Tutorial 9 Solutions

STAT 3013/4027/8027

1. SI Example 4.8. Simply write out the steps as outlined in the example. This is the classic t-test. **the solution is in the textbook.**

2. Consider a Poission regression model using the canonical link function (how do we determine the canonical link function?):

$$
\begin{aligned}
Y_1, \ldots, Y_n &\overset{\text{indep.}}{\sim} \text{Poisson}(\lambda_i) \\
\log(\lambda_i) &= \beta_0 + \beta_1 x_i + \beta_2 x_i^2 \\
&\text{for } i = 1, \ldots, n.
\end{aligned}
$$

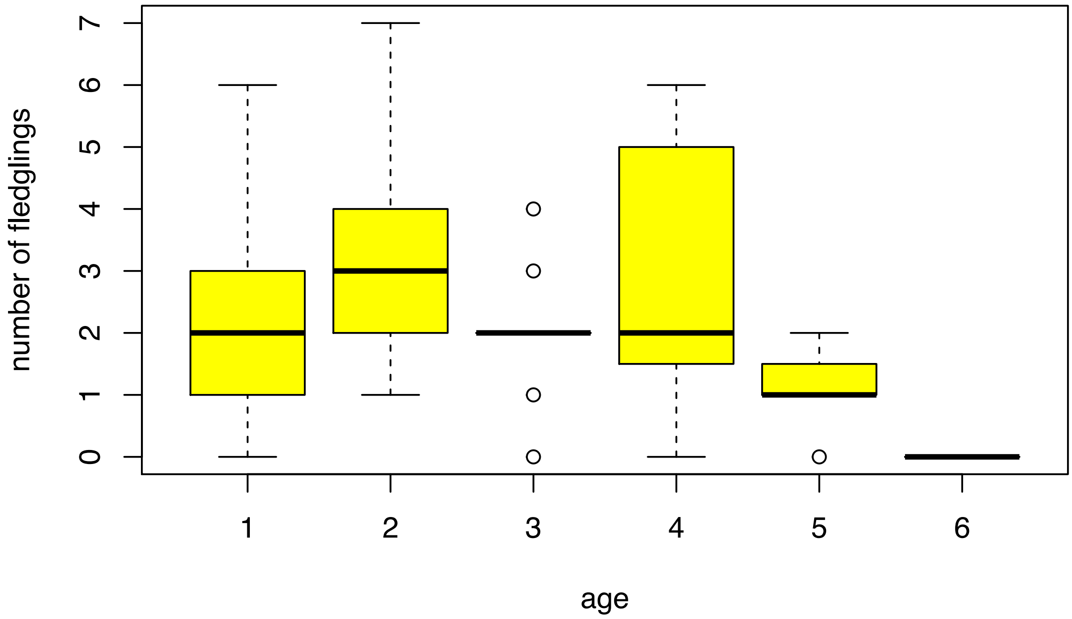
- Using `optim()` and the data on the website to find the MLEs: $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$, as well as their estimated asymptotic variances.
- Additionally, using the bootstrap procedure discussed in class, provide the estimated biases and variances for the parameters.
- Data: A sample from a population of 52 female song sparrows was studied over the course of a summer and their reproductive activities were recorded. In particular, the age and number of new offspring were recorded for each sparrow (Arcese et al, 1992). Let Y = fledged (number of offspring), and X = age (age of mother).

Ans. Let's first examine the data:

```
##
D <- read.table("Data", header=TRUE)
summary(D)
```

```
##     fledged          age
##  Min.   :0.000   Min.   :1.000
##  1st Qu.:1.000   1st Qu.:2.000
##  Median :2.000   Median :3.000
##  Mean   :2.404   Mean   :3.077
##  3rd Qu.:3.000   3rd Qu.:4.000
##  Max.   :7.000   Max.   :6.000
```

```
n <- nrow(D)

y <- D[,1]
x <- D[,2]

##
boxplot(y ~ x, col="yellow", xlab="age", ylab="number of fledglings")
```



It appears that there is some curvature, so using a quadratic term seems reasonable in order to allow for some flexibility.

- The likelihood for the model is:

$$
\begin{aligned}
L(\beta_0, \beta_1, \beta_2, |\boldsymbol{y}) &= \prod_{i=1}^{n} \left(\frac{e^{-\lambda_i} \lambda_i^{-y_i}}{y_i!} \right) \\
log(\lambda_i) &= \beta_0 + \beta_1 x_i + \beta_2 x_i^2
\end{aligned}
$$

- We will use `optim()` to get the MLEs:

```
##
log.lik <- function(theta){

  beta0 <- theta[1]
  beta1 <- theta[2]
  beta2 <- theta[3]

  out <- sum(dpois(y, exp(beta0 + beta1*x + beta2*x^2), log=TRUE))
  return(out)
  }

##
theta.start <- c(0,0,0)
out <- optim(theta.start, log.lik, hessian = TRUE,
             control = list(fnscale=-1), method="BFGS")

##
beta.hat <- out$par
beta.hat
```

```
## [1]  0.2762872  0.6821119 -0.1345834
```

- To get the estimated standard errors, recall that asymptotically:

$$\hat{\boldsymbol{\theta}} \sim \text{multivaritate normal}_3(\boldsymbol{\theta}, I(\theta)^{-1})$$

$$\hat{\theta} = \{\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2\}$$

$I(\theta)$ is a 3×3 Fisher information matrix. To get an estimate we note $[-\text{Hessian matrix}] \rightarrow I(\hat{\boldsymbol{\theta}})$.

```
var.beta.hat <- diag(solve(-out$hessian))
var.beta.hat
```

```
## [1] 0.195488995 0.114536655 0.003345583
```

```
se.beta.hat <- sqrt(var.beta.hat)
se.beta.hat
```

```
## [1] 0.44214137 0.33843264 0.05784102
```

- Of course R already has a nice function to do all of this. But now you know how to code up non-standard likelihoods and estimate the parameters:

```
## We can do the same thing using the glm() function
mod <- glm(y ~ x + I(x^2), family="poisson")
summary(mod)
```

```
##
## Call:
## glm(formula = y ~ x + I(x^2), family = "poisson")
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.4650  -0.6355  -0.2298   0.4937   2.0429
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.27662    0.44219   0.626   0.5316
## x            0.68174    0.33850   2.014   0.0440 *
## I(x^2)      -0.13451    0.05786  -2.325   0.0201 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 76.081  on 51  degrees of freedom
## Residual deviance: 67.837  on 49  degrees of freedom
## AIC: 198.78
##
## Number of Fisher Scoring iterations: 5
```

- Now let's consider using the bootstrap (**WE will talk about the bootstrap later in the course.**):
    1. Randomly sample the data with replacement.
    2. Estimate the parameters through `optim()` or `glm()` function. **Note: this fitting procedure is a statistic.**
    3. Store the estimated parameters.
    4. Repeat steps 1-3 2,000 times.

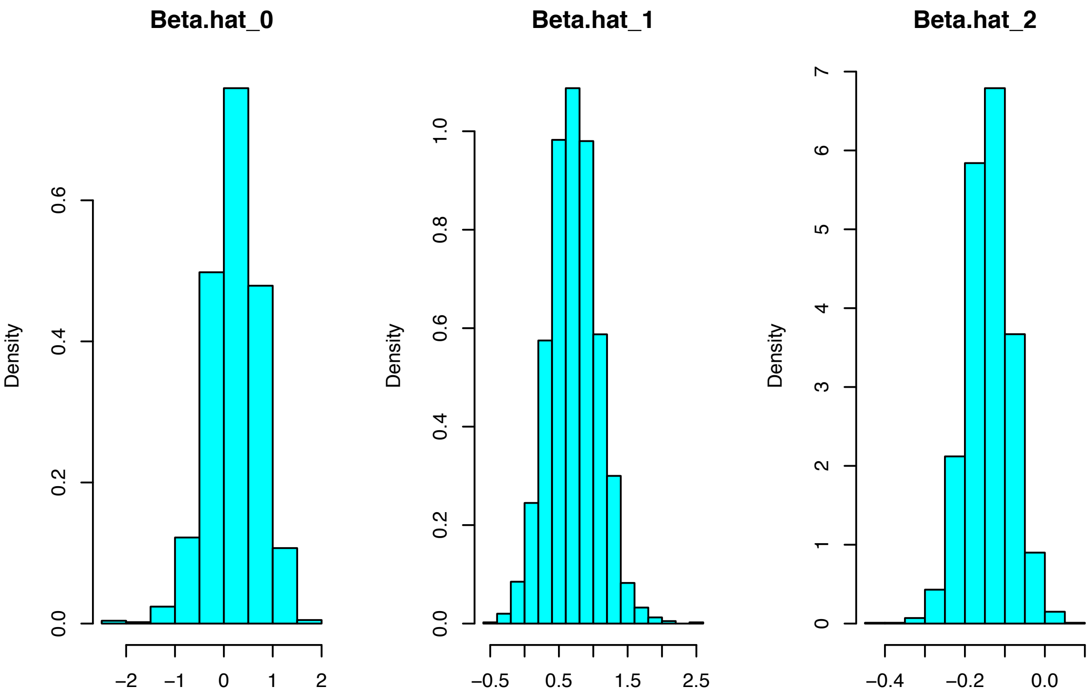
```r
set.seed(1001)
S <- 2000
Out <- matrix(0, S, 3)
n <- nrow(D)


for(s in 1:S){

  Sam <- sample(1:n, n, replace=TRUE)
  D.b <- D[Sam,]
  y <- D.b$fledged
  x <- D.b$age
  mod <- glm(y ~ x + I(x^2), family="poisson")
  Out[s, ] <- mod$coef

}

  y <- D$fledged
  x <- D$age

  bias.boot <- apply(Out, 2, mean) - glm(y ~ x + I(x^2), family="poisson")$coef
  bias.boot
```

```
##  (Intercept)            x       I(x^2)
## -0.059474960  0.032231969 -0.004655707
```

```r
  v.boot <- apply(Out, 2, var)
  sd.boot <- sqrt(v.boot)
  sd.boot
```

```
## [1] 0.51035201 0.35713958 0.05631127
```

```r
  par(mfrow=c(1,3))
  hist(Out[,1], main="Beta.hat_0", col="cyan", xlab="", prob=TRUE)
  hist(Out[,2], main="Beta.hat_1", col="cyan", xlab="", prob=TRUE)
  hist(Out[,3], main="Beta.hat_2", col="cyan", xlab="", prob=TRUE)
```



- Note that the ‘sd.boot’ values are somewhat similar to the values under the assumption of asymptotic normality. We can see from the plot that the histograms for $\hat{\beta}_0$ and $\hat{\beta}_1$ are just a bit skewed, but can be considered approximately normal.

3. SI 4.19, 5.1.

GJJ Q 4.19

$X_1, \dots, X_n \overset{iid}{\sim} \text{Poisson}(\theta)$

$H_0: \theta = \theta_0 \quad H_1: \theta \neq \theta_0$

$\Rightarrow$ We will construct Score & Wald Tests.

$$L(\theta) = \frac{e^{-n\theta} \theta^{\Sigma x_i}}{\prod_{i=1}^{n} x_i!}$$

$$\ell(\theta) = -n\theta + \Sigma x_i \log(\theta) - \Sigma \log(x_i!)$$

$$\ell'(\theta) = U(\theta) = \frac{\Sigma x_i}{\theta} - n$$

$$\ell''(\theta) = -\frac{\Sigma x_i}{\theta^2} \; ; \quad I(\theta) = -E(\ell''(\theta))$$

$$\Rightarrow -E\left(-\frac{\Sigma x_i}{\theta^2}\right) = \frac{1}{\theta^2} E(\Sigma x_i)$$

$$= \frac{n\theta}{\theta^2} = \frac{n}{\theta}$$

a.) Score Test:

matrix form

$$S = U_{\theta_0}^T I_{\theta_0}^{-1} U_{\theta_0}$$

$$= \left(\frac{\Sigma x_i}{\theta_0} - n\right) \frac{\theta_0}{n} \left(\frac{\Sigma x_i}{\theta_0} - n\right)$$

$$= \left(\frac{\Sigma x_i}{\theta_0} - n\right)^2 \frac{\theta_0}{n} \,\dot{\sim}\, \chi^2_1$$

b.) Wald test:

$$\hat{\theta} = \bar{x} \quad \Rightarrow \quad I_{\hat{\theta}} = \frac{n}{\bar{x}}$$

$$W = (\hat{\theta} - \theta_0)^T I_{\hat{\theta}} (\hat{\theta} - \theta_0)$$

$$= (\bar{x} - \theta_0)^2 \frac{n}{\bar{x}} \sim \chi^2_1$$

GJJ Q 5.1 $\quad X_1, \ldots, X_n \overset{iid}{\sim} f(x) = \frac{1}{\theta}\exp(-x/\theta)$

$$Y = \frac{2\sum x_i}{\theta} = \frac{2n\bar{x}}{\theta} \sim \chi^2_{2n}$$

Can you show this?

$$P\left(\chi^2_{\alpha/2,2n} \le \frac{2n\bar{x}}{\theta} \le \chi^2_{1-\alpha/2,2n}\right) = 1-\alpha$$

$$\Rightarrow P\left(\frac{\chi^2_{\alpha/2,2n}}{2n\bar{x}} \le \frac{1}{\theta} \le \frac{\chi^2_{1-\alpha/2,2n}}{2n\bar{x}}\right) = 1-\alpha$$

$$\Rightarrow P\left(\frac{2n\bar{x}}{\chi^2_{1-\alpha/2,2n}} \le \theta \le \frac{2n\bar{x}}{\chi^2_{\alpha/2,2n}}\right) = 1-\alpha$$

$$\therefore \left[\frac{2n\bar{x}}{\chi^2_{1-\alpha/2,2n}}, \frac{2n\bar{x}}{\chi^2_{\alpha/2,2n}}\right]$$