**1) Modeling the distribution of times between two consecutive tram stops**

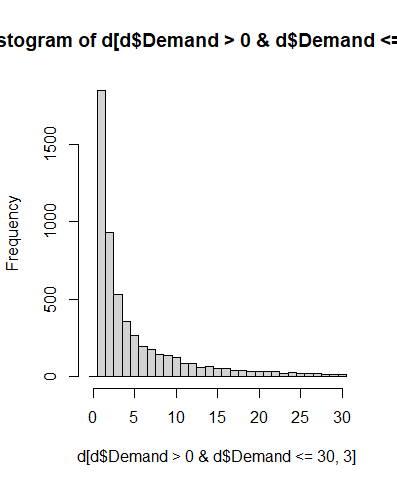
Let be a random variable that returns the time between two tram stops. From the Cracow data we have the following distribution of :

|  |  |  |  |
| --- | --- | --- | --- |
| time | # of edges | probability |  |
| 1 | 253 | 0,4729 | 47.3% |
| 2 | 223 | 0,4168 | 41.7% |
| 3 | 43 | 0,0804 | 8% |
| 4 | 13 | 0,0243 | 2.4% |
| 6 | 3 | 0,0056 | 0.6% |

We should use the same probability distribution when generating travel time for each edge in the generated graph

2**) Modeling the OD matrix values**

Let be a random variable that models the demand value for a randomly chosen cell in the OD matrix. We assume the exponential distribution for the number of passengers in each OD matrix cell (discretized). We treat the 0 case separately. In the Cracow OD matrix 35.6% of the cells was empty (demand = 0). For the demand values of 1 and more the distribution looks as follows:



For these values we fit the parameter of the exponential distribution using the MLE. It is known that the MLE estimator of for exponential distribution is . Since we excluded 0, in our case we have .

Therefore, the procedure of generating random data for the OD matrix is as follows:

For each cell of the OD matrix  
 with 35.6% probability enter 0  
ELSE  
 generate a random number from the exponential distribution , where and return )

We return because the MLE is computed for x = 0, 1, 2, … but these values represent in fact the values x = 1, 2, 3, … since we exluded 0 from the domain of and treated it with a separate case; ceiling guarantees that the returned values will be greater than 0

The script in R:

d <- read.csv("demand\_cracow.csv", header = T, sep=";")

zero <- nrow(d[d$Demand==0,]) # treat 0 separately

pr0 <- zero / nrow(d) # pr0 = 0.356

d <- d[d$Demand > 0, ] # calculate for X = 1, 2, 3, ...

t <- as.data.frame(table(d$Demand))

t$Var1 <- as.numeric(levels(t$Var1))

nrow(d) # 5805 elements

lambdaE <- sum(nrow(d)) / (sum(d$Demand-1))

# MLE for exponential distribution; lambdaE=0.1257174

#example:

x <- ceiling(rexp(6000, lambdaE)) # generate 6000 entries for the OD matrix

table(x)

#distribution (736 entries with value 1, 587 entries with value 2 etc.):

#1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25

#736 587 585 489 420 398 308 275 250 234 194 176 152 159 105 102 96 73 77 81 52 54 50 44 30

#26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 51

#23 31 19 21 23 14 14 17 12 13 10 9 9 3 10 8 4 4 6 3 3 1 3 2 1

#52 53 58 59 63 67 70

#2 1 1 2 1 2 1

# final generator for a single OD matrix entry:

if (runif(1)<=0.356) { # should it be 0?

print(0)

} else { # not 0, so generate from exponential distribution

print(ceiling(rexp(1, 0.1257174)))

}

**3) Modeling the distribution of the lengths of paths in the network graph**

In the Cracow network graph there are 149 tram stops (nodes):

14 of them have degree 1 (they are the final stops)

102 of them have degree 2 (they are on the path)

22 of them have degree 3

11 of them have degree 4

This means that a random node in a generated graph should have degree:

1 – with probability 14/149 = 9.4%

2 – with probability 102/149 = 68.5%

3 – with probability 22/149 = 14.8%

4 – with probability 11/149 = 7.3%

This information can help us in generating paths. The problem is to generate a graph with a given degree distribution. There are some methods for this, for example see <https://mathinsight.org/generating_networks_desired_degree_distribution>