

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

$$\mathbf{s}(\boldsymbol{\beta}) = \sum_{i=1}^n \frac{y_i - \Phi(\eta_i)}{\Phi(\eta_i)(1 - \Phi(\eta_i))} \phi(\eta_i) \mathbf{x}_i = \mathbf{X}^T D \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu}),$$

where $D = \text{diag}(\phi(\eta_i))$ and $\Sigma = \text{diag}(\text{Var}(y_i)) = \text{diag}(\Phi(\eta_i)(1 - \Phi(\eta_i)))$. Next, we find the expected Fisher information, $F(\beta)$. 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

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

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

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

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

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

F.inv = NULL
eta = NULL

while (rel.err > tol & iter < max_iter){
  # Calculate eta.
  eta <- X %*% beta

  # Calculate score.
  D <- diag(as.vector(phi(eta)), n, n)
  Sigma <- diag(as.vector(Phi(eta)*(1 - Phi(eta))), n, n)
  mu.vec <- as.vector(Phi(eta))
  score = t(X) %*% D %*% solve(Sigma) %*% (y - mu.vec)

  # Calculate Fisher information and its inverse.
  W <- diag(as.vector(phi(eta)^2 / (Phi(eta)*(1-Phi(eta)))), n, n)
  F <- t(X) %*% W %*% X
  F.inv <- solve(F)

  # Update beta.
  beta.new <- beta + F.inv %*% score

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

# Calculating std.errors and deviance.

```

```

std.Error <- sqrt(diag(F.inv))
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
##   -1.388441    2.723233

# se for beta
summary(model)

##
## Call:
## glm(formula = y ~ x, family = binomial(link = "probit"), data = df)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.0794  -0.8456  -0.4285   0.8347   2.1520
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.38844    0.09533  -14.56  <2e-16 ***
## x           2.72323    0.16842   16.17  <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: 1385.9  on 999  degrees of freedom
## Residual deviance: 1080.0  on 998  degrees of freedom
## AIC: 1084
##
## Number of Fisher Scoring iterations: 4

# vcov
vcov(model)

##              (Intercept)          x
## (Intercept)  0.009087329 -0.01424267

```

```
## x -0.014242665 0.02836526
```

```
# deviance
```

```
model$deviance
```

```
## [1] 1080.042
```

```
### fit using myglm
```

```
mymodel <- myglm(y ~ x, data = df)
```

```
# beta
```

```
mymodel$coefficients
```

```
## y std.Error
```

```
## (Intercept) -1.388441 0.09532785
```

```
## x 2.723234 0.16842073
```

```
# vcov
```

```
mymodel$vcov
```

```
## (Intercept) x
```

```
## (Intercept) 0.009087398 -0.01424280
```

```
## x -0.014242796 0.02836554
```

```
# deviance
```

```
mymodel$deviance
```

```
## [1] 1080.042
```

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 $\beta_1 = 1/\sigma$. Or equivalently:

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

Due to the [functional invariance of the maximum likelihood estimator](#), we can write the MLEs of σ and μ as

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

where $\hat{\beta}_0$ and $\hat{\beta}_1$ denote the MLEs of β_0 and β_1 , respectively. Thus, the maximum likelihood estimates for this data set can be computed as

```
mod.probit <- glm(menarche ~ age, family = binomial(link = 'probit'), data = juul.girl)
b <- mod.probit$coefficients
mle.mu <- -b[1]/b[2]
mle.sigma <- 1/b[2]
mle.mu
```

```
## (Intercept)
##      13.18563
```

```
mle.sigma
```

```
##      age
##  1.159653
```

The standard errors of $\hat{\mu}$ and $\hat{\sigma}$ can then be estimated using the delta method. A first order Taylor expansion of $\hat{\mu}$ gives

$$\hat{\mu} \approx \hat{\mu}(\mathbf{b}) + \nabla \hat{\mu}(\mathbf{b})^T (\hat{\beta} - \mathbf{b}),$$

where we have used the notation $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)^T$ and expanded around $\hat{\beta} = \mathbf{b} = (b_0, b_1)^T$. Next, we take the variance of this linear approximation, such that

$$\begin{aligned} \text{Var}(\hat{\mu}) &\approx \text{Var}\left(\nabla \hat{\mu}(\mathbf{b})^T \hat{\beta}\right) \\ &= \nabla \hat{\mu}(\mathbf{b})^T \text{Var}(\hat{\beta}) \nabla \hat{\mu}(\mathbf{b}). \end{aligned}$$

Using $\nabla \hat{\mu} = (-1/\hat{\beta}_1, \hat{\beta}_0/\hat{\beta}_1^2)^T$, we can calculate the standard error as in the code below:

```
mod.probit <- glm(menarche ~ age, family = binomial(link = 'probit'), data = juul.girl)
b <- mod.probit$coefficients
grad.mu <- c(-1/b[2], b[1]/b[2]^2)
se.mu <- t(grad.mu) %*% vcov(mod.probit) %*% grad.mu
se.mu
```

```
##           [,1]
## [1,] 0.01404909
```

That is, $\widehat{SE}(\hat{\mu}) = 0.0140491$. We follow the same procedure to estimate the standard error of $\hat{\sigma}$. First, we approximate it by a first order Taylor series expansion:

$$\hat{\sigma} \approx \hat{\sigma}(\mathbf{b}) + \nabla \hat{\sigma}(\mathbf{b})^T (\hat{\beta} - \mathbf{b}),$$

which implies that the variance can be approximated as

$$\begin{aligned} \text{Var}(\hat{\sigma}) &\approx \text{Var}(\nabla \hat{\sigma}(\mathbf{b})^T \hat{\beta}) \\ &= \nabla \hat{\sigma}(\mathbf{b})^T \text{Var}(\hat{\beta}) \nabla \hat{\sigma}(\mathbf{b}). \end{aligned}$$

Using that $\nabla \hat{\sigma} = (0, -1/\beta_1^2)^T$, we calculate the standard error as in the code below.

```
grad.sigma <- c(0, -1/b[2]^2)
se.sigma <- t(grad.sigma) %*% vcov(mod.probit) %*% grad.sigma
se.sigma
```

```
##           [,1]
## [1,] 0.01188363
```

That is, $\widehat{SE}(\hat{\sigma}) = 0.0118836$.

Q: are we estimating the standard error (SE) or is the SE an estimate in itself? i.e. we don't need the hat?

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 β_0 and β_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 on $\ln t_i$:

```
mod.lognorm <- glm(menarche ~ log(age), family = binomial(link = "probit"), data = juul.girl)
mu.hat <- -mod.lognorm$coefficients[1]/mod.lognorm$coefficients[2]
sigma.hat <- 1/mod.lognorm$coefficients[2]
summary(mod.lognorm)
```

```
##
## Call:
## glm(formula = menarche ~ log(age), family = binomial(link = "probit"),
##      data = juul.girl)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.28617  -0.11683   0.00315   0.10715   2.54738
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -29.060      2.704  -10.75  <2e-16 ***
## log(age)       11.287      1.049   10.76  <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: 719.39  on 518  degrees of freedom
## Residual deviance: 198.05  on 517  degrees of freedom
## AIC: 202.05
##
## Number of Fisher Scoring iterations: 8
```


Exactly as in 2b), due to the functional invariance of MLEs, we can estimate the mean of T_i as

$$\exp\left(\hat{\mu} + \frac{\hat{\sigma}^2}{2}\right) = 13.1797035,$$

and we can estimate the standard deviation as

$$\sqrt{[\exp(\hat{\sigma}^2) - 1] \exp(2\hat{\mu} + \hat{\sigma}^2)} = 0.201604.$$

The formulas for mean and standard deviation of the log-normal distribution are gathered from [Wikipedia](#).