# i Cover Page

**Examination paper for TMA4315 Generalized Linear Models**
**Date: December 1st 2025**
**Time: 15:00 - 19:00**

**Course contact: Bob O'Hara**
**Present at the exam location: No**

**Permitted examination support material: C**
- Tabeller og formler i statistikk (Tapir forlag, Fagbokforlaget),
- one A4 sheet with your own notes,
- specified calculator

**OTHER INFORMATION**

**Read the questions carefully** and make your own assumptions. In your answers, explain clearly what assumptions you have made and how you have understood or limited the assignment

If there are direct errors or omissions in the assignment set and you cannot make your own assumptions, please refer to the information about complaints regarding formal errors on the NTNU website "Explanation of grades and appeals".

**SPECIFIC INFORMATION FOR YOUR COURSE**

**Paper drawings**:
For question 4 you are meant to answer on handwritten sheets. Other questions must be answered directly in Inspera. 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.
**Please note that only blue or black ballpoint pens are allowed.**

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 incorrectly completed sheets will be added to your assignment.

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

**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 Inspera. Be aware that it may take a working day until any hand-written material is available in "previous tests".

A random variable $Y_i$ from a member of the exponential family has the following probability density function:

$$f(y_i) = \exp\left(\frac{y_i\theta_i - b(\theta_i)}{\phi} w_i + c(y_i, \phi, w_i)\right)$$

where

- $\theta_i$ is the natural parameter,
- $\phi$ is the dispersion parameter.
- $w_i$ is a known weight

## 1 E(Y)

What is $E(Y_i)$?
**Select one alternative:**

- $b'(\theta)\frac{\phi}{w_i}$
- $b'(\theta)$
- $b''(\theta)\frac{\phi}{w_i}$
- $c''(y_i, \phi, w_i)$
- $c(y_i, \phi, w_i)$
- $b''(\theta)$
- $b(\theta)$
- $b(\theta)\frac{\phi}{w_i}$
- $c'(y_i, \phi, w_i)$

Maximum marks: 1

## 2 Var(Y)

What is $\text{Var}(y_i)$?
**Select one alternative:**

- X $b''(\theta)\frac{\phi}{w_i}$
- $c'(y_i, \phi, w_i)$
- $b'(\theta)$
- X $b'(\theta)\frac{\phi}{w_i}$
- $c(y_i, \phi, w_i)$
- $b(\theta)$
- $c''(y_i, \phi, w_i)$
- $b(\theta)$
- $b''(\theta)$

---

Maximum marks: 1

## 3 GLM Properties

Match the properties with the parts of a GLM that they are necessary for

**Please match the values:**

| | Random | Link | Systematic |
|---|---|---|---|
| Invertible | | X | |
| Exponential family | X | | |
| Twice differentiable | | X | |
| Linear | | | X |

Maximum marks: 4

## 4 Poisson to Multinomial

Assume we make independent counts, $y_{ic}$,of *C* classes (e.g. $c = 1, \dots, C$ species at $i = 1, \dots, N$ sites). For each site *i* we could condition on $m = \sum_{c=1}^{C} y_c$ and model the vector of observations as coming from a multinomial distribution with *C* classes, so that

$$f(\mathbf{y}_i) = \frac{m_i!}{y_{i1}! \cdots y_{i,C-1}!(m_i - y_{i1} - \cdots - y_{i,C-1})!} \pi_1^{y_{i1}} \pi_2^{y_{i2}} \cdots \pi_{C-1}^{y_{C-1}} (1 - \pi_1 - \pi_2 - \cdots - \pi_{C-1})^{m_i - y_i 1 - y_i 2 - \cdots - y_{i,C-1}}$$

where $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{iC})$ is a vector of counts.

Alternatively, we could assume that each $y_{ic}$ follows a Poisson distribution with mean $\lambda_{ic}$.
(1) **(4 marks)** Show that if we condition on the sum $\sum_{c=1}^{C} y_{ic}$, and assume each $y_{ic}$ follows a Poisson distribution, the vector $\mathbf{y_i}$ still follows a multinomial distribution.
(2) **(2 marks)** if we write our Poisson so that as a GLM, so that $\log(\lambda_{ic}) = \mathbf{x}'_{ic}\beta$, show that when considering this as a multinomial distribution (i.e. focussing on $\pi_{ic}$ rather than $\lambda_{ic}$), some of the parameters are not needed.
(3) **(2 marks)** What is the score for the multinomial distribution?
(4) **(2 marks)** What is the observed Fisher information for the multinomial distribution?

(hints: (1) consider $P(\mathbf{y}_i | \sum_c y_{ic}) = \frac{\prod_c P(y_{ic})}{P(\sum_c y_{ic})}$, (2) $\sum_{c=1}^{C} y_{ic}$ also follows a Poisson distribution).

Maximum marks: 10

# Question 4

## 4: Poisson to Multinomial

(Note: I have given far more than was needed for full marks)

**(1) (4 marks) Show that if we condition on the sum $\sum_{c=1}^{C} y_{ic}$, and assume each $y_{ic}$ follows a Poisson distribution, the vector $\mathbf{y}_i$ still follows a multinomial distribution.**

We want to find $P(\mathbf{y}_i | \sum_c y_{ic})$. The numerator is a product of Poisson distribution, and the hint nicely tells us that the denomninator is also a Poisson distribution. So we have

$$\begin{aligned}
P(\mathbf{y}_i | \sum_c y_{ic}) &= \frac{\prod_c P(y_{ic})}{P(\sum_c y_{ic})} \\
&= \frac{\prod_c \lambda_{ic}^{y_{ic}} e^{-\lambda_{ic}} / y_{ic}!}{(\sum_c \lambda_{ic})^{\sum_c y_{ic}} e^{-\sum_c \lambda_{ic}} / (\sum_c y_{ic})!} \\
&= \frac{(\sum_c y_{ic})!}{\prod_c y_{ic}!} \frac{\prod_c \lambda_{ic}^{y_{ic}} e^{-\lambda_{ic}}}{(\sum_c \lambda_{ic})^{\sum_c y_{ic}} e^{-\sum_c \lambda_{ic}}} \\
&= \frac{(\sum_c y_{ic})!}{\prod_c y_{ic}!} \frac{e^{-\sum_c \lambda_{ic}} \prod_c \lambda_{ic}^{y_{ic}}}{(\sum_c \lambda_{ic})^{\sum_c y_{ic}} e^{-\sum_c \lambda_{ic}}} \\
&= \frac{(\sum_c y_{ic})!}{\prod_c y_{ic}!} \frac{\prod_c \lambda_{ic}^{y_{ic}}}{\prod_c (\sum_c \lambda_{ic})^{y_{ic}}} \\
&= \frac{(\sum_c y_{ic})!}{\prod_c y_{ic}!} \prod_c \left( \frac{\lambda_{ic}}{\sum_c \lambda_{ic}} \right)^{y_{ic}}
\end{aligned}$$

The key things are (1) the exponentials cancel, and (2) you can find $\lambda_{ic} / \sum_c \lambda_i c = p_{ic}$

**(2) (2 marks) if we write our Poisson so that as a GLM, so that $\log(\lambda_{ic}) = \mathbf{x}'_{ic}\beta$, show that when considering this as a multinomial distribution (i.e. focussing $\pi_{ic}$ on rather than $\pi_{ic}$), some of the parameters are not needed.**

From the above, we would use a log link function, i.e. $\log \lambda_{ic} = \mathbf{x}_{ic}^T\beta$. Then

$$p_{ic} = \lambda_{ic} / \sum_c \lambda_i c = \frac{e^{\mathbf{x}_{ic}^T}}{\sum_c e^{\mathbf{x}_{ic}^T\beta}}$$

Now, we can write $\mathbf{x}_{ic}^T\beta = \sum_j x_{icj} = \alpha_i + \kappa_c + \gamma_{ic}$, where $\alpha_i$, $\kappa_c$ and $\gamma_{ic}$ are linear combinations, we can see that

$$p_{ic} = \frac{e^{\alpha_i+\kappa_c+\gamma_{ic}}}{\sum_c e^{\alpha_i+\kappa_c+\gamma_{ic}}} = \frac{e^{\alpha_i}e^{\kappa_c+\gamma_{ic}}}{e^{\alpha_i}\sum_c e^{\kappa_c+\gamma_{ic}}} = \frac{e^{\kappa_c+\gamma_{ic}}}{\sum_c e^{\kappa_c+\gamma_{ic}}}$$

In other words, any terms involving the sample $i$ cancel.

One way of looking at this is that in the multinomial formulation you fix $Y_i = \sum_c y_{ic}$, in the Poisson you estimate $\alpha_i$. Either way you have to "use" $N$ degrees of freedom.

(Note made before grading: I did not expect anyone to see all of this. Noticing that some of the terms cancel is enough in the exam!)

**(3) (2 marks) What is the score for the multinomial distribution?**

From Module 6,

$$\mathbf{s}(\beta) = \sum_{i=1}^{G} \mathbf{X}_i^T \mathbf{D}_i \Sigma_i^{-1}(\mathbf{y}_i - n_i\pi_i)$$

where

- $\mathbf{D}_i = \frac{\partial h(\eta)}{\partial \eta}|_{\eta=\eta_i}$ has dimension $c \times c$ with diagonal.

$$\Sigma_i = \mathrm{Cov}(\mathbf{Y}) = m \begin{pmatrix} \pi_1(1-\pi_1) & -\pi_1\pi_2 & \cdots & -\pi_1\pi_c \\ -\pi_2\pi_1 & \pi_2(1-\pi_2) & \cdots & -\pi_2\pi_c \\ \vdots & \vdots & \ddots & \vdots \\ -\pi_c\pi_1 & -\pi_c\pi_2 & \cdots & \pi_c(1-\pi_c) \end{pmatrix}$$

We can't go much further than this, unless we assume a canonical link when we know the score is (because $\frac{\partial h(\eta)}{\partial \eta} = b''(\theta) = Var(Y)w/\phi$, because $\eta = \theta$):

$$\mathbf{s}(\beta) = \sum_{i=1}^{G} \mathbf{X}_i^T(\mathbf{y}_i - n_i\pi_i)$$

We can also write the score as a score from a Poisson distribution:

$$\mathbf{s}(\beta) = \sum_{i=1}^{N}\sum_{c=1}^{C}(y_{ic} - \lambda_{ic})\mathbf{x}_i = \sum_{i=1}^{N}\sum_{c=1}^{C}(y_{ic} - n_i\pi_{ic})\mathbf{x}_i$$

Note, however, $\beta$ should not include the terms we have noted are not needed.

### (4) (2 marks) What is the observed Fisher information for the multinomial distribution

$$F(\beta) = \sum_{i=1}^{G} \mathbf{X}_i^T \mathbf{W}_i \mathbf{X}_i$$

where $\mathbf{W}_i$ is given as $\mathbf{D}_i\Sigma_i^{-1}\mathbf{D}_i^T$.

Which, again, is about as far as we can get unless we assume a canonical link, when we get

$$F(\beta) = \sum_{i=1}^{G} \mathbf{X}_i^T \mathbf{D}_i \mathbf{X}_i$$

# 5 Fitting Multinomial Distributions

This maths suggests that multinomial distributions could be fitted to data by specifying the random component as a Poisson distribution. Explain briefly (a) how you would do this in R, and (b) how you would interpret the parameters, compared to a multinomial model.

(you do not need to go into details about writing the code, just outline your approach)

**Write your answer in the box below. Changes are saved automatically.**

(1) In R we would format the data so that the response is a column with the response stacked. We also have to stack the matrix of covariates for each site. We then need to add a column for Site (i.e. sample) and Species (i.e. observation within sample), and make both factors. Then we can fit a model glm(Y ~ factor(Site) + Species*…, family=poisson()).
(2) Most of the parameters are interpreted in the same way as the multinomial. Because Species is a factor, it is a contrast to a reference level, just as the multinomial model has a reference level. The only difference is the Site effect, which can be ignored.

---

Maximum marks: 4

We have data from the 2025 Norwegian Eliteserien football league. Rather than analyse the goals scores, we can simply look at the results, as ordinal data (0=away win, 1=draw, 2=home win).

Our design matrix is set up so that each column is a team, and each row is a game. For each row the home team is 1, and away team has -1: the other teams have a zero. Haugesund (whether home or away) is given a value of 0.

When we fit the model we get the following output from summary():

```
Call:
vglm(formula = Result ~ Team, family = cumulative(parallel = T))

Coefficients:
                       Estimate Std. Error z value Pr(>|z|)
(Intercept):1          -0.99878    0.16686  -5.986 2.15e-09 ***
(Intercept):2           0.09637    0.15161   0.636 0.525012
TeamStrømsgodset       -0.75369    0.63112  -1.194 0.232397
TeamFredrikstad        -2.20244    0.62286  -3.536 0.000406 ***
TeamVålerenga          -2.03990    0.62096  -3.285 0.001020 **
TeamHamKam             -1.95134    0.62451  -3.125 0.001780 **
TeamKFUM               -2.01216    0.62142  -3.238 0.001204 **
TeamMolde              -1.88238    0.62096  -3.031 0.002434 **
TeamTromsø             -2.84915    0.64244  -4.435 9.21e-06 ***
TeamBryne              -1.59110    0.63272  -2.515 0.011913 *
TeamBodø/Glimt         -3.60227    0.67915  -5.304 1.13e-07 ***
TeamRosenborg          -2.23594    0.62395  -3.584 0.000339 ***
TeamBrann              -2.87209    0.64278  -4.468 7.89e-06 ***
TeamKristiansund       -1.85347    0.62073  -2.986 0.002827 **
TeamSandefjord Fotball -2.25797    0.62661  -3.603 0.000314 ***
TeamSarpsborg 08       -2.06457    0.62314  -3.313 0.000922 ***
TeamViking             -3.63921    0.68139  -5.341 9.25e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Names of linear predictors: logitlink(P[Y<=1]), logitlink(P[Y<=2])

Residual deviance: 393.6269 on 431 degrees of freedom

Log-likelihood: -196.8134 on 431 degrees of freedom

Number of Fisher scoring iterations: 5

No Hauck-Donner effect found in any of the estimates
```

## 6 Home win

If two teams are equally matched (e.g. if Haugesund play Haugesund), what is the probability of a home win?

$\boxed{0.48}$ =1 - exp(0.09637)/(1+exp(0.09637))

(answer to 2 decimal places. Use . for a decimal)

Maximum marks: 1

(for the numerical problems, we allow for some rounding error)

## 7 Home win confidence interval

What is the 95% confidence interval for this probability?

Lower: 0.40 . =1 - exp(0.09637+1.96*0.16686)/(1+exp(0.09637+1.96*0.16686))

Upper: 0.56 . =1 - exp(0.09637-1.96*0.16686)/(1+exp(0.09637-1.96*0.16686))

(answer to 2 decimal places. Use . for a decimal)

Maximum marks: 2

## 8 Away Win

If two teams are equally matched (e.g. if Haugesund play Haugesund), what is the probability of an away win?

0.27 =exp(-0.99878)/(1+exp(-0.99878))

(answer to 2 decimal places. Use . for a decimal)

Maximum marks: 1

## 9 What no Haugesund?

There is no effect for Haugesund in the output. Why not? (the choice of Haugesund was arbitrary: another team could have been used instead)

**Write your answer in the box below. Changes are saved automatically.**

If we included Haugesund , the design matrix would not be full rank, so we could not estimate all of the parameters. Thus we use Haugesund as a reference level in the factor: all of the estimates are contrasts to htem, i.e. how much better that team is than Haugesund.

Maximum marks: 2

# 10 Model fit

Does the model explain the data well? Give the statistics you use to come to this conclusion,

**Write your answer in the box below. Changes are saved automatically.**

The residual deviance is 363.6 on 431 degrees of freedom. We can test this against a chi-squared distribution. This gives a p-value of 0.99, suggesting the model is (if anything) fitting too well.

---

Maximum marks: 4

In an experiment to look at learning, psychologists asked participants to repeatedly play rock-paper-scissors against an AI that was using a non-optimal strategy. They wanted to know if participants would learn to improve their strategy, and increase their chances of winning.

For our purposes we do not need to know the details of the game. We only use data where the (human) participant or AI won, and look at whether the probability of winning increased over time.

Our first model is a GLM, assuming a binomial distribution with the canonical link and round as a covariate: if the participants are learning to beat the AI, their probability of success should increase as round increases. We get the following summary:

```
> summary(m1)

Call:
glm(formula = win ~ round, family = "binomial", data = rps2)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)    
(Intercept) 0.174938   0.100236   1.745   0.0809 .  
round       0.026906   0.003592   7.491 6.85e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2286.8  on 1872  degrees of freedom
Residual deviance: 2228.8  on 1871  degrees of freedom
AIC: 2232.8

Number of Fisher Scoring iterations: 4
```

## 11 The Canonical GLM question

What link function was used?
**Select one alternative:**

- cloglog
- ✗ logit
- identity
- negative inverse
- log
- probit

Maximum marks: 1

## 12 Hypothesis Testing

**Without using the coefficient and standard error**, test whether there is an effect of round on the probability of success and state your conclusion.

**Write your answer in the box below. Changes are saved automatically.**

We can compare the null and residual deviances: 2286.8 - 2228.8 = 58.0.
We compare this to a chi squared with one degree of freedom (1872-1871), which becomes stupidly significant.

Maximum marks: 4

One question is whether there is variation in the rate at which participants learn. This was examined by extending the analysis to a GLMM. The model was specified in R as

**win ~ round + (round|id)**

So that the model for the $i^{th}$ individual (for all rounds) is $\eta_i = \mathbf{X}_i\beta + \mathbf{U}_i\gamma_i$, where $\mathbf{X}_i$ is the design matrix for the fixed effects for individual *i*, and $\mathbf{U}_i$ is the design matrix for random effects. The vector of random effects, $\gamma_i$, is assumed to follow a multivariate normal distribution with covariance matrix Q.

## 13 GLMM vs GLM

Why is a GLMM better than a GLM for this question?

**Write your answer in the box below. Changes are saved automatically.**

Because with so many participants there are a lot of parameters, so a GLMM will be easier to interpret: we can look at the variation in learning between participants. In addition, a mixed model will help us if we want to predict the performance of a new participant

Maximum marks: 2

## 14 GLM vs GLMM

Why is a GLM better than a GLMM for this question?

**Write your answer in the box below. Changes are saved automatically.**

If we want to test whether there is variation in performance, a GLM is better because the distribution of the test statistic is known (a chi-squared distribution).

Maximum marks: 2

## 15 How big is The Matrix?

What are the dimensions of **X**? (the concatenation of all of the $\mathbf{X}_i$s)

Rows: 1873 .

Columns: 2

Maximum marks: 2

## 16 u Column

How many columns does each $\mathbf{U}_i$ have?

2

Maximum marks: 1

## 17 Q Who

What is the structure of Q?

**Select one alternative:**

X $\begin{pmatrix} \tau_0^2 & \rho\tau_0\tau_1 \\ \rho\tau_0\tau_1 & \tau_1^2 \end{pmatrix}$

- $\begin{pmatrix} \tau_0 & \rho \\ \rho & \tau_1 \end{pmatrix}$

- $\begin{pmatrix} \tau_0 & 0 \\ 0 & \tau_1 \end{pmatrix}$

- $\begin{pmatrix} \tau_0 & \rho\tau_0\tau_1 \\ \rho\tau_0\tau_1 & \tau_1 \end{pmatrix}$

---

Maximum marks: 1

When the GLMM was fitted, the following summary was obtained:

```
> summary(m3)
Generalized linear mixed model fit by maximum likelihood (Laplac
 [glmerMod]
 Family: binomial  ( logit )
Formula: win ~ round + (round | id)
   Data: rps2

      AIC       BIC    logLik -2*log(L)  df.resid
   2025.5    2053.2   -1007.8    2015.5      1868

Scaled residuals:
    Min      1Q  Median      3Q     Max
-6.3859 -0.8990  0.3278  0.6669  1.6593

Random effects:
 Groups Name        Variance Std.Dev. Corr
 id     (Intercept) 0.41418  0.64357
        round       0.00212  0.04604  -0.45
Number of obs: 1873, groups:  id, 52

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) 0.026858   0.141119   0.190    0.849
round       0.041599   0.007887   5.274 1.33e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
      (Intr)
round -0.605
```

## 18 Learning about rounds

Does the conclusion about the overall effect of round change?

**Select one alternative:**

- ✗ No
- ○ Yes
- ○ Not clear to me

Maximum marks: 1

## 19 Learning about Learning Data

How does this new analysis change how you view the data? What new things have you learned?

**Write your answer in the box below. Changes are saved automatically.**

The new analysis tells us about variation in particiapnts: we can view the data as being more than just the overall learning. It shows us that there is variation in initial skill as well as in learning: some participants learn faster than others. There is also a negative correlation between intercept & slope: those who were better at the start tended not to learn as much

Maximum marks: 4

## 20 Failing to learn

Based on this model, what is the probability that a participant would get worse over time?

Explain your reasoning.

**Write your answer in the box below. Changes are saved automatically.**

From the model, the effect of learning is Gaussian with a mean of 0.042, and a variance of 0.0021. Thus we can ask Pr(Y<0 I Y~ N(0.042, 0.0021)), which gives a probability of 0.18.

Maximum marks: 5

## 21 Prediction

Explain the differences between a marginal model to predict success at beating the computer, and a conditional model. Explain when you would use each one in prediction.

**Write your answer in the box below. Changes are saved automatically.**

A conditional model conditions on the levels of the random effect, i.e. on the participant. A marginal model marginalises (=integrates) over the distribution of the random effect. So, a conditional model will predict success for a specified participant, whereas a marginal model predicts the range of values for an unknown participant.

Maximum marks: 4