1. (Poisson Regression) The independent random variables $Y_i, i = 1, 2, ..n$,represent the outcomes of a Poisson experiment where the mean μ_i is propotional to the value of x_i. That is, $Y_i \sim Poisson(\mu_i)$ and $\mu_i = \gamma x_i$, Assume that the x_i, values are known constants.

a) Find the MLE of γ

$$L(\gamma) = \prod_{i=1}^{n} (\frac{\mu_i^{y_i}}{y_i!} e^{-\mu_i}) = \prod_{i=1}^{n} \frac{(\gamma x_i)^{y_i} e^{-\gamma x_i}}{y_i!} = \frac{\gamma^{\sum_{i=1}^{n} y_i} \prod_{i=1}^{n} x_i^{y_i}}{\prod_{i=1}^{n} y_i!} e^{-\gamma \sum_{i=1}^{n} x_i}, \quad y_i \in 0, 1, 2..$$

$$l(\gamma) = \ln \gamma \sum_{i=1}^{n} y_i + \sum_{i=1}^{n} x_i^{y_i} - \sum_{i=1}^{n} \ln y_i! - \gamma \sum_{i=1}^{n} x_i$$

$$l'(\gamma) = \frac{\sum_{i=1}^{n} y_i}{\gamma} - \sum_{i=1}^{n} x_i \overset{\text{set}}{=} 0$$

$$\hat{\gamma}_{MLE} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i}$$

---

b) Find the mean and variance of $\hat{\gamma}_{MLE}$

For x_i are known constants. $Y_i \sim Poisson(\mu_i)$, $E[y_i] = Var[y_i] = \mu_i = \gamma x_i$,

$$E[\hat{\gamma}_{MLE}] = \frac{E[\sum_{i=1}^{n} y_i]}{\sum_{i=1}^{n} x_i} = \frac{\sum_{i=1}^{n} E[y_i]}{\sum_{i=1}^{n} x_i} = \frac{\sum_{i=1}^{n} \gamma x_i}{\sum_{i=1}^{n} x_i} = \gamma$$

For Y_i are independent random variables, $Cov(y_i, y_j) = 0, i \neq j$, $Var[\sum_{i=1}^{n} y_i] = \sum_{i=1}^{n} Var[y_i]$

$$Var[\hat{\gamma}_{MLE}] = \frac{Var[\sum_{i=1}^{n} y_i]}{(\sum_{i=1}^{n} x_i)^2} = \frac{\sum_{i=1}^{n} Var[y_i]}{(\sum_{i=1}^{n} x_i)^2} = \frac{\sum_{i=1}^{n} \gamma x_i}{(\sum_{i=1}^{n} x_i)^2} = \frac{\gamma}{\sum_{i=1}^{n} x_i}$$

---

2. Consider the regression model $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, $i = 1, .., n$. Find the maximum likelihood estimates of the paramiters if:

a)$\varepsilon_i \sim N(0, \sigma^2 x_i^2)$, independent for $i = 1, .., n$.

For $E[\varepsilon_i] = 0$, $Var[\varepsilon_i] = \sigma^2 x_i^2$, x_i and ε_i are independents,

$E[y_i] = E[\beta_0 + \beta_1 x_i] + E[\varepsilon_i] = \beta_0 + \beta_1 x_i$

$Var[y_i] = Var[\beta_0 + \beta_1 x_i] + Var[\varepsilon_i] = \sigma^2 x_i^2$

From Slutsky's theorem? C.L.T.?

$Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2 x_i^2)$

$f_Y(y_i) = \frac{1}{x_i \sqrt{2\pi\sigma^2}} e^{\frac{-1}{2\sigma^2 x_i^2}(y_i - \beta_0 - \beta_1 x_i)^2}$

$$L(\sigma) = \prod_{i=1}^{n} (x_i \sqrt{2\pi\sigma^2})^{-1} e^{\sum_{i=1}^{n} \frac{-1}{\sigma^2 x_i^2}(y_i - \beta_0 - \beta_1 x_i)^2} = (2\pi\sigma^2)^{-\frac{n}{2}} (\prod_{i=1}^{n} x_i)^{-1} e^{\sum_{i=1}^{n} \frac{-1}{\sigma^2 x_i^2}(y_i - \beta_0 - \beta_1 x_i)^2}$$

$$l(\sigma) = -n\ln\sigma^2 - \frac{n}{2}\ln(2\pi) - \ln(\prod_{i=1}^{n} x_i) - \sum_{i=1}^{n}\frac{1}{\sigma^2 x_i^2}(y_i - \beta_0 - \beta_1 x_i)^2$$

$$l'(\sigma) = -\frac{n}{\sigma^2} + \frac{1}{\sigma^4}\sum_{i=1}^{n}\frac{1}{x_i^2}(y_i - \beta_0 - \beta_1 x_i)^2 \stackrel{\text{set}}{=} 0$$

$$\hat{\sigma}_{MLE} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i^2}(y_i - \beta_0 - \beta_1 x_i)^2$$

---

b) $\varepsilon_i \sim i.i.d.\ f(\varepsilon;\lambda) = \frac{\lambda}{2}e^{-\lambda|x|}$.

$\varepsilon_i = y_i - \beta_0 - \beta_1 x_i +$

$f_Y(y_i) = \frac{\lambda}{2}e^{-\lambda|y_i - \beta_0 - \beta_1 x_i|}$

$$L(\lambda) = \prod_{i=1}^{n}(\frac{\lambda}{2}e^{-\lambda|y_i-\beta_0-\beta_1 x_i|}) = \lambda^n 2^{-n} e^{-\lambda\sum_{i=1}^{n}|y_i-\beta_0-\beta_1 x_i|}$$

$$l(\lambda) = n\ln\lambda - n\ln 2 - \lambda\sum_{i=1}^{n}|y_i - \beta_0 - \beta_1 x_i|$$

$$l'(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^{n}|y_i - \beta_0 - \beta_1 x_i| \stackrel{\text{set}}{=} 0$$

$$\hat{\lambda}_{MLE} = \frac{n}{\sum_{i=1}^{n}|y_i - \beta_0 - \beta_1 x_i|}$$

---

3. Finde the finite breakdown point and the infinite breakdown point for

a) the Mean Absolute Deviation, or $\frac{1}{n}\sum_{i=1}^{n}|X_i - \bar{X}_i|$.

The finite breakdown point is the smallest proportion m/n of the sample values such that $|\hat{\theta}^* - \hat{\theta}|$ can be made arbitarily large by corrupting m data values and computing $\hat{\theta}^*$, where n is the samle size, $\hat{\theta}$ is the estimator. The limit as $n \to \infty$ is called the breakdown point.

Replace X_i with X_i^*

$|\hat{\theta}^* - \hat{\theta}| = |\frac{1}{n}\sum_{i=1}^{n}|X_n^* - \bar{X}_i| - \frac{1}{n}\sum_{i=1}^{n}|X_n - \bar{X}_i|| = \frac{1}{n}|X_n^* - X_n| = \frac{1}{n}$

The breakdown point $= \frac{1}{n}$

$$\lim_{n\to\infty}|\hat{\theta}^* - \hat{\theta}| = \lim_{n\to\infty}\frac{1}{n} = 0$$

---

b) the Median Absolute Deviation, or Median$\{(X_1 - \bar{X}_i), .., (X_n - \bar{X}_i)\}$.

When n is even,

$|\hat{\theta}^* - \hat{\theta}| = |(X^*_{\frac{n}{2}} - \bar{X}_i) - (X_{\frac{n}{2}} - \bar{X}_i)| = |X^*_{\frac{n}{2}} - X_{\frac{n}{2}}|$

The breakdown point $= \frac{n/2}{n} = \frac{1}{2}$

$$\lim_{n\to\infty} |\hat{\theta}^* - \hat{\theta}| = \lim_{n\to\infty} (\frac{1}{2}) = \frac{1}{2}$$

When n is odd,

$|\hat{\theta}^* - \hat{\theta}| = |(X^*_{\frac{n+1}{2}} - \bar{X}_i) - (X_{\frac{n+1}{2}} - \bar{X}_i)| = |X^*_{\frac{n+1}{2}} - X_{\frac{n+1}{2}}|$

The breakdown point $= \frac{(n+1)/2}{n} = \frac{1}{2} + \frac{1}{2n}$

$$\lim_{n\to\infty} |\hat{\theta}^* - \hat{\theta}| = \lim_{n\to\infty} (\frac{1}{2} + \frac{1}{2n}) = \frac{1}{2}$$

---

4. Assume that $X_1, X_2, ..X_n$ are i.i.d. $Uniform(a, b)$. Find the asymptotic relative efficiency of the sample median to the sample mean.

For $X \sim Unif(a, b)$, $E[X] = \frac{a+b}{2}$, $Var[X] = \frac{(b-a)^2}{12}$, $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$,

$E[\bar{X}] = E[\frac{1}{n}\sum_{i=1}^{n} x_i] = \frac{1}{n}\sum_{i=1}^{n} E[x_i] = \frac{1}{n}\sum_{i=1}^{n} \frac{a+b}{2} = \frac{a+b}{2}$

For X_i are independent,

$Var[\bar{X}] = Var[\frac{1}{n}\sum_{i=1}^{n} x_i] = \frac{1}{n^2}\sum_{i=1}^{n} Var[x_i] = \frac{1}{n^2}\sum_{i=1}^{n} \frac{(b-a)^2}{12} = \frac{(b-a)^2}{12n}$

From Slutsky's theorem? C.L.T.?

$$\bar{X} \sim N(\frac{a+b}{2}, \frac{(b-a)^2}{12n})$$

For large n, the sample median $m_n \approx N(M, \frac{1}{4nf^2(M)})$, where M is the population median, $f(x)$ is the p.d.f. of X

$E[m_n] = M$

$Var[m_n] = \frac{1}{4nf^2(M)} = \frac{(b-a)^2}{4n}$

The asymptotic relative efficiency of m_n to $\bar{X}$

$$= \frac{Var[\bar{X}]}{Var[m_n]} = \frac{\frac{(b-a)^2}{12n}}{\frac{(b-a)^2}{4n}} = \frac{1}{3}$$

Therefore, the sample mean is asymptotic more efficiency than sample median.