

# TMA4315: Project 1

Jim Totland, Martin Gudahl Tufte

9/8/2021

## Problem 1

a)

Since the response variables  $y_i \sim \text{Bernoulli}(\pi_i)$ , where  $\pi_i = \Pr(y_i = 1 \mid \mathbf{x}_i)$ . The conditional mean is given by  $Ey_i = \pi_i$ , which is connected to the covariates via the following relationship:

$$\mathbf{x}_i^T \boldsymbol{\beta} =: \eta_i = \Phi^{-1}(\pi_i),$$

or equivalently:  $\pi_i = \Phi(\eta_i)$ . This results in the likelihood function

$$\begin{aligned} L(\boldsymbol{\beta}) &= \prod_{i=1}^n \pi_i^{y_i} (1 - \pi_i)^{1-y_i} \\ &= \prod_{i=1}^n \Phi(\eta_i)^{y_i} (1 - \Phi(\eta_i))^{1-y_i}. \end{aligned}$$

Thus, the log-likelihood becomes

$$l(\boldsymbol{\beta}) := \ln(L(\boldsymbol{\beta})) = \sum_{i=1}^n \underbrace{y_i \ln(\Phi(\eta_i)) + (1 - y_i) \ln(1 - \Phi(\eta_i))}_{=l_i(\boldsymbol{\beta})} = \sum_{i=1}^n l_i(\boldsymbol{\beta}).$$

To find the score function, we calculate

$$\begin{aligned} \frac{\partial l_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= \frac{y_i}{\Phi(\eta_i)} \frac{\partial \Phi(\eta_i)}{\partial \boldsymbol{\beta}} - \frac{1 - y_i}{1 - \Phi(\eta_i)} \frac{\partial \Phi(\eta_i)}{\partial \boldsymbol{\beta}} \\ &= \frac{y_i}{\Phi(\eta_i)} \phi(\eta_i) \mathbf{x}_i - \frac{1 - y_i}{1 - \Phi(\eta_i)} \phi(\eta_i) \mathbf{x}_i \\ &= \frac{y_i(1 - \Phi(\eta_i)) - (1 - y_i)\Phi(\eta_i)}{\Phi(\eta_i)(1 - \Phi(\eta_i))} \phi(\eta_i) \mathbf{x}_i \\ &= \frac{y_i - \Phi(\eta_i)}{\Phi(\eta_i)(1 - \Phi(\eta_i))} \phi(\eta_i) \mathbf{x}_i. \end{aligned}$$

Consequently, the score function is given by

$$\mathbf{s}(\boldsymbol{\beta}) = \sum_{i=1}^n \frac{y_i - \Phi(\eta_i)}{\Phi(\eta_i)(1 - \Phi(\eta_i))} \phi(\eta_i) \mathbf{x}_i.$$

Next, we find the expected Fisher information,  $F(\boldsymbol{\beta})$ . We find it by using the result

$$\begin{aligned}
F(\beta) &= \text{Var}(\mathbf{s}(\beta)) = \text{Var}\left(\sum_{i=1}^n \frac{y_i - \Phi(\eta_i)}{\Phi(\eta_i)(1 - \Phi(\eta_i))} \phi(\eta_i) \mathbf{x}_i\right) \\
&= \sum_{i=1}^n \left[ \frac{\phi(\eta_i)}{\Phi(\eta_i)(1 - \Phi(\eta_i))} \right]^2 \text{Var}(y_i \mathbf{x}_i) = \sum_{i=1}^n \left[ \frac{\phi(\eta_i)}{\Phi(\eta_i)(1 - \Phi(\eta_i))} \right]^2 \mathbf{x}_i \text{Var}(y_i) \mathbf{x}_i^T \\
&= \sum_{i=1}^n \left[ \frac{\phi(\eta_i)}{\Phi(\eta_i)(1 - \Phi(\eta_i))} \right]^2 \pi_i(1 - \pi_i) \mathbf{x}_i \mathbf{x}_i^T = \sum_{i=1}^n \frac{\phi(\eta_i)^2}{\Phi(\eta_i)(1 - \Phi(\eta_i))} \mathbf{x}_i \mathbf{x}_i^T,
\end{aligned}$$

Where in the third equality we have used that the  $y_i$ 's are independent. The expected Fisher information can also be verified to have this expression by the general relationship

$$F(\beta) = \sum_{i=1}^n \frac{h'(\eta_i)^2}{\text{Var}(y_i)} \mathbf{x}_i \mathbf{x}_i^T,$$

where  $h'(\eta_i) = \Phi'(\eta_i) = \phi(\eta_i)$  and  $\text{Var}(y_i) = \pi_i(1 - \pi_i) = \Phi(\eta_i)(1 - \Phi(\eta_i))$ .

b)

The expected Fisher information is given by

$$F(\beta) = \sum_{i=1}^n \frac{\phi(\eta_i)^2}{\Phi(\eta_i)(1 - \Phi(\eta_i))} \mathbf{x}_i \mathbf{x}_i^T = \mathbf{x}^T W \mathbf{x},$$

where  $W = \text{diag}\left(\frac{\phi(\eta_i)^2}{\Phi(\eta_i)(1 - \Phi(\eta_i))}\right)$ .

The Fisher scoring algorithm states that the next iterate is given by

$$\beta^{(t+1)} = \beta^{(t)} + F(\beta^{(t)})^{-1} \mathbf{s}(\beta^{(t)}).$$

Inserting the expected Fisher information and the score function we get **Jeg tror det du beskriver under er Iterated Reweighted Least Squares, som er en variant av Fisher Scoring algo. Tror bi kun trenger formelen over for å implementere myglm.**

$$\beta^{(t+1)} = (\mathbf{x}^T W^{(t)} \mathbf{x})^{-1} \mathbf{x}^T W^{(t)} \tilde{\mathbf{y}}^{(t)},$$

where the working response vector  $\tilde{\mathbf{y}}^{(t)}$  has element  $i$  given by

$$\tilde{y}_i^{(t)} = \mathbf{x}_i^T \beta^{(t)} + \frac{y_i - h(\mathbf{x}_i^T \beta^{(t)})}{h'(\mathbf{x}_i^T \beta^{(t)})} = \eta_i^{(t)} + \frac{y_i - \Phi(\eta_i^{(t)})}{\phi(\eta_i^{(t)})}.$$

**litt om deviance her. Litt ullent for meg hva en saturated model er, men dette gir riktig svar i c).** We also need the deviance, which is defined as

$$D = 2(l_{\text{saturated}} - l(\hat{\beta})).$$

When we fit a parameter for each data point (which is the case for the saturated model), the result for the Bernoulli distribution is that  $\hat{\pi}_i = y_i$ . This means that the likelihood function of the saturated model is given by

$$L_{\text{saturated}} = \prod_{i=1}^n \hat{\pi}_i^{y_i} (1 - \hat{\pi}_i)^{1-y_i} = \prod_{i=1}^n y_i^{y_i} (1 + y_i)^{1-y_i} = 1,$$

Where we have used  $0^0 = 1$ . Consequently, the log-likelihood  $l_{\text{saturated}} = \ln(1) = 0$  and the deviance becomes  $-2l(\hat{\beta})$ . Next follows the Implementation of `myglm` in R:

```
Phi <- function(x) return (pnorm(x))
phi <- function(x) return (dnorm(x))

myglm <- function(formula, data, start = NULL){

  # response variable
  resp <- all.vars(formula)[1]
  y <- as.matrix( data[resp] )

  # model matrix
  X <- model.matrix(formula, data)
  n <- dim(X)[1]
  p <- dim(X)[2]

  # starting beta
  if (is.null(start)){
    beta = rep(0, p)
  }
  else {
    beta = start
  }

  # Fisher scoring algorithm
  max_iter <- 50
  tol <- 1e-10
  iter <- 0
  rel.err <- Inf

  while (rel.err > tol & iter < max_iter){
    # calculate eta, y tilde, W
    eta <- X %*% beta
    y.tilde <- eta + (y - Phi(eta)) / (phi(eta))
    W <- diag( as.vector(phi(eta)^2 / (Phi(eta)*(1-Phi(eta)))), n, n )

    # update beta
    A <- t(X) %*% W %*% X
    b <- t(X) %*% W %*% y.tilde
    beta.new <- solve(A, b)

    iter <- iter + 1
    rel.err <- max(abs(beta.new - beta) / abs(beta.new))
    beta <- beta.new
  }

  # Calculating vcov and deviance.
  F.inv <- solve(A)
```

```

std.Error <- sqrt(diag(F.inv))

eta = X %*% beta
deviance = -2 * sum(y*log(pnorm(eta)) + (1 - y)*log(1 - pnorm(eta)))

return (list("coefficients" = data.frame(beta, std.Error),
          "deviance" = deviance,
          "vcov" = F.inv))
}

```

c)

Simulation of 1000 Bernoulli draws with a random probability.

```

# probability
x = runif(1000, 0, 1)
# draw n bernoulli with prob x
y <- rbinom(1000, 1, x)
df <- data.frame(y, x)
### fit using glm
model <- glm(y ~ x, family = binomial(link = "probit"), data = df)
# beta
model$coefficients

## (Intercept)          x
##   -1.492948    2.992932

# se for beta
summary(model)

##
## Call:
## glm(formula = y ~ x, family = binomial(link = "probit"), data = df)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.2229  -0.7936  -0.3981   0.8152   2.2736
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.49295    0.09524  -15.68  <2e-16 ***
## x            2.99293    0.17543   17.06  <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: 1383.8  on 999  degrees of freedom
## Residual deviance: 1029.2  on 998  degrees of freedom
## AIC: 1033.2
##
## Number of Fisher Scoring iterations: 4

# vcov
vcov(model)

```

```
##           (Intercept)          x
## (Intercept)  0.009070212 -0.01473482
## x           -0.014734823  0.03077485
```

```
# deviance
model$deviance
```

```
## [1] 1029.161
```

```
### fit using myglm
mymodel <- myglm(y ~ x, data = df)
# beta
mymodel$coefficients
```

```
##           beta  std.Error
## (Intercept) -1.492948  0.09523798
## x           2.992932  0.17542858
```

```
# vcov
mymodel$vcov
```

```
##           (Intercept)          x
## (Intercept)  0.009070273 -0.01473496
## x           -0.014734955  0.03077519
```

```
# deviance
mymodel$deviance
```

```
## [1] 1029.161
```

## Problem 2

a)

```
#install.packages("ISwR")
library(ISwR) # Install the package if needed
data(juul)
juul$menarche <- juul$menarche - 1
juul.girl <- subset(juul, age>8 & age<20 & complete.cases(menarche))

mod.probit <- glm(menarche ~ age, family=binomial(link="probit"), data= juul.girl)
anova(mod.probit, test = "Chisq")
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: probit
##
## Response: menarche
##
## Terms added sequentially (first to last)
##
##           Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                518      719.39
## age    1           522      517      197.39 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The low p-value suggests that age has an effect on the response variable.

b)

Relating to the `juul` data set, we define for each observation/individual

$$y_i = \begin{cases} 0, & \text{if menarche has occurred.} \\ 1, & \text{if menarche has not occurred.} \end{cases}$$

and  $t_i$  as the age at the time of examination, which corresponds to `age` in the data set. Let  $T_i \sim \mathcal{N}(\mu, \sigma^2)$ , where  $T_i$  is the time until menarche occurs for the  $i$ 'th individual. Furthermore, let

$$\begin{aligned} \pi_i &:= P(y_i = 1) = P(T_i \leq t_i) \\ &= P\left(\frac{T_i - \mu}{\sigma} \leq \frac{t_i - \mu}{\sigma}\right) = \Phi\left(\frac{t_i - \mu}{\sigma}\right) \end{aligned}$$

This, in turn, gives

$$\Phi^{-1}(\pi_i) = -\frac{\mu}{\sigma} + \frac{1}{\sigma}t_i = \beta_0 + \beta_1 t_i,$$

where  $\beta_0 = -\mu/\sigma$  and  $\beta_1 = 1/\sigma$ .

Let  $\mu = \mu(\beta_0, \beta_1)$  and  $\sigma = \sigma(\beta_0, \beta_1)$ . If  $\hat{\beta}_0$  and  $\hat{\beta}_1$  denotes the MLEs of  $\beta_0$  and  $\beta_1$ , respectively, then  $\hat{\mu} = \mu(\hat{\beta}_0, \hat{\beta}_1)$  and  $\hat{\sigma} = \sigma(\hat{\beta}_0, \hat{\beta}_1)$  is the MLEs of  $\mu$  and  $\sigma$ . It follows that

$$\sigma = \frac{1}{\beta_1}, \quad \mu = -\frac{\beta_0}{\beta_1},$$

so the MLEs are

$$\hat{\sigma} = \frac{1}{\hat{\beta}_1}, \quad \hat{\mu} = -\frac{\hat{\beta}_0}{\hat{\beta}_1}.$$

The corresponding standard errors of these estimators can then be calculated using the delta method. A first order Taylor expansion of  $\mu(\beta_0, \beta_1)$  gives

$$\mu(\beta_0, \beta_1) \approx \hat{\mu} + \frac{\partial \mu}{\partial \beta_0} (\beta_0 - \hat{\beta}_0) + \frac{\partial \mu}{\partial \beta_1} (\beta_1 - \hat{\beta}_1),$$

so

$$\begin{aligned} \text{Var}(\mu) &\approx \left(\frac{\partial \mu}{\partial \beta_0}\right)^2 \text{Var}(\beta_0) + \left(\frac{\partial \mu}{\partial \beta_1}\right)^2 \text{Var}(\beta_1) + 2 \left(\frac{\partial \mu}{\partial \beta_0}\right) \left(\frac{\partial \mu}{\partial \beta_1}\right) \text{Cov}(\beta_0, \beta_1) \\ &= \left(-\frac{1}{\beta_1}\right)^2 \text{Var}(\beta_0) + \left(\frac{\beta_0}{\beta_1^2}\right)^2 \text{Var}(\beta_1) - \frac{2\beta_0}{\beta_1^3} \text{Cov}(\beta_0, \beta_1). \end{aligned}$$

A similar derivation for  $\sigma$  yields

$$\begin{aligned} \text{Var}(\sigma) &\approx \left(\frac{\partial \sigma}{\partial \beta_0}\right)^2 \text{Var}(\beta_0) + \left(\frac{\partial \sigma}{\partial \beta_1}\right)^2 \text{Var}(\beta_1) + 2 \left(\frac{\partial \sigma}{\partial \beta_0}\right) \left(\frac{\partial \sigma}{\partial \beta_1}\right) \text{Cov}(\beta_0, \beta_1) \\ &= \left(-\frac{1}{\beta_1^2}\right)^2 \text{Var}(\beta_1). \end{aligned}$$

Thus, the standard error of the MLE estimators are

$$\widehat{\text{SE}}(\hat{\sigma}) = \frac{1}{\hat{\beta}_1^2} \text{SE}(\hat{\beta}_1), \quad \widehat{\text{SE}}(\hat{\mu}) = \frac{1}{\hat{\beta}_1} \sqrt{\text{Var}(\hat{\beta}_0) + \left(\frac{\hat{\beta}_0}{\hat{\beta}_1}\right) \text{Var}(\hat{\beta}_1)},$$

where it was used (assumed?) that  $\text{Cov}(\beta_0, \beta_1) = 0$ . In R we find the  $\beta$ 's and their standard error using the summary-function.

```
mod.probit <- glm(menarche ~ age, family = binomial(link = 'probit'), data = juul.girl)
s <- summary(mod.probit)
df <- s$coefficients
df
```

```
##              Estimate Std. Error   z value    Pr(>|z|)
## (Intercept) -11.3703291 1.06345963 -10.69183 1.111563e-26
## age          0.8623271 0.08106225  10.63784 1.986860e-26
```

```
beta0 <- df[1,1]
beta1 <- df[1,2]
se0 <- df[2,1]
se1 <- df[2,2]
```

From the readout we get  $\hat{\beta}_0 = -11.3703291$  with a standard error of 0.8623271 and  $\hat{\beta}_1 = 1.0634596$  with a standard error of 0.0810622. Thus  $\hat{\mu} = 10.6918296$  and  $\hat{\sigma} = 0.9403272$ . Their standard errors are  $\widehat{\text{SE}}(\hat{\sigma}) = 0.0716765$  and  $\widehat{\text{SE}}(\hat{\mu}) = 0.7716134$ .

c)

```
mod.logit <- glm(menarche ~ age, family = binomial(link = 'logit'), data = juul.girl)
mod.logit$coefficients[2]
```

```
##      age
## 1.517289
```

To show find the distribution of the  $T_i$ 's, we start with the cumulative distribution:

$$\Pr(T_i \leq t_i) = \Pr(y_i = 1 \mid t_i) = \pi_i = \frac{1}{1 + e^{-\eta_i}}.$$

The pdf of  $T_i$  is then given as

$$\begin{aligned} f_{T_i}(t_i) &= \frac{d}{dt_i} \left( \frac{1}{1 + e^{-\eta_i}} \right) = \frac{\beta_1 e^{-\beta_0 - \beta_1 t_i}}{(1 + e^{-\beta_0 - \beta_1 t_i})^2} \\ &= \frac{e^{-(t_i - (-\beta_0/\beta_1))/(1/\beta_1)}}{1/\beta_1 (1 + e^{-(t_i - (-\beta_0/\beta_1))/(1/\beta_1)})^2} = \frac{e^{-(t_i - \mu)/s}}{s(1 + e^{-(t_i - \mu)/s})^2}. \end{aligned}$$

This is the logistic distribution, with parameters  $\mu = -\beta_0/\beta_1$  and  $s = 1/\beta_1$ , where we have used the parametrization from [Wikipedia](#). We compute estimates of the mean and variance from the estimates of  $\beta_0$  and  $\beta_1$  in the output above. An estimate of the mean is then given by  $E(T_i) = -\beta_0/\beta_1 \approx 13.1901147$  and an estimate of the variance is given by  $\text{Var}(T_i) = s^2 \pi^2/3 = \pi^2/(3\beta_1^2) \approx 1.4290323$ .

d)

We now assume that the latent ages follow a log-normal distribution, i.e.

$$T_i \sim \text{Lognormal}(\mu, \sigma^2).$$

This is equivalent to stating that  $\ln T_i \sim \mathcal{N}(\mu, \sigma^2)$ . Now we can follow the same approach as in 2b):

$$\begin{aligned}
\pi_i &:= \Pr(y_i = 1) = \Pr(T_i \leq t_i) = \Pr(\ln T_i \leq \ln t_i) \\
&= \Pr\left(\frac{\ln T_i - \mu}{\sigma} \leq \frac{\ln t_i - \mu}{\sigma}\right) = \Phi\left(\frac{\ln t_i - \mu}{\sigma}\right)
\end{aligned}$$

This, in turn, gives

$$\Phi^{-1}(\pi_i) = -\frac{\mu}{\sigma} + \frac{1}{\sigma} \ln t_i = \beta_0 + \beta_1 \ln t_i,$$

where  $\beta_0 = -\mu/\sigma$  and  $\beta_1 = 1/\sigma$ . Consequently, we fit GLM with a probit link-function and...