

i Cover Page

Examination paper for TMA4315 - Generalized Linear Models

Course contact: Bob O'Hara

Present at the exam location: NO

Permitted examination support material: C

- Tabeller og formel i statistikk (Tapir forlag, Fagbokforlaget),
- one yellow A4 sheet with your own notes (stamped by the Department of Mathematical Sciences),
- specified calculator

OTHER INFORMATION

Read the questions carefully and make your own assumptions. Specify in your answer which assumptions you have used as a basis for interpreting/defining the assignment.

The academic person is only contacted in case of errors or insufficiencies in the question set. Address an invigilator if you suspect errors or insufficiencies. Write down the question in advance.

SPECIFIC INFORMATION FOR YOUR COURSE

Hand drawings:

For question 9 you are meant to answer on handwritten sheets. Other questions must be answered directly in Inspira. At the bottom of the question, you will find a seven-digit code. Fill in this code in the top left corner of the sheets you wish to submit.

We recommend that you do this during the exam. If you require access to the codes after the examination time ends, click "Show submission".

You are responsible for filling in the correct codes on the handwritten sheets. Therefore, read the cover sheet carefully. The Examination Office cannot guarantee that that incorrectly completed sheets will be added to your assignment.

Weighting: The maximum achievable score for each question is given with the question.

Notifications:

Any messages during the exam (e.g., in case of errors in the exam set) will be sent out via notifications in Inspira. A notification will appear as a dialog box on the screen. You can find the notification again by clicking on the bell icon at the top right

Withdrawing from the exam:

If you wish to submit a blank test/withdraw from the exam for another reason, go to the menu in the top right-hand corner and click "Submit blank". This cannot be undone, even if the test is still open.

Access to your answers:

After the exam, you can find your answers under previous tests in Inspira. Be aware that it may take a working day until any hand-written material is available in the archive.

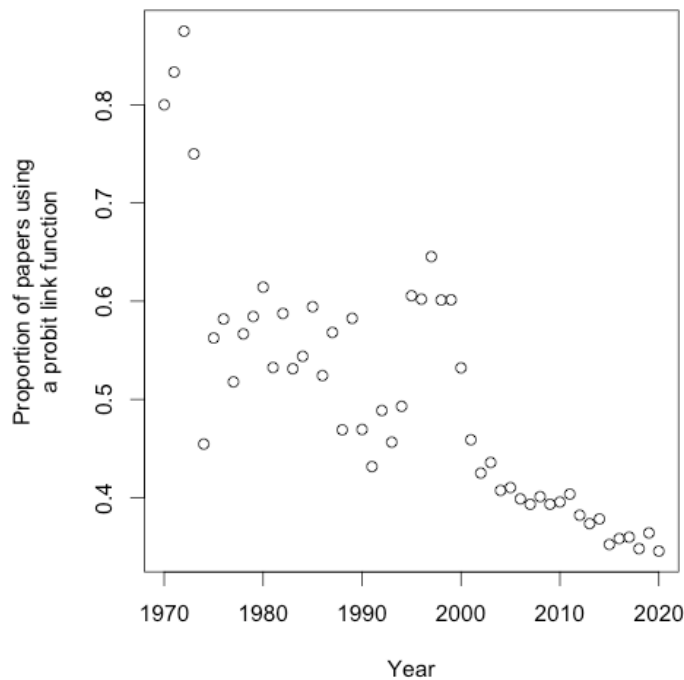
1 Copy of Parts of a GLM

List and explain the 3 parts of a Generalised Linear Model (1-2 sentences for each element)

Fill in your answer here

1. The random component, which is a distribution from the exponential family.
2. The link function, which transforms the systematic component to be the expected value of each variable in the random component (or the response function, which is the inverse of the link function). It should be invertible and twice differentiable.
3. The systematic component, which is a linear predictor, $X^T \beta$

Maximum marks: 6



Three link function are commonly used for the binomial distribution: the logit, probit and cloglog. Here we will compare logits and probits.

We can look at how the use of the logit and probit links has changed over time. The plot is of the proportion of papers mentioning 'probit' out of all papers mentioning 'probit' or 'logit' in each year.

The data can be downloaded from FigShare:
<https://doi.org/10.6084/m9.figshare.28033454.v2>
 Code and some analysis here:
<https://rpubs.com/oharar/cloglogftw>

2 What is a logit?

If η is a linear predictor, what is the response function for the logit link?

$$\frac{e^{\eta}}{1+e^{\eta}}$$

Maximum marks: 1

3 Describe the Data

Describe the general features of the data. Does it look like a model with a linear effect of Year (on the link scale) will fit well?

Fill in your answer here

The proportion of probits is decreasing over time. But there is an odd jump in the mid-1990s. The variance in the proportions also seems to be lower later in time (this is probably because there are more papers more recently).

Note that we expect it to be curved if it is linear on the logit/probit scale.

Maximum marks: 3

Two models were fitted, one with a logit link and another with a probit link. These are the output from the summary() calls.

Note that YearC = Year - 1970, so 1970 is Year 0.

1. logit

Call:

```
glm(formula = cbind(Data$probit, Data$logit) ~ YearC,
     family = binomial("logit"), data = Data)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.621080	0.050908	12.20	<2e-16 ***
YearC	-0.025317	0.001224	-20.69	<2e-16 ***

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 626.91 on 54 degrees of freedom
 Residual deviance: 194.35 on 53 degrees of freedom
 AIC: 509.25

Number of Fisher Scoring iterations: 3

1. probit

Call:

```
glm(formula = cbind(Data$probit, Data$logit) ~ YearC,
     family = binomial("probit"), data=Data)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.3856193	0.0317087	12.16	<2e-16 ***
YearC	-0.0157446	0.0007603	-20.71	<2e-16 ***

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

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 626.91 on 54 degrees of freedom
 Residual deviance: 194.79 on 53 degrees of freedom
 AIC: 509.69

Number of Fisher Scoring iterations: 3

4 Year Confidence Interval

Calculate an approximate confidence interval for the change in the proportion of papers using logits over time, using the model with the logit link function.

(give your answer to 3 decimal places)

Lower: .

Upper: .

Maximum marks: 4

5 What has been going on?

What can you conclude about how logits and probits have been used, and how this has changed over time?

(you can also use this space to show your working for the previous problem)

Fill in your answer here

In 1970 probits were more prevalent ($e^{0.385}/(1+e^{0.385}) = 60\%$ probits), but this has decreased by about 2.5% each year, on average. These estimates are very unlikely to be by chance if there was no change in reality.

Maximum marks: 3

6 Why no p-value?

Why can't we use one of the usual significance tests to get a p-value to test if one model is better than the other?

Fill in your answer here

The usual tests of significance assume that one model is nested within another, so we know the (asymptotic) distribution of the test statistic. Here nothing is nested, so we do not know how to calculate the statistic.
(for what it is worth, a non-parametric bootstrap could be used)

Maximum marks: 2

7 Which is better?

Although we cannot do a significance test using these outputs, we can still compare the model fits. Which of the models explains the data better, and why do you conclude this?

Fill in your answer here

The probit model has a slightly lower residual deviance (by 0.44), or equivalently a higher AIC by the same amount. Thus it explains the data better. But the difference is so small the models are equivalent in practice.

Maximum marks: 3

8 Overdispersion

Is there evidence that the data are overdispersed in either model? Explain how you come to that conclusion.

Fill in your answer here

Yes, there is evidence of overdispersion. For the probit model the residual deviance is 194.79 on 53 degrees of freedom. If we use a chi squared test, this is very significant ($p < 10^{-17}$), so there is very very strong evidence for overdispersion.

Maximum marks: 2

9 Getting the Score (Function)

Note: Answers for this question must be written on handwritten sheets.

The binomial distribution can be written in this form

$$f(r_i | n_i, p_i) = \frac{n_i!}{r_i!(n_i - r_i)!} p_i^{r_i} (1 - p_i)^{n_i - r_i}$$

The probit response function is

$$p_i = \Phi(\eta_i) = \frac{1}{2\pi} \int_{-\infty}^{\eta_i} e^{-\frac{1}{2}z^2} dz$$

so the link function is just written as $\eta_i = \Phi^{-1}(p_i)$. We should write this as a member of the exponential family:

$$f(y|\theta) = \exp\left(\frac{y\theta - b(\theta)}{\phi} w_i + c(y, w_i, \phi)\right)$$

- Show that $\theta = \log \frac{\Phi(\eta_i)}{1 - \Phi(\eta_i)}$
- Derive $E(Y)$ and $\text{Var}(Y)$ as a function of η
- Show that the score can be written as

$$s_i(\beta) = \frac{(y_i - \Phi(x_i' \beta)) x_i f(x_i' \beta)}{\Phi(x_i' \beta)(1 - \Phi(x_i' \beta))}$$

- Write down the expression for the asymptotic distribution of the parameters, and explain how it would be calculated by the computer

Answers on the next 4 pages

Maximum marks: 15

10 Getting the MLEs

Explain what you need to derive the maximum likelihood estimates, and outline how you would calculate them.

Fill in your answer here

We need to be able to calculate the score, i.e. the first derivative of the likelihood with respect to the parameters. We can use iterated weighted least squares: we start with some working responses and working weights, and use weighted least squares to estimate the parameters. From these we recalculate the working responses and weights and do fit a new model. We do this until convergence. (and yes, some equations would be helpful, but as one student pointed out, it would take time, and can be found with a simple search)

Maximum marks: 3

Q9

Getting the Score (Function)

Note: Answers for this question must be written on handwritten sheets.

The binomial distribution can be written in this form

$$f(r_i|n_i, p_i) = \frac{n_i!}{r_i!(n_i - r_i)!} p_i^{r_i} (1 - p_i)^{n_i - r_i}$$

The probit response function is

$$p_i = \Phi(\eta_i) = \frac{1}{2\pi} \int_{-\infty}^{\eta_i} e^{\frac{1}{2}z^2} dz$$

so the link function is just written as $\eta_i = \Phi^{-1}(p_i)$. We should write this as a member of the exponential family:

$$f(y|\theta) = \exp \left(\frac{y\theta - b(\theta)}{\phi} w_i + c(y, w_i, \phi) \right)$$

a. Show that $\theta = \log \frac{\Phi(\eta_i)}{1 - \Phi(\eta_i)}$

We know that for a binomial GLM we have

$$l(y_i|p_i) = y_i \log \frac{\pi_i}{1 - \pi_i} - \log(1 - \pi_i)$$

So $\theta_i = \log \frac{\pi_i}{1 - \pi_i}$. And $\pi_i = \Phi(\eta_i)$ (see above), so

$$\theta_i = \log \frac{\Phi(\eta_i)}{1 - \Phi(\eta_i)}$$

b. Derive $E(Y)$ and $Var(Y)$ as a function of η

(some of this could be short-cutted as “we already know that...”)

We know that $E(Y) = b'(\theta)$, and we have just shown how θ and η relate to each other. $b(\theta) = \log(1 - p) = \log(1 + e^\theta)$. If we let $w = 1 + e^\theta$ we can use the chain rule

$$\begin{aligned} E(Y) = b'(\theta) &= \frac{\partial b(\theta)}{\partial \theta} = \frac{\partial b(\theta)}{\partial w} \frac{\partial w}{\partial \theta} \\ &= \frac{1}{w} e^\theta \\ &= \frac{e^\theta}{1 + e^\theta} \end{aligned}$$

(you don't lose marks for just remembering this result!)

Now plugging in $\theta_i = \log \frac{\Phi(\eta_i)}{1 - \Phi(\eta_i)}$

$$E(Y) = \frac{\frac{\Phi(\eta_i)}{1 - \Phi(\eta_i)}}{1 + \frac{\Phi(\eta_i)}{1 - \Phi(\eta_i)}}$$

The denominator is just $\frac{(1 - \Phi(\eta_i)) + \Phi(\eta_i)}{1 - \Phi(\eta_i)} = \frac{1}{1 - \Phi(\eta_i)}$, so $E(Y) = \Phi(\eta_i)$

Next, $Var(Y) = b''(\theta)\phi/w$. For the binomial we have $\phi = 1$ and $w = 1$. We can remember that $\frac{\partial u/v}{\partial x} = \frac{vu' - uv'}{v^2}$, so

$$\begin{aligned} Var(Y) = b''(\theta) &= \frac{\partial b'(\theta)}{\partial \theta} = \frac{\partial \frac{e^\theta}{1 + e^\theta}}{\partial \theta} \\ &= \frac{(1 + e^\theta)e^\theta - e^\theta e^\theta}{(1 + e^\theta)^2} = \frac{e^\theta}{1 + e^\theta} \frac{1 + e^\theta - e^\theta}{1 + e^\theta} \\ &= \frac{e^\theta}{1 + e^\theta} \frac{1}{1 + e^\theta} = p(1 - p) \end{aligned}$$

Again, plugging in $\theta_i = \log \frac{\Phi(\eta_i)}{1 - \Phi(\eta_i)}$

$$\begin{aligned}
\text{Var}(Y) &= \frac{e^\theta}{1+e^\theta} \frac{1}{1+e^\theta} \\
&= \frac{\frac{\Phi(\eta_i)}{1-\Phi(\eta_i)}}{(1 + \frac{\Phi(\eta_i)}{1-\Phi(\eta_i)})^2} \\
&= \frac{\frac{\Phi(\eta_i)}{1-\Phi(\eta_i)}}{(\frac{1-\Phi(\eta_i)+\Phi(\eta_i)}{1-\Phi(\eta_i)})^2} \\
&= \Phi(\eta_i)(1-\Phi(\eta_i))
\end{aligned}$$

c. Show that the score can be written as

$$s_i(\beta) = \frac{(y_i - \Phi(x'_i\beta))x_i f(x'_i\beta)}{\Phi(x'_i\beta)(1 - \Phi(x'_i\beta))}$$

There are at least 2 approaches. The first uses the general theory for GLMs, the second recreates it:

Plan A: Use the theory

The score for a GLM is

$$s_i(\beta) = \frac{(y_i - \mu_i)x_i h'(\eta_i)}{\text{Var}(Y_i)}$$

So we need μ_i , $h'(\eta_i)$ and $\text{Var}(Y_i)$:

$$\begin{aligned}
\mu_i &= \Phi(x'_i\beta) \\
h'(\eta) &= \partial\Phi(\eta)/\partial\eta = \phi(\eta) = f(x'_i\beta) \\
\text{Var}(Y_i) &= \Phi(x'_i\beta)(1 - \Phi(x'_i\beta))
\end{aligned}$$

Noting that $f(x'_i\beta)$ is the pdf for a normal distribution. Plugging these in gives us

$$s_i(\beta) = \frac{(y_i - \Phi(x'_i\beta))x_i f(x'_i\beta)}{\Phi(x'_i\beta)(1 - \Phi(x'_i\beta))}$$

Plan B: the long way round.

This uses the chain rule, as so often in the course. We can write the score as

$$s_i(\beta) = \frac{\partial l(\beta)}{\partial \beta} = \frac{\partial l(\theta)}{\partial \theta} \frac{\partial \theta}{\partial \eta} \frac{\partial \eta}{\partial \beta}$$

Then, $\frac{l(\theta)}{\partial\theta} = (y_i - b'(\theta))\frac{w_i}{\phi}$ and $\frac{w_i}{\phi} = 1$ and $b'(\theta) = \Phi(\theta)$, so

$$\frac{l(\theta)}{\partial\theta} = (y_i - \Phi(x'_i\beta))$$

The second term is a bit more involved, because $\theta_i = \log \frac{\Phi(\eta_i)}{1-\Phi(\eta_i)}$. We can write $\zeta = \frac{\Phi(\eta_i)}{1-\Phi(\eta_i)}$, and define $\Phi'(x) = f(x)$. Now let the fun begin:

$$\begin{aligned} \frac{\partial\theta}{\partial\eta} &= \frac{\partial\theta}{\partial\zeta} \frac{\partial\zeta}{\partial\eta} = \frac{1}{\zeta} \frac{\partial\zeta}{\partial\eta} \\ &= \frac{1}{\zeta} \frac{(1 - \Phi(\eta_i))f(\eta_i) - \Phi(\eta_i)(-f(\eta_i))}{(1 - \Phi(\eta_i))^2} \\ &= \frac{1 - \Phi(\eta_i)}{\Phi(\eta_i)} \frac{f(\eta_i)}{(1 - \Phi(\eta_i))^2} \\ &= \frac{f(\eta_i)}{\Phi(\eta_i)(1 - \Phi(\eta_i))} \end{aligned}$$

Finally, $\eta = x'_i\beta$, so $\frac{\partial\eta}{\partial\beta} = x_i$. Now we can put these together:

$$\begin{aligned} s_i(\beta) &= \frac{\partial l(\theta)}{\partial\theta} \frac{\partial\theta}{\partial\eta} \frac{\partial\eta}{\partial\beta} \\ &= (y_i - \Phi(x'_i\beta)) \frac{f(\eta_i)}{\Phi(\eta_i)(1 - \Phi(\eta_i))} x_i \\ &= \frac{(y_i - \Phi(x'_i\beta))x_i f(x'_i\beta)}{\Phi(x'_i\beta)(1 - \Phi(x'_i\beta))} \end{aligned}$$

(in case any of you are wondering, ζ is zeta. I used it because we already had beta, theta and eta, so why not?o)

d. Write down the expression for the asymptotic distribution of the parameters, and explain how it would be calculated by the computer

Asymptotically, $\hat{\beta} \sim N(\beta, F^{-1}(\hat{\beta}))$, where $F^{-1}(\hat{\beta})$ is the expected Fisher information. We can estimate $\hat{\beta}$ by iteratively weighted least squares (i.e. Newton-Raphson with the observed Fisher information), and $F^{-1}(\hat{\beta})$ can be calculated either directly or setting it equal to the observed Fisher information. Note that the observed and expected Fisher information are not the same, because the link is not canonical.

We can also use the expected Fisher information in the Newton-Raphson algorithm, which would be Fisher scoring.

The use of logits and probits might be different between different scientific fields, e.g. because a Big Name in the field said that one was better. So we can add Field to the model as a categorical covariate.

After removing the fields that don't use logits or probits enough, there are 26 fields. We add it as a random effect, leading to this model:

```
glmm1 <- glmer(Resp ~ YearC + (1|Field) + (1|OD), family=binomial("logit"),
data=MostData)
```

where

- **Resp** is the response, with logit as a success and probit as failure.
- **YearC** is the year - 1970,
- **Field** is a factor for the field of study
- **OD** is an overdispersion term, i.e. it is a factor with one level for every year/Field combination.

This gave the following summary

```
Generalized linear mixed model fit by maximum likelihood
(Laplace Approximation) ['glmerMod']
```

```
Family: binomial ( logit )
```

```
Formula: Resp ~ YearC + (1 | Field) + (1 | OD)
```

```
Data: MostData
```

AIC	BIC	logLik	deviance	df.resid
3806.6	3827.7	-1899.3	3798.6	1426

```
Scaled residuals:
```

Min	1Q	Median	3Q	Max
-3.2466	-0.4978	0.0000	0.4824	2.7261

```
Random effects:
```

Groups	Name	Variance	Std.Dev.
--------	------	----------	----------

OD	(Intercept)	0.0194	0.1393
----	-------------	--------	--------

Field	(Intercept)	0.5646	0.7514
-------	-------------	--------	--------

```
Number of obs: 1430, groups: OD, 1430; Field, 26
```

```
Fixed effects:
```

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.030925	0.152663	-0.203	0.839
YearC	0.013580	0.001731	7.846	4.29e-15 ***

```
---
```

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

```
Correlation of Fixed Effects:
```

```
(Intr)
```

```
YearC -0.220
```

11 The Equation

Which of these is the equations best represents the linear predictor of the fitted model?

Select one alternative:

- ☒ $\eta_{ij} = \beta_0 + \beta_1 x_{ij} + \gamma_j + \epsilon_{ij}$
- ☐ $\eta_{ij} = \beta_0 + \beta_1 x_{ij} + \gamma_j x_{ij} + \epsilon_{ij}$
- ☐ $\eta_{ij} = \beta x_i + \gamma_j + \epsilon_{ij}$
- ☐ $\eta_{ij} = \beta_0 + \beta_1 x_{ij} + \gamma_j$

Maximum marks: 1

12 Intraclass Correlation

We can get some idea of how much variation there is between fields compared to within fields by looking at the intraclass correlation. If S is the between study variance and R is the residual variance (= overdispersion here), how is the intraclass correlation calculated?

$$S/(S+R) = 1 - R/(S+R)$$

(if you struggle to use the maths tool, use the answer to a later question or the paper to write it out)

Maximum marks: 1

13 Intraclass Correlation Calculation

We can calculate an intra-class correlation (ICC) as we can in a linear mixed model, although the interpretation is slightly different.

What is the ICC for this model?

0.97 .

Maximum marks: 1

14 Comparison of Effects

Based on this, comment on the importance of the Year, Field, and overdispersion in the data. (note that the standard deviation of Year is 15.9)

Fill in your answer here

The overdispersion is now really small, so there is much more variation between field than within fields. The year effect is lower than before, and it would take about 10 years to change the logit of probits (!) by the same as 1 standard deviation in the overdispersion.

Thus most of the variation in the data is variation between fields:
some really like logits, some don't. Pragmatically, Year is still important, in that there is still an effect.
(there are several statistics that could be used to compare the Year, Field and OD)

Maximum marks: 4

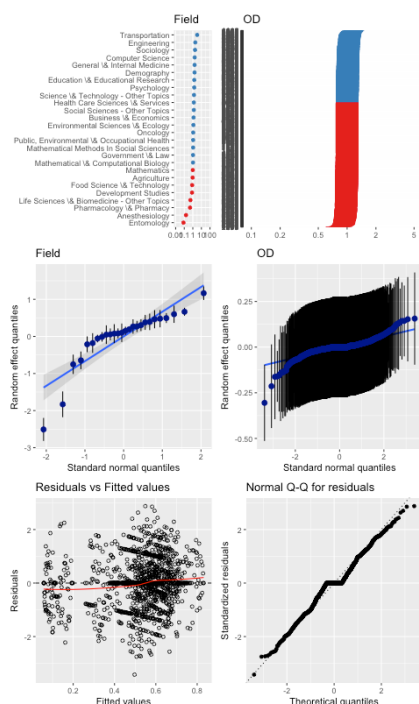
15 How the Model was fitted

The model fitting produced warnings. We will not look at them, but what are the general problems with fitting a model like this, i.e. why is it more difficult than fitting a GLM or a linear mixed model?

Fill in your answer here

With a GLM we can use IWLS to fit the model, and this is quick and robust. LMMs can be fitted with REML, which is still quick and fairly robust. But a GLMM is more difficult. The problem is that we have to integrate over the random effects, and this has to be done numerically, and using approximations to the marginal likelihood.

Maximum marks: 2



Here are some plots we can use to assess the fit of the GLMM, from the logit model.

Top row: random effects for Field (left) and overdispersion (right)

Middle row: normal probability plots for Field (left) and overdispersion (right)

Bottom row: residual plot (left), normal probability plot for residuals (right)

16 Model Fit

Comment on how well the model fits the data. Include comments on what assumptions the seem to be reasonable, and what does not.

Fill in your answer here

The model largely fits OK, but not brilliantly:

1. The overdispersion terms have thicker tails than a normal distribution, and a couple of Fields are slight outliers: entomologists and anesthatolgists don't like logits (note: yes, we did flip what is counted as a "success"!), aside from those the tails still look slightly too thick.
2. The residual plot looks OK: there is a cluster on the left, which is entomology and anaesthatolgy, and there is some sign of heteroscedasticity but this isn't strong.
3. Finally, the QQ plot for the residuals looks OK, except for a weird kink at 0 (which is actually caused by a coding mistake: not removing Year/Field combinations with no data)

Maximum marks: 4