02424 - Assignment 3

Spring 22, 02424
Assignment 3
October 20, 2023

Nicolaj Hans Nielsen, s184335
Anton Ruby Larsen, s174356

Danmarks
Tekniske Universitet

Contents

A | Part 1, A

A.1 Exercise 1

This is the third and final assignment in this course. In the first part the this assignment, we will work with concrete strength. In the following, we will present the available data. We are given a dataset of 49 observation which each consist of 5 variables. The variables are further described in table A.1.

Variable	Domain	Description
date	Ordinal	YYYY-MM-DD
y7	Continuous	Concrete strength at day 7 [MPa]
y28	Continuous	Concrete strength at day 28 [MPa]
batch	Factor	Batch number for a given sample
air.temp	Continuous	Air temperate [°C]

Table A.1 – Description of concrete data.

We are to model the strength of the concrete given the other variables. We will first inspect how the measured strength at day 7 and day 28 differs with the measured time.

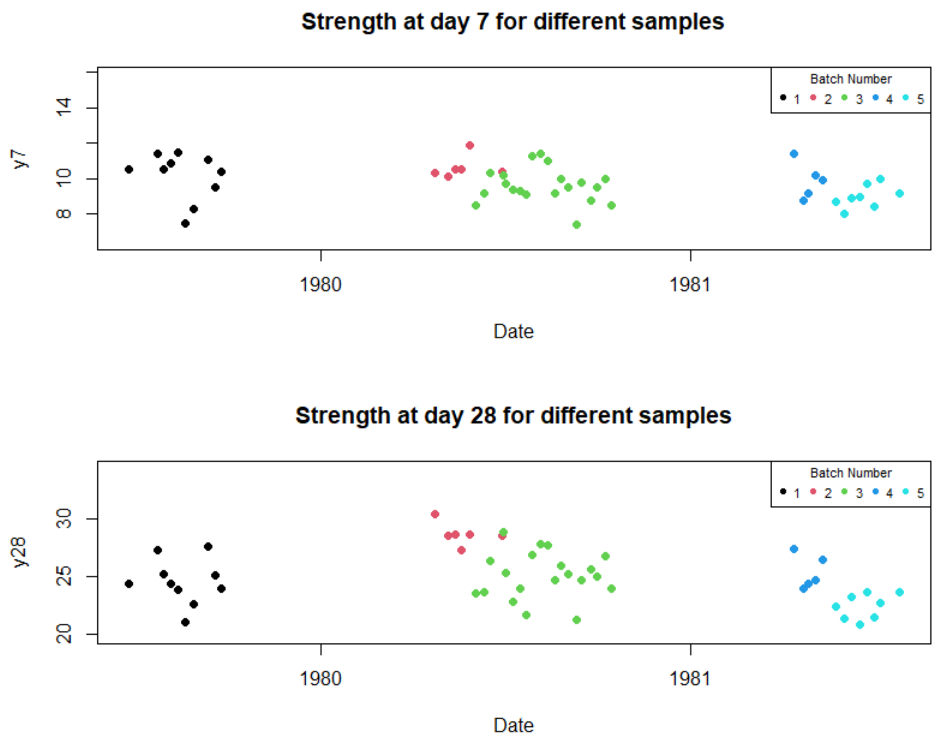


Figure A.1 – Sample time plotted versus the concrete strength at day 7 and 28.

In figure A.1 we see that the strength does not seem to vary that much through time but we might see some batch variation both in day 7 and day 28 measurements. We hence try to plot the different variables against the batches to see if the batch variation is consistent.

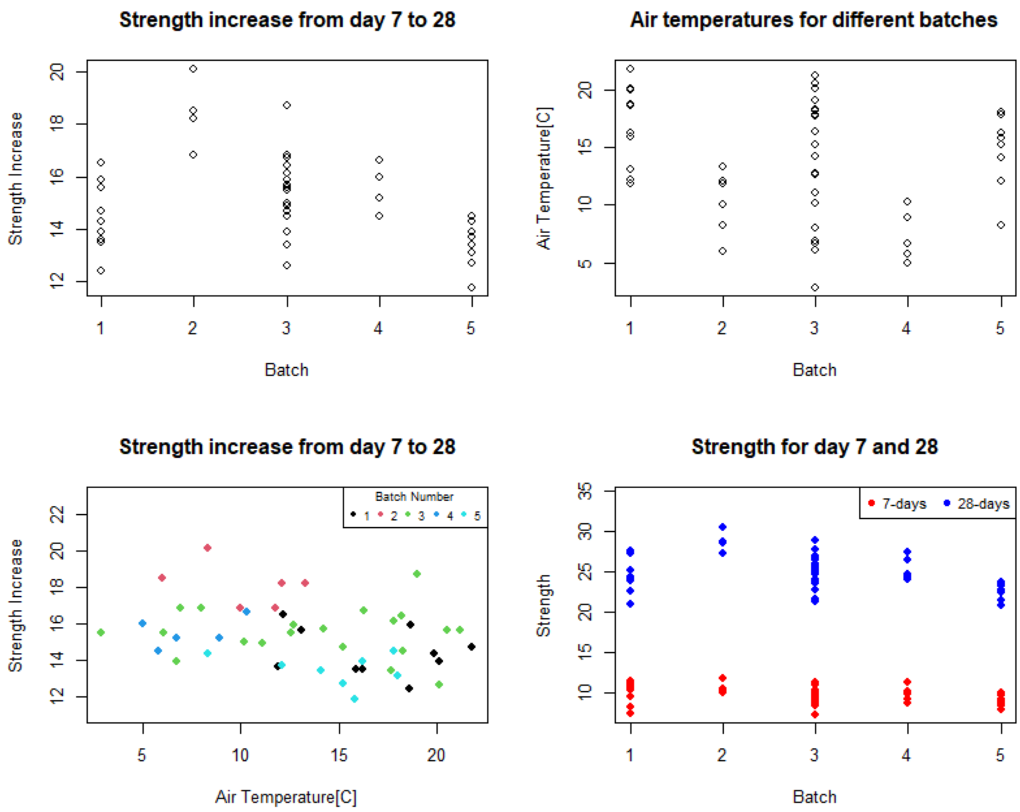


Figure A.2 – Different plots to investigate if batch variation is present.

We see from the bottom left plot in figure A.2 that there is a slight negative trend between air temperature and the strength increase from day 7 to day 28. On the other hand, if we look at each batch individually, this trend does not seem to be as apparent. In the top left plot, we see some variation in the strength increase but again we cannot partition the variance easily and say that the variance is only from the batch or some variance could be described by the air temperature. If we compare the top right plot with the top left plot, we see that the air temperature seems to be low when the strength is high and vice versa. Therefore, we should definitely investigate how the air temperature affects the concrete strength and assess this relative to the batch variance. The bottom right plot shows that the strength increase is quite consistent from day 7 to 28.

A.2 Exercise 2

We will now estimate the mean concrete strengths for day 7 and day 28. The estimates can be seen in table A.2.

	batch 1	batch 2	batch 3	batch 4	batch 5
$\overline{y7}$	10.16	10.62	9.61	9.90	8.99
$\overline{y28}$	24.55	28.72	25.09	25.40	22.41

Table A.2 – Batch means for day 7 and day 28 concrete strengths.

We see that the means varies quite a bit from batch to batch but as we also saw in exercise

A.1 we do not know if this variation is a batch variation or a air temperature variation.

A.3 Exercise 3

We will now try to fit a generalized mixed effect model to predict the day 28 strength using the other variables. To fit a mixed effect model, we will use the model building framework given in slide 51 from lecture 9 which we have repeated here in figure A.3.

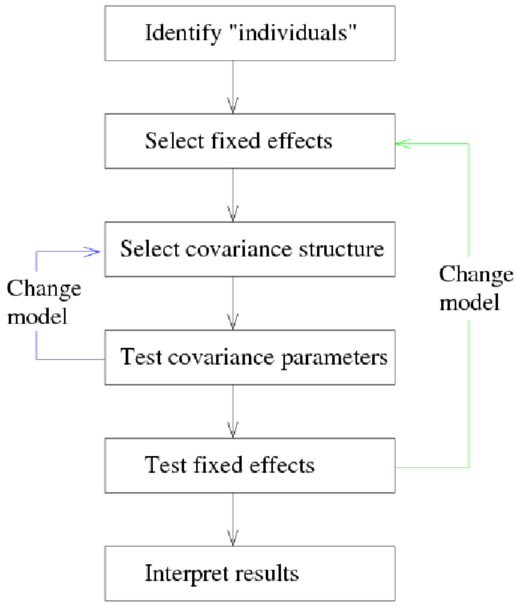


Figure A.3 – Mixed model fitting framework

We will use a linear mixed model where both the fixed and random effects are Gaussian distributed. Such a model has the general form given in A.3.1.

$$\begin{aligned} \boldsymbol{Y} \mid \boldsymbol{U} = \boldsymbol{u} &\sim N(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{u}, \boldsymbol{\Sigma}) \\ \boldsymbol{U} &\sim N(\boldsymbol{0}, \boldsymbol{\Psi}) \end{aligned} \tag{A.3.1}$$

For our initial model the parameters in A.3.1 are given as

$$\boldsymbol{X} = \begin{bmatrix} \mathbf{1}_N & \text{y7} & \text{air.temp} & (\text{y7*air.temp}) \end{bmatrix} \tag{A.3.2}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \mu \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} \tag{A.3.3}$$

$$\boldsymbol{Z} = \begin{bmatrix} \mathbb{1}(i=1)_N & \mathbb{1}(i=2)_N & \mathbb{1}(i=3)_N & \mathbb{1}(i=4)_N & \mathbb{1}(i=5)_N \end{bmatrix} \tag{A.3.4}$$

$$\boldsymbol{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{bmatrix} \tag{A.3.5}$$

$$\boldsymbol{\Sigma} = \sigma^2 I_N \tag{A.3.6}$$

$$\boldsymbol{\Psi} = \sigma_u^2 I_k \tag{A.3.7}$$

where $\mathbf{1}_N$ is a N long column vector of ones and $\mathbb{1}(i = i')_N$ is a N long column vector which is one if the observation is from group i' and 0 otherwise. Before we can reduce our initial model to only include necessary terms, we need to test if our distributional assumptions are correct. From [1] we know that we make two major assumptions when we choose a linear mixed model where both the fixed and random effects are Gaussian distributed.

> *Assumption 1* - the within-group errors are independent and identically normally distributed, with mean zero and variance $\boldsymbol{\Sigma}$, and they are independent of the random effects.
>
> *Assumption 2* - the random effects are normally distributed, with mean zero and covariance matrix $\boldsymbol{\Psi}$ (not depending on the group) and are independent for different groups;

From section 4.3 in [1], we know we can test assumptions 1 by plotting the residuals against the fitted values and explanatory values and subsequently use a QQ plot for the residuals. For assumption 2 we can plot a QQ plot for the estimated random effects which should be normal distributed.

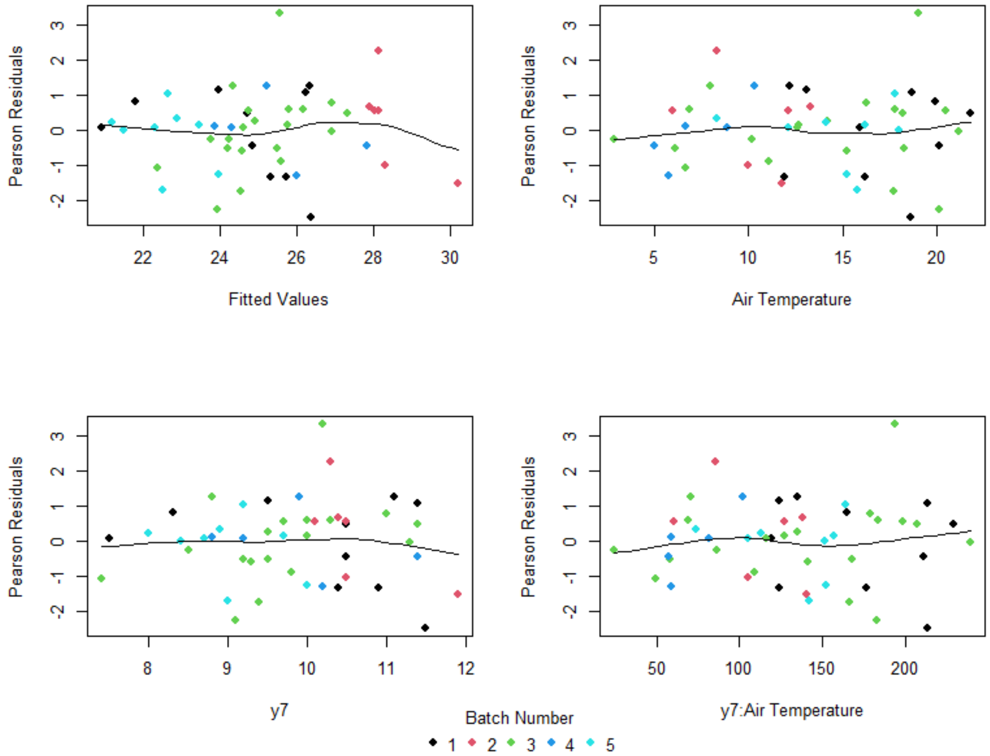


Figure A.4 – Pearson residuals plotted against the fitted values and explanatory variables.

In figure A.4, we see that the there seems to be no alarming issues in the residuals for the fixed effects. They seems quite normal, no huge outliers or increase in variance. Further, there also do not seem to be any huge batch differences.

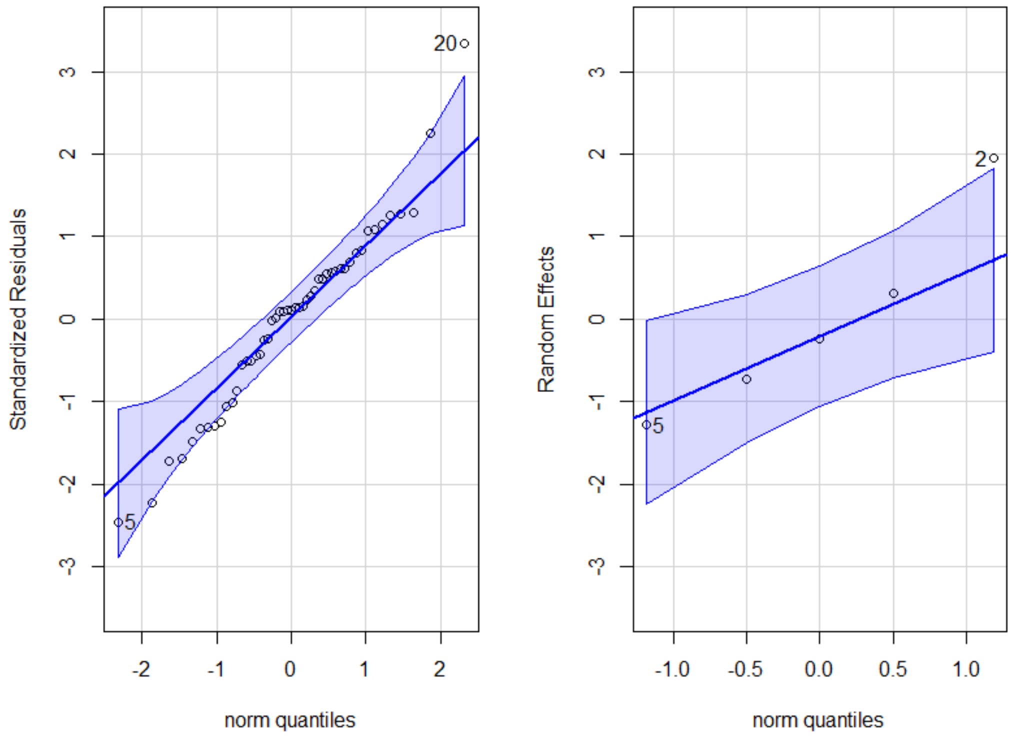


Figure A.5 – QQ plots for the Pearson residuals and the random effects.

From the left plot in figure A.5, we see that the Pearson residuals look normal. For the right plot we see that the random effects also seem normal though it is hard to say much based on 5 points.

Fitting the model

To test covariance parameters we will perform the likelihood ratio test described in Theorem 5.2 in [2] and to test fixed effects we will the package `lmerTest` which implements the Satterthwaite's method. We will use the Satterthwaite's method within a type II selection framework. We start by testing the random effect in the full model.

Df	χ^2_{obs}	p-value
1	11.545	0.00068

Table A.3 – Likelihood ratio test for the random effect in the full model.

We hence keep the random effect and proceed to testing the fixed effects.

Interaction	Numerator Df	Denominator Df	F value	p value
air.temp:y7	1	42.396	0.0244	0.8767

Table A.4 – Type II backward selection of the full model

Hence we can drop the interaction 'air.temp:y7' and obtain the new model we have written in R notation in A.3.8.

$$\texttt{y28} \sim \texttt{y7 + air.temp + (1|batch)} \tag{A.3.8}$$

According the test framework given in figure A.3, we need to test the random effect again.

Df	χ^2_{obs}	p-value
1	11.596	0.00066

Table A.5 – Likelihood ratio test for the random effect in the full model.

We see that the random effect is still significant and proceed to testing the fixed effects.

Interaction	Numerator Df	Denominator Df	F value	p value
air.temp	1	45.575	2.9035	0.0952

Table A.6 – Type II backward selection of the model A.3.8.

Hence we can drop the mixed term 'air.temp' and obtain the new model we have written in R notation in A.3.9.

$$\texttt{y28} \sim \texttt{y7 + (1|batch)} \tag{A.3.9}$$

We then again test the random effect.

Df	χ^2_{obs}	p-value
1	18.096	$2.1e-05$

Table A.7 – Likelihood ratio test for the random effect for the model A.3.9.

Hence the random effect is still significant and we can test the fixed effect.

Interaction	Numerator Df	Denominator Df	F value	p value
y7	1	44.427	54.426	3.036e-09

Table A.8 – Type II backward selection of the model A.3.9

Hence we have found all the sufficient and necessary terms and conclude at the model given in A.3.9. To connect the R notation in A.3.9 with the general model given A.3.1 we will give the model parameters for the final model as we did for the full model.

$$\boldsymbol{X} = \begin{bmatrix} \mathbf{1}_N & \text{y7} \end{bmatrix} \tag{A.3.10}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \mu \\ \beta_1 \end{bmatrix} \tag{A.3.11}$$

$$\boldsymbol{Z} = \begin{bmatrix} \mathbb{1}(i=1)_N & \mathbb{1}(i=2)_N & \mathbb{1}(i=3)_N & \mathbb{1}(i=4)_N & \mathbb{1}(i=5)_N \end{bmatrix} \tag{A.3.12}$$

$$\boldsymbol{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{bmatrix} \tag{A.3.13}$$

$$\boldsymbol{\Sigma} = \sigma^2 I_N \tag{A.3.14}$$

$$\boldsymbol{\Psi} = \sigma_u^2 I_k \tag{A.3.15}$$

and from REML we have obtained the following parameter estimates given with their profile confidence interval.

	2.5 %	β	97.5 %
μ	8.299	12.154	15.838
β_1	0.978	1.326	1.700
σ	0.969	1.191	1.477
σ_u	0.694	1.520	3.058

Table A.9 – Parameter estimates for A.3.9.

A.4 Exercise 4

We saw in section A.1 that maybe the air temperature had an effect on the concrete strength but in section A.3 that the air temperature was not significant in the mixed effect model. Hence we must conclude that the concrete strength at day 28 is only dependent on the strength at day 7 and not the air temperature as we can see from table A.4 and A.6. To visualize the model we will plot confidence and prediction intervals for the five conditional expectation and the one unconditional expectation. We have used the library `ciTools` to find the intervals which is described here, [3].

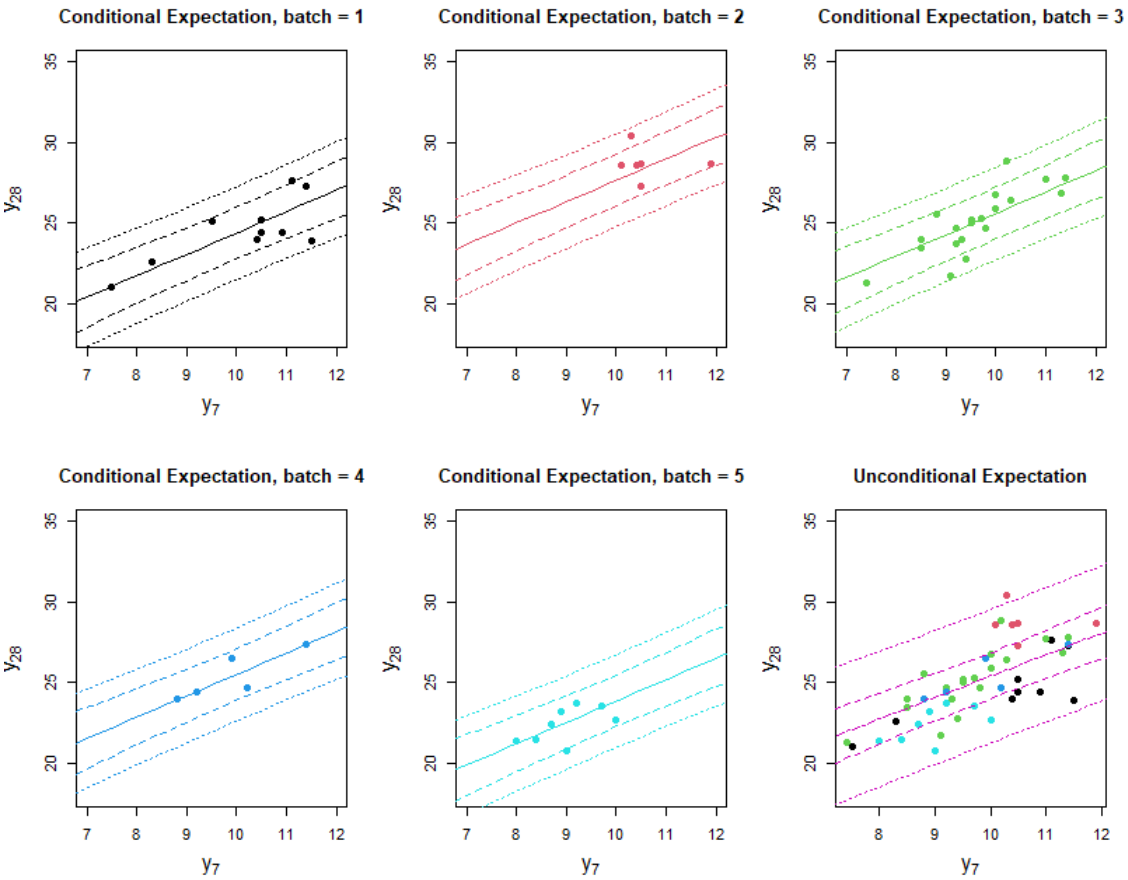


Figure A.6 – Confidence and prediction intervals for the five conditional expectation and the one unconditional expectation.

B Part 1, B

B.1 Exercise 5

We will now try to model both the 7 and 28 day strength simultaneously. Therefore, we need to use a multivariate mixed effect model. We will model the strengths by assuming a common intercept as fixed effect and the batch as random effect. Hence we know from equation 5.84 in [2] that the model will be of the form

$$\boldsymbol{Y}_{ij} = \boldsymbol{\mu} + \boldsymbol{u}_i + \boldsymbol{\epsilon}_{ij}; \quad i = 1, \dots 5; \quad j = 1, \dots n_i \tag{B.1.1}$$

where $\boldsymbol{Y}_{ij} = \begin{bmatrix} y_{7,i,j} & y_{28,i,j} \end{bmatrix}^T$, $\boldsymbol{\mu} = \begin{bmatrix} \mu_7 & \mu_{28} \end{bmatrix}^T$, $\boldsymbol{u}_i \overset{i.i.d}{\sim} \mathcal{N}_2(0, \boldsymbol{\Sigma}_0)$ and $\boldsymbol{\epsilon}_{ij} \overset{i.i.d}{\sim} \mathcal{N}_2(0, \boldsymbol{\Sigma})$.

B.2 Exercise 6

For the model B.1.1 we have three model parameters, $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_0$.From Theorem 5.11 in [2] we know that

$$\begin{aligned} \tilde{\mu} &= \bar{y}_{++} \\ \tilde{\Sigma} &= \frac{1}{N-k}\mathbf{SSE} \\ \tilde{\Sigma}_0 &= \frac{1}{n_0}\left(\frac{\mathbf{SSB}}{k-1} - \tilde{\boldsymbol{\Sigma}}\right) \end{aligned} \tag{B.2.1}$$

We need to estimate the total mean of y7 and y28, $\bar{y}_{++}$, the within group variation, **SSE**, the between group variation, **SSB**, and weighted average group size, n_0. From equation 5.86 in [2] we know that $\bar{y}_{++}$ is given as

$$\bar{y}_{++} = \sum_{i=1}^{k}\sum_{j=1}^{n_i} \frac{Y_{ij}}{N} = \begin{bmatrix} 9.77 \\ 25.02 \end{bmatrix} \tag{B.2.2}$$

From equation 5.87 and 5.88, we know that **SSE** and **SSB** are given as

$$\mathbf{SSE} = \sum_{i=1}^{k}\sum_{j=1}^{n_i} \left(Y_{ij} - \bar{Y}_{i+}\right)\left(Y_{ij} - \bar{Y}_{i+}\right)^T = \begin{bmatrix} 43.07 & 55.46 \\ 55.46 & 132.33 \end{bmatrix} \tag{B.2.3}$$

$$\mathbf{SSB} = \sum_{i=1}^{k} n_i \left(\bar{Y}_{i+} - \bar{Y}_{++}\right)\left(\bar{Y}_{i+} - \bar{Y}_{++}\right)^T = \begin{bmatrix} 11.35 & 33.27 \\ 33.27 & 139.43 \end{bmatrix}. \tag{B.2.4}$$

where $\bar{Y}_{i+}$ are the group means. Lastly we have n_0 which is given in equation 5.29 as

$$n_0 = \frac{N - \frac{\sum_{i=1}^{k} n_i^2}{N}}{k-1} = 9.06 \tag{B.2.5}$$

We can now estimate the parameters and obtain

$$\tilde{\mu} = \begin{bmatrix} 9.77 \\ 25.02 \end{bmatrix} \tag{B.2.6}$$

$$\tilde{\Sigma} = \begin{bmatrix} 0.98 & 1.26 \\ 1.26 & 3.01 \end{bmatrix} \tag{B.2.7}$$

$$\tilde{\Sigma}_0 = \begin{bmatrix} 0.21 & 0.78 \\ 0.78 & 3.51 \end{bmatrix} \tag{B.2.8}$$

The total mean, $\tilde{\mu}$, is easy to understand but the two variation terms are a bit trickier to grasp. We have hence in figure B.1 plotted the total variation together with the two sources of variation, $\tilde{\Sigma}_0$ and $\tilde{\Sigma}$.

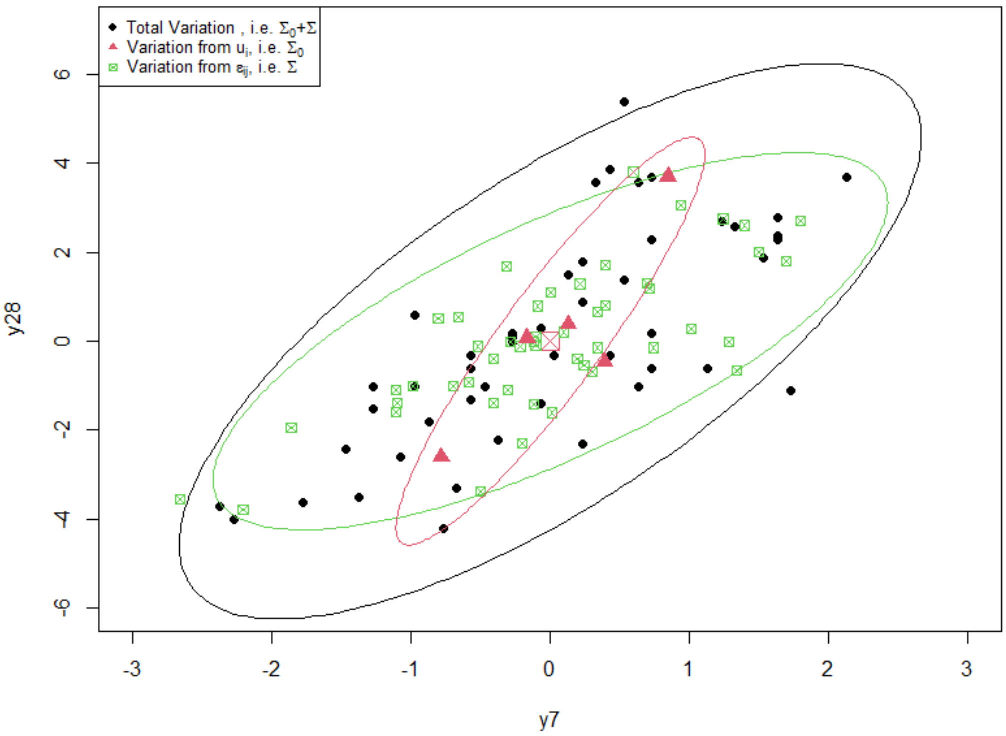


Figure B.1 – The total variation of model B.1.1 together with the two sources of variation, $\tilde{\Sigma}_0$ and $\tilde{\Sigma}$.

B.3 Exercise 7

In last exercise of part 1 we need to estimate the correlation between day 7 and day 28 including a confidence interval. We will understand the correlation between day 7 and day 28 as how they are correlated taking all variation into account, i.e. the total variation where we assume day 7 and day 28 is from the same observation. Hence we know that the marginal distribution for $Y = [y7, y28]^T$ is

$$Y \sim \mathcal{N}_2(\mu, \tilde{\Sigma}_0 + \tilde{\Sigma}) \tag{B.3.1}$$

We know that because Y is normal distributed, all of the correlation between $y7$ and $y28$ will be described by Pearson correlation coefficient. It is given as

$$r_{y7,y28} = \frac{Cov(y7, y28)}{\sigma_{y7}\sigma_{y28}} = \frac{2.04}{1.08 \cdot 2.55} = 0.73 \tag{B.3.2}$$

From [4] we know that the confidence interval for $r_{y7,y28}$ we can obtain from the Fisher transformation. Fisher proposed the transformation $z = arctanh(r)$, which is the inverse hyperbolic tangent function. When r is the sample correlation for bivariate normal data and z = arctanh(r) then the following statements are true.

- The distribution of z is approximately normal and tends to normality rapidly as the sample is increased
- The standard error of z is approximately $1/\sqrt{N-3}$, which is independent of the value of the correlation.

We can hence use this transformation to calculate a normal confidence interval in the z-domain and the transform it back. We calculate the confidence interval by

$$\begin{aligned}&\rho_{y7,y28} \in \left[\tanh\left(\operatorname{arctanh}(r_{y7,y28}) \pm \Phi^{-1}\left(\alpha/2, \operatorname{arctanh}(r_{y7,y28}), 1/\sqrt{N-3}\right)\right)\right] \Rightarrow \\ &\rho_{y7,y28} \in [0.57, 0.84]\end{aligned} \tag{B.3.3}$$

We depict the transformation in figure B.2, where we see that the correlation in the transformed domain is normal and can take values on the whole real line. After we transform back all density in mapped to the interval $[0, 1]$.

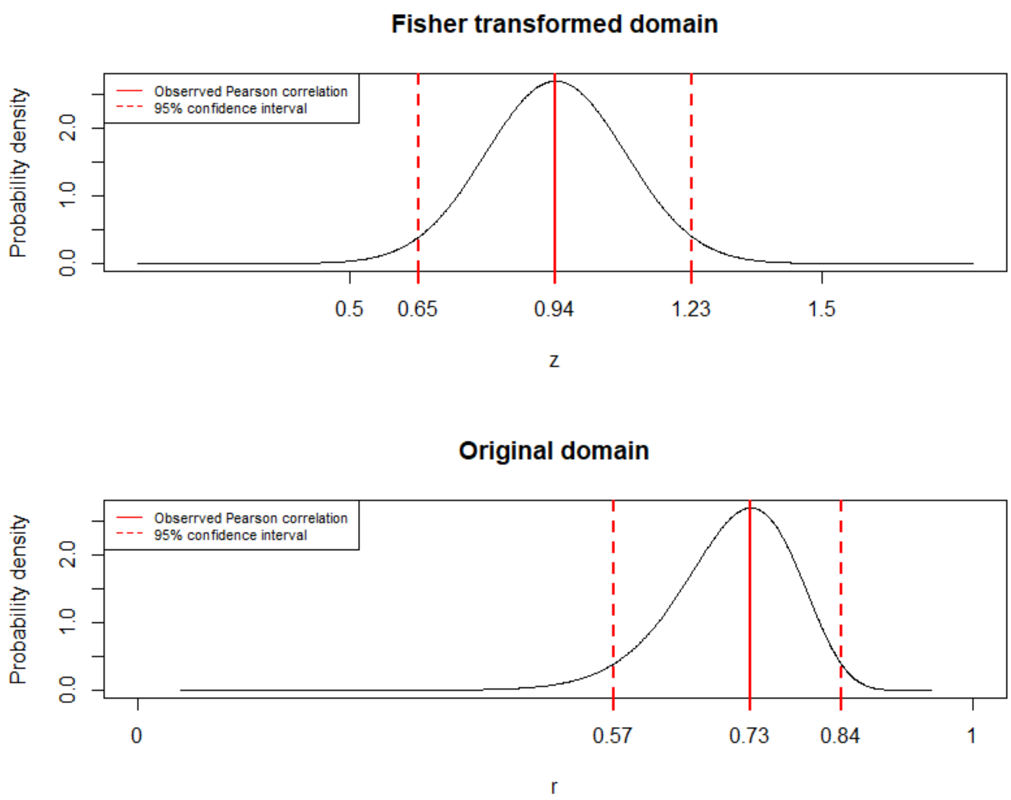


Figure B.2 – Illustration of the mapping back and forth from the Fisher transformed domain.

Hence if we consider the confidence of the correlation, the total variation of the data follow some normal distribution between the red and green 95% ellipse depicted in B.3.

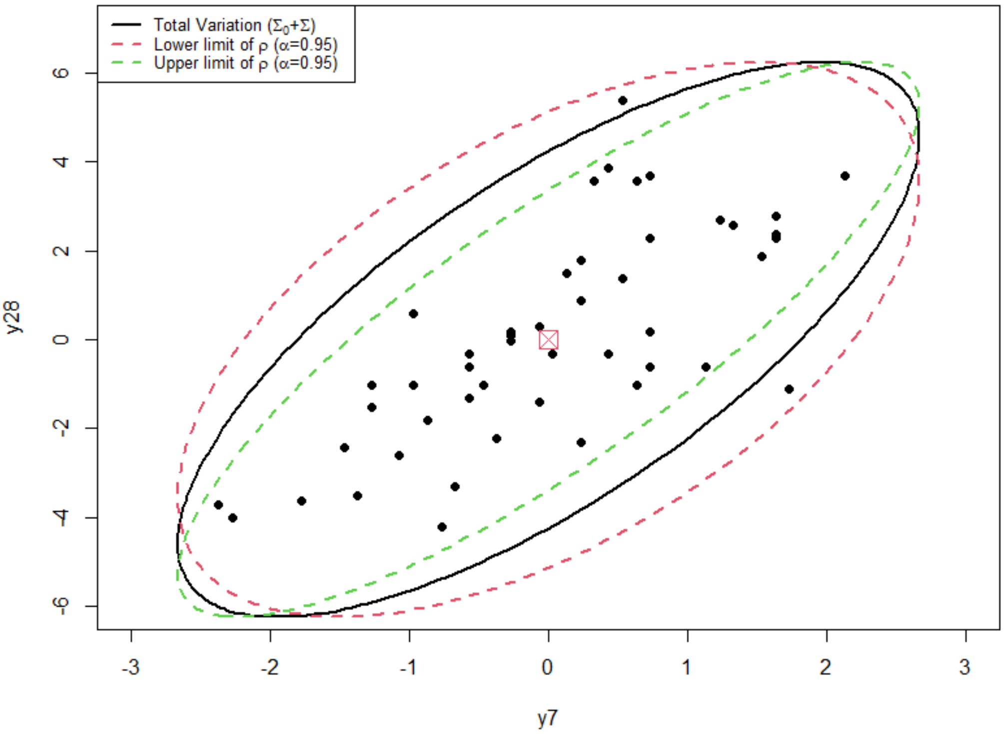


Figure B.3 – Total variation distribution for the confidence interval of the correlation.

C | Part 2, A

In this part of the assignment, we will work with the clothing insulation data that we also analyzed in assignment 2. Here we have the following variables.

Variable	Type	Description
clo	Integer	Number of changes
t0ut	Continuous	Outdoor temperature
tIn0p	Continuous	Indoor operating temperature
sex	Factor	Sex of the subject
subjId	Factor	Identifier for subject
time	Continuous	Total measurement time
day	Factor	Day (within the subject)
nobs	Integer	Measurement number (within the day)

(C.0.1)

We refer to assignment 2 for a thorough introduction of all the variables and their relation to the response variable *clo*. We will focus our analysis on the differences between subjects to avoid a complete replica of the descriptive analysis from assignment 2. Then we will construct two generalized mixed effect models with the following distributional assumptions of the residuals; binomial and Poisson.

C.1 Data Analysis with Focus on Difference Between Subjects

We are to predict the clothing level and hence we will now consider how it is distributed across subject IDs.

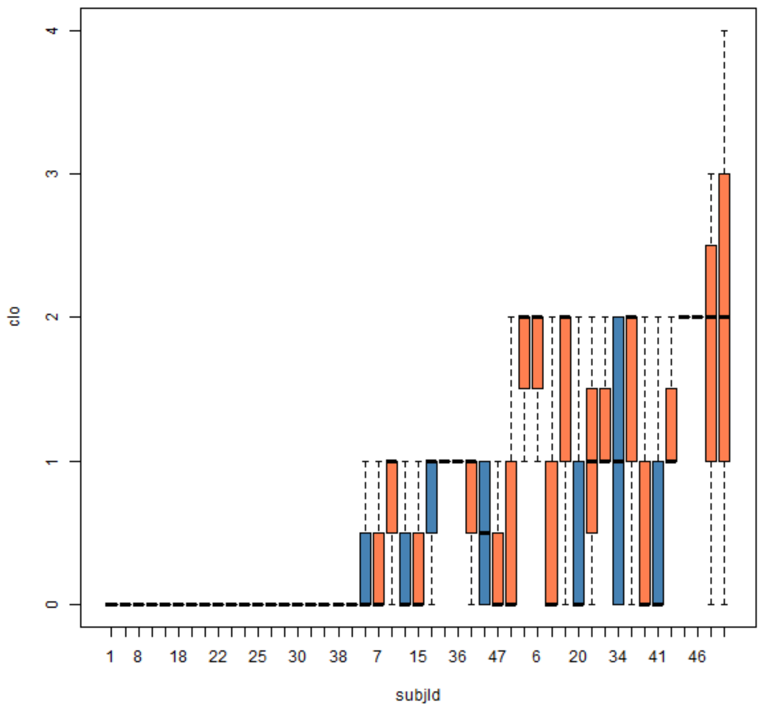


Figure C.1 – Boxplots for the clothing number of cloth changes for each subject id where we have sorted the Ids according to the max number of changes at one day. The orange cones are feamles and the blue are males.

In figure C.1, we see the boxplot for each subject ID and their number of changes. The essential thing to notice is that many subjects do not change cloth at all on any of the days. Further, we observe that a couple of the subjects changes cloth between 0 and 2 times, some always change two times and lastly there are two people that changes cloth between 0 and 3-4 times. There seems to be quite a difference in the variation in spread and central tendency of the changes in clothing level across subject Ids. For now, it also seems that there is a slightly larger variation for the females but we will come back to this. We now consider the mean values for each subject:

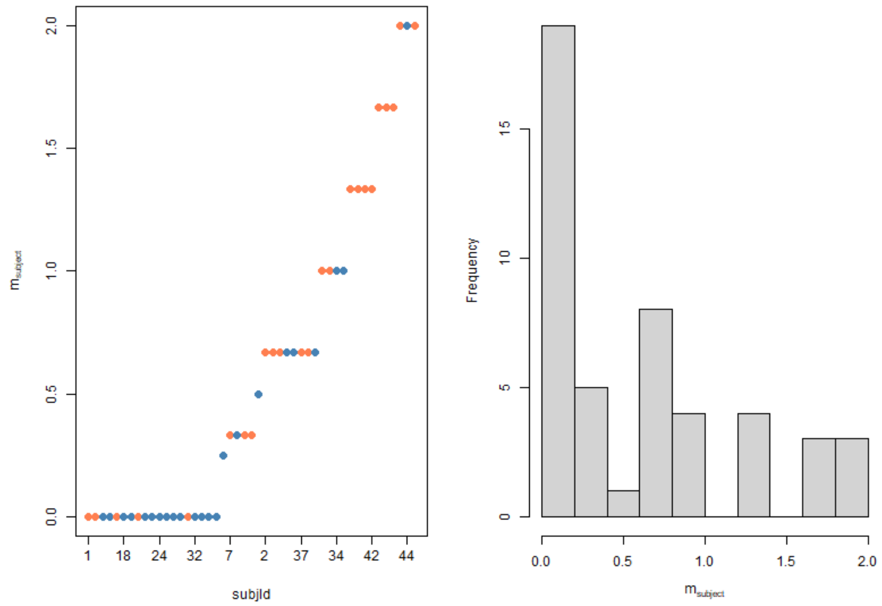


Figure C.2 – Distribution of mean values. To the left we have ordered the mean values increasingly. To the right we see the histogram of the mean value parameters.

In figure C.2, we see that there is a great difference in the mean number of clothing changes across subjects. As we saw before many do not change at all, and then we see a small accumulation around 0.66 times each day which we see as this second mode.
We are to model the clothing level and from assignment 2, we know that sex is a very important predictor. Therefore, we consider the empirical PMF for each sex.

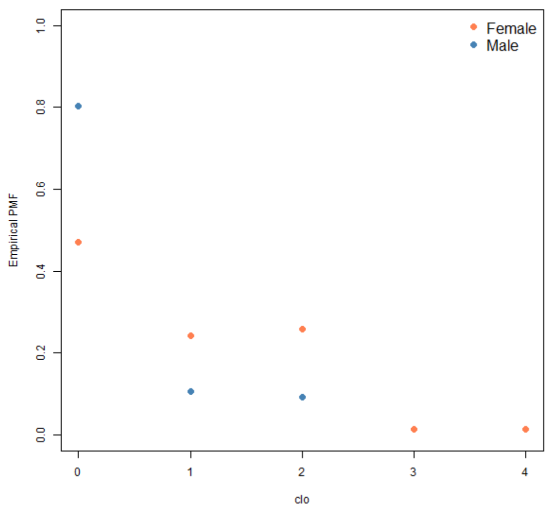


Figure C.3 – Empirical probability mass function for each sex.

In figure C.3, we understand why we are asked to fit a binomial and Poisson distribution; the responses are positive integers. We also see that there seems to be some variance that might be difficult to capture for clo changes around 1 and 2, and of cause this zero inflation could also be hard to capture. Notice also that the response is distributed over only a limited set of values.

C.2

In the following, we will use the glmmTMB framework to develop a generalized mixed effect model with a Binomial and Poisson distribution.
In the glmmTMB framework, we formulate hierarchical models using the following generic equation:

$$\begin{aligned} \boldsymbol{\eta} &= \boldsymbol{X}\theta + \boldsymbol{U} \\ \boldsymbol{U} &\sim N(\mathbf{0}, \boldsymbol{\Psi}) \\ E[\boldsymbol{Y} \mid \boldsymbol{U}] &= g^{-1}(\boldsymbol{\eta}) \end{aligned} \tag{C.2.1}$$

Explicitly for each of the models we have:

$$\begin{aligned} \boldsymbol{Y}|\boldsymbol{U} = \boldsymbol{u} &\sim \text{Pois}\left(\boldsymbol{X}\theta + \boldsymbol{u}\right) \\ \boldsymbol{U} &\sim N(\mathbf{0}, \boldsymbol{\Psi}) \end{aligned} \tag{C.2.2}$$

and

$$\begin{aligned} \boldsymbol{Y}|\boldsymbol{U} = \boldsymbol{u} &\sim \text{Binomial}\left(n_i, \boldsymbol{X}\theta + \boldsymbol{u}\right) \\ \boldsymbol{U} &\sim N(\mathbf{0}, \boldsymbol{\Psi}) \end{aligned} \tag{C.2.3}$$

where we will use the canonical link function i.e. log for Poisson and logit for the binomial. In the following, we will use different modelling approaches approaches:

- forward selection approach. Here we will built on intuition from assignment 2.
- backward selection approach. Here we will create a generic framework and do backward selection.

C.2.1 Forward Selection Approach

From assignment 2, we found the best generalized linear model to describe the data using either Poisson or binomial was

$$\texttt{clo} \sim \texttt{sex} \tag{C.2.4}$$

Using the pipeline from in figure A.3, we should hence suggest a structure for the random effect and test if it is significant. We only have a limited set of data hence we would not suggest a hugely complex structure. Further, for the model to be useful, it would be natural to select a random effect on the intercept of subject ids. Therefore, we now consider the following model,

$$\texttt{clo} \sim \texttt{sex + (1|subjId).} \tag{C.2.5}$$

We now make a likelihood ratio test on the random effect for both models:

Df	χ^2_{obs}	p-value
1	12.735	0.00036

Table C.1 – Likelihood ratio test for the random effect in the binomial

Df	χ^2_{obs}	p-value
1	6.93111	0.0085

Table C.2 – Likelihood ratio test for the random effect in the Poison

In both cases, we see that the random effect is significant. The parameter estimates and explicit model formulation will be displayed below.

C.2.2 Backward Selection Approach

In this approach, we use the results of the data exploration in assignment 2 but propose a larger initial model structure and reduce the fixed effects using a sequence of type-II tests. Here we will again use the model building framework depicted in figure A.3. We have a limited set of data hence we will not consider second order interaction terms and only random effects on the intercepts.We will therefore initialize with the model:

$$\begin{aligned}\texttt{clo} \sim \texttt{factor(sex)} + \texttt{tOut + tInOp + factor(sex):tOut} \\ + \texttt{factor(sex):tInOp - 1 + (1|subjId)}\end{aligned} \tag{C.2.6}$$

Note that `tOut:tInOp` is not included as it gave rise to numerical instability.
The idea is now that we fix the structure in the fixed effect model but introduce the random effect and test if the random effect is significant. If it is significant, we keep it. Then we fix the random effect but reduce the fixed effect model structure. The idea is inspired also from the model selection procedure introduced in lmerTest [5].
In table G.1 each type-II test can be found for each reduction for the Poisson and in table G.2 the test can be found for the binomial. In both cases, we obtain the same linear predictor where we have a different intercept for each sex:

$$\boldsymbol{\eta} = \boldsymbol{X}\theta + \boldsymbol{u} \tag{C.2.7}$$

where

$$\boldsymbol{X} = \begin{bmatrix}\mathbb{1}(j = \text{'Female'})_N & \mathbb{1}(j = \text{'Male'})_N\end{bmatrix}, \qquad \boldsymbol{\theta} = \begin{bmatrix}\beta_{Female} \\ \beta_{Male}\end{bmatrix} \tag{C.2.8}$$

and in our case $\mathbf{u} \sim \mathcal{N}(0, \psi^2)$ is just a vector where ψ is the standard deviation.

C.2.3 Parameter Estimates

The following parameter estimates are obtained for Poisson with Wald confidence intervals:

	2.5 %	Estimate	97.5 %
β_{Female}	-0.75	-0.34	0.07
β_{Male}	-2.11	-1.50	-0.88
ψ	0.38	0.68	1.21

Table C.3 – Parameter estimates for Poisson.

The following estimates are obtained for the binomial:

	2.5 %	Estimate	97.5 %
β_{Female}	-2.28	-1.79	-1.29
β_{Male}	-3.86	-3.14	-2.42
ψ	0.55	0.91	1.49

Table C.4 – Parameter estimates for the Binomial.

With the backward and forward selection approach we find that we should include sex hence this model is deemed appropriate. However, we should note that for the Poisson model, zero is included in the confidence interval for β_{female}. We will comment on this later when we consider model performance.

C.2.4 Residual Diagnostics

We will now check the residuals. As in assignment 2, we should again note that as the responses are integers, the usual residual checking is not appropriate as seen in figure C.4.

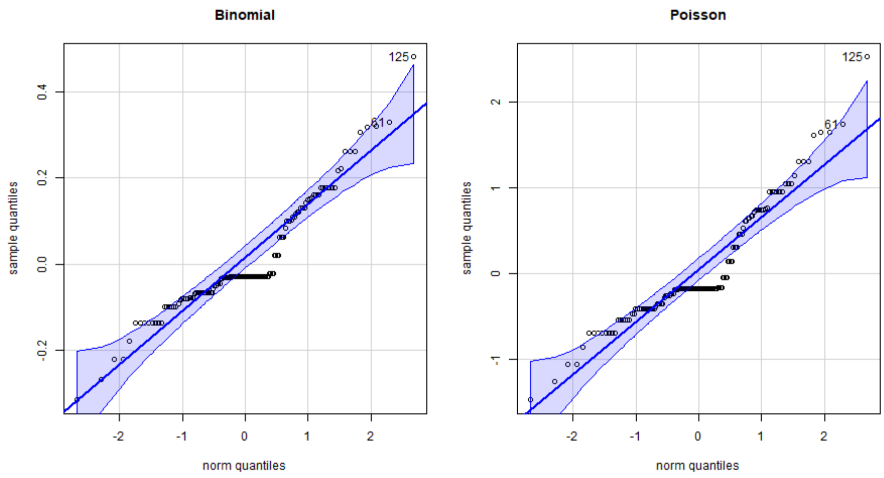


Figure C.4 – QQ-Plots for the reponse residuals of the model.

We could plot the Pearson or deviance residuals but these are also not appropriate. We will instead use the package DHARMa - residual diagnostics for hierarchical (multi-level/mixed) regression models, [6]. Here they introduced scaled or quantile residuals. The concept is like a parametric boostrap:

1. We simulate data with the fitted model. Here we take into account the predictor combination of each observation.
2. For each of our observation, we can now calculate a empirical CDF based on the simulated data.
3. The scaled residual is now found as the value of the empirical density function at the value of the observed data.

Intuitively, we simulate a great deal of samples from the model for each predictor combination. Then we compare each of our observations with the simulated data. A scaled residual of 0.5 means that half of the simulated data is above the observation and the other half is below. We therefore expect the scaled residuals to be uniform if we have formulated an adequate model. We can assess this in figure C.5:

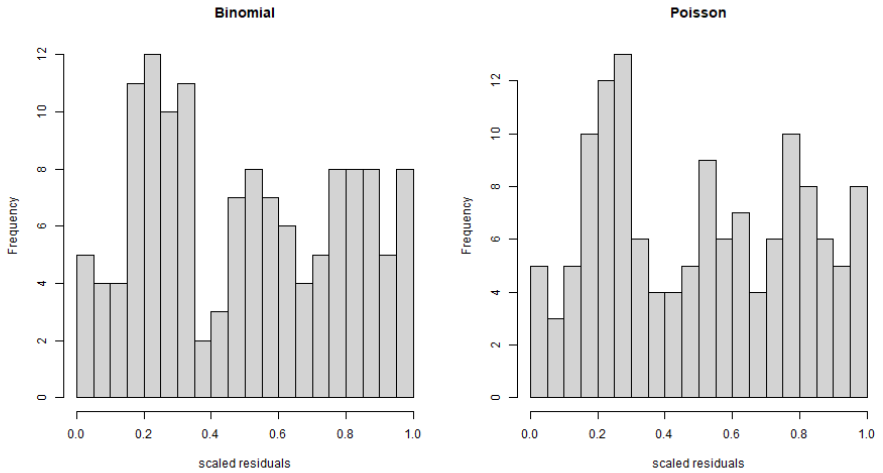


Figure C.5 – The scaled residuals. An adequate model would have uniformly distributed scaled residuals.

In figure C.5 we see that the scaled residuals are not perfectly uniform. Of cause, we have to acknowledge that we work with finite samples and hence will have to adjust our visual expectations. In appendix in figure G.1 we plot the qqplot for a uniform distribution where we also see that it look reasonable. To transform this into a well-known domain, we will use the inverse normal cdf. Let $S \sim U\left[0,1\right]$ be a uniform random variable, then we can transform this directly to a standard normal using the inverse CDF of a standard normal, Φ^{-1}:

$$S_{\mathcal{N}} = \Phi^{-1}\left(S\right). \tag{C.2.9}$$

Hence indeed $S_{\mathcal{N}} \sim \mathcal{N}\left(0,1\right)$ if $S \sim U\left[0,1\right]$. This idea makes us able to transform this into a well known domain of checking if a standard normal is seen:

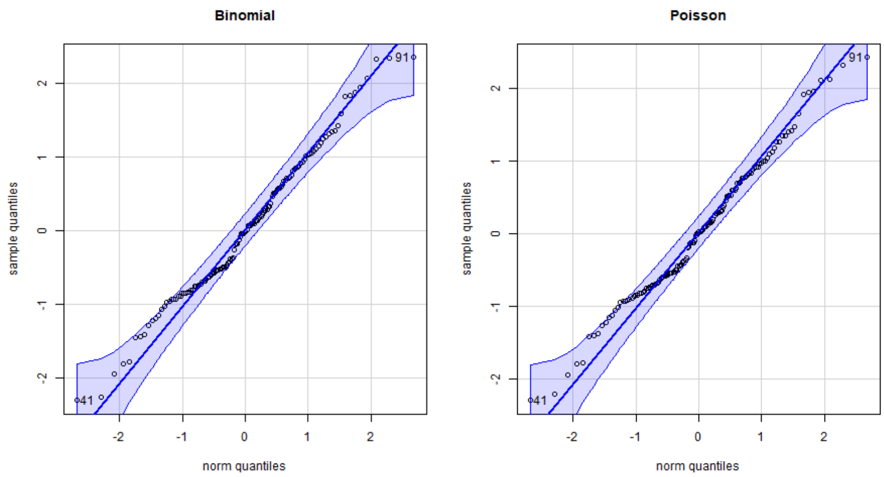


Figure C.6 – QQ-plots for the scaled residuals transformed to a normal domain.

In figure C.6, we see that the residuals looks reasonable only with slight deviations. We conclude that the residuals are ok.

C.2.5 Model Performance

To see how the model performs, we plot the predictions on the number of changes which can be seen in figure C.7.

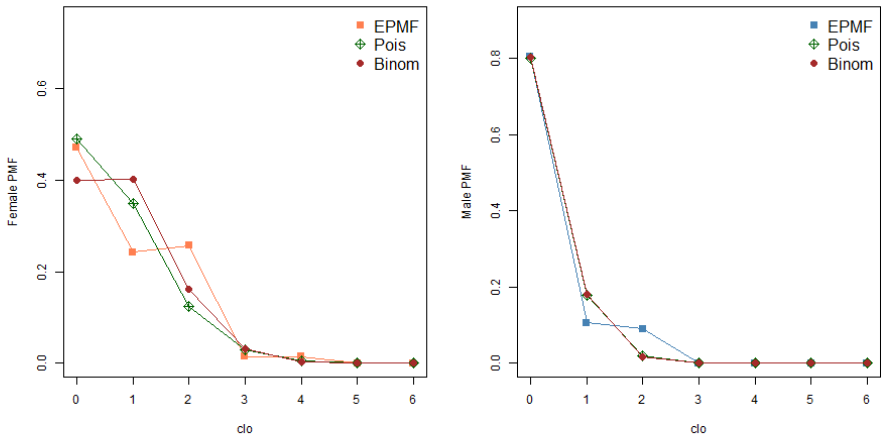


Figure C.7 – The predictions on number of clothing changes with the hierarchical Poisson and Binomial for each sex. Note the different scales in each figure.

In figure C.7, we notice that both models perform reasonably well on the male dataset while they both have problems with the high variability for clothing changes around 1-2 for the females.
It seems that the models perform equally well and to quantify this, we present information criteria and log likelihood below.

	Binomial	Poisson
AIC	263.37	266.10
BIC	272.10	274.83
log Lik	-128.68	-130.05

We see that with a binomial distributional assumption both the AIC and BIC is smaller which indicates that the Binomial is a better model. Likewise the log likelihood is slightly larger.

C.2.6 Sanity Check

In the following, we will work with new priors for the hierarchical Poisson model. Before we do that, we will make a sanity check to see if we have any benefits of using the hierarchical models. Therefore we compare it with a model with the fixed effects only. For the fixed effect model, we obtain an AIC of 271.027 and BIC of 276.8523 hence there is a benefit of using the hierarchical model. Graphically, we have:

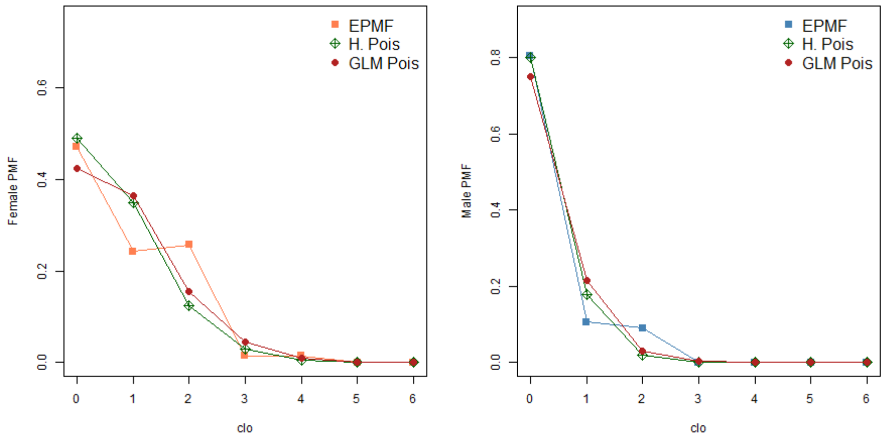


Figure C.8 – The predictions on number of clothing changes with the hierarchical Poisson and a generalised linear model with the same fixed effects.

In the AIC and BIC we saw a slight improvement but in figure C.8 it is hard to see any obvious substantial improvements.

D Part 2, B

D.1 Exercise 1

We now extend our framework to non-Gaussian distributed random effects. We will consider the Poisson-Gamma model which can be written as

$$\begin{aligned} Y|u &\sim \text{Pois}\left(\mu\left(Zu\right)\right) \\ U &\sim \text{Gamma}\left(\alpha, \frac{1}{\alpha}\right) \\ \mu &= \exp\left(X\beta\right) \end{aligned} \tag{D.1.1}$$

Normally when we want to fit a mixed effect model we integrate the random effects out of the joint likelihood to obtain the marginal likelihood as shown in D.1.2.

$$\begin{aligned} L_M\left(\theta; y\right) &= \int_{\mathbb{R}^k} f_{Y|u}\left(y; u, \beta\right) f_U\left(u; \alpha\right) du \\ &= \int_{\mathbb{R}^k} L\left(y; u, \theta\right) du, \end{aligned} \tag{D.1.2}$$

where $u \in \mathbb{R}^k$ and $\theta = \{\beta, \alpha\}$. This can though often become analytical intractable when other distributions than the Gaussian is used for the fixed or random effects. Hence one must resort to numerical approximation such as Monte Carlo methods or the Laplace approximation as we will use here. The Laplace approximation makes a 2nd order Taylor expansion of the joint likelihood at the point where u maximizes the joint likelihood.

$$\ell(\theta, u, y) \approx \ell\left(\theta, \hat{u}_\theta, y\right) - \frac{1}{2}\left(u - \hat{u}_\theta\right)^T \left(- \ell''_{uu}(\theta, u, y)\big|_{u=\hat{u}_\theta}\right)\left(u - \hat{u}_\theta\right) \tag{D.1.3}$$

where $\ell(\theta, u, y)$ denotes the log joint likelihood and,

$$\hat{u}_\theta = \arg\max_u L\left(y; u, \theta\right). \tag{D.1.4}$$

This means that we can approximate the integral D.1.2 by the new integral, D.1.5.

$$\begin{aligned} L_M(\theta, y) &\approx \int_{\mathbb{R}^k} e^{\ell(\theta, \hat{u}_\theta, y) - \frac{1}{2}(u-\hat{u}_\theta)^T \left(- \ell''_{uu}(\theta, u, y)|_{u=\hat{u}_\theta}\right)(u-\hat{u}_\theta)} du \\ &= L\left(\theta, \hat{u}_\theta, y\right) \int_{\mathbb{R}^k} e^{-\frac{1}{2}(u-\hat{u}_\theta)^T \left(- \ell''_{uu}(\theta, u, y)|_{u=\hat{u}_\theta}\right)(u-\hat{u}_\theta)} du \end{aligned} \tag{D.1.5}$$

We recognize that the kernel of a Gaussian in the integrand and hence it is possible to do the integration analytically. We hence obtain the following approximation of the marginal likelihood.

$$L_M(\theta, y) \approx L\left(\theta, \hat{u}_\theta, y\right) \sqrt{\frac{(2\pi)^k}{\det\left(- \ell''_{uu}(\theta, u, y)\big|_{u=\hat{u}_\theta}\right)}} \tag{D.1.6}$$

We see that the Laplace approximation consists of two nested optimizations. We want to optimize $\theta = \{\beta, \alpha\}$ in the marginal likelihood but in every iteration of the optimization we need first to approximate the marginal likelihood. In every approximation we need to find $\hat{u}_\theta = \arg\max_u L(y; u, \theta)$ and then calculate D.1.6 where the primary work is to find the Hessian $\ell''_{uu}(\theta, u, y)|_{u=\hat{u}_\theta}$. From slide 10, lecture 11 we have the following procedure for the Laplace approximation.

0. Initialize θ to some arbitrary value θ_0
1. With current value for θ optimize the joint likelihood w.r.t. u to get $\hat{u}_\theta$ and corresponding Hessian
2. Use $\hat{u}_\theta$ and $H(\hat{u}_\theta)$ to approximate $\ell_M(\theta)$
3. Compute value and gradient of $\ell_M(\theta)$
4. If the gradient is " $> \epsilon$" update θ and go to 1.

We have implemented the procedure in R and try to fit the model D.1.1 where

$$\boldsymbol{X} = \begin{bmatrix} \mathbb{1}(j =' Female')_N & \mathbb{1}(j =' Male')_N \end{bmatrix} \tag{D.1.7}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_{Female} \\ \beta_{Male} \end{bmatrix} \tag{D.1.8}$$

$$\boldsymbol{Z} = \begin{bmatrix} \mathbb{1}(i = 1)_N & \mathbb{1}(i = 2)_N & \cdots & \mathbb{1}(i = 46)_N & \mathbb{1}(i = 47)_N \end{bmatrix} \tag{D.1.9}$$

$$\boldsymbol{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_{46} \\ u_{47} \end{bmatrix} \tag{D.1.10}$$

$$\boldsymbol{\alpha} = \alpha \tag{D.1.11}$$

We initialize the optimization at $\theta = \{\beta_{Female}, \beta_{Male}, \alpha\} = \{0.1, 0.1, 1\}$ and after 12 seconds of computing we obtain the convergence message `"False Convergence"` which means either that the function we optimize over is discontinuous or the gradient of the function is ill defined. This is weird because everything should be smooth but these problems stems from numerical instabilities. The origin of the problem is the dimensionality of the integral D.1.2 which we try to approximate. We have 47 random effects and hence the integral will be 47-dimensional which leads to a 47 dimensional optimization. This cause instabilities but we can fix it. We can use the independence of the subject IDs to rewrite D.1.2 as

$$L_M(\theta; y) = \prod_{i=1}^{47} \int_{\mathbb{R}^1} L(y; u_i, \theta)\, du_i, \tag{D.1.12}$$

which means the approximation reduces to

$$L_M(\theta, y) \approx \prod_{i=1}^{47} L(\theta, \hat{u}_{i,\theta}, y) \sqrt{\frac{(2\pi)^1}{\det\left(-\ \ell''_{u_i u_i}(\theta, u_i, y)\big|_{u_i=\hat{u}_{i,\theta}}\right)}} \tag{D.1.13}$$

and

$$\hat{u}_{i,\theta} = \arg\max_{u_i} L\left(y; u_i, \theta\right). \tag{D.1.14}$$

This makes the optimization much faster and robust. We again initialize the optimization at $\theta = \{\beta_{Female}, \beta_{Male}, \alpha\} = \{0.1, 0.1, 1\}$ and after 5.85 seconds we obtain the convergence message `"Relative Convergence"` which means the solver has converged. The numerical problems are solved and the obtained parameters are

$$\theta = \{\beta_{Female}, \beta_{Male}, \alpha\} \tag{D.1.15}$$

$$= \{-0.153, -1.233, 3.594\} \tag{D.1.16}$$

D.2 Exercise 2

We calculate the AIC and BIC for the Laplace approximation. The results are shown in table D.1 and we see that actually the Poisson-Gamma model fitted by use of the Laplace approximation is worse than the two other models.

	Binomial	Poisson	Laplace
AIC	263.37	266.10	268.91
BIC	272.10	274.83	277.64

Table D.1 – Results for all models until now in part 2.

E | Part 2, C

E.1

In the following, we will consider the same model structure as in section D.

$$\begin{aligned} Y_{ij} \mid U_i &\sim \text{Pois}\left(\mu_{ij} U_i\right) \\ U_i &\sim G(\alpha, \beta) \\ \mu_{ij} &= e^{x_{ij}^T \theta} \end{aligned} \tag{E.1.1}$$

As in section D we will only use sex as explanatory variable and hence we could write $\mu_{ij} = e^{\mathbb{1}_{female}\theta_{female} + \mathbb{1}_{male}\theta_{male}}$.

E.2

Consider the conditional distribution of Y given U_i, $f_{Y|U_i}(y; \mu_{i,j} U_i)$, and the probability density function of U, $f_U(U; \alpha, \beta)$:

$$f_{Y|U_i}(y_{i,j}; \mu_{i,j} U_i) = \frac{(\mu_{i,j} U_i)^{y_{i,j}}}{y_{i,j}!} \exp\left(-\mu_{i,j} U_i\right) \tag{E.2.1}$$

$$f_U(U; \alpha, \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} U_i^{\alpha-1} \exp(-U_i/\beta) \tag{E.2.2}$$

We note that as $\mathbb{E}\left[U_i\right] = 1$, we can use the alternative parameterization of α and β, remark 6.5 [2], to simplify the expression for the gamma density. We choose to do that in the end to make the derivation more aligned with the steps we did in appendix G.1 with the framework without the $\mu_{i,j}$ where we could use slide 13, lecture 12 for each step. In general we will make the steps in the derivation quite explicit for our own later reference. The marginal distribution of Y can be found as:

$$\begin{aligned} g_Y(y_{i,j}; \mu_{i,j}, \alpha, \beta) &= \int_{U_i=0}^{\infty} f_{Y|U_i}(y_{i,j}; \mu_{i,j} U_i) f_U(U_i; \alpha, \beta) \mathrm{d}U_i \\ &= \int_{U_i=0}^{\infty} \frac{(\mu_{i,j} U_i)^{y_{i,j}}}{y_{i,j}!} \exp\left(-\mu_{i,j} U\right) \frac{1}{\beta^\alpha \Gamma(\alpha)} U_i^{\alpha-1} \exp(-U_i/\beta) \mathrm{d}U_i \\ &= \frac{1}{y_{i,j}! \Gamma(\alpha) \beta^\alpha} \int_{U_i=0}^{\infty} (\mu_{i,j} U_i)^{y_{i,j}} U_i^{\alpha-1} \exp\left\{-U_i \left(\mu_{i,j} + 1/\beta\right)\right\} \mathrm{d}U_i \\ &= \frac{\mu_{i,j}^{y_{i,j}}}{y_{i,j}! \Gamma(\alpha) \beta^\alpha} \int_{U_i=0}^{\infty} U_i^{y_{i,j}+\alpha-1} \exp\left\{-U_i \left(\mu_{i,j} + 1/\beta\right)\right\} \mathrm{d}U_i \end{aligned} \tag{E.2.3}$$

We can reduce this equation using that integrating over the support of a PDF is 1. Recall that the density for a gamma function with shape parameter κ and scale parameter θ is $f(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}}$. We recognize that the integrand looks partly like the density function of a gamma function with shape parameter $y_{i,j}+\alpha$ and scale parameter $1/\left(\mu_{i,j}+1/\beta\right)$. We need only a constant to obtain the density hence we multiply and divide by the constant $\Gamma(y_{i,j}+\alpha)\left(1/\left[\mu_{i,j}+1/\beta\right]\right)^{y_{i,j}+\alpha}$ to obtain

$$
\begin{aligned}
g_Y(y_{i,j};\mu_{i,j},\alpha,\beta) =& \frac{\mu_{i,j}^{y_{i,j}}\Gamma(y_{i,j}+\alpha)\left(\frac{1}{\mu_{i,j}+1/\beta}\right)^{y_{i,j}+\alpha}}{y_{i,j}!\Gamma(\alpha)\beta^\alpha} \\
& \int_{U_i=0}^{\infty} \frac{1}{\Gamma(y_{i,j}+\alpha)\left(\frac{1}{\mu_{i,j}+1/\beta}\right)^{y_{i,j}+\alpha}} U_i^{y_{i,j}+\alpha-1} \exp\left\{-U_i\left(\mu_{i,j}+1/\beta\right)\right\} \mathrm{d}U_i
\end{aligned} \tag{E.2.4}
$$

The integral would now be 1 hence we have integrated out the random effect:

$$
\begin{aligned}
g_Y(y_{i,j};\mu_{i,j},\alpha,\beta) &= \frac{\Gamma(y_{i,j}+\alpha)\mu_{i,j}^{y_{i,j}}}{y_{i,j}!\Gamma(\alpha)\beta^\alpha}\left(\frac{1}{\mu_{i,j}+1/\beta}\right)^{y_{i,j}+\alpha} \\
&= \frac{\Gamma(y_{i,j}+\alpha)\mu_{i,j}^{y_{i,j}}}{y_{i,j}!\Gamma(\alpha)\beta^\alpha}\beta^{y_{i,j}+\alpha}\left(\frac{1}{\mu_{i,j}\beta+1}\right)^{y_{i,j}+\alpha} \\
&= \frac{\Gamma(y_{i,j}+\alpha)\left(\mu_{i,j}\beta\right)^{y_{i,j}}}{y_{i,j}!\Gamma(\alpha)}\left(\frac{1}{\mu_{i,j}\beta+1}\right)^{y_{i,j}}\left(\frac{1}{\mu_{i,j}\beta+1}\right)^{y_{i,j}+\alpha} \\
&= \frac{\Gamma(y_{i,j}+\alpha)}{y_{i,j}!\Gamma(\alpha)}\left(\frac{\mu_{i,j}\beta}{\mu_{i,j}\beta+1}\right)^{y_{i,j}}\left(\frac{1}{\mu_{i,j}\beta+1}\right)^{\alpha}
\end{aligned} \tag{E.2.5}
$$

We will now introduce the reparametrization from remark 6.5 [2]:

$$
\begin{aligned}
\mathbb{E}\left[U_i\right] =& \alpha\beta \\
\gamma =& \beta.
\end{aligned} \tag{E.2.6}
$$

Here γ is the signal to noise ratio. We have $\mathbb{E}\left[U_i\right] = 1$ and we can hence simplify the expression using the reparameterization. To be consistent with the previous section, we will write $\alpha = \alpha$ and $\beta = 1/\alpha$.

$$
g_Y(y_{i,j};\mu_{i,j},\alpha) = \frac{\Gamma(y_{i,j}+\alpha)}{y_{i,j}!\Gamma(\alpha)}\left(\frac{\mu_{i,j}}{\mu_{i,j}+\alpha}\right)^{y_{i,j}}\left(\frac{\alpha}{\mu_{i,j}+\alpha}\right)^{\alpha} \tag{E.2.7}
$$

This is the marginal distribution and it is in a form that makes us able to identify it as a negative binomial which we will show in the next section.

E.3

In the following, we will first consider the density of a negative binomial distribution, $Z \sim \mathrm{NB}\left(r, p\right)$. We will use a parameterization that is slightly different from the normal which is introduce on p. 265 [2]. Consult this page for further specifications and note that the probability mass function is:

$$
g(y) = \binom{y+r-1}{y}\left(1-p\right)^y p^r. \tag{E.3.1}
$$

Using the result on p. 265 [2], we know for $r \in \mathbb{R}_+$ and $y \in \mathbb{Z}$, we can write the binomial coefficient in terms of gamma functions as:

$$\binom{y+r-1}{y} = \frac{\Gamma(r+y)}{\Gamma(r)\, y!}. \tag{E.3.2}$$

We can substitute this result into Equation G.2.1 to obtain the result:

$$g(y) = \frac{\Gamma(r+y)}{\Gamma(r)\, y!} (1-p)^y p^r. \tag{E.3.3}$$

Compare Equation E.3.3 with the marginal distribution found in Equation E.2.7. We observe immediately that $r = \alpha$ and now we postulate $p = \frac{\alpha}{\mu_{i,j}+\alpha}$. To convince our-self, consider

$$1 - p = 1 - \frac{\alpha}{\mu_{i,j}+\alpha} = \frac{\mu_{i,j}+\alpha}{\mu_{i,j}+\alpha} - \frac{\alpha}{\mu_{i,j}+\alpha} = \frac{\mu_{i,j}}{\mu_{i,j}+\alpha}. \tag{E.3.4}$$

Therefore, we see directly that we have a negative binomial for the marginal distribution.

$$g_Y(y_{i,j}; \mu_{i,j}, \alpha) = \mathrm{NB}\left(\alpha, \frac{\alpha}{\mu_{i,j}+\alpha}\right). \tag{E.3.5}$$

E.4

From the density, we know that we can formulate the marginal likelihood as the product of the densities.

$$L(Y_{i,j}|\mu_{i,j}, \alpha) = \prod_{i=0}^{N} g_Y(y_{i,j}; \mu_{i,j}, \alpha) \quad = \prod_{i=0}^{N} \frac{\Gamma(y_{i,j}+\alpha)}{y_{i,j}!\Gamma(\alpha)} \left(\frac{\mu_{i,j}\beta}{\mu_{i,j}\beta+1}\right)^{y_{i,j}} \left(\frac{1}{\mu_{i,j}\beta+1}\right)^{\alpha}, \tag{E.4.1}$$

where N is the number of data points. If we plug in the explicit formulation with only the two cases with males and females, we would have $\mu_{i,j} = e^{\mathbb{1}_{female}\theta_{female} + \mathbb{1}_{male}\theta_{male}}$ we could divide the product into that of males and females. As y is distributed over only a limited set of values $Y \in \{0, 1, 2, 4\}$, we would be able to factor the product into sub-products. In the end, instead of computing a large product of all data points, we could instead count the number of data points that falls into each category like the thunderstorm example, example 6.2 [2]. We will however use the density for each sample

In practice, we will of cause use the log-likelihood instead hence we would have

$$\begin{aligned} \ell(Y_{i,j}|\mu_{i,j}, \alpha) &= \sum_{i=0}^{N} \log g_Y(y_{i,j}; \mu_{i,j}, \alpha) \\ &= \sum_{i=0}^{N} \log\left[\frac{\Gamma(y_{i,j}+\alpha)}{y_{i,j}!\Gamma(\alpha)} \left(\frac{\mu_{i,j}\beta}{\mu_{i,j}\beta+1}\right)^{y_{i,j}} \left(\frac{1}{\mu_{i,j}\beta+1}\right)^{\alpha}\right], \end{aligned} \tag{E.4.2}$$

E.5

To implement this, we create a design matrix and parameter vector.

$$\mathbf{X} = \left[\mathbb{1}(x_{i,j} = male)_N \quad \mathbb{1}(x_{i,j} = female)\right] \quad \theta = \left[\beta_{male}, \beta_{female}\right]^{\intercal} \tag{E.5.1}$$

We will of cause also optimize α. We then make a function to compute the negative log-likelihood and use built-in PORT optimization framework nlminb to optimize the parameters. When we do this, we obtain the following parameters:

$$\alpha^* = 3.966, \qquad \beta^*_{male} = -1.245, \qquad \beta^*_{female} = -0.154 \tag{E.5.2}$$

We see that these parameter estimates are quite close to the results obtained for the Laplace approximation. The information criteria and obtain likelihood are:

	Negative Binomial
AIC	272.0442
BIC	280.7822
log Lik	-133.0221

Table E.1 – Note that we have used the name Negative Binomial for this closed form solution for the hierarchical Poisson-Gamma

We see that the AIC and BIC are higher than the results obtain using the Poisson with Normal distribution as prior.

E.6

In the following, we will find the conditional distribution of the group means which we can do using Bayes rule:

$$g_{U_i}\left(\mu_{i,j}U_i|Y = y_{i,j}\right) = \frac{f_{Y|U_i}(y_{i,j};\mu_{i,j}U_i)f_{U_i}(U_i;\alpha,\beta)}{g_Y(y_{i,j};\mu_{i,j},\alpha)} \tag{E.6.1}$$

Consider first the numerator

$$\begin{aligned} f_{Y|U_i}(y_{i,j};\mu_{i,j}U_i)f_{U_i}(U_i;\alpha,\beta) &= \frac{(\mu_{i,j}U_i)^{y_{i,j}}}{y_{i,j}!}\exp\left(-\mu_{i,j}U\right)\frac{1}{\beta^\alpha\Gamma(\alpha)}U_i^{\alpha-1}\exp(-U_i/\beta) \\ &= \frac{\alpha}{\Gamma\left(\alpha\right)y_{i,j}!}\left(\mu_{i,j}U_i\right)^{y_{i,j}}\left(U_i\alpha\right)^{\alpha-1}\exp\left(-U_i\left(\alpha+\mu_{i,j}\right)\right). \end{aligned} \tag{E.6.2}$$

We found an expression for the denominator in the previous section:

$$g_Y(y_{i,j};\mu_{i,j},\alpha) = \frac{\Gamma(y_{i,j}+\alpha)}{y_{i,j}!\Gamma(\alpha)}\left(\frac{\mu_{i,j}}{\mu_{i,j}+\alpha}\right)^{y_{i,j}}\left(\frac{\alpha}{\mu_{i,j}+\alpha}\right)^{\alpha} \tag{E.6.3}$$

We now divide by this expression an collect terms with the same exponent:

$$
\begin{aligned}
g_{U_i}(\cdot) &= \frac{\alpha}{\Gamma(y_{i,j}+\alpha)} \left(\frac{\mu_{i,j} U_i (\mu_{i,j}+\alpha)}{\mu_{i,j}}\right)^{y_{i,j}} (U_i\alpha)^{\alpha-1} \left(\frac{\mu_{i,j}+\alpha}{\alpha}\right)^{\alpha} \exp\left(-U_i\left[\alpha+\mu_{i,j}\right]\right) \\
&= \frac{1}{\Gamma(y_{i,j}+\alpha)} \left(U_i\left[\mu_{i,j}+\alpha\right]\right)^{y_{i,j}} U_i^{\alpha-1}\alpha^{\alpha} \left(\frac{\mu_{i,j}+\alpha}{\alpha}\right)^{\alpha} \exp\left(-U_i\left[\alpha+\mu_{i,j}\right]\right) \\
&= \frac{1}{\Gamma(y_{i,j}+\alpha)} \left(U_i\left[\mu_{i,j}+\alpha\right]\right)^{y_{i,j}} \left(U_i\left[\mu_{i,j}+\alpha\right]\right)^{\alpha-1} (\alpha+\mu_{i,j}) \exp\left(-U_i\left[\alpha+\mu_{i,j}\right]\right) \\
&= \frac{1}{\Gamma(y_{i,j}+\alpha)} U_i^{y_{i,j}+\alpha-1} (\mu_{i,j}+\alpha)^{y_{i,j}+\alpha} \exp\left(-U_i\left[\alpha+\mu_{i,j}\right]\right) \\
&= \frac{1}{\Gamma(y_{i,j}+\alpha)} U_i^{y_{i,j}+\alpha-1} (\mu_{i,j}+\alpha)^{y_{i,j}+\alpha} \exp\left(-U_i\left[\alpha+\mu_{i,j}\right]\right) \\
&= \frac{1}{\Gamma(y_{i,j}+\alpha)} U_i^{y_{i,j}+\alpha-1} \left(\frac{1}{\mu_{i,j}+\alpha}\right)^{-y_{i,j}-\alpha} \exp\left(-U_i\left[\alpha+\mu_{i,j}\right]\right) \\
&= \frac{1}{\Gamma(y_{i,j}+\alpha)\left(\frac{1}{\mu_{i,j}+\alpha}\right)^{+y_{i,j}+\alpha}} U_i^{y_{i,j}+\alpha-1} \exp\left(-U_i\left[\alpha+\mu_{i,j}\right]\right)
\end{aligned}
\tag{E.6.4}
$$

We now recognize this as a gamma distribution with shape parameter $\kappa = y_{i,j} + \alpha$ and scale parameter $\psi = \frac{1}{\alpha+\mu_{i,j}}$. Hence we have found that:

$$
U_i|Y = y_{i,j} = \mathrm{G}\left(y_{i,j}+\alpha, \frac{1}{\alpha+\mu_{i,j}}\right) \tag{E.6.5}
$$

We know that we can find the first two moments of of a gamma as:

$$
\begin{aligned}
\mathbb{E}\left[U_i|Y = y_{i,j}\right] &= \kappa\psi \\
&= \frac{y_{i,j}+\alpha}{\alpha+\mu_{i,j}} \\
\mathrm{Var}\left[U_i|Y = y_{i,j}\right] &= \kappa\psi^2 \\
&= \frac{y_{i,j}+\alpha}{(\alpha+\mu_{i,j})^2}.
\end{aligned}
\tag{E.6.6}
$$

We can now compare with the found conditional means and variances from the Laplace estimation. The conditional mean and variance in the Laplace approximation is given by

$$
\mathbb{E}\left[U_i|Y\right] = \arg\max_u L(y,u,\theta) \tag{E.6.7}
$$

$$
\mathbb{V}\left[U_i|Y\right] = -\left(\ell''_{uu}(\theta,u,y)\big|_{u=\hat{u}_\theta}\right)^{-1} \tag{E.6.8}
$$

We have plotted the found expectations in figure E.1 and variances in E.2. We see that the analytical mean is a bit higher in the start both for females and males and then it matches quite nicely for females but under shoots for males. The pattern is the same for the variance.

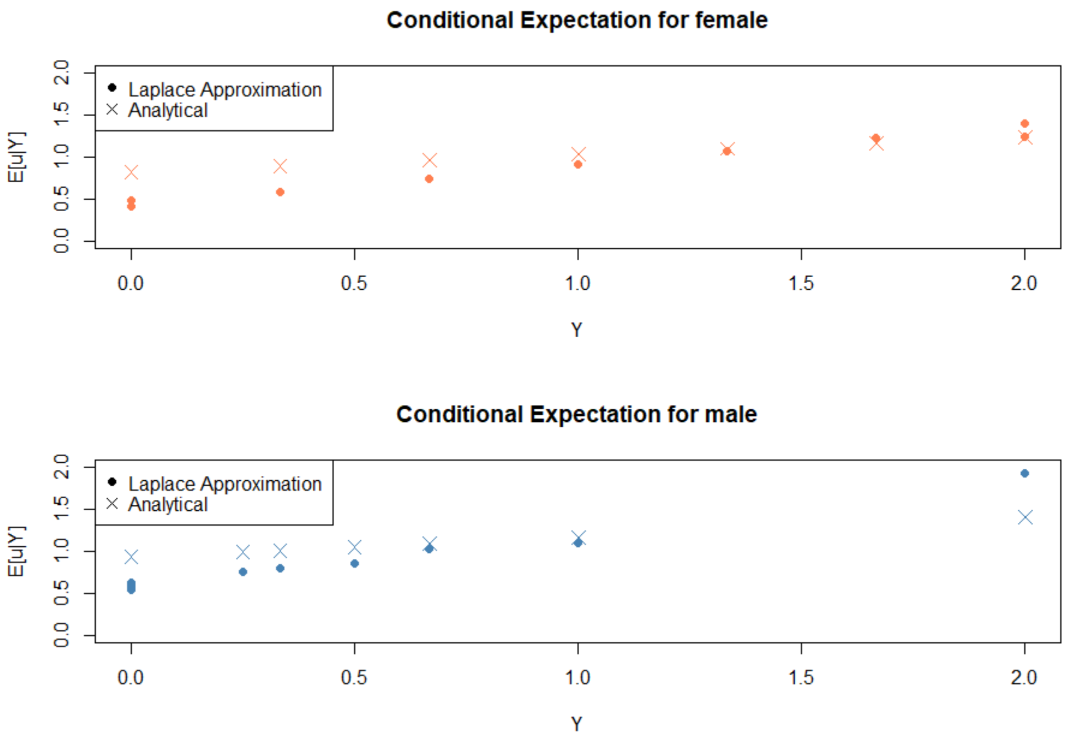


Figure E.1 – Conditional expectation for the random effects.

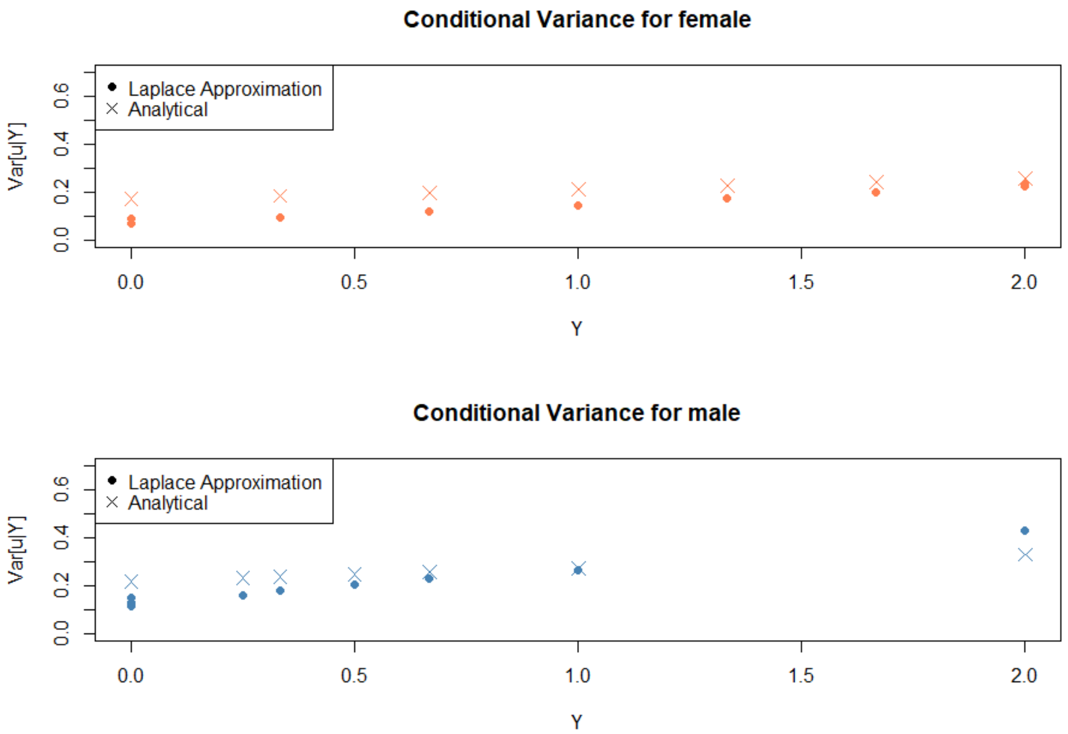


Figure E.2 – Conditional variance for the random effects.

F | Part 2, D

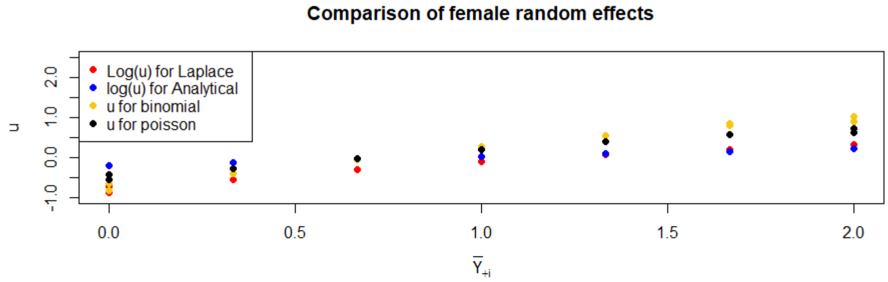
F.1

We have now fitted 4 models. Two with normal random effects where the conditional means were respectively Binomial and Poisson distributed. Next we fitted a Poisson-Gamma model by use of Laplace approximation and lastly due to conjugacy we could also solve the Poisson-Gamma model analytically. This resulted in a negative Binomial distributed marginal likelihood. The AIC and BIC for the four models are shown in table F.2 where we see that the Binomial with normal distributed random effects is actually the best model.

	Binomial	Poisson	Laplace	Negative Binomial
AIC	263.37	266.10	268.91	272.0442
BIC	272.10	274.83	277.64	280.7822

Table F.1 – Results for all models in part 2.

In figure F.1 we have plotted the random effects for all the models. For the two Poisson-Gamma models we have taken the log of the random effects such that all the random effects are added on the fixed effects. We see that the random effects for the binomial model are the largest followed by the Poisson. Hence we can conclude that the normal distributed random effects are a bit larger than the gamma distributed random effects.



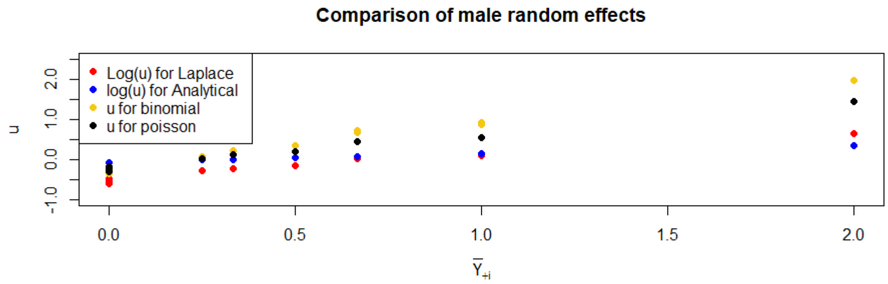


Figure F.1 – Random effects for all models.

In the table F.2 parameters for all models with Poisson distributed fixed effects are shown. We see that the Laplace approximation and the analytical variant are almost equal. Further we see that the Poisson with normal random effects predicts a bit fewer clothing changes compared to the models with Gamma random effects.

	Poisson Normal	Laplace	Negative Binomial
β_{Female}	-0.34	-0.153	-0.154
β_{Male}	-1.50	-1.233	-1.245
Random Effect Parameter	0.68	3.594	3.966

Table F.2 – Results for all models in part 2.

F.2 Conclusion

In the sections above we have two major findings. The first is that there seems to be different parameter estimates when we impose different prior distributions and include the random component in different ways. We see that the parameters are different when we compare the parameters of a hierarchical Poisson model with a normal prior to a hierarchical Poisson model with a gamma prior. The second finding is that when we use a Laplace approximation instead of a closed form solution, the parameters are quite close. There are minor differences but the obtain AIC are almost identical.

G | Appendix

Appendix A

Appendix for Part 2 A

Iteration	Interaction	LRT	p value
1	factor(sex):tOut	0.11562	0.7338
2	tOut	0.01227	0.9118
3	factor(sex):tInOp	2.3683	0.1238
4	tInOp	0.5387	0.462955

Table G.1 – Type II backward selection for the Hierarchical Poisson.

Appendix for Part 2 A

Iteration	Interaction	LRT	p value
1	factor(sex):tOut	0.06799	0.7943
2	tOut	0.02911	0.8645
3	factor(sex):tInOp	2.4186	0.1199
4	tInOp	0.8147	0.366722

Table G.2 – Type II backward selection for the Hierarchical Binomial.

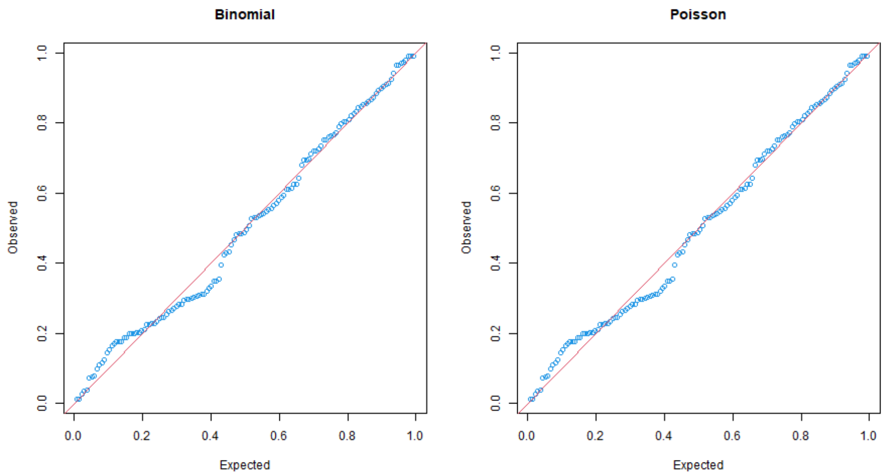


Figure G.1 – QQ-plots for the scaled residuals transformed with uniform quantiles.

G.1

Consider the conditional distribution of Y given μ, $f_{Y|\mu}(y;\mu)$, and the probability density function of μ, $f_\mu(\mu;\alpha,\beta)$:

$$f_{Y|\mu}(y;\mu) = \frac{\mu^y}{y!}\exp(-\mu) \tag{G.1.1}$$

$$f_\mu(\mu;\alpha,\beta) = \frac{1}{\beta\Gamma(\alpha)}\left(\frac{\mu}{\beta}\right)^{\alpha-1}\exp(-\mu/\beta) \tag{G.1.2}$$

The marginal distribution of Y can be found as:

$$\begin{aligned} g_Y(y;\alpha,\beta) &= \int_{\mu=0}^{\infty} f_{Y|\mu}(y;\mu)f_\mu(\mu;\alpha,\beta)\mathrm{d}\mu \\ &= \int_{\mu=0}^{\infty}\frac{\mu^y}{y!}\exp(-\mu)\frac{1}{\beta\Gamma(\alpha)}\left(\frac{\mu}{\beta}\right)^{\alpha-1}\exp(-\mu/\beta)\mathrm{d}\mu \\ &= \frac{1}{\beta^\alpha y!\Gamma(\alpha)}\int_{\mu=0}^{\infty}\mu^{y+\alpha-1}\exp(-\mu(\beta+1)/\beta)\mathrm{d}\mu \end{aligned} \tag{G.1.3}$$

There are multiple ways to do the next step. We observe that we can multiple and divide by $\Gamma(y+\alpha)\left(\frac{\beta}{(\beta+1)}\right)^{y+\alpha}$. This means that now

$$\begin{aligned} g_Y(y;\alpha,\beta) &= \frac{1}{\beta^\alpha y!\Gamma(\alpha)}\int_{\mu=0}^{\infty}\mu^{y+\alpha-1}\exp(-\mu(\beta+1)/\beta)\frac{\Gamma(y+\alpha)\left(\frac{\beta}{(\beta+1)}\right)^{y+\alpha}}{\Gamma(y+\alpha)\left(\frac{\beta}{(\beta+1)}\right)^{y+\alpha}}\mathrm{d}\mu \\ &= \frac{\Gamma(y+\alpha)\left(\frac{\beta}{(\beta+1)}\right)^{y+\alpha}}{\beta^\alpha y!\Gamma(\alpha)}\int_{\mu=0}^{\infty}\frac{1}{\Gamma(y+\alpha)\left(\frac{\beta}{(\beta+1)}\right)^{y+\alpha}}\mu^{y+\alpha-1}\exp(-\mu(\beta+1)/\beta)\mathrm{d}\mu \end{aligned} \tag{G.1.4}$$

Recall that the density for a gamma function with shape parameter κ and scale parameter θ is $f(x) = \frac{1}{\Gamma(k)\theta^k}x^{k-1}e^{-\frac{x}{\theta}}$. We recognize that the integral is the density function of a gamma function with shape parameter $y+\alpha$ and sale parameter $\beta/(\beta+1)$. As we integrate over the entire support of the function, it must integrate to 1 hence:

$$\begin{aligned} g_Y(y;\alpha,\beta) &= \frac{\Gamma(y+\alpha)\left(\frac{\beta}{(\beta+1)}\right)^{y+\alpha}}{\beta^\alpha y!\Gamma(\alpha)} \\ &= \frac{\Gamma(y+\alpha)\beta^y}{y!\Gamma(\alpha)(\beta+1)^{y+\alpha}}. \end{aligned} \tag{G.1.5}$$

G.2

In the following, we will first consider the density of a negative binomial distribution, $Z \sim \mathrm{NB}(r,p)$. We will use a parameterization that is slightly different from the normal which is introduce on p. 265 [2]. Consult this page for further specifications and note that the probability mass function is:

$$g(y) = \binom{y+r-1}{y}(1-p)^y p^r. \tag{G.2.1}$$

Using the result on p. 265 [2], we know for $r \in \mathbb{R}_+$ and $y \in \mathbb{Z}$, we can write the binomial coefficient in terms of beta functions as:

$$\binom{y+r-1}{y} = \frac{\Gamma(r+y)}{\Gamma(r)\,y!}. \tag{G.2.2}$$

We can substitute this result into Equation G.2.1 to obtain the result:

$$g(y) = \frac{\Gamma(r+y)}{\Gamma(r)\,y!}(1-p)^y p^r \tag{G.2.3}$$

This is if we use the usual parameterization of the negative binomial. However, as introduced on p. 265 [2], we can introduce a second parameterization Compare Equation G.2.3 with the marginal distribution found in Equation G.1.5. We observe immediately that $r=\alpha$ and now we postulate $p = 1/(1+\beta)$. To convince our-self, consider $1-p = 1-1/(1+\beta) = \frac{1+\beta}{1+\beta} - \frac{1}{1+\beta} = \frac{\beta}{1+\beta}$. Therefore, we substitute into Equation G.2.3,

$$\begin{aligned} g(y) &= \frac{\Gamma(\alpha+y)}{\Gamma(\alpha)\,y!}\left(\frac{\beta}{1+\beta}\right)^y\left(\frac{1}{1+\beta}\right)^\alpha \\ &= \frac{\Gamma(y+\alpha)\beta^y}{y!\Gamma(\alpha)(\beta+1)^{y+\alpha}} \end{aligned} \tag{G.2.4}$$

hence we see that inded we have a negative binomial with parameters $r=\alpha$ and $p = 1/(1+\beta)$.

Bibliography

[1] J. Pinheiro and D. Bates, *Mixed-effects models in S and S-PLUS*. Springer science & business media, 2006.

[2] H. Madsen and P. Thyregod, *Introduction to general and generalized linear models*. CRC Press, 2010.

[3] M. Avery. Mixed models with citools. [Online]. Available: https://cran.r-project.org/web/packages/ciTools/vignettes/citools-merMod-vignette.html

[4] R. Wicklin. Fisher's transformation of the correlation coefficient. [Online]. Available: https://blogs.sas.com/content/iml/2017/09/20/fishers-transformation-correlation.html

[5] A. Kuznetsova, P. B. Brockhoff, and R. H. Christensen, "lmertest package: tests in linear mixed effects models," *Journal of statistical software*, vol. 82, pp. 1–26, 2017.

[6] F. Hartig, "Dharma: residual diagnostics for hierarchical (multi-level/mixed) regression models," *R package version 0.3*, vol. 3, 2020.