## GLM Exercise for Week 6 (EFD, MLE, Binary Regression)

### Question 1 (MLE example)

An insurance company reported 75 claim amounts, and you are given the following:

- $\sum_{i=1}^{75} y_i = 91158.4$
- $\sum_{i=1}^{75} \ln(y_i) = 393.5$
- $\sum_{i=1}^{75} (\ln(y_i))^2 = 2353.6$

Assume claims came from a lognormal distribution with parameters $\mu$ and $\sigma$.
Estimate the parameters of the lognormal distribution using the MLE (maximum likelihood estimation) method.

Solution:

- Relevant concept: Workshop Week 6 Slides MLE for univariate variable.
- Let $Y \sim LN(\mu, \sigma^2)$
  - We have $f(y) = \frac{1}{y_i \times \sigma \times \sqrt{2\pi}} \times \exp\left(-\frac{1}{2\sigma^2}(\ln(y_i) - \mu)^2\right)$.
  - Then $L = \prod_{i=1}^{75} f(y_i) = \prod_{i=1}^{75} \left(\frac{\exp\left(-\frac{1}{2\sigma^2}(\ln(y_i)-\mu)^2\right)}{y_i\sqrt{2\pi\sigma^2}}\right)$.
  - Hence $l = \ln(L) = -\frac{1}{2}\sum_{i=1}^{75}\left(\frac{\ln(y_i)-\mu}{\sigma}\right)^2 - 75\ln\sigma - 75\ln\sqrt{2\pi} - \sum_{i=1}^{75}\ln(y_i)$.
- Take the first-order derivative and set up the equations to 0.
  - $\frac{\partial l}{\partial \mu} = \frac{\sum_{i=1}^{75}(\ln(y_i)-\mu)}{\sigma^2} = 0.$
  - $\frac{\partial l}{\partial \sigma} = \frac{\sum_{i=1}^{75}(\ln(y_i)-\mu)^2}{\sigma^3} - \frac{75}{\sigma} = 0.$
- Finally, we have
  - $\hat{\mu} = \frac{\sum_{i=1}^{75}\ln(y_i)}{75} = \frac{393.5}{75} = 5.25.$
  - $\widehat{\sigma^2} = \frac{\sum_{i=1}^{75}(\ln(y_i))^2 + 75\times\mu^2 - 2\times\mu\times\sum_{i=1}^{75}\ln(y_i)}{75} = \frac{2353.6+75\times5.25^2-2\times5.25\times393.5}{75} = 3.85.$

**Question 2 (MLE example from past exam)**

The government is introducing an unemployment program to help people that is currently unemployed. There are 1000 participants in the program. Let $y_i$ denote the unemployment benefits (in dollars) that each program participant receives and given the following:

- $\sum_{i=1}^{1000} \ln(y_i) = 6294.681$, $\sum_{i=1}^{1000} \{\ln(y_i)\}^2 = 39662.75$, $\bar{y} = 552.5599$, $s_y = 111.5146$
- The 25th, 50th, and 95th percentiles of $y_i$ are 475.9673, 540.2550, and 615.9487 respectively.

a) Assumes $y_i$ came from a **gamma** distribution with shape parameter $\alpha$ and rate parameter $\beta$, where $\beta > 0$ and $\alpha$ equals 2. The probability density function is $f(y_i) = \frac{\beta^\alpha}{\Gamma(\alpha)} \times y_i^{\alpha-1} \times \exp(-\beta \times y_i)$, and $y_i > 0$. Estimate the parameter $\beta$ of the gamma distribution using the MLE (maximum likelihood estimation) method.

b) Assumes $y_i$ came from a **Pareto** distribution with shape parameter $\alpha$ and scale parameter $k$, where $\alpha > 1$ and $k$ equals 276.1067. The probability density function is $f(y_i) = \frac{\alpha \times k^\alpha}{y_i^{\alpha+1}}$, and $y_i \geq k$. Estimate the parameter $\alpha$ of the Pareto distribution using the MLE (maximum likelihood estimation) method.

Solution:

a)

- Relevant concept: Workshop Week 6 Slides MLE for univariate variable.
- Given that
  - $Y \sim Gamma(\alpha, \beta)$, $f(y_i) = \frac{\beta^\alpha}{\Gamma(\alpha)} \times y_i^{\alpha-1} \times \exp(-\beta \times y_i)$.
  - $n = 1000$, $\bar{y} = 552.5599$ and $\alpha = 2$.
- Step 1: Likelihood $L = \prod_{i=1}^{n=1000} f(y_i)$.
- Step 2: Log-likelihood $l = \ln(L) = \ln\left(\prod_{i=1}^{n=1000} f(y_i)\right) = \sum_{i=1}^{n=1000} \ln\left(f(y_i)\right) = \sum_{i=1}^{n=1000} \left\{\alpha \times \ln(\beta) - \ln\left(\Gamma(\alpha)\right) + (\alpha - 1) \times \ln(y_i) - \beta \times y_i \times \ln(e)\right\} = \sum_{i=1}^{n=1000} \left\{\alpha \times \ln(\beta) - \ln\left(\Gamma(\alpha)\right) + (\alpha - 1) \times \ln(y_i) - \beta \times y_i\right\} = n \times \alpha \times \ln(\beta) - n \times \ln\left(\Gamma(\alpha)\right) + (\alpha - 1) \times \ln\left(\prod_{i=1}^{n=1000} y_i\right) - \beta \times \sum_{i=1}^{1000} y_i$.
- Step 3: Take the first-order derivative with respect to $\beta$ (given that $\alpha = 2$) and set up the equation to be 0: $\frac{\partial l}{\partial \beta} = \frac{n \times \alpha}{\beta} - \sum_{i=1}^{n=1000} y_i = 0$. Therefore, $\hat{\beta} = \frac{\alpha}{\frac{\sum_{i=1}^{n=1000} y_i}{n}} = \frac{2}{\bar{y}} = \frac{2}{552.5599} =$ $\mathbf{0.0036195}$.

b)

- Relevant concept: Workshop Week6 Slides MLE for univariate variable.
- Given that
  - $Y \sim Pareto(\alpha, k)$, $f(y_i) = \frac{\alpha \times k^\alpha}{y_i^{\alpha+1}}$.
  - $n = 1000$, $\sum_{i=1}^{n=1000} \ln(y_i) = 6294.681$ and $k = 276.1067$.

- Step 1: Likelihood $L = \prod_{i=1}^{n=1000} f(y_i)$.
- Step 2: Log-likelihood $l = \ln(L) = \ln(\prod_{i=1}^{n=1000} f(y_i)) = \sum_{i=1}^{n=1000} \ln(f(y_i)) = \sum_{i=1}^{n=1000}\{\ln(\alpha) + \alpha \times \ln(k) - (\alpha+1) \times \ln(y_i)\} = n \times \ln\alpha + n \times \alpha \times \ln(k) - (\alpha+1) \times \sum_{i=1}^{n=1000} \ln(y_i)$.

Step 3: Take the first order derivative with respect to $\alpha$ (given that $k = 276.1067$) and set up the equation to be 0: $\frac{\partial l}{\partial \alpha} = \frac{n}{\alpha} + n \times \ln(k) - \sum_{i=1}^{n=1000} \ln(y_i) = 0$. Therefore, $\hat{\alpha} = \frac{n}{\sum_{i=1}^{n=1000} \ln(y_i) - n \times \ln(k)} = \frac{1000}{6294.681 - 1000 \times \ln(276.1067)} = \mathbf{1.4839.}$

**Question 3 (EFD example)**

Lily wants to build a regression model to predict the binary labour force outcome ($Y$). She assumes this response variable follows a ***Bernoulli*** distribution. The covariates (predictors, $X$) she uses include an individual's age, gender, and the labour force history.

The probability density function of $Y|X$ (use $Y$ for short) is $f_y(y|\pi) = \pi^y(1-\pi)^{1-y}$, for $y =$ 0 or 1, and where $\pi$ (a constant known) is the parameter for $Y$ with $0 \leq \pi \leq 1$. Assume that $a(\phi) = \phi = 1$, and the ***canonical*** link function $g(\mu) = \ln\frac{\pi}{1-\pi} = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k$.

It is known that if the conditional distribution of $Y|X$ (use $Y$ for short) follows an exponential family distribution, then its density can be written as $f_y(y|\theta,\phi) = \exp\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\}$ for some specified function $a(\cdot), b(\cdot)$ and $c(\cdot)$, where $b(\theta)$ is the cumulant generating function.

a) Show that the distribution of the random variable $Y|X$ (use $Y$ for short) belongs to exponential family distribution (EFD). In particular,
   - express $a(\phi)$ based on $\phi$ and $b(\theta)$ based on $\theta$ respectively.
   - find $\phi, \theta$ based on $\pi$.
   - express $c(y,\phi)$ based on $y$ and $\phi$.

b) Show that $E(Y) = \pi$, and $V(Y) = \pi(1-\pi)$.

c) Calculate the variance function $V(\mu_y)$.

Solution:

a)
- Relevant concept: Workshop Week 6 Slides EFD math proof.
- $f_y(y|\pi) = \pi^y(1-\pi)^{1-y}$.
  - By taking the log: $\ln\left(f_y(y|\pi)\right) = \ln(\pi^y(1-\pi)^{1-y}) = y\ln(\pi) + (1-y)\ln(1-\pi)$.
  - By taking exponential: $f_y(y|\pi) = \exp(y\ln(\pi) + (1-y)\ln(1-\pi)) = \exp(y\ln(\pi) + \ln(1-\pi) - y\ln(1-\pi)) = \exp\left(y\ln\left(\frac{\pi}{1-\pi}\right) + \ln(1-\pi)\right)$.
- Compare this to the general density $f_y(y|\theta,\phi) = \exp\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\}$.
  - Given $a(\phi) = \phi = 1$, and the **canonical** link function $g(\mu) = \ln\frac{\pi}{1-\pi} = \theta$ (by definition).
  - Therefore, the general density $f_y(y|\theta,\phi) = \exp\left\{\frac{y\times\theta - b(\theta)}{1} + c(y,\phi)\right\} = \exp\left(y\ln\left(\frac{\pi}{1-\pi}\right) + \ln(1-\pi)\right) = \exp\{\frac{y\ln\left(\frac{\pi}{1-\pi}\right) - \{-\ln(1-\pi)\}}{1} + 0\}$.
- Equate each component in both
  - $f_y(y|\theta,\phi) = \exp\left\{\frac{y\times\theta - b(\theta)}{1} + c(y,\phi)\right\}$, and
  - $f_y(y|\pi) = \exp\{\frac{y\ln\left(\frac{\pi}{1-\pi}\right) - \{-\ln(1-\pi)\}}{1} + 0\}$
- We have
  - $a(\phi) = \phi = 1$ (given).

- $\theta = \ln\frac{\pi}{1-\pi} \rightarrow \pi = \frac{\exp(\theta)}{1+\exp(\theta)}$.
- $c(y, \phi) = c(y, 1) = 0$.
- $b(\theta) = -\ln(1-\pi) = -\ln\left(1 - \frac{\exp(\theta)}{1+\exp(\theta)}\right) = -\ln\left(\frac{1+\exp(\theta)-\exp(\theta)}{1+\exp(\theta)}\right) = -\ln\left(\frac{1}{1+\exp(\theta)}\right) = \ln(1+\exp(\theta))$.

b)

- We have
  - $b(\theta) = \ln(1+\exp(\theta))$, and
  - $\theta = \ln\frac{\pi}{1-\pi}$.
- Therefore, by definition:
  - $E(Y) = b'(\theta) = \frac{\exp(\theta)}{1+\exp(\theta)} = \pi$.
  - $V(Y) = a(\phi) \times b''(\theta) = 1 \times b''(\theta) = \left(\frac{\exp(\theta)}{1+\exp(\theta)}\right)' = \pi(1-\pi)$.

c)

- Variance function: $V(\mu_y) = b''(\theta) = \pi(1-\pi)$.

**Question 4 (EFD example from past exam)**

Monika is an actuary working for a general insurer which specializes in selling car insurance policies. Monika wants to build a regression model to predict the claim size (i.e., the amount of claim, $Y$) from the policyholders. She assumes this response variable follows a distribution called Machop. The ***possible*** covariates (predictors, $X$) she uses include the policyholder's age, gender, residence location, claim history, and the car type the policyholder drives.

The probability density function for a Machop random variable $Y|X$ (use $Y$ for short) is $f_y(y|\mu,\lambda) = \left(\frac{\lambda}{2\pi y^3}\right)^{\frac{1}{2}} \times \exp\left\{\frac{-\lambda(y-\mu)^2}{2\mu^2 y}\right\}$, for $y > 0$ (continuous), and where $\mu$ (mean parameter) and $\lambda$ (shape parameter) are two parameters for $Y$ with $\mu > 0, \lambda > 0$. Assume that $a(\phi) = -2\phi$, and the ***canonical*** link function $g(\mu) = \frac{1}{\mu^2} = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k$.

It is known that if the conditional distribution of $Y|X$ (use $Y$ for short) follows an exponential family distribution, then its density can be written as $f_y(y|\theta,\phi) = \exp\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\}$ for some specified function $a(\cdot), b(\cdot)$ and $c(\cdot)$, where $b(\theta)$ is the cumulant generating function.

a) Find an expression for the *inverse* canonical link function of a Machop response in terms of $X$ and $\beta$.
b) Show whether the Machop distribution belongs to the exponential family or not. (*Hint*: you may assume that $\mu$ and $\lambda$ are two known constants.)
c) Show that for this Machop distribution, $E(Y) = \mu$. Express the variance function in terms of $\mu$.

## Solutions:

- Relevant concept: Workshop Week 6 Slides EFD math proof.

a)

- The **canonical** link function $g(\mu) = \frac{1}{\mu^2} = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k \rightarrow$ Solve for $\mu$.
- The inverse canonical link function is therefore $\mu = \frac{1}{(\beta_0+\beta_1 X_1+\cdots+\beta_k X_k)^{\frac{1}{2}}}$.

b)

- Yes.
- $f_y(y|\mu,\lambda) = \left(\frac{\lambda}{2\pi y^3}\right)^{\frac{1}{2}} \times \exp\left\{\frac{-\lambda(y-\mu)^2}{2\mu^2 y}\right\}$.
  - By taking the log: $\ln\left(f_y(y|\mu,\lambda)\right) = \frac{1}{2}\ln(\lambda) - \frac{1}{2}\ln(2\pi y^3) - \frac{\lambda y}{2\mu^2} + \frac{\lambda}{\mu} - \frac{\lambda}{2y}$.
  - By taking exponential: $f_y(y|\mu,\lambda) = \exp\left\{\frac{1}{2}\ln(\lambda) - \frac{1}{2}\ln(2\pi y^3) - \frac{\lambda y}{2\mu^2} + \frac{\lambda}{\mu} - \frac{\lambda}{2y}\right\} = \exp\left\{-\frac{\lambda y}{2\mu^2} + \frac{\lambda}{\mu} - \frac{\lambda}{2y} + \frac{1}{2}\ln(\lambda) - \frac{1}{2}\ln(2\pi y^3)\right\}$.

- Compare this to the general density $f_y(y|\theta,\phi) = \exp\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\}$.
  - Given $a(\phi) = -2\phi$, and the **canonical** link function $g(\mu) = \frac{1}{\mu^2} = \theta$ (by definition).
  - Therefore, the general density $f_y(y|\theta,\phi) = \exp\{\frac{y\frac{1}{\mu^2} - b(\theta)}{-2\phi} + c(y,\phi)\}$.
- Equate each component in both
  - $f_y(y|\mu,\lambda) = \exp\left\{-\frac{\lambda y}{2\mu^2} + \frac{\lambda}{\mu} - \frac{\lambda}{2y} + \frac{1}{2}\ln(\lambda) - \frac{1}{2}\ln(2\pi y^3)\right\}$, and
  - $f_y(y|\theta,\phi) = \exp\{\frac{y\frac{1}{\mu^2} - b(\theta)}{-2\phi} + c(y,\theta)\}$.
- Therefore, we have
  - $-\frac{\lambda y}{2\mu^2} = \frac{y\frac{1}{\mu^2}}{-2\phi} \rightarrow \boldsymbol{\phi = \frac{1}{\lambda}} \rightarrow \boldsymbol{a(\phi)} = -2\phi = \boldsymbol{\frac{-2}{\lambda}}$.
  - $\frac{\lambda}{\mu} = \frac{-b(\theta)}{-2\phi} \rightarrow \boldsymbol{b(\theta)} = \boldsymbol{\frac{2}{\mu}}$ (since $\phi = \frac{1}{\lambda}$) $= 2 \times \theta^{\frac{1}{2}}$ (since $\frac{1}{\mu^2} = \theta$).
  - $-\frac{\lambda}{2y} + \frac{1}{2}\ln(\lambda) - \frac{1}{2}\ln(2\pi y^3) = c(y,\phi) \rightarrow$ express in terms of $\phi = \frac{1}{\lambda} \rightarrow$ $\boldsymbol{c(y,\phi) = -\frac{1}{2y\phi} + \frac{1}{2}\ln\left(\frac{1}{\phi}\right) - \frac{1}{2}\ln(2\pi y^3)}$.
- As a result:
  - $\boldsymbol{\phi = \frac{1}{\lambda}, \frac{1}{\mu^2} = \theta}$
  - $\boldsymbol{a(\phi) = -2\phi = \frac{-2}{\lambda}, b(\theta) = 2 \times \theta^{\frac{1}{2}} = \frac{2}{\mu}, c(y,\phi) = -\frac{1}{2y\phi} + \frac{1}{2}\ln\left(\frac{1}{\phi}\right) - \frac{1}{2}\ln(2\pi y^3)}$.

c)

- Given $b(\theta) = 2 \times \theta^{\frac{1}{2}}, \frac{1}{\mu^2} = \theta \rightarrow b'(\theta) = \theta^{-\frac{1}{2}}$ and $b''(\theta) = -\frac{1}{2}\theta^{-\frac{3}{2}} = -\frac{1}{2}\mu^3$.
- Therefore
  - $E(Y) = b'(\theta) = \theta^{-\frac{1}{2}} = \mu$ (since $\frac{1}{\mu^2} = \theta$).
  - Variance function $= b''(\theta) = -\frac{1}{2}\theta^{-\frac{3}{2}} = -\frac{1}{2}\mu^3$.

**Question 5 (Binary Regression example from past exam)**

CBR bank is interested in predicting whether a customer will default on his or her credit card payment, based on the monthly credit card balance. Let $Y_i = 1$ denote the customer will default, and $Y_i = 0$ denote the customer will not default. Let $x$ denote each customer's monthly credit card balance (in **thousand dollars**). When $x > 0$, there is an outstanding credit card balance and the customer needs to repay the credit. When $x < 0$, the customer has an extra deposit in CBR bank. Given that $P(Y_i = 1|X_i = x) = \pi_i(x) = \mu_Y(X = x)$.

CBR bank decides to use a logistic (i.e., logit) model to fit the credit card balance and default data. The fitted logistic link function is $logit\left(\widehat{\pi_i(x)}\right) = -4 + 3.6x$.

a) A customer has an outstanding credit card balance of \$1300. Predict whether this customer will default or not.
b) A customer has an outstanding credit card balance of \$3000. Calculate the fitted odds ratio of this customer will not default. Predict whether this customer will default or not using this odds ratio.

**Solutions:**

a)

- Relevant concept: Workshop Week 7 Slides Binary regression prediction.
- **Method 1**
  - $Y_i|X_i = x \sim Bern(\pi_i(x))$.
  - Therefore, $x = 1.3, \pi_i(\widehat{x = 1.3}) = P(Y_i = \widehat{1|x} = 1.3) = \mu_Y(\widehat{x = 1}.3) = \frac{exp(-4+3.6\times1.3)}{1+exp(-4+3.6\times1.3)} \approx 0.6637$ (this is the probability that the customer will default)
  - Since this probability > 0.5, the customer will default.
- **Method 2**
  - $Y_i|X_i = x \sim Bern\left(\pi_i(x)\right)$
  - Therefore $x = 1.3$, we have fitted odds of **default**= $\frac{P(Y_i = \widehat{1|X_i} = x)}{P(Y_i = \widehat{0|X_i} = x)} =$ $\exp(-4 + 3.6x) = \exp(-4 + 3.6 \times 1.3) = \exp(0.68) \approx \mathbf{1.973878 > 1} \rightarrow$ $P(Y_i = \widehat{1|x} = 1.3) > P(Y_i = \widehat{0|x} = 1.3) \rightarrow Y_i$ is more likelky to be 1 (hence the customer will default).
- **Method 3**
  - $Y_i|X_i = x \sim Bern\left(\pi_i(x)\right)$
  - Therefore $x = 1.3$, we have
    - $\pi_i(\widehat{x = 1}.3) = P(Y_i = \widehat{1|x} = 1.3) = \frac{\exp(-4+3.6\times1.3)}{1+\exp(-4+3.6\times1.3)} \approx 0.6637$ (probability of default, a success)
    - $P(Y_i = \widehat{0|x} = 1.3) = 1 - P(Y_i = \widehat{1|x} = 1.3) = 1 - 0.6637 = 0.3363$ (probability of not default, a failure)
  - Therefore $P(Y_i = \widehat{1|x} = 1.3) > P(Y_i = \widehat{0|x} = 1.3) \rightarrow Y_i$ is more likelky to be 1 (hence the customer will default).

b)

- Relevant concept: Workshop Week 7 Slides Binary regression odds ratio and prediction.
- Fitted odds of **not default**= $\frac{\widehat{P(Y_i = 0|X_i = x)}}{\widehat{P(Y_i = 1|X_i = x)}} = \frac{1}{Fitted\ Odds\ of\ \boldsymbol{default}} = \frac{1}{\frac{\widehat{P(Y_i = 1|X_i = x)}}{\widehat{P(Y_i = 0|X_i = x)}}} = \frac{1}{\exp(-4+3.6x)} = \frac{1}{\exp(-4+3.6\times 3)} \approx \mathbf{0.0011}$.
- Hence, fitted Odds of **not default**= $\frac{\widehat{P(Y_i = 0|X_i = x)}}{\widehat{P(Y_i = 1|X_i = x)}} = \frac{1-\pi_i\widehat{(x=3)}}{\pi_i\widehat{(x=3)}} = 0.0011 < 1 \rightarrow 1 - \pi_i\widehat{(x = 3)} < \pi_i\widehat{(x = 3)} \rightarrow \pi_i\widehat{(x = 3)} > 0.5 \rightarrow \widehat{P(Y_i = 1|X_i = x)} > 0.5 \rightarrow \widehat{Y_i} = 1$.
  **This customer will default.**