UNC BIOS 762

Summary

Creating my own project

Author:
Mingwei Fei

Special colaborator:
Bonnie

Jan, 2023

Contents

Chapter 1

Matrix

1.1 Orthogonal Projection Matrix

(i) $\Lambda^T = P^T X$, generally the matrix that is right product will keep the left matrix column space. So $\rho' X, P' X$ is the form.

(ii) The F statistics is actually a quadratic form, that we need to use $(MP)^T(MP)$, here M is the orthogonal projection matrix of $\Lambda^{'}\beta$.

$$\begin{aligned}
M_{MP} &= (MP)^T[(MP)^T(MP)]^{-1}(MP) \\
Y' M_{MP} Y &= Y'(MP)^T[(MP)^T(MP)]^{-1}(MP)Y \\
&= Y'P'M[P'MP]^{-1}MPY \\
&= (MPY)^T[P'MP]^{-1}MPY \\
&= (\Lambda^T\beta)^T[P'X(X'X)^{-}X'P]^{-1}(\Lambda^T\beta) \\
&= \beta^T\Lambda[\Lambda'(X'X)^{-}\Lambda]^{-1}\Lambda^T\beta
\end{aligned}$$

1.1.1 Spectral Decomposition

Why do we need to do spectral decomposition? It is mainly used in quadratic form of non-centrality chi-square distribution.

Definition 1.1.1. The spectral decomposition allows the representation of any symmetric matrix in terms of an orthogonal matrix and a diagonal matrix of eigenvalues.

Example:

1.2 Generalized Least Squares

Consider the linear model $Y = X\beta + Z\gamma + \epsilon$, where $E(\epsilon) = 0$ and $Cov(\epsilon) = V$, V is assumed known and positive definite, and (β, γ) are unknown. Further, let $A =$

$X(X'V^-X)^-X'V^{-1}$, X is $n \times p$, Z is $n \times q$, and both X and Z may be less than full rank. Let C(H) denote the usual label for the column space of an arbitrary matrix H.

If we want to do hypothesis testing, we will need to construct F test by breaking down the quadratic forms into chi-square terms (independent/orthogonal), so the variance term needs to be identical and independent.

(a) Show that $(I-A)'V^{-1}(I-A) = (I-A)'V^{-1} = V^{-1}(I-A)$.
Lets review the o.p.o onto C(X), and a projection $A = X(X'V^-X)^-X'V^{-1}$.

$$V = QQ^T$$
$$P = Q^{-1}X[(Q^{-1}X)'(Q^{-1}X)]^-(Q^{-1}X)^T = Q^{-1}X[X'V^{-1}X]^-X^TQ^{-1}$$

We need to transform P to A, so by definition, $AX = X$. we would like to prove $PQ^{-1}X = Q^{-1}X$

$$PQ^{-1}X = Q^{-1}X[X'V^{-1}X]^-X^TQ^{-1}Q^{-1}X = Q^{-1}X, \qquad X[X'V^{-1}X]^-X^TQ^{-1}Q^{-1} = A$$

Then we have $AX = X$, then A is a projection onto C(X). **Need attention that A is not a o.p.o. with respective to $x'y$ (not symmetric) it is only a projection. But is o.p.o to $x'V^{-1}y$.**
For $\rho'AY = \rho'MY$, that $M = X(X'X)^-X^T$, we need to have $C(X) = C(VX), C(X) = C(V^{-1}X)$, then $A = X(X'V^-X)^-X'V^{-1}$.
Show that A, I-A are projections

$$A^2 = A$$
$$X(X'V^-X)^-X'V^{-1}X(X'V^-X)^-X'V^{-1} = X(X'V^-X)^-X'V^{-1}$$
$$(I-A)^2 = I-A$$
$$(I-A)(I-A) = I-A-A+A^2 = I-A$$

Then transform the equation

$$(I-A)'V^{-1}(I-A)(I-A) = (I-A)'V^{-1}(I-A),$$
$$(I-A)'V^{-1}(I-A) = (I-A)'V^{-1}$$

To show $(I-A)'V^{-1}(I-A) = V^{-1}(I-A)$

$$(I-A)'(I-A)' = [(I-A)(I-A)]^T = (I-A)^T$$
$$(I-A)'(I-A)'V^{-1}(I-A) = (I-A)'V^{-1}(I-A)$$
$$(I-A)'V^{-1}(I-A) = V^{-1}(I-A)$$

(b) Show that A is the projection operator onto C(X) along $C(V^{-1}X)^{\perp}$. We need to show $AX = X, x \in C(X)$, and $Aw = 0, w \in C(V^{-1}X)^{\perp}, C(V^{-1}X) = C(X)$ We have shown that $AX = X$ in above (a), then let $w \in C(V^{-1}X)^{\perp}$

$$(V^{-1}X)'w = 0, \qquad X'V^{-1}w = 0$$
$$Aw = w, \qquad X'V^{-1}Aw = 0$$
$$X'V^{-1}A = 0, \qquad A \perp C(V^{-1}X)$$

(c) Let B denote the projection operator onto C(X,Z) along $C(V^{-1}(X,Z))^{\perp}$. Assume that all matrix inverses exist. Show that

$$B = A + (I - A)Z[Z'(I - A)'V^{-1}(I - A)Z]^{-1}Z'(I - A)'V^{-1}$$

To show B is a projection, $B^2 = B$, but if need to show projection onto C(X,Z), then show $BX = X, x \in C(X,Z)$. If need to show projection along, then show$Bw = 0, w \in C(V^{-1}(X,Z))^{\perp}$.

$$\begin{aligned} B^2 &= \left(A + (I - A)Z[Z'(I - A)'V^{-1}(I - A)Z]^{-1}Z'(I - A)'V^{-1}\right)\left(A + (I - A)Z[Z'(I - A)'V^{-1}(I - \right. \\ &= \left(A + (I - A)Z[Z'(I - A)'V^{-1}(I - A)Z]^{-1}Z'(I - A)'V^{-1}\right) = B \end{aligned}$$

$$A(I - A) = 0$$

Show $B(X,Z) = (BX, BZ) = (X,Z)$.

$$\begin{aligned} BX &= \left(A + (I - A)Z[Z'(I - A)'V^{-1}(I - A)Z]^{-1}Z'(I - A)'V^{-1}\right) X \\ &= AX + (I - A)Z[Z'(I - A)'V^{-1}(I - A)Z]^{-1}Z'(I - A)'V^{-1}(I - A)X = X, \qquad \text{part (a)} \\ BZ &= \left(A + (I - A)Z[Z'(I - A)'V^{-1}(I - A)Z]^{-1}Z'(I - A)'V^{-1}\right) Z \\ &= AZ + (I - A)Z[Z'(I - A)'V^{-1}(I - A)Z]^{-1}Z'(I - A)'V^{-1}(I - A)Z = AZ + (I - A)Z = Z, \quad \text{part} \end{aligned}$$

Next show projection along $C(V^{-1}(X,Z))^{\perp}$.

$$(V^{-1}(X,Z))'w = 0, \qquad X'V^{-1}w = 0, Z'V^{-1}w = 0$$

$$\begin{aligned} Bw &= \left(A + (I - A)Z[Z'(I - A)'V^{-1}(I - A)Z]^{-1}Z'(I - A)'V^{-1}\right) w = 0 \\ &= Aw + (I - A)Z[Z'(I - A)'V^{-1}(I - A)Z]^{-1}Z'(I - A)'V^{-1}w = 0 \end{aligned}$$

$$(X,Z)'V^{-1}B = 0, \qquad B \perp C(V^{-1}(X,Z))$$

(d) Show that $(\hat{\gamma}, \hat{\beta})$ are generalized BLUE's for the linear model, where $(\hat{\gamma}, \hat{\beta})$ satisfy

$$\begin{aligned} \hat{\gamma} &= [Z'(I - A)'V^{-1}(I - A)Z]^{-1}Z'(I - A)'V^{-1}(I - A)Y \\ X\hat{\beta} &= A(Y - Z\hat{\gamma}) \end{aligned}$$

To show projection operator onto C(Z) is $[Z'(I-A)'V^{-1}(I-A)Z]^{-1}Z'(I-A)'V^{-1}$, also it is along $C(V^{-1}Z)^{\perp}$.
By Gauss-Markov theorem, the least square estimates of γ, β are BLUE's.

$$\begin{aligned} Y &= X\beta + Z\gamma \\ AY &= AX\beta + AZ\gamma = X\beta + AZ\gamma \\ (I - A)Y &= (I - A)Z\gamma \end{aligned}$$

We can construct projection operation regarding C(Z).

$$Q^{-1}(I-A)Y = Q^{-1}(I-A)Z\gamma$$
$$P_z = Q^{-1}(I-A)Z[(Q^{-1}(I-A)Z)'(Q^{-1}(I-A)Z)]^{-}(Q^{-1}(I-A)Z)'$$
$$P_z Q^{-1}(I-A)Z = Q^{-1}(I-A)Z$$

$$[Z'(I-A)'V^{-1}(I-A)Z]^{-}Z'(I-A)'V^{-1}Z = Z, \qquad \text{also use part (a)}$$
$$C = [Z'(I-A)'V^{-1}(I-A)Z]^{-}Z'(I-A)'V^{-1}, \qquad CZ$$

We have C a projection operator onto C(Z), also we can prove orthogonality to $C(V^{-1}Z)^{\perp}$. Then we have proved.

(e) Suppose that $\epsilon \sim N_n(0,V)$ and V is known. Further, suppose that γ, β are both estimable. From first principles, derive the likelihood ratio test for the hypothesis $H_0 : \gamma = 0$, where γ, β are both unknown, and state the exact distribution of the test statistic under the null and alternative hypotheses.
To derive the likelihood ratio test, we link with F-test and construct independent chi-square statistics. The nominator is the square difference between two models, and the denominator is MSE.

$$Y = X\beta + Z\gamma + \epsilon, \qquad Y = W\delta + \epsilon,$$
$$W = (X,Z), \delta = (\beta,\gamma)^T, \qquad V = QQ^T, X_{n\times p}, Z_{n\times q},$$

Assume rank(W) = p+q, **Pay attention to define matrix rank**

$$Q^{-1}Y = W = Q^{-1}X\beta + Q^{-1}Z\gamma \sim N(0,I)$$
$$Q^{-1}Y = \tilde{Y}, Q^{-1}Z = \tilde{Z}, \qquad Q^{-1}X = \tilde{X}$$
$$\tilde{Y} = \tilde{W}\delta + \tilde{\epsilon}, \qquad \tilde{\epsilon} \sim N(0,I)$$

The likelihood ratio test

$$LRT = \frac{Sup_{H_0} L(\delta|W)}{Sup_{H_1} L(\delta|W)}$$

The likelihood under H_0

$$\delta_0 = (\beta_0, 0)', \qquad W_0 = (\tilde{X})$$
$$L(\beta_0|W) = \frac{1}{\sqrt{2\pi}} exp\left(\frac{-1}{2}(\tilde{Y} - W_0\delta_0)'(\tilde{Y} - W_0\delta_0)\right)$$
$$= \frac{1}{\sqrt{2\pi}} exp\left(\frac{-1}{2}(\tilde{Y} - \tilde{X}\hat{\beta})'(\tilde{Y} - \tilde{X}\hat{\beta})\right)$$
$$\tilde{X}\hat{\beta} = M_0 W_0, \qquad \hat{\beta} \text{ satisfy}$$
$$M_0 = \tilde{X}[(\tilde{X})'(\tilde{X})]^{-1}\tilde{X}'$$

Thus the numerator is

$$L(\beta_0|W) = \frac{1}{\sqrt{2\pi}} exp\left(\frac{-1}{2}(\tilde{Y}'(I - M_0)\tilde{Y}\right)$$

The likelihood under H_1

$$\begin{aligned}
L(\delta|W) &= \frac{1}{\sqrt{2\pi}} exp\left(\frac{-1}{2}(\tilde{Y} - W\delta)'(\tilde{Y} - W\delta)\right) \\
\tilde{W}\hat{\delta} &= M\tilde{W} \qquad \hat{\delta} \text{ satisfy} \\
M &= \tilde{W}[(\tilde{W})'(\tilde{W})]^{-1}\tilde{W}'
\end{aligned}$$

Thus the denominator is

$$L(\delta|W) = \frac{1}{\sqrt{2\pi}} exp\left(\frac{-1}{2}(\tilde{Y}'(I - M)\tilde{Y}\right)$$

The LRT test

$$-2\Lambda = \frac{\left(\tilde{Y}'(I - M_0)\tilde{Y}\right)}{\left(\tilde{Y}'(I - M)\tilde{Y}\right)} = \left(\tilde{Y}'(M - M_0)\tilde{Y}\right)$$

Thus we reject the H_0, if $\left(\tilde{Y}'(M - M_0)\tilde{Y}\right) > c$. $M - M_0$ is an orthogonal projection as $(M - M_0)(M - M_0) = (M - M_0)$ and $M - M_0$ is symmetric. So $r(M - M_0) = q$ when X and Z are full rank.

$$\begin{aligned}
\left(\tilde{Y}'(M - M_0)\tilde{Y}\right) &\sim (H_0)\chi^2(q) \\
\left(\tilde{Y}'(M - M_0)\tilde{Y}\right) &\sim (H_1)\chi^2(q, \theta) \\
\theta &= \frac{\|(M - M_0)\tilde{W}\delta\|}{2}
\end{aligned}$$

Thus we reject H_0 at level α if $\left(\tilde{Y}'(M - M_0)\tilde{Y}\right) > \chi^2_\alpha(q)$

(f) Suppose that $\epsilon \sim N_n(0, \sigma^2 R)$, where R is known and positive definite, and β, γ, σ^2 are all unknown. Further, assume that β, γ are both estimable. Derive an exact joint 95% confidence region for β, γ, σ^2.
To write the distribution of β, γ, σ^2, such as F distribution or Chi-square (if σ^2 is known) distribution $M_\beta/\sigma^2, M_\gamma/\sigma^2$.

$$\begin{aligned}
Y &= X\beta + Z\gamma + \epsilon, \qquad Y = W\delta + \epsilon, \\
W &= (X, Z), \delta = (\beta, \gamma)^T, \qquad R = QQ^T, X_{n\times p}, Z_{n\times q},
\end{aligned}$$

Assume rank(W) = p+q, **Pay attention to define matrix rank**

$$
\begin{aligned}
Q^{-1}Y = W &= Q^{-1}X\beta + Q^{-1}Z\gamma \sim N(0, \sigma^2 I) \\
Q^{-1}Y = \tilde{Y}, Q^{-1}Z &= \tilde{Z}, \qquad Q^{-1}X = \tilde{X} \\
\tilde{Y} &= \tilde{W}\delta + \tilde{\epsilon}, \qquad \tilde{\epsilon} \sim N(0, \sigma^2 I)
\end{aligned}
$$

We have the F test for β

$$
\begin{aligned}
\lambda' &= (1, 0, 0), \qquad \rho'\tilde{W} = \lambda', \\
F_\beta &= \frac{(\tilde{Y} - \tilde{W}b)' M_{MP} (\tilde{Y} - \tilde{W}b)/r(M_{MP})}{\sigma^2} \\
&= \frac{(\tilde{Y} - \tilde{W}b)'(M\rho)[\rho' M\rho]^-(M\rho)'(\tilde{Y} - \tilde{W}b)/r(M_{MP})}{\sigma^2} \\
&= \frac{(\rho' M\tilde{Y} - \rho' M\tilde{W}b)'[\rho' M\rho]^-(\rho' M\tilde{Y} - \rho' M\tilde{W}b/r(M_{MP})}{\sigma^2} \\
&= \frac{(\lambda'\hat{\delta} - \lambda'\delta)'[\rho' M\rho]^-(\lambda'\hat{\delta} - \lambda'\delta)/r(M_{MP})}{\sigma^2} \\
&= \frac{(\hat{\beta} - \beta)'[\rho' M\rho]^-(\hat{\beta} - \beta)/r(M_{MP})}{\sigma^2}, \qquad r(M_{MP}) = p
\end{aligned}
$$

We have the chi-square distribution for σ^2

$$
\frac{Y'(I - M)Y}{\sigma^2}\chi^2(n - p - q)
$$

The degrees of freedom is $n - p - q$ since both (β, γ) are estimable.
Furthermore, we have the chi-square distribution for δ, because $M \perp (I - M)$

$$
\begin{aligned}
&\frac{\|MY - \tilde{W}\delta\|}{\sigma^2} \sim \chi^2(p + q) \\
&P\{\chi^2_{a/2}(p + q) \leq \frac{\|MY - \tilde{W}\delta\|}{\sigma^2} \leq \chi^2_{1-a/2}(p + q), \quad \chi^2_{b/2}(n - p - q) \leq \frac{\|(I - M)Y\|}{\sigma^2} \leq \chi^2_{1-b/2}(n - p \\
&\quad = 1 - \alpha
\end{aligned}
$$

Such that $(1 - a)(1 - b) = 1 - \alpha$.
The region is given

$$
\begin{aligned}
&\{(\delta, \sigma^2) : \chi^2_{a/2}(p + q) \leq \frac{\|MY - \tilde{W}\delta\|}{\sigma^2} \leq \chi^2_{1-a/2}(p + q), \quad \chi^2_{b/2}(n - p - q) \leq \frac{\|(I - M)Y\|}{\sigma^2} \leq \chi^2_{1-b/2} \\
&\quad = 1 - \alpha
\end{aligned}
$$

1.2.1 Segmented linear regression

Assume one function in a certain range of X and another in a different range. For a general segmented linear regression:

$$
\begin{aligned}
Y_i &= \beta_0 + \beta_1 x_i + \epsilon_i, && x_i \le t \\
Y_i &= \alpha_0 + \alpha_i x_i + \epsilon_i, && x_i > t
\end{aligned}
$$

with independent, identically distributed normal errors, $\epsilon \sim N(0, \sigma^2), i = 1, ..n$.

The model is continuous if $\beta_0 + \beta_1 t = \alpha_0 + \alpha_1 t$, and discontinuous otherwise. Assume that t is known, Let $I(z)$ be the indicator function, and define $(z)_+ = max(0, z)$. Consider fitting the regression model:

$$
Y_i = \delta_0 + \delta_1 x_i + \delta_2 I(x_i - t) + \delta_3 (x_i - t)_+ \epsilon_i, \tag{1.1}
$$

with independent, identically distributed normal errors, $\epsilon \sim N(0, \sigma^2), i = 1, ..n$.

(a) Give the relations between $\beta_0, \beta_1, \alpha_0, \alpha_1$, and $\delta_0, ...\delta_3$.
Compare the model at $X_i > t, X_i \le t$, we can find the relations.

(b) Write the model (3.1) for the form $Y = X\beta + \epsilon$, clearly identify all the components and **derive** the UMVUE of $(\beta_1, \alpha_1)'$ using appropriate projections.

$$
\begin{aligned}
Y &= (Y_1, Y_2, ..Y_n) \\
\beta &= (\delta_0, \delta_1, \delta_2, \delta_3)
\end{aligned}
$$

$$
X = \begin{bmatrix} 1 & x_1 & I(x_1 - t) & (x_1 - t)_+ \\ 1 & x_2 & I(x_2 - t) & (x_2 - t)_+ \\ . & . & . & . \\ 1 & x_n & I(x_n - t) & (x_n - t)_+ \end{bmatrix}
$$

Derive UMVUE, we need to write the likelihood function:

$$
L(\beta, \sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} exp\{\frac{-(y_i - x_i\beta)^2}{2\sigma^2}\}
$$

or we can write in matrix form

$$
\begin{aligned}
L(\beta, \sigma) &= \frac{1}{\sqrt{2\pi}\sigma}^n exp\{\frac{-(Y - X\beta)^T (Y - X\beta)}{2\sigma^2}\} \\
lnL(\beta, \sigma) &= n - ln(\sqrt{2\pi}\sigma) + \frac{-(Y - X\beta)^T (Y - X\beta)}{2\sigma^2} \\
\frac{\partial lnL}{\partial \beta} &= \frac{X^T (Y - X\beta)}{\sigma^2} = 0 \\
X^T Y &= X^T X\beta \\
\frac{\partial lnL}{\partial \sigma} &= \frac{-n}{\sigma} + \frac{(Y - X\beta)^T (Y - X\beta)}{\sigma^3} = 0 \\
\sigma^2 &= \frac{(Y - X\beta)^T (Y - X\beta)}{n}
\end{aligned}
$$

Assume X is full rank, then we have

$$\hat{\beta} = (X^T X)^{-1} X^T Y = (\hat{\delta_0}, \hat{\delta_1}, \hat{\delta_2}, \hat{\delta_3})$$

Since $X^T Y$ is sufficient and complete statistics for β based on exponential family form of the likelihood, so $\hat{\beta}$ is the UMVUE of β. And hence, the UMVUE of α_1, β_1 is

$$\begin{aligned}
\hat{\beta_1} &= \hat{\delta_1}, \\
\hat{\alpha_1} &= \hat{\delta_1} + \hat{\delta_3}
\end{aligned}$$

(c) Use model (3.1) and appropriate projections to derive the F test for the null hypothesis:

(i) H_0 : continuity of the segmented regression: $\beta_0 + \beta_1 t = \alpha_0 + \alpha_1 t$.

$$\begin{aligned}
H_0 : \beta_0 + \beta_1 t &= \alpha_0 + \alpha_1 t, \\
\beta_0 - \alpha_0 &= (\alpha_1 - \beta_1)t \\
\delta_0 - (\delta_0 + \delta_2 - \delta_3 t) &= (\delta_1 + \delta_3 - \delta_1)t \\
\delta_2 &= 0
\end{aligned}$$

So the $H_0 : \delta_2 = 0, \lambda^T \beta = 0, \lambda^T = (0, 0, 1, 0)$
The F-test takes the form:

$$\begin{aligned}
F - test &= \frac{[\lambda^T \hat{\beta} - 0]'(\lambda^T (X^T X)^- \lambda)^{-1} [\lambda^T \hat{\beta} - 0]/r(M_\lambda)}{MSE}, \\
M_\lambda = MP[(MP)'(MP)]^- P'M &= MP[P'MP]^- P'M = MP[P'X(X'X)^- X'P]^{-1} P'M \\
&= MP(\lambda^T (X^T X)^- \lambda)^{-1} P'M \\
&= \frac{[\lambda^T (X^T X)^{-1} X^T Y]^T (\lambda^T (X^T X)^- \lambda)^{-1} [\lambda^T (X^T X)^{-1} X^T Y]}{\sigma^2} \\
&= \frac{[\lambda^T (X^T X)^{-1} X^T Y]^T [\lambda^T (X^T X)^{-1} X^T Y]}{\sigma^2 (\lambda^T (X^T X)^- \lambda)} = \frac{Y^T [X (X^T X)^{-1} \lambda \lambda^T (X^T X)^{-1} X^T] Y}{\sigma^2 (\lambda^T (X^T X)^- \lambda)} \\
\lambda \lambda^T &= \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} (0, 0, 1, 0) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}
\end{aligned}$$

Note that $M_\Lambda = [X(X^T X)^{-1} \lambda \lambda^T (X^T X)^{-1} X^T]$ is orthogonal projection with rank 1, since $\lambda \lambda^T$ is rank 1.

$$\begin{aligned}
F - test &= Y^T M_\Lambda Y, \\
\sigma^2 &= \frac{Y^T (I - M) Y}{n - 4}
\end{aligned}$$

$M = X(X'X)^{-1}X'$ is opo onto $C(X)$, we claim

$$F - test = \frac{Y^T M_\Lambda Y/1}{Y^T(I-M)Y/(n-4)} \sim F(1, n-4, \gamma),$$
$$\gamma = \frac{\mu' M_\Lambda \mu}{2\sigma^2}$$

To show independence in F-test, we need to prove $M_\Lambda(I-M) = 0$.

(ii) H_0 : identity of the two segments: $\beta_0 = \alpha_0, \beta_1 = \alpha_1$.
The null hypothesis could be written as:

$$\begin{aligned} H_0 : \beta_0 = \alpha_0, &\qquad \delta_0 = \delta_0 + \delta_2 - \delta_3 t \\ \beta_1 = \alpha_1, &\qquad \delta_1 = \delta_1 + \delta_3 \\ H_0 : \delta_2 - \delta_3 t = 0, &\qquad \delta_3 = 0 \\ \Lambda^T = &\begin{bmatrix} 0 & 0 & 1 & -t \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{aligned}$$

The orthogonal projection operator for H_0 we can write $\Lambda^T = P^T X$. Let

$$\begin{aligned} M_{MP} = MP[(MP)'(MP)]^{-1}(MP)' &= MP[P'MP]^{-1}P'M \\ F - test &= (\Lambda\hat{\beta})^T Cov(\Lambda\hat{\beta})^{-1}(\Lambda\hat{\beta}) \\ &= \frac{Y' M_{MP} Y / r(M_{MP})}{MSE}, \end{aligned}$$

(d) Derive a joint 95% confidence region for $(\beta_1 - \alpha_1, \beta_0 + \beta_1, \sigma^2)$

$$\begin{aligned} \beta_1 - \alpha_1 = \delta_1 - (\delta_1 + \delta_3) &= -\delta_3 \\ \beta_0 + \beta_1 &= \delta_0 + \delta_1 \end{aligned}$$

Thus we want a joint confidence region for

$$\begin{aligned} (-\delta_3, \delta_0 + \delta_1, \sigma^2)' &= (\Lambda^T \beta, \sigma^2)^T \\ \Lambda^T &= \begin{bmatrix} 0 & 0 & 0 & -1 \\ 1 & 1 & 0 & 0 \end{bmatrix} \\ \hat{\beta} &= (X'X)^{-1}X'Y \end{aligned}$$

$\Lambda^T\hat{\beta}$ is the UMVUE of $\Lambda^T\beta$ by Gauss-Markov theorem.
Since $\Lambda^T\beta$ is estimable, we can find a matrix P so that

$$\begin{aligned} \Lambda^T &= P'X \\ \Lambda^T\beta = P'X\beta = P'MY, &\qquad M = X(X'X)^{-1}X' \\ M_{MP} &= (MP)[(MP)^T(MP)]^{-1}(MP)^T \end{aligned}$$

M_{MP} is the appropriate opo for testing $H_0 : \Lambda^T\beta = 0$.

$$\begin{aligned} M_{MP}(Y - X\beta) &\sim N(0, \sigma^2 M_{MP}) \\ \frac{\|M_{MP}(Y - X\beta)\|}{\sigma^2} &\sim \chi^2(r(M_{MP})) \\ r(M_{MP}) &= 2 \end{aligned}$$

$M_{MP}Y$ is the UMVUE of $M_{MP}X\hat{\beta}$
The UMVUE of σ^2 is

$$\begin{aligned} \frac{\|(I - M)Y\|}{\sigma^2} &\sim \chi^2(n-4) \\ r(M_{MP}) &= 2 \end{aligned}$$

Thus $M_{MP} \perp (I - M)$
A joint 95% confidence region for $(\Lambda^T\beta, \sigma^2)^T$

$$\begin{aligned} P\{\chi_a^2(2) < \frac{\|M_{MP}(Y - X\beta)\|}{\sigma^2} < \chi_{1-a}^2(2)\}, \qquad P\{\chi_b^2(n-4) < \frac{\|(I - M)Y\|}{\sigma^2} < \chi_{1-b}^2(n-4)\} \\ = P\{\chi_a^2(2) < \frac{\|M_{MP}(Y - X\beta)\|}{\sigma^2} < \chi_{1-a}^2(2)\} P\{\chi_b^2(n-4) < \frac{\|(I - M)Y\|}{\sigma^2} < \chi_{1-b}^2(n-4)\} \\ = 1 - \alpha \end{aligned}$$

where a and b are chosen so that $(1 - 2a)(1 - 2b) = 1 - \alpha$

(e) Under model (3.1), derive a 95% prediction region for a future $q \times 1$ future response vector taken at the $q \times 4$ future covariate matrix X_f.
The prediction could be considered as bayesian rule, so the $\sigma_{new}^2 = \sigma^2(1 + X_f(X'X)^{-1}X_f^T)$.

$$\begin{aligned} Cov(X_f\beta) = Cov(X_f(X'X)^{-1}X'Y) &= [X_f(X'X)^{-1}X']Cov(Y)[X_f X(X'X)^{-1}]^T \\ &= [X_f(X'X)^{-1}X']Cov(Y)[X(X'X)^{-1}X_f^T] = \sigma^2[X_f(X'X)^{-1}X_f^T] \\ \sigma_{new}^2 &= \sigma^2(1 + X_f(X'X)^{-1}X_f^T) \\ \sigma^2 &= \frac{\|(I - M)Y\|}{(n-4)} \end{aligned}$$

Thus a 95% prediction region for a future $q \times 1$ vector Y_f

$$\{Y_f : \frac{((\hat{Y}_f - Y_f)^T(X_f(X'X)^{-1}X_f^T)^{-1}(\hat{Y}_f - Y_f))}{MSE} \le F(.95, q, n-4)\}$$

1.3 MLE in linear model

Consider the usual linear model $Y = X\beta + \epsilon, \epsilon \sim N(0, \sigma^2 I), rank(X) = r$.

Using the property of MVN, $Y \sim N_n(X\beta, \sigma^2 I)$, the joint density $Y = (Y_1, .., Y_n)'$

$$P(Y) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\Big\{-\frac{(Y - X\beta)^T(Y - X\beta)}{\sigma^2}\Big\}$$

We note that $|\sigma^2 I|^{-1/2} = \sigma^{-n}$. We could generally write

$$L(\beta, \sigma^2) \propto (\sigma^2)^{-\frac{n}{2}} \exp\Big\{-\frac{(Y - X\beta)^T(Y - X\beta)}{\sigma^2}\Big\}$$

by dropping the $(2\pi)^{-n/2}$ term,

$$l(\beta, \sigma^2) = -n \log \sigma - \frac{(Y - X\beta)^T(Y - X\beta)}{\sigma^2}$$

Now maximizing $l(\beta, \sigma^2)$ with respect to β is equivalent to minimizing $g(\beta) = (Y - X\beta)^T(Y - X\beta)$ with respect to β. This is just the least squares criterion. Thus the MLE of β for the usual linear model, denoted $\hat{\beta}_{ml}$ satisfies

$$X\hat{\beta}_{ml} = MY$$

where $M = X(X'X)^{-1}X'$ is the orthogonal projection operator onto C(X).

To get the estimate of σ^2, we substitute the MLE of $\hat{\beta}_{ml}$. However here is a catch, it does not consider the loss of degrees of freedom when estimating β, and treating $\hat{\beta}_{ml}$ is known when estimating σ^2. The MLE of σ^2 is not unbiased.

The MLE of σ^2 is $\frac{Y'(I-M)Y}{n}$,

$$E[\hat{\sigma}^2] = \frac{n-r}{n}\sigma^2$$

1.4 Sampling distribution in linear model

Consider the usual linear model $Y = X\beta + \epsilon, \epsilon \sim N(0, \sigma^2 I), rank(X) = r$. Using the property of MVN, $Y \sim N_n(X\beta, \sigma^2 I)$, the joint density $Y = (Y_1, .., Y_n)'$

$\Lambda'\beta$ is an estimable vector of linear function of β, where Λ is a $p \times s$ matrix.

$$\begin{aligned}
\Lambda' &= P'X \\
P'X\beta &= P'MY \\
P'MP &= P'X(X'X)^{-1}X'P = \Lambda'(X'X)^{-1}\Lambda
\end{aligned}$$

Pay attention that $P'MP$ is the covariance matrix of $Cov(P'MY) = (P'M)Cov(Y)(P'M)' = P'M\sigma^2 IMP = \sigma^2 P'MP$

$$P'MY \sim N_s(\Lambda'\beta, \sigma^2\Lambda'(X'X)^{-1}\Lambda)$$

Chapter 2

Generalized Linear Model

Likelihood Ratio Test based on likelihood

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.

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

Conditional Probability

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 $p(n_{11}, n_{12}|n_{13}, ...n_{IJ}, n)$.

2.1 Practice

2.1.1 Contingency table parameters

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

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

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 $(1, \pi_{00}, \pi_{01}, \pi_{10}, \pi_{11})$.

$$
\begin{aligned}
Z_1 &= I[(X, Y) = (0, 0)] \\
Z_2 &= I[(X, Y) = (0, 1)] \\
Z_3 &= I[(X, Y) = (1, 0)] \\
Z_4 &= I[(X, Y) = (1, 1)] \\
p(\theta) &= \prod_k \pi_k^{I(Z_k=1)} \\
M_Z(t) &= E[exp(t^T Z)] = E[exp(t^T(Z_1 + Z_2 + ...Z_n))] = E[exp(t^T Z_1 + t^T Z_2 + ...t^T Z_n)] \\
&= E[\prod_{i=1}^{n} exp(t^T Z_i)] \\
&= \prod_{i=1}^{n} E[exp(t^T Z_i)] \qquad \text{(by independence)} \\
&= \prod_{i=1}^{n} M_{Z_i}(t) = \prod_{i=1}^{n} P(Z_i = 1)e^{tz_i} \qquad \text{by MGF of discrete variable } Z_i \\
&= \left(\sum_{j=1}^{J} \pi_j exp(t_j) \right)^n \qquad \text{by MGF of multinoulli}
\end{aligned}
$$

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

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

(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 $(\pi_{1+}\pi_{0+})^{-1} + (\pi_{+1}\pi_{+0})^{-1} \leq (\pi_{1+}\pi_{+0})^{-1} + (\pi_{+1}\pi_{0+})^{-1}$.

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

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

2.2 Exercise

Consider pairs of independent random variables $(y_{i1}, y_{i2}), i = 1,, n$ such that both y_{i1} and y_{i2} follow a $N(\mu_i, \psi)$ distribution. Let ψ be the parameter of interest and the μ_i are nuisance parameters.

(a) Show that the maximum likelihood estimate of ψ is inconsistent.
The joint density of y_{i1}, y_{i2}

$$P(y_{i1}, y_{i2}) = \frac{1}{2\pi\psi} exp\left(-\frac{(y_{i1} - \mu_i)^2 + (y_{i2} - \mu_i)^2}{2\psi}\right)$$
$$P(y_1, y_2) = \prod_{i=1}^{n} \frac{1}{(2\pi\psi)^n} exp\left(-\sum_{i=1}^{n} \frac{(y_{i1} - \mu_i)^2 + (y_{i2} - \mu_i)^2}{2\psi}\right)$$

The log-likelihood function

$$ln(y_1, y_2) = -nlog(2\pi) - nlog\psi - \sum_{i=1}^{n} \frac{(y_{i1} - \mu_i)^2 + (y_{i2} - \mu_i)^2}{2\psi}$$

Obtain MLE of μ_i, ψ

$$\partial_{\mu_i} ln = -1/(2\psi) \sum_{i=1}^{n} -2(y_{i1} - \mu_i + y_{i2} - \mu_i) = 0, \qquad \hat{\mu_i}$$
$$\mu_i = 1/2(y_{i1} + y_{i2})$$
$$\partial_{\psi} ln = -n/\psi + \frac{\sum_{i=1}^{n}[(y_{i1} - \mu_1)^2 + (y_{i2} - \mu_2)^2]}{2\psi^2} = 0$$
$$\hat{\psi} = 1/2n \left(\sum_{i=1}^{n}[(y_{i1} - \mu_1)^2 + (y_{i2} - \mu_2)^2]\right)$$
$$= \frac{1}{4n} \sum_{i=1}^{n} (y_{i1} - y_{i2})^2$$

As $E(y_{i1} - y_{i2}) = 0, Var(y_{i1} - y_{i2}) = 2\psi$

$$Var(y_{i1} - y_{i2}) = E(y_{i1} - y_{i2})^2 - [E(y_{i1} - y_{i2})]^2 = 2\psi, \qquad E(y_{i1} - y_{i2})^2 = 2\psi$$

By WLLN,

$$\hat{\psi} = \frac{1}{4n} \sum_{i=1}^{n} (y_{i1} - y_{i2})^2 \xrightarrow{n\to\infty} 1/4E(y_{i1} - y_{i2})^2 = \psi/2 \neq \psi$$

So MLE of ψ is not consistent.

(b) Construct a consistent estimate for ψ based on the available information.
From part(a), we can construct $\tilde{\psi} = 2\hat{\psi} = \frac{1}{2n}\sum_{i=1}^{n}(y_{i1} - y_{i2})^2$. By WLLN, the

$$\tilde{\psi} = \frac{1}{2n}\sum_{i=1}^{n}(y_{i1} - y_{i2})^2 \xrightarrow[n\to\infty]{p} = \psi$$

(c) Assume that y_{i1} and y_{i2} follow a $N(\mu_i, \psi_i)$ distribution for $i = 1,,n$, where $\mu_i = \beta_0 + \beta_1(x_i - \bar{x})$ and $\psi_i = exp(\alpha_0 + \alpha_1(x_i - \bar{x}))$, in which x_i is a covariate of interest and $\bar{x}$ is the mean of the x_is. Derive the score test statistic for testing homogeneous variance.
The hypothesis are

$$\begin{aligned} H_0 &: \alpha_1 = 0 \\ H_1 &: \alpha_1 \neq 0 \end{aligned}$$

The log-likelihood function

$$\begin{aligned}
\xi &= (\beta_0, \beta_1, \alpha_0, \alpha_1)^T \\
ln(y_1, y_2, \mu_i, \psi_i) &= -nlog(2\pi) - \sum_{i=1}^{n} log\psi_i - \sum_{i=1}^{n}\frac{(y_{i1} - \mu_i)^2 + (y_{i2} - \mu_i)^2}{2\psi_i} \\
ln(y_1, y_2, \xi) &= -nlog(2\pi) - \sum_{i=1}^{n}(\alpha_0 + \alpha_1(x_i - \bar{x})) \\
&\quad - \sum_{i=1}^{n}\frac{(y_{i1} - \beta_0 - \beta_1(x_i - \bar{x}))^2 + (y_{i2} - \beta_0 - \beta_1(x_i - \bar{x}))^2}{2exp(\alpha_0 + \alpha_1(x_i - \bar{x}))}, \qquad \sum x_i - \bar{x} = 0 \\
&= -nlog(2\pi) - n\alpha_0 - 1/2\sum_{i=1}^{n}\frac{(y_{i1} - \beta_0 - \beta_1(x_i - \bar{x}))^2 + (y_{i2} - \beta_0 + \beta_1(x_i - \bar{x}))^2}{exp(\alpha_0 + \alpha_1(x_i - \bar{x}))}
\end{aligned}$$

We will get the score function and Fisher information for ξ

$$\begin{aligned}
\frac{\partial ln(\xi)}{\partial \alpha_0} &= -n + 1/2\sum_{i=1}^{n}\frac{(y_{i1} - \beta_0 - \beta_1(x_i - \bar{x}))^2 + (y_{i2} - \beta_0 - \beta_1(x_i - \bar{x}))^2}{exp(\alpha_0 + \alpha_1(x_i - \bar{x}))} \\
&= -n + 1/2\sum_{i=1}^{n}\psi_i^{-1}[(y_{i1} - \mu_i)^2 + (y_{i2} - \mu_i)^2] \\
\frac{\partial^2 ln(\xi)}{\partial \alpha_0^2} &= -1/2\sum_{i=1}^{n}\psi_i^{-1}[(y_{i1} - \mu_i)^2 + (y_{i2} - \mu_i)^2]
\end{aligned}$$

$$\begin{aligned}
\frac{\partial ln(\xi)}{\partial \alpha_1} &= 1/2\sum_{i=1}^{n}\psi_i^{-1}[(y_{i1} - \mu_i)^2 + (y_{i2} - \mu_i)^2](x_i - \bar{x}) \\
\frac{\partial^2 ln(\xi)}{\partial \alpha_1^2} &= -1/2\sum_{i=1}^{n}\psi_i^{-1}[(y_{i1} - \mu_i)^2 + (y_{i2} - \mu_i)^2](x_i - \bar{x})^2
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial ln(\xi)}{\partial \beta_0} &= \sum_{i=1}^{n} \psi_i^{-1}[(y_{i1} - \beta_0 - \beta_1(x_i - \bar{x})) + (y_{i2} - \beta_0 - \beta_1(x_i - \bar{x}))] \\
\frac{\partial^2 ln(\xi)}{\partial \beta_0^2} &= -2\sum_{i=1}^{n} \psi_i^{-1}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial ln(\xi)}{\partial \beta_1} &= \sum_{i=1}^{n} \psi_i^{-1}[(y_{i1} - \beta_0 - \beta_1(x_i - \bar{x})) + (y_{i2} - \beta_0 - \beta_1(x_i - \bar{x}))](x_i - \bar{x}) \\
\frac{\partial^2 ln(\xi)}{\partial \beta_1^2} &= -2\sum_{i=1}^{n} \psi_i^{-1}(x_i - \bar{x})^2
\end{aligned}
$$

Other derivatives

$$
\begin{aligned}
\frac{\partial^2 ln(\xi)}{\partial \alpha_0 \alpha_1} &= -1/2\sum_{i=1}^{n} \psi_i^{-1}[(y_{i1} - \mu_i)^2 + (y_{i2} - \mu_i)^2](x_i - \bar{x}) \\
\frac{\partial^2 ln(\xi)}{\partial \alpha_0 \beta_0} &= -\sum_{i=1}^{n} \psi_i^{-1}[(y_{i1} - \mu_i)^2 + (y_{i2} - \mu_i)^2](x_i - \bar{x}) \\
\frac{\partial^2 ln(\xi)}{\partial \alpha_0 \beta_1} &= -\sum_{i=1}^{n} \psi_i^{-1}[(y_{i1} - \mu_i)^2 + (y_{i2} - \mu_i)^2](x_i - \bar{x}) \\
\frac{\partial^2 ln(\xi)}{\partial \alpha_1 \beta_0} &= -\sum_{i=1}^{n} \psi_i^{-1}[(y_{i1} - \mu_i) + (y_{i2} - \mu_i)](x_i - \bar{x}) \\
\frac{\partial^2 ln(\xi)}{\partial \alpha_1 \beta_1} &= -\sum_{i=1}^{n} \psi_i^{-1}[(y_{i1} - \mu_i) + (y_{i2} - \mu_i)](x_i - \bar{x})^2 \\
\frac{\partial^2 ln(\xi)}{\partial \beta_0 \beta_1} &= -2\sum_{i=1}^{n} \psi_i^{-1}(x_i - \bar{x})
\end{aligned}
$$

Taking expectation as $I(\xi) = -E(\partial^2 \xi)$

$$
\begin{aligned}
&E(y_{i1}-\mu_i)^2 = \psi_i, \qquad E(y_{i1}) = E(y_{i2}) = \mu_i, \qquad \sum_{i=1}^{n} x_i - n\bar{x} = 0 \\
&E[\frac{\partial^2 ln(\xi)}{\partial \alpha_0^2}] = -1/2 \sum_{i=1}^{n} \psi_i^{-1}[E(y_{i1}-\mu_i)^2 + E(y_{i2}-\mu_i)^2] = -n \\
&E[\frac{\partial^2 ln(\xi)}{\partial \alpha_1^2}] = -\sum_{i=1}^{n}(x_i-\bar{x})^2 \\
&E[\frac{\partial^2 ln(\xi)}{\partial \beta_0^2}] = -2\sum_{i=1}^{n} \psi_i^{-1} \\
&E[\frac{\partial^2 ln(\xi)}{\partial \beta_1^2}] = -2\sum_{i=1}^{n} \psi_i^{-1}(x_i-\bar{x})^2 \\
&E[\frac{\partial^2 ln(\xi)}{\partial \alpha_0 \alpha_1}] = -1/2 \sum_{i=1}^{n} \psi_i^{-1}[E(y_{i1}-\mu_i)^2 + E(y_{i2}-\mu_i)^2]E(x_i-\bar{x}) = 0 \\
&E[\frac{\partial^2 ln(\xi)}{\partial \alpha_0 \beta_0}] = 0, \qquad E[\frac{\partial^2 ln(\xi)}{\partial \alpha_0 \beta_1}] = 0 \\
&E[\frac{\partial^2 ln(\xi)}{\partial \alpha_1 \beta_0}] = 0, \qquad E[\frac{\partial^2 ln(\xi)}{\partial \alpha_1 \beta_1}] = 0 \\
&E[\frac{\partial^2 ln(\xi)}{\partial \beta_0 \beta_1}] = -2\sum_{i=1}^{n} \psi_i^{-1}(x_i-\bar{x})
\end{aligned}
$$

Then

$$
I(\xi) = -E(\partial^2 \xi) = \begin{bmatrix} n & 0 & 0 & 0 \\ 0 & \sum_{i=1}^{n}(x_i-\bar{x})^2 & 0 & 0 \\ 0 & 0 & 2\sum_{i=1}^{n}\psi_i^{-1} & 2\sum_{i=1}^{n}\psi_i^{-1}(x_i-\bar{x}) \\ 0 & 0 & 2\sum_{i=1}^{n}\psi_i^{-1}(x_i-\bar{x}) & 2\sum_{i=1}^{n}\psi_i^{-1}(x_i-\bar{x})^2 \end{bmatrix}
$$

Under null hypothesis, we have score test statistics follows a chi-square distribution

$$
\frac{\partial ln}{\partial \tilde{\xi}}^T I(\tilde{\xi})^{-1} \frac{\partial ln}{\partial \tilde{\xi}} \sim \chi^2(1)
$$

So we have $\tilde{\psi} = exp(\tilde{\alpha_0})$, then $\tilde{\alpha_0} = ln(\tilde{\psi})$.
From part (a) which ψ is constant, we have $\psi = \frac{1}{4n}\sum_{i=1}^{n}(y_{i1}-y_{i2})^2$ and then,

$$
\begin{aligned}
\hat{\mu_i} &= 1/2(y_{i1}+y_{i2}) \\
\hat{\psi} &= \frac{1}{4n}\sum_{i=1}^{n}(y_{i1}-y_{i2})^2
\end{aligned}
$$

then the score function under $\tilde{\xi}$

$$\dot{l}(\xi) = \begin{bmatrix} \partial_{\alpha_0} l(\xi) & = -n + 1/2\sum_{i=1}^n \tilde{\psi}^{-1}[(y_{i1}-\mu_i)^2 + (y_{i2}-\mu_i)^2] = 0 \\ \partial_{\alpha_1} l(\xi) & = 1/2\sum_{i=1}^n \tilde{\psi}^{-1} 1/2 (y_{i1}-y_{i2})^2(x_i-\bar{x}) = \frac{1}{4\tilde{\psi}}\sum_{i=1}^n (y_{i1}-y_{i2})^2(x_i-\bar{x}) \\ \partial_{\beta_0} l(\xi) & = \sum_{i=1}^n \tilde{\psi}^{-1}[(y_{i1}-\beta_0-\beta_1(x_i-\bar{x})) + (y_{i2}-\beta_0-\beta_1(x_i-\bar{x}))] = 0 \\ \partial_{\beta_1} l(\xi) & = \sum_{i=1}^n \tilde{\psi}^{-1}[(y_{i1}-\beta_0-\beta_1(x_i-\bar{x})) + (y_{i2}-\beta_0-\beta_1(x_i-\bar{x}))](x_i-\bar{x}) = 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ \frac{1}{4\tilde{\psi}}\sum_{i=1}^n (y_{i1}-y_{i2})^2(x_i-\bar{x}) \\ 0 \\ 0 \end{bmatrix}$$

Under null hypothesis, $2\sum_{i=1}^n \psi_i^{-1}(x_i-\bar{x}) = 0$, then

$$I_n(\tilde{\xi}) = \begin{bmatrix} n & 0 & 0 & 0 \\ 0 & \sum_{i=1}^n (x_i-\bar{x})^2 & 0 & 0 \\ 0 & 0 & 2\sum_{i=1}^n \tilde{\psi}^{-1} & 0 \\ 0 & 0 & 0 & 2\sum_{i=1}^n \tilde{\psi}^{-1}(x_i-\bar{x})^2 \end{bmatrix}$$

The score test statistics

$$SCn = \frac{\partial ln}{\partial \tilde{\xi}}^T I_n(\tilde{\xi})^{-1} \frac{\partial ln}{\partial \tilde{\xi}} = (0, \frac{1}{4\tilde{\psi}}\sum_{i=1}^n (y_{i1}-y_{i2})^2(x_i-\bar{x}), 0, 0)$$

$$\begin{bmatrix} n & 0 & 0 & 0 \\ 0 & \sum_{i=1}^n (x_i-\bar{x})^2 & 0 & 0 \\ 0 & 0 & 2\sum_{i=1}^n \tilde{\psi}^{-1} & 0 \\ 0 & 0 & 0 & 2\sum_{i=1}^n \tilde{\psi}^{-1}(x_i-\bar{x})^2 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ \frac{1}{4\tilde{\psi}}\sum_{i=1}^n (y_{i1}-y_{i2})^2(x_i-\bar{x}) \\ 0 \\ 0 \end{bmatrix}$$

$$= \frac{\left[\frac{1}{4\tilde{\psi}}\sum_{i=1}^n (y_{i1}-y_{i2})^2(x_i-\bar{x})\right]^2}{\sum_{i=1}^n (x_i-\bar{x})^2}$$

With $\tilde{\psi} = \frac{1}{4n}\sum_{i=1}^n (y_{i1}-y_{i2})^2$, we have

$$SCn = \frac{\left[n^2\sum_{i=1}^n (y_{i1}-y_{i2})^2(x_i-\bar{x})\right]^2}{[\sum_{i=1}^n (y_{i1}-y_{i2})^2]^2 \sum_{i=1}^n (x_i-\bar{x})^2} \sim \chi^2(1)$$

We will reject the H_0 if $SCn > \chi^2(1, 1-\alpha)$.

2.2.1 e

Suppose that the vector $Y = (Y_0; Y_1; Y_2)^T$ follows a multinomial distribution with total count m and probability vector $(\gamma_0; \gamma_1; \gamma_2)^T$ with

$$\gamma_j = \binom{2}{j}\pi^j(1-\pi)^{2-j}\theta^{-j(2-j)}/f(\pi, \theta), \qquad j = 0, 1, 2$$

where

$$f(\pi, \theta) = \sum_{k=0}^{2} \binom{2}{k} \pi^k (1-\pi)^{2-k} \theta^{-k(2-k)}$$

and $0 \leq \pi \leq 1, \theta > 0$ are parameters. Furthermore, define $\lambda = log\frac{\pi}{1-\pi}$ and $\psi = log\theta$.

(a) Derive a sufficient statistic for λ assuming $\psi = \psi_0$ is known. Derive a conditional likelihood for ψ.
Write the joint distribution of Y

$$\begin{aligned}
P(Y) &= \binom{m}{y_0, y_1, y_2} \gamma_1^{y_1} \gamma_2^{y_2} \gamma_0^{y_0} \\
&= exp\left[log\binom{m}{y_0, y_1, y_2} + y_0 log\gamma_0 + y_1 log\gamma_1 + y_2 log\gamma_2\right]
\end{aligned}$$

$$\begin{aligned}
\gamma_0 &= \binom{2}{0} \pi^0 (1-\pi)^2 \theta^0 / f(\pi, \theta) = (1-\pi)^2 / f(\pi, \theta) \\
\gamma_1 &= \binom{2}{1} \pi^1 (1-\pi)^1 \theta^{-1} / f(\pi, \theta) = 2\pi(1-\pi)\theta^{-1} / f(\pi, \theta) \\
\gamma_2 &= \binom{2}{2} \pi^2 (1-\pi)^0 \theta^0 / f(\pi, \theta) = \pi^2 / f(\pi, \theta)
\end{aligned}$$

$$\begin{aligned}
logP(Y) &= log\binom{m}{y_0, y_1, y_2} + y_0[2log(1-\pi) - logf(\pi, \theta)] \\
&+ y_1[log2\pi(1-\pi) - log\theta - logf(\pi, \theta)] + y_2[2log\pi - logf(\pi, \theta)] \\
f(\pi, \theta) &= \binom{2}{0}\pi^0(1-\pi)^2\theta^0 + \binom{2}{1}\pi^1(1-\pi)^1\theta^{-1} + \binom{2}{2}\pi^2(1-\pi)^0\theta^0 \\
logf(\pi, \theta) &= 2log(1-\pi) + log2\pi(1-\pi) - log\theta + 2log\pi \\
logP(Y) &= log\binom{m}{y_0, y_1, y_2} + (2y_0 + y_1)log(1-\pi) \\
&- (y_0 + y_1 + y_2)logf(\pi, \theta) + (y_1 + 2y_2)log\pi + y_1 log2 - y_1 log\theta \\
m &= y_0 + y_1 + y_2, \qquad y_1 = m - y_0 - y_2 \\
logP(Y) &= log\binom{m}{y_0, y_1, y_2} + (m + y_0 - y_2)log(1-\pi) - mlogf(\pi, \theta) \\
&+ (m - y_0 + y_2)log\pi + y_1 log2 - y_1 log\theta \\
&= log\binom{m}{y_0, y_1, y_2} + mlog\left[\frac{e^\lambda}{1+e^\lambda}\frac{1}{1+e^\lambda}\frac{(1+e^\lambda)^2}{1+2e^{\lambda-\psi}+e^{2\lambda}}\right] \\
&- (y_0 - y_2)\lambda + y_1 log2 - y_1\psi
\end{aligned}$$

If assume $\psi = \psi_0$ is known, then a sufficient statistics is $m, y_0 - y_2$.

$$logP(Y) = log\binom{m}{y_0, y_1, y_2} + mlog\left[\frac{e^{\lambda}}{1 + 2e^{\lambda-\psi} + e^{2\lambda}}\right] - (y_0 - y_2)\lambda + y_1 log2 - y_1\psi$$

Let $y_2 - y_0 = t$,

$$P(t) = \sum_t \binom{m}{y_0, y_1, y_2}\left[\frac{e^{\lambda}}{1 + 2e^{\lambda-\psi} + e^{2\lambda}}\right]^m exp(\lambda t)2^{y_1} exp(-\psi y_1)$$

$$P(y_1|t) = \frac{P(t,Y)}{P(t)} = \frac{\binom{m}{y_0,y_1,y_2}\left[\frac{e^{\lambda}}{1+2e^{\lambda-\psi}+e^{2\lambda}}\right]^m exp(\lambda t)2^{y_1}exp(-\psi y_1)}{\sum_t \binom{m}{y_0,y_1,y_2}\left[\frac{e^{\lambda}}{1+2e^{\lambda-\psi}+e^{2\lambda}}\right]^m exp(\lambda t)2^{y_1}exp(-\psi y_1)}$$

$$= \frac{\frac{1}{y_0!y_1!y_2!}2^{y_1}exp(-\psi y_1)}{\sum_{y_2'-y_0'=t}\frac{1}{y_0'!y_1'!y_2'!}2^{y_1'}exp(-\psi y_1')}$$

The conditional distribution for ψ

$$P(y_1, \psi|t) = \frac{\frac{1}{y_0!y_1!y_2!}2^{y_1}exp(-\psi y_1)}{\sum_{y_2'-y_0'=t}\frac{1}{y_0'!y_1'!y_2'!}2^{y_1'}exp(-\psi y_1')}$$

(b) The data $y_0 = 3; y_1 = 0; y_2 = 2$ were observed. Based on the conditional likelihood of Part (a), compute the exact one-sided p-value for testing $H0 : \theta = 1$ against $H_0 : \theta > 1$ with λ unspecified.

The null hypothesis could be written as

$$H_0 : \psi = 0 \qquad vs. \qquad H_1 : \psi \neq 0$$

From $y_0 = 3; y_1 = 0; y_2 = 2$, we have $t = y_2 - y_0 = -1, m = 5$. There are possible 3 combinations that t=-1 as below

y_1	y_2	y_0	t	case
0	2	3	-1	1
2	1	2	-1	2
4	0	1	-1	3

So under H_0, the conditional probability for y_1 in the above 3 cases are

$$\begin{aligned} denominator &= \frac{1}{0!2!3!}2^0 exp(-\psi 0) + \frac{1}{1!2!2!}2^2 exp(-\psi 2) + \frac{1}{0!4!1!}2^4 exp(-\psi 4) \\ &= 2/3exp(-4\psi) + exp(-2\psi) + 1/12 = 21/12 \end{aligned}$$

$$P(y_1 = 0, \psi|t = -1) = \frac{\frac{1}{0!2!3!}2^0 exp(0)}{\sum_{y_2'-y_0'=t}\frac{1}{y_0'!y_1'!y_2'!}2^{y_1'}exp(-\psi y_1')} = \frac{1/12}{21/12} = 1/21$$

$$P(y_1 = 2, \psi|t = -1) = \frac{\frac{1}{1!2!2!}2^2 exp(0)}{\sum_{y_2'-y_0'=t}\frac{1}{y_0'!y_1'!y_2'!}2^{y_1'}exp(-\psi y_1')} = \frac{1/12}{21/12} = 12/21$$

$$P(y_1 = 4, \psi|t = -1) = \frac{\frac{1}{0!4!1!}2^4 exp(0)}{\sum_{y_2'-y_0'=t}\frac{1}{y_0'!y_1'!y_2'!}2^{y_1'}exp(-\psi y_1')} = \frac{1/12}{21/12} = 8/21$$

We will reject H_0 if $P(y_1|t=-1) < 0.05$. Under the current sample, one sided test p-value for $P(y_1 = 0|t=-1) = 1/21 = 0.0476$, that $\psi \neq 0$.

2.2.2 b

Consider the following

(a) For an arbitrary model, consider the conditional score statistic

$$U_\psi(\xi) = \frac{\partial l_c(\xi, \psi_0)}{\partial \psi}|_{\psi_0=\psi}$$

Show that the conditional score statistic for any model can be written as

$$U_\psi(\xi) = \partial_\psi logp(Y|\xi) - E[\partial_\psi logp(Y|\xi)|s_\lambda(\psi_0)]|_{\psi_0=\psi}$$

The conditional score statistic is the derivative of the conditional distribution

$$U_\psi(\xi) = \frac{\partial l_c(\xi, \psi_0)}{\partial \psi}|_{\psi_0=\psi}$$
$$p(\mathbf{Y}|\xi) = p(\mathbf{Y}|s_\lambda(\psi_0), \xi)p(s_\lambda(\psi_0)|\xi), \qquad p(\mathbf{Y}|s_\lambda(\psi_0), \xi) = \frac{p(\mathbf{Y}|\xi)}{p(s_\lambda(\psi_0)|\xi)}$$
$$l_c(\xi, \psi_0) = logp(\mathbf{Y}|s_\lambda(\psi_0), \xi) = logp(\mathbf{Y}|\xi) - logp(s_\lambda(\psi_0)|\xi)$$

Then we need to prove

$$U_\psi(\xi) = \frac{\partial l_c(\xi, \psi_0)}{\partial \psi}|_{\psi_0=\psi} = \partial_\psi logp(\mathbf{Y}|\xi) - \partial_\psi logp(s_\lambda(\psi_0)|\xi)$$
$$\partial_\psi logp(s_\lambda(\psi_0)|\xi) = E[\partial_\psi logp(Y|\xi)|s_\lambda(\psi_0)]|_{\psi_0=\psi}$$

We can write

$$logp(\mathbf{Y}|\xi) = logp(\mathbf{Y}|s_\lambda(\psi_0), \xi) + logp(s_\lambda(\psi_0)|\xi)$$
$$E\left(\partial_\psi[logp(\mathbf{Y}|\xi)|s_\lambda]\right) = E\left(\partial_\psi[logp(\mathbf{Y}|s_\lambda(\psi_0), \xi)|s_\lambda]\right) + E\left(\partial_\psi[logp(s_\lambda(\psi_0), \xi)|s_\lambda]\right)$$

in which, the integral and expectation can switch, then we have

$$E\left(\partial_\psi[logp(\mathbf{Y}|s_\lambda(\psi_0), \xi)|s_\lambda]\right) = \partial_\psi E\left([logp(\mathbf{Y}|s_\lambda(\psi_0), \xi)|s_\lambda]\right) = \partial_\psi E\left([logp(\mathbf{Y}|\xi)]\right) = 0$$

So,

$$E\left(\partial_\psi[logp(\mathbf{Y}|\xi)|s_\lambda]\right) = \partial_\psi logp(s_\lambda(\psi_0), \xi)$$

Then we show

$$U_\psi(\xi) = \partial_\psi logp(Y|\xi) - E[\partial_\psi logp(Y|\xi)|s_\lambda(\psi_0)]|_{\psi_0=\psi}$$

(b) Suppose that $y_1;..y_n$ are independent and y_i follows a Poisson distribution with mean $exp(\lambda_0 + \lambda_1 x_{i1} + \psi x_{i2})$, where $(x_{i1}; x_{i2})$ are covariates, $\lambda = (\lambda_0; \lambda_1)$ is the nuisance parameter vector and ψ is the parameter of interest. Derive the conditional likelihood of ψ and show that this conditional likelihood is free of λ.
The joint distribution of $(y_1, , y_n)$ is given by

$$P(Y|\lambda, \psi) = exp\left(\sum_{i=1}^{n} y_i(\lambda_0 + \lambda_1 x_{i1} + \psi x_{i2}) - \sum_{i=1}^{n} exp(\lambda_0 + \lambda_1 x_{i1} + \psi x_{i2}) - logy_i!\right)$$

Thus, $S_0 = \sum_{i=1}^{n} y_i$ is the sufficient and complete statistics for λ_0, and $S_1 = \sum_{i=1}^{n} y_i x_{i1}$ is the sufficient and complete statistics for λ_1.
The conditional distribution of ψ given S_0, S_1 is given by

$$\begin{aligned}
p(\mathbf{Y}, \psi|S = (S_0, S_1)) &= \frac{exp\left(\sum_{i=1}^{n} y_i(\lambda_0 + \lambda_1 x_{i1} + \psi x_{i2}) - \sum_{i=1}^{n} exp(\lambda_0 + \lambda_1 x_{i1} + \psi x_{i2}) - logy_i!\right)}{\sum_{y' \in S} exp\left(\sum_{i=1}^{n} y_i'(\lambda_0 + \lambda_1 x_{i1} + \psi x_{i2}) - \sum_{i=1}^{n} exp(\lambda_0 + \lambda_1 x_{i1} + \psi x_{i2}) - logy_i'!\right)} \\
&= \frac{exp\left(S_1\lambda_0 + S_2\lambda_1 + S_3\psi\right) - \sum_{i=1}^{n} exp(\lambda_0 + \lambda_1 x_{i1} + \psi x_{i2}) - logy_i!)}{\sum_{y' \in S} exp\left(S_1'\lambda_0 + S_2'\lambda_1 + S_3'\psi\right) - \sum_{i=1}^{n} exp(\lambda_0 + \lambda_1 x_{i1} + \psi x_{i2}) - logy_i'!)} \\
&= \frac{exp\left(S_3\psi - logy_i!\right)}{\sum_{y' \in S} exp\left(S_3'\psi - logy_i'!\right)}, \qquad S_3 = \sum_{i=1}^{n} y_i x_{i2}, S_3' = \sum_{i=1}^{n} y_i' x_{i2}
\end{aligned}$$

which is independent of λ.

(c) Derive the conditional score statistic for part (b) and write out a Newton-Raphson algorithm for obtaining the conditional maximum likelihood estimate of ψ based on $U_\psi(\xi)$.
The log likelihood of the conditional distribution is

$$l_c(\psi) = S_3\psi - logy_i! - log\left[\sum_{y' \in S} exp\left(S_3'\psi - logy_i'!\right)\right], \qquad S_3 = \sum_{i=1}^{n} y_i x_{i2}, S_3' = \sum_{i=1}^{n} y_i' x_{i2}$$

The score function and observed fisher information is

$$\begin{aligned}
U_\psi(\xi) &= \frac{\partial l_c(\xi, \psi_0)}{\partial \psi}|_{\psi_0 = \psi} \\
&= \psi - \frac{\sum_{y' \in S} S_3' exp\left(S_3'\psi - logy_i'!\right)}{\sum_{y' \in S} exp\left(S_3'\psi - logy_i'!\right)}
\end{aligned}$$

$$\frac{\partial^2 l_c(\xi, \psi_0)}{\partial \psi^2} = \left[\frac{\sum_{y' \in S} S_3' exp\left(S_3'\psi - logy_i'!\right)}{\sum_{y' \in S} exp\left(S_3'\psi - logy_i'!\right)}\right]^2 - \frac{\sum_{y' \in S} S_3'^2 exp\left(S_3'\psi - logy_i'!\right)}{\sum_{y' \in S} exp\left(S_3'\psi - logy_i'!\right)}$$

The newton-Raphson algorithm

$$\psi^{k+1} = \psi^k - \left[\frac{\partial^2 l_c(\psi^k)}{\partial \psi^2}\right]^{-1} U_\psi(\psi^k)$$

where $\frac{\partial^2 l_c(\psi^k)}{\partial \psi^2}, U_\psi(\psi^k)$ are from above equations.

(d) Now suppose that we only have two random variables $y_1 \sim Poisson(\mu_1)$ and $y_2 \sim Poisson(\mu_2)$, where y_1 and y_2 are independent. We are interested in making inferences on the ratio $\psi = \mu_1/\mu_2$. Let $\xi = (\psi, \lambda)$, where λ represents the nuisance parameter.

(i) Show that the log-likelihood function of ξ can be written as

$$l(\xi) = (y_1 + y_2)\lambda + y_1 log(\psi) - exp(\lambda)(1 + \psi)$$

where λ is a function of μ_2. Explicitly state what λ is.
Write the joint distribution of y_1, y_2

$$\begin{aligned}
P(y_1, y_2) &= \frac{\mu_1^{y_1} e^{-\mu_1}}{y_1!} \frac{\mu_2^{y_2} e^{-\mu_2}}{y_2!} \\
logP(y_1, y_2) &= y_1 log\mu_1 - \mu_1 + y_2 \log \mu_2 - \mu_2 - logy_1! - logy_2! \\
&= y_1 log\frac{\mu_1}{\mu_2} + y_1 log\mu_2 + y_2 log\mu_2 - \mu_1 - \mu_2 - logy_1! - logy_2! \\
&= y_1 log\frac{\mu_1}{\mu_2} + (y_1 + y_2) log\mu_2 - \mu_2(\mu_1/\mu_2 + 1) - logy_1! - logy_2!
\end{aligned}$$

where

$$\begin{aligned}
\psi &= log\frac{\mu_1}{\mu_2} \\
\lambda &= log\mu_2
\end{aligned}$$

(ii) Derive the conditional likelihood of ψ and write out a Newton-Raphson algorithm for obtaining the conditional maximum likelihood estimate of ψ .
From part (a), we see $y_1 + y_2$ is the sufficient statistics for λ, while $y_1 + y_2 \sim Poission(\mu_1 + \mu_2)$ then we have conditional distribution of ψ condition on $S = y_1 + y_2$.

$$\begin{aligned}
Y(\psi|S = y_1 + y_2, \lambda) &= \frac{exp\left[y_1\psi + (y_1 + y_2)\lambda - exp(\lambda)(\psi + 1) - logy_1! - logy_2!\right]}{exp\left[(y_1 + y_2)log(\mu_1 + \mu_2) - (\mu_1 + \mu_2) - log(y_1 + y_2)!\right]} \\
&= \frac{exp\left[y_1\psi + S\lambda - exp(\lambda)(\psi + 1) - logy_1! - logy_2!\right]}{exp\left[S(\lambda + log(\psi + 1)) - exp(\lambda)(\psi + 1) - logS!\right]} \\
&= \frac{exp\left[y_1\psi - logy_1! - logy_2!\right]}{exp\left[(y_1 + S - y_1)log(\psi + 1)) - logS!\right]} \\
&= \binom{S}{y_1} \left(\frac{\psi}{1 + \psi}\right)^{y_1} \left(\frac{1}{1 + \psi}\right)^{S - y_1}
\end{aligned}$$

The conditional distribution is a binomial, $B(S, \psi/(1 + \psi))$.

The score function and observed fisher information

$$
\begin{aligned}
logY(\psi|S,\lambda) &= y_1 log\psi - Slog(1+\psi) + log\binom{S}{y_1} \\
\partial_\psi logY(\psi|S,\lambda) &= \frac{y_1}{\psi} - \frac{S}{1+\psi} = 0, \qquad \hat{\psi} = y_1/(S-y_1) \\
\partial_\psi^2 logY(\psi|S,\lambda) &= -\frac{y_1}{\psi^2} + \frac{S}{(1+\psi)^2}
\end{aligned}
$$

The $CMLE = \hat{\psi} = y_1/(S-y_1)$. And the newton-Raphson equation

$$
\begin{aligned}
\psi^{k+1} &= \psi^k - \left[\frac{\partial^2 l_c(\psi^k)}{\partial\psi^2}\right]^{-1} U_\psi(\psi^k) \\
&= \psi^k - \left[-\frac{y_1}{\psi^2} + \frac{S}{(1+\psi)^2}\right]^{-1}\left[\frac{y_1}{\psi} - \frac{S}{1+\psi}\right]|_{\psi=\psi^k} \\
&= \psi^k + \frac{y_1/\psi^k - S/(1+\psi^k)}{y_1/{\psi^k}^2 - S/(1+\psi^k)^2}
\end{aligned}
$$

2.2.3 a

Suppose that $y_1; ...y_n$ are independent Bernoulli random variables, where $y_i \sim Bernoulli(\pi)$, and we consider a logistic regression so that $logit(\pi) = x_i'\beta$, where $\beta = (\beta_1; ...\beta_p)$. Our interest is inference on $(\beta_1; \beta_2)$, with all other parameters being treated as nuisance.

(a) Derive the conditional likelihood of $(\beta_1; \beta_2)$ and express it in the simplest possible form.
The joint distribution of $y_1; ...y_n$

$$
\begin{aligned}
p(Y) &= \prod_{i=0}^{n} p_i^{y_i}(1-p_i)^{(1-y_i)} \\
logp(Y) &= \sum_{i=0}^{n} y_i logp_i + (1-y_i)log(1-p_i) = \sum_{i=0}^{n} y_i log\frac{p_i}{1-p_i} + log(1-p_i) \\
logit(pi) &= log\frac{p_i}{1-p_i} = x_i'\beta, \qquad p_i = \frac{exp(x_i'\beta)}{1+exp(x_i'\beta)} \\
logp(Y) &= \sum_{i=0}^{n} y_i x_i'\beta - log(1+exp(x_i'\beta)) \\
&= \sum_{i=0}^{n} y_i(x_{i1}\beta_1 + x_{i2}\beta_2 + x_{i3}\beta_3 + ..x_{ip}\beta_p) - log(1+exp(x_i'\beta))
\end{aligned}
$$

We can see that $\sum_{i=0}^{n} x_{i1}y_i$ is a sufficient and complete statistics for β_1. When only $(\beta_1; \beta_2)$ are the interest, and all other parameters being treated as nuisance. Then $s_j = \sum_{i=0}^{n} y_i x_{ij}$ is sufficient statistics for β_j. Let $S = (s_3, s_4, ..s_p)$

$$
\begin{aligned}
P(\beta_1,\beta_2|S) &= \frac{exp\left[\sum_{i=0}^{n}(y_i x_{i1})\beta_1 + (y_i x_{i2})\beta_2 + ..(y_i x_{ip})\beta_p - log(1+exp(x_i'\beta))\right]}{\sum_{t\in S} exp\left[(t_i x_{i1})\beta_1 + (t_i x_{i2})\beta_2 + ...(t_i x_{ip})\beta_p - log(1+exp(x_i^T\beta)\right]} \\
&= \frac{exp\left(\sum_{i=0}^{n}(y_i x_{i1})\beta_1 + (y_i x_{i2})\beta_2)\right)}{\sum_{t\in S} exp\left((t_i x_{i1})\beta_1 + (t_i x_{i2})\beta_2)\right)} \\
&= \frac{exp\left(S_1\beta_1 + S_2\beta_2)\right)}{\sum_{S'} exp\left(S_1'\beta_1 + S_2'\beta_2)\right)}, \qquad S_j = \sum_{i=0}^{n}(y_i x_{ij}), S_j' = \sum_{i=0}^{n}(t_i x_{ij})
\end{aligned}
$$

(b) Derive the score equations for $(\beta_1;\beta_2)$ based on the conditional likelihood derived in part (a).
The log conditional distribution is

$$
\begin{aligned}
l_c(\beta_1,\beta_2|S) &= logp(Y,\xi) - logp(s,\lambda,\psi_0) = logP(\beta_1,\beta_2|S) \\
l_c(\beta_1,\beta_2|S) &= log\frac{exp\left(S_1\beta_1 + S_2\beta_2)\right)}{\sum_{S'} exp\left(S_1'\beta_1 + S_2'\beta_2)\right)} = S_1\beta_1 + S_2\beta_2 - log\sum_{S'} exp\left(S_1'\beta_1 + S_2'\beta_2)\right) \\
\frac{\partial l_c}{\partial\beta_1} &= S_1 - \frac{\sum_{S'} S_1' exp\left(S_1'\beta_1 + S_2'\beta_2)\right)}{\sum_{S'} exp\left(S_1'\beta_1 + S_2'\beta_2)\right)} \\
\frac{\partial l_c}{\partial\beta_2} &= S_2 - \frac{\sum_{S'} S_2' exp\left(S_1'\beta_1 + S_2'\beta_2)\right)}{\sum_{S'} exp\left(S_1'\beta_1 + S_2'\beta_2)\right)}
\end{aligned}
$$

The score equations are setting the score function to 0

$$
SCn = 0 = \begin{bmatrix} S_1 - \frac{\sum_{S'} S_1' exp\left(S_1'\beta_1 + S_2'\beta_2)\right)}{\sum_{S'} exp\left(S_1'\beta_1 + S_2'\beta_2)\right)} \\ S_2 - \frac{\sum_{S'} S_2' exp\left(S_1'\beta_1 + S_2'\beta_2)\right)}{\sum_{S'} exp\left(S_1'\beta_1 + S_2'\beta_2)\right)} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}
$$

(c) Derive the asymptotic covariance matrix of the conditional maximum likelihood estimates of $(\beta_1;\beta_2)$.
The Fisher information of $(\beta_1;\beta_2)$

$$
\begin{aligned}
\frac{\partial^2 l_c}{\partial\beta_1^2} &= \left[\frac{\sum_T T_1 exp\left(T_1\beta_1 + T_2\beta_2\right)}{\sum_T exp\left(T_1\beta_1 + T_2\beta_2\right)}\right]^2 - \frac{\sum_T T_1^2 exp\left(T_1\beta_1 + T_2\beta_2\right)}{\sum_T exp\left(T_1\beta_1 + T_2\beta_2\right)} \\
\frac{\partial^2 l_c}{\partial\beta_2^2} &= \left[\frac{\sum_T T_2 exp\left(T_1\beta_1 + T_2\beta_2\right)}{\sum_T exp\left(T_1\beta_1 + T_2\beta_2\right)}\right]^2 - \frac{\sum_T T_2^2 exp\left(T_1\beta_1 + T_2\beta_2\right)}{\sum_T exp\left(T_1\beta_1 + T_2\beta_2\right)} \\
\frac{\partial^2 l_c}{\partial\beta_1\beta_2} &= \frac{\left[\sum_T T_1 exp\left(T_1\beta_1 + T_2\beta_2\right)\right]\left[\sum_T T_2 exp\left(T_1\beta_1 + T_2\beta_2\right)\right]}{\left[\sum_T exp\left(T_1\beta_1 + T_2\beta_2\right)\right]^2} - \frac{\sum_T T_1 T_2 exp\left(T_1\beta_1 + T_2\beta_2\right)}{\sum_T exp\left(T_1\beta_1 + T_2\beta_2\right)}
\end{aligned}
$$

Thus the asymptotic covariance matrix $Cov(\beta_1, \beta_2)$ is

$$
\begin{aligned}
Cov(\beta_1, \beta_2) &= I(\beta_1, \beta_2)^{-1} \\
I(\beta_1, \beta_2) &= -E\left[\frac{\partial^2 l_c}{\partial \beta^2}\right] = -\lim_{n\to\infty} \frac{I_n(\beta)}{n} \\
I_n(\beta) &= -\begin{bmatrix} \frac{\partial^2 l_c}{\partial \beta_1^2} & \frac{\partial^2 l_c}{\partial \beta_1 \beta_2} \\ \frac{\partial^2 l_c}{\partial \beta_1 \beta_2} & \frac{\partial^2 l_c}{\partial \beta_2^2} \end{bmatrix}
\end{aligned}
$$

(d) Derive the conditional score test for testing $H_0 : \beta_1 = \beta_2 = 0$.

$$SCn = \frac{\partial l_c}{\partial \tilde{\beta}}^T I_n(\tilde{\beta})^{-1} \frac{\partial l_c}{\partial \tilde{\beta}} \sim \chi^2(1)$$

SCn is estimated under $H_0, \beta_1 = \beta_2 = 0$. The SCn quadratic form is rank 1, so the degrees of freedom is 1.
We will reject H_0 if $SCn > \chi^2(1, \alpha)$.

Chapter 3

Exponential Family

3.1 The Standard Exponential Distribution

The standard exponential distribution family

$$p(y|\theta) = \phi\Big[\exp\Big(y\theta - b(\theta)\Big) - c(y)\Big] - \frac{1}{2}s(y,\phi)$$

We will explore the fun characteristics of the exponential family

(i) Mean and Variance by derivatives

$$\begin{aligned}
log \int p(y|\theta) = log \int \phi\Big[\exp\Big(y\theta - b(\theta)\Big) - c(y)\Big] - \frac{1}{2}s(y,\phi)dv = 0 \\
log \int \exp\{(y\theta)\}h(y)v(dy) = b(\theta) \\
\partial_\theta log \int \exp\{(y\theta)\}h(y)v(dy) = \partial_\theta b(\theta)
\end{aligned}$$

To proceed we need to move the gradient past the integral sign. In general derivatives can not be moved past integral signs (both are certain kinds of limits, and sequences of limits can differ depending on the order in which the limits are taken). However it turns out that the move is justified in this case by an appeal to the dominated convergence theorem.

$$
\begin{aligned}
\partial_\theta b(\theta) &= \partial_\theta log \int \exp\{(y\theta)\}h(y)v(dy) \\
&= \frac{\int y \exp\{(y\theta)\}h(y)v(dy)}{\int \exp\{(y\theta)\}h(y)v(dy)} \\
&= \int y \exp\{y\theta - b(\theta)\}h(x)v(dy) \\
&= E[y]
\end{aligned}
$$

Also we can see that the first derivative of $b(\theta)$ is equal to the mean of the sufficient statistics. Similar for the variance.

Another proof is to use the Bartlett's identities

Suppose that differentiation and integration are exchangeable and all the necessary expectations are finite. We have the following results:

$$
\begin{aligned}
E\_\xi\Big(\partial_j l_n\Big) &= 0, \\
E_\xi\Big(\partial^2_{j,k} l_n\Big) + E_\xi\Big(\partial_j l_n \partial_k l_n\Big) &= 0
\end{aligned}
$$

By the above two equations, we can get the expectation and variance.

3.2 The Bernoulli Distribution

The standard exponential distribution family

$$
p(y|\theta) = \phi\Big[\exp\Big(y\theta - b(\theta)\Big) - c(y)\Big] - \frac{1}{2}s(y, \phi)
$$

For Bernoulli distribution,

$$
\begin{aligned}
p(x|\pi) &= \pi^x(1-\pi)^{1-x} \\
&= \exp\{\log\Big(\frac{\pi}{1-\pi}\Big)x + \log(1-\pi)\}
\end{aligned}
$$

We see that Bernoulli distribution is an exponential family distribution with

$$
\begin{aligned}
\theta &= \log\Big(\frac{\pi}{1-\pi}\Big) \\
b(\theta) &= -\log(1-\pi) = \log\Big(1 + \exp(\theta)\Big)x \\
\phi &= 1
\end{aligned}
$$

3.2.1 Mean and Variance

For a univariate random variable Y, in this case, all the Y_i have the same π

$$\begin{aligned}\frac{\partial b(\theta)}{\partial\theta} &= \frac{\exp(\theta)}{1+\exp(\theta)} = \frac{1}{1+\exp(-\theta)} = \mu = E(Y)\\ \frac{\partial^2 b(\theta)}{\partial\theta\,\partial\theta} &= \frac{\exp(\theta)}{\Big[1+\exp(\theta)\Big]^2} = \mu(1-\mu) = Var(Y)\end{aligned}$$

In regression model, $logit(\pi) = X\beta$, which β is a vector, then we will use the chain rule. And each individual y_i has its own equation that π_i is different.

$$\begin{aligned}\theta &= X\beta, \qquad \theta_i = x_i^T\beta\\ \partial_\beta b(\theta_i) &= \partial_{\theta_i} b(\theta_i)\partial_\beta\theta_i\\ &= \frac{\exp(\theta_i)}{1+\exp(\theta_i)}x_i = \frac{1}{1+\exp(-\theta)}x_i = \mu_i x_i\\ \partial^2_\beta b(\theta_i) &= \frac{\exp(\theta_i)}{\Big[1+\exp(\theta_i)\Big]^2}x_i^{\otimes 2} = \mu_i(1-\mu_i)x_i^{\otimes 2}\end{aligned}$$

And we will need to connect this with the Fisher Information or Newton-Raphson algorithm

$$\begin{aligned}\theta_i &= k\Big(x_i^T\beta\Big) = x_i^T\beta\\ \xi &= (\beta,\phi)\\ ln(\xi) &= \sum_{i=1}^n \phi\Big[y_i k\Big(x_i^T\beta\Big) - b\Big(k\Big(x_i^T\beta\Big)\Big) - c(y_i)\Big] - \frac{1}{2}s(y_i,\phi)\\ \dot{ln}(\beta) &= \frac{\partial ln(\beta)}{\partial\beta} = \phi\sum_{i=1}^n\Big[y_i - \dot{b}\Big(k\Big(x_i^T\beta\Big)\Big)\Big]\dot{k}\Big(x_i^T\beta\Big)x_i\\ &= \sum_{i=1}^n\Big[y_i-\mu_i\Big]x_i\\ \ddot{ln}(\beta) &= \frac{\partial^2 ln(\beta)}{\partial\beta\,\partial\beta} = -\phi\sum_{i=1}^n \ddot{b}\Big(k(x_i^T\beta)\Big)\dot{k}(x_i^T\beta)^2 x_i x_i^T + \phi\sum_{i=1}^n\Big[y_i-\dot{b}(k(x_i^T\beta))\Big]\ddot{k}(x_i^T\beta)x_i x_i^T\\ &= -\sum_{i=1}^n \ddot{b}\Big(\theta_i\Big)x_i x_i^T = -\sum_{i=1}^n V(\theta_i)x_i x_i^T, \qquad \partial^2_\beta b(\theta_i) = V(\theta_i)\end{aligned}$$

let

$$
\begin{aligned}
V(\theta) &= diag\{V(\theta_i)\}, \qquad e_i = y_i - \mu_i \\
\sum_{i=1}^{n} V(\theta_i)x_i x_i^T &= XV(\theta)V^T \\
\mu_i &= \dot{b}(\theta_i), \qquad v_i = \ddot{b}(\theta_i) \\
\dot{\theta}_i &= \partial_\beta \theta_i = \dot{k}(x_i^T\beta)x_i, \qquad \ddot{\theta}_i = \partial_\beta^2\theta_i = \ddot{k}(x_i^T\beta)x_i x_i^T \\
\dot{b}(\theta_i) &= \partial_\theta b(\theta)\Big|_{\theta=\theta_i}, \dot{k}(\eta) = \partial_\eta k(\eta), \ddot{k}(\eta) = \partial_\eta^2(\eta)
\end{aligned}
$$

So

$$
E\Big[-\ddot{l}n(\beta)\Big] = \phi \sum_{i=1}^{n} v_i \dot{\theta}_i^{\otimes 2}
$$

Another set is to use $E(y_i), Var(y_i)$ which is also used commonly as that are the information we generally get. It is used a lot in GEE.

$$
\begin{aligned}
\partial_\mu\theta &= \partial_\theta\mu^{-1}, \qquad \partial_\mu\mu = \partial_\theta\mu\partial_\mu\theta = 1 \\
\partial_\theta\mu &= \partial_\theta b(\theta) = \ddot{b}(\theta) \\
\partial_\mu\theta &= \Big(\partial_\theta\mu\Big)^{-1} = \ddot{b}(\theta)^{-1}
\end{aligned}
$$

Then we have the connection between the two system

$$
\begin{aligned}
\partial_\beta\theta &= \partial_\beta\mu_i\partial_{\mu_i}\theta_i = \partial_\beta\mu_i\Big[\ddot{b}(\theta_i)\Big]^{-1} \\
\partial_\beta^2\theta_i &= \Big(\partial_{\mu_i}^2\theta_i\Big)\Big(\partial_\beta\mu_i\Big)^{\otimes 2} + \partial_{\mu_i}\theta_i\Big(\partial_\beta^2\mu_i\Big) \\
&= -\dddot{b}(\theta_i)\ddot{b}(\theta_i)^{-3}\Big(\partial_\beta\mu_i\Big)^{\otimes 2} + \Big[\ddot{b}(\theta_i)\Big]^{-1}\Big(\partial_\beta^2\mu_i\Big)
\end{aligned}
$$

The generalized estimation model

$$
\begin{aligned}
V(\beta) &= \mathrm{diag}\Big(v_1(\beta), \ldots, v_n(\beta)\Big) \\
e(\beta) &= (y_1 - \mu_1(\beta), \ldots, y_n - \mu_n(\beta))' \\
D_\theta(\beta)' &= \Big(\partial_\beta\beta_1(\beta), \ldots, \partial_\beta\beta_n(\beta)\Big)_{p\times n} \\
D(\beta)^T &= \Big(\partial_\beta\mu_1(\beta), \ldots, \partial_\beta\mu_n(\beta)\Big)_{p\times n} \\
\dot{l}_n(\beta) &= \phi D_\theta(\beta)^T e(\beta) = \phi D(\beta)' V(\beta)^{-1} e(\beta) \\
E\Big[-\ddot{l}_n(\beta)\Big] &= \phi D_\theta(\beta)' V D_\theta(\beta) = \phi D(\beta)' V(\beta)^{-1} D(\beta)
\end{aligned}
$$

Chapter 4

Likelihood Functions

4.1 Hypergeometric Distribution

(i) How to get the likelihood function of drawing objects without replacement?

Definition 4.1.1. Hypergeometric is drawing objects without replacement, while binomial is drawing objects with replacement. So the binomial distribution has the probability p for each drawn, while the hypergeometric probability mass function:

$$p_X(k) = p(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

Here we already know how many K objects totally, and we would like to draw k out of K. The difference of replacement vs. no replacement is the probability. With replacement, we could get $p = \frac{K}{N}$ for each time, while without replacement, we get the probability by using the designed drawn divided by all possible drawns. That's how we derive the concepts of nuisance parameters as well.

Hypergeometric distribution is derived from binomial, $\binom{a}{b}$ is a binomial coefficients, which means two outcomes, yes or no.

Hypergeometric distribution focuses on the range of k, and the sum of all probabilities of p(k) is 1, so we need to know all possible values for k. $0 \leq k \leq min(n, K)$.

4.1.1 Non-central Hypergeometric distribution

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

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

$$p(n_{ij}) = \prod_{i=1}^{2}\prod_{j=1}^{2}\frac{exp(-\mu_{ij})\mu_{ij}^{n_{ij}}}{n_{ij}!}$$

$$= \frac{exp(-\mu_{11})\mu_{11}^{n_{11}}}{n_{11}!}\frac{exp(-\mu_{12})\mu_{12}^{n_{12}}}{n_{12}!}\frac{exp(-\mu_{21})\mu_{21}^{n_{21}}}{n_{21}!}\frac{exp(-\mu_{22})\mu_{22}^{n_{22}}}{n_{22}!}$$

$$n_{12} = n_{1+} - n_{11}, \qquad n_{21} = n_{+1} - n_{11},$$

$$n_{22} = n - n_{12} - n_{21} - n_{11} = n - n_{1+} - n_{+1} + n_{11}$$

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

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 $n_{11} >= -n + n_{1+} + n_{+1}$.

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

$$= \frac{exp(-\sum_{i=1}^{2}\sum_{j=1}^{2}\mu_{ij})\left(\frac{\mu_{11}\mu_{22}}{\mu_{12}\mu_{21}}\right)^{n_{11}}\left(\frac{\mu_{12}}{\mu_{22}}\right)^{n_{1+}}\left(\frac{\mu_{21}}{\mu_{22}}\right)^{n_{+1}}\mu_{22}^{n}}{n_{11}!(n_{1+}-n_{11})!(n_{+1}-n_{11})!(n-n_{1+}-n_{+1}+n_{11})!}$$

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

So we can have

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

$$= \frac{exp(-\sum_{i=1}^{2}\sum_{j=1}^{2}\mu_{ij})\left(\frac{\mu_{11}\mu_{22}}{\mu_{12}\mu_{21}}\right)^{n_{11}}\left(\frac{\mu_{12}}{\mu_{22}}\right)^{n_{1+}}\left(\frac{\mu_{21}}{\mu_{22}}\right)^{n_{+1}}\mu_{22}^{n}}{n_{11}!(n_{1+}-n_{11})!(n_{+1}-n_{11})!(n-n_{1+}-n_{+1}+n_{11})!}$$

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

Which we can rewrite

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

Dervie the hypergeometric distribution:

For a fixed sample size n, the joint distribution of the cell counts in the 2×2 table is given by

$$p = \frac{n!}{n_{11}!n_{12}!n_{21}!n_{22}!}\pi_{11}^{n_{11}}\pi_{12}^{n_{12}}\pi_{21}^{n_{21}}\pi_{22}^{n_{22}}$$
$$\psi = \frac{\pi_{11}\pi_{22}}{\pi_{12}!\pi_{21}!}$$

Let ψ be the parameter of interest, and π_{21}, π_{12} are the nuisance parameters. By looking at the sufficient statistics of π_{12}, π_{21}, which is $n_{12} = n_{1+} - n_{11}, n_{21} = n_{+1} - n_{11}$. We have a distribution of n_{11} which is the parameter of interest.

There are two ways to get the distribution of conditional probability, one is directly use the conditional probability definition, while the other is to use the conditional log-likelihood formula. Which way should we go will depend on the situation.

If it is easier to get the log-likelihood, then go with the log-likelihood function. But for hypergeometric distribution, it is easier to just use definition as the binomial coefficient is not easy to deal with in log form.

Method 2: Use multinomial distribution definition. When we fixed n_{1+}, n_{+1} which is equal to fix n_{12}, n_{21}

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

The marginal distribution of $p(n_{1.}, n_{.1}|n)$

$$p(n_{1.}, n_{.1}|n) = \sum_{N_{11}\in \max(0,-n+n_{1+}+n_{+1})}^{\min(n_{1+},n_{+1})} \frac{n!}{n_{11}!n_{12}!n_{21}!n_{22}!}\psi^{n_{11}}\pi_{12}^{n_{1+}}\pi_{21}^{n_{+1}}\pi_{22}^{n-n_{1+}-n_{+1}}$$

We don't change n_{11} in this formula in order to construct the hypergeometric coefficients in the conditional probability. Most of the terms could be canceled and left n_{11}

$$
\begin{aligned}
p(n_{11}|n_{1.},n_{.1},n) &= \frac{p(n_{11},n_{1.},n_{.1}|n)}{p(n_{1.},n_{.1}|n)} \\
&= \frac{n!}{n_{11}!n_{12}!n_{21}!n_{22}!}\psi^{n_{11}}\pi_{12}^{n_{1+}}\pi_{21}^{n_{+1}}\pi_{22}^{n-n_{1+}-n_{+1}} \\
&\Bigg/ \sum_{x\in\max(0,-n+n_{1+}+n_{+1})}^{\min(n_{1+},n_{+1})} \frac{n!}{n_{11}!n_{12}!n_{21}!n_{22}!}\psi^{x}\pi_{12}^{n_{1+}}\pi_{21}^{n_{+1}}\pi_{22}^{n-n_{1+}-n_{+1}} \\
&= \binom{n_{1.}}{n_{11}}\binom{n-n_{1.}}{n_{.1}-n_{11}}\psi^{n_{11}} \Bigg/ P_0(\psi) \\
P_0(\psi) &= \sum_{x\in\max(0,-n+n_{1+}+n_{+1})}^{\min(n_{1+},n_{+1})} \binom{n_{1.}}{x}\binom{n-n_{1.}}{n_{.1}-x}\psi^{x}
\end{aligned}
$$

The n_{1+}, n_{+1} could be considered as the K from above PDF, that we already know the total row and column, and see what the probability is in each draw.

4.1.2 HG distribution exponential family

Although the hypergeometric distribution looks ugly, the characteristics are the same as other distribution. Here need to be aware that, the ψ is the random variable, while n_{11} is y.

$$
\begin{aligned}
P(n_{11}|n_{1+},n_{+1},n,\psi) &= exp\{n_{11}log\psi - logP_0(\psi) + const\} \\
\theta = log\psi, \qquad \phi &= 1, \qquad b(\theta) = logP_0(\psi) = logP_0(exp(\theta))
\end{aligned}
$$

M(t), K(t)

$$
\begin{aligned}
M(t) = E[exp(ty)] &= \int_{\max(0,-n+n_{1+}+n_{+1})}^{\min(n_{1+},n_{+1})} exp(ty)exp\{ylog\psi - logP_0(\psi) + c\}dy \\
&= \int_{\max(0,-n+n_{1+}+n_{+1})}^{\min(n_{1+},n_{+1})} exp\{y(\theta+t) - logP_0(exp(\theta)) + c\}dy \\
&= \int_{\max(0,-n+n_{1+}+n_{+1})}^{\min(n_{1+},n_{+1})} \\
&\quad exp\{y(\theta+t) - logP_0(exp(\theta+t)) + logP_0(exp(\theta+t)) - logP_0(exp(\theta)) + c\}dy \\
M(t) &= exp\{\phi b(\theta+t/\phi) - b(\theta)\} = exp(b(\theta+t) - b(\theta)) \\
&= P_0(exp(\theta+t))/P_0(exp(\theta)) \\
M(t) &= P_0(exp(t)exp(\theta))/P_0(exp(\theta)) = P_0(exp(t)\psi)/P_0(\psi)
\end{aligned}
$$

The cumulant moment generating function

$$
K(t) = logM(t) = logP_0(exp(t)\psi) - logP_0(\psi)
$$

μ, σ^2

$$\mu = \partial_t K(t) = \left.\frac{P_0(exp(t)\psi)'}{P_0(exp(t)\psi)}\right|_{t=0} = \frac{P_1(\psi)}{P_0(\psi)}$$

$$\sigma^2 = \left.\frac{\partial \partial_t K(t)}{\partial t}\right|_{t=0} = \frac{P_2(\psi)}{P_0(\psi)} - \mu_i^2$$

$$P_j(\psi) = \int_{x \in \max(0, -n+n_{1+}+n_{+1})}^{\min(n_{1+}, n_{+1})} \binom{n_{1+}}{x}\binom{n-n_{1+}}{n_{+1}-x}\psi^x x^j$$

4.2 Multinomial distribution

Get the covariance matrix for cross-sectional, prospective, retrospective sampling method.

4.2.1 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:

$$p(\theta) = \prod_{i=1}^{n}\prod_{j=1}^{J} \pi_j^{I(Y_i=j)}, \qquad \theta = (\pi_1, \pi_2, ...\pi_J)'$$

$$lnp(\theta) = \sum_{i=1}^{n}\sum_{j=1}^{J} I(Y_i = j)log(\pi_j) = \sum_{j=1}^{J} n_j log(\pi_j)$$

$$\begin{aligned}
M_X(t) &= E[exp(t^T X)] = E[exp(t^T(Y_1 + Y_2 + ...Y_n))] = E[exp(t^T Y_1 + t^T Y_2 + ...t^T Y_n)] \\
&= E[\prod_{i=1}^{n} exp(t^T Y_i)] \\
&= \prod_{i=1}^{n} E[exp(t^T Y_i)] \qquad \text{(by independence)} \\
&= \prod_{i=1}^{n} M_{Y_i}(t) = \prod_{i=1}^{n} P(Y_i = 1)e^{ty_i} \qquad \text{by MGF of discrete variable } Y_i \\
&= \left(\sum_{j=1}^{J} \pi_j exp(t_j)\right)^n \qquad \text{by MGF of multinoulli}
\end{aligned}$$

The MGF for bernoulli distribution

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

For multinomial distribution

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

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?

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

4.2.2 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 $\hat{\pi} = \frac{n_i}{n}$.

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

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.

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

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 $n \to \infty$.

4.2.3 Odds ratio

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

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

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 $\hat{R}$.

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

4.3 Contingency Table

We can use either multinomial distribution, poisson distribution to model the contingency table.

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.

$n_{ij} \sim Poisson(\mu_{ij})$, n_{ij} are independent, $i = 1, ..I, j = 1, ..J$.

By poisson distribution of each cell counts, $n = \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij} \sim Poisson(\sum_{i=1}^{I}\sum_{j=1}^{J} \mu_{ij})$

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

Then the likelihood function

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

The joint distribution is Multinomial $(n, \pi_{11}, \pi_{12}, ..\pi_{IJ})$, where

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

$n_{i.} = \sum_{j=1}^{I} n_{ij} \sim Poisson(\sum_{j=1}^{J} \mu_{ij}), i = 1, ..I$. The conditional distribution will be built based on fixed row margins (the denominator will be the supposed known).

$$\begin{aligned}
n_{i+} &= \sum_{j=1}^{J} n_{ij} \\
n_{i+} &\sim Poisson(\sum_{j=1}^{J} \mu_{ij}) \\
p(n_{11}, ..n_{ij}|n_{i+}) &= \frac{P(n_{11})P(n_{12})..P(n_{IJ})}{P(n_{1.})P(n_{2.})\ldots P(n_{I.})} \\
&= \prod_{i=1}^{I}\prod_{j=1}^{J} \frac{\exp(-\mu_{ij}){\mu_{ij}}^{n_{ij}}}{n_{ij}!} \Bigg/ \prod_{i=1}^{I} \frac{\exp(-\mu_i)\mu_i^{n_{i+}}}{n_{i+}!} \\
&= \prod_{i=1}^{I} \binom{n_{i+}}{n_{ij}} \prod_{i=1}^{I}\prod_{j=1}^{J} \left(\frac{\mu_{ij}}{\sum_{j=1}^{J} \mu_{ij}}\right)^{n_{ij}}
\end{aligned}$$

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

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

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.

(d) Conditional distribution of n_{11}

Need to see that the variable transformation in this case. One skill I need to develop is to construct the probability distribution or likelihood function for each scenario.

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 $n_{1+} = n_{11} + n_{12}, n_{+1} = n_{11} + n_{21}$.

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

The Jacobian transformation matrix

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

$$\|J\| = 1$$

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

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

Thus $n_{11}|n_{1,}, n_{.1}, n$ is a non-central hypergeometric distribution.

(e) Let π_{ij} denote the cell probability for the ith row and jth column of the table and assume that n is fixed. Consider testing the hypothesis $H_0 : \pi_{ij} = \pi_{i.}\pi_{.j}, i = 1, .., I, j = 1, .., J$. Derive the MLE's of π_{ij} under H_0.

We need to derive the likelihood function of cell probability. Under the H_0, get the MLE of $\pi_{ij} = \pi_{i.}\pi_{.j}$. We could use the likelihood function to get the $\pi_{i.}, \pi_{.j}$ MLE estimates, then by the invariance of MLE, $\hat{p_{ij}} = \hat{\pi_{i.}}\hat{\pi_{.j}}$.

From part (a) we have the likelihood function of π_{ij}

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

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

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

Under H_0, the log-likelihood

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

By Lagrangian multiplier theorem,

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

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

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

Using Lagrange multiplier λ, the log-likelihood of $\pi_{1.}, .., \pi_{I.}$ is

$$
\begin{aligned}
l(\pi_{1.}, ..\pi_{I.}|n) &= \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij} \log\left(\pi_{i.}\pi_{.j}\right) + \lambda\Big(1 - \sum_{i=1}^{I} \pi_{i.}\Big) \\
\dot{l}(\pi_{i.}) &= \frac{\sum_{j=1}^{J} n_{ij}}{\pi_{i.}} - \lambda \underset{set}{=} 0 \\
\hat{\pi_{i.}} &= \frac{\sum_{j=1}^{J} n_{ij}}{\lambda}, \qquad \lambda = \sum_{i=1}^{I}\sum_{j=1}^{J} n_{ij} = n \\
\hat{\pi}_{i.} &= \frac{n_{i+}}{n}
\end{aligned}
$$

(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.

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

The degrees of freedom refers to the number of random variables that used to estimate. When the mean of row or column are calculated, the mean will be used to center the data, then the free (random) data would be less 1. Here both the row and column will be less 1 when under H_0.

4.4 Sampling Method

cross-sectional, prospective, retrospective sampling method.

4.4.1 Retrospective

In retrospective study, we know the response and further select patients accordingly, in other words, the columns margin is fixed. So we know the percentage of X in the Y categories.

For retrospective study, the Y is fixed. The likelihood function is constructed as below:

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

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

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

By CLT,

$$\sqrt{n}\left(\begin{pmatrix}\theta\\ \gamma\end{pmatrix} - \begin{pmatrix}\hat{\theta}\\ \hat{\gamma}\end{pmatrix}\right) \xrightarrow{d} N(0,\Sigma)$$

The covariance matrix, binomial distribution variance is $np(1-p)$

$$
\begin{aligned}
\sqrt{n}\left(\theta - \hat{\theta}\right) &\xrightarrow{d} N(0,\Sigma) \\
\Sigma &= \begin{bmatrix}\theta(1-\theta) & 0 \\ 0 & \gamma(1-\gamma)\end{bmatrix}
\end{aligned}
$$

The Fisher Information matrix is inverse of Covariance matrix.

$$
\begin{aligned}
I(\theta,\gamma) &= \Sigma^{-1} \\
&= \begin{bmatrix}\theta^{-1}(1-\theta)^{-1} & 0 \\ 0 & \gamma^{-1}(1-\gamma)^{-1}\end{bmatrix}
\end{aligned}
$$

Then get covariance matrix by delta method for odds ratio,

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

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

In prospective study, the row margin is fixed. We will need to figure out how many parameters are needed to construct the likelihood function.

We have

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

So the likelihood function

$$
P(\theta, \gamma) = \theta^{n_{11}}(1-\theta)^{n_{12}}\gamma^{n_{21}}(1-\gamma)^{n_{22}}
$$

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.

4.4.4 Odds ratio

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

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

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 $\hat{R}$.

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

4.5 Conditional Probability

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 $p(n_{11}, n_{12} | n_{13}, ...n_{IJ}, n)$.

4.5.1 Contingency table

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

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

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 $(1, \pi_{00}, \pi_{01}, \pi_{10}, \pi_{11})$.

$$\begin{aligned}
Z_1 &= I[(X,Y) = (0,0)] \\
Z_2 &= I[(X,Y) = (0,1)] \\
Z_3 &= I[(X,Y) = (1,0)] \\
Z_4 &= I[(X,Y) = (1,1)] \\
p(\theta) &= \prod_k \pi_k^{I(Z_k=1)} \\
M_Z(t) &= E[exp(t^T Z)] = E[exp(t^T(Z_1+Z_2+...Z_n))] = E[exp(t^T Z_1 + t^T Z_2 + ...t^T Z_n)] \\
&= E[\prod_{i=1}^{n} exp(t^T Z_i)] \\
&= \prod_{i=1}^{n} E[exp(t^T Z_i)] \qquad \text{(by independence)} \\
&= \prod_{i=1}^{n} M_{Z_i}(t) = \prod_{i=1}^{n} P(Z_i = 1)e^{tz_i} \qquad \text{by MGF of discrete variable } Z_i \\
&= \left(\sum_{j=1}^{J} \pi_j exp(t_j)\right)^n \qquad \text{by MGF of multinoulli}
\end{aligned}$$

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

$$\begin{aligned}
E(Z_1Z_2) &= \frac{\partial^2 M_Z(t)}{\partial Z_i \partial Z_j}|_{t_i=t_j=0} \\
&= \frac{\partial \left(n(\pi_i e^{t_i})(\sum_{k=1}^K \pi_k e^{t_k})^{n-1}\right)'}{\partial t_j} \\
&= n(n-1)(\sum_{k=1}^K \pi_k e^{t_k})^{n-2}\pi_i\pi_j|_{t_i=t_j=0} = n(n-1)\pi_i\pi_j \\
E(X_i) &= n\pi_i \\
Cov(Z_i, Z_j) &= E(Z_iZ_2) - E(Z_1)E(Z_j) = n(n-1)\pi_i\pi_j - n^2\pi_i\pi_j = -n\pi_i\pi_j \\
Var(Z_i) &= E(Z_i^2) - E(Z_i)^2 \\
E(Z_i^2) &= \frac{\partial \left(n(\pi_i e^{t_i})(\sum_{k=1}^K \pi_k e^{t_k})^{n-1}\right)'}{\partial t_i} \\
&= n(\sum_{k=1}^K \pi_k e^{t_k})^{n-1}\pi_i e^{t_i} + n(n-1)(\sum_{k=1}^K \pi_k e^{t_k})^{n-2}\pi_i\pi_i e^{2t_i}|_{t_i=0} \\
&= n\pi_i + n(n-1)\pi_i^2 = n\pi_i(1-\pi) \\
Var(Z_i/n) &= \frac{1}{n^2}Var(Z_i) = \frac{1}{n}\pi_i(1-\pi_i)
\end{aligned}$$

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,

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

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.

$$\begin{gathered}
f = exp(g) = R, \qquad f(g)' = R \\
\sqrt{n}(f(\hat{g}) - f(g)) \xrightarrow{d} N\left(0, f(g)'(\frac{1}{\pi_{11}} + \frac{1}{\pi_{12}} + \frac{1}{\pi_{21}} + \frac{1}{\pi_{22}})f(g)'^T\right) \\
\sqrt{n}(\hat{R} - R) \xrightarrow{d} N\left(0, R^2(\frac{1}{\pi_{11}} + \frac{1}{\pi_{12}} + \frac{1}{\pi_{21}} + \frac{1}{\pi_{22}})\right) \\
(\hat{R} - R) \xrightarrow{d} N\left(0, \frac{1}{n}R^2(\frac{1}{\pi_{11}} + \frac{1}{\pi_{12}} + \frac{1}{\pi_{21}} + \frac{1}{\pi_{22}})\right)
\end{gathered}$$

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 $(\hat{\alpha}; \hat{\beta})$.

$$\begin{aligned}
\pi_{01} + \pi_{11} &= \frac{exp(\alpha)}{1+\exp(\alpha)} \\
exp(\alpha) &= \frac{\pi_{10}+\pi_{11}}{\pi_{01}+\pi_{00}}, \qquad \alpha = log\left(\frac{\pi_{10}+\pi_{11}}{\pi_{01}+\pi_{00}}\right) \\
\pi_{10} + \pi_{11} &= \frac{exp(\alpha+\beta)}{1+\exp(\alpha+\beta)} \\
\alpha+\beta &= log\left(\frac{\pi_{01}+\pi_{11}}{\pi_{10}+\pi_{00}}\right) \\
\beta &= log\left(\frac{\pi_{01}+\pi_{11}}{\pi_{10}+\pi_{00}}\right) - log\frac{\pi_{10}+\pi_{11}}{\pi_{01}+\pi_{00}}, \qquad \beta \ = log\left(\frac{(\pi_{01}+\pi_{11})(\pi_{01}+\pi_{00})}{(\pi_{10}+\pi_{00})(\pi_{10}+\pi_{11})}\right)
\end{aligned}$$

By invariance of MLE,

$$\begin{aligned}
\hat{\alpha} &= log\left(\frac{\hat{\pi_{10}}+\hat{\pi_{11}}}{\hat{\pi_{01}}+\hat{\pi_{00}}}\right) = log\left(\frac{n_{10}+n_{11}}{n_{01}+n_{00}}\right) \\
\hat{\beta} &= log\left(\frac{(\hat{\pi}_{01}+\hat{\pi}_{11})(\hat{\pi}_{01}+\hat{\pi}_{00})}{(\hat{\pi}_{10}+\hat{\pi}_{00})(\hat{\pi}_{10}+\hat{\pi}_{11})}\right) = log\left(\frac{(n_{01}+n_{11})(n_{01}+n_{00})}{(n_{10}+n_{00})(n_{10}+n_{11})}\right)
\end{aligned}$$

(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 $(\pi_{1+}\pi_{0+})^{-1} + (\pi_{+1}\pi_{+0})^{-1} \leq (\pi_{1+}\pi_{+0})^{-1} + (\pi_{+1}\pi_{0+})^{-1}$.

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

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

Bibliography

[1] Armstrong, M.A. *Basic Topology.* England: Editorial Board, 2000.

[2] Coxeter, H.S.M. *Introduction to Geometry.* Toronto: John Wiley and Sons Inc, 1969.

[3] Chronister, James. *Blender Basics.* Estados Unidos: 2004.

[4] Engelking, Ryszard. *General Topology.* Berlin: Hildermann, 1989.

[5] Engelking, Ryszard y Karol Sieklucki. *Topology: A Geometric Approach.* Berlin: Hildermann, 1992.

[6] Hatcher, Allen. *Algebraic Topology.* Cambridge: Cambridge University Press, 2002.

[7] Hoffman, Kenneth y Ray Kunze. *Linear Algebra, second edition.* New Jersey: Prentice Hall. Inc, 1961.

[8] Kackzynski, T., K. Mischainkow y M. Mrozen. *Algebraic Topology: A Computational Approach.* 2000.

[9] Kosniowski, Czes. *Topología Algebraica.* Barcelona: Editorial Revert, 1986.

[10] Lefschetz, Solomon. *Algebraic Topology.* Rhode Island: American Mathematical Society, 1991.

[11] May, J. P. *A Concise Course in Algebraic Topology.* Chicago.

[12] Mill, J. Van. *Infinite-Dimensional Topology.* Amsterdam: Elsevier Science Publications B. V, 1989.

[13] Munkres, James R. *Topology.* New York: Prentice Hall Inc, 2000.

[14] Roseman, Dennis. *Elementary Topology.* New Jersey: Prentice Hall Inc, 1999.

[15] Sato, Hajime. *Algebraic Topology: An Intuituve Approach.* Rhode Island: American Mathematical Society, 1996.

[16] Zomorodian, Afra J. *Topology for Computing.* United States: Cambridge University Press, 2005.