MTH5126 Statistics for Insurance 2022-23

Assessed Coursework - R-based

Instructions:

1. Attempt and submit all questions on QMPlus (see QMPlus for deadline).
2. Put your student ID at the top of your answer script.
3. Include the R code, that you have used to generate your solutions, together with the main R output produced, in your answer script (i.e., ***PDF document***). Where the R code is missing from a particular question part, no marks will be awarded even if the output (e.g., a graph) is included.
4. Outputs (e.g., graphs) should be labelled where appropriate. Appropriate commentary should be provided alongside the R code and R output in the answer script, where relevant, to fully demonstrate your understanding.
5. You should turn your answer script into a pdf document before you upload it on QMPlus.

Questions

1. Before answering this question, construct the function, rpareto, in R using the following code:

```
rpareto = function(n, alpha, lambda){
rp = lambda * ((1 - runif(n))^(-1/alpha) - 1)
rp}
```

   The *rpareto* function generates a random sample of size n from a two-parameter Pareto distribution with parameters α and λ.

   (i) Generate, using *rpareto*, a random sample of size 25,000 from the two-parameter Pareto distribution with parameters $\alpha = 3$ and $\lambda = 1$, assigning the simulated values to a vector called *A_vec*. You should set a random number generator seed of NNN

before generating *A_vec*, where NNN is the last three digits of your student ID. Use the R function, head(), to display the first five values of *A_vec* in your answer script. **[20]**

R codes:

(with NNN=123 as an example, thereafter the same. You can use NNN=123 to check if your codes can be correctly run.)

```
set.seed (123)
A_vec =
rpareto(n = 25000, alpha = 3, lambda = 1)
head(A_vec, 5)
```

Output:

0.11966335 0.67788900 0.19160373 1.04468423 1.56103061

(ii) Construct an R function, called *A_exceed_u*, with two arguments, A and u, that returns only the non-zero entries of a vector, E, where vector E is of length 25,000, with entries defined as:

$$E_i = \max\{A_i - u, 0\}$$

Use the R function, head(), to display in your answer script the first five values of *A_exceed_u* when $A = A\_vec$ and $u = 1$. **[20]**

R codes:

```
A_exceed_u =
function(A, u){
E = pmax(A - u, 0)
output = E[E!=0]
output}
head(A_exceed_u(A = A_vec, u = 1), 5)
```

Output:

0.04468423 0.56103061 0.10259132 0.85069357 0.15317919

(iii) Construct another R function, called *F_u*, with one argument, *A_greater_than_u*, that returns a vector of length 101, containing the probabilities:

$$P(A_i – u \leqslant x | A_i > u), \text{where } x \geqslant 0,$$

for values of x from 0 to 10 inclusive, at intervals of 0.1.

Use the R function, head(), to display in your answer script the first five values of

F_u when $A\_greater\_than\_u = A\_exceed\_u$ with arguments $A = A\_vec$ and $u = 1$. **[20]**

R codes:

$F\_u =$

function($A\_greater\_than\_u$) {

y = vector(length = 101)

for (i in 1:101) {

y[i] =

length($A\_greater\_than\_u$[$A\_greater\_than\_u$ <= 0.1 * (i-1)]) /

length($A\_greater\_than\_u$)

}

y}

head(F_u(A_exceed_u($A = A\_vec$, u = 1)), 5)

Output:

0.0000000 0.1463087 0.2526846 0.3439597 0.4201342

(iv) Plot, on a single graph, four line graphs of the values of F_u against x for the values of x specified in part (iii) when $A\_greater\_than\_u = A\_exceed\_u$, $A = A\_vec$ and u = 1, 2, 3 and 4. You should use separate colours to identify each line graph and you should clearly specify which value of u is represented by each of the four line graphs. **[20]**

R codes:

x = seq(from = 0, to = 10, by = 0.1)

plot(

x,

F_u($A\_greater\_than\_u = A\_exceed\_u$(A = A_vec, u = 1)),

type = "l",

main = "Values of F_u against x for u = 1, 2, 3 and 4",

col = "red",

ylab = "F_u")

lines(x, F_u($A\_greater\_than\_u = A\_exceed\_u$(A = A_vec, u = 2), col = "yellow")

lines(x, F_u($A\_greater\_than\_u = A\_exceed\_u$(A = A_vec, u = 3), col = "blue")

lines(x, F_u($A\_greater\_than\_u = A\_exceed\_u$(A = A_vec, u = 4), col = "green")

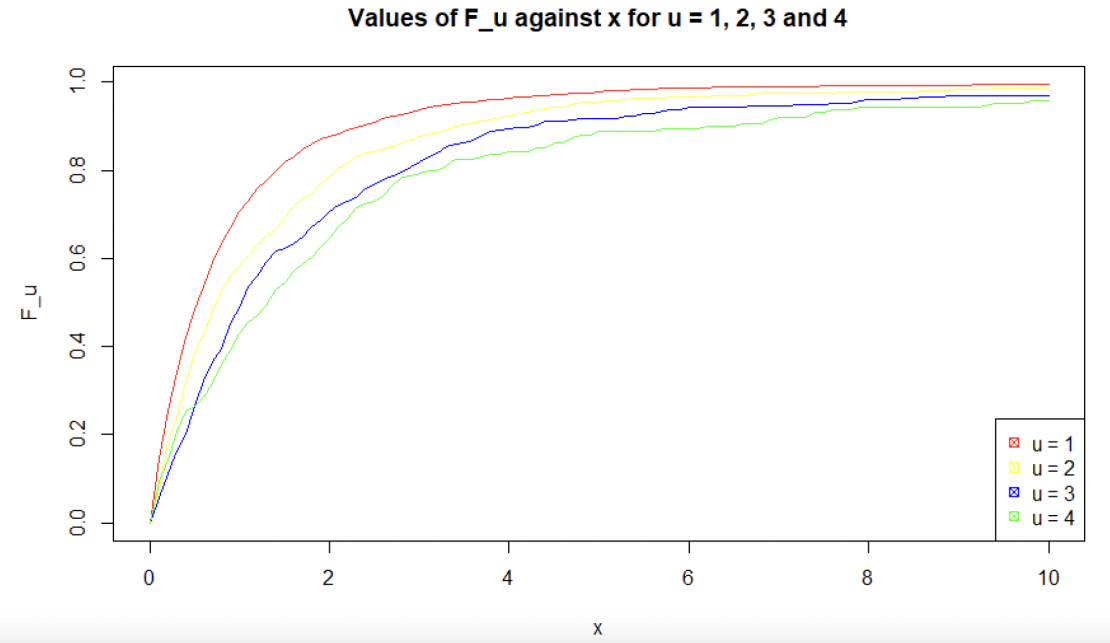
legend("bottomright",

legend = c("u = 1", "u = 2", "u = 3", "u = 4"),

= c("red", "yellow", "blue", "green"),

= 7)

Output:



(v) Comment on the graph produced in part (iv).

Hint: You are given that if $X \sim Pareto(\alpha, \lambda)$, then the threshold exceedance $X – u | X > u$ is distributed as $Pareto(\alpha, \lambda + u)$. **[20]**

Answer:

For all but the smallest values of x, $F\_u(x)$ decreases as u increases.

This is consistent with the result that if $X \sim Pareto(\alpha, \lambda)$, then the threshold exceedance $X – u | X > u$ is distributed as $Pareto(\alpha, \lambda + u)$.

There is some irregularity caused by sampling variation due to low data volumes above the higher values of u.

As a result of this irregularity, the curves for u = 3 and u = 4 cross over for small values of x.