STAT 563 - Final

Maggie Buffum

June 10, 2019

Problem 1

(Poisson Regression) The independent random variables Y_i, $i = 1, \dots, n$, represent the outcomes of a Poisson experiment where the mean μ_i is proportional 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.

Using $\mu_i = \gamma x_i$, we first find the likehood function:

$$L(\gamma) = \prod_{i=1}^{n} \left(\frac{e^{-x_i \gamma}(x_i \gamma)^{y_i}}{y_i!} \right) = e^{-\gamma \sum_{i=1}^{n} x_i} \prod_{i=1}^{n} \left(\frac{x_i^{y_i} \gamma^{y_i}}{y_i!} \right)$$

and the log-likelihood function:

$$l(\gamma) = \ln \left[e^{-\gamma \sum_{i=1}^{n} x_i} \prod_{i=1}^{n} \left(\frac{x_i^{y_i} \gamma^{y_i}}{y_i!} \right) \right] = -\gamma \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i \ln(x_i) + \ln(\gamma) \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} y_i!$$

We want to maximize the log-likelihood as a function of γ. The derivative of the log-likelihood with respect to γ is $-\sum_{i=1}^{n} x_i + \frac{1}{\gamma} \sum_{i=1}^{n} y_i$. Setting equal to 0 and solving for γ, we have

$$\frac{1}{\gamma} \sum_{i=1}^{n} y_i = \sum_{i=1}^{n} x_i \implies \hat{\gamma} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i} \implies \hat{\gamma} = \frac{\bar{y}}{\bar{x}}$$

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

We can derive the variance of the MLW directly and by noting that the variance of the Y_is is equal to their means, γx_i, and that the Y_is are independent:

$$Var(\hat{\gamma}_{MLE}) = Var\left(\frac{\bar{y}}{\bar{x}}\right) = \frac{1}{n^2 \bar{x}^2} Var\left(\sum_{i=1}^{n} y_i\right) = \frac{1}{n^2 \bar{x}^2} \sum_{i=1}^{n} Var(y_i) = \frac{\gamma}{n^2 \bar{x}^2} \sum_{i=1}^{n} x_i = \frac{\gamma}{n \bar{x}}$$

Problem 2

Consider the regression model $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$, $i = 1, \dots, n$. Find the maximum likelihood estimates of the parameters if the following hold.

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

We need to determine the likelihood function of β_0, β_1, and σ^2 given the Y_is. We know from the model that $Y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2 x_i^2)$. Therefore, the likelihood function is

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

and the log-likelihood function

$$\begin{aligned} l(\beta_0, \beta_1, \sigma^2 | Y_i) &= \ln\left[\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2 x_i^2}} e^{-\frac{1}{2\sigma^2 x_i^2}(y_i - \beta_0 - \beta_1 x_i)^2}\right] = \ln\left[\frac{1}{(\sqrt{2\pi\sigma^2})^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\frac{1}{x_i^2}(y_i - \beta_0 - \beta_1 x_i)^2} \prod_{i=1}^{n} \frac{1}{x_i}\right] \\ &= \ln\left(\frac{1}{(2\pi\sigma^2)^{n/2}}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 + \sum_{i=1}^{n} \ln\left(\frac{1}{x_i}\right) \end{aligned}$$

Taking the derivative with respect to β_0 and setting equal to 0, we have

$$\begin{aligned} \frac{dl(\beta_0)}{\beta_0} &= -2\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)(-1) = \text{ (set) } 0 \\ &\implies \frac{1}{\sigma^2}\sum_{i=1}^{n}\beta_0 = \frac{1}{\sigma^2}\sum_{i=1}^{n} y_i - \frac{1}{\sigma^2}\sum_{i=1}^{n}\beta_1 x_i \\ &\implies \hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} \end{aligned}$$

Similarly, we find the MLE of β_1:

$$\begin{aligned} \frac{dl(\beta_1)}{\beta_1} &= -2\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)(-x_i) = \text{ (set) } 0 \\ &\implies \beta_1\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i x_i - \beta_0\sum_{i=1}^{n} x_i \\ &\implies \beta_1\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i x_i - (\bar{y} - \beta_1\bar{x})\sum_{i=1}^{n} x_i \\ &\implies \beta_1\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i x_i - \bar{y}\sum_{i=1}^{n} x_i + \beta_1\bar{x}\sum_{i=1}^{n} x_i \\ &\implies \beta_1\sum_{i=1}^{n} x_i^2 - n\beta_1\bar{x}^2 = \sum_{i=1}^{n} y_i x_i - n\bar{y}\bar{x} \\ &\implies \hat{\beta}_1 = \frac{\sum_{i=1}^{n} y_i x_i - n\bar{y}\bar{x}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} \end{aligned}$$

We use the same method to derive the MLE of σ^2:

$$\begin{aligned} \frac{dl(\beta_1)}{\beta_1} &= -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 = \text{ (set) } 0 \\ &\implies \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 = \frac{n}{2\sigma^2} \\ &\implies \frac{n\sigma^2}{2} = \frac{1}{2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 \\ &\implies \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \end{aligned}$$

(b) $\epsilon_i \sim iid\ f(\epsilon;\lambda) = \frac{\lambda}{2}e^{-\lambda|\epsilon|}$.

From the model we know that $\epsilon_i \sim \text{LaPlace}(\mu = 0, b = \frac{1}{\lambda})$. Noting that $g^{-1}(y_i) = y_i - (\beta_0 + \beta_1 x_i)$ and its derivative is 1, we see that $Y_i \sim \text{LaPlace}(\mu_i = \beta_0 + \beta_1 x_i, b = \frac{1}{\lambda})$. Now we can find the likelihood function:

$$L(\beta_0, \beta_1, \lambda; \mathbf{y}) = \prod_{i=1}^{n} \frac{\lambda}{2} e^{-\lambda|y_i - (\beta_0 + \beta_1 x_i)|} = \left(\frac{\lambda}{2}\right)^n e^{-\lambda \sum |y_i - (\beta_0 + \beta_1 x_i)|}$$

and the log-likelihood function:

$$l(\beta_0, \beta_1, \lambda; \mathbf{y}) = \ln\left[\left(\frac{\lambda}{2}\right)^n e^{-\lambda \sum |y_i - (\beta_0 + \beta_1 x_i)|}\right] = n\ln(\lambda) - n\ln(2) - \lambda \sum_{i=1}^{n} |y_i - (\beta_0 + \beta_1 x_i)|$$

Consider the MLE of β_0 and β_1. We know that to maximize the log-likelihood as a function of either β_0 or β_1 requires us to minimize $\sum_{i=1}^{n} |y_i - (\beta_0 + \beta_1 x_i)|$. This is an example of Least Absolute Deviation (LAD) regression, the solutions for which have no closed form and must be evaluated numerically. However, given what we know about the absolute value function, namely that the solution for m to minimize the function $|y_i - m|$ is the median of y_i, we expect that the solution for $\beta_0 + \beta_1 x_i$ should be close to the median of y_i across the observed x_is. Methods such as the EM algorithm can be used to approximate the solutions. Let's call the solution $\hat{y}_{MLE}$.

Now let's estimate λ by setting the derivative of the log-likelihood function equal to zero:

$$\frac{dl}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} |y_i - (\beta_0 + \beta_1 x_i)| = \text{ (set) } 0 \implies \frac{1}{\hat{\lambda}_{MLE}} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_{MLE}|.$$

Problem 3

Find the finite breakdown point and the infinite breakdown point for the following.

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

Let $Y_i = |X_i - \bar{X}|$. Consider the case where we perturb one of the sampled X_is. We know that perturbing one sample point in a random sample is enough to change the sample mean, so we know that the new sample mean, $\bar{X}^*$, is different from the original sample mean, $\bar{X}$. Because each Y_i is a function of $\bar{X}$, pertubring one of the X_is actually perturbs every single sample point in the Mean Absolute Deviation. Since again, we already know that perturbing one sample point will change the mean, the finite breakdown point for the Mean Absolute Deviation is $\frac{1}{n}$, and the infinite breakdown point is $\lim_{n\to\infty} \frac{1}{n} = 0$.

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

Similar to part (a), we know that perturbing one of the samlped X_is is enough to change the sample mean. Therefore we know that it only takes one corrupt sample point to corrupt every element of the sequence across which we want to take the median. We already know that to corrupt the median, we need to perturb at least 50% of the sample points – clearly by corrupting one of the sampled X_is we've corrupted all i of the sampled $Y_i = |X_i - \bar{X}|$, and changed the sample median. Therefore, the finite breakdown point for the Median Absolute Deviation is also $\frac{1}{n}$, and the infinite breakdown point is $\lim_{n\to\infty} \frac{1}{n} = 0$.

Problem 4

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

Recall that the sample median (m) is distributed $N(M, 1/(4nf^2(\mu)))$ for large n, where M is the population median and μ is the population mean. Because the uniform distribution is constant across all X, $f(\mu) = 1/(b-a)$. Therefore, the relative efficiency of the sample median to the sample mean is

$$\frac{Var(\bar{X})}{Var(m)} = \frac{\sigma^2/n}{1/(4nf^2(\mu))} = \frac{\sigma^2 4nf^2(\mu)}{n} = \frac{(b-a)^2 4}{12(b-a)^2} = \frac{1}{3},$$

for large n. Clearly, for large n, the asymptotic efficiency is equal to the relative efficiency.
