STK3100 Exercises, Week 8

Vinnie Ko, Jonas Moss

October 18, 2018

Exercise 4.20

The pmf of Poisson distribution is

$$f(y) = \frac{\mu^y}{y!} e^{-\mu}$$

So,

$$\begin{aligned}
L(\boldsymbol{y}) &= \log \mu \sum_{i=1}^{n} y_i - n\mu - \sum_{i}^{n} \log y_i! \\
\frac{\partial L(\boldsymbol{y})}{\partial \mu} &= \frac{1}{\mu} \sum_{i=1}^{n} y_i - n = -n + \frac{n\overline{y}}{\mu} \\
H &= \frac{\partial^2 L(\boldsymbol{y})}{\partial \mu^2} = -\frac{n\overline{y}}{\mu^2} \\
\mathcal{J} &= \mathrm{E}\left[-\frac{\partial^2 L(\boldsymbol{Y})}{\partial \mu^2}\right] = \frac{n}{\mu^2} \mathrm{E}\left[\overline{Y}\right] = \frac{n}{\mu}.
\end{aligned}$$

Fisher scoring gives

$$\mu^{(t+1)} = \mu^{(t)} + \left(\mathcal{J}^{(t)}\right)^{-1} u^{(t)} = \mu^{(t)} + \frac{\mu^{(t)}}{n} \left(-n + \frac{n\overline{y}}{\mu^{(t)}}\right) = \overline{y}.$$

Newton-Raphson gives

$$\mu^{(t+1)} = \mu^{(t)} + \left(H^{(t)}\right)^{-1} u^{(t)} = \mu^{(t)} - \frac{\left(\mu^{(t)}\right)^2}{n\overline{y}} \left(-n + \frac{n\overline{y}}{\mu^{(t)}}\right) = \mu^{(t)} \left(2 - \frac{\mu^{(t)}}{\overline{y}}\right).$$

Note that if $\mu^{(t)} = \overline{y}$, then $\mu^{(t+1)} = \overline{y}$.

Exercise 4.26

i: Find the AIC formula)

The log likelihood of the normal model is

$$\begin{aligned}
\log \prod_{i=1}^{n} \phi\left(x_i; \mu, \sigma^2\right) &= -\sum_{i=1}^{n} \left[\frac{1}{2} \log 2\pi + \log \sigma + \frac{1}{2\sigma^2} \left(x_i - \mu\right)^2\right] \\
&= -\left[n \frac{1}{2} \log 2\pi + n \log \sigma + \frac{n}{2\sigma^2} \frac{1}{n} \sum_{i=1}^{n} \left(x_i - \mu\right)^2\right]. \qquad (1)
\end{aligned}$$

Plug in the maximum likelihood estimates $\widehat{\mu_{ML}}$ and $\widehat{\sigma^2_{ML}} = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \widehat{\mu_{ML}}\right)$. Then the normal log likelihood (1) simplifies to

$$-\left[n\frac{1}{2}\log 2\pi + n\log\widehat{\sigma_{ML}} + \frac{1}{2}n\right].$$

Since AIC is defined by $-2\text{loglik} + 2\left(p+1\right)$ when we have $p+1$ parameters, we get

$$\begin{aligned}\text{AIC} &= n\log 2\pi + n\log\widehat{\sigma^2_{ML}} + n + 2p \\ &= n\left[\log 2\pi\widehat{\sigma^2_{ML}} + 1\right] + 2\left(p+1\right).\end{aligned} \tag{2}$$

ii: How to Compare AICs

We will simply $\text{AIC}_2 < \text{AIC}_1$. To that end, denote the estimated standard deviation in either model by $\widehat{\sigma_i}$. Using the identity (2), $\text{AIC}_2 < \text{AIC}_1$ is equivalent to

$$n\left[\log 2\pi\widehat{\sigma^2_2} + 1\right] + 2\left(p+q+1\right) < n\left[\log 2\pi\widehat{\sigma^2_1} + 1\right] + 2\left(p+1\right).$$

This can be simplified to $\log\left[\widehat{\sigma^2_2}/\widehat{\sigma^2_1}\right] < -2q/n$ by shuffling of terms and contraction of logarithms. Take the exponential on both sides to get $\widehat{\sigma^2_2}/\widehat{\sigma^2_1} < \exp\left(-2q/n\right)$. Now use the factthat $\widehat{\sigma^2_2}/\widehat{\sigma^2_1} = \text{SSE}_2/\text{SSE}_1$, to obtain the desired $\text{SSE}_2/\text{SSE}_1 < \exp\left(-2q/n\right)$.

Exercise 4.27

a: A Solution Sketch

Let h be a differentiable function. Now we use a first order Taylor approximation on $h\left(Y\right)$ around its mean μ:

$$h\left(Y\right) = h\left(\mu\right) + h'\left(\mu\right)\left(Y-\mu\right) + o\left(Y-\mu\right)$$

We can use this to obtain

$$\begin{aligned}Eh\left(Y\right) &= h\left(\mu\right) + h'\left(\mu\right)E\left(Y-\mu\right) + E\left(o\left[Y-\mu\right]\right) \\ &\approx h\left(\mu\right)\end{aligned}$$

$$\begin{aligned}\text{Var}h\left(Y\right) &= \text{Var}\left[h\left(\mu\right) + h'\left(\mu\right)\left(Y-\mu\right) + o\left(Y-\mu\right)\right] \\ &= \text{Var}\left[h'\left(\mu\right)Y + o\left(Y-\mu\right)\right] \\ &= \text{Var}\left[Y\right]h'\left(\mu\right) + h'\left(\mu\right)\text{Cov}\left\{Y, o\left(Y-\mu\right)\right\} + \text{Var}\left[o\left(Y-\mu\right)\right] \\ &\approx \sigma^2 h'\left(\mu\right)^2\end{aligned}$$

We want to *stabilize the variance*. We do this by identifying the function h such that $\sigma^2 h'\left(\mu\right)^2$ is constant. When the standard deviation is proportional to the mean, $\sigma^2 = a^2\mu^2$. Hence we need to solve $\mu^2 h'\left(\mu\right)^2 = 1$. which is equivalent to solving $h'\left(\mu\right)^2 = \mu^{-2}$. Since μ is positive, this is equivalent to solving $h'\left(\mu\right) = \mu^{-1}$, which has the solution $h\left(\mu\right) = \log\left(\mu\right)$. Provided the error terms go to zero as $\mu \to 0$, which appear reasonable since the standard deviation is proportional to the mean, we are done.

b: The Logarithm of the Expectation

When $\log Y \sim N\left(\mu, \sigma^2\right)$, the expectation of Y is $E \exp^Z$ where $Z \sim N\left(\mu, \sigma^2\right)$. This is the moment generating function of a normal variate evaluated at $t = 1$. The moment generating function is

$$\int \exp\left(xt\right) \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}\left(x^2 - 2x\mu + \mu^2\right)\right] dx,$$

which equals

$$\begin{aligned} \int \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}\left(x^2 - 2x\left(\mu + t\sigma^2\right) + \mu^2\right)\right] dx &= \\ \int \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}\left(x^2 - 2x\left(\mu + t\sigma^2\right) + \left(\mu + t\sigma^2\right)^2\right)\right] dx \exp\left(\frac{2\mu t\sigma^2 + t\sigma^4}{2\sigma^2}\right) &= \\ \exp\left(\mu t + \frac{1}{2}t\sigma^2\right) \end{aligned}$$

Now let $t = 1$ and $E \exp^Z = \exp\left(\mu + \frac{1}{2}\sigma^2\right)$. The logarithm of this is $\mu + \frac{1}{2}\sigma^2$ and since $E \log Z = \mu$ we are finished.

c: The Median of the Log-normal distribution

The variates y_i are log-normally distributed with some parameters μ_i, σ. The median of the log-normal distribution is equal to e^{μ_i}, because quantiles are carried over by monotone transformations. (Recall that the median of the normal is μ_i.) This means that the desired parameters μ_i are a simple transformation of the median. What's more, the mean is $\exp\left(\mu_i + \frac{1}{2}\sigma^2\right)$, which is unnaturally pertubed by the standard deviation. Moreover, as the log-normal is fat-tailed, the median is more natural as a measure of centrality.

Exercise 17

a)

We are given that

$$Y_i = \beta_1 x_{i,1} + \cdots + \beta_p x_{i,p} + \varepsilon_i$$

where $\varepsilon_i \sim N\left(0, \frac{\sigma^2}{w_i}\right)$.
By multiplying with $\sqrt{w_i}$ on both sides we get

$$\begin{aligned} \sqrt{w_i} Y_i &= \beta_1 \sqrt{w_i} x_{i,1} + \cdots + \beta_p \sqrt{w_i} x_{i,p} + \sqrt{w_i}\varepsilon_i \\ Y_i^* &= \beta_1 x_{i,1}^* + \cdots + \beta_p x_{i,p}^* + \varepsilon_i^* \end{aligned}$$

where $\varepsilon_i^* = \sqrt{w_i}\varepsilon_i \sim \sqrt{w_i} N\left(0, \frac{\sigma^2}{w_i}\right) = N\left(0, \sigma^2\right)$

b)

By using normal equation, we obtain

$$\widehat{\boldsymbol{\beta}} = \left(\left(\boldsymbol{X}^*\right)^{\mathrm{T}} \boldsymbol{X}^*\right)^{-1} \left(\boldsymbol{X}^*\right)^{\mathrm{T}} \boldsymbol{Y}^*.$$

Let $\boldsymbol{W}^{\frac{1}{2}} = \operatorname{diag}\{\sqrt{w_i}\}$, then

$$\begin{aligned}
\widehat{\boldsymbol{\beta}} &= \left((\boldsymbol{X}^*)^{\mathrm{T}} \boldsymbol{X}^*\right)^{-1} (\boldsymbol{X}^*)^{\mathrm{T}} \boldsymbol{Y}^* \\
&= \left(\left(\boldsymbol{W}^{\frac{1}{2}}\boldsymbol{X}\right)^{\mathrm{T}} \boldsymbol{W}^{\frac{1}{2}}\boldsymbol{X}\right)^{-1} \left(\boldsymbol{W}^{\frac{1}{2}}\boldsymbol{X}\right)^{\mathrm{T}} \boldsymbol{W}^{\frac{1}{2}}\boldsymbol{Y} \\
&= \left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}^{\frac{1}{2}}\boldsymbol{W}^{\frac{1}{2}}\boldsymbol{X}\right)^{-1} \boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}^{\frac{1}{2}}\boldsymbol{W}^{\frac{1}{2}}\boldsymbol{Y} \\
&= \left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{X}\right)^{-1} \boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{Y}.
\end{aligned}$$

Exercise 18

a)

$$\begin{aligned}
f(y) &= \frac{1}{\sigma\sqrt{2\pi y^3}} \exp\left[-\frac{1}{2y}\left(\frac{y-\mu}{\mu\sigma}\right)^2\right] \\
&= \exp\left[-\frac{1}{2}\left(\frac{y}{\mu^2\sigma^2} - \frac{2}{\mu\sigma^2} + \frac{1}{y\sigma^2}\right) - \log\left(\sigma\sqrt{2\pi y^3}\right)\right] \\
&= \exp\left[-\frac{1}{2}\left(\frac{y}{\mu^2\sigma^2} - \frac{2}{\mu\sigma^2}\right) - \frac{1}{2y\sigma^2} - \log\left(\sigma\sqrt{2\pi y^3}\right)\right] \\
&= \exp\left[\frac{-\frac{y}{2\mu^2} + \frac{1}{\mu}}{\sigma^2} - \frac{1}{2y\sigma^2} - \log\left(\sigma\sqrt{2\pi y^3}\right)\right] \\
&= \exp\left[\frac{\theta y - b(\theta)}{a(\phi)} + c(y,\phi)\right]
\end{aligned}$$

where $\theta = -\frac{1}{2\mu^2}$, $b(\theta) = -\sqrt{-2\theta} = -\frac{1}{\mu}$, $a(\phi) = \phi = \sigma^2$, $c(y,\phi) = -\frac{1}{2y\phi} - \log\left(\sqrt{2\pi y^3\phi}\right) = -\frac{1}{2y\sigma^2} - \log\left(\sigma\sqrt{2\pi y^3}\right)$.
So, $f(y)$ is within the exponential dispersion family.

b)

$$\begin{aligned}
\mathrm{E}[Y] &= b'(\theta) = \frac{1}{\sqrt{-2\theta}} = \mu \\
\mathrm{Var}(Y) &= b''(\theta)a(\phi) = -\frac{-2}{2}(-2\theta)^{-\frac{3}{2}}\phi = (-2\theta)^{-\frac{3}{2}}\phi = \mu^3\sigma^2
\end{aligned}$$

c)

Canonical link function $g(\cdot)$ is a link function such that $\theta = g\left(\mathrm{E}[Y]\right)$. (p.123 of the book).

$$\theta = -\frac{1}{2\mu^2} = g(\mu)$$

Thus, the canonical link function for inverse Gaussian distribution is $g(\mu) = -\frac{1}{2\mu^2}$.

d)

The log-likelihood function is

$$L(\boldsymbol{y}, \widehat{\boldsymbol{\mu}}, \boldsymbol{\sigma}^2) = -\frac{1}{2}\sum_{i=1}^{n} \log\left(2\pi\sigma^2 y_i^3\right) - \frac{1}{2}\sum_{i=1}^{n} \frac{1}{y_i}\left(\frac{y_i - \mu_i}{\mu_i \sigma}\right)^2.$$

So, the scaled deviance is

$$\begin{aligned}
\frac{D(\boldsymbol{y}, \widehat{\boldsymbol{\mu}})}{\phi} &= -2\left(L\left(\widehat{\boldsymbol{\mu}}, \boldsymbol{y}\right) - L\left(\boldsymbol{y}, \boldsymbol{y}\right)\right) \\
&= -2\left(-\frac{1}{2}\sum_{i=1}^{n} \log\left(2\pi\sigma^2 y_i^3\right) - \frac{1}{2}\sum_{i=1}^{n} \frac{1}{y_i}\left(\frac{y_i - \widehat{\mu}_i}{\widehat{\mu}_i \sigma}\right)^2 + \frac{1}{2}\sum_{i=1}^{n} \log\left(2\pi\sigma^2 y_i^3\right)\right) \\
&= \sum_{i=1}^{n} \frac{1}{y_i}\left(\frac{y_i - \widehat{\mu}_i}{\widehat{\mu}_i \sigma}\right)^2 \\
&= \frac{1}{\sigma^2}\sum_{i=1}^{n} \frac{1}{y_i}\left(\frac{y_i - \widehat{\mu}_i}{\widehat{\mu}_i}\right)^2.
\end{aligned}$$

Hence, the deviance is

$$D(\boldsymbol{y}, \widehat{\boldsymbol{\mu}}) = \sum_{i=1}^{n} \frac{1}{y_i}\left(\frac{y_i - \widehat{\mu}_i}{\widehat{\mu}_i}\right)^2.$$

Exercise 20

a)

i)

The probability mass function of A is

$$f(A) = \binom{A+C}{A}\pi(1)^A(1-\pi(1))^{A+C-A}.$$

Then the log-likelihood is

$$L(\pi(1)) = \log\binom{A+C}{A} + A\log\pi(1) + C\log(1-\pi(1))$$

We find the MLE of $\pi(1)$

$$\frac{\partial L(\pi(1))}{\partial\pi(1)} = \frac{A}{\pi(1)} - \frac{C}{1-\pi(1)} = 0$$

Hence $\widehat{\pi}(1) = \dfrac{A}{A+C}$.

By repeating the same for $B \sim \text{Bin}(B+D, \pi(1))$, we obtain $\widehat{\pi}(0) = \dfrac{B}{B+D}$.

ii)

We find the maximim likelihood estimator of the odds ratio by using the invariance principle:

$$\widehat{\mathrm{OR}} = \frac{\frac{\widehat{\pi}(1)}{1-\widehat{\pi}(1)}}{\frac{\widehat{\pi}(0)}{1-\widehat{\pi}(0)}} = \frac{\widehat{\pi}(1)}{\widehat{\pi}(0)} \cdot \frac{1-\widehat{\pi}(0)}{1-\widehat{\pi}(1)} = \frac{A(B+D)}{B(A+C)} \cdot \frac{D(A+C)}{C(B+D)} = \frac{AD}{BC}.$$

b)

i)

This follows from the invariance principle of the maximum likelihood estimator.

ii)

By CLT, we know that

$$\widehat{\pi}(j) \ \xrightarrow{d}\ N\left(\pi(j), \mathcal{I}_{\pi(j)}^{-1}\right)$$

where

$$\mathcal{I}_{\pi(j)} = \mathrm{E}\left[-\frac{\partial^2 L(\pi(j))}{\partial \pi(j)^2}\right] = \begin{cases} \mathrm{E}\left[\dfrac{B}{\pi(0)^2} + \dfrac{D}{(1-\pi(0))^2}\right] & \text{if } j=0 \\ \mathrm{E}\left[\dfrac{A}{\pi(1)^2} + \dfrac{C}{(1-\pi(1))^2}\right] & \text{if } j=1 \end{cases}.$$

Multivariate Delta method

Suppose that

$$\sqrt{n}(\boldsymbol{X}_n - \boldsymbol{\eta}) \ \xrightarrow{d}\ N_r\left(\mathbf{0}, \boldsymbol{\Sigma}\right),$$

for some r-dimensional random vector $\boldsymbol{X}_n$ depending on n. If S is a function $\mathbb{R}^r \to \mathbb{R}^s$ which is once differentiable at $\boldsymbol{\eta}$ and has Jacobian matrix $\dot{\boldsymbol{S}}(\boldsymbol{\eta})$, then

$$\sqrt{n}\left(S(\boldsymbol{X}_n) - S(\boldsymbol{\eta})\right) \ \xrightarrow{d}\ N_s\left(\mathbf{0},\ \dot{\boldsymbol{S}}(\boldsymbol{\eta})^{\mathrm{T}}\, \boldsymbol{\Sigma}\, \dot{\boldsymbol{S}}(\boldsymbol{\eta})\right),$$

provided that $\dot{\boldsymbol{S}}(\boldsymbol{\eta})^{\mathrm{T}}\, \boldsymbol{\Sigma}\, \dot{\boldsymbol{S}}(\boldsymbol{\eta})$ is positive definite.

Applying delta method with $\theta(j) = \log\left(\dfrac{\pi(j)}{1-\pi(j)}\right)$ gives

$$\begin{aligned} \widehat{\theta}(j) \ \xrightarrow{d}\ & N\left(\theta(j), \left(\frac{\partial \theta(j)}{\partial \pi(j)}\right)^2 \mathcal{I}_{\pi(j)}^{-1}\right) \\ &= N\left(\theta(j), \left(\mathcal{I}_{\pi(j)} \pi(j)^2 \left(1-\pi(j)\right)^2\right)^{-1}\right). \end{aligned}$$

We can estimate $\left(\mathcal{I}_{\pi(j)}\pi(j)^2\left(1-\pi(j)\right)^2\right)^{-1}$ by replacing $\pi(j)$ with $\widehat{\pi}(j)$ and $\mathcal{I}_{\pi(j)}$ by $\mathcal{J}_{\pi(j)}$ (observed information).

When $j=0$,

$$\begin{aligned}
\widehat{\operatorname{Var}}\left(\widehat{\theta}(0)\right) &= \left(\mathcal{J}_{\widehat{\pi}(0)}\widehat{\pi}(0)^2\left(1-\widehat{\pi}(0)\right)^2\right)^{-1} \\
&= \left(\left(\frac{B}{\widehat{\pi}(0)^2}+\frac{D}{(1-\widehat{\pi}(0))^2}\right)\cdot\widehat{\pi}(0)^2\left(1-\widehat{\pi}(0)\right)^2\right)^{-1} \\
&= \left(B\left(1-\widehat{\pi}(0)\right)^2+D\,\widehat{\pi}(0)^2\right)^{-1} \\
&= \left(\frac{BD^2}{(B+D)^2}+\frac{B^2D}{(B+D)^2}\right)^{-1} \\
&= \left(\frac{BD}{B+D}\right)^{-1} \\
&= \frac{B+D}{BD} \\
&= \frac{1}{B}+\frac{1}{D}.
\end{aligned}$$

Similarly, when $j=1$,

$$\widehat{\operatorname{Var}}\left(\widehat{\theta}(1)\right)=\frac{1}{A}+\frac{1}{C}.$$

c)

i)

$$\begin{aligned}
\operatorname{SE}\left(\log\widehat{\mathrm{OR}}\right)^2 &= \widehat{\operatorname{Var}}\left(\log\widehat{\mathrm{OR}}\right) \\
&= \widehat{\operatorname{Var}}\left(\widehat{\theta}(1)-\widehat{\theta}(0)\right) \\
&= \widehat{\operatorname{Var}}\left(\widehat{\theta}(1)\right)+\widehat{\operatorname{Var}}\left(\widehat{\theta}(0)\right) && (\because A \perp\!\!\!\perp B) \\
&= \frac{1}{A}+\frac{1}{C}+\frac{1}{B}+\frac{1}{D} \\
&= \frac{1}{A}+\frac{1}{B}+\frac{1}{C}+\frac{1}{D}
\end{aligned}$$

ii)

A 95% confidence interval of $\log\mathrm{OR}$:

$$\log\widehat{\mathrm{OR}}\pm z_{0.975}\cdot\operatorname{SE}\left(\log\widehat{\mathrm{OR}}\right).$$

By taking *exp* on the both side we obtain a 95% confidence interval of OR:

$$\widehat{\mathrm{OR}}\cdot\exp\left[\pm z_{0.975}\cdot\operatorname{SE}\left(\log\widehat{\mathrm{OR}}\right)\right].$$

d)

```
> # Enter the data
> diabetes = as.data.frame(matrix(c(377, 17864, 336, 20099), 2, 2))
> rownames(diabetes) = c("diseased","healthy")
> colnames(diabetes) = c("male","female")
> show(diabetes)
male female
diseased   377    336
healthy  17864  20099
>
> # Perform chi-square test
> chisq.test(diabetes)

Pearson's Chi-squared test with Yates' continuity correction

data:  diabetes
X-squared = 9.277, df = 1, p-value = 0.00232
```

p-value: $0.00232 < 0.05$. So, we conclude that there is significance difference between the occurrence of diabetes between men and women.

e)

By using the result from c), we obtain the 95% confidence interval of odds ratio for diabetes between men and women.

```
> # Estimated odds ratio
> OR.hat = diabetes[1,1]*diabetes[2,2]/(diabetes[1,2]*diabetes[2,1])
> show(OR.hat)
[1] 1.262402
> # Standard error of log of odds ratio
> std.error = sqrt(sum(1/diabetes))
>
> # 95% confidence interval of odds ratio
> low.CI = OR.hat*exp(-qnorm(0.975)*std.error)
> upp.CI = OR.hat*exp(qnorm(0.975)*std.error)
> c(low.CI, upp.CI)
[1] 1.088277 1.464387
```

Thus, $\widehat{\mathrm{OR}} = 1.2624$ and 95% confidence interval of OR: $[1.0883, 1.4644]$.

f)

i)

This follows from the fact that $\mathrm{OR} = 1$ if and only if $\pi(0) = \pi(1) \neq 0$.

ii)

$H_0 : \pi(0) = \pi(1)$. and we know $\widehat{\pi}(j) \overset{\text{approx}}{\sim} N\left(\pi(j), \mathcal{I}^{-1}_{\pi(j)}\right)$.

Since $\mathrm{OR} = 1 \iff \pi(0) = \pi(1)$, we can check whether 1 is in the 95% confidence interval of OR from e):

$$1 \notin [1.0883, 1.4644].$$

So, we reject the null hypothesis and conclude that there is significance difference between the occurrence of diabetes between men and women.

g)

If we fit logistic regression with `sex` as covariate, we have

$$\log \widehat{\mathrm{OR}} = \log\left(\frac{\frac{\widehat{\pi}(1)}{1-\widehat{\pi}(1)}}{\frac{\widehat{\pi}(0)}{1-\widehat{\pi}(0)}}\right) = \log\left(\frac{\widehat{\pi}(1)}{1-\widehat{\pi}(1)}\right) - \log\left(\frac{\widehat{\pi}(0)}{1-\widehat{\pi}(0)}\right) = \widehat{\beta}_1.$$

We can use this relationship to convert the confidence interval of $\widehat{\beta}_1$ into the confidence interval of $\widehat{\mathrm{OR}}$.

```
> # Modify data such that it's suitable for logistic regression
> diabetes.glm.form = data.frame(
+   y = as.numeric(diabetes["diseased",]),
+   n = as.numeric(colSums(diabetes)),
+   sex = c(1,0))
> diabetes.glm.form[,"sex"] = as.factor(diabetes.glm.form[,"sex"])
> head(diabetes.glm.form)
y     n sex
1 377 18241   1
2 336 20435   0
>
> # Fit logistic regerssion
> diabetes.model.1 = glm(cbind(y, n - y) ~ sex, family = binomial(link = "logit"),
    data = diabetes.glm.form)
> summary(diabetes.model.1)

Call:
glm(formula = cbind(y, n - y) ~ sex, family = binomial(link = "logit"),
data = diabetes.glm.form)

Deviance Residuals:
[1]  0  0

Coefficients:
Estimate Std. Error z value Pr(>|z|)
(Intercept) -4.09131    0.05501 -74.376  < 2e-16 ***
sex1         0.23302    0.07573   3.077  0.00209 **
---
Signif. codes:  0 ?***? 0.001 ?**? 0.01 ?*? 0.05 ?.? 0.1 ? ? 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance:  9.4891e+00  on 1  degrees of freedom
Residual deviance: -2.7651e-12  on 0  degrees of freedom
AIC: 19.389

Number of Fisher Scoring iterations: 2

>
>
> # Confidence interval based on logistic regression
> beta.1.hat = summary(diabetes.model.1)$coefficients["sex1","Estimate"]
> std.error.beta.1 = summary(diabetes.model.1)$coefficients["sex1","Std. Error"]
```

```
> low.CI = exp(beta.1.hat - qnorm(0.975)*std.error.beta.1)
> upp.CI = exp(beta.1.hat + qnorm(0.975)*std.error.beta.1)
> c(low.CI, upp.CI)
[1] 1.088278 1.464387
```

As expected, this matches the confidence interval we obtained from e).