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

MT 5019
SOLUTIONS
February 13 2024

Solutions for Examination Categorical Data Analysis, February 13, 2024

Problem 1

a. We have that

$$P(Y=1|X=x)=\pi(x)=\frac{\exp(\alpha+\beta x)}{1+\exp(\alpha+\beta x)}. \tag{1}$$

b. The predicted probability of bankruptcy for a person with income $I=\mu/8$ and predictor $X=\log_2(I/\mu)=-3$, is

$$\hat{\pi}(-3)=\frac{\exp(\hat{\alpha}-3\hat{\beta})}{1+\exp(\hat{\alpha}-3\hat{\beta})}=\frac{\exp(-4.7+3\cdot 0.85)}{1+\exp(-4.7+3\cdot 0.85)}=\frac{\exp(-2.15)}{1+\exp(-2.15)}=0.104.$$

c. We estimate the variance

$$\begin{aligned}\mathrm{Var}(\hat{\alpha}-3\hat{\beta}) &= \mathrm{Var}(\hat{\alpha})-2\cdot 3\cdot \mathrm{Cov}(\hat{\alpha},\hat{\beta})+3^2\cdot \mathrm{Var}(\hat{\beta})\\ &= \mathrm{Var}(\hat{\alpha})-6\cdot \mathrm{Cov}(\hat{\alpha},\hat{\beta})+9\cdot \mathrm{Var}(\hat{\beta})\end{aligned}$$

by the squared standard error

$$\begin{aligned}\mathrm{SE}^2 &= \widehat{\mathrm{Var}}(\hat{\alpha}-3\hat{\beta})\\ &= \widehat{\mathrm{Var}}(\hat{\alpha})-6\cdot \widehat{\mathrm{Cov}}(\hat{\alpha},\hat{\beta})+9\cdot \widehat{\mathrm{Var}}(\hat{\beta})\\ &= 0.015+6\cdot 0.003+9\cdot 0.005\\ &= 0.078.\end{aligned}$$

This gives an approximate 95% confidence interval

$$\begin{aligned}&(\hat{\alpha}-3\hat{\beta}-1.96\cdot \mathrm{SE},\hat{\alpha}-3\hat{\beta}+1.96\cdot \mathrm{SE})\\ =\ &\left(-2.15-1.96\cdot\sqrt{0.078},-2.15+1.96\cdot\sqrt{0.078}\right)\\ =\ &(-2.697,-1.603)\end{aligned}$$

for $\mathrm{logit}[\pi(-3)]=\alpha-3\beta$, and an approximate 95% confidence interval

$$\left(\frac{\exp(-2.677)}{1+\exp(-2.677)},\frac{\exp(-1.603)}{1+\exp(-1.603)}\right)=(0.0631,0.1676)$$

for $\pi(-3)$.

d. Let I_1 and $I_2 = I_1/2$ refer to Adam's and Ben's annual incomes. The corresponding predictors are $x_1 = \log_2(I_1/\mu)$ and $x_2 = \log_2(I_2/\mu) = x_1 - 1$. Hence, the odds ratio of bankruptcy between Adam and Ben, is

$$\mathrm{OR} = \frac{\pi(x_1)/(1-\pi(x_1))}{\pi(x_2)/(1-\pi(x_2))} = \frac{\exp(\alpha+\beta x_1)}{\exp(\alpha+\beta x_2)} = \exp(\beta).$$

Similarly as in part 1c) one shows that an approximate 95% confidence interval for $\log(\mathrm{OR}) = \beta$ is

$$\begin{aligned} & \left(\hat{\beta} - 1.96\cdot\sqrt{\widehat{\mathrm{Var}}(\hat{\beta})}, \hat{\beta} + 1.96\cdot\sqrt{\widehat{\mathrm{Var}}(\hat{\beta})}\right) \\ = & \left(-0.85 - 1.96\cdot\sqrt{0.005}, -0.85 + 1.96\cdot\sqrt{0.005}\right) \\ = & (-0.989, -0.711). \end{aligned}$$

The corresponding confidence interval for the odds ratio is

$$(\exp(-0.989), \exp(-0.711)) = (0.3721, 0.4910).$$

Consequently, with (approximate) probability 95%, the odds of bankruptcy for Adam is between 37% and 49% of the odds of bankruptcy for Ben.

Problem 2

a. This is a case-control design where the column sums are fixed to 500. The two columns have independent binomial distributions. That is, the number of individuals in the high income group among the cases and controls are independent, with binomial distributions.

b. The odds that a case belongs to the high income group is

$$P(X=1|Y=1)/P(X=0|Y=1),$$

and for a control individual, the odds of belonging to this income group is

$$P(X=1|Y=0)/P(X=0|Y=0).$$

This gives an odds ratio

$$\mathrm{OR}^* = \frac{P(X=1|Y=1)/P(X=0|Y=1)}{P(X=1|Y=0)/P(X=0|Y=0)}. \tag{2}$$

c. In order to test H_0 we check if $\hat{\beta} - \hat{\beta}^*$ is significantly different from zero. For this we will use that the two parameter estimates are approximately normally distributed, and from separate studies. They can therefore be regarded as independent random variables, so that $\hat{\beta} - \hat{\beta}^*$ is approximately normal, with a standard error

$$\mathrm{SE} = \sqrt{\widehat{\mathrm{Var}}(\hat{\beta}) + \widehat{\mathrm{Var}}(\hat{\beta}^*)} = \sqrt{0.005 + 0.007} = 0.1095.$$

This gives a Wald statistic

$$z = \frac{\hat{\beta} - \hat{\beta}^*}{\mathrm{SE}} = \frac{-0.85 - (-0.76)}{0.1095} = -0.822.$$

Since $|z| \leq 1.96$, we do not reject H_0.

d. From Bayes' Theorem we deduce

$$P(X=i|Y=j) = \frac{P(Y=j|X=i)P(X=i)}{P(Y=j)}.$$

Inserting this formula into (2), we find that

$$\begin{aligned}
\exp(\beta^*) &= \mathrm{OR}^* \\
&= \frac{\frac{P(Y=1|X=1)P(X=1)}{P(Y=1)} / \frac{P(Y=1|X=0)P(X=0)}{P(Y=1)}}{\frac{P(Y=0|X=1)P(X=1)}{P(Y=0)} / \frac{P(Y=0|X=0)P(X=0)}{P(Y=0)}} \\
&= \frac{P(Y=1|X=1)/P(Y=0|X=1)}{P(Y=1|X=0)/P(Y=0|X=0)} \\
&= \frac{\pi(1)/(1-\pi(1))}{\pi(0)/(1-\pi(0))} \\
&= \frac{\exp(\alpha+\beta)}{\exp(\alpha)} \\
&= \exp(\beta),
\end{aligned} \tag{3}$$

where in the fourth step the logistic regression model (1) of Problem 1 was assumed.

Problem 3

a. The loglinear parametrization of (XY, Z) is

$$\mu_{ijk} = \exp(\lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY}) \tag{4}$$

for $1 \leq i, j, k \leq 2$. Assume that $X = 2$, $Y = 2$ and $Z = 2$ are chosen as baseline levels. Then all loglinear parameters are put to zero for which at least one index i, j or k equals 2. The remaining parameters are

$$\boldsymbol{\beta} = (\lambda, \lambda_1^X, \lambda_1^Y, \lambda_1^Z, \lambda_{11}^{XY}). \tag{5}$$

b. It follows from (4) that

$$\mu_{ijk} = A_{ij}B_k,$$

with $A_{ij} = \exp(\lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY})$ and $B_k = \exp(\lambda_k^Z)$. Then

$$\begin{aligned}
\mu_{ij+} &= A_{ij}B_+, \\
\mu_{++k} &= A_{++}B_k, \\
\mu_{+++} &= A_{++}B_+.
\end{aligned}$$

Consequently,

$$\frac{\mu_{ij+}\mu_{++k}}{\mu_{+++}} = \frac{A_{ij}B_+ \cdot A_{++}B_k}{A_{++}B_+} = A_{ij}B_k = \mu_{ijk}.$$

An alternative solution uses cell probabilities

$$\pi_{ijk} = \frac{\mu_{ijk}}{\mu_{+++}}$$

of the multinomial model, obtained by conditioning the Poisson model on the total cell count n_{+++}. Since Z is independent of X, Y, we have that

$$\mu_{ijk} = \mu_{+++} \cdot \pi_{ijk} = \mu_{+++} \cdot \pi_{ij+}\pi_{++k} = \mu_{+++} \cdot \frac{\mu_{ij+}}{\mu_{+++}} \cdot \frac{\mu_{++k}}{\mu_{+++}} = \frac{\mu_{ij+}\mu_{++k}}{\mu_{+++}},$$

as was to be proved.

c. The ML-estimates

$$\hat{\mu}_{ijk} = \frac{n_{ij+}n_{++k}}{n}$$

of all expected cell counts of model (XY, Z) are found by replacing μ_{ij+}, μ_{++k} and μ_{+++} in the definition of μ_{ijk} by their corresponding observed values n_{ij+}, n_{++k} and $n = n_{+++}$. By summing data from the two partial tables we get the following marginal table for X and Y:

Values of n_{ij+}

	$j = 1$	$j = 2$
$i = 1$	85	74
$i = 2$	48	53

Since the total number of observations of the two partial tables are $n_{++1} = 174$ and $n_{++2} = 86$, and the total number of observations is $n = 174 + 86 = 260$, we get

$$\hat{\mu}_{111} = \frac{n_{11+}n_{++1}}{n} = \frac{85 \cdot 174}{260} = 56.88,$$

for cell $(1, 1, 1)$. A similar calculation of all other $\hat{\mu}_{ijk}$ gives the following result:

Values of $\hat{\mu}_{ij1}$:

	$j = 1$	$j = 2$
$i = 1$	56.88	49.52
$i = 2$	32.12	35.47

Values of $\hat{\mu}_{ij2}$:

	$j = 1$	$j = 2$
$i = 1$	28.11	24.48
$i = 2$	15.88	17.53

d. Let M refer to a loglinear model that data is drawn from. The chisquare statistic for testing the null hypothesis $H_0 : M = (XY, Z)$ against the alternative hypothesis $H_a : M = (XYZ)$ but not $M = (XY, Z)$, is

$$\begin{aligned} X^2(XY, Z) &= \textstyle\sum_{ijk} \frac{(n_{ijk} - \hat{\mu}_{ijk})^2}{\hat{\mu}_{ijk}} \\ &= \frac{(65-56.88)^2}{56.88} + \ldots + \frac{(15-17.53)^2}{17.53} \\ &= 8.419 \\ &> \chi^2_3(0.05) = 7.81. \end{aligned} \tag{6}$$

Therefore we reject H_0 at level 5%. In the last step of (6) we used that the number of degrees of freedom is

$$\mathrm{df} = p(XYZ) - p(XY, Z) = 8 - 5 = 3,$$

since the saturated model has one parameter for each cell, and there are $2 \times 2 \times 2 = 8$ cells in the table. From (5) we also know that the parameter vector $\boldsymbol{\beta}$ of (XY, Z) contains 5 parameters.

Problem 4

a. Submodel $(X * W + Z * W)$ has one intercept, three types of main effects (X, Z and W) and two types of second order interactions (XW and ZW). It follows that

$$P(Y = 1|X = i, Z = k, W = h) = \frac{\exp(\alpha + \beta_i^X + \beta_k^Z + \beta_h^W + \beta_{ih}^{XW} + \beta_{kh}^{ZW})}{1 + \exp(\alpha + \beta_i^X + \beta_k^Z + \beta_h^W + \beta_{ih}^{XW} + \beta_{kh}^{ZW})}.$$

b. The number of parameters of $(X * W + Z * W)$ is

$$p = 1 + (3 - 1) + (3 - 1) + (3 - 1) + (3 - 1)(3 - 1) + (3 - 1)(3 - 1) = 15,$$

where the first term corresponds to an intercept, each main effect contributes with $3 - 1 = 2$ parameters (one per level; excluding the baseline level), and each second order interaction adds $(3 - 1)(3 - 1) = 4$ parameters (one for each pair of levels, none of which is a baseline level).

c. Reasoning as in 4a), each main effect, second order interaction and third order interaction adds $3 - 1 = 2$, $(3 - 1)^2 = 4$ and $(3 - 1)^3 = 8$ parameters. Since each model is balanced, we know how many main effects, second order interactions and third order interactions there are. This gives the following completion of the given table:

M	$G^2(M)$	$p(M)$
$(X * Z * W)$	0	27
$(X * Z + X * W + Z * W)$	7.70	19
$(X * Z + X * W)$	15.27	15
$(X * Z + Z * W)$	31.76	15
$(X * W + Z * W)$	20.43	15
$(X + Z * W)$	36.11	11
$(Z + X * W)$	24.57	11
$(W + X * Z)$	38.61	11
$(X + Z + W)$	41.57	7
None	117.78	1

d. The deviance $G^2(M) = 2\,[L(X * Z * W) - L(M)]$ of M is a log likelihood ratio test statistic when M is tested against the saturated model $(X * Z * W)$. We can therefore write Akaike's information criterion as

$$\text{AIC}(M) = -2L(M) + 2p(M) = -2L(X * Z * W) + G^2(M) + 2p(M).$$

Since $-2L(X * Z * W)$ does not depend on the model M, minimizing $\text{AIC}(M)$ with respect to M is equivalent to minimizing the sum of the second and twice the third columns of the above table. This gives

M	$G^2(M) + 2p(M)$
$(X * Z * W)$	54.00
$(X * Z + X * W + Z * W)$	45.70
$(X * Z + X * W)$	**45.27**
$(X * Z + Z * W)$	61.76
$(X * W + Z * W)$	50.43
$(X + Z * W)$	58.11
$(Z + X * W)$	46.57
$(W + X * Z)$	60.61
$(X + Z + W)$	55.57
None	119.78

The chosen model, with lowest AIC(M) is therefore $(X * Z + X * W)$.

e. In forward inclusion (FI), we start to test the "None" model against $(X + Z + W)$, the one with three main effects. This gives a likelihood ratio statistic

$$\begin{aligned} G^2(\text{None}, X + Z + W) &= G^2(\text{None}) - G^2(X + Z + W) \\ &= 117.78 - 41.57 = 76.21 \\ &> \chi^2_{7-1}(0.05) = 12.59, \end{aligned}$$

where "None" is rejected. We continue to test $(X+Z+W)$ against all three models with one second order interaction effect, and obtain

$$\begin{aligned} G^2(X + Z + W, W + X * Z) &= 41.57 - 38.61 = 2.96 < \chi^2_{11-7}(0.05) = 9.49, \\ G^2(X + Z + W, Z + X * W) &= 41.57 - 24.57 = 17.00 > \chi^2_{11-7}(0.05) = 9.49, \\ G^2(X + Z + W, X + Z * W) &= 41.57 - 36.11 = 5.46 < \chi^2_{11-7}(0.05) = 9.49, . \end{aligned}$$

Only in one of the three tests is $(X + Z + W)$ rejected, when tested against $(Z + X * W)$. In the next step this model is tested against the two models with one additional second order interaction:

$$\begin{aligned} G^2(Z + X * W, X * W + Z * W) &= 24.57 - 20.43 = 4.14 < \chi^2_{15-11}(0.05) = 9.49, \\ G^2(Z + X * W, X * Z + X * W) &= 24.57 - 15.27 = 9.30 < \chi^2_{15-11}(0.05) = 9.49. \end{aligned}$$

Since $Z + X * W$ is not rejected in any of the two tests, this model is eventually selected by the FI scheme. Notice that this model is smaller than the one obtained with the AIC procedure.

Problem 5

a. The number of accidents for drivers in bonus class i, has probability function

$$f_{Y_i}(y_i) = \exp(-t_i\mu_i)\frac{(t_i\mu_i)^{y_i}}{y_i!}.$$

Since these Poisson variables are independent, it follows that the log likelihood is

$$L(\alpha, \beta) = \sum_{i=1}^{4} \log\left[f_{Y_i}(y_i)\right] = \sum_{i=1}^{4} \left[y_i \log(\mu_i) - t_i\mu_i\right] + \text{const},$$

where the constant is independent of α and β.

b. Since $\partial\mu_i/\partial\alpha = \mu_i$ and $\partial \log(\mu_i)/\partial\alpha = 1$, the score function component for α is

$$u_1(\alpha,\beta) = \frac{\partial L(\alpha,\beta)}{\partial\alpha} = \sum_{i=1}^{4}(y_i - t_i\mu_i).$$

The maximum likelihood estimator $\hat{\alpha}_0$ of α under the assumption $\beta = 0$, is obtained from

$$u_1(\hat{\alpha}_0, 0) = 0 \Longleftrightarrow \sum_{i=1}^{4}\left[y_i - t_i\exp(\hat{\alpha}_0)\right].$$

This equation has the explicit solution

$$\hat{\alpha}_0 = \log\frac{\sum_{i=1}^{4} y_i}{\sum_{i=1}^{4} t_i} = 4.032.$$

c. Using that $\partial\mu_i/\partial\beta = x_i\mu_i$ and $\partial\log(\mu_i)/\partial\beta = x_i$, we first obtain the score component

$$u_2(\alpha,\beta) = \frac{\partial L(\alpha,\beta)}{\partial\beta} = \sum_{i=1}^{4} x_i(y_i - t_i\mu_i)$$

for β. Evaluation of the score vector components at $(\hat{\alpha}_0, 0)$ gives

$$\begin{aligned} u_1(\hat{\alpha}_0,0) &= 0,\\ u_2(\hat{\alpha}_0,0) &= \textstyle\sum_i x_i(y_i - t_i\exp(\hat{\alpha}_0))\\ &= \textstyle\sum_i x_i y_i - \frac{\sum_i x_i t_i \sum_i y_i}{\sum_i t_i}\\ &= -378.61. \end{aligned}$$

Differentiating u_1 and u_2 with respect to α and β, we obtain the elements

$$\begin{aligned} H_{11}(\alpha,\beta) &= \partial^2 L/\partial\alpha^2 = -\textstyle\sum_i t_i\mu_i,\\ H_{12}(\alpha,\beta) = H_{21}(\alpha,\beta) &= \partial^2 L/(\partial\alpha\partial\beta) = -\textstyle\sum_i x_i t_i\mu_i,\\ H_{22}(\alpha,\beta) &= \partial^2 L/\partial\beta^2 = -\textstyle\sum_i x_i^2 t_i\mu_i, \end{aligned}$$

of the Hessian matrix. Evaluation of these elements at $(\hat{\alpha}_0, 0)$ gives

$$\begin{aligned} H_{11}(\hat{\alpha}_0,0) &= -\exp(\hat{\alpha}_0)\textstyle\sum_i t_i = -\sum_i y_i = -2812,\\ H_{12}(\hat{\alpha}_0,0) = H_{21}(\hat{\alpha}_0,0) &= -\exp(\hat{\alpha}_0)\textstyle\sum_i x_i t_i = -\sum_i x_i t_i \sum_i y_i/\sum_i t_i = -7684,\\ H_{22}(\hat{\alpha}_0,0) &= -\exp(\hat{\alpha}_0)\textstyle\sum_i x_i^2 t_i = -\sum_i x_i^2 t_i \sum_i y_i/\sum_i t_i = -24077. \end{aligned}$$

Since the Newton-Raphson procedure in each step maximizes a second order Taylor expansion of $L(\alpha,\beta)$ around the previous iterate,

$$(\alpha^{(1)},\beta^{(1)}) = (\hat{\alpha}_0, 0) - (u_1(\hat{\alpha}_0,0), u_2(\hat{\alpha}_0,0))\begin{pmatrix} H_{11}(\hat{\alpha}_0,0) & H_{12}(\hat{\alpha}_0,0)\\ H_{21}(\hat{\alpha}_0,0) & H_{22}(\hat{\alpha}_0,0)\end{pmatrix}^{-1}.$$

Inserting numerical values, we find that

$$(\alpha^{(1)},\beta^{(1)}) = (4.032, 0) - (0, -378.61)\begin{pmatrix} -2812 & -7684\\ -7684 & -24077\end{pmatrix}^{-1} = (4.3679, -0.1229).$$

d. The Fisher scoring algorithm differs from the Newton-Raphson algorithm in that the Hessian matrix elements $H_{ij}(\alpha, \beta)$ are replaced by the negative Fisher information elements $-J_{ij}(\alpha, \beta) = E\left[H_{ij}(\alpha, \beta)\right]$ at all places. But since the Hessian matrix does not depend on data Y_i it is non-stochastic, so that $-J_{ij}(\alpha, \beta) = H_{ij}(\alpha, \beta)$. Therefore the Fisher scoring algorithm is identical to Newton-Raphson (due to the fact that this model uses a log link, which is canonical for Poisson data). In particular, $(\alpha^{(1)}, \beta^{(1)}) = (4.3679, -0.1229)$.