Tutorial 7

STAT 3013/4027/8027

1. [based on Q 3.17]. Suppose $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{gamma}(a, b)$ where $E[X] = ab$. Suppose that the data on on gamma-rays (measuring the interarrival times of 3,935 photons (units in seconds) can be modeled by a gamma distribution.

   a. Make a histogram of the data. The data are on Wattle.

   b. Determine the method of moments estimators for a and b.

   c. Determine the MLEs for a and b. You will need to use the `digamma()` and `trigamma()` functions in `R`.

   d. Place the fitted densities (based on both estimators) on the histogram.

2. Consider a random sample of twins pairs. Twin pairs may be identical or fraternal. Let u of these pairs consist of male pairs, v consist of female pairs, and w consist of opposite sex pairs. A simple model for these data is based on a Bernoulli distribution for each pair dictating whether it consists of identical or fraternal twins. Suppose that identical twins occur with probability p and fraternal twins with probability $1 - p$. Once the decision is made as to whether the twins are identical or not, then sexes are are assigned to the twins. If the twins are identical, then one assignment of sex is made. If the twins are non-identical, then two independent assignments of sex are made. Suppose boys are chosen with probablity q and girls with probability $1 - q$.

   a. Write out the likelihood for these data.

   b. Formulate a "missing data" model and specify both steps of the EM algorithm to obtain the MLEs for p and q.

   c. Carry out the algorithm for $u = 22$, $v = 21$, and $w = 25$.

4.1

3. SI question 4.1 (a, b)

(a). $X_1,\dots,X_n \sim Pois(\theta)$

$H_0: \theta=\theta_0$

$H_1: \theta=\theta_1\ (\theta_1>\theta_0)$

$$L(\theta)=\prod_{i=1}^n e^{-\theta}\frac{\theta^{x_i}}{x_i!} = \frac{e^{n\theta}\,\theta^{\sum x_i}}{\prod_{i=1}^n x_i!}$$

$$\lambda = \frac{L(\theta_0|\vec{x})}{L(\theta_1|\vec{x})} = \frac{e^{n\theta_0}\,\theta_0^{\sum x_i}}{e^{n\theta_1}\,\theta_1^{\sum x_i}} = e^{n(\theta_1-\theta_0)}\left(\frac{\theta_0}{\theta_1}\right)^{\sum x_i}$$

∴ critical region is

$C: \{\lambda<k\}$ s.t. $P(C)=0.05$

$$=\{e^{n(\theta_1-\theta_0)}\left(\frac{\theta_0}{\theta_1}\right)^{\sum x_i}<k\}$$

$$=\{\underbrace{n(\theta_1-\theta_0)}_{\text{positive}}+(\sum x_i)\underbrace{\log\left(\frac{\theta_0}{\theta_1}\right)}_{\text{negative}}<\log k\}$$

$=\{\sum x_i>k^*\}$ where $k^*=\dfrac{\log k - n(\theta_1-\theta_0)}{\log(\theta_0/\theta_1)}$

$x_i \overset{iid}{\sim} Pois(\theta_0) \Rightarrow \sum x_i \overset{iid}{\sim} Pois(n\theta_0)$ by mgf

Check: we know that

$P(C)=P[\sum x_i>k^*]=0.05$

std. normal.

$\Rightarrow P[\bar X>k^{**}]=0.05$ then by CLT, we have $\Rightarrow P\left[\frac{\bar X-\theta_0}{\sqrt{\theta_0/n}}>k^{***}\right]=0.05$ $\quad k^{***}=1.64$

(b). $X_1,\dots,X_n\sim exp(\theta)$; $E(X)=\theta$

$H_0:\theta=\theta_0$, $H_1:\theta=\theta_1\,(>\theta_0)$

$$L(\theta)=\prod_{i=1}^n \frac{1}{\theta}e^{-x_i/\theta}=\frac{1}{\theta^n}e^{-\sum x_i/\theta}$$

$$\lambda=\frac{L(\theta_0|\vec x)}{L(\theta_1|\vec x)}=\frac{1/\theta_0^n\; e^{-\sum x_i/\theta_0}}{1/\theta_1^n\; e^{-\sum x_i/\theta_1}}=\left(\frac{\theta_1}{\theta_0}\right)^n e^{\sum x_i\left(\frac{1}{\theta_1}-\frac{1}{\theta_0}\right)}$$

critical region

$$C=\{\lambda<k\}=\{\left(\frac{\theta_1}{\theta_0}\right)^n e^{\sum x_i\left(\frac{1}{\theta_1}-\frac{1}{\theta_0}\right)}<k\}$$

$\boxed{\alpha=0.05=P(C)}$

$$=\{\underbrace{n\log\left(\frac{\theta_1}{\theta_0}\right)}_{+}+\underbrace{\sum x_i\left(\frac{1}{\theta_1}-\frac{1}{\theta_0}\right)}_{-}<\log k\}$$

$$=\{\sum x_i>k^*\}$$

Now $X_i\overset{iid}{\sim}Exp(\theta_0)\Rightarrow \sum x_i \overset{iid}{\sim} Gamma[n,\theta_0]$ by mgf

$\therefore P(C)=0.05$

$\Rightarrow P[\sum_{i=1}^{n} x_i > k^{*}]=0.05$

$\Rightarrow P[\overline{X_i} > k^{**}]=0.05$

$\because$ By CLT $\quad P[\frac{\overline{X}-\theta_0}{\sqrt{\theta_0/n}} = Z > k^{***}] = 005 \Rightarrow k^{***} = 1.64$