randomLCA Examples

Ken Beath

July 12, 2008

1 Introduction

Following are two examples of using randomLCA for latent class analysis. Some aspects will certainly change but most code should still work. Two things that will change are the use of accessor functions and better labelling of results.

2 Example 1

This example demonstrates the fitting of data from Rindskopf and Rindskopf (1986), where latent class analysis is used to determine diagnostic classifications based on medical tests. Although this example is for medical data, the model is simply standard latent class so the methods can be applied to data from other areas.

A series of latent class models for 1 to 4 classes can be fitted using the commands

```
> myocardial.lca1 <- randomLCA(myocardial[, 1:4],
+     freq = myocardial$freq, nclass = 1)
> myocardial.lca2 <- randomLCA(myocardial[, 1:4],
+     freq = myocardial$freq, nclass = 2)
> myocardial.lca3 <- randomLCA(myocardial[, 1:4],
+     freq = myocardial$freq, nclass = 3)
```

The BIC values may be extracted from the fitted objects and are shown in Table 1.

```
> bic.data <- data.frame(classes = 1:3, bic = c(myocardial.lca1$bic,
+     myocardial.lca2$bic, myocardial.lca3$bic))
```

classes	bic
1	524.7
2	402.3
3	421.1

Table 1: BIC by class.

Using BIC as a selection method, this selects the 2 class model, indicating a nice beakdown into diseased and nondiseased, which it is assumed represent those with and without myocardial infarction. The true nature of classes is always debateable.

The outcome probabilities can then be obtained, and are shown in Table 2.

```
> outcomep.data <- myocardial.lca2$outcomep
> rownames(outcomep.data) <- c("Class 1", "Class 2")
> colnames(outcomep.data) <- names(myocardial)[1:4]
```

	Q.wave	History	LDH	CPK
Class 1	0.767	0.791	0.828	1.000
Class 2	0.000	0.195	0.027	0.196

Table 2: Outcome Probabilities.

This gives some interesting information. In Class 2, those without myocardial infarction, will have abscence of Q.wave but in those with myocardial infarction it will only be present in 76.7% of those with myocardial infarction. The class probabilities can be obtained as `myocardial.lca2$classp` of 0.46 and 0.54 for Class 1 and 2 respectively.

One aspect of latent class is that no subject is uniquely allocated to a given class, although in some cases a subject may have an extremely high probability.

The class probs can be obtained as

```
> classprobs <- cbind(myocardial.lca2$patterns,
+     myocardial.lca2$classprob)
> colnames(classprobs) <- c(names(myocardial)[1:4],
+     "Class 1", "Class 2")
```

with results shown in Table 3. This shows subjects with 3 or 4 positive tests to be strongly classified as having myocardial infarction, and even some with 2, depending on which to to be well classified. Having only one positive test makes it unlikely that it is myocardial infarction.

Outcome probabilities are shown in Figure 1.

3 Example 2

This example shows the fitting of the dentistry data from Qu, Tan and Kutner (1996). The data consists of the results of five dentists evaluating x-rays for presence or absence of caries. As there is no gold standard, the latent class method is to assume two classes, diseased and non-diseased which are identified from the data.

4 Latent Class

A series of latent class models for 1 to 4 classes can be fitted using the commands

```
> dentistry.lca1 <- randomLCA(dentistry[, 1:5],
+     freq = dentistry$freq, nclass = 1)
> dentistry.lca2 <- randomLCA(dentistry[, 1:5],
```

Q.wave	History	LDH	CPK	Class 1	Class 2
1	1	1	1	1.000	0.000
0	1	1	1	0.992	0.008
1	0	1	1	1.000	0.000
0	0	1	1	0.889	0.111
1	1	0	1	1.000	0.000
0	1	0	1	0.419	0.581
1	0	0	1	1.000	0.000
0	0	0	1	0.044	0.956
1	1	1	0	0.990	0.010
0	1	1	0	0.000	1.000
1	0	1	0	0.865	0.135
0	0	1	0	0.000	1.000
1	1	0	0	0.366	0.634
0	1	0	0	0.000	1.000
1	0	0	0	0.036	0.964
0	0	0	0	0.000	1.000

Table 3: Class Probabilities.

```
+     freq = dentistry$freq, nclass = 2)
> dentistry.lca3 <- randomLCA(dentistry[, 1:5],
+     freq = dentistry$freq, nclass = 3)
> dentistry.lca4 <- randomLCA(dentistry[, 1:5],
+     freq = dentistry$freq, nclass = 4)
```

The BIC values may be extracted from the fitted objects and are shown in Table 4. This indicates the presence of 3 classes. A possible interpretation is that there is a class of subjects with moderate disease, or the alternative of heterogeneous disease which will be covered in the next section. Outcome probabilities are shown in Figure 2.

```
> bic.data <- data.frame(classes = 1:4, bic = c(dentistry.lca1$bic,
+     dentistry.lca2$bic, dentistry.lca3$bic, dentistry.lca4$bic))
```

classes	bic
1	17531.1
2	15021.6
3	14962.9
4	15000.0

Table 4: BIC by class.

5 Latent Class with Random Effects

The method used in Qu, Tan and Kutner (1996) is to introduce a random effect to model heterogeneity within classes. In their model the probabilities are

```
> trellis.par.set(col.whitebg())
> print(plot(myocardial.lca2, type = "l", xlab = "Test",
+     ylab = "Outcome Probability", scales = list(x = list(at = 1:4,
+         labels = names(myocardial)[1:4]))))
```

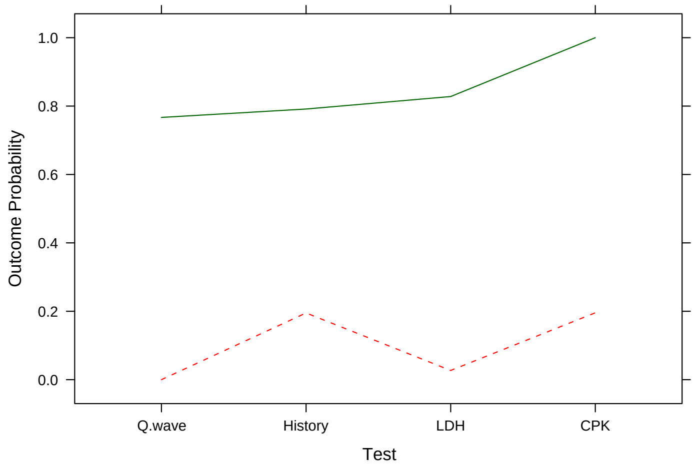

Figure 1: Outcome probabilities for 2 Class Latent Class model.

```
> trellis.par.set(col.whitebg())
> print(plot(dentistry.lca3, type = "l", xlab = "Dentist",
+     ylab = "Outcome Probability"))
```

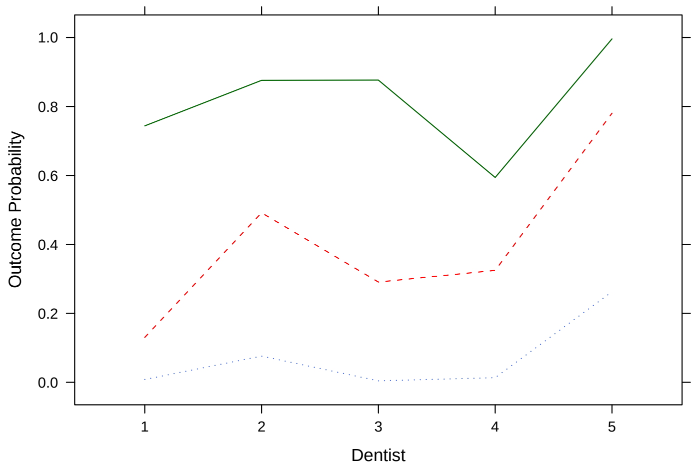

Figure 2: Outcome probabilities for 3 Class Latent Class model.

transformed to the probit scale and then a normal random effect introduced. In practice it usually makes little difference if a probit or logit transform is used.

```
> dentistry.lca2random <- randomLCA(dentistry[,
+     1:5], freq = dentistry$freq, initmodel = dentistry.lca2,
+     nclass = 2, random = TRUE, probit = TRUE)
```

The BIC is reduced to 14944.7 showing an improvement over any of the latent class models. An alternative model is to allow the variance of the random effect to vary by outcome (dentist). This can be performed using the `blocksize` parameter. This allows the structure of the data to be set as a series of blocks, and within each block each outcome has a different loading.

```
> dentistry.lca2random1 <- randomLCA(dentistry[,
+     1:5], freq = dentistry$freq, initmodel = dentistry.lca2random,
+     nclass = 2, random = TRUE, probit = TRUE,
+     blocksize = 5)
```

This increases the BIC to 14949.4, and is the 2LCR model obtained by Qu, Tan and Kutner (1996). It appears that the simpler model is more appropriate.

A further extension is to allow the loading or random effect variance to vary by class.

```
> dentistry.lca2random2 <- randomLCA(dentistry[,
+     1:5], freq = dentistry$freq, initmodel = dentistry.lca2random1,
+     nclass = 2, random = TRUE, probit = TRUE,
+     blocksize = 5, byclass = TRUE, quadpoints = 41)
```

The BIC increases to 14987.6. It is not surprising that this model isn't an improvement, there are now 21 parameters fitted to 32 observations. This also may give problems with the fitting algorithm so the number of quadrature points is increase to 41.

The observed and fitted values can be obtained and are shown in Table 5. Differences from the Que et al paper result from different approximation methods.

```
> obs.data <- data.frame(dentistry.lca2random1$patterns,
+     dentistry.lca2random1$observed, dentistry.lca2$fitted,
+     dentistry.lca2random1$fitted)
> names(obs.data) <- c("V1", "V2", "V3", "V4", "V5",
+     "Obs", "Exp 2LC", "Exp 2LCR")
```

The marginal outcome probabilities, obtained by integrating over the random effect can be plotted, as in Figure 3.

```
> trellis.par.set(col.whitebg())
> print(plot(dentistry.lca2random1, graphtype = "marginal",
+     type = "l", xlab = "Dentist", ylab = "Marginal Outcome Probability"))
```

```
NULL
```

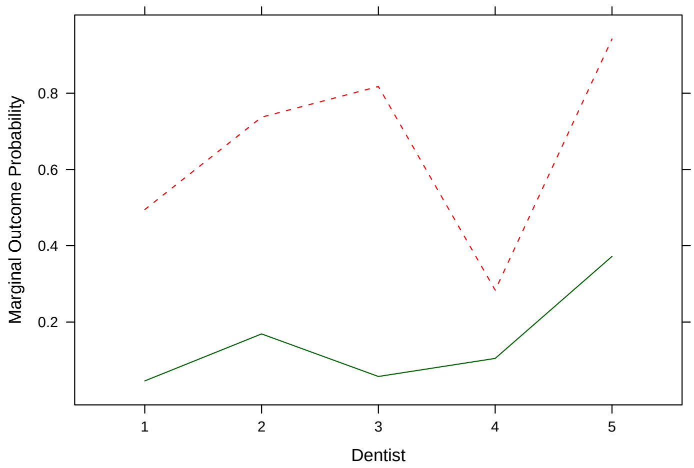

Figure 3: Marginal Outcome Probabilities for 2 Class Latent Class with Random Effect (2LCR) model.

V1	V2	V3	V4	V5	Obs	Exp 2LC	Exp 2LCR
0	0	0	0	0	1880	1836.3	1882.6
0	0	0	0	1	789	830.4	784.7
0	0	0	1	0	43	61.9	38.2
0	0	0	1	1	75	49.6	79.7
0	0	1	0	0	23	28.6	24.2
0	0	1	0	1	63	47.5	63.8
0	0	1	1	0	8	4.0	6.8
0	0	1	1	1	22	35.1	25.8
0	1	0	0	0	188	213.9	184.7
0	1	0	0	1	191	152.2	192.5
0	1	0	1	0	17	12.1	23.1
0	1	0	1	1	67	61.0	67.2
0	1	1	0	0	15	11.2	12.5
0	1	1	0	1	85	91.6	87.4
0	1	1	1	0	8	8.1	7.1
0	1	1	1	1	56	86.4	50.8
1	0	0	0	0	22	21.2	18.5
1	0	0	0	1	26	25.2	27.9
1	0	0	1	0	6	2.1	4.8
1	0	0	1	1	14	16.1	16.0
1	0	1	0	0	1	2.5	2.3
1	0	1	0	1	20	24.7	19.7
1	0	1	1	0	2	2.2	1.8
1	0	1	1	1	17	23.5	14.5
1	1	0	0	0	2	6.0	7.3
1	1	0	0	1	20	42.0	19.8
1	1	0	1	0	6	3.7	4.7
1	1	0	1	1	27	39.3	22.4
1	1	1	0	0	3	5.7	2.7
1	1	1	0	1	72	61.1	69.6
1	1	1	1	0	1	5.4	3.2
1	1	1	1	1	100	58.4	103.0

Table 5: Observed and expected frequencies