**Odds Ratios**

Odds Ratios are made up of odds, which are themselves a ratio of probabilities

\(\text{Odds} = \frac{P(X\_1=1)}{P(X\_1=0)}.\)  
Since we chose to put $P(X\_1=1)$ in the numerator, we interpret these odds as the “odds being in favor of $X\_1=1$”. For example, if $X\_1$ is the symptom sleep problems which takes the values 0 (no sleep problems) with probability 0.75 and the value 1 (sleep problems) with probability 0.25, then the odds of having sleep problems are 1 to 4.

However, these odds may be different in different circumstances. Let’s say these circumstances are captured by variable $X\_2$ which takes values in ${0,1}$. In our example, those circumstances could be whether you live next to a busy street (1) or not (0). If the odds indeed depend on $X\_2$ then we have

\[\text{Odds}\_{X\_2=1} = \frac{P(X\_1=1 \mid X\_2=1)}{P(X\_1=0 \mid X\_2=1)} \neq  
\frac{P(X\_1=1 \mid X\_2=0)}{P(X\_1=0 \mid X\_2=0)} = \text{Odds}\_{X\_2=0}  
.\]

A way to quantify the degree to which the odds are different depending whether we set $X\_2=1$ or $X\_2=0$ is to divide the odds in those two situations

\[\text{Odds Ratio} = \frac{\text{Odds}\_{X\_2=1}}{\text{Odds}\_{X\_2=0}}  
,\]

which gives rise to an odds ratio (OR).

How do we interpret this odds ratio? If the OR is equal to 1, then $X\_2$ has no influence on the odds between $P(X\_1=1)$ and $P(X\_1=0)$; if OR > 1, $X\_2=1$ *increases* the odds compared to $X\_2=0$; and if OR < 1, $X\_2=1$ *decreases* the odds compared to $X\_2=0$. In our example from above, an OR = 4 would imply that the odds of sleep problems (vs. no sleep problems) are four times larger when living next to a busy street (vs. not living next to a busy street).

In the remainder of this blog post, I will illustrate how to compute such odds ratios based on MGMs estimated with the R-package [*mgm*](https://cran.r-project.org/web/packages/mgm/index.html).

**Loading Example Data**

We use a data set on Autism Spectrum Disorder (ASD) which contains $n=3521$ observations of seven variables, gender (1 = male, 2 = female), IQ (continuous), Integration in Society (3 categories ordinal), Number of comorbidities (count), Type of housing (1 = supervised, 2 = unsupervised), working hours (continuous), and satisfaction with treatment (continuous).

library(mgm) # data is loaded with mgm package (version 1.2-11)

head(autism\_data$data)

## Gender IQ Integration in Society No of Comorbidities Type of Housing Workinghours

## 1 1 6.00 1 1 1 0

## 2 2 6.00 2 1 1 0

## 3 1 5.00 2 0 1 0

## 4 1 6.00 1 0 1 10

## 5 1 5.00 1 1 1 0

## 6 1 4.49 1 1 1 0

## Satisfaction: Treatment

## 1 3.00

## 2 2.00

## 3 4.00

## 4 3.00

## 5 1.00

## 6 1.75

**Fitting MGM**

We model gender, integration in society and type of housing as categorical variables (with 2, 3 and 2 categories), and all remaining variables as Gaussian variables

set.seed(1)

mod <- mgm(data = autism\_data$data,

type = c("c", "g", "c", "g", "c", "g", "g"),

level = c(2, 1, 3, 1, 2, 1, 1),

lambdaSel = "CV",

ruleReg = "AND",

pbar=FALSE)

## Note that the sign of parameter estimates is stored separately; see ?mgm

and we visualize the dependencies of the resulting model using the qqgraph package:

library(qgraph)

qgraph(mod$pairwise$wadj,

nodeNames = autism\_data$colnames,

edge.color = mod$pairwise$edgecolor,

# edge.labels = TRUE,

legend = TRUE,

layout = "spring")

**Example Calculation of Odds Ratio**

We consider the simplest possible example of calculating the OR involving two binary variables. Note that we calculate the OR between two variables within a multivariate model. This means that the OR is conditional on all other variables in the model which implies that it can be different from the OR calculated based on those two variables alone (specifically, this will always happen when at least one additional variable is connected to both binary variables.)

Some of you might know that there is a simple relationship between the OR and the coefficients in logistic regression. Since multinomial regression with two outcomes is equivalent to logistic regression, we could use this simple rule in this specific example. Here, however, we start out with the definition of OR and show how to calculate it in the general case of $\geq 2$ categories. Along the way, we’ll also derive the simple relationship between parameters and ORs for the binary case.

In our data set we use the two binary variables type of housing and gender for illustration. Specifically, we look at how the odds of type of housing change as a function of gender. Corresponding to the column numbers of the two variables, let $X\_5$ be type of housing, and $X\_1$ gender.

The definition of ORs above shows that we need to calculate four conditional probabilities. We first calculate $P(X\_5=1 \mid X\_1=0)$ and $P(X\_5=0 \mid X\_1=0)$ in the numerator. To compute these probabilities, we need the estimated parameters of the multinomial regression on $X\_5$. In the standard parameterization of multinomial regression one of the response categories serves as the reference category. The regularization used within the mgm package allows more direct parameterization in which the probability of each response category can be modeled directly. We therefore get a set of parameters for *each* response category.

We can find those parameters in mod$nodemodels[[5]] which contains the parameters of the multinomial regression on variable $X\_5$:

coefs <- mod$nodemodels[[5]]$model

coefs

## $`1`

## 8 x 1 sparse Matrix of class "dgCMatrix"

## 1

## (Intercept) 0.62372320

## V1.2 -0.38147442

## V2. -0.42186120

## V3.2 0.13918915

## V3.3 -0.04001268

## V4. 0.01414994

## V6. -0.22024227

## V7. 0.04346463

##

## $`2`

## 8 x 1 sparse Matrix of class "dgCMatrix"

## 1

## (Intercept) -0.62372320

## V1.2 0.38147442

## V2. 0.42186120

## V3.2 -0.13918915

## V3.3 0.04001268

## V4. -0.01414994

## V6. 0.22024227

## V7. -0.04346463

The first set of parameters in coefs[[1]] models $P(X\_5 = 0 \mid \dots)$ and the second set of parameters in coefs[[2]] models $P(X\_5 = 1 \mid \dots)$.

The data set contains seven variables, which means that we have six variables that predict variable 5. However, since variable 3 (Integration in Society) is a categorical variable with three categories, it is represented by two dummy variables that code for its 2nd and 3rd category. This is why we have a total of 5\*1+2=7 predictor variables and 7 associated parameters.

Now, back to computing those probabilities in the enumerator. We would like to compute $P(X\_5=1 \mid X\_1=0)$ and $P(X\_5=0 \mid X\_1=0)$, however we see that the probability of $P(X\_5)$ not only depends on $X\_1$ but also on all other variables (none of the parameters are zero). We therefore need to fix all variables to some value in order to obtain a conditional probability. That is, we actually have to write the conditional probabilities as $P(X\_5=1 \mid X\_1, X\_2, X\_3, X\_4, X\_6, X\_7)$ and $P(X\_5=0 \mid X\_1, X\_2, X\_3, X\_4, X\_6, X\_7)$. Here we will fix all other variables to 0, but we will see later that it does not matter for our OR calculation to which value we set all these variables, as long as we choose the same values in all of the four probabilities.

The probability of $P(X\_5=0 \mid \dots)$ is calculated by dividing the potential for this category by the sum of all (here two) potentials:

\(P(X\_5=0 \mid \dots) = \frac{  
\text{Potential}(X\_5=0 \mid \dots)  
}{  
\text{Potential}(X\_5=0 \mid \dots) + \text{Potential}(X\_5=1 \mid \dots)  
}\)  
If $X\_5$ would have $m$ categories, there would be $m$ terms in the denominator.

The potentials are specified by the estimated parameters

\[\text{Potential}(X\_5=1 \mid X\_1=0, X\_2=0, \dots, X\_7=0) =  
\exp \{  
\beta\_{0} + \beta\_{0,1.2} \mathbb{I}(X\_1=1) + \dots + \beta\_{0,7} X\_7  
\}  
,\]

where $\beta\_{0}$ is the intercept and the remaining seven parameters are the ones associated with the predictor terms in the model, and $\mathbb{I}(X\_1=1)$ is the indicator function (or dummy variable) for category $X\_1=1$.

Notice that we set all variables to zero, which means that the above potential simplifies to

\[\text{Potential}(X\_5=1 \mid \dots) =  
\exp \{\beta\_{0} \} \approx \exp \{ 0.624 \}\]

where I took the intercept parameter $\beta\_{0}$ from coefs[[1]][1,1]. Similarly, we have

\(\text{Potential}(X\_5=0 \mid \dots) =  
\exp \{\beta\_{1} \} \approx \exp \{ -0.624 \}\)  
taken from coefs[[1]][1,1].

Using the two potentials, we can compute the probabilities. We now do this in R:

Potential0 <- exp(coefs[[1]][1,1])

Potential1 <- exp(coefs[[2]][1,1])

Prob0 <- Potential0 / (Potential0 + Potential1)

Prob1 <- Potential1 / (Potential0 + Potential1)

Prob0

## [1] 0.7768575

Prob1

## [1] 0.2231425

We calculated that the probability of $P(X\_5=0 \mid \dots)$ (supervised housing) is $\approx 0.78$ and the probability of $P(X\_5=0 \mid \dots)$ (unsupervised housing) is $\approx 0.22$.

Now we can compute the odds $\text{Odds}\_{X\_2=0}$:

odds\_x1\_0 <- Prob1 / Prob0

odds\_x1\_0

## [1] 0.2872373

The odds are smaller than one, which means that it is more likely that an individual lives in supervised housing.

Note that when computing the odds, the denominator in the calculations for the probabilities cancel out, which means we could have immediately computed the odds with the potentials:

Potential1 / Potential0

## [1] 0.2872373

So far, we computed the numerator of the formula of the Odds Ratio. We now compute the denominator, which includes the same conditional probabilities as above, except that we set $X\_1=1$ instead of $X\_1=0$. As discussed above, all other variables are kept constant at 0. To keep things short, I only show the R code for this second case:

Potential0 <- exp(coefs[[1]][1,1] + coefs[[1]][2,1] \* 1)

Potential1 <- exp(coefs[[2]][1,1] + coefs[[2]][2,1] \* 1)

odds\_x1\_1 <- Potential1 / Potential0

Similar to above, coefs[[1]][1,1] contains the intercept and coefs[[1]][2,1] contains the parameter associated with predictor $X\_1$ for probability $P(X\_5 = 0 \mid \dots)$. coefs[[2]] contains the corresponding parameters for probability $P(X\_5 = 1 \mid \dots)$.

We can now compute the OR:

OR <- odds\_x1\_1 / odds\_x1\_0

OR

## [1] 2.144591

We see that for females (coded 1) the odds of living in unsupervised housing (coded 1) are about twice as high as for males.

In a similar way, ORs can be calculated when the predicted variable or the predictor variable of interest has more than two categories. One can also compute ORs that combine the effect of several variables, for example by setting several variables to 1 in the numerator, and setting them all to 0 in the denominator.

**Does it matter to which value we fix the other variables?**

In the above calculation we fixed all other variables to zero. Would the OR have been different if we had fixed them to a different value? We will show with some basic calculations that the answer is no. Along the way, we will also derive a much simpler way to compute the OR from the parameter estimates for the special case in which the response variable is binary.

To keep the notation manageable, we only consider one third variable $X\_3$ instead of the five in the empirical example above. However, we would reach the same conclusion with any number of additional variables that are kept constant.

We start out with the definition of the odds ratio:

\[\text{Odds Ratio} =  
\frac{\text{Odds}\_{X\_2=1}}{\text{Odds}\_{X\_2=0}}  
=  
\frac{  
\frac{P(X\_1=1 \mid X\_2=1,X\_3=x\_3)}{P(X\_1=0 \mid X\_2=1,X\_3=x\_3)}  
}{  
\frac{P(X\_1=1 \mid X\_2=0,X\_3=x\_3)}{P(X\_1=0 \mid X\_2=0,X\_3=x\_3)}  
}  
.\]

The question is whether it matters what we fill in for $x\_3$. We will show that the terms associated with $x\_3$ cancel out and it therefore does not matter to which value we fix $x\_3$.

\[\frac{  
\frac{P(X\_1=1 \mid X\_2=1,X\_3=x\_3)}{P(X\_1=0 \mid X\_2=1,X\_3=x\_3)}  
}{  
\frac{P(X\_1=1 \mid X\_2=0,X\_3=x\_3)}{P(X\_1=0 \mid X\_2=0,X\_3=x\_3)}  
}  
=  
\frac{  
\frac{\exp\{\beta\_1 + \beta\_{21}1 + \beta\_{31}x\_3\}  
}{\exp\{\beta\_0 + \beta\_{20}1 + \beta\_{30}x\_3\}}  
}{  
\frac{\exp\{\beta\_1 + \beta\_{21}0 + \beta\_{31}x\_3\}  
}{  
\exp\{\beta\_0 + \beta\_{20}0 + \beta\_{30}x\_3\}  
}  
}  
=  
\frac{  
\frac{\exp\{\beta\_1 + \beta\_{21} + \beta\_{31}x\_3\}  
}{\exp\{\beta\_0 + \beta\_{20} + \beta\_{30}x\_3\}}  
}{  
\frac{\exp\{\beta\_1 + \beta\_{31}x\_3\}  
}{  
\exp\{\beta\_0 + \beta\_{30}x\_3\}  
}  
}\]

In the first step we fixed $X\_1=1$ in the numerator and $X\_1=0$ in the denominator and simplified. The parameter $\beta\_{21}$ refers to the coefficient associated with $X\_2$ in the equation modeling $P(X\_1=1 \mid \dots)$, while $\beta\_{20}$ refers to the coefficient associated with $X\_2$ in the equation modeling $P(X\_1=0 \mid \dots)$.

We further rearrange

\[\frac{  
\frac{\exp\{\beta\_1 + \beta\_{21} + \beta\_{31}x\_3\}  
}{\exp\{\beta\_0 + \beta\_{20} + \beta\_{30}x\_3\}}  
}{  
\frac{\exp\{\beta\_1 + \beta\_{31}x\_3\}  
}{  
\exp\{\beta\_0 + \beta\_{30}x\_3\}  
}  
}  
=  
\frac{\exp\{\beta\_1 + \beta\_{21} + \beta\_{31}x\_3\}  
}{\exp\{\beta\_0 + \beta\_{20} + \beta\_{30}x\_3\}}  
\frac{\exp\{\beta\_0 + \beta\_{30}x\_3\}  
}{  
\exp\{\beta\_1 + \beta\_{31}x\_3\}  
}  
,\]

which is equal to

\[\exp\{\beta\_1 + \beta\_{21} + \beta\_{31}x\_3 + \beta\_0 + \beta\_{30}x\_3 – (\beta\_0 + \beta\_{20} + \beta\_{30}x\_3 + \beta\_1 + \beta\_{31}x\_3)\}  
.\]

We collect all the terms with $x\_3$

\[\exp\{  
(\beta\_1 + \beta\_{21} + \beta\_0 – \beta\_0 – \beta\_{20} – \beta\_1)  
+  
(\beta\_{31}x\_3 + \beta\_{30}x\_3 – \beta\_{30}x\_3 – \beta\_{31}x\_3)  
\}\]

and we see that the terms including $x\_3$ add to zero, which shows that no matter what number we fill in for $x\_3$, we will always obtain the same OR.

Actually, we can further simplify and get

\[\exp\{  
(\beta\_1 + \beta\_{21} + \beta\_0 – \beta\_0 – \beta\_{20} – \beta\_1)  
+ 0  
=  
\exp\{  
\beta\_{21} – \beta\_{20}  
\}\]

which reveals a simpler way to compute the OR. We can verify this with our estimated coefficients

exp(coefs[[2]][2,1] - coefs[[1]][2,1])

## [1] 2.144591

and indeed we get the same OR.

If we wanted the OR with the numerator and denominator swapped, one would calculate:

exp(coefs[[1]][2,1] - coefs[[2]][2,1])

## [1] 0.4662894

One could verify this by repeating the derivation above with swapped numerator/denominator or by using the general approach of calculating ORs that we used above.

This reflects the relation between multiple regression parameters and the OR

\[\exp\{\beta\_x\}=\text{OR}\_x\]

since the relation between logistic regression parameterization and the symmetric multinomial regression parameterization used here is $\beta\_x = 2 \beta\_{x1}$, where $\beta\_{x1}$ is the parameter corresponding to $\beta\_x$ in the equation modeling $P(X\_1=1 \mid \dots)$.

**Is the OR “significant”?**

The model has been estimated with $\ell\_1$-regularized regression, in which the regularization parameters have been selected with 10-fold cross-validation with the goal that the parameter estimates generalize to new samples. Thus, variable selection has already been performed and it is not necessary to perform an additional hypothesis test on the OR or the underlying variables.

**An Alternative: Changes in Predicted Probabilities**

An alternative to ORs is to report change in predicted probabilities of $X\_5$ depending on which value we fill in for $X\_1$. When considering only two variables such a change in probabilities is perhaps easier to interpret than ORs. However, we will see that changes in predicted probabilities do not have the nice property of ORs that it doesn’t matter to which value we fix all other variables. This makes this alternative less attractive for models that include more than two variables.

When looking at changes in predicted probabilities we are interested in the difference

\[P(X\_5=1 \mid X\_1=1, \dots) – P(X\_5=0 \mid X\_1=0, \dots)  
.\]

When calculating these probabilities we are again required to fix all other variables (“…”) to some value, which we again choose to be 0. In the interest of brevity I only show the R-code for this calculation.

We compute the probabilities for the case $X\_1=0$

Potential0 <- exp(coefs[[1]][1,1])

Potential1 <- exp(coefs[[2]][1,1])

Prob1\_x10 <- Potential1 / (Potential0 + Potential1)

and for the case $X\_1=1$:

Potential0 <- exp(coefs[[1]][1,1] + coefs[[1]][2,1] \* 1)

Potential1 <- exp(coefs[[2]][1,1] + coefs[[2]][2,1] \* 1)

Prob1\_x11 <- Potential1 / (Potential0 + Potential1)

We see a change in probability of

Prob1\_x11 - Prob1\_x10

## [1] 0.1580482

That is, the probability of living in unsupervised housing is $\approx 0.16$ higher for females, which is consistent with the OR > 1 calculated above.

However, now let’s set $X\_3=1$ instead of $X\_3=0$:

Potential0 <- exp(coefs[[1]][1,1] + coefs[[1]][3,1]\*1)

Potential1 <- exp(coefs[[2]][1,1] + coefs[[2]][3,1]\*1)

Prob1\_x10 <- Potential1 / (Potential0 + Potential1)

Potential0 <- exp(coefs[[1]][1,1] + coefs[[1]][2,1] \* 1 + coefs[[1]][3,1]\*1)

Potential1 <- exp(coefs[[2]][1,1] + coefs[[2]][2,1] \* 1 + coefs[[2]][3,1]\*1)

Prob1\_x11 <- Potential1 / (Potential0 + Potential1)

We now get an increase of probability of

Prob1\_x11 - Prob1\_x10

## [1] 0.1884348

We see that changes in the predicted probabilities as a function of $X\_5$ depend on where we fixed the other variables. So while changes in probabilities are maybe easier to interpret, they have the downside that the changes depend on which values we fix the other variables. However, this may be acceptable in situations in which we are interested in some specific state of all other variables.

**Summary**

Starting out with the definition of odds ratios, I showed how to compute them in a general setting and how to compute them from the output of the [*mgm*](https://cran.r-project.org/web/packages/mgm/index.html)-package. We looked into whether the OR depends on the specific values at which we fix the other variables (it doesn’t). Proving this fact revealed a simpler formula for calculating ORs for the special case of having a binary response variable. Finally, we considered predicted probabilities as an alternative for ORs.