last change: 25.11.2012@12.00

Exercises

---

1. log-linear regression models

$$\mathbb{P}\left(y_i \mid \alpha, \beta, \mathbf{x}_i\right) \sim \mu_i^{y_i} \exp\left(-\mu_i\right)$$
$$\mu\left(\mathbf{x}\right) = \exp\left(\alpha + \sum_j \beta^{(j)} x^{(j)}\right)$$
$$\log\left(\mu\left(\mathbf{x}\right)\right) = \alpha + \sum_j \beta^{(j)} x^{(j)}$$

data

1	y_1	$x_1^{(1)}$	$\ldots$	$x_1^{(k)}$
$\vdots$				
n	y_n	$x_n^{(1)}$		$x_n^{(k)}$

Likelihood:

$$\log L\left(\alpha, \beta \mid y_i, \mathbf{x}_i\right) = \sum_{i=1}^{n} y_i \left(\alpha + \sum_{j=1}^{k} \beta^{(j)} x_i^{(j)}\right) - \exp\left(\alpha + \sum_{j=1}^{k} \beta^{(j)} x_i^{(j)}\right)$$

log-linear models and contingency tables

for the following data:

age-group	cat1$\left(y^{(1)}\right)$	cat2$\left(y^{(2)}\right)$	cat3$\left(y^{(3)}\right)$
1	$y_1^{(1)}$	$y_1^{(2)}$	$y_1^{(3)}$
2	$y_2^{(1)}$	$y_2^{(2)}$	$y_2^{(3)}$
$\vdots$			
I	$y_I^{(1)}$	$y_I^{(2)}$	$y_I^{(3)}$

a) specify Poisson reg model and interpret the coefficients

solution:

$$\log \mu_{ij} = \mu_0 + \alpha_i + \beta_j + (\alpha\beta)_{ij}$$
$$\alpha_1 = \beta_1 = (\alpha\beta)_{1j} = (\alpha\beta)_{i1} = 0$$

b) write likelihood function

solution: given data in form of a 2-dim contingency table $\left(y_i^{(j)}\right),\ i = 1, \ldots, I,\ j = 1, \ldots, J$

get

$$\mathbb{L}\left(\alpha, \beta, \alpha\beta \mid y_i^{(j)}\right) = \prod_{j=1}^{J} \prod_{i=1}^{I} \mathbb{P}\left(y_i^{(j)} \middle| \alpha, \beta, \alpha\beta\right)$$
$$\mathbb{P}\left(y_i^{(j)} \middle| \alpha, \beta, \alpha\beta\right) \sim \mu_{ij}^{y_i^{(j)}} \exp\left(-\mu_{ij}\right)$$
$$\mu_{ij} = \exp\left(\mu_0 + \alpha_i + \beta_j + (\alpha\beta)_{ij}\right)$$
$$\alpha_1 = \beta_1 = (\alpha\beta)_{1j} = (\alpha\beta)_{i1} = 0$$

c) write likelihood-ratio test for hypothesis of age–group independence of cat

reduced model

$$\mathbb{L}\left(\alpha \mid y_i^{(j)}\right) = \prod_{j=1}^{J} \prod_{i=1}^{I} \mathbb{P}\left(y_i^{(j)} \middle| \alpha\right)$$
$$\mathbb{P}\left(y_i^{(j)} \middle| \alpha\right) \sim \mu_{ij}^{y_i^{(j)}} \exp\left(-\mu_{ij}\right)$$
$$\mu_{ij} = \exp\left(\mu_0 + \alpha_i\right)$$
$$\alpha_1 = 0$$

LR-test:

$$-2\left(\log \mathbb{L}\left(\alpha \mid y_i^{(j)}\right) - \mathbb{L}\left(\alpha, \beta, \alpha\beta \mid y_i^{(j)}\right)\right) \sim \chi^2\left(df_{full} - df_{red}\right)$$

$df_{full} = 3I,\ df_{red} = 3$

d) discuss equivalence of the log-linear and multinomial models for the given grouped data

solution: notation: $\mathbb{P}\left(y = j \mid \text{group}\,(i)\right) = p_{j|i}$

multinomial:

$$\log\left(p_{j|i} \,/\, p_{1|i}\right) = \alpha_j^* + \beta_{ij}^*$$

log linear: $\mu = \sum_{ij} \mu_{ij},\ \mu_{ij} = p_{ij}\mu$
get

$$\begin{aligned}\log\left(p_{j|i} \,/\, p_{1|i}\right) &= \alpha_j^* + \beta_{ij}^* \\ &= \log\left(\mu p_{ij} \,/\, \mu p_{i1}\right) \\ &= \log\left(\mu_{ij} \,/\, \mu_{i1}\right) \\ &= \alpha_i + \beta_j + (\alpha\beta)_{ij} - \alpha_i \\ &= \beta_j + (\alpha\beta)_{ij}\end{aligned}$$

interpretation of parameters:

$$\log\left(p_{j|1} \,/\, p_{1|1}\right) = \beta_j$$

$\Longrightarrow$

$$(\alpha\beta)_{ij} = \log\left(p_{j|i} \,/\, p_{1|i}\right) - \log\left(p_{j|1} \,/\, p_{1|1}\right) = \log \frac{p_{j|i} \,/\, p_{1|i}}{p_{j|1} \,/\, p_{1|1}}$$

---

2. three-dimensional contingency tables and log-linear models

a) specify a 3-dimensional contingency table in (group) variables (factors) R, C, L and write an equivalent (full) log-linear model

solution:

$$\log \mu_{ijk} = \mu_0 + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk}$$

b) interpret triple interaction as differences between pairwise interactions of any two variables at different levels of the third variable

solution: fix third variable (k_0)

$$\begin{aligned}\log \mu_{ijk_0} &= \mu_0 + \alpha_i + \beta_j + \gamma_{k_0} + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik_0} + (\beta\gamma)_{jk_0} + (\alpha\beta\gamma)_{ijk_0} \\ &= (\mu_0 + \gamma_{k_0}) + \left(\alpha_i + (\alpha\gamma)_{ik_0}\right) + \left(\beta_j + (\beta\gamma)_{jk_0}\right) + \left((\alpha\beta)_{ij} + (\alpha\beta\gamma)_{ijk_0}\right) \\ &= \mu_0^* + \alpha_i^* + \beta_j^* + (\alpha\beta)_{ij}^*\end{aligned}$$

where

1. $\alpha_i^*,\ \beta_j^*,\ (\alpha\beta)_{ij}^*$ are main effect and interactions between the different levels of the first and second factor on a fixed level of the third factor.

2. if $(\alpha\beta\gamma)_{ijk} = 0$ for all i, j, k then the interactions between the different levels of the first and second factor are the same regardless of any level of the third factor. if interactions are interpreted as log-odds ratios then the log odds ratios between any pair of levels of the first and second factor are iqual for every level of the third factor. interchanging the factors one could say that *in the case of zero triple interactions the log-odds ratios between any pair of levels of any pair of factors are the same for every level of the third factor.* the model without triple interactions is called the **model of homogenuous odds ratios**.

3. If $(\alpha\beta\gamma)_{ijk_0} \neq 0$ for some k_0 this triple interaction is the change in the interactions between the different levels of the first and the second factor for different levels of the third factor.

c) interpret the pairwise interactions in terms of (conditional) odds
solution:

$$\begin{aligned}\log \mu_{ijk} - \log \mu_{i1k} &= \mu_0 + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} \\ &\quad - (\mu_0 + \alpha_i + \gamma_k + (\alpha\gamma)_{ik}) \\ &= \beta_j + (\alpha\beta)_{ij} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk}\end{aligned}$$

consider any fixed level of the third factor (k). Then

$$\log \mu_{ijk} - \log \mu_{i1k} = \log\left(\mu_{ijk} \,/\, \mu_{ij1}\right) = \log\left(p_{ijk} \,/\, p_{ij1}\right)$$

d) describe the LR-test for the hypothesis that triple interactions are negligible.

in the homogenuos odds ratio model:

e) interpret the pairwise interactions in terms of odds.

f) discuss the relation between pairwise interactions and conditional independence properties.

g) discuss the relations between the reduced model and a multinomial regression of first on second and third factors (hint: Table 7.2)

---

3)

corresponding to the data

```
data
      A   B   C   D
cat1 42  44  40 126
cat2 33 133  43  37
cat3 37  43 127  46
cat4 48  42  38 121
```

consider the following outputs:

a) model

$$\boxed{\log \mu_i = \mu_0 + \log \frac{p_i}{p_1}}$$

where $\mu_0 = \log \lambda + \log p_1$

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.68888    0.07906  46.661  < 2e-16 ***
xB           0.49317    0.10033   4.915 8.86e-07 ***
xC           0.43825    0.10140   4.322 1.55e-05 ***
xD           0.72392    0.09633   7.515 5.71e-14 ***
```

(check:)

```
> (rs<-apply(n,2,sum))
  A   B   C   D
160 262 248 330
> log(rs[-1]/rs[1])
        B         C         D
0.4931707 0.4382549 0.7239188
```

also, note that the mle of the intercept ($\widehat{\mu}_0$)turns out to be

```
log(40)=log(rs[1]/4)=3.68888
```

i.e., the 2-dimsional data are replaced by their row sum averages

b) model

$$\boxed{\log \mu_i j = \mu_0 + \log \frac{p_{i\cdot}}{p_{1\cdot}} + \log \frac{p_{\cdot i}}{p_{\cdot 1}}}$$

where $\mu_0 = \log \lambda + \log p_{1\cdot} + \log p_{\cdot 1}$

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.69685    0.09601  38.504  < 2e-16 ***
xB           0.49317    0.10033   4.915 8.86e-07 ***
```

```
xC           0.43826    0.10140   4.322 1.55e-05 ***
xD           0.72392    0.09633   7.515 5.71e-14 ***
ycat2       -0.02410    0.08963  -0.269    0.788
ycat3        0.00396    0.08900   0.044    0.965
ycat4       -0.01198    0.08936  -0.134    0.893
```

(check:)

```
> (rs<-apply(n,2,sum))
> log(rs[-1]/rs[1])
        B         C         D
0.4931707 0.4382549 0.7239188
> (cs<-apply(n,1,sum))
cat1 cat2 cat3 cat4
 252  246  253  249
> log(cs[-1]/cs[1])
        cat2         cat3         cat4
-0.024097552  0.003960401 -0.011976191
```

also, note that the mle of the intercept is not obvious in this case

c) model

$$\boxed{\log \mu_{ij} = \mu_0 + \log \frac{p_{i1}}{p_{11}} + \log \frac{p_{1j}}{p_{11}} + \log \frac{p_{ij} \ / p_{i1}}{p_{1j} \ / p_{11}}}$$

where $\mu_0 = \log \lambda + \log p_{11}$

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  3.73767    0.15430  24.223  < 2e-16 ***
xB           0.04652    0.21572   0.216 0.829264
xC          -0.04879    0.22093  -0.221 0.825216
xD           1.09861    0.17817   6.166 7.01e-10 ***
ycat2       -0.24116    0.23262  -1.037 0.299868
ycat3       -0.12675    0.22547  -0.562 0.574002
ycat4        0.13353    0.21129   0.632 0.527396
xB:ycat2     1.34732    0.29045   4.639 3.50e-06 ***
xC:ycat2     0.31348    0.31995   0.980 0.327192
xD:ycat2    -0.98420    0.29846  -3.298 0.000975 ***
xB:ycat3     0.10376    0.31116   0.333 0.738779
xC:ycat3     1.28206    0.28933   4.431 9.37e-06 ***
xD:ycat3    -0.88089    0.28375  -3.104 0.001906 **
xB:ycat4    -0.18005    0.30196  -0.596 0.550991
xC:ycat4    -0.18482    0.30977  -0.597 0.550743
xD:ycat4    -0.17402    0.24667  -0.706 0.480498
```

(check:)

```
> log(n[1,-1]/n[1,1])
          B           C           D
 0.04652002 -0.04879016  1.09861229
> log(n[-1,1]/n[1,1])
      cat2       cat3       cat4
-0.2411621 -0.1267517  0.1335314
> for(i in 2:4){
+ for(j in 2:4){
+ print(log(n[i,j]/n[i,1])-log(n[1,j]/n[1,1]))
+ }
+ }
[1]  1.347322
[1]  0.3134827
[1] -0.984202
[1]  0.1037622
[1]  1.282059
[1] -0.8808888
[1] -0.1800514
[1] -0.1848247
[1] -0.1740228
also, note that log(42)=log(n[1,1])=3.73767
```
