Statistical Modeling 1
Problem set 4 2024

1. (a) Let us assume $Y_i \sim IG(\mu_i, \phi)$. Consider the model

$$\log(\mu_i) = \beta_0 + \beta_1 \log(x_i).$$

Let the estimates of the parameters β_0, β_1, ϕ be as $\hat{\beta}_0 = 1, \hat{\beta}_1 = 0.5, \tilde{\phi} = 0.05$, when

$$\mathbf{X} = \begin{pmatrix} 1 & \log(3) \\ 1 & \log(3) \\ 1 & \log(3) \\ 1 & \log(6) \\ 1 & \log(6) \\ 1 & \log(6) \\ 1 & \log(9) \\ 1 & \log(9) \\ 1 & \log(9) \end{pmatrix}.$$

Calculate the estimated covariance matrix $\widehat{\mathrm{Cov}}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\widehat{\mathbf{W}}\mathbf{X})^{-1}$.

(2 points)

Solution:

```
> phi<-0.05
> betahat<-t(t(c(1,0.5)))
> X<-cbind(1,log(rep(c(3,6,9), each=3)))
> link.hat<-X%*%betahat
> mu.hat<-exp(link.hat)
> D<-diag(as.numeric(mu.hat))
> Var.Y<-phi*mu.hat^3
> V<-diag(as.numeric(Var.Y))
> W<-D%*%solve(V)%*%D
> W
          [,1]     [,2]     [,3]     [,4]     [,5]     [,6]    [,7]    [,8]    [,9]
 [1,] 4.247906 0.000000 0.000000 0.000000 0.000000 0.000000 0.00000 0.00000 0.00000
 [2,] 0.000000 4.247906 0.000000 0.000000 0.000000 0.000000 0.00000 0.00000 0.00000
 [3,] 0.000000 0.000000 4.247906 0.000000 0.000000 0.000000 0.00000 0.00000 0.00000
 [4,] 0.000000 0.000000 0.000000 3.003723 0.000000 0.000000 0.00000 0.00000 0.00000
 [5,] 0.000000 0.000000 0.000000 0.000000 3.003723 0.000000 0.00000 0.00000 0.00000
 [6,] 0.000000 0.000000 0.000000 0.000000 0.000000 3.003723 0.00000 0.00000 0.00000
 [7,] 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 2.45253 0.00000 0.00000
 [8,] 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.00000 2.45253 0.00000
 [9,] 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.00000 0.00000 2.45253
> cov.betahat<-solve(t(X)%*%W%*%X)
> cov.betahat
           [,1]       [,2]
[1,]  0.4453876 -0.2583823
[2,] -0.2583823  0.1624215
```

(b) Let $Y_i \sim Poi(\mu_i)$. Then the probability density function of the random variable Y_i is

$$f(y_i|\mu_i) = \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!}.$$

Show first that Y_i belongs to the exponential family of distributions, and then show that

$$\begin{aligned} \mathrm{E}(Y_i) &= \mu_i, \\ \mathrm{Var}(Y_i) &= \mu_i. \end{aligned}$$

Hint! There is no dispersion parameter ϕ in Poisson distribution and hence you may consider it as $\phi = 1$.

(2 points)

Solution:

In Poisson distribution, the density function has the form

$$\begin{aligned} f(y_i|\mu_i) &= \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!} = \frac{\exp(-\mu_i + y_i \log(\mu_i))}{y_i!} \\ &= \exp\left(\frac{y_i \log(\mu_i) - e^{\log(\mu_i)}}{1} - \log(y_i!)\right) \\ &= \exp\left(\frac{y_i\Theta_i - b(\Theta_i)}{a(\phi)} + c(y_i, \phi)\right) \end{aligned}$$

where

$$\Theta_i = \log(\mu_i),\ a(\phi) = \phi,\ \phi = 1,\ b(\Theta_i) = e^{\Theta_i},\ c(y_i, \phi) = -\log(y_i!).$$

Since for distributions belonging to Exponential Family of Distributions, the expected value is $\mathrm{E}(Y_i) = b'(\Theta_i)$ and the variance is $\mathrm{Var}(Y_i) = b''(\Theta_i)a(\phi)$, we got

$$\begin{aligned} \mathrm{E}(Y_i) &= b'(\Theta_i) = e^{\Theta_i} = e^{\log(\mu_i)} = \mu_i, \\ \mathrm{Var}(Y_i) &= b''(\Theta_i)a(\phi) = e^{\Theta_i} \cdot 1 = e^{\log(\mu_i)} = \mu_i. \end{aligned}$$

(c) Consider the simple Gamma model with

$$\begin{aligned} Y_i &\sim Gamma(\mu_i, \phi), \\ \mu_i &= \eta_i = \beta_0. \end{aligned}$$

Construct the $100(1-\alpha)\%$ prediction interval for the new observation Y_f.

(2 points)

Solution:

Based on the weekly problem set 3, the question 3(c), we know that $\hat{\beta}_0 = \frac{(\sum_{i=1}^n y_i)}{n} = \bar{y}$, and hence

$$\begin{aligned} \tilde{\phi} &= \frac{\sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{v(\hat{\mu}_i)}}{n - \mathrm{rank}(\mathbf{X})} = \frac{\sum_{i=1}^n \frac{(y_i - \hat{\beta}_0)^2}{\hat{\beta}_0^2}}{n-1} = \frac{1}{\bar{y}^2} \cdot \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1} = \frac{s_y^2}{\bar{y}^2}, \\ \widehat{\mathrm{Var}}(Y_i) &= \tilde{\phi}\hat{\mu}_i^2 = \tilde{\phi}\hat{\beta}_0^2 = \frac{s_y^2}{\bar{y}^2} \cdot \bar{y}^2 = s_y^2, \end{aligned}$$

since $\mathbf{X} = \mathbf{1}$. Furthermore, since $\frac{\partial \mu_i}{\partial \eta_i} = 1$, we have

$$\widehat{\mathbf{D}} = \mathbf{I}, \quad \widehat{\mathbf{V}} = s_y^2 \mathbf{I}, \quad \widehat{\mathbf{W}} = \widehat{\mathbf{D}}\widehat{\mathbf{V}}^{-1}\widehat{\mathbf{D}} = \frac{1}{s_y^2}\mathbf{I},$$

and thus

$$\widehat{\mathrm{Cov}}(\hat{\boldsymbol{\beta}}) = (\mathbf{1}'\widehat{\mathbf{W}}\mathbf{1})^{-1} = \left(\frac{1}{s_y^2}\mathbf{1}'\mathbf{1}\right)^{-1} = \frac{s_y^2}{n}.$$

Estimated variance of the prediction error is now

$$\begin{aligned} \widehat{\mathrm{Var}}(e_f) &= \widehat{\mathrm{Var}}(Y_f) + \left(\frac{\partial \hat{\mu}_f}{\partial \hat{\eta}_f}\right)^2 \cdot \mathbf{x}_f' \widehat{\mathrm{Cov}}(\hat{\boldsymbol{\beta}})\mathbf{x}_f \\ &= s_y^2 + (1)^2 \cdot 1 \cdot \frac{s_y^2}{n} \cdot 1 = s_y^2 + \frac{s_y^2}{n}, \end{aligned}$$

and the prediction interval is

$$\begin{gathered} \left[\hat{\beta}_0 - z_{\alpha/2}\sqrt{\widehat{\mathrm{Var}}(e_f)}, \hat{\beta}_0 + z_{\alpha/2}\sqrt{\widehat{\mathrm{Var}}(e_f)}\right], \\ \left[\bar{y} - z_{\alpha/2}\sqrt{s_y^2 + \frac{s_y^2}{n}}, \bar{y} + z_{\alpha/2}\sqrt{s_y^2 + \frac{s_y^2}{n}}\right], \end{gathered}$$

where $z_{\alpha/2}$ such that $P(Z > z_{\alpha/2}) = \alpha/2$, when $Z \sim N(0, 1)$.