Exercises: Log linear models[a, b]

Lecturer: Georgios Karagiannis georgios.karagiannis@durham.ac.uk

[a]Author: Georgios P. Karagiannis.
[b]Acknowledgments to students in 2018 for spotting typos.

The next exercise is from Homework 2

Exercise 1

Assume model (X, Y, Z), under Poisson sampling scheme. Consider corner points constraints.

1. Find the Likelihood equations.
2. Express the Log linear coefficients with respect to the expected counts
3. Find the MLEs of the Log linear coefficients

The next exercise is from Homework 2

Exercise 2

Consider the dataset in Table 4.

Gender (Z)	Information Opinion (X)	Health Opinion (Y)	
		Support	Oppose
Male	Support	76	160
	Oppose	6	25
Female	Support	114	181
	Oppose	11	48

Table 1: The 1988 General Social Survey compiled by the National Opinion Research Center asked: "Do you support or oppose the following measures to deal with AIDS? (1) Have the government pay all of the health care costs of AIDS patients; (2) Develop a government information program to promote safe sex practices, such as the use of condoms. This Table summarizes opinions about health care costs (Y) and the information program (X), classified also by the respondent's gender (Z). [Source: 1988 General Social Survey, National Opinion Research Center.]

Use suitable tools in order to compare the following models:

1. [Y , X , Z are independent] vs. [X and Y are conditionally independent on Z]
2. [X and Z are jointly independent from Y] vs. [X and Y are jointly independent from Z]

Any inference based on hypothesis tests should be performed at sig. level 5%.

Exercise 3

Consider a $I \times J \times K$ table n_{ijk} generated from a multinomial sampling scheme; i.e.

$$(n_{111}, ..., n_{IJK}) \sim \text{Mult}(n, \underset{\sim}{\pi})$$

where $\underset{\sim}{\pi} = (\pi_{111}, ..., \pi_{IJK})$, and $n = \sum_{ijk} n_{ijk}$. Show that the rejection area of the Goodness of fit test for model M_0 based on the deviance/likelihood ratio statistic is

$$\text{RA} = \{\underbrace{2 \sum_{\forall i,j,k} n_{ijk} \log(\frac{n_{ijk}}{\hat{\mu}_{ijk}})}_{=G^2} > q\}$$

for some value q, where $\mu_{ijk} = n\pi_{ijk}$, and $\hat{\mu}_{ijk}$ are the MLE of μ_{ijk} under modelM_0.

- The likelihood is
$$L(\underset{\sim}{\pi}) \propto \prod_{\forall i,j,k} \pi_{ijk}^{n_{ijk}} \propto \prod_{\forall i,j,k} \mu_{ij,k}^{n_{ij,k}}$$
- Under the saturated model the MLE (aka maximum likelihood estimate) is $\hat{\mu}_{ijk}^{\text{satur.}} = n_{ijk}$ and under the model M_0 the MLE is $\hat{\mu}_{ijk}$ (this is assumed).
- Then from (**??**), I get
$$\{\log(\frac{L(\underset{\sim}{\pi} = \underset{\sim}{\hat{\mu}}/n_{+++})}{L(\underset{\sim}{\pi} = \underset{\sim}{\hat{\mu}}^{\text{satur.}}/n_{+++})}) < q'\} \Longrightarrow \{\underbrace{2 \sum_{\forall i,j,k} n_{ijk} \log(\frac{n_{ijk}}{\hat{\mu}_{ijk}})}_{=G^2} > q\}$$

The next exercise is from Homework 2

Exercise 4

Consider a $I \times J \times K$ table n_{ijk} generated from a Poisson sampling scheme; i.e.

$$n_{ijk} \sim \text{Poi}(\mu_{ijk})$$

1. Show that the rejection area of the Goodness of fit test for model M_0 based on the deviance/likelihood ratio statistic is

$$\text{RA} = \{2 \sum_{\forall i,j,k} n_{ijk} \log(\frac{n_{ijk}}{\hat{\mu}_{ijk}}) > q\} \tag{1}$$

   for some value q, where $\hat{\mu}_{ijk}$ are the MLE of μ_{ijk} under modelM_0.

2. In an exercise/example in the Handout [Handouts: The Log-linear model], you where asked to compute the rejection area of the LR statistic under the Multinational sampling scheme, and in fact the resulting rejection area was the the same as in (1). State the assumption, we secretly took in order to get (1) under the Poisson sampling scheme in order to make the likelihood ratio under the Poisson sampling scheme to look like that under the Multinational sampling scheme.

Exercise 5

Consider a $2 \times 2 \times 2$ contingency table, with classification variables X, Y, Z.

1. State the equation of the Log-linear model of the model describing the dependency type (XZ, XY)
2. Apply the two types of the non-identifiability constraints (corner points and sum-to-zero)
3. Write down the number of the free parameters, and say how you calculated them
4. Consider the corner points only;
   (a) express the log ratio of $\log(\frac{\pi_{1|j,k}}{\pi_{2|j,k}})$ as a function of the linear model coefficients (aka the λ's).
   (b) express the log coditional odds ratio $\log(\theta^{XY}_{(k)})$ as a function of the linear model coefficients (aka the λ's). and give a sort interpretation about λ^{XY}_{11} based on this.

PJC exercise (modified to Poisson sampling)

Exercise 6

Consider log-linear modelling of four-dimensional contingency tables, i.e. $d = 4$. Suppose that the table has R rows, C columns and S slices and H hyper-slices (values of l). Consider that the observations are collected based on a Multinomial sampling scheme.

- Write down the full model, for $\log(\pi_{ijkl}) = \log(\frac{\mu_{ijkl}}{n_{++++}}) = ...$

  Hint: In multinational sampling, the intercept coefficient in the linear predictor function is set to zero.

- How many parameters (degrees of freedom) are associated with each term in the full model?

- To what value must they all sum?

- How many different possible log-linear models for $d = 4$ are there if we don't require that models are hierarchical?

Exercise 7

The following 2×3 contingency table shows data reported by Fox et al. (1993). In this question, you will use it to work through the Iterative Proportional Fitting process for the independence model. Note that IPF is not necessary in this model as we have closed-form maximum likelihood estimates but it is being used as a simple example.

Anti-emetic response data after 2 days:

	3 Level of response		
	1 None	1 Partial	1 Complete
Control	12	3	7
Treatment	3	7	12

Find the values of $\hat{p}_i$ $(i = 1, 2)$ and $\hat{p}_j$ $(j = 1, 2, 3)$ for this table.

Starting from the "approximation" $\hat{p}_{ij}^{(0)} = 1$ for all i and j, carry out one full iteration of the IPF algorithm.

Why is there no point in carrying out further iterations?

Exercise 8

The R output at the end of the question shows the result of forwards and backwards step-wise model selection using BIC for a 4-dimensional contingency table starting from the independence table. The column labelled AIC is in fact the BIC value in the form where lower values are preferred to higher values.

1. At the third stage of the procedure (begins with `Step:  AIC=1486.01`), explain why the particular model terms were considered for addition and deletion.

2. What models were considered along the way that were nearly as good as the model selected (say with BIC values at most 4 higher than the final model).

3. Interpret the final model considering (in turn) each of hs, phs and fol as a response variable of interest.

```
> step(loglm(f~., minn38), scope=list(lower=~., upper=~.^4), k=log(sum(minn38$f)))
Start:  AIC=3836.09
f ~ hs + phs + fol + sex

          Df    AIC
+ phs:fol 18 2543.0
+ hs:phs   6 2779.1
+ hs:sex   2 3433.9
+ phs:sex  3 3448.1
+ hs:fol  12 3792.6
+ fol:sex  6 3819.7
<none>       3836.1

Step:  AIC=2542.99
f ~ hs + phs + fol + sex + phs:fol

          Df    AIC
+ hs:phs   6 1486.0
+ hs:sex   2 2140.8
+ phs:sex  3 2155.0
+ hs:fol  12 2499.5
+ fol:sex  6 2526.6
<none>       2543.0
- phs:fol 18 3836.1

Step:  AIC=1486.01
f ~ hs + phs + fol + sex + phs:fol + hs:phs

          Df    AIC
+ hs:sex   2 1083.8
+ phs:sex  3 1098.0
+ fol:sex  6 1469.6
<none>       1486.0
+ hs:fol  12 1569.0
- hs:phs   6 2543.0
```

```
- phs:fol 18 2779.1

Step:  AIC=1083.83
f ~ hs + phs + fol + sex + phs:fol + hs:phs + hs:sex

          Df     AIC
+ phs:sex  3  718.53
+ fol:sex  6 1043.04
<none>       1083.83
+ hs:fol  12 1166.83
- hs:sex   2 1486.01
- hs:phs   6 2140.80
- phs:fol 18 2376.92

Step:  AIC=718.53
f ~ hs + phs + fol + sex + phs:fol + hs:phs + hs:sex + phs:sex

             Df     AIC
+ fol:sex     6  714.66
<none>           718.53
+ hs:phs:sex  6  767.92
+ hs:fol     12  801.54
- phs:sex     3 1083.83
- hs:sex      2 1098.04
- hs:phs      6 1752.83
- phs:fol    18 2011.63

Step:  AIC=714.66
f ~ hs + phs + fol + sex + phs:fol + hs:phs + hs:sex + phs:sex +
    fol:sex

              Df     AIC
<none>              714.66
- fol:sex      6  718.53
+ hs:phs:sex   6  764.04
+ hs:fol      12  793.14
+ phs:fol:sex 18  826.08
```

```
- phs:sex      3 1043.04
- hs:sex       2 1094.17
- hs:phs       6 1748.96
- phs:fol     18 1995.25
Call:
loglm(formula = f ~ hs + phs + fol + sex + phs:fol + hs:phs +
    hs:sex + phs:sex + fol:sex, data = minn38, evaluate = FALSE)

Statistics:
                      X^2  df     P(> X^2)
Likelihood Ratio 256.1798 120 6.849299e-12
Pearson          262.6046 120 1.165845e-12
```

The next exercise was addressed in the Last Lecture

Exercise 9

In 1968, 715 blue collar workers, selected from Danish Industry, were asked a number of questions concerning their job satisfaction. Some of these questions were summarized in a measure of job satisfaction. Based on similar questions the job satisfaction of the supervisors were measured. Also included in the investigation was an external evaluation of the quality of management for each factory. Table 2 shows the 715 workers distributed on the three variables

Y: Own job satisfaction,

X: Supervisors job satisfaction,

Z: Quality of management.

Quality of management (Z)	Supervisors job satisfaction(X)	Own Job satisfaction(Y)	
		Low	High
Bad	Low	103	87
	High	32	42
Good	Low	59	109
	High	78	205

Table 2: Own job satisfaction, supervisors job satisfaction and the quality of management for 715 blue collar workers in Denmark in 1968. (Source: Petersen (1968), table M/7)

1. Is the own job satisfaction and the supervisors satisfaction conditionally independent of the quality management?
2. Test the hypothesis that the Quality of management, Supervisors job satisfaction, and Own Job satisfaction are mutually independent, against the hypothesis that the own job satisfaction and the supervisors satisfaction conditionally independent of the quality management
3. [Y and X are conditionally independent on Z] vs. [X and Y jointly independent on Z and Y]

Any inference based on hypothesis tests should be performed at sig. level 5%.

PJC exercise

Exercise 10

In log-linear modelling of contingency tables, we represent the logarithms of the cell probabilities $p_{i_1 \ldots i_p}$ linearly in terms of tables involving subsets of the indices; we constrain those tables by requiring that the sum over any index in one of those tables yields zero.

For example, when $p = 2$, the full model is

$$\log p_{ij} = \eta_{ij} = \lambda + \lambda_i^{(R)} + \lambda_j^{(C)} + \lambda_{ij}^{(RC)} \tag{2}$$

and the constraints are $\sum_i \lambda_i^{(R)} = 0$, $\sum_j \lambda_j^{(C)} = 0$, $\sum_i \lambda_{ij}^{(RC)} = 0$ for each j and $\sum_j \lambda_{ij}^{(RC)} = 0$ for each i.

1. Show that, for any table of cell probabilities p_{ij}, we can find λ, one-dimensional tables $\lambda^{(R)}$, $\lambda^{(C)}$ and two-dimensional table $\lambda^{(RC)}$ satisifying the constraints and so that (2) holds for all i, j.
2. Show also that the resulting values of the various different kinds of "λ" are unique.
3. Without doing detailed calculations, describe how the argument generalises to $p = 3$ and beyond.

I have used the superscripts (R), (C) and (RC) in the interests of clarity but you may omit them provided you are careful about your use of letters!

PJC exercise

Exercise 11

In the lectures, maximum likelihood estimates were derived using Lagrange multipliers. We considered hierarchical log-linear models for $d = 2$ (independence and full/saturated versions).

Some key general features of the process turn out to be that:

- one of the constraints is always that the sum of the cell probabilities is 1 and we always find that the Lagrange multiplier for that constraint is the total number of data n;
- all other Lagrange multipliers turn out to be zero;
- for any log-linear model coefficient "β" (in the Lectures denoted as λ's), the estimates of the table of probabilities having the same indices are simply the corresponding data proportions, i.e. for the $d = 2$ independence model, $\hat{p}_i = y_i/n$ and $\hat{p}_j = y_j/n$ and in the full model $\hat{p}_{ij} = y_{ij}/n$;
- the method used to deduce that $\hat{p}_i = y_i/n$ and $\hat{p}_j = y_j/n$ for $d = 2$ generalises to all single-index marginals
- for $d = 2$, the equations for the saturated model are the same as those for the independence model except that an additional equation is introduced involving β_{ij} and the solution of the additional equation for β_{ij} is obtained by (i) summing the equation with respect to i and exploiting the solution to the equation for β_i to show that the Lagrange multipliers, associated with the constraints that $\sum_i \beta_{ij} = 0$ for each j, all take the same value; (ii) summing with respect to j in order to show that the Lagrange multipliers for the row-sum constraints on β_{ij} all take the same value; (iii) summing with respect to both i and j to show that the sum of the row-sum and column-sum Lagrange multipliers is 0 for each i and j. This generalises when $d > 2$ to all pair-wise interactions present in the model and also generalizes to more complex terms in the model.

Now consider $d = 3$ and the hierarchical model

$$\log p_{ijk} = \eta_{ijk} = \beta + \beta_i + \beta_j + \beta_k + \beta_{ij} + \beta_{ik} + \beta_{jk}$$

Recall that the data here are y_{ijk} and the model is for cell probabilities p_{ijk} and that we use notation involving fewer indices (or alternatively "dots" as subscripts) to mean that we have summed over the indices which have been dropped. For example $y_k = y_{..k} = \sum_{i,j} y_{ijk}$ and $p_{ik} = \sum_j p_{ijk}$.

1. Write down the log-likelihood to be maximised and the constraints to be imposed.
2. Hence write down the expression involving Lagrange multipliers for which a turning point must be found in order to obtain the maximum likelihood estimates.

3. Show that the methods from $d = 2$ can be used hierarchically to derive the following equations satisfied by the maximum likelihood estimates: $\hat{p}_{ij} = y_{ij}/n$, $\hat{p}_{ik} = y_{ik}/n$ and $\hat{p}_{jk} = y_{jk}/n$. Note that you will also show $\hat{p}_i = y_i/n$, $\hat{p}_j = y_j/n$ and $\hat{p}_k = y_k/n$ but these are redundant as they follow from the equations for $\hat{p}_{ij}$ etc.

4. [This is too advanced...] Show that knowing all p_{ij}, p_{ik} and p_{jk} uniquely determines the values of all the "βs". In other words, we have uniquely determined the m.l.e. in the previous part of the question.

For the full model, show how the argument from $d = 2$ generalises to use the solution of the equations for β_{ij}, β_{ik} and β_{jk} to help to show that $\hat{p}_{ijk} = y_{ijk}/n$.

PJC exercise

Exercise 12

Consider log-linear modelling of three-dimensional contingency tables, i.e. $p = 3$. Suppose that the table has R rows, C columns and S slices (values of k).

- Write down the full model.
- How many parameters (degrees of freedom) are associated with each term in the full model?

  Hint: remember that summing over any index of a "λ" yields zero; therefore the number of parameters for λ_i is $R - 1$ since we can determine the value for the first row from the values for the other rows.
- Enumerate all possible hierarchical log-linear models for $p = 3$ which include all main-effect terms.

The constant term λ has no degrees of freedom. It exists purely to ensure that $\sum_{i,j,k} p_{ijk} = 1$ and its value is determined by the values of all the other terms in the model.

For a term involving a single variable, there is just one constraint which is that all the entries sum to zero. This means that the final entry in the table is determined by the values of the remainder and so the number of parameters is one less than the number of entries. So λ_i, λ_j and λ_k have respectively $R - 1$, $C - 1$ and $S - 1$ parameters.

For a "λ" involving two variables, we have a two-dimensional table of entries in which each row and column sums to zero. So if we put arbitrary numbers in all except the final row and column, we can make the table satisfy the constraints by filling in the final column (except for the last row) so each row except the last sums to zero and then fill in the final row so that each column sums to one; note that these numbers in the final row and column are determined by the original arbitrary selection which occupied a rectangle with one less row and one less column. Therefore λ_{ij}, λ_{ik} and λ_{jk} have respectively $(R-1)(C-1)$, $(R-1)(S-1)$ and $(C-1)(S-1)$ parameters.

Finally, for λ_{ijk}, do the same thing by putting arbitrary numbers in all except the final row, column and slice of the three-dimensional table and note that the remaining entries are uniquely determined to satisfy the constraints by first filling in the final row in all except the final column and slice, then filling in the final column in all except the final slices and finishing by filling in the final slice. The arbitrary starting point occupied a three-dimensional table with one less row, one less column and one less slice. Therefore the number of parameters is $(R-1)(C-1)(S-1)$.

Note that $(R-1)+(C-1)+(S-1)+(R-1)(C-1)+(R-1)(S-1)+(C-1)(S-1)+(R-1)(C-1)(S-1) = RCS-1$ is the number of freely varying parameters and this is clearly the number required to specify a discrete probability distribution having RCS distinct possible outcomes.

PJC exercise

Exercise 13

All possible hierarchical log-linear models were fitted to a contingency table for three factors: hs (3 levels), phs (4 levels) and fol (7 levels). In total there were 7861 data in the table.

The following table summarises the results of fitting each model by maximum likelihood estimation. The "log-likelihood" column is the value of the log-likelihood at the maximum. The "deviance" column is twice the difference between the maximum of the log-likelihood for the full model and the maximum of the log-likelihood for the model specified.

Model	Deviance	Log-likelihood
Full	0	40293
hs+phs+fol+hs:phs+hs:fol+phs:fol	87	40249.5
hs+phs+fol+phs:fol+hs:fol	637	39974.5
hs+phs+fol+phs:fol+hs:phs	125	40230.5
hs+phs+fol+hs:phs+hs:fol	787	39899.5
hs+phs+fol+hs:phs	899	39843.5
hs+phs+fol+hs:fol	1412	39587.0
hs+phs+fol+phs:fol	750	39918.0
Independence (hs+phs+fol)	1524	39531.0

1. Compute the number of free parameters (degrees of freedom) for each model in the table.
2. Which model would be chosen using AIC?
3. Which model would be chosen using BIC?

4. For the model chosen in part (c), what would we have learned if were interested in treating "phs" as the response variable?

5. For the model chosen in part (c), carry out the likelihood ratio test for that model versus the Independence model.

Revision

Exercise 14

Consider a sample of 261 individuals, for three factors: Y (2 levels), X (3 levels), and Z (4 levels). The sampling scheme that was implemented was the Poisson sampling scheme. The deviance and the log-likelihood of a number of models are given in Table 3.

Model	Deviance	Log-likelihood	BIC	free parameters	AIC
$[X, Y, Z]$	133.37	-60.63			
$[XY]$	126.62	-54.01			
$[XZ]$	132.91	-54.59			
$[YZ]$	57.12	-44.46			
$[XY, XZ]$	126.16	-53.96			
$[XY, YZ]$	50.37	-42.95			
$[XZ, YZ]$	56.66	-44.36			
$[XY, XZ, YZ]$	49.91	-42.84			
$[XYZ]$		-42.00			

Table 3:

Clarification: Regarding the notation in the table above: In 3 way tables with classification variables X , Y, and Z, the notation $[XY]$, $[XZ]$, and $[YZ]$ is a simplified equivalent of the notation $[Z, XY]$, $[Y, XZ]$, and $[X, YZ]$ respectively. In other words, $[XY]$, $[XZ]$, $[YZ]$ and $[Z, XY]$, $[Y, XZ]$, $[X, YZ]$ represent the same models / independency types respectively. For example, the log-linear model

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

(and hence the type of dependency it represents) is denoted equivalently as both $[XY]$ and $[Z, XY]$.

1. Define the Akaike Information Criterion (AIC), and Bayes Information Criterion (BIC). Fill in Table 3. Which model is selected by AIC, and BIC?

2. Intuitively discuss how AIC and BIC work to address model selection problems.

Exercise 15

During a particular summer, an experiment was conducted to find out the preference between two types of beverages: soda and lemonade. The data was drawn from two locations: city and rural. In each location, the gender and the choice of drinks were collected. The results are summarized in Table 4.

Location (Z)	Gender (X)	Preferred drink (Y)	
		Lemonade	Soda
City	Female	9	1
	Male	70	20
Rural	Female	30	60
	Male	2	8

Table 4: Dataset

1. Calculate the marginal contingency table of the observed counts with classifiers Gender and Preferred drink. Calculate a 95% confidence interval for the marginal odds ratio of the Gender and Drink. Interpret the result.
2. Compute and interpret the results of:
   - (a) conditional odds ratio of the Gender and Drink at each Location level.
   - (b) marginal odds ratio of the Gender and Location.
   - (c) marginal odds ratio of the Location and Drink.
3. Compare the results above, investigate the phenomenon, and explain why this might happen.
4. Specify the log-linear model equations for the dependency types:
   - (a) Gender, Drink, and Location are mutually independent
   - (b) Gender and the Drink are conditionally independent given the Location

   as well as specify the number of parameters for associated log-linear models that result after setting non-identifiability constraints.