

# A Tutorial on fitting Cumulative Link Models with the **ordinal** Package

Rune Haubo B Christensen

August 30, 2011

## Abstract

It is shown by example how a cumulative link mixed model is fitted with the `clm` function in package **ordinal**. Model interpretation and inference is briefly discussed.

## 1 Introduction

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:

```
> data(wine)
> 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

> 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)`.

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

## 2 Fitting Cumulative Link Models

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

$$\text{logit}(P(Y_i \leq j)) = \theta_j - \beta_1(\text{temp}_i) - \beta_2(\text{contact}_i) \quad (1)$$

$$i = 1, \dots, n, \quad j = 1, \dots, J - 1$$

This is a model for the cumulative probability of the  $i$ th rating falling in the  $j$ th category or below, where  $i$  index all observations and  $j = 1, \dots, J$  index the response categories ( $J = 5$ ).  $\{\theta_j\}$  are known as threshold parameters, intercepts or cut-points.

This model is also known as the *proportional odds model*, a *cumulative logit model*, and an *ordered logit model*.

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

```
> fm1 <- clm(rating ~ temp + contact, data = wine)
> fm1
```

```
formula: rating ~ temp + contact
data:    wine
```

```
link threshold nobs logLik AIC      niter max.grad
logit flexible  72   -86.49 184.98 6(0)  4.01e-12
```

Coefficients:

```
tempwarm contactyes
  2.503      1.528
```

Threshold coefficients:

```
1|2  2|3  3|4  4|5
-1.344 1.251 3.467 5.006
```

Additional information is provided with the `summary` method:

```
> summary(fm1)
```

```
formula: rating ~ temp + contact
data:    wine
```

```
link threshold nobs logLik AIC      niter max.grad cond.H
logit flexible 72   -86.49 184.98 6(0)  4.01e-12 2.7e+01
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
tempwarm	2.5031	0.5287	4.735	2.19e-06	***
contactyes	1.5278	0.4766	3.205	0.00135	**

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Threshold coefficients:

	Estimate	Std. Error	z value
1 2	-1.3444	0.5171	-2.600
2 3	1.2508	0.4379	2.857
3 4	3.4669	0.5978	5.800
4 5	5.0064	0.7309	6.850

The primary result is the coefficient table with parameter estimates, standard errors and Wald (or normal) based  $p$ -values for tests of the parameters being zero. The maximum likelihood estimates of the parameters are:

$$\hat{\beta}_1 = 2.50, \quad \hat{\beta}_2 = 1.53, \quad \{\hat{\theta}_j\} = [-1.34, 1.25, 3.47, 5.01]. \quad (2)$$

The number of Newton-Raphson iterations is given below **niter** with the number of step-halvings in parenthesis. A large number of iteration, say, larger than approximately 10 indicates that the model was difficult to fit.

**max.grad** is the maximum absolute gradient of the log-likelihood function with respect to the parameters. A small absolute gradient is a necessary condition for convergence of the model. The iterative procedure will declare convergence whenever the maximum absolute gradient is below

```
> clm.control()$gradTol
```

```
[1] 1e-06
```

which may be altered — see **help(clm.control)**.

The condition number of the Hessian (**cond.H**) 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 could 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

```
> exp(coef(fm1)[5])
```

```
tempwarm
12.22034
```

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

```
> fm2 <- clm(rating ~ temp, data = wine)
> anova(fm2, fm1)
```

Likelihood ratio tests of cumulative link models:

	formula:	link:	threshold:
fm2	rating ~ temp	logit	flexible
fm1	rating ~ temp + contact	logit	flexible

  

	no.par	AIC	logLik	LR.stat	df	Pr(>Chisq)
fm2	5	194.03	-92.013			
fm1	6	184.98	-86.492	11.043	1	0.0008902 ***

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

which in this case produce a slightly lower  $p$ -value. Equivalently we can use `drop1` to obtain likelihood ratio tests of the explanatory variables while *controlling* for the remaining variables:

```
> drop1(fm1, test = "Chi")
```

Single term deletions

Model:  
rating ~ temp + contact

	Df	AIC	LRT	Pr(Chi)
<none>		184.98		
temp	1	209.91	26.928	2.112e-07 ***
contact	1	194.03	11.043	0.0008902 ***

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Likelihood ratio tests of the explanatory variables while *ignoring* the remaining variables are provided by the `add1` method:

```
> fm0 <- clm(rating ~ 1, data = wine)
> add1(fm0, scope = ~temp + contact, test = "Chi")
```

Single term additions

Model:  
rating ~ 1

	Df	AIC	LRT	Pr(Chi)
<none>		215.44		
temp	1	194.03	23.4113	1.308e-06 ***
contact	1	209.91	7.5263	0.00608 **

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

In this case these latter tests are not as strong as the tests controlling for the other variable.

Confidence intervals are provided by the `confint` method:

```
> confint(fm1)
```

2.5 % 97.5 %

```
tempwarm 1.5097627 3.595225
contactyes 0.6157925 2.492404
```

These are based on the profile likelihood function and generally fairly accurate. Less accurate, but simple and symmetric confidence intervals based on the standard errors of the parameters (so-called Wald confidence intervals) can be obtained with

```
> confint.default(fm1)

                2.5 %      97.5 %
1|2          -2.3578848 -0.330882
2|3           0.3925794  2.109038
3|4           2.2952980  4.638476
4|5           3.5738541  6.438954
tempwarm      1.4669081  3.539296
contactyes     0.5936345  2.461961
```

In addition to the logit link, which is the default, the probit, log-log, complementary log-log and cauchit links are also available. For instance, a proportional hazards model for grouped survival times is fitted using the complementary log-log link:

```
> fm.cll <- clm(rating ~ contact + temp, data = wine, link = "cloglog")
```

The cumulative link model in (1) assume that the thresholds,  $\{\theta_j\}$  are constant for all values of the remaining explanatory variables, here `temp` and `contact`. This is generally referred to as the *proportional odds assumption* or *equal slopes assumption*, while both names are not entirely adequate since the assumption is also made for other links than the logit and for categorical variables as well as continuous ones. We can relax that assumption in two general ways: with nominal effects and scale effects which we will now discuss in turn:

### 3 Nominal Effects

The CLM in (1) specifies a structure in which the regression parameters,  $\beta$  are not allowed to vary with  $j$  as are the threshold parameters,  $\theta$ . Nominal effects relax this assumption by allowing one or more regression parameters to vary with  $j$ . In the following model we allow the regression parameter for `contact` to vary with  $j$ :

$$\text{logit}(P(Y_i \leq j)) = \theta_j - \beta_1(\text{temp}_i) - \beta_{2j}(\text{contact}_i) \quad (3)$$

$$i = 1, \dots, n, \quad j = 1, \dots, J - 1$$

This means that there is one estimate of  $\beta_2$  for each  $j$ , i.e., for each of the five thresholds in addition to the estimates of  $\theta_j$  for each  $j$ . This model is specified as follows in `clm`:

```
> fm.nom <- clm(rating ~ temp, nominal = ~contact, data = wine)
> summary(fm.nom)
```

```
formula: rating ~ temp
nominal: ~contact
data:    wine
```

```
link threshold nobs logLik AIC      niter max.grad cond.H
logit flexible  72   -86.21 190.42 6(0)  1.64e-10 4.8e+01
```

Coefficients:

```
Estimate Std. Error z value Pr(>|z|)
```

```
tempwarm    2.519      0.535   4.708  2.5e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Threshold coefficients:

	Estimate	Std. Error	z value
1 2.(Intercept)	-1.3230	0.5623	-2.353
2 3.(Intercept)	1.2464	0.4748	2.625
3 4.(Intercept)	3.5500	0.6560	5.411
4 5.(Intercept)	4.6602	0.8604	5.416
1 2.contactyes	-1.6151	1.1618	-1.390
2 3.contactyes	-1.5116	0.5906	-2.559
3 4.contactyes	-1.6748	0.6488	-2.581
4 5.contactyes	-1.0506	0.8965	-1.172

As can be seen from the output of `summary` there is no regression coefficient estimated for `contact`, but there are two sets of threshold parameters estimated.

The first five threshold parameters have `.(Intercept)` appended their names indicating that these are the estimates of  $\theta_j$ . The following five threshold parameters have `.contactyes` appended their name indicating that these parameters are differences between the threshold parameters at the two levels of contact. This interpretation corresponds to the default treatment contrasts; if other types of contrasts are specified, the interpretation is a little different. As can be seen from the output, the effect of `contact` is almost constant across thresholds and around 1.5 corresponding to the estimate from `fm1` on page 3, so probably there is not much evidence that the effect `contact` varies with  $j$ .

We can perform a likelihood ratio test of the equal slopes or proportional odds assumption for `contact` by comparing the likelihoods of models (1) and (3) as follows:

```
> anova(fm1, fm.nom)
```

Likelihood ratio tests of cumulative link models:

	formula:	nominal:	link:	threshold:
fm1	rating ~ temp + contact	~1	logit	flexible
fm.nom	rating ~ temp	~contact	logit	flexible

  

	no.par	AIC	logLik	LR.stat	df	Pr(>Chisq)
fm1	6	184.98	-86.492			
fm.nom	9	190.42	-86.209	0.5667	3	0.904

There is only little difference in the log-likelihoods of the two models, so the test is insignificant. There is therefore no evidence that the proportional odds assumption is violated for `contact`.

It is not possible to estimate both  $\beta_2$  and  $\beta_{2j}$  in the same model. Consequently variables that appear in `nominal` cannot enter elsewhere as well. For instance not all parameters are identifiable in the following model:

```
> fm.nom2 <- clm(rating ~ temp + contact, nominal = ~contact, data = wine)
```

We are made aware of this when summarizing or printing the model:

```
> summary(fm.nom2)
```

```
formula: rating ~ temp + contact
nominal: ~contact
```

```

data:      wine

link threshold nobs logLik AIC      niter max.grad cond.H
logit flexible 72    -86.21 190.42 6(0)  1.64e-10 4.8e+01

Coefficients: (1 not defined because of singularities)
              Estimate Std. Error z value Pr(>|z|)
tempwarm      2.519      0.535    4.708 2.5e-06 ***
contactyes     NA         NA        NA      NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Threshold coefficients:
              Estimate Std. Error z value
1|2.(Intercept) -1.3230      0.5623  -2.353
2|3.(Intercept)  1.2464      0.4748   2.625
3|4.(Intercept)  3.5500      0.6560   5.411
4|5.(Intercept)  4.6602      0.8604   5.416
1|2.contactyes  -1.6151      1.1618  -1.390
2|3.contactyes  -1.5116      0.5906  -2.559
3|4.contactyes  -1.6748      0.6488  -2.581
4|5.contactyes  -1.0506      0.8965  -1.172

```

## 4 Scale Effects

Scale effects are usually motivated from the latent variable interpretation of a CLM. Assume the following model for a latent variable:

$$S_i = \alpha + \mathbf{x}_i^T \boldsymbol{\beta}^* + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^{*2}) \quad (4)$$

If the ordinal variable  $Y_i$  is observed such that  $Y_i = j$  is recorded if  $\theta_{j-1}^* < S_i \leq \theta_j^*$ , where

$$-\infty \equiv \theta_0^* < \theta_1^* < \dots < \theta_{J-1}^* < \theta_J^* \equiv \infty \quad (5)$$

then we have the cumulative link model for  $Y_i$ :

$$P(Y_i \leq j) = \Phi(\theta_j - \mathbf{x}_i^T \boldsymbol{\beta})$$

where we have used  $\theta_j = (\theta_j^* + \alpha)/\sigma$  and  $\boldsymbol{\beta} = \boldsymbol{\beta}^*/\sigma$ , and  $\Phi$  is the inverse probit link and denotes the standard normal CDF. Other assumptions on the distribution of the latent variable,  $S_i$  lead to other link functions.

If the scale (or dispersion) of the latent distribution is described by a log-linear model such that  $\log(\sigma_i) = \mathbf{z}_i^T \boldsymbol{\zeta}$ , then the resulting CLM reads:

$$P(Y_i \leq j) = \Phi\left(\frac{\theta_j - \mathbf{x}_i^T \boldsymbol{\beta}}{\sigma_i}\right) \quad (6)$$

Thus allowing for scale effects corresponds to modelling not only the location of the latent distribution, but also the scale. If the link function is the **probit**, the corresponding latent distribution is Gaussian and the location equals the mean and scale or dispersion equals

the spread or standard deviation. Just as the absolute location ( $\alpha$ ) is not identifiable, the absolute scale ( $\sigma$ ) is not identifiable either in the CLM.

We can allow the scale of the latent distribution to depend on temperature in the following way:

$$\text{logit}(P(Y_i \leq j)) = \frac{\theta_j - \beta_1(\text{temp}_i) - \beta_2(\text{contact}_i)}{\exp(\zeta_1(\text{temp}_i))} \quad (7)$$

$$i = 1, \dots, n, \quad j = 1, \dots, J - 1$$

We can estimate this model with

```
> fm.sca <- clm(rating ~ temp + contact, scale = ~temp, data = wine)
> summary(fm.sca)
```

```
formula: rating ~ temp + contact
scale:    ~temp
data:     wine
```

```
link threshold nobis logLik AIC      niter max.grad cond.H
logit flexible  72    -86.44 186.88 8(0)  5.25e-09 1.0e+02
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
tempwarm	2.6294	0.6860	3.833	0.000127	***
contactyes	1.5878	0.5301	2.995	0.002743	**

---

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

log-scale coefficients:

	Estimate	Std. Error	z value	Pr(> z )
tempwarm	0.09536	0.29414	0.324	0.746

Threshold coefficients:

	Estimate	Std. Error	z value
1 2	-1.3520	0.5223	-2.588
2 3	1.2730	0.4533	2.808
3 4	3.6170	0.7774	4.653
4 5	5.2982	1.2027	4.405

Notice that both location and scale effects of `temp` are identifiable. Also notice that the scale coefficient for `temp` is given on the log-scale, where the Wald test is more appropriate. The absolute scale of the latent distribution is not estimable, but we can estimate the scale at warm conditions relative to cold conditions. Therefore the estimate of  $\kappa$  in the relation  $\sigma_{warm} = \kappa \sigma_{cold}$  is given by

```
> exp(fm.sca$zeta)
```

```
tempwarm
1.100054
```

However, the scale difference is not significant in this case as judged by the  $p$ -value in the summary output. `confint` and `anova` apply with no change to models without scale, but `drop1`, `add1` and `step` methods will only drop or add terms to the (location) formula and not to scale.



## 5 Predictions

Fitted values are extracted with e.g., `fitted(fm1)` and produce fitted probabilities, i.e., the  $i$ th fitted probability is the probability that the  $i$ th observation falls in the response category that it did. The predictions of which response class the observations would be most likely to fall in given the model are obtained with:

```
> predict(fm1, type = "class")

$fit
 [1] 2 2 3 3 3 3 4 4 2 2 3 3 3 3 4 4 2 2 3 3 3 3 4 4 2 2 3 3 3 3 4 4 2 2 3 3 3 3
[39] 4 4 2 2 3 3 3 3 4 4 2 2 3 3 3 3 4 4 2 2 3 3 3 3 4 4 2 2 3 3 3 3 4 4
Levels: 1 2 3 4 5
```

Say we just wanted the predictions for the four combinations of `temp` and `contact`. The probability that an observation falls in each of the five response categories based on the fitted model is given by:

```
> newData <- expand.grid(temp = levels(wine$temp), contact = levels(wine$contact))
> cbind(newData, predict(fm1, newdata = newData)$fit)

  temp contact      1      2      3      4      5
1 cold      no 0.206790132 0.57064970 0.1922909 0.02361882 0.00665041
2 warm      no 0.020887709 0.20141572 0.5015755 0.20049402 0.07562701
3 cold     yes 0.053546010 0.37764614 0.4430599 0.09582084 0.02992711
4 warm     yes 0.004608274 0.05380128 0.3042099 0.36359581 0.27378469
```

Standard errors and confidence intervals of predictions are also available, for example, the predictions, standard errors and 95% confidence intervals for the first six observations are given by

```
> head(do.call("cbind", predict(fm1, se.fit = TRUE, interval = TRUE)))

      fit      se.fit      lwr      upr
[1,] 0.57064970 0.08683884 0.56100370 0.58024235
[2,] 0.19229094 0.06388672 0.18751955 0.19715427
[3,] 0.44305990 0.07939754 0.43393644 0.45222195
[4,] 0.09582084 0.04257593 0.09403264 0.09763938
[5,] 0.20049402 0.06761012 0.19528829 0.20580304
[6,] 0.20049402 0.06761012 0.19528829 0.20580304
```

The confidence level can be set with the `level` argument and other types of predictions are available with the `type` argument.

Model identifiability:

If we attempt to test the proportional odds assumption for `temp`, some peculiarities show up:

```
> fm.nom2 <- clm(rating ~ contact, nominal = ~temp, data = wine)
> summary(fm.nom2)
```

```
formula: rating ~ contact
nominal: ~temp
data:    wine
```

```
link threshold nobs logLik AIC      niter max.grad cond.H
logit flexible  72    -84.90 187.81 16(0) 6.45e-07 2.8e+08
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
contactyes	1.4652	0.4688	3.125	0.00178 **

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Threshold coefficients:

	Estimate	Std. Error	z value
1 2.(Intercept)	-1.2656	0.5149	-2.458
2 3.(Intercept)	1.1040	0.4388	2.516
3 4.(Intercept)	3.7657	0.8046	4.680
4 5.(Intercept)	19.8964	2145.3568	0.009
1 2.tempwarm	-16.0954	1245.0152	-0.013
2 3.tempwarm	-2.1530	0.5898	-3.650
3 4.tempwarm	-2.8733	0.8174	-3.515
4 5.tempwarm	-17.5500	2145.3569	-0.008

Observe that several of the threshold coefficients are extremely large with huge standard errors. Also the condition number of the Hessian is very large. These are all indications that the model is not identifiable. The problem is revealed with

```
> with(wine, table(temp, rating))
```

	rating				
temp	1	2	3	4	5
cold	5	16	13	2	0
warm	0	6	13	10	7

where the zeros indicate the problem. Nevertheless we can still test the proportional odds assumption for temp:

```
> anova(fm1, fm.nom2)
```

Likelihood ratio tests of cumulative link models:

	formula:	nominal:	link:	threshold:
fm1	rating ~ temp + contact	~1	logit	flexible
fm.nom2	rating ~ contact	~temp	logit	flexible

	no.par	AIC	logLik	LR.stat	df	Pr(>Chisq)
fm1	6	184.98	-86.492			
fm.nom2	9	187.81	-84.904	3.175	3	0.3654

## References

- Randall, J. (1989). The analysis of sensory data by generalised linear model. *Biometrical journal* 7, pp. 781–793.
- Tutz, G. and W. Hennevogl (1996). Random effects in ordinal regression models. *Computational Statistics & Data Analysis* 22, pp. 537–557.