

# Overdispersion, and how to deal with it in R and JAGS

(requires R-packages AER, coda, lme4, R2jags, DHARMA/devtools)

Carsten F. Dormann

07 December, 2016

## Contents

<b>1</b>	<b>Introduction: what is overdispersion?</b>	<b>1</b>
<b>2</b>	<b>Recognising (and testing for) overdispersion</b>	<b>1</b>
<b>3</b>	<b>“Fixing” overdispersion</b>	<b>5</b>
3.1	Quasi-families . . . . .	5
3.2	Different distribution (here: negative binomial) . . . . .	6
3.3	Observation-level random effects (OLRE) . . . . .	8
<b>4</b>	<b>Overdispersion in JAGS</b>	<b>12</b>

## 1 Introduction: what is overdispersion?

Overdispersion describes the observation that variation is higher than would be expected. Some distributions do not have a parameter to fit variability of the observation. For example, the *normal distribution* does that through the parameter  $\sigma$  (i.e. the standard deviation of the model), which is constant in a typical regression. In contrast, the *Poisson distribution* has no such parameter, and in fact the variance increases with the mean (i.e. the variance and the mean have the same value). In this latter case, for an expected value of  $E(y) = 5$ , we also expect that the variance of observed data points is 5. But what if it is not? What if the observed variance is much higher, i.e. if the data are overdispersed? (Note that it could also be lower, underdispersed. This is less often the case, and not all approaches below allow for modelling underdispersion, but some do.)

Overdispersion arises in different ways, most commonly through “clumping”. Imagine the number of seedlings in a forest plot. Depending on the distance to the source tree, there may be many (hundreds) or none. The same goes for shooting stars: either the sky is empty, or littered with shooting stars. Such data would be overdispersed for a Poisson distribution. Also, overdispersion arises “naturally” if important predictors are missing or functionally misspecified (e.g. linear instead of non-linear).

Overdispersion is often mentioned together with zero-inflation, but it is distinct. Overdispersion also includes the case where none of your data points are actually 0. We’ll look at zero-inflation later, and stick to overdispersion here.

## 2 Recognising (and testing for) overdispersion

May we should start with an example to get the point visualised. Note that we manually set the breaks to 1-unit bins, so that we can see the 0s as they are, not pooled with 1s, 2s, etc.

```
library(lme4)
data(grouseticks)
summary(grouseticks)
```

# Package ‘RVAideMemoire’

January 27, 2021

**Encoding** latin1  
**Type** Package  
**Title** Testing and Plotting Procedures for Biostatistics  
**Version** 0.9-79  
**Date** 2021-01-27  
**Author** Maxime Hervé  
**Maintainer** Maxime Hervé <maxime.herve@univ-rennes1.fr>

**biocViews**

**Imports** ade4 (>= 1.7-8), boot, car, FactoMineR, graphics, grDevices, lme4 (>= 1.0-4), MASS, mixOmics, nnet, pls, pspearman, stats, utils, vegan (>= 2.4-3)

**Suggests** ape, dgof, emmeans, EMT, FSA, labdsv, mvnormtest, ordinal, RGCCA, statmod, survival

**Description** Contains miscellaneous functions useful in biostatistics, mostly univariate and multivariate testing procedures with a special emphasis on permutation tests. Many functions intend to simplify user's life by shortening existing procedures or by implementing plotting functions that can be used with as many methods from different packages as possible.

**License** GPL-2

**LazyLoad** yes

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2021-01-27 08:00:02 UTC

**R topics documented:**

RVAideMemoire-package . . . . .	4
adonis.II . . . . .	5
Anova.clm . . . . .	6
back.emmeans . . . . .	7
bootstrap . . . . .	8

Estimation of overdispersion with [glmer](#) models**Description**

Estimates residual deviance and residual degrees of freedom to check for overdispersion with [glmer](#) models. This function is directly coming from <http://glmm.wikidot.com/faq>.

**Usage**

```
overdisp.glmer(model)
```

**Arguments**

`model`  
a model fitted by [glmer](#).

**Author(s)**

Ben Bolker

**See Also**

[glmer](#)

**Examples**

```
require(lme4)

# Example from the 'glmer' function
gm1 <- glmer(cbind(incidence, size-incidence)~period+(1|herd),
  family="binomial", data=cbpp)
overdisp.glmer(gm1)
```

R version 4.0.2 (2020-06-22) -- "Taking Off Again"  
##### Use **bovine herd example**, lecture 3 item 2a

```
> library(lme4)
Loading required package: Matrix
> data(cbpp)
> gml <- glmer(cbind(incidence,size-incidence)~period+(1|herd),
+ family="binomial",data=cbpp)
> summary(gml)
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: cbind(incidence, size - incidence) ~ period + (1 | herd)
Data: cbpp
```

AIC	BIC	logLik	deviance	df.resid
194.1	204.2	-92.0	184.1	51

Scaled residuals:

Min	1Q	Median	3Q	Max
-2.3816	-0.7889	-0.2026	0.5142	2.8791

Random effects:

Groups	Name	Variance	Std.Dev.
herd	(Intercept)	0.4123	0.6421

Number of obs: 56, groups: herd, 15

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-1.3983	0.2312	-6.048	1.47e-09 ***
period2	-0.9919	0.3032	-3.272	0.001068 **
period3	-1.1282	0.3228	-3.495	0.000474 ***
period4	-1.5797	0.4220	-3.743	0.000182 ***

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

Correlation of Fixed Effects:

	(Intr)	perid2	perid3
period2	-0.363		
period3	-0.340	0.280	
period4	-0.260	0.213	0.198

```
> plot(gml)
> #gives residuals (pearson) vs fitted plot (shows no structure)
```

**for overdispersion function need**  
> install.packages("RVAideMemoire")  
##this is one of those that installs alot of packages, be aware

```
package 'base64enc' successfully unpacked and MD5 sums checked
package 'yaml' successfully unpacked and MD5 sums checked
package 'lazyeval' successfully unpacked and MD5 sums checked
package 'later' successfully unpacked and MD5 sums checked
package 'htmltools' successfully unpacked and MD5 sums checked
package 'htmlwidgets' successfully unpacked and MD5 sums checked
package 'jsonlite' successfully unpacked and MD5 sums checked
package 'crosstalk' successfully unpacked and MD5 sums checked
package 'promises' successfully unpacked and MD5 sums checked
package 'pixmap' successfully unpacked and MD5 sums checked
package 'DT' successfully unpacked and MD5 sums checked
package 'ellipse' successfully unpacked and MD5 sums checked
package 'flashClust' successfully unpacked and MD5 sums checked
package 'leaps' successfully unpacked and MD5 sums checked
package 'scatterplot3d' successfully unpacked and MD5 sums checked
package 'ggrepel' successfully unpacked and MD5 sums checked
package 'permute' successfully unpacked and MD5 sums checked
package 'ade4' successfully unpacked and MD5 sums checked
package 'FactoMineR' successfully unpacked and MD5 sums checked
package 'pls' successfully unpacked and MD5 sums checked
package 'p Spearman' successfully unpacked and MD5 sums checked
package 'vegan' successfully unpacked and MD5 sums checked
package 'RVAideMemoire' successfully unpacked and MD5 sums checked
```

The downloaded binary packages are in  
C:\Users\rag\AppData\Local\Temp\RtmpEF0gYF\downloaded\_packages  
> library(RVAideMemoire)  
\*\*\* Package RVAideMemoire v 0.9-79 \*\*\*

Attaching package: 'RVAideMemoire'

The following object is masked from 'package:lme4':

dummy

Warning message:

package 'RVAideMemoire' was built under R version 4.0.5

## after all that

> **overdisp.glmer(gml)**

Residual deviance: 73.474 on 51 degrees of freedom (ratio: 1.441)

> # significant overdispersion, as best we can see if you trust the chi-square approximation

> qchisq(.95, 51)

[1] 68.66929

> pchisq(73.47, 51)

[1] 0.9786524