized).

By Central limit theorem and delta method,

$$\begin{aligned}
\xi &= (\alpha, \beta)^T \\
g(\xi) &= \{log\left(\frac{\pi_{10}+\pi_{11}}{\pi_{01}+\pi_{00}}\right), log\left(\frac{(\pi_{01}+\pi_{11})(\pi_{01}+\pi_{00})}{(\pi_{10}+\pi_{00})(\pi_{10}+\pi_{11})}\right)\}^T \\
\sqrt{n}(g(\hat{\xi}) - g(\xi)) &\xrightarrow{d} N\left(0, \Sigma^N\right) \\
\Sigma^N &= \frac{\partial g(\xi)}{\partial \pi}\Sigma\frac{\partial g(\xi)}{\partial \pi}^T
\end{aligned}$$

Σ^N is calculated by delta method,

$$\begin{aligned}
\frac{\partial g(\alpha)}{\partial \pi_{00}} &= -\frac{1}{(\pi_{01}+\pi_{00})} = -\frac{1}{\pi_{0+}} \\
\frac{\partial g(\alpha)}{\partial \pi_{01}} &= -\frac{1}{(\pi_{01}+\pi_{00})} = -\frac{1}{\pi_{0+}} \\
\frac{\partial g(\alpha)}{\partial \pi_{10}} &= \frac{1}{(\pi_{10}+\pi_{11})} = \frac{1}{\pi_{1+}} \\
\frac{\partial g(\alpha)}{\partial \pi_{11}} &= \frac{1}{(\pi_{10}+\pi_{11})} = \frac{1}{\pi_{1+}} \\
\frac{\partial g(\beta)}{\partial \pi_{00}} &= \frac{(\pi_{10}-\pi_{01})}{(\pi_{01}+\pi_{00})(\pi_{00}+\pi_{10})} = -\frac{1}{(\pi_{10}+\pi_{00})} + \frac{1}{(\pi_{01}+\pi_{00})} = -\frac{1}{\pi_{+0}} + \frac{1}{\pi_{0+}} \\
\frac{\partial g(\beta)}{\partial \pi_{01}} &= \frac{1}{(\pi_{01}+\pi_{11})} + \frac{1}{(\pi_{01}+\pi_{00})} \\
\frac{\partial g(\beta)}{\partial \pi_{10}} &= -\frac{1}{(\pi_{10}+\pi_{00})} - \frac{1}{(\pi_{10}+\pi_{11})} \\
\frac{\partial g(\beta)}{\partial \pi_{11}} &= \frac{(\pi_{10}-\pi_{01})}{(\pi_{10}+\pi_{11})(\pi_{01}+\pi_{11})} = -\frac{1}{(\pi_{10}+\pi_{11})} + \frac{1}{(\pi_{01}+\pi_{11})} \\
\frac{\partial g(\xi)}{\partial \pi} &= \begin{bmatrix} -\frac{1}{\pi_{0+}} & -\frac{1}{\pi_{0+}} & \frac{1}{\pi_{1+}} & \frac{1}{\pi_{1+}} \\ \frac{1}{\pi_{0+}} - \frac{1}{\pi_{+0}} & \frac{1}{\pi_{0+}} + \frac{1}{\pi_{+1}} & -\frac{1}{\pi_{+0}} - \frac{1}{\pi_{1+}} & \frac{1}{\pi_{+1}} - \frac{1}{\pi_{1+}} \end{bmatrix} \\
\Sigma^N &= \frac{\partial g(\xi)}{\partial \pi}\Sigma\frac{\partial g(\xi)}{\partial \pi}^T \\
&= \left(\frac{1}{\pi_{11}} + \frac{1}{\pi_{12}} + \frac{1}{\pi_{21}} + \frac{1}{\pi_{22}}\right)
\end{aligned}$$

(f) Under the model of part (d), show that $(\pi_{1+}\pi_{0+})^{-1} + (\pi_{+1}\pi_{+0})^{-1} \le (\pi_{1+}\pi_{+0})^{-1} + (\pi_{+1}\pi_{0+})^{-1}$.

$$
\begin{aligned}
&(\pi_{1+}\pi_{+0})^{-1} + (\pi_{+1}\pi_{0+})^{-1} - (\pi_{1+}\pi_{0+})^{-1} - (\pi_{+1}\pi_{+0})^{-1} \\
&= \frac{\pi_{0+} - \pi_{+0}}{\pi_{1+}\pi_{+0}\pi_{0+}} + \frac{\pi_{+0} - \pi_{0+}}{\pi_{+1}\pi_{0+}\pi_{+0}} \\
&= \frac{(\pi_{0+} - \pi_{+0})(\pi_{+1} - \pi_{1+})}{\pi_{1+}\pi_{+0}\pi_{0+}\pi_{+1}} \\
&= \frac{(\pi_{01} - \pi_{10})^2}{\pi_{1+}\pi_{+0}\pi_{0+}\pi_{+1}} \geq 0
\end{aligned}
$$

From above, we have $(\pi_{1+}\pi_{0+})^{-1} + (\pi_{+1}\pi_{+0})^{-1} \leq (\pi_{1+}\pi_{+0})^{-1} + (\pi_{+1}\pi_{0+})^{-1}$.