

# Birth Data - Loglinear Models

February 1, 2012

```
> library(catdata)
> data(birth)
> attach(birth)
```

In the following loglinear models are fitted with the binary variables Sex, Membranes, Cesarean and Induced from the "birth" data. As an overview a contingency table is plotted.

```
> table1 <- table(Sex, Membranes, Cesarean, Induced)
> ftable(table1)
```

			Induced	0	1
Sex	Membranes	Cesarean			
1	0	0		177	45
		1		37	18
	1	0		104	16
		1		9	7
2	0	0		137	53
		1		24	12
	1	0		74	15
		1		8	2

Now we start fitting the models. The goal is to find a model with good fit but sparse parametrization.

First the saturated model is fitted, then the model with all 3-factor interactions and the model with all 2-factor interactions, and finally the independence model. To control for model fit we look at the corresponding deviances and degrees of freedom.

```
> m4 <- loglin(table1, margin=list(c(1,2,3,4)), fit=TRUE)
```

```
2 iterations: deviation 0
```

```
> cat("deviance(m4)=", m4$lrt, "df(m4)=", m4$df, "\n")
```

```
deviance(m4)= 0 df(m4)= 0
```

```
> m3 <- loglin(table1, margin=list(c(1,2,3), c(1,2,4), c(1,3,4), c(2,3,4)), fit=TRUE)
```

```
3 iterations: deviation 0.02260795
```

```
> cat("deviance(m3)=", m3$lrt, "df(m3)=", m3$df, "\n")
```

```

deviance(m3)= 0.8337201 df(m3)= 1

> m2 <- loglin(table1, margin=list(c(1,2), c(1,3), c(1,4), c(2,3), c(2,4), c(3,4)), fit=TRUE)
3 iterations: deviation 0.09069387

> cat("deviance(m2)=", m2$lrt, "df(m2)=", m2$df, "\n")

deviance(m2)= 4.765221 df(m2)= 5

> m1 <- loglin(table1, margin=list(c(1), c(2), c(3), c(4)), fit=TRUE)
2 iterations: deviation 1.136868e-13

> cat("deviance(m1)=", m1$lrt, "df(m1)=", m1$df, "\n")

deviance(m1)= 28.91474 df(m1)= 11

```

In order to see if a model or rather the reduction of a model is appropriate we use chi-square tests.

```

> (df34 <- m3$df - m4$df)

[1] 1

> (dev34 <- m3$lrt - m4$lrt)

[1] 0.8337201

> 1-pchisq(dev34, df34)

[1] 0.361199

> (df23 <- m2$df - m3$df)

[1] 4

> (dev23 <- m2$lrt - m3$lrt)

[1] 3.931501

> 1-pchisq(dev23, df23)

[1] 0.4153555

> (df12 <- m1$df - m2$df)

[1] 6

> (dev12 <- m1$lrt - m2$lrt)

[1] 24.14952

> 1-pchisq(dev12, df12)

[1] 0.000490195

```

Since model "m2" fits the data well but model "m1" is definitely rejected we fit submodels of "m2" by leaving out one of the 2-factor interactions.

```
> m2.GM <- loglin(table1, margin=list(c(1,3), c(1,4), c(2,3), c(2,4), c(3,4)), fit=TRUE)
3 iterations: deviation 0.08317297

> cat("deviance(m2.GM)=", m2.GM$lrt, "df(m2.GM)=", m2.GM$df, "\n")
deviance(m2.GM)= 5.243597 df(m2.GM)= 6

> m2.MC <- loglin(table1, margin=list(c(1,2), c(1,3), c(1,4), c(2,4), c(3,4)), fit=TRUE)
3 iterations: deviation 0.05209252

> cat("deviance(m2.MC)=", m2.MC$lrt, "df(m2.MC)=", m2.MC$df, "\n")
deviance(m2.MC)= 9.964879 df(m2.MC)= 6

> m2.CI <- loglin(table1, margin=list(c(1,2), c(1,3), c(1,4), c(2,3), c(2,4)), fit=TRUE)
3 iterations: deviation 0.07241194

> cat("deviance(m2.CI)=", m2.CI$lrt, "df(m2.CI)=", m2.CI$df, "\n")
deviance(m2.CI)= 12.16702 df(m2.CI)= 6

> m2.GI <- loglin(table1, margin=list(c(1,2), c(1,3), c(2,3), c(2,4), c(3,4)), fit=TRUE)
3 iterations: deviation 0.05997258

> cat("deviance(m2.GI)=", m2.GI$lrt, "df(m2.GI)=", m2.GI$df, "\n")
deviance(m2.GI)= 6.971228 df(m2.GI)= 6

> m2.GC <- loglin(table1, margin=list(c(1,2), c(1,4), c(2,3), c(2,4), c(3,4)), fit=TRUE)
4 iterations: deviation 0.002251441

> cat("deviance(m2.GC)=", m2.GC$lrt, "df(m2.GC)=", m2.GC$df, "\n")
deviance(m2.GC)= 6.565801 df(m2.GC)= 6

> m2.MI <- loglin(table1, margin=list(c(1,2), c(1,3), c(1,4), c(2,3), c(3,4)), fit=TRUE)
4 iterations: deviation 0.0106861

> cat("deviance(m2.MI)=", m2.MI$lrt, "df(m2.MI)=", m2.MI$df, "\n")
deviance(m2.MI)= 10.09911 df(m2.MI)= 6
```

These six models all have 6 degrees of freedom so that the difference of degrees of freedom corresponding to model "m2" is 1 in each case.

```
> 1 - pchisq(m2.GM$lrt - m2$lrt, 1)
```

```

[1] 0.4891588
> 1 - pchisq(m2.MC$lrt - m2$lrt, 1)
[1] 0.02259133
> 1 - pchisq(m2.CI$lrt - m2$lrt, 1)
[1] 0.006515878
> 1 - pchisq(m2.GI$lrt - m2$lrt, 1)
[1] 0.1374742
> 1 - pchisq(m2.GC$lrt - m2$lrt, 1)
[1] 0.1796424
> 1 - pchisq(m2.MI$lrt - m2$lrt, 1)
[1] 0.02091462

```

Testing of the 2-factor interactions shows that the interactions "MC", "CI" and "MI" should be kept in the model. In the next step the model that contains these interactions, G|MC|MI|CI, is fitted.

```

> m2.GM.GI.GC<- loglin(table1, margin=list(c(1), c(2,3), c(2,4), c(3,4)), fit=TRUE)
3 iterations: deviation 0.06111802
> cat("deviance(m2.GM.GI.GC)=", m2.GM.GI.GC$lrt, "df(m2.GM.GI.GC)=", m2.GM.GI.GC$df, "\n")
deviance(m2.GM.GI.GC)= 8.910131 df(m2.GM.GI.GC)= 8
> 1 - pchisq(m2.GM.GI.GC$lrt - m2$lrt, m2.GM.GI.GC$df - m2$df)
[1] 0.2462357

```

Comparison with model "m2" shows that reduction is possible. However, reduction to a model in which the main effect "G" is omitted is rejected.

```

> m2.G<- loglin(table1, margin=list(c(2,3), c(2,4), c(3,4)), fit=TRUE)
3 iterations: deviation 0.06111802
> cat("deviance(m2.G)=", m2.G$lrt, "df(m2.G)=", m2.G$df, "\n")
deviance(m2.G)= 19.42836 df(m2.G)= 9
> 1 - pchisq(m2.G$lrt - m2$lrt, m2.G$df - m2$df)
[1] 0.005453381

```