NATIONAL UNIVERSITY OF SINGAPORE
Department of Statistics and Applied Probability

2018/19 Semester 2 **ST2137 Computer Aided Data Analysis** **Tutorial 9**

Note: This tutorial will be discussed in the lecture on 18 April 2019

1. Write an R function to compute the cdf (i.e. $F_X(x) = \Pr(X \le x)$) of the t-distribution with n degrees of freedom, which has a pdf.

$$f_X(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \frac{1}{\sqrt{n\pi}} \frac{1}{\left(1+\frac{x^2}{n}\right)^{(n+1)/2}}, \quad \text{for } -\infty < x < \infty$$

$$F_X(x) = \int_{-\infty}^{x} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \frac{1}{\sqrt{n\pi}} \frac{1}{\left(1+\frac{t^2}{n}\right)^{(n+1)/2}}\, dt, \quad \text{for } -\infty < x < \infty$$

   (a) Suppose $n = 5$. What is the cdf when $x = 2.5$?
   (b) Compare your result in part (a) to the result from the R function "pt" with $n = 4$.
   [You may use the built-in R function "gamma(k)" to compute $\Gamma(k)$ and "pi" to get the value for π.]

. Suppose $X_1, X_2, \cdots, X_n$ are independent and identically distributed random variables from a Beta distribution with parameters α and β. The pdf of Beta(α, β) is given by

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)+\Gamma(\beta)}\, x^{\alpha-1}(1-x)^{\beta-1}, \quad \text{for } 0 < x < 1$$

It can be shown that the log-likelihood function is given by

$$\log L(\alpha, \beta | x_1, \cdots, x_n) = n(\log\Gamma(\alpha+\beta) - \log\Gamma(\alpha) - \log\Gamma(\beta)) + (\alpha-1)\sum_{i=1}^{n} \log x_i + (\beta-1)\sum_{i=1}^{n} \log(1-x_i)$$

A random sample of 30 observations from a beta distribution is recorded in the file "beta30.txt". Find the MLE of α and β using the 2-dimensional optimization R function "optim".

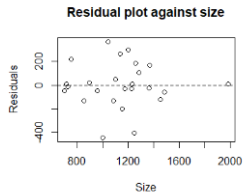
3. An agent for a residential real estate company in a large city would like to be able to predict the monthly rental cost for apartments, based on the size of the apartment as defined by square footage. A sample of 25 apartments in a particular residential neighbourhood was selected. The information gathered is given in the data file "rent.txt" (free format with space as delimiter).

   Use SAS, R and SPSS to do parts (a) to (f).
   (a) Fit a simple linear regression model $y = \beta_0 + \beta_1 x + \epsilon$ to the data, where y and x are the monthly rental cost and the size of the apartment respectively.
   (b) At the 0.05 level of significance, is there evidence of a linear relationship between the size of the apartment and the monthly rent?
   (c) Determine the coefficient of determination R^2.
   (d) Plot the residual against the fitted value. Comment on the plot.
   (e) Check the normality assumption.
   (f) Predict the monthly rental cost for an apartment that has 1,000 square feet.

(g) Your friends Jim and Jennifer are considering signing a lease for an apartment in this residential neighborhood. They are trying to decide between two apartments, one with 1,000 square feet, for a monthly rent of $1,275, and the other with 1,200 square feet, for a monthly rent of $1,425. What would you recommend to them? Why?

Answers/hints to selected questions

1. 0.9030482 (R code: p13.8)
2. $\hat{\alpha} = 2.22095, \hat{\beta} = 3.58868$ (Your answer may be slightly different from these answers.) (R code: p13.22-13.23)
3. (a) $\widehat{Rent} = 177.12 + 1.0651size$
   (b) Yes. p-value $= 7.5(10)^{-8}$, hence reject H_0: $\beta_1 = 0$
   (c) $R^2 = 0.7226$
   (d) No obvious pattern



   (e) No evidence against normality assumption, KS-test with p-value = 0.6493
   (f) $1242.265
   (g) 1200 sq ft apartment as estimate of $E(Rent|size = 1000) = 1242.265 < 1275$ and estimate of $E(Rent|size = 1200) = 1455.294 > 1425$.

   SAS code (Refer to p14.22, p14.30, p14.38-39)
   R code (Refer to p14.25-26, p14.32-33, p14.38-39)
   SPSS (Refer to p14.27-28, p14.34-37, p14.38-39)