**NATIONAL UNIVERSITY OF SINGAPORE**
**Department of Statistics and Applied Probability**

**2018/19 Semester 2** **ST2137 Computer Aided Data Analysis** **Tutorial 8**

1. The probability density function of Pareto distribution is given as follows.

$$f(x) = \frac{\alpha\beta^{\alpha}}{x^{\alpha+1}}, \qquad x > \beta, \qquad \alpha, \beta > 0$$

It can be shown that the distribution function of Pareto distribution is given by

$$F(x) = 1 - \left(\frac{\beta}{x}\right)^{\alpha}, \qquad x > \beta, \qquad \alpha, \beta > 0$$

(a) Using the inversion method to generate a random sample of size 30 from the Pareto distribution with $\alpha = 4$ and $\beta = 2$ using R.
(b) Repeat part (a) using SAS.

2. Suppose $X_1$ and $X_2$ follow Chi-square distributions with degrees of freedom $2\alpha$ and $2\beta$ respectively. Let $Y = \frac{X_1}{X_1+X_2}$. Then $Y$ follows a Beta distribution with parameters $\alpha$ and $\beta$. Use R to do parts (a) to (c).
(a) Generate a sample of size 1000 from $\chi^2(6)$ distribution and another sample of size 1000 from $\chi^2(10)$ distribution.
(b) Generate a sample of size 1000 from $Beta(3,5)$ using the two random samples obtained in part (a).
(c) Construct a histogram for the data from part (b) with the true density superimposed on it. Comment on the quality of the generated random numbers.
[Note: The p.d.f. of $Beta(\alpha,\beta)$ is given by

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

"dbeta($x$,α,β)" in R can be used to calculate $f(x)$.]
(d) Repeat parts (a) to (c) using SAS.

3. Suppose that $X_1, X_2, \cdots, X_{25}$ is a random sample from $N(\mu, 2^2)$. We want to test H$_0$: $\mu = 0$ against H$_1$: $\mu \neq 0$. Consider test statistic $T = \frac{\bar{X}-0}{S/\sqrt{n}}$, where $n$, $\bar{X}$ and $S$ are respectively the sample size, the sample mean and sample standard deviation, is used. H$_0$ will be rejected if $|t| > t_{0.025;\, n-1}$.
(a) Use R and SAS to do a simulation study to find the power of the test with $n = 9$ when $\mu = -1$, $\mu = -0.5$, $\mu = 0.5$ and $\mu = 1$. Comment on the results.
(b) Repeat part (a) for with $n = 25$.
(c) Comment on the results of parts (a) and (b)
[Note: Power of a test when $\mu = \mu_1$ is given by Pr(Rejecting H$_0$ when $\mu = \mu_1$) = $\Pr\left(\left|\frac{\bar{X}}{S/\sqrt{n}}\right| > t_{0.025;n-1} \middle| \mu = \mu_1\right)$ ]

4. A random sample of 50 pieces of real estate sold in Seattle during 2002, as recorded by the county assessor is selected. The selling prices of these pieces of real estate were recorded in the file "price.txt". The distribution of the selling price is far from normal. The 25% trimmed mean is used to estimate the centre of the distribution.

Selling prices for Seattle real estate, 2002 ($1000s)

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 142 | 175 | 197.5 | 149.4 | 705 | 232 | 50 | 146.5 | 155 | 1850 |
| 132.5 | 215 | 116.7 | 244.9 | 290 | 200 | 260 | 449.9 | 66.407 | 164.95 |
| 362 | 307 | 266 | 166 | 375 | 244.95 | 210.95 | 265 | 296 | 335 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 335 | 1370 | 256 | 148.5 | 987.5 | 324.5 | 215.5 | 684.5 | 270 | 330 |
| 222 | 179.8 | 257 | 252.95 | 149.95 | 225 | 217 | 570 | 507 | 190 |

(a) Find a bootstrap estimate of the bias of the 25% trimmed mean
(b) Find a bootstrap estimate of the standard error of the 25% trimmed mean
(c) Construct a 95% basic bootstrap confidence interval for the 25% trimmed mean
(d) Obtain a 95% confidence interval for the 25% trimmed mean using the bootstrap percentile method.

(You may use the functions in R to get the answers)

Answers/hints to selected questions

1. Let $U = F_X(x) = 1 - \left(\frac{\beta}{x}\right)^\alpha$. Then $X = F^{-1}(U) = \frac{\beta}{(1-U)^{1/\alpha}}$.

   Part of R code:
```
rpareto <- function(n=1,a=1,b=2){
u <- runif(n); x <- b/(1-u)^(1/a); return(x)}
rpareto(30,4,2)
```
   Part of SAS code:
```
data paretoran;
  a=4; b=2;
do i=1 to 30;
  u=rand("uniform"); x=b/u**(1/a); output;
end;
keep x;
proc print data=paretoran;
```
2. (a) & (b) Part of R code: `x1 <- rchisq(1000,6); x2 <- rchisq(1000,10); y <- x1/(x1+x2)`

   (c) `hist(y, freq=T, breaks=seq(0,1,0.05));`

   (d) Part of the SAS code: `do i = 1 to 1000; x1=rand("chisquare",2*a); x2=rand("chisquare",2*b); y=x1/(x1+x2); output; end; proc univariate; histogram y/midpoints=0 to 1 by 0.05 beta(theta=0, sigma=1, alpha=3, beta=5); run;`

   (For specifying the parameters of beta distribution in SAS, refer to the following website. http://support.sas.com/documentation/cdl/en/procstat/67528/HTML/default/viewer.htm#procstat_univariate_details62.htm

3. (a) Power of $t$-test for $\mu_1 = -1, -0.5, 0.5, 1$ are 0.2627, 0.1018, 0.1018, 0.2627 respectively. (R code: p11.25-11.26. SAS code: p11.27-11.30.)

   (b) Power of $t$-test for $\mu_1 = -1, -0.5, 0.5, 1$ are 0.6697, 0.2245, 0.2245, 0.6697 respectively. (The above are theoretical answers. Your answers from simulations should be close to these figures.)

   (c) For a fixed $\mu_1$ in $H_1$, the power increases as $n$ increases.

4. (a) ~ 0.51; (b) ~ 17.57; (c) ~ (207, 275); (d) ~ (213, 281) (Your answers may be different from these approximate answers, especially part (a).)
```
library(boot)
tm25 <- function(x,b){return(mean(x[b],0.25))}
boot.tm25 <- boot(price, statistic=tm25, R=10000); boot.tm25
boot.ci(boot.tm25,type=c("basic","perc"))
```