LNM_AS2

Zhilin Yang

2023-02-05

Q1

$$Q_1 \quad f(y) = \sqrt{\frac{\lambda}{2\pi y^3}} \exp\left\{-\frac{\lambda(y-u)^2}{2yu^2}\right\}$$

(a)

$$\log(f(y)) = \frac{1}{2}\log\lambda - \frac{1}{2}\log(2\pi y^3) - \frac{\lambda(y^2+u^2-2yu^2)}{2yu^2}$$

$$= \frac{1}{2}\log\frac{\lambda}{2\pi y^3} - \frac{\lambda y}{2u^2} + \frac{\lambda}{u} - \frac{\lambda}{2y}$$

$$\frac{\theta y}{a(\theta)} = \frac{-\lambda}{2u^2} y$$

$$\frac{\theta y}{a(\phi)} = \frac{-\lambda y}{2u^2} \qquad \text{Let } \theta = -\frac{1}{u^2} \Rightarrow a(\phi) = -2/\lambda$$

$$\Rightarrow \phi = 1/\lambda$$

$$\theta = \frac{1}{u^2} \Rightarrow \theta^{\frac{1}{2}} = \frac{1}{u}$$

$$f(x) = \exp\{\log(f(y))\}$$

$$= \exp\left\{-\frac{1}{2}\log\phi - \frac{1}{2}\log(2\pi y^3) + \frac{\theta y}{2\phi} + \frac{\theta^{\frac{1}{2}}}{\phi} - \frac{1}{\phi(2y)}\right\}$$

$$= \exp\left\{\frac{\theta y - 2\theta^{\frac{1}{2}}}{-2\phi} - \frac{1}{2\phi y} - \frac{1}{2}\log(2\pi\phi y^3)\right\}$$

2

Therefore, $\theta = \frac{1}{u^2}$, $\phi = 1/\lambda$, $a(\phi) = -2/\lambda$

$b(\theta) = 2\theta^{\frac{1}{2}} = \frac{2}{u}$ $\quad c(y,\phi) = -\frac{1}{2}\left\{\frac{1}{\theta y} + \log(2\pi\phi y^3)\right\}$

(b)

$$\frac{\partial}{\partial\theta} b(\theta) = \theta^{-\frac{1}{2}} = \frac{1}{\theta^{\frac{1}{2}}} = u \quad \Rightarrow \frac{1}{u} = \theta^{\frac{1}{2}} \quad \frac{1}{u^2} = \theta$$

The canonical Link Function is $g(u) = \frac{1}{u^2}$

(c) Please search on the Internet to find one recent application for the Inverse Gaussian distribution. You need to provide a brief description and the reference link of the application.

One application of the Inverse Gaussian Distribution is applied in an insurance and economic data. The data is skewed to the right and is also leptokurtic. The Inverse Gaussian Distribution is suitable for modeling data has such properties (nonnegative positively skewed data).

The Reference Link:
https://www.tandfonline.com/doi/full/10.1080/02664763.2018.1542668

Q2

2-) For each of the situations below, identify the response and explanatory variables, variable types, and the generalized linear model that is well-suited to model the data. Write down the linear predictor as well as the link function. Make sure to justify your answer.

(a) The effect of age, sex, daily food intake and minutes of daily exercise on a person's BMI.

Response Variable: BMI

Explanatory Variables: age, sex, daily food intake and minutes of daily exercise.

Numerical Variables: age,daily food intake and minutes of daily exercise.

categorical variables: sex

The generalized linear model that is well-suited to model the data:

Gamma Distribution. The Gamma distribution has no negative values. In addition , the distribution of predictors tend to be normal.

Linear Predictor: y = β0 + β1*x(age) + β2*x(daily_food_intake) + β3*x(minutes_of_daily_exercise) + β4*x(sex)

Link Function: Xβ = -1/μ

(b) The effect of LNM course grade, sex, age, GPA, major, prior years of work experience, and prior income levels on whether a MScA students finds an employment upon graduation

Response Variable: whether a MScA students finds an employment upon graduation

Explanatory Variables: LNM course grade, sex, age, GPA, major, prior years of work experience, and prior income levels.

Numerical Variables: LNM course grade, age, GPA, prior years of work experience, prior income levels

categorical variables: sex, major

The generalized linear model that is well-suited to model the data: Bernoulli distribution. As the response variable is binary, and we do not need to repeat the experiment for every input.

Linear Predictor: Y = β0+β1*x(LNM_course_grade)+β2*x(age)+β3*x(GPA)+β4*x(prior_years_of_work_experience)+β5*x(prior_income_levels)+β6*x(sex)+β7*x(major)

Link Function: Xβ =ln(μ/1-μ)

(c) The number of mortgage loan defaults in a given year by different counties across the United States. For each household/borrower information on income, loan interest rate, age, debt, loan to value at origination are available.

Response Variable: The number of mortgage loan defaults in a given year by different counties across the United States.

Explanatory Variables: income, loan interest rate, age, debt, loan to value at origination

Numerical Variables: :income, loan interest rate, age, debt,loan to value at origination

categorical variables: None

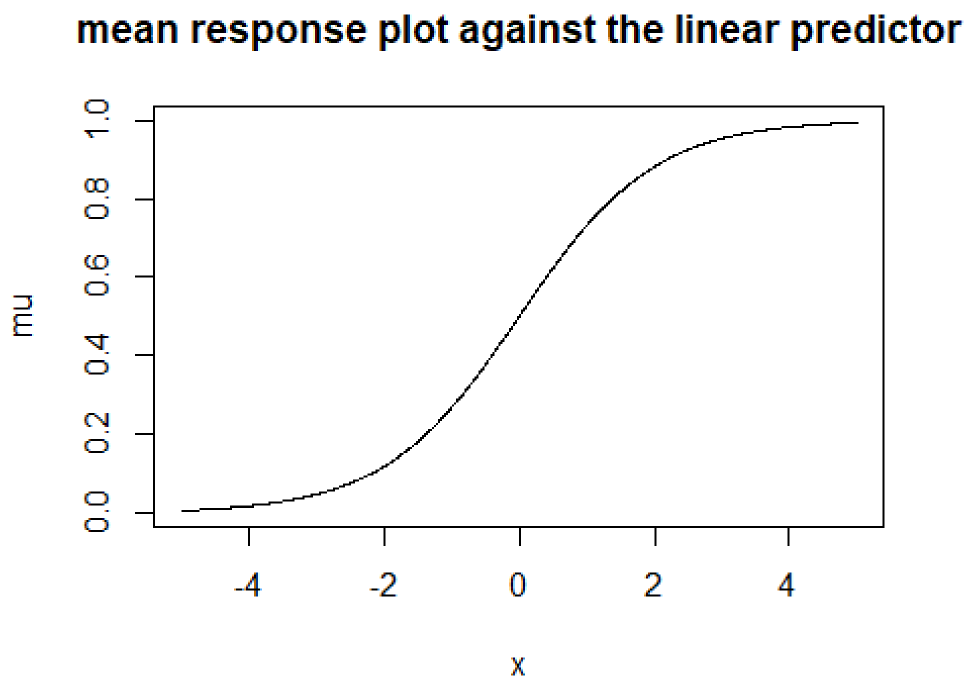
The generalized linear model that is well-suited to model the data: Poisson Distribution Model. The Poisson Distribution refers to the number of occurrences in a given time.

Linear Predictor: Y = β0 + β1*x(income) + β2*x(loan_interest_rate) + β3*x(age) + β4*x(debt) + β5*x(loan_to_value_at_origination)

Link Function: y=ln(μ)

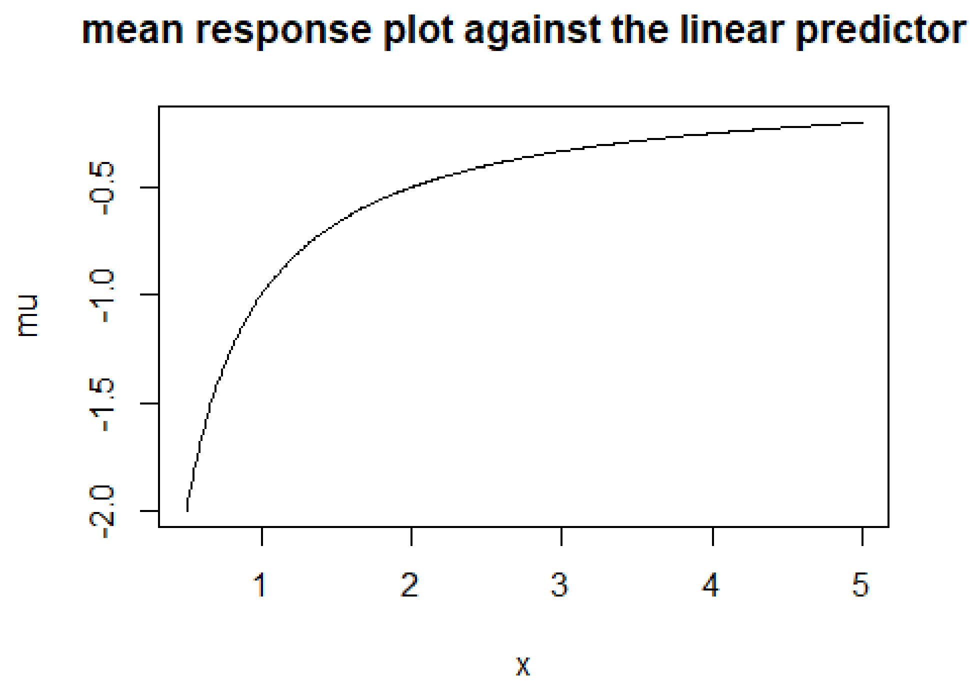
Q3

```
x=seq(-5,5,by=0.01)
mean.response= function(x) exp(x)/(1+exp(x))
mu= mean.response(x)
plot(x,mu,type='l',main='mean response plot against the linear predicto
r')
```



Q4

```
x=seq(0.5,5,by=0.01)
mean.response= function(x) -1/x
mu= mean.response(x)
plot(x,mu,type='l',main='mean response plot against the linear predicto
r')
```



Q5:

Q_5 $\quad f(y;\theta) = \theta e^{-\theta y}$

$$= \exp\{\log\theta - \theta y\}$$

The exponential family Distribution has general form

$$f(x) = \exp\left\{\frac{\theta y - b(\theta)}{a(\phi)} + c(y,\phi)\right\}$$

$\Rightarrow$ $\theta = -\theta$ $\quad b(\theta) = -\log\theta$

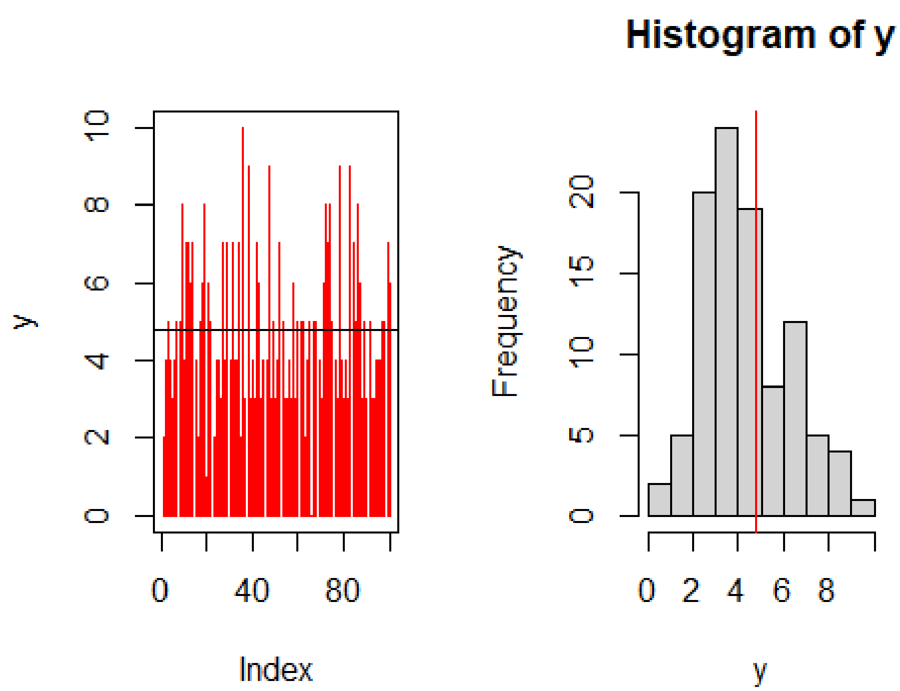
$a(\phi) = 1$ $\quad c(y,\phi) = 0$

Q6

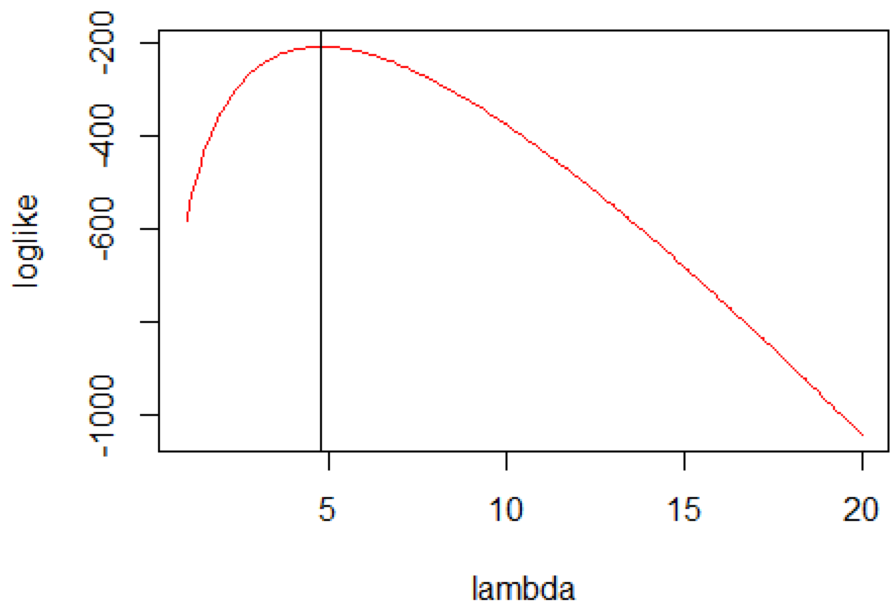
```
data = read.csv('C:/Users/Richa/Desktop/LNM/poisson_mle.csv')
head(data)
```

```
##   data
## 1    2
## 2    4
## 3    5
## 4    4
```

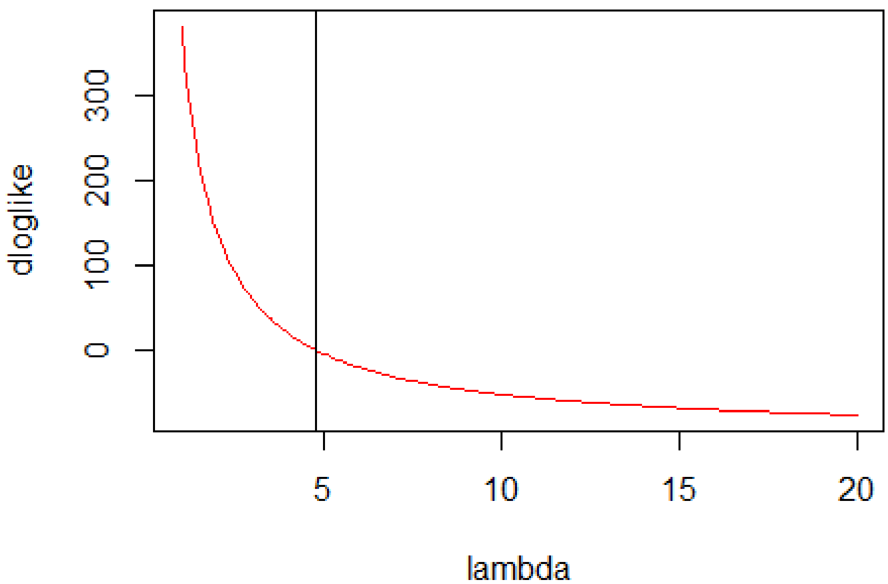
```
## 5    3
## 6    4

lb <- seq(1, 20, by=0.1)

y=data[,1]

#Calculate the log likelihood for various values of p
loglike <- log(lb)*sum(y) - length(y)*lb - sum(log(factorial(y)))

#Calculate the derivative of the log likelihood for various values of p
dloglike <- sum(y)/lb - length(y)

#Calculate the MLE esimtate for p
lb_hat <- mean(y)
lb_hat

## [1] 4.8

#Plot the results
par(mfrow=c(1,2))

plot(y,type="h", col="red")
abline(h=lb_hat)

hist(y)
abline(v=lb_hat, col="red")
```



```
#Plot the log likelihood with a vertical line showing the MLE estimate f
or p
plot(x = lb, y = loglike, type='l', col = 'red', xlab = 'lambda')
abline(v = lb_hat, col='black')
```




```
#Plot the derivative of the log likelihood with a vertical line showing
the MLE estimate for p
plot(x = lb, y = dloglike, type='l', col = 'red', xlab = 'lambda')
abline(v = lb_hat, col='black')
```



As we can see from the plot, the log likelihood is maximized when lambda = 4.8, which matches the result from the sequence generated.

Q7 $\quad f(y;\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} \exp\{-\beta y\}$

$$\log(f(y)) = \alpha\log\beta - \log(\Gamma(\alpha)) + (\alpha-1)\log y - \beta y$$

$$f(x) = \exp(\log(f(y))) = \exp\{-\beta y + \alpha\log\beta + (\alpha-1)\log y - \log\Gamma(\alpha)\}$$

$$= \exp\left\{\frac{-\beta/\alpha\cdot y + \frac{\alpha}{\alpha}\log\beta}{1/\alpha} + (\alpha-1)\log y - \log\Gamma(\alpha)\right\}$$

$$= \exp\left\{\frac{-\beta/\alpha\cdot y - \log\beta}{-1/\alpha} + (\alpha-1)\log y - \log\Gamma(\alpha)\right.$$

$\Rightarrow$ let $\theta = \frac{\beta}{\alpha} \qquad \phi = 1/\alpha \qquad a(\phi) = -1/\alpha$

$\beta = \theta\alpha = \theta/\phi$

$\log\beta = \log\frac{\theta}{\alpha} = \log\theta - \log\alpha$

$$\Rightarrow f(x) = \exp\left\{\frac{\theta y - \log\theta}{-\phi} + \frac{\log\phi}{\phi} + \left(\frac{1}{\phi}-1\right)\log y - \log\left(\Gamma\left(\frac{1}{\phi}\right)\right)\right\}$$

$$= \exp\left\{\frac{\theta y - \log\theta}{-\phi} + c(y,\phi)\right\}$$

which is in the form of $\exp\left\{\frac{y\theta - b(\theta)}{\alpha(\theta)} + c(y,\phi)\right\}$

Therefore, the gamma distribution

satisfy the required form of exponential family of distributions

Since $b(\theta) = \log\theta \qquad \alpha(\phi) = -\phi$

$$\Rightarrow E(Y) = \frac{\partial}{\partial\theta} b(\theta) = \frac{\partial}{\partial\theta}\log\theta = \frac{1}{\theta}$$

$$Var(Y) = a(\phi)\cdot\frac{\partial^2}{\partial\theta^2} b(\theta) = -\phi\cdot\frac{\partial}{\partial\theta}\frac{1}{\theta}$$

$$= -\phi\cdot(-\theta^{-2})$$

$$= \phi\theta^{-2}$$

Since $\theta = \beta/\alpha \qquad \phi = 1/\alpha$

$\Rightarrow E(Y) = \frac{1}{\theta} = \beta/\alpha \qquad Var(Y) = \phi\theta^{-2} = \frac{1}{\alpha}\frac{\beta}{\alpha} = \frac{\beta}{\alpha^2}$