A Tutorial on fitting Cumulative Link Mixed Models with `clmm2` from the ordinal Package

Rune Haubo B Christensen

November 16, 2022

Abstract

It is shown by example how a cumulative link mixed model is fitted with the `clmm2` function in package ordinal. Model interpretation and inference is briefly discussed. A tutorial for the more recent `clmm` function is work in progress.

We will consider the data on the bitterness of wine from Randall (1989) presented in Table 1 and available as the object `wine` in package ordinal. The data were also analyzed with mixed effects models by Tutz and Hennevogl (1996). The following gives an impression of the wine data object:

```
R> data(wine)
R> head(wine)

  response rating temp contact bottle judge
1       36      2 cold      no      1     1
2       48      3 cold      no      2     1
3       47      3 cold     yes      3     1
4       67      4 cold     yes      4     1
5       77      4 warm      no      5     1
6       60      4 warm      no      6     1

R> str(wine)

'data.frame':        72 obs. of  6 variables:
 $ response: num  36 48 47 67 77 60 83 90 17 22 ...
 $ rating  : Ord.factor w/ 5 levels "1"<"2"<"3"<"4"<..: 2 3 3 4 4 4 5 5 1 2 ...
 $ temp    : Factor w/ 2 levels "cold","warm": 1 1 1 1 2 2 2 2 1 1 ...
 $ contact : Factor w/ 2 levels "no","yes": 1 1 2 2 1 1 2 2 1 1 ...
 $ bottle  : Factor w/ 8 levels "1","2","3","4",..: 1 2 3 4 5 6 7 8 1 2 ...
 $ judge   : Factor w/ 9 levels "1","2","3","4",..: 1 1 1 1 1 1 1 1 2 2 ...
```

The data represent a factorial experiment on factors determining the bitterness of wine with $1 = $ "least bitter" and $5 = $ "most bitter". Two treatment factors (temperature and contact) each have two levels. Temperature and contact between juice and skins can be controlled when crushing grapes during wine production. Nine judges each assessed wine from two bottles from each of the four treatment conditions, hence there are 72 observations in all. For more information see the manual entry for the wine data: `help(wine)`.

We will fit the following cumulative link mixed model to the wine data:

$$\text{logit}(P(Y_i \leq j)) = \theta_j - \beta_1(\texttt{temp}_i) - \beta_2(\texttt{contact}_i) - u(\texttt{judge}_i)$$
$$i = 1, \ldots, n, \quad j = 1, \ldots, J-1 \tag{1}$$

1

Table 1: Ratings of the bitterness of some white wines. Data are adopted from Randall (1989).

Temperature	Contact	Bottle	Judge 1	2	3	4	5	6	7	8	9
cold	no	1	2	1	2	3	2	3	1	2	1
cold	no	2	3	2	3	2	3	2	1	2	2
cold	yes	3	3	1	3	3	4	3	2	2	3
cold	yes	4	4	3	2	2	3	2	2	3	2
warm	no	5	4	2	5	3	3	2	2	3	3
warm	no	6	4	3	5	2	3	4	3	3	2
warm	yes	7	5	5	4	5	3	5	2	3	4
warm	yes	8	5	4	4	3	3	4	3	4	4

This is a model for the cumulative probability of the ith rating falling in the jth category or below, where i index all observations and $j = 1, \ldots, J$ index the response categories ($J = 5$). $\{\theta_j\}$ are known as threshold parameters or cut-points. We take the judge effects to be random, and assume that the judge effects are IID normal: $u(\texttt{judge}_i) \sim N(0, \sigma_u^2)$.

We fit this model with the `clmm2` function in package ordinal. Here we save the fitted `clmm2` model in the object `fm1` (short for `fitted` `model` `1`) and `print` the model by simply typing its name:

```
R> fm1 <- clmm2(rating ~ temp + contact, random=judge, data=wine)
R> fm1

Cumulative Link Mixed Model fitted with the Laplace approximation

Call:
clmm2(location = rating ~ temp + contact, random = judge, data = wine)

Random effects:
        Var Std.Dev
judge 1.279455 1.13113

Location coefficients:
  tempwarm contactyes
  3.062993   1.834883

No Scale coefficients

Threshold coefficients:
      1|2       2|3       3|4       4|5
-1.623664  1.513364  4.228525  6.088770

log-likelihood: -81.56541
AIC: 177.1308
```

Maximum likelihood estimates of the parameters are provided using the Laplace approximation to compute the likelihood function. A more accurate approximation is provided by the adaptive Gauss-Hermite quadrature method. Here we use 10 quadrature nodes and use the `summary` method to display additional information:

```
R> fm2 <- clmm2(rating ~ temp + contact, random=judge, data=wine,
```

```
            Hess=TRUE, nAGQ=10)
R> summary(fm2)

Cumulative Link Mixed Model fitted with the adaptive Gauss-Hermite
quadrature approximation with 10 quadrature points

Call:
clmm2(location = rating ~ temp + contact, random = judge, data = wine,
    Hess = TRUE, nAGQ = 10)

Random effects:
          Var  Std.Dev
judge 1.287741 1.134787

Location coefficients:
          Estimate Std. Error z value Pr(>|z|)
tempwarm    3.0619   0.5951     5.1454 2.6698e-07
contactyes  1.8334   0.5122     3.5797 0.00034396

No scale coefficients

Threshold coefficients:
    Estimate Std. Error z value
1|2 -1.6235   0.6834    -2.3757
2|3  1.5128   0.6044     2.5028
3|4  4.2271   0.8090     5.2252
4|5  6.0862   0.9719     6.2619

log-likelihood: -81.53246
AIC: 177.0649
Condition number of Hessian: 27.62091
```

The small changes in the parameter estimates show that the Laplace approximation was in fact rather accurate in this case. Observe that we set the option `Hess = TRUE`. This is needed if we want to use the `summary` method since the Hessian is needed to compute standard errors of the model coefficients.

The results contain the maximum likelihood estimates of the parameters:

$$\hat{\beta}_1 = 3.06, \quad \hat{\beta}_2 = 1.83, \quad \hat{\sigma}_u^2 = 1.29 = 1.13^2, \quad \{\hat{\theta}_j\} = [-1.62, \ 1.51, \ 4.23, \ 6.09]. \quad (2)$$

Observe the number under `Std.Dev` for the random effect is **not** the standard error of the random effects variance, `Var`. Rather, it is the standard deviation of the random effects, i.e., it is the square root of the variance. In our example $\sqrt{1.29} \simeq 1.13$.

The condition number of the Hessian measures the empirical identifiability of the model. High numbers, say larger than 10^4 or 10^6 indicate that the model is ill defined. This would indicate that the model can be simplified, that possibly some parameters are not identifiable, and that optimization of the model can be difficult. In this case the condition number of the Hessian does not indicate a problem with the model.

The coefficients for `temp` and `contact` are positive indicating that higher temperature and more contact increase the bitterness of wine, i.e., rating in higher categories is more likely. The odds ratio of the event $Y \geq j$ is $\exp(\beta_{\text{treatment}})$, thus the odds ratio of bitterness being rated in category j or above at warm relative to cold temperatures is

```
R> exp(coef(fm2)[5])
```

```
tempwarm
21.36805
```

The p-values for the location coefficients provided by the `summary` method are based on the so-called Wald statistic. More accurate test are provided by likelihood ratio tests. These can be obtained with the `anova` method, for example, the likelihood ratio test of `contact` is

```
R> fm3 <- clmm2(rating ~ temp, random=judge, data=wine, nAGQ=10)
R> anova(fm3, fm2)
```

```
Likelihood ratio tests of cumulative link models

Response: rating
                Model Resid. df -2logLik    Test    Df LR stat.
1          temp |  |         66 177.4090
2 temp + contact |  |         65 163.0649 1 vs 2     1 14.34409
       Pr(Chi)
1
2 0.0001522572
```

which in this case is slightly more significant. The Wald test is not reliable for variance parameters, so the `summary` method does not provide a test of σ_u, but a likelihood ratio test can be obtained with `anova`:

```
R> fm4 <- clm2(rating ~ temp + contact, data=wine)
R> anova(fm4, fm2)
```

```
Likelihood ratio tests of cumulative link models

Response: rating
                Model Resid. df -2logLik    Test    Df LR stat.
1 temp + contact |  |         66 172.9838
2 temp + contact |  |         65 163.0649 1 vs 2     1 9.918925
      Pr(Chi)
1
2 0.001635879
```

showing that the judge term is significant. Since this test of $\sigma_u = 0$ is on the boundary of the parameter space (a variance cannot be negative), it is often argued that a more correct p-value is obtained by halving the p-value produced by the conventional likelihood ratio test. In this case halving the p-value is of little relevance.

A profile likelihood confidence interval of σ_u is obtained with:

```
R> pr2 <- profile(fm2, range=c(.1, 4), nSteps=30, trace=0)
R> confint(pr2)
```

```
          2.5 %   97.5 %
stDev 0.5048597 2.272327
```

The profile likelihood can also be plotted:

```
R> plot(pr2)
```

The result is shown in Fig. 1 where horizontal lines indicate 95% and 99% confidence bounds. Clearly the profile likelihood function is asymmetric and symmetric confidence intervals would be inaccurate.
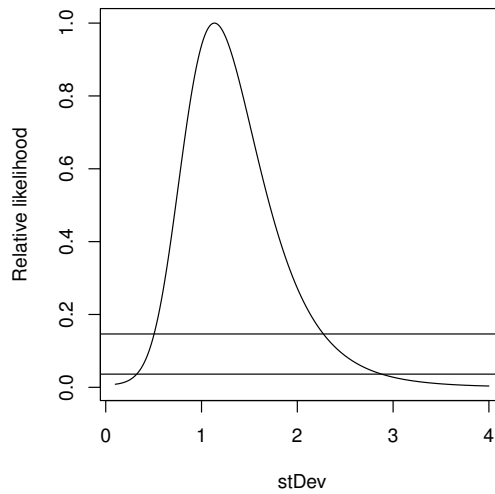
Figure 1: Profile likelihood of σ_u.

The judge effects, $u(\texttt{judge}_i)$ are not parameters, so they cannot be *estimated* in the conventional sense, but a "best guess" is provided by the *conditional modes*. Similarly the *conditional variance* provides an uncertainty measure of the conditional modes. These quantities are included in `clmm2` objects as the `ranef` and `condVar` components. The following code generates the plot in Fig. 2 illustrating judge effects via conditional modes with 95% confidence intervals based on the conditional variance:

```
R> ci <- fm2$ranef + qnorm(0.975) * sqrt(fm2$condVar) %o% c(-1, 1)
R> ord.re <- order(fm2$ranef)
R> ci <- ci[order(fm2$ranef),]
R> plot(1:9, fm2$ranef[ord.re], axes=FALSE, ylim=range(ci),
      xlab="Judge", ylab="Judge effect")
R> axis(1, at=1:9, labels = ord.re)
R> axis(2)
R> for(i in 1:9) segments(i, ci[i,1], i, ci[i, 2])
R> abline(h = 0, lty=2)
```

The seventh judge gave the lowest ratings of bitterness while the first judge gave the highest ratings of bitterness. The significant judge effect indicate that judges perceived the bitterness of the wines differently. Two natural interpretations are that either a bitterness of, say, 3 means different things to different judges, or the judges actually perceived the bitterness of the wines differently. Possibly both effects play their part.

The fitted or predicted probabilites can be obtained with the judge effects at their conditional modes or for an average judge ($u = 0$). The former are available with `fitted(fm)` or with `predict(fm)`, where `fm` is a `fitted` model object. In our example we get

```
R> head(cbind(wine, fitted(fm2)))
```

```
  response rating temp contact bottle judge fitted(fm2)
1       36      2 cold      no      1     1   0.4188998
2       48      3 cold      no      2     1   0.4723982
3       47      3 cold     yes      3     1   0.5499191
4       67      4 cold     yes      4     1   0.2607227
```
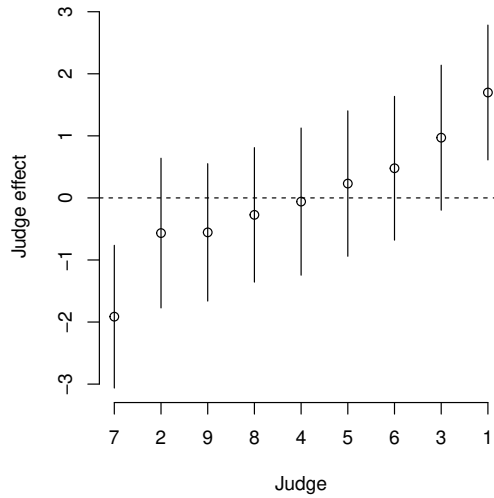
5

Figure 2: Judge effects given by conditional modes with 95% confidence intervals based on the conditional variance.

```
5        77       4 warm       no         5       1   0.4203948
6        60       4 warm       no         6       1   0.4203948
```

Predicted probabilities for an average judge can be obtained by including the data used to fit the model in the `newdata` argument of `predict`:

```
R> head(cbind(wine, pred=predict(fm2, newdata=wine)))
```

```
  response rating temp contact bottle judge       pred
1       36      2 cold      no      1     1 0.6547512
2       48      3 cold      no      2     1 0.1661397
3       47      3 cold     yes      3     1 0.4958192
4       67      4 cold     yes      4     1 0.0696343
5       77      4 warm      no      5     1 0.1913894
6       60      4 warm      no      6     1 0.1913894
```

Model (1) says that for an average judge at cold temperature the cumulative probability of a bitterness rating in category j or below is

$$P(Y_i \leq j) = \text{logit}^{-1}[\theta_j - \beta_2(\texttt{contact}_i)]$$

since u is set to zero and $\beta_1(\texttt{temp}_i) = 0$ at cold conditions. Further, $\text{logit}^{-1}(\eta) = 1/[1 + \exp(\eta)]$ is the cumulative distribution function of the logistic distribution available as the `plogis` function. The (non-cumulative) probability of a bitterness rating in category j is $\pi_j = P(Y_i \leq j) - P(Y_i \leq j - 1)$, for instance the probability of a bitterness rating in the third category at these conditions can be computed as

```
R> plogis(fm2$Theta[3] - fm2$beta[2]) -
   plogis(fm2$Theta[2] - fm2$beta[2])

      3|4
0.4958192
```

This corresponds to the third entry of `predict(fm2, newdata=wine)` given above.

6

Judge effects are random and normally distributed, so an average judge effect is 0. Extreme judge effects, say 5th and 95th percentile judge effects are given by

```
R> qnorm(0.95) * c(-1, 1) * fm2$stDev

[1] -1.866558  1.866558
```

At the baseline experimental conditions (cold and no contact) the probabilites of bitterness ratings in the five categories for a 5th percentile judge is

```
R> pred <-
   function(eta, theta, cat = 1:(length(theta)+1), inv.link = plogis)
 {
   Theta <- c(-1e3, theta, 1e3)
   sapply(cat, function(j)
          inv.link(Theta[j+1] - eta) - inv.link(Theta[j] - eta) )
 }
R> pred(qnorm(0.05) * fm2$stDev, fm2$Theta)

          1|2            2|3            3|4            4|5
0.5604689972 0.4065839967 0.0306948517 0.0019005715 0.0003515829
```

We can compute these probabilities for average, 5th and 95th percentile judges at the four experimental conditions. The following code plots these probabilities and the results are shown in Fig. 3.

```
R> mat <- expand.grid(judge = qnorm(0.95) * c(-1, 0, 1) * fm2$stDev,
                      contact = c(0, fm2$beta[2]),
                      temp = c(0, fm2$beta[1]))
R> pred.mat <- pred(eta=rowSums(mat), theta=fm2$Theta)
R> lab <- paste("contact=", rep(levels(wine$contact), 2), ", ",
               "temp=", rep(levels(wine$temp), each=2), sep="")
R> par(mfrow=c(2, 2))
R> for(k in c(1, 4, 7, 10)) {
   plot(1:5, pred.mat[k,], lty=2, type = "l", ylim=c(0,1),
        xlab="Bitterness rating scale", axes=FALSE,
        ylab="Probability", main=lab[ceiling(k/3)], las=1)
   axis(1); axis(2)
   lines(1:5, pred.mat[k+1, ], lty=1)
   lines(1:5, pred.mat[k+2, ], lty=3)
   legend("topright",
          c("avg. judge", "5th %-tile judge", "95th %-tile judge"),
          lty=1:3, bty="n")
 }
```

At constant experimental conditions the odds ratio for a bitterness rating in category j or above for a 95th percentile judge relative to a 5th percentile judge is

```
R> exp(2*qnorm(0.95) * fm2$stDev)

   judge
41.80921
```

The differences between judges can also be expressed in terms of the interquartile range: the odds ratio for a bitterness rating in category j or above for a third quartile judge relative to a first quartile judge is

```
R> exp(2*qnorm(0.75) * fm2$stDev)

   judge
```
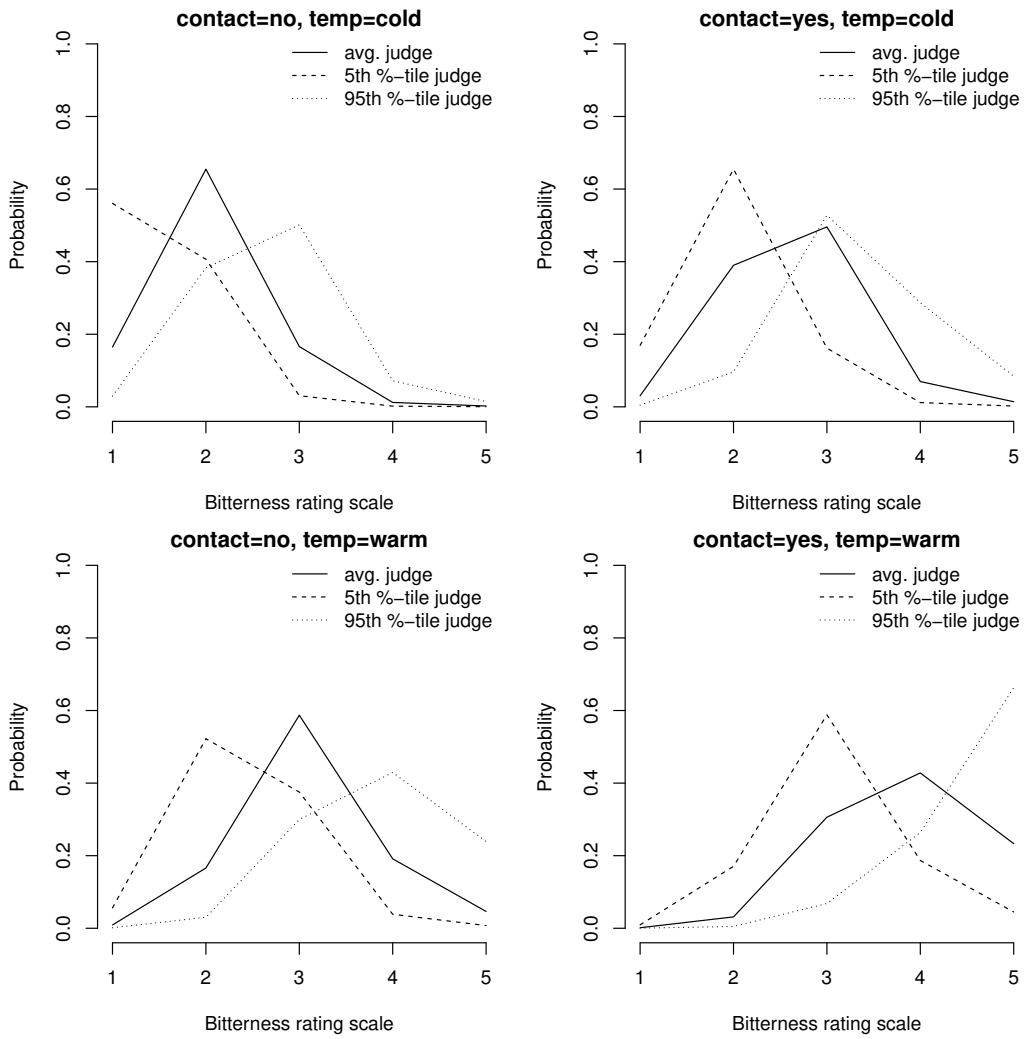
Figure 3: Rating probabilities for average and extreme judges at different experimental conditions.

4.621893

References

Randall, J. (1989). The analysis of sensory data by generalised linear model. *Biometrical journal 7*, 781–793.

Tutz, G. and W. Hennevogl (1996). Random effects in ordinal regression models. *Computational Statistics & Data Analysis 22*, 537–557.