R computing for Business Data Analytics

Homework 3 (Due date: November 14, 2014)

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

The email title must be R_HW3_GroupName. NO late homework will be accepted.

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

The first recursion exists in p_k:

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

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

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%) Random Numbers and Monte-Carlo Simulation

(a) Implement the algorithm in the bottom of page 10 in lecture 6. Write a function *runi.congru* that has five arguments (N, A, B, m, seed) (N is the number of random variates to be simulated). After that, generate five uniform random numbers using A=1217, B=0, m=32767, and seed=1. Save the five numbers as a vector u in *R* and show me the five numbers.

(b) Implement the *inverse transformation* method in the bottom of page 11 in lecture 6. Write a function *rbinom.invtran* that has four arguments (N, n, p, uni) and returns N binomial random numbers. Set n=3, p=0.5, and use the vector u in part (a) as inputs to uni to generate five binom random variates. Show me the simulated numbers too.

(c) Now, use the function *runi.congru* in (a) to simulate another 50 numbers by setting seed=2 (A=1217, B=0, m=32767 still) and store the 50 numbers in a vector U.

(d) For the zero-truncated Poission distribution

$$P(X = x \mid \lambda, x > 0) = \frac{\lambda^x e^{-\lambda}}{x!(1-e^{-\lambda})}$$

Following the logic of inverse transformation, write a function *rztpois.invtran* that has three arguments (N, lambda, uni) (lambda for λ) and returns N zero-truncated Poisson random numbers. Set lambda=4 and use the vector U (part (c)) as inputs to uni to generate 50 non-zero Poisson random variates. Show me the simulated numbers too.

(e) Assume the simulated numbers $(x_1, x_2, ..., x_{50})$ are *counts of arrived customers per group* for a restaurant. Let's say all $x_i \le 2$ will be assigned to two-people tables, all $2 < x_i \le 4$ will be assigned to four-people tables, $4 < x_i \le 6$ will be assigned to six-people tables, and finally, all $6 < x_i$ will be put into private rooms.

Now, based on the simulated 50 groups $(x_1, x_2, ..., x_{50})$, how many groups of customers will be seated in two-people tables, four-people tables, six-people tables, and private rooms respectively?