

# Marginalized predictions in the Generalized Mixed Models using **marPredict** package (v0.0.2)

Maciej J. Dańko<sup>\*1</sup>

<sup>1</sup>*Max Planck Institute for Demographic Research*

## Abstract

to be added

## 1 Introduction

### 1.1 Generalized linear mixed effect models

The linear mixed models are defined as

$$y = X\beta + Zu + \epsilon \quad (1)$$

where,  $y$  is a vector of observations (dependent variable),  $\beta$  and  $u$  are vectors of unknown the fixed effect and the random effect coefficients,  $X$  and  $Z$  are design matrices relating the observed  $y$  to the  $\beta$  and the  $u$ , and  $\epsilon$  is an unknown vector of the random terms. It is assumed that expected values of  $u$  and  $\epsilon$  are equal to zero ( $E(u) = E(\epsilon) = 0$ ), hence the expectation of the response  $E(y)$  is given by:

$$E(y) = X\beta + Zu \quad (2)$$

In the generalized linear mixed models, the response variable  $y$  may have different distribution than normal. The models use a link function  $g$  to transform the expectation of the response  $E(y)$  to the linear predictor  $\eta$ :

$$g(E(y)) = \eta = X\beta + Zu \quad (3)$$

Similarly as in eq. 1,  $\eta$  is a linear function of regressors and random effects. Any link function must be invertible, so that  $E(y)$  can be calculated.

$$E(y) = g^{-1}(X\beta + Zu) \quad (4)$$

### 1.2 Predictions in generalized linear mixed models

In the generalized linear models making the predictions is straightforward. Once the vector of coefficients  $\hat{\beta}$  is estimated it can be used together with a new design matrix  $X_p$  to predict the marginal responses at particular value of covariates:

$$E(y_p) = g^{-1}(X_p\hat{\beta}) \quad (5)$$

In the generalized linear mixed models the predictions are more problematic because of the presence of  $Z_p$ .

---

<sup>\*</sup>Corresponding author: danko@demogr.mpg.de, maciej.danko@gmail.com

$$E(y_p) = g^{-1}(X_p\hat{\beta} + Z_p\hat{u}) \quad (6)$$

The typical approach used across different packages is either to (1) ignore  $Z_p\hat{u}$  (set it to zero) or (2) construct matrix  $Z_p$  for a particular grouping level of random variables such that matrices  $X_p$  and  $Z_p$  have the same number of rows.

Let us consider an exemplary data for a clinical trial of two competing oral antifungal treatments for the toenail infection (HSAUR2 package; Everitt and Hothorn (2017)), with `patientID` as random effect. The data can be fitted using `glmer` function of the `lme4` package (Bates et al. 2015)

```
# install packages
if(!require(lme4)) {install.packages('lme4'); library(lme4)}
if(!any(installed.packages()[,1]=='HSAUR2')) install.packages('HSAUR2')

# load the clinical trial data for the toenail infection
data1 <- HSAUR2::toenail

# fit the generalized linear mixed model to binary outcome data
model1 <- glmer(outcome ~ treatment * visit+(1|patientID),
               data=data1, family=binomial, nAGQ=25)
```

As stated above, one of the way of model predictions is to exclude random terms. This can be done by calling a generic `predict` function with parameter `re.form = NA` for a model fitted with the `glmer` function (`lme4` package), or with parameter `level = 0` for model fitted with the `lme` function (`nlme` package (???); gaussian family only). The syntax may differ for other packages.

```
newdata <- expand.grid(visit=c(1,4,7), treatment=levels(data1$treatment))
data.frame(newdata, predict = predict(model1, newdata=newdata,
                                     type='response', re.form=NA))
```

	visit	treatment	predict
1	1	itraconazole	0.2240253938
2	4	itraconazole	0.0261602678
3	7	itraconazole	0.0024933054
4	1	terbinafine	0.2107740970
5	4	terbinafine	0.0120927345
6	7	terbinafine	0.0005607344

Here I consider 1<sup>st</sup>, 4<sup>th</sup>, and 7<sup>th</sup> visit in combination of two treatments. The predictions are performed on the scale of response variable (`type = 'response'`).

The alternative approach is to include the random effects for particular subjects or group of subjects. In `glmer` it is done by setting `re.form = NULL` or by specifying the random effect formula explicitly.

```
newdata <- data.frame(newdata, patientID=1:2)
data.frame(newdata, predict = predict(model1, newdata=newdata,
                                     type='response', re.form=~(1|patientID)))
```

	visit	treatment	patientID	predict
1	1	itraconazole	1	0.915633988
2	4	itraconazole	2	0.182607465
3	7	itraconazole	1	0.085893643
4	1	terbinafine	2	0.689537664
5	4	terbinafine	1	0.315145593
6	7	terbinafine	2	0.004644215

### 1.3 Marginalized model predictions conditional to the estimated population random effects

The marginalized predictions are alternative option

Here I present the alternative approach, which calculates model predictions conditionally to the estimated population random effects  $\hat{R} = Z\hat{u}$ . In this approach I define  $\hat{R}_c$  as a vector of all elements of  $\hat{R}$  that belongs to a specific combination  $c$  of categorical variable levels. Each  $c$  correspond to a single row of  $X_p$  design matrix. The mechanism of grouping  $\hat{R}$  into  $\hat{R}_c$  is demonstrated in **Tab. 1**.

**Table 1.** Demonstration of grouping  $\hat{R}$  into  $\hat{R}_c$  for a data consistent of two 3-level categorical variables (X1 and X2). The estimated random effects are grouped according to combinations of variable levels  $c$ .  $n_c$  is a number of individual random effects in a combination  $c$  and  $\hat{P}_c$  is a element of the vector  $\hat{P} = X_p\hat{\beta}$  corresponding to a particular  $c$ .

X1 level	X2 level	$c$	$\hat{R}_c$	$n_c$	$\hat{P}_c$
a	A	aA	{0.4567, 0.1456, -0.4537, -1.4561}	4	2.4567
b	A	bA	{0.1572, 1.2645}	2	3.8761
c	A	cA	{0.4464, 0.1455, -0.4537, -1.4561}	4	5.1411
a	B	aB	{0.4567, 0.1456, -1.4561}	3	1.0808
b	B	bB	{0.1572, 1.2645, -0.4537, 0.4567}	4	3.6432
c	B	cB	{0.1572, 1.2645}	2	2.2221
a	C	aC	{0.4464}	1	3.2512
b	C	bC	{0.4567, 0.1456, -0.4537, -1.4561}	4	3.1235
c	C	cC	{0.4567, 1.2645, -0.4537, -1.4561}	4	3.5378

The next step is to calculate averaged fixed effects for each  $c$ :

$$E(y_p(c)) = \sum_i^{n_c} g^{-1} \left( \hat{P}_c + \hat{R}_c(i) \right) / n_c \quad (7)$$

If there is no categorical variable in  $X_p$  then  $P_c = P$  and  $R_c = R$ .

### 1.4 Averaging fixed effects conditionally to the missing variables

The `predict` function requires user to specify all variables used for model fitting in the `newdata` data frame. Yet, I propose an alternative approach, which use the idea of averaging presented in previous section. The effect of missing variables is coded as  $S = X\hat{\beta}_m$ , where  $X$  is design matrix for a total population, and  $\hat{\beta}_m$  is a vector of model coefficients, which has inserted 0 for every coefficient related to variables included in the `newdata`.

**Table 2.** Demonstration of calculation of the  $\hat{\beta}_m$  and its complement  $\hat{\beta}_d$  for a model  $Y \sim X1 + X2 + C1 + X1:X2$ , where X1 and X2 are two-level categorical variables, C1 is a continuous variable, and ‘:’ denotes interaction. It is assumed that only X1 and C1 variables are present in the `newdata` data frame.

	(Intercept)	X1b	X2B	C1	X1b:X2B
$\hat{\beta}$	4.0122	0.1234	2.0567	0.1645	-0.8876
$\hat{\beta}_m$	0	0	2.0567	0	-0.8876
$\hat{\beta}_d$	4.0122	0.1234	0	0.1645	0

The complement  $\hat{\beta}_d$  is then used to calculate predictions for variables present in the **newdata**:  $\hat{Q} = X_p \hat{\beta}_d$ . Finally both predictions can be combined in similar way as it was done for random effects (see eq. 7 and **Tab. 1**):

$$E(y_p(c)) = \sum_i^{n_c} g^{-1} \left( \hat{Q}_c + \hat{S}_c(i) \right) / n_c \quad (8)$$

## 1.5 Offset and rates

Offset  $\xi$  is a explanatory variable of a known effect. The offset is not estimated, but its values are simply added to the linear predictor.

$$g(E(y)) = \eta = X\beta + \xi \quad (9)$$

The offset is especially useful in GLMM when dependent variable is a count. In such a case offset variable is used to handle the rates and an offset  $\xi_i$  of observation  $i$  is related to its exposure  $e_i$  via link function  $\xi_i = g(e_i)$ , where  $g$  is typically a  $\log$  function.

When the offset is used during estimation of a model, it should be also included in the model predictions. To include the offset for each combination of factors and their levels  $c$ , the classification method described in previous sections can be adopted.

$$E(y_p(c)) = \sum_i^{n_c} g^{-1} \left( \hat{P}_c + \xi_c(i) \right) / n_c \quad (10)$$

To calculate rates, sum of counts are dived by sum of exposures for each  $c$ .

$$E(y_p(c)) = \sum_i^{n_c} g^{-1} \left( \hat{P}_c + \xi_c(i) \right) / \sum_i^{n_c} g^{-1} \xi(i) \quad (11)$$

## 1.6 Fast simulation of confidence intervals

The similar approach is used in Dańko (2019) and Knowles and Frederick (2019) package

## 2 The marPredict package

The **marPredict** package delivers a function for model predictions that extends functionality of the default **predict** function by adding:

- Marginalized random effects: incorporating conditional average random effects to model predictions, which are unbiased both on the predictor and response scale.
- Marginalized variable effects: enabling possiblty for addindg average effect of the variables not included in the **newdata** data frame
- Marginalized offset: incorporating offset averaging to model predictions and allowing the correct prediction of rates.
- using fast bootstrap methods for calculating confidence intervals

In this chapter, I present 5 examples showing how to use the package

## 2.1 Installing and loading the package

The package is currently not available on CRAN

to install the developer version of the package please run

```
if(!require(devtools)) {install.packages('devtools'); library(devtools)}
devtools::install_github('MaciejDanko/marPredict')
```

You may need to install RTools on Windows system before. <https://cran.r-project.org/bin/windows/Rtools/>

## 2.2 Example 1: Biased and unbiased model predictions

Let us consider a artificially generated data from a simple two-level Poisson model. The model assumes that a hypothetical variable Y depends on the two covariates X1 and X2, their interaction, and subject identification number ID.

```
head(marPredict_e1)
```

	Y	linkY	X1	X2	ID	IDv
1	1349	7.183494	c	bb	ID1	0.6834944
2	332	5.783494	b	cc	ID1	0.6834944
3	285	5.683494	a	aa	ID1	0.6834944
4	350	5.883494	c	cc	ID1	0.6834944
5	1297	7.183494	c	bb	ID1	0.6834944
6	158	4.983494	a	cc	ID1	0.6834944

where Y is a count response,  $\log Y$  is its logarithm (latent scale), X1 and X2 are factors, ID is a identifier of an individual, and IDv is a contribution of an individual to the intercept on the latent scale.  $\log Y$  and IDv have only informative property - they are not used to fit models. Knowing the formula that generated the data, we can use it to fit glmer model.

```
# fit the model
fit1 <- lme4::glmer(Y ~ X1 * X2 + (1|ID), family=poisson(),
                  data = marPredict_e1)
```

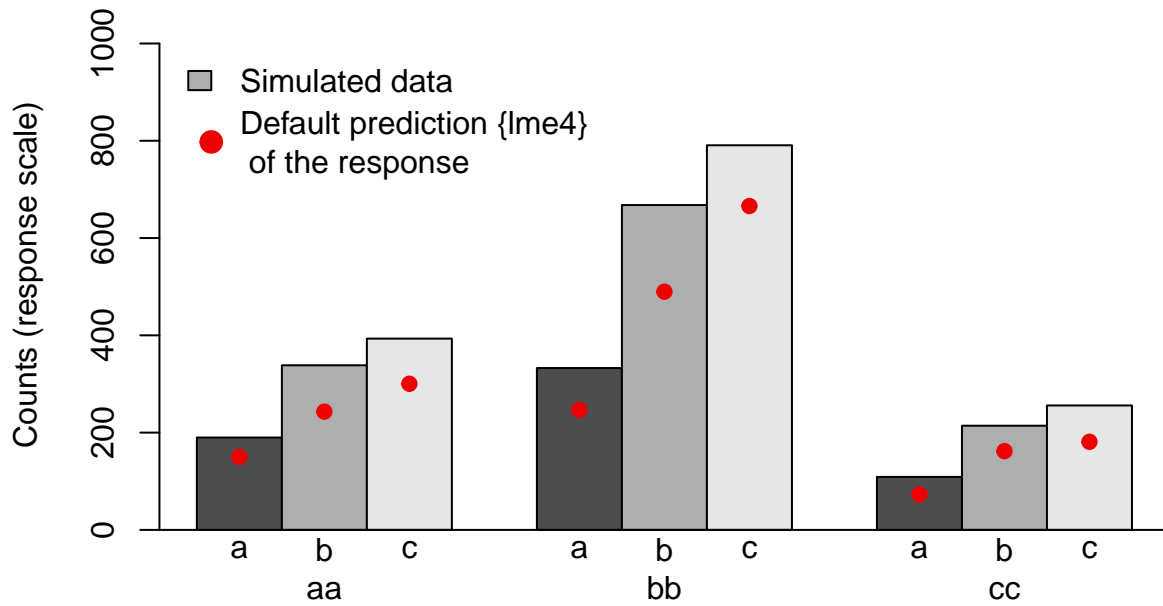
We can now perform model predictions using the generic predict function (or the lme4::predict.merMod equivalently).

```
# calculate and plot data (means)
DATA <- with(marPredict_e1, tapply(Y, data.frame(X1=X1,X2=X2), mean))
posX <- as.vector(barplot(DATA,beside=TRUE,ylim=c(0,1050),
                          ylab='Counts (response scale)'))
text(posX,-40,rep(rownames(DATA),3),xpd=TRUE);
axis(1,at=c(-5,100),labels = FALSE)

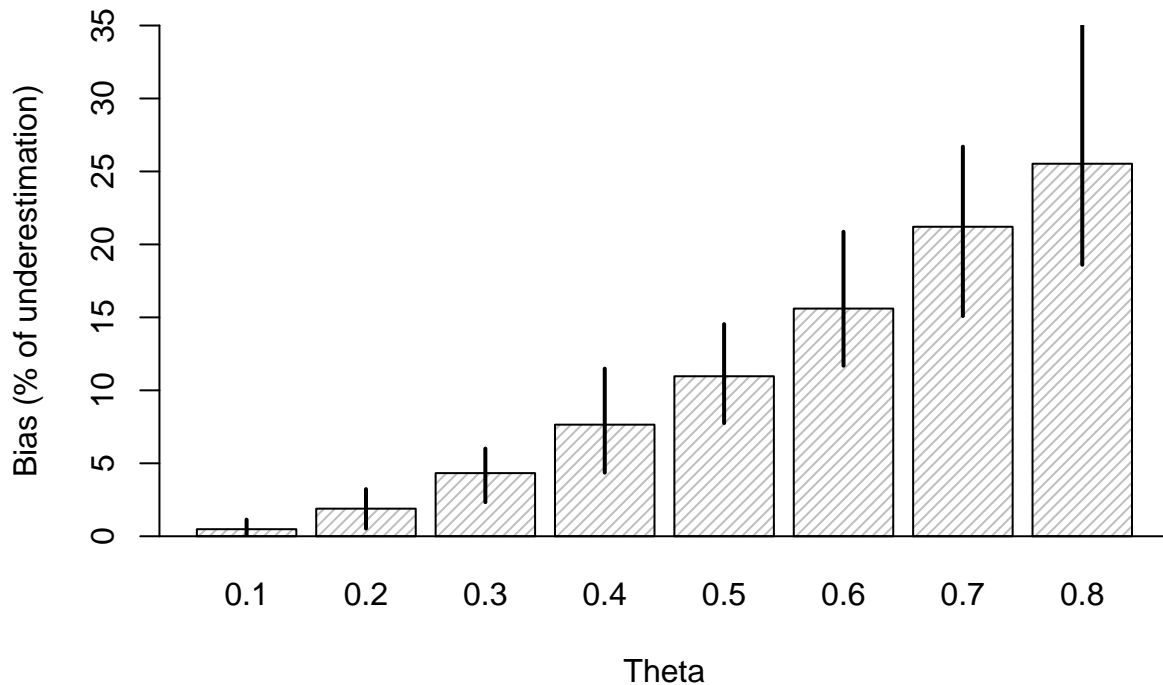
# calculate the model predictions
newdata <- expand.grid(X1=levels(marPredict_e1$X1), X2=levels(marPredict_e1$X2))
posY <- predict(fit1,
               newdata=newdata,
               type='response',
               re.form = NA)
```

```
# plot the predicted responses
lines(posX,posY,pch=20,type='p',col='red2',cex=1.5)

# plot the legend
legend('topleft',c('Simulated data','Default prediction {lme4}\n of the response'),
      pch=c(NA,19),pt.cex=c(1,1.5),
      fill=c(gray.colors(3)[2],NA),
      col=c(NA,'red2'),bty='n',border=c('black',NA))
```



**Figure 1.** Artificially generated Poisson data and the response predicted by the fitted `glmer` model. The results show that predictions on the response scale are highly underestimated. It is easy to show that that the bias depends on the magnitude of the theta parameter.



**Figure 2.** Estimate of the bias between predictions on the response scale and the real data. The confidence intervals were calculated according to the bootstrap percentile method for 100 replicates.

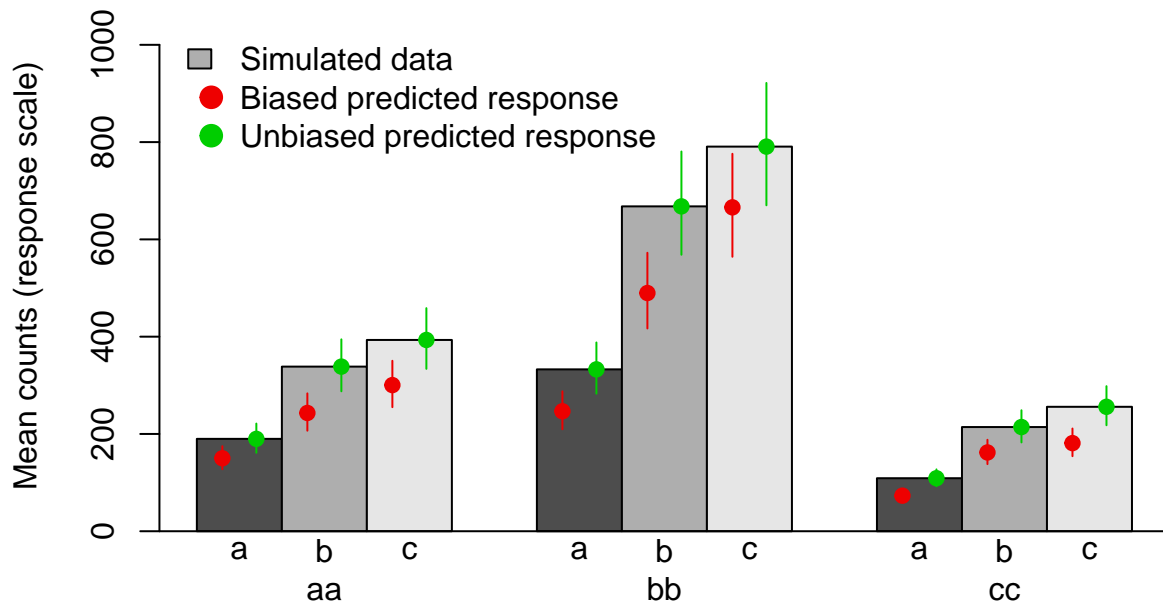
The unbiased predictions on response scale can be achieved using the `marPredict` function.

```
# plot the data (means)
posX <- as.vector(barplot(DATA,beside=TRUE,ylim=c(0,1050),
                          ylab='Mean counts (response scale)'))
text(posX,-40,rep(rownames(DATA),3),xpd=TRUE)
axis(1,at=c(-5,100),labels = FALSE)

# calculate the biased and the unbiased predictions on the response scale
cY <- marPredict(object=fit1,
                  newdata=as.data.frame(newdata),
                  type='response')

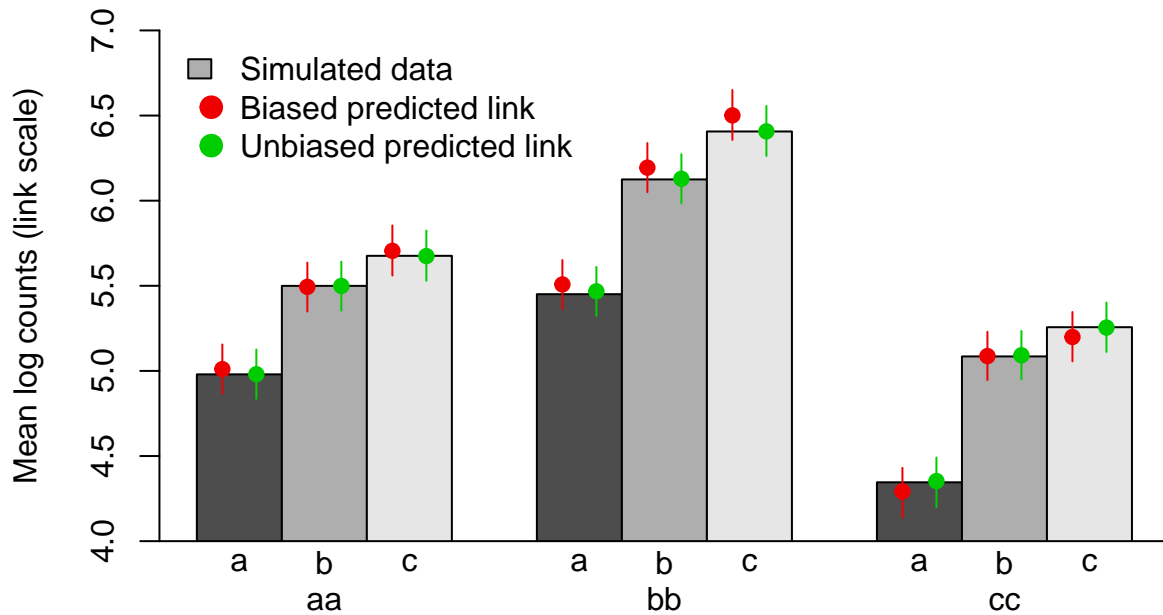
# plot the predicted responses
lines(posX-0.2,cY$fit$biased,pch=20,type='p',col='red2',cex=1.5)
lines(posX+0.2,cY$fit$unbiased,pch=20,type='p',col='green3',cex=1.5)

# plot the confidence intervals (the bootstrap percentile method)
for (i in seq_along(posX)) lines(c(posX[i],posX[i]+0.2,
                                   cY$CI.fit$unbiased[i,], col='green3')
                                )
for (i in seq_along(posX)) lines(c(posX[i],posX[i]-0.2,
                                   cY$CI.fit$biased[i,], col='red2')
                                )
legend('topleft',c('Simulated data','Biased predicted response',
                  'Unbiased predicted response'),
      pch=c(NA,19,19),pt.cex=c(1,1.5,1.5),
      fill=c(gray.colors(3)[2],NA,NA),
      col=c(NA,'red2','green3'), bty='n', border=c('black',NA,NA))
```



It must be noted that the individual contributions to the random effects are assumed to be fixed and are not drawn during the bootstrap method.

The bias is visible also on the link (latent) scale, however the magnitude is much smaller



## 2.3 Example 2: GAMM

The **marPredict** package supports two GAMM related packages **mgcv** and **gamm4**. Let us consider generate a Poisson data with one categorical and one continuous variable.

```
head(marPredict_e2)
```

	Y	linkY	X1	C1	ID	IDv
1	41	3.563412	a	0.44246177	ID1	0.5918332
2	6	2.237577	a	-0.09471397	ID1	0.5918332
3	9	2.644807	a	-0.23040374	ID1	0.5918332
4	21	2.833768	a	0.28028334	ID1	0.5918332
5	2	1.469079	c	-0.06203028	ID1	0.5918332
6	11	2.653087	b	-0.05314235	ID1	0.5918332

The artificially generated `marPredict_e2` can be fitted using the `gamm4` function of the **gamm4** package.

```
# fitting the model
fit2 <- gamm4::gamm4(Y ~ X1 + s(C1), random = ~(1|ID),
  family = poisson(), data = marPredict_e2)

# making the predictions
newmarPredict_e2 <- expand.grid(C1=seq(-0.5,0.5,0.05),
  X1=levels(marPredict_e2$X1))
cY2 <- marPredict(fit2, newmarPredict_e2, type = 'response', ci.fit = TRUE)

# plotting the results
ina <- which(newmarPredict_e2$X1=='a'); inb <- which(newmarPredict_e2$X1=='b')
inc <- which(newmarPredict_e2$X1=='c')
xa <- newmarPredict_e2$C1[ina]
xb <- newmarPredict_e2$C1[inb]
xc <- newmarPredict_e2$C1[inc]
```

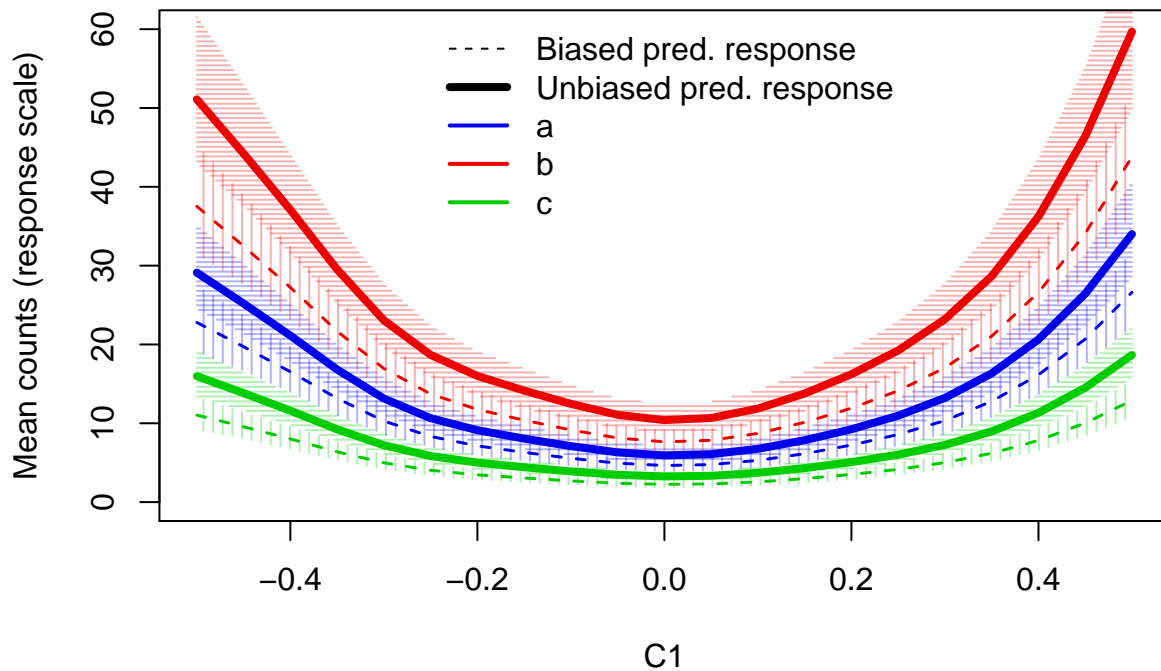


```

CIplot(xa, cY2$fit$unbiased[ina], cY2$CI.fit$unbiased[ina,],
      density=40, angle=0, first = TRUE, ylim=c(0,60), col='blue2', lwd=4,
      xaxt='s', yaxt='s', ylab='Mean counts (response scale)', xlab='C1')
CIplot(xb, cY2$fit$unbiased[inb], cY2$CI.fit$unbiased[inb,],
      density=40, angle=0, col='red2', lwd=4)
CIplot(xc, cY2$fit$unbiased[inc], cY2$CI.fit$unbiased[inc,],
      density=40, angle=0, col='green3', lwd=4)
CIplot(xa, cY2$fit$biased[ina], cY2$CI.fit$biased[ina,],
      density=20, angle=90, col='blue2', lty=2)
CIplot(xb, cY2$fit$biased[inb], cY2$CI.fit$biased[inb,],
      density=20, angle=90, col='red2', lty=2)
CIplot(xc, cY2$fit$biased[inc], cY2$CI.fit$biased[inc,],
      density=20, angle=90, col='green3', lty=2)

legend('top', c('Biased pred. response', 'Unbiased pred. response',
               'a', 'b', 'c'), lty=c(2,1,1,1,1), lwd=c(1,4,2,2,2),
      col=c(1,1,'blue2','red2','green3'), bty='n')

```



## 2.4 Example 3: Multi-level logistic regression

```
head(marPredict_e3)
```

	Y	linkY	X1	C1	ID	IDv
1	1	2.0951233	a	0.44246177	ID1	0.7101998
2	0	1.0207718	a	-0.09471397	ID1	0.7101998
3	0	0.7493923	a	-0.23040374	ID1	0.7101998
4	1	1.7707665	a	0.28028334	ID1	0.7101998
5	1	-0.1638608	c	-0.06203028	ID1	0.7101998
6	1	2.6039151	b	-0.05314235	ID1	0.7101998

```

# fitting the model
fit3 <- lme4::glmer(Y ~ X1 + C1 + (1|ID), family=binomial(), data=marPredict_e3, nAGQ = 20)

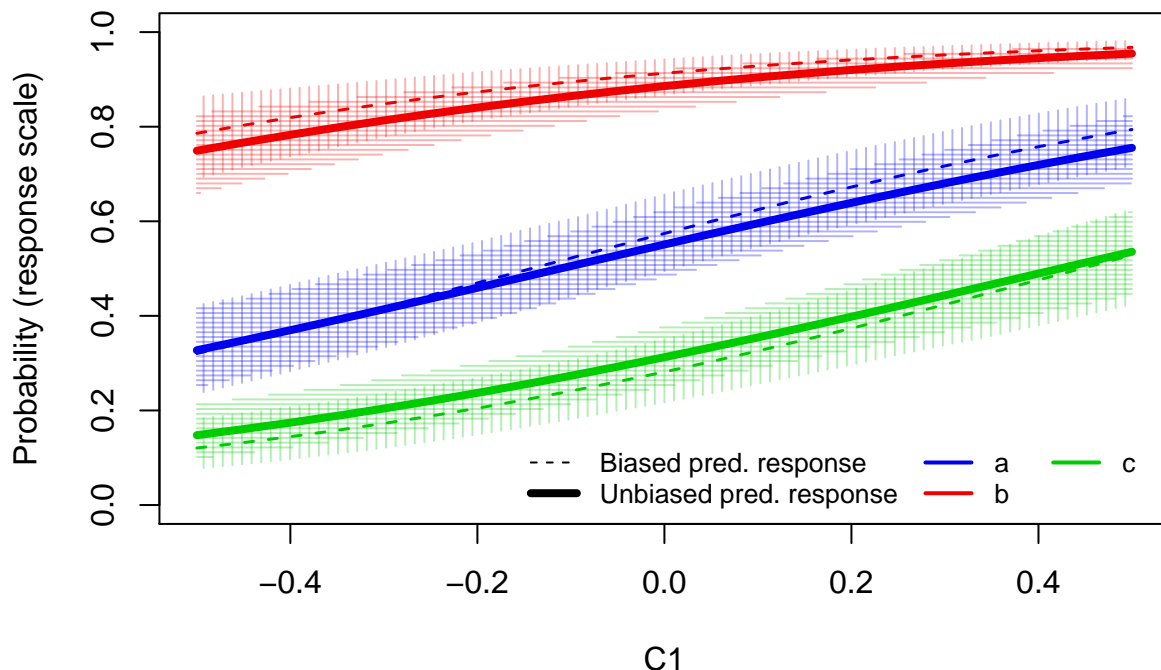
# making the predictions
newmarPredict_e3 <- expand.grid(C1=seq(-0.5,0.5,0.05), X1=levels(marPredict_e3$X1))
cY3 <- marPredict(fit3, newmarPredict_e3, n.sim = 1000,
                  type = 'response', ci.fit = TRUE)

# plotting the results
ina <- which(newmarPredict_e3$X1=='a'); inb <- which(newmarPredict_e3$X1=='b')
inc <- which(newmarPredict_e3$X1=='c')
xa <- newmarPredict_e3$C1[ina]; xb <- newmarPredict_e3$C1[inb]; xc <- newmarPredict_e3$C1[inc]

CIplot(xa, cY3$fit$unbiased[ina], cY3$CI.fit$unbiased[ina,],
       density=40, angle=0, first = TRUE, ylim=c(0,1), col='blue2', lwd=4,
       xaxt='s', yaxt='s', ylab='Probability (response scale)', xlab='C1')
CIplot(xb, cY3$fit$unbiased[inb], cY3$CI.fit$unbiased[inb,],
       density=40, angle=0, col='red2', lwd=4)
CIplot(xc, cY3$fit$unbiased[inc], cY3$CI.fit$unbiased[inc,],
       density=40, angle=0, col='green3', lwd=4)
CIplot(xa, cY3$fit$biased[ina], cY3$CI.fit$biased[ina,],
       density=20, angle=90, col='blue2', lty=2)
CIplot(xb, cY3$fit$biased[inb], cY3$CI.fit$biased[inb,],
       density=20, angle=90, col='red2', lty=2)
CIplot(xc, cY3$fit$biased[inc], cY3$CI.fit$biased[inc,],
       density=20, angle=90, col='green3', lty=2)

legend('bottomright', c('Biased pred. response', 'Unbiased pred. response',
                        'a', 'b', 'c', ''), lty=c(2,1,1,1,1,NA), lwd=c(1,4,2,2,2,NA),
      col = c(1,1,'blue2','red2','green3',NA), bty='n',
      cex = 0.8, text.width = strwidth('a')*c(7,7,2,2,2,NA), ncol = 3)

```



The data generated in this example has assumed very high `theta` parameter. It is thus clear that binomial logistic regression is more resistant to the bias resulting from ignoring random effects.

## 2.5 Example 4: Count data with offset and rates

```
head(marPredict_e4)
```

	Y	linkY	X1	X2	ID	IDv	exposures
1	1188	7.023201	b	bb	ID1	1.323201	376.96592
2	528	6.323201	c	cc	ID1	1.323201	112.78009
3	689	6.523201	b	cc	ID1	1.323201	327.38154
4	625	6.523201	d	bb	ID1	1.323201	71.18416
5	133	4.923201	d	cc	ID1	1.323201	76.66934
6	651	6.523201	d	bb	ID1	1.323201	120.42189

```
# fit the model
fit4 <- lme4::glmer(Y~X1*X2+(1|ID), family=poisson(),
                  offset=log(marPredict_e4$exposures), data=marPredict_e4, nAGQ = 20)

newmarPredict_e4 <- expand.grid(X1=levels(marPredict_e4$X1), X2=levels(marPredict_e4$X2))

# calculate the biased and the unbiased predictions on the response scale
cY4 <- marPredict(object = fit4,
                  newdata = newmarPredict_e4,
                  type = 'response',
                  use.offset = TRUE,
                  as.rate = TRUE,
                  ci.fit = TRUE,
                  method = 'at.each.cat')

DATA.counts <- with(marPredict_e4, tapply(Y, data.frame(X1=X1, X2=X2), sum))
DATA.exposures <- with(marPredict_e4, tapply(exposures, data.frame(X1=X1,X2=X2), sum))
DATA.rates <- DATA.counts/DATA.exposures

posX <- as.vector(barplot(DATA.rates,beside=TRUE, ylim=c(0,3),
                        ylab='Rates'))
text(posX,-0.1,rep(rownames(DATA.counts),3),xpd=TRUE)
axis(1,at=c(-5,100),labels = FALSE)

lines(posX-0.2,cY4$fit$biased,pch=20,type='p',col='red2',cex=1.5)
lines(posX+0.2,cY4$fit$unbiased,pch=20,type='p',
      col='green3',cex=1.5)

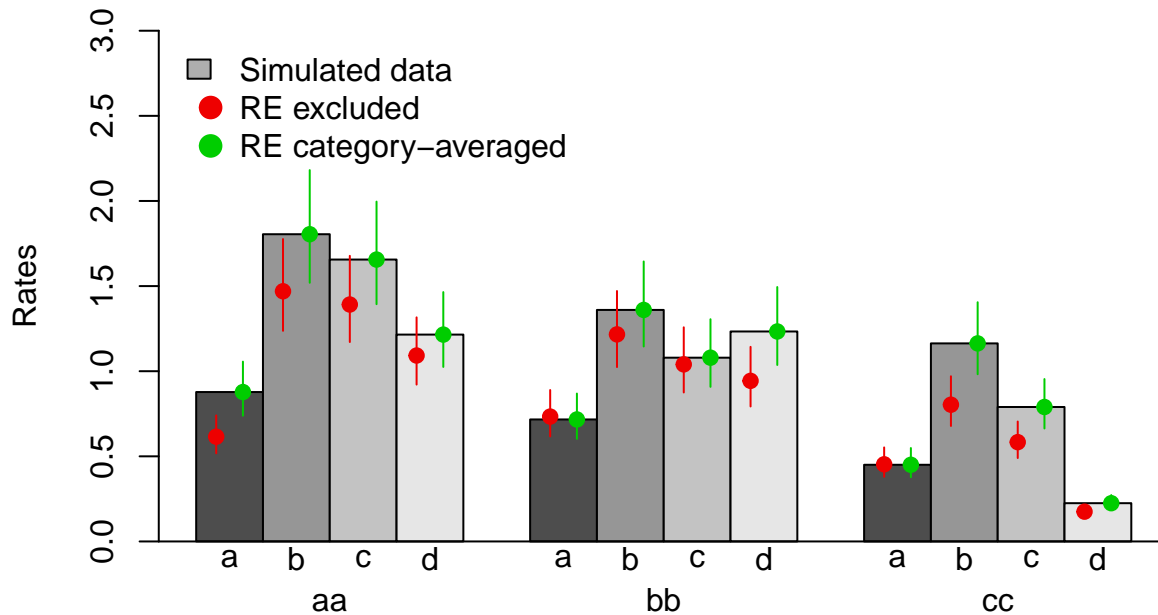
# plot the confidence intervals (the bootstrap percentile method)
for (i in seq_along(posX)) lines(c(posX[i],posX[i])+0.2,
                                cY4$CI.fit$unbiased[i,],
                                col='green3')
for (i in seq_along(posX)) lines(c(posX[i],posX[i])-0.2,
                                cY4$CI.fit$biased[i,], col='red2')

legend('topleft',c('Simulated data','RE excluded',
```

```

      'RE category-averaged'),
pch=c(NA,19,19),pt.cex=c(1,1.5,1.5),
fill=c(gray.colors(3)[2],NA,NA),
col=c(NA,'red2','green3'),
bty='n', border=c('black',NA,NA))

```



## 2.6 Example 5: Predictions with variables missing in the newdata

```
head(marPredict_e5)
```

```

      Y   linkY X1 X2      C1 ID   IDv
1 319 5.778722 c cc -0.14833941 ID1 1.856213
2 167 5.315066 a aa  0.02743115 ID1 1.856213
3 136 5.063320 a cc -0.27140510 ID1 1.856213
4 171 4.936572 c cc  0.41309378 ID1 1.856213
5 155 5.066377 a aa  0.19322383 ID1 1.856213
6  60 4.077087 a cc  0.38608379 ID1 1.856213

```

```

fit5 <- lme4::glmer(Y ~ X1 * X2 + C1 + (1|ID), family=poisson(),
  data = marPredict_e5)
DATA <- with(marPredict_e5, tapply(Y, data.frame(X1=X1), mean))

```

```
newmarPredict_e5 <- expand.grid(X1=levels(marPredict_e5$X1))
```

```

# error due to missing variables in newdata
cY5 <- try(marPredict(object=fit5,
  newdata = newmarPredict_e5,
  type='response',
  ci.fit = TRUE,
  method = 'at.each.cat'))

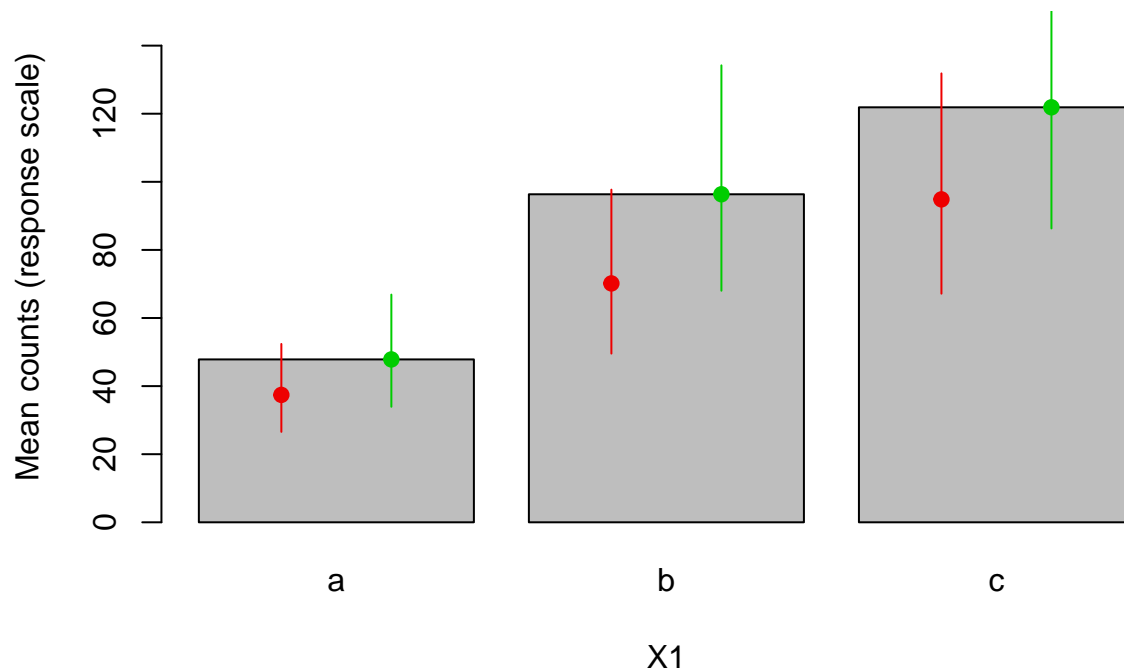
```

Error : Missing variable(s) in newdata: X2, C1.  
 Add missing variables or set average.missing = TRUE.

*# averaging missing variables*

```
cY5 <- try(marPredict(object = fit5,
  newdata = newmarPredict_e5,
  type='response',
  average.missing = TRUE,
  ci.fit = TRUE,
  method = 'at.each.cat'))
```

```
posX <- as.vector(barplot(DATA,beside=TRUE,ylim=c(0,150),
  ylab='Mean counts (response scale)',
  xlab='X1'))
lines(posX-0.2,cY5$fit$biased, type='p',col='red2',pch=20,cex=1.5)
lines(posX+0.2,cY5$fit$unbiased,pch=20,type='p', col='green3',cex=1.5)
for (i in seq_along(posX)) lines(c(posX[i],posX[i]+0.2,
  cY5$CI.fit$unbiased[i,],col='green3')
for (i in seq_along(posX)) lines(c(posX[i],posX[i]-0.2,
  cY5$CI.fit$biased[i,], col='red2')
```



## 2.7 use of the approximation

## 2.8 Confidence intervals

- comparison biased ci with classic se
- comparison LME4 CI with REsim == TRUE
- comparison with merTools

### 3 Conclusions

### References

- Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. “Fitting Linear Mixed-Effects Models Using lme4.” *Journal of Statistical Software* 67 (1): 1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Dańko, Maciej J. 2019. “hopit: An R Package for Analysis of Reporting Behavior Using Generalized Ordered Probit Models.” *Journal of Open Source Software* x (xx): 1481. <https://doi.org/10.21105/joss.01481>.
- Everitt, Brian S., and Torsten Hothorn. 2017. *HSAUR2: A Handbook of Statistical Analyses Using R (2nd Edition)*. <https://CRAN.R-project.org/package=HSAUR2>.
- Knowles, Jared E., and Carl Frederick. 2019. *merTools: Tools for Analyzing Mixed Effect Regression Models*. <https://CRAN.R-project.org/package=merTools>.