STOCKHOLM UNIVERSITY
DEPT OF MATHEMATICS
Div. of Mathematical statistics
Ola Hössjer

MT 5006
SOLUTIONS
February 3, 2021

Solutions for Examination Categorical Data Analysis, February 3, 2021

Problem 1

a. Under binomial rows sampling we have that

$$\begin{array}{rcl} N_{00} & \sim & \text{Bin}(7, \pi_0), \\ N_{10} & \sim & \text{Bin}(8, \pi_1) \end{array}$$

are independent and binomially distributed. Therefore the likelihood $l(\pi_0, \pi_1)$ is given by the joint distribution of N_{01} and N_{11}, i.e.

$$\begin{array}{rcl} l(\pi_0, \pi_1) & = & P(N_{00} = 5, N_{10} = 2) \\ & = & \binom{7}{5}\pi_0^5(1-\pi_0)^2 \cdot \binom{8}{2}\pi_1^2(1-\pi_1)^6 \\ & = & 588 \cdot \pi_0^5(1-\pi_0)^2\pi_1^2(1-\pi_1)^6. \end{array} \tag{1}$$

b. The null hypothesis and the alternative hypothesis correspond to

$$\begin{array}{rl} H_0: & \pi_0 = \pi_1, \\ H_a: & \pi_0 > \pi_1, \end{array} \tag{2}$$

respectively. Introducing the odds ratio

$$\theta = \frac{\pi_0/(1-\pi_0)}{\pi_1/(1-\pi_1)}, \tag{3}$$

we find that (2) is equivalent to

$$\begin{array}{rl} H_0: & \theta = 1, \\ H_a: & \theta > 1. \end{array}$$

c. Let n_{ij} be the observed cell counts. If we condition on the two row sums n_{i+} and the two column sums n_{+j}, then N_{00} has a hypergeometric distribution under H_0, i.e.

$$\begin{array}{l} P_{H_0}(N_{00} = k | N_{0+} = 7, N_{1+} = 8, N_{+0} = 7, N_{+1} = 8) \\ \quad = \binom{7}{k}\binom{8}{7-k}/\binom{15}{7} \end{array} \tag{4}$$

for $0 \leq k \leq 7$.

The null hypothesis is rejected for large values of N_{00}, since then it is more likely that H_a holds. Denote the conditioning above by three dots $(\ldots)$. Since $n_{00} = 5$, we find that

$$\begin{aligned}
P\text{-value} &= P_{H_0}(N_{11} = 5|\ldots) + P_{H_0}(N_{00} = 6|\ldots) + P_{H_0}(N_{00} = 7|\ldots) \\
&= \binom{7}{5}\binom{8}{2}/\binom{15}{7} + \binom{7}{6}\binom{8}{1}/\binom{15}{7} + \binom{7}{7}\binom{8}{0}/\binom{15}{7} \\
&= (21 \cdot 28 + 7 \cdot 8 + 1 \cdot 1)/6435 \\
&= 645/6435 \\
&= 0.100.
\end{aligned}$$

Hence we cannot reject the null hypothesis, that the lady guesses at random, at level 5%.

d. Starting with the joint distribution of N_{00} and N_{10}, as in (1), we condition on the columns sums as well. Since we already condition on row sums in (1), and since $N_{+1} = 15 - N_{+0}$, we only need to write out N_{+0} in the conditioning. This gives

$$\begin{aligned}
P(N_{00} = k|N_{+0} = 7) &= P(N_{00} = k, N_{10} = 7-k)/P(N_{+0} = 7) \\
&\propto P(N_{00} = k, N_{10} = 7-k) \\
&= \binom{7}{k}\pi_0^k(1-\pi_0)^{7-k} \cdot \binom{8}{7-k}\pi_1^{7-k}(1-\pi_1)^{8-(7-k)} \\
&\propto \binom{7}{k}\binom{8}{7-k}\theta^k,
\end{aligned} \tag{5}$$

for $k = 0, 1, \ldots, 7$, where the odds ratio (3) was used in the fourth step. Expressions to the right and left of a proportionality sign $\propto$ in (5) differ by a multiplicative constant, not depending on k. The proportionality constant of the last step is chosen so that all probabilities sum to one. This gives a non-central hypergeometric distribution

$$P(N_{00} = k|N_{+0} = 7) = \frac{\binom{7}{k}\binom{8}{7-k}\theta^k}{\sum_{l=0}^{7}\binom{7}{l}\binom{8}{7-l}\theta^l},$$

for $0 \le k \le 7$. The special case $\theta = 1$ is identical to the hyptergeometric distribution (4).

Problem 2

a. Because of independent binomial rows sampling, the log likelihood of the dataset is

$$\begin{aligned}
L(\pi, \Delta) &= \log\binom{n_{0+}}{n_{01}} + n_{00}\log(1-\pi-\Delta) + n_{01}\log(\pi+\Delta) \\
&+ \log\binom{n_{1+}}{n_{11}} + n_{10}\log(1-\pi) + n_{11}\log(\pi),
\end{aligned} \tag{6}$$

with $n_{00} = 2350$, $n_{01} = 42$, $n_{10} = 2417$, and $n_{11} = 53$.

b. Inserting the numbers of the table into the definitions of $\hat{\Delta}$ and $\hat{\pi}$, we find that

$$\begin{aligned}
\hat{\Delta} &= 42/2392 - 53/2470 = 0.00390, \\
\hat{\pi} &= 95/4862 = 0.0195.
\end{aligned}$$

This gives a score statistic

$$z_S = \frac{0.00390}{\sqrt{(\frac{1}{2392} + \frac{1}{2470}) \cdot 0.0195(1 - 0.0195)}} = -0.983.$$

Since $z_S > -1.645$ we conclude that H_0 cannot be rejected at signficance level 5%.

c. By differentiating (6) with respect to π and Δ, we find that

$$\begin{aligned} u_\pi(\pi, \Delta) &= n_{01}/(\pi + \Delta) - n_{00}/(1 - \pi - \Delta) + n_{11}/\pi - n_{10}/(1 - \pi), \\ u_\Delta(\pi, \Delta) &= n_{01}/(\pi + \Delta) - n_{00}/(1 - \pi - \Delta). \end{aligned} \tag{7}$$

d. We start by finding the elements of the Hessian matrix $\boldsymbol{H}(\pi, \Delta)$. That is, we differentiate (7) with respect to π and Δ, and obtain

$$\begin{aligned} H_{\pi\pi}(\pi, \Delta) &= \partial u_\pi(\pi, \Delta)/\partial\pi \\ &= -n_{01}/(\pi + \Delta)^2 - n_{00}/(1 - \pi - \Delta)^2 \\ &\quad -n_{11}/\pi^2 - n_{10}/(1 - \pi)^2, \\ H_{\pi\Delta}(\pi, \Delta) &= \partial u_\pi(\pi, \Delta)/\partial\Delta \\ &= -n_{01}/(\pi + \Delta)^2 - n_{00}/(1 - \pi - \Delta)^2, \\ H_{\Delta\Delta}(\pi, \Delta) &= \partial u_\Delta(\pi, \Delta)/\partial\Delta \\ &= -n_{01}/(\pi + \Delta)^2 - n_{00}/(1 - \pi - \Delta)^2. \end{aligned} \tag{8}$$

Since the rows of the table have independent binomial distributions, it follows that the expected cell counts are $E(N_{00}) = n_{0+}(1 - \pi - \Delta)$, $E(N_{01}) = n_{0+}(\pi + \Delta)$, $E(N_{10}) = n_{1+}(1 - \pi)$, and $E(N_{11}) = n_{1+}\pi$. Inserting these expectations into (8), and changing sign, we find that the elements of the Fisher information matrix are given by

$$\begin{aligned} J_{\pi\pi}(\pi, \Delta) &= -E[H_{\pi\pi}(\pi, \Delta)] = n_{0+}/[(\pi + \Delta)(1 - \pi - \Delta)] + n_{1+}/[\pi(1 - \pi)], \\ J_{\pi\Delta}(\pi, \Delta) &= -E[H_{\pi\Delta}(\pi, \Delta)] = n_{0+}/[(\pi + \Delta)(1 - \pi - \Delta)], \\ J_{\Delta\Delta}(\pi, \Delta) &= -E[H_{\Delta\Delta}(\pi, \Delta)] = n_{0+}/[(\pi + \Delta)(1 - \pi - \Delta)]. \end{aligned} \tag{9}$$

e. It is convenient to introduce $\hat{\pi}_0 = n_{01}/n_{0+}$ and $\hat{\pi} = \hat{\pi}(0) = n_{+1}/n$. In view of (7), the numerator of the score test is

$$\begin{aligned} u(\hat{\pi}, 0) &= n_{01}/\hat{\pi} - n_{00}/(1 - \hat{\pi}) \\ &= n_{0+}[\hat{\pi}_0/\hat{\pi} - (1 - \hat{\pi}_0)/(1 - \hat{\pi})] \\ &= n_{0+}(\hat{\pi}_0 - \hat{\pi})/[\hat{\pi}(1 - \hat{\pi})] \\ &= n_{0+}n_{1+}\hat{\Delta}/[n\hat{\pi}(1 - \hat{\pi})], \end{aligned} \tag{10}$$

where in the last step we used that $\hat{\pi}_0 - \hat{\pi} = (n_{1+}/n)\hat{\Delta}$. On the other hand, making use of (9) and the hint, we find that

$$\begin{aligned} \text{Var}[u(\hat{\pi}, 0)] &= J_{\Delta\Delta}(\hat{\pi}, 0) - J_{\pi\Delta}(\hat{\pi}, 0)^2/J_{\pi\pi}(\hat{\pi}, 0) \\ &= n_{0+}/[\hat{\pi}(1 - \hat{\pi})] - \{n_{0+}/[\hat{\pi}(1 - \hat{\pi})]\}^2/\{n/[\hat{\pi}(1 - \hat{\pi})]\} \\ &= n_{0+}n_{1+}/[n\hat{\pi}(1 - \hat{\pi})]. \end{aligned} \tag{11}$$

Finally, by taking the ratio of (10) and the square root of (11), we obtain the sought for expression

$$z_S = \frac{\hat{\Delta}}{\sqrt{\frac{n}{n_{0+}n_{1+}}\hat{\pi}(1-\hat{\pi})}} = \frac{\hat{\Delta}}{\sqrt{(\frac{1}{n_{0+}}+\frac{1}{n_{1+}})\hat{\pi}(1-\hat{\pi})}}$$

of the score statistic.

Problem 3

a. The loglinear parametrization of (XZ, YZ) is

$$\mu_{ijk} = \exp(\lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}) \tag{12}$$

for $0 \le i, j, k \le 1$. Assume that $X = 0$, $Y = 0$ and $Z = 0$ are chosen as baseline levels. Then those loglinear parameters are put to zero for which at least one index i, j or k equals 0. The remaining parameters are

$$\boldsymbol{\beta} = (\lambda, \lambda_1^X, \lambda_1^Y, \lambda_1^Z, \lambda_{11}^{XZ}, \lambda_{11}^{YZ}). \tag{13}$$

b. It follows from (12) that

$$\mu_{ijk} = A_k B_{ik} C_{jk}, \tag{14}$$

with $A_k = \exp(\lambda + \lambda_k^Z)$, $B_{ik} = \exp(\lambda_i^X + \lambda_{ik}^{XZ})$ and $C_{jk} = \exp(\lambda_j^Y + \lambda_{jk}^{YZ})$. Then, summing over one of i or j, or over both indeces simultaneously in (14), we find that

$$\begin{aligned} \mu_{i+k} &= A_k B_{ik} C_{+k}, \\ \mu_{+jk} &= A_k B_{+k} C_{jk}, \\ \mu_{++k} &= A_k B_{+k} C_{+k}. \end{aligned}$$

Consequently,

$$\frac{\mu_{i+k}\mu_{+jk}}{\mu_{++k}} = \frac{A_k B_{ik} C_{+k} \cdot A_k B_{+k} C_{jk}}{A_k B_{+k} C_{+k}} = A_k B_{ik} C_{jk} = \mu_{ijk}.$$

Alternatively, we may work directly with the cell probabilities $\pi_{ijk} = \mu_{ijk}/\mu_{+++}$. Since X and Y are conditionally independent given Z for model (XZ, YZ), it follows that

$$\pi_{ijk} = \pi_{++k}\pi_{ij|k} = \pi_{++k}\pi_{i+|k}\pi_{+j|k} = \pi_{++k} \cdot \frac{\pi_{i+k}}{\pi_{++k}} \cdot \frac{\pi_{+jk}}{\pi_{++k}} = \frac{\pi_{i+k}\pi_{+jk}}{\pi_{++k}},$$

and hence

$$\mu_{ijk} = \mu_{+++}\pi_{ijk} = \mu_{+++} \cdot \frac{\frac{\mu_{i+k}}{\mu_{+++}} \cdot \frac{\mu_{+jk}}{\mu_{+++}}}{\frac{\mu_{++k}}{\mu_{+++}}} = \frac{\mu_{i+k}\mu_{+jk}}{\mu_{++k}}.$$

c. The maximum likelihood estimates

$$\hat{\mu}_{ijk} = \frac{n_{i+k}n_{+jk}}{n_{++k}}$$

of the expected cell counts are obtained by replacing μ_{i+k}, μ_{+jk} and μ_{++k} by estimates n_{i+k}, n_{+jk} and n_{++k} respecitvely. From the given marginals of the two partial tables we can read off all n_{i+k}, n_{+jk} and n_{++k}. Applying this for $i = j = k = 1$, we find that

$$\hat{\mu}_{111} = \frac{n_{1+1}n_{+11}}{n_{++1}} = \frac{49 \cdot 51}{98} = 25.5,$$

which agrees with the value in cell $(i, j, k) = (1, 1, 1)$, in the rightmost partial table of Appendix B.

d. The chisquare goodness-of-fit statistic for testing (XZ, YZ), against the saturated model (XYZ), is

$$\begin{aligned} X^2 &= \sum_{ijk}(n_{ijk} - \hat{\mu}_{ijk})^2/\hat{\mu}_{ijk} \\ &= (841 - 838.2)^2/838.2 + \ldots + (29 - 25.5)^2/25.5 \\ &= 9.36 \\ &> \chi^2_2(0.05) = 5.99, \end{aligned}$$

where in the last step we used that df $= 8 - 6 = 2$, since the saturated model has $2 \times 2 \times 2 = 8$ parameters, whereas the conditional independence model (XZ, YZ) has 6 parameters according to (13). Therefore we reject conditional independence between X and Y given Z at level 5%. This suggests there might be other common risk factors for mothers and children.

e. From the two partial tables we obtain the following estimated conditional odds ratios:

$$\begin{aligned} \hat{\theta}_{XY(0)} &= (841 \cdot 4)/(27 \cdot 30) = 4.153, \\ \hat{\theta}_{XY(1)} &= (27 \cdot 29)/(22 \cdot 20) = 1.779. \end{aligned}$$

Since $\hat{\theta}_{XY(0)}$ and $\hat{\theta}_{XY(1)}$ are both larger than 1, this indicates other possible common (genetic or shared environmental) risk factors, whereas model (XY, YZ) has $\theta^{XY}_{(0)} = \theta^{XY}_{(1)} = 1$. Since $\hat{\theta}_{XY(0)}$ is larger than $\hat{\theta}_{XY(1)}$, this indicates that there is no homogeneous association $\theta^{XY}_{(0)} = \theta^{XY}_{(1)}$ between X and Y given Z, as for model (XZ, YZ). It rather indicates that there is not only a second order association term between X and Y, but also a third order association term between X, Y and Z.

Problem 4

a. The loglinear parametrization for (XY, XZ, YZ) requires addition of an XY-interaction term compared to (12). This gives

$$\mu_{ijk} = \exp(\lambda + \lambda^X_i + \lambda^Y_j + \lambda^Z_k + \lambda^{XY}_{ij} + \lambda^{XZ}_{ik} + \lambda^{YZ}_{jk}). \tag{15}$$

b. Let $\pi_{ijk} = \mu_{ijk}/\mu_{+++} = P(X = i, Y = j, Z = k)$ be the cell probabilities, and $\pi_{i+k} = P(X = i, Z = k)$ the corresponding marignal probability for X and Z..

Using (15) we find that

$$\begin{aligned}
\text{logit}[P(Y=1|X=i,Z=k)] &= \log[P(Y=1|X=i,Z=k)/P(Y=0|X=i,Z=k)] \\
&= \log[(\pi_{i1k}/\pi_{i+k})/(\pi_{i0k}/\pi_{i+k})] \\
&= \log(\pi_{i1k}/\pi_{i0k}) \\
&= \log(\mu_{i1k}/\mu_{i0k}) \\
&= (\lambda + \lambda_i^X + \lambda_1^Y + \lambda_k^Z + \lambda_{i1}^{XY} + \lambda_{ik}^{XZ} + \lambda_{1k}^{YZ}) \\
&- (\lambda + \lambda_i^X + \lambda_0^Y + \lambda_k^Z + \lambda_{i0}^{XY} + \lambda_{ik}^{XZ} + \lambda_{0k}^{YZ}) \\
&= \alpha + \beta_i^X + \beta_k^Z,
\end{aligned}$$

where in the last step we used that

$$\begin{aligned}
\alpha &= \lambda_1^Y - \lambda_0^Y, \\
\beta_i^X &= \lambda_{i1}^{XY} - \lambda_{i0}^{XY}, \\
\beta_k^Z &= \lambda_{1k}^{YZ} - \lambda_{0k}^{YZ}.
\end{aligned}$$

If $X = 0$ and $Z = 0$ are chosen as baseline levels, then any loglinear parameter with $i = 0$ or $k = 0$ among it indeces is zero, which implies $\beta_0^X = \beta_0^Z = 0$. The only remaining parameters are $(\alpha, \beta_1^X, \beta_1^Z)$.

c. Since there is no third order interaction XYZ in the model, the conditional odds ratio between X and Y does not depend on the level k of the conditioning variable Z. We find that

$$\begin{aligned}
\log(\theta_{XY(k)}) &= \text{logit}[P(Y=1|X=1,Z=k)] - \text{logit}[P(Y=1|X=0,Z=k)] \\
&= \alpha + \beta_1^X + \beta_k^Z - (\alpha + \beta_0^X + \beta_k^Z) \\
&= \beta_1^X - \beta_0^X \\
&= \beta_1^X.
\end{aligned}$$

A Wald type approximate 95% confidence interval for $\log(\theta_{XY(k)})$ is

$$\begin{aligned}
& (\hat{\beta}_1^X - 1.96\sqrt{\widehat{\text{Var}}(\hat{\beta}_1^X)}, \hat{\beta}_1^X + 1.96\sqrt{\widehat{\text{Var}}(\hat{\beta}_1^X)}) \\
=\ & (0.8347 - 1.96\sqrt{0.1255}, 0.8347 + 1.96\sqrt{0.1255}) \\
=\ & (0.1404, 1.5290),
\end{aligned}$$

and the one for $\theta_{XY(k)}$ is

$$I = (\exp(0.1404), \exp(1.5290)) = (1.15, 4.61).$$

Since $1 \notin I$, this indicates (weakly) that there are additional common risk factors for the mother and child apart from Z. (Notice however, from the solution of Problem 3e), that a third order interaction between X, Y, and Z should possibly be included in the model as well.)

d. Let $\pi(i,k) = P(Y=1|X=i,Z=k)$. Our goal is to find a confidence interval for $\pi(0,1)$. We will apply the delta method, based on the logit transformation. Recall from Problem 4b) that

$$\text{logit}[\pi(0,1)] = \text{logit}[P(Y=1|X=0,Z=1)] = \alpha + \beta_1^Z.$$

In order to find a confidence interval for logit$[\pi(0,1)]$, we notice that the standard error is

$$\begin{aligned}
\text{SE} &= \sqrt{\widehat{\text{Var}}(\hat{\alpha}+\hat{\beta}_1^Z)} \\
&= \sqrt{\widehat{\text{Var}}(\hat{\alpha})+2\widehat{\text{Cov}}(\hat{\alpha},\hat{\beta}_1^Z)+\widehat{\text{Var}}(\hat{\beta}_1^Z)} \\
&= \sqrt{0.0342-2\cdot 0.0295+0.0977} \\
&= \sqrt{0.0792} \\
&= 0.27.
\end{aligned}$$

This gives a Wald type 95% confidence interval

$$\begin{aligned}
&(\hat{\alpha}+\hat{\beta}_1^Z-1.96\cdot\text{SE}, \hat{\alpha}+\hat{\beta}_1^Z+1.96\cdot\text{SE}) \\
&= (-3.3818+3.0497-1.96\cdot 0.27, -3.3818+3.0497+1.96\cdot 0.27) \\
&(-0.861, 0.197)
\end{aligned}$$

for logit$[\pi(0,1)]$, which we transform to a confidence interval

$$\left(\frac{\exp(-0.861)}{1+\exp(-0.861)}, \frac{\exp(0.197)}{1+\exp(0.197)}\right) = (0.297, 0.549)$$

for $\pi(0,1)$.

Problem 5

a. By differentiating the density/probability function formula for the exponential dispersion family (EDF) twice with respect to the natural parameter θ_i of observation i, we find that the score function and Hessian of this observation are

$$u_i(y) = \frac{\partial \log f(y;\theta,\omega_i,\phi)}{\partial\theta_i} = \frac{\omega_i(y-b'(\theta_i))}{\phi} \tag{16}$$

and

$$H_i(y) = \frac{u_i(y)}{\partial\theta_i} = -\frac{\omega_i b''(\theta_i)}{\phi} \tag{17}$$

respectively.

b. Since $nY_i \sim \text{Bin}(n_i,\pi_i)$, we have that

$$\begin{aligned}
P(Y_i=y) &= P(n_iY_i=n_iy) \\
&= \binom{n_i}{n_iy}\pi^{n_iy}(1-\pi)^{n_i-n_iy} \\
&= \binom{n_i}{n_iy}[(\pi_i/(1-\pi_i)]^{n_iy}(1-\pi_i)^{n_i} \\
&= \exp\left\{[y\log(\pi_i/1-\pi_i)-\log(1-\pi)^{-1}]/(1/n_i)+\log(\tbinom{n_i}{n_iy})\right\}.
\end{aligned}$$

This corresponds to an EDF with

$$\begin{aligned}
\theta_i &= \log[\pi_i/(1-\pi_i)], \\
\omega_i &= n_i, \\
\phi &= 1, \\
b(\theta_i) &= \log[1/(1-\pi_i)] = \log(1+e^{\theta_i}), \\
c(y,\phi) &= \log(\tbinom{n_i}{n_iy}).
\end{aligned} \tag{18}$$

From this it follows that

$$\begin{aligned} b'(\theta_i) &= e^{\theta_i}/(1+e^{\theta_i}) = \pi_i, \\ b''(\theta_i) &= e^{\theta_i}/(1+e^{\theta_i})^2 = \pi_i(1-\pi_i). \end{aligned} \tag{19}$$

Making use of (16)-(17) and (18)-(19), we find that

$$\begin{aligned} u_i &= n_i(y-\pi_i) \\ H_i &= -n_i\pi_i(1-\pi_i). \end{aligned} \tag{20}$$

c. Recall that $E(Y_i|\boldsymbol{x}_i) = \pi_i$ and that the natural parameter is θ_i. Since a canonoical link function g is used, it follows from the first equation of (18) that

$$g(\pi_i) = \log(\frac{\pi_i}{1-\pi_i}) = \theta_i = \boldsymbol{x}_i\boldsymbol{\beta} = \sum_{j=1}^{p} x_{ij}\beta_j. \tag{21}$$

Combining (20) and (21), we find that component j of the score function is

$$\begin{aligned} u_j(\boldsymbol{\beta}) &= \tfrac{\partial L(\boldsymbol{\beta})}{\partial \beta_j} \\ &= \textstyle\sum_{i=1}^n \tfrac{\partial \log f(y_i)}{\partial \beta_j} \\ &= \textstyle\sum_{i=1}^n \tfrac{\partial \log f(y_i)}{\partial \theta_j} \cdot \tfrac{\partial \theta_i}{\partial \beta_j} \\ &= \textstyle\sum_{i=1}^n n_i(y_i-\pi_i)x_{ij}. \end{aligned}$$

Similarly, component (j,k) of the Hessian matrix is

$$\begin{aligned} H_{jk}(\boldsymbol{\beta}) &= -\tfrac{\partial^2 L(\boldsymbol{\beta})}{\partial \beta_j \partial \beta_k} \\ &= \textstyle\sum_{i=1}^n \sum_{i=1}^n \tfrac{\partial^2 \log f(y_i)}{\partial \beta_j \partial \beta_k} \\ &= \textstyle\sum_{i=1}^n \tfrac{\partial^2 \log f(y_i)}{\partial^2 \theta_j} \cdot \tfrac{\partial \theta_i}{\partial \beta_j} \cdot \tfrac{\partial \theta_i}{\partial \beta_k} \\ &= -\textstyle\sum_{i=1}^n n_i\pi_i(1-\pi_i)x_{ij}x_{ik}. \end{aligned} \tag{22}$$

Since the components of the Hessian matrix do not depend on data $Y_1, \ldots, Y_n$, they equal their expected values. From (22) we deduce that the components of the Fisher information matrix are

$$\begin{aligned} J_{jk}(\boldsymbol{\beta}) &= -E[H_{jk}(\boldsymbol{\beta})] \\ &= -H_{jk}(\boldsymbol{\beta}) \\ &= \textstyle\sum_{i=1}^n n_i\pi_i(1-\pi_i)x_{ij}x_{ik}. \end{aligned}$$