R computing for Business Data Analytics

Homework 4 (Due date: 23:59PM on December 11, 2016)

Please e-mail your homework (.pdf) and the associated *R* code (.*R*) to hchuang.om@gmail.com.

The email title must be R_HW4_GroupName. NO late homework will be accepted.

Q1. (20%) The probability density $P(X=k)$ of a random variable X ~ Poisson-Tweedie family (α, β, γ) can be calculated through a *double recursion algorithm*.

The first recursion exists in p_k:

$$p_0 = \begin{cases} e^{\beta[(1-\gamma)^\alpha - 1]/\alpha}, \alpha \neq 0 \\ (1-\gamma)^\beta \quad , \alpha = 0 \end{cases}, \; p_1 = \beta\gamma * p_0,$$

$$p_{k+1} = \frac{1}{k+1}\left(\beta\gamma * p_k + \sum_{j=1}^{k} j * r_{k+1-j} p_j\right), \; k=1,2,...$$

The second recursion exists in r_j:

$$r_1 = (1-\alpha)\gamma \;\; , \; r_{j+1} = \left(\frac{j-1+\alpha}{j+1}\right)\gamma * r_j, \; j=1,2,...$$

$$\text{where } \alpha \in (-\infty, 1], \; \beta \in (0, +\infty], \; \gamma \in [0,1]$$

Write a function *PTF* that has four inputs (k, a, b, g) (a for α, b for β, and g for γ) and returns p_k. Use the function to calculate PTF(9, -3, 2, 0.5) (The answer should be close to 0.04235).

Q2. (20%) MLE and Simulation

(a) For the Bernoulli distribution $P(X = x_i \mid p) = p^{x_i}(1-p)^{1-x_i}$, derive $\hat{p}_{MLE}$.

(Hint: Solve p for $\ell'(p) = 0$ where $\ell(p) = \log(\prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i})$)

(b) Continuing (a), given p=0.5, simulate n=50, n=5000, n=500,000 Bernoulli random numbers. For each simulated sample of size n, calculate $\hat{p}_{MLE}$ from the sample and compare $\hat{p}_{MLE}$ to the TRUE p=0.5, what have you observed?

(c) For the exponential distribution $f(x_i) = \lambda e^{-\lambda x_i}$, derive $\hat{\lambda}_{MLE}$ and prove the Markov Property:

$P(X > s+t \mid X > s) = P(X > t)$ (Hint: $F(x) = 1 - e^{-\lambda x}$)

(d) Continuing (c), given λ=0.5, simulate n=50, n=5000, n=500,000 exponential random numbers. For each simulated sample of size n, calculate $\hat{\lambda}_{MLE}$ from the sample and compare $\hat{\lambda}_{MLE}$ to the TRUE λ=0.5, what have you observed?

Q3. (20%) Binomial and Poisson Distributions

(a) The table below records the historical number of car accidents/week in a district. Create a vector called *car.accident* that stores 109 zeros, 65 ones, 22 twos, 3 threes, and 1 four.

(b) Apply the *fitdistr*( ) function in *R* (load the *MASS* library first) to fit the *car.accident* data with the Poisson distribution. What is the value of estimated λ? What is the log-likelihood?

(c) Given the estimated λ, use *R* to do the computation and finish the 3rd column of table below. Are the predicted frequencies close to the actual frequency?

(d) Given the estimated λ, finish the 4th column of table below using *R*.

Car Accidents	Frequency	Poisson(λ=???)	Binomial(n=200, p= λ/200)
0	109	200*$P(X=0\|\lambda)$=	200*$P(X=0\|n, p)$=
1	65	200*$P(X=1\|\lambda)$=	200*$P(X=1\|n, p)$=
2	22	200*$P(X=2\|\lambda)$=	200*$P(X=2\|n, p)$=
3	3	200*$P(X=3\|\lambda)$=	200*$P(X=3\|n, p)$=
4	1	200*$P(X=4\|\lambda)$=	200*$P(X=4\|n, p)$=
>4	0	200*$P(X>4\|\lambda)$=	200*$P(X>4\|n, p)$=

Q4. (20%) Binomial, Poisson, and Normal Distributions

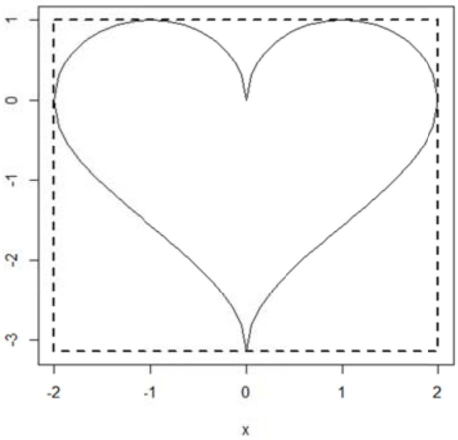
(a) Finish the second, third, and fourth column of the table below using *R*.

(b) Based on the probability densities you calculate, generate a plot in which the x-axis is 0:8 and the y-axis lies between [0, 0.25]. The plot should have three lines with *different* width and colors. (red thin line for Binomial, green thick line for Poisson, blue thicker line for Normal).

Defectives (x)	Binomial(n=20, p=0.2) P(X=x)	Poisson(λ=4) P(X=x)	Normal ($\mu = 4, \sigma^2 = 4$) P(x-0.5<X<x+0.5)
0			
1			
2			
3			
4			
5			
6			
7			
8			

Q5. (20%) Two mathematical functions $\sqrt{1-(|x|-1)^2}$ and $\text{acos}(1-|x|)-\pi$ can be used to draw a beautiful heart with the following *R* code.

```
> heart_up = function(x) {sqrt(1-(abs(x)-1)^2)}
> heart_lo = function(x) {acos(1-abs(x))-pi}
> x=seq(-2,2,0.05)
> plot(x, heart_lo(x), ylim=c(heart_lo(0),1), type= 'l')
> lines(x, heart_up(x))
```



The **exact area** of this heart can be obtained through *integration*. Use functions for integration in *R* and show me what the area is?