CSCI E-106: Data Modeling

Assignment 0

Due: January, 28 2019 at 7pm EST

Instructions: Students should submit their reports on Canvas. The report needs to contain copies of the questions, step-by-step walk-through solutions, and final answers clearly indicated. For the questions 1 to 7, please solve it by hand; for questions 8 to 10, please submit two files: (1) a R Markdown file (.Rmd extension) and (2) a PDF document generated using knitr.

For the questions 1 to 3, calculate the mean and variance for each of the following probability distributions

1. $f_X(x) = ax^{a-1},\ \ 0 < x < 1,\ \ a > 0$
2. $f_X(x) = \frac{1}{n},\ \ x = 1, \cdots, n, \quad n > 0\ an\ integer$
3. $f_X(x) = \frac{3}{2}(x-1)^2\ \ ,\ \ 0 < x < 2$

Questions for Joint Distribution:

4. Let X_1 and X_2 be independent standard normal distribution random variables $N(0,1)$. Find the PDF of $\frac{(X_1-X_2)^2}{2}$
5. Let X_1 and X_2 be independent gamma distribution random variables with $gamma(\alpha_1,1)$ and $gamma(\alpha_2,1)$. Find the marginal distributions of $\frac{X_1}{X_1+X_2}$ and $\frac{X_2}{X_1+X_2}$

For the following questions, find the Maximum Likelihood Estimation (MLE).

6. Let $X_1,\ldots, X_n$ be a random sample from a gamma(α,β) population. Find the MLE of β assuming α is known.

7. Let $X_1,\ldots X_n$ be a random sample from the PDF. Find the MLE of θ:

$$f(x|\theta) = \theta X^{-2}\ ,\ \ 0 < \theta \le x < \infty$$

R Programming:

For the following questions, let matrix $X = \begin{bmatrix} 10 & 1 & 9 \\ 3 & 8 & 7 \\ 5 & 2 & 4 \end{bmatrix}$ $Y = \begin{bmatrix} 2 & 8 & 3 \\ 5 & 1 & 12 \\ 13 & 4 & 7 \end{bmatrix}$

(Please note that A^T is a transpose of a matrix A, a matrix such that $[A^T]_{ij} = A_{ji}$)

8. For matrix *X and Y* above, calculate $(X + Y)$

9. For matrix X *and Y* above, calculate $(X^TX)^{-1}X^TY$

10. Write R code to draw 10,000 random samples from uniform distribution and calculate the 99% percentile.

Note: μ = mean σ^2 = variance

1. $f_X(x) = a x^{a-1}, \quad 0 < x < 1, \quad a > 0$

$$\mu = E(f_X(x)) = \int x f_X(x)\, dx$$

$$= \int_0^1 x \cdot a x^{a-1}\, dx = a \int_0^1 x^{\cancel{a+1}}\, dx = a \frac{x^{a+1}}{a+1} \Bigg|_{x=0}^{1}$$

$$\boxed{\mu = \boxed{\frac{a}{a+1}}} \quad -①$$

$$\sigma^2 = E[(X-\mu)^2] = E(x^2) - [E(x)]^2$$

$$\int_0^1 x^2 \cdot a x^{a-1}\, dx$$

$$= a\int_0^1 x^{a+1}\, dx = a \frac{x^{a+2}}{a+2}\Bigg|_{x=0}^{1}$$

$$= a\frac{1^{a+2}}{a+2} - a\frac{0^{a+2}}{a+2}$$

$$= \frac{a}{a+2}$$

$$E(x) = \frac{a}{a+1}$$

$$[E(x)]^2 = \left(\frac{a}{a+1}\right)^2$$

$$= \frac{a^2}{a^2+2a+1}$$

$$\boxed{\therefore \sigma^2 = E(x^2) - [E(x)]^2 = \frac{a}{a+2} - \frac{a^2}{a^2+2a+1}}$$

(2.) $f_X(x) = \frac{1}{n}$, $x = 1, \cdots, n$, $n > 0$ integer

$\mu = E(f_X(x)) = \int x \cdot f_X(x)\,dx$

$\mu = \int_1^n x \cdot \frac{1}{n}\,dx$

$= \frac{1}{n}\int_1^n x\,dx = \frac{1}{n}\left(\frac{x^2}{2}\right)_{x=1}^{n}$

$= \frac{1}{n}\left(\frac{n^2}{2} - \frac{1^2}{2}\right) = \frac{n}{2} - \frac{1}{2n}$

$\boxed{\mu = \frac{n}{2} - \frac{1}{2n}}$

$\sigma^2 = E(x^2) - [E(x)]^2$

$E(x) = \mu = \frac{n}{2} - \frac{1}{2n}$

$[E(x)]^2 = \left(\frac{n}{2} - \frac{1}{2n}\right)^2$

$E(x^2) = \int_1^n x^2 \cdot \frac{1}{n}\,dx$

$= \frac{1}{n}\int_1^n x^2\,dx = \frac{1}{n}\left(\frac{x^3}{3}\right)_{x=1}^{n}$

$= \frac{1}{n}\left(\frac{n^3}{3} - \frac{1}{3}\right)$

$= \frac{n^2}{3} - \frac{1}{3n}$

$\boxed{\therefore \sigma^2 = \frac{n^2}{3} - \frac{1}{3n} - \left(\frac{n}{2} - \frac{1}{2n}\right)^2}$

③ $f_X(x) = \frac{3}{2}(x-1)^2 \quad , 0 < x < 2$

$\mu = E[f_X(x)] = \int_0^2 x \cdot f_X(x)\,dx$

$\mu = \int_0^2 x \frac{3}{2}(x-1)^2 dx = \frac{3}{2}\int_0^2 x(x^2-2x+1)\,dx$

$= \frac{3}{2}\int_0^2 x^3 - 2x^2 + x \quad dx$

$= \frac{3}{2}\left(\frac{x^4}{4} - \frac{2x^3}{3} + \frac{x^2}{2}\right)_{x=0}^{2}$

$= \frac{3}{2}\left(\frac{2^4}{4} - \frac{2(2^3)}{3} + \frac{2^2}{2}\right)$

$\boxed{\mu = 1}$

$\sigma^2 = E[(x-\mu)^2] = E(x^2) - [E(x)]^2$

$[E(x)]^2 = \mu^2 = 1$

$E(x^2) = \int_0^2 x^2 \frac{3}{2}(x-1)^2 dx$

$\frac{3}{2}\int_0^2 x^2(x^2-2x+1)\,dx$

$= \frac{3}{2}\int_0^2 x^4 - 2x^3 + x^2 \; dx$

$= \frac{3}{2}\left(\frac{x^5}{5} - \frac{2x^4}{4} + \frac{x^3}{3}\right)_{x=0}^{2}$

$= \frac{3}{2}\left(\frac{2^5}{5} - \frac{2(2^4)}{4} + \frac{2^3}{3}\right)$

$= 6.83$

$\sigma^2 = 6.83 - 1$

$\boxed{\sigma^2 = 5.83}$

Scanned by CamScanner

4. $X_1, X_2 \sim N(0,1)$. Find PDF of $\frac{(x_1-x_2)^2}{2}$

$$f(x_1) = f(x_2) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$$

$$f(x_1, x_2) = \frac{(x_1-x_2)^2}{2} = \frac{\left(\frac{e^{-x_1^2/2} - e^{-x_2^2/2}}{\sqrt{2\pi}}\right)^2}{2}$$

$$= \frac{\frac{e^{-x_1^2} - e^{-x_2^2}}{\cancel{2}\pi^2 4}}{\cancel{2}}$$

PDF: $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{e^{-x_1^2} - e^{-x_2^2}}{4\pi^2} \, dx_1 \, dx_2$

$$= \frac{1}{4\pi^2}\int_{-\infty}^{\infty} \sqrt{\pi} - x_1 e^{-x_2^2} \, dx_2$$

$$= \frac{1}{4\pi^2}\left(\sqrt{\pi}\, X_2 - X_1\sqrt{\pi}\right) = \frac{x_2}{4\pi^{3/2}} - \frac{x_1}{4\pi^{3/2}}$$

oops... $\cancel{\boxed{u = \frac{x_2 - x_1}{4\pi^{3/2}}}}$

$$\boxed{\text{PDF}: \frac{x_2 - x_1}{4\pi^{3/2}}}$$

⑤. Let $X_1 = \Gamma(\alpha_1, 1)$, $X_2 = \Gamma(\alpha_2, 2)$

Find marginal distributions of $\frac{X_1}{X_1+X_2}$ and $\frac{X_2}{X_1+X_2}$

$$X_1 = \frac{1}{\Gamma(\alpha)} X_1^{\alpha-1} e^{-X_1}$$

$$X_2 = \frac{1}{\Gamma(\alpha)} X_2^{\alpha-1} e^{-X_2}$$

$$\frac{X_1}{X_1+X_2} = \boxed{\frac{\frac{1}{\Gamma(\alpha)} X_1^{\alpha-1} e^{-X_1}}{\frac{1}{\Gamma(\alpha)} X_1^{\alpha-1} e^{-X_1} + \frac{1}{\Gamma(\alpha)} X_2^{\alpha-1} e^{-X_2}}}$$

m.d.f. $\frac{X_1}{X_1+X_2} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \quad dX_1\, dX_2$

vice versa for $\frac{X_2}{X_1+X_2}$?

Not sure how to continue, and how I began ...

(6.) $X_1, \cdots, X_n$ random sample from $\Gamma(a, \beta)$ population.

Find β assuming α is known

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x_i^{\alpha-1} e^{-x_i/\beta}$$

$$L(\beta|x) = (x_1, \cdots x_n)^{n-1} \left(\frac{1}{\Gamma(\alpha)\beta^{\alpha}}\right)^n e^{\frac{-\sum_{i=1}^n x_i}{\beta}}$$

$$\log L(\beta|x) = n(-\alpha)\log\Gamma(\alpha) - n(a)\log(\beta) + (a-1)\sum_{i=1}^n \log(x_i) - \frac{\sum_{i=1}^n x_i}{\beta}$$

$$\frac{\partial \log L(\beta|x)}{\partial \beta} = \frac{-n\alpha}{\beta} + \frac{\sum_{i=1}^n x_i}{\beta^2} = 0$$

$$\hat{\beta} = \frac{\sum_{i=1}^n x_i/n}{\alpha} = \boxed{\frac{\bar{x}}{\alpha}}$$

Note: MLE of $\beta = \hat{\beta}$

(7.) $f(x|\theta) = \theta x^{-2}$, $0 < \theta \leq x < \infty$, $x_1, \cdots x_n$

Find MLE θ $(\hat{\theta})$

$f(x|\theta) = \theta^n (x_i, \cdots x_n)^{-2}$

$L(x|\theta) = \theta^n (x_i, \cdots x_n)^{-2}$

Given $0 < \theta \leq x < \infty$ constraint, $L(x|\theta)$ is greatest when n is at smallest value (as x^n sits in denominator of $f(x|\theta)$).

$\boxed{\therefore \hat{\theta} = x_{n_{min}} = x_1}$

YK_Assignment0

Yinan Kang

1/29/2019

R Markdown

```r
X <- matrix(c(10,1,9,3,8,7,5,2,4), nrow=3, ncol=3, byrow = TRUE)

Y <- matrix(c(2,8,3,5,1,12,13,4,7), nrow=3, ncol=3, byrow = TRUE)

print(X)
```

```
##      [,1] [,2] [,3]
## [1,]   10    1    9
## [2,]    3    8    7
## [3,]    5    2    4
```

```r
print(Y)
```

```
##      [,1] [,2] [,3]
## [1,]    2    8    3
## [2,]    5    1   12
## [3,]   13    4    7
```

Problem 8

```r
print(X+Y)
```

```
##      [,1] [,2] [,3]
## [1,]   12    9   12
## [2,]    8    9   19
## [3,]   18    6   11
```

Problem 9

```r
print(solve(t(X)*X)*t(X)*Y)
```

```
##             [,1]        [,2]       [,3]
## [1,] -0.38365304 -0.32360300  0.9862385
## [2,] -0.06741729  0.07876934  0.7033639
## [3,]  7.69266055  0.82059123 -4.1457696
```

Problem 10

```r
sample <- runif(10000)
quantile(sample, .99)
```

```
##       99%
## 0.9905336
```

The 99th percentile in this uniform distribution of random variables is 0.9905336