Netherlands School of Economics

ECONOMETRIC INSTITUTE

Report 7117

SIMULATING INTER-ARRIVAL TIMES WITH TIME-DEPENDENT ARRIVAL RATES

by F. van Doeland and J. van Daal

August 20, 1971

Provisional and Confidential

NETHERLANDS SCHOOL OF ECONOMICS
Econometric Institute

Errata to report 7117

p. 1: line 3: point (.) after the last word has to be deleted.

p. 2: line 13: after the last word the word "and" should be placed.
footnote "watering" should be "water-way"

p. 3: line 18: ". with" has to be deleted.

p. 5: line 8 and 9: "the reason why" should be read as: "explained by the fact that".

p. 6: line 12: "subrontime" should be "subrontine"

p. 8: line 1: ". With" should be "with".

SIMULATING INTER-ARRIVAL TIMES WITH TIME-DEPENDENT ARRIVAL RATES

by F. van Doeland and J. van Daal

Contents

1. INTRODUCTION

Many simulation studies of queueing problems require the generating of inter-arrival times as input of the system. Starting for instance at instant 0 [1)] the first drawing X_1 from our distribution of inter-arrival times. provides us with the first instant of arrival: $A_1 = X_1 + 0$; the second drawing X_2 results in $A_2 = A_1 + X_2$, etc.

Apart from empirical frequency functions several classes of theoretical distribution functions can be used for generating the inter-arrival times; among which the class of the exponential distribution functions plays an important role. All these functions show parameters like the λ in the exponential density function

$$f(u) = \lambda e^{-\lambda u}$$

In most simulation studies these parameters are kept constant over time; they are never constant, however, during a longer interval of time. In most cases this is harmless because the mean

---

1) The word "instant" indicates a point on the time axis. The word "time" means an interval on the time axis.

inter-arrival times, hence the generated times, are small compared with the time over which the parameters alter substantially. As an example we mention the simulation of a road-traffic situation during rush-hours; the mean inter-arrival times of cars are fractions of minutes approximately constant as long as the flow of traffic continues i.e. approximately during one hour.

There are, however, situations in which the arrivals follow a Poisson distribution while nevertheless it is impossible to generate inter-arrival times with the help of an exponential distribution with a constant parameter. We have come across such situations in studying the behaviour of big ships [2)] sailing on a major Dutch river.

One can roughly distinguish three categories of ships, viz.:

1. Ships sailing (continuously) 24 hours a day, e.g. pusher-tugs coasters,
2. Ships sailing from 4 a.m. to 8 p.m. (approximately), e.g. pusher-tugs sailing "semi-continuously" and ships carrying sand and gravel,
3. Ships sailing from 7 a.m. to 6 p.m. (approximately).

Suppose one can expect λ_i ships per day on an average of category i (i = 1, 2, 3) coming from one direction (e.g. upstream) and passing a fixed point of the river. Suppose further that in each category the ships will arrive at our observation point following a Poisson distribution. The probability that k ships of any of our three categories will arrive during a certain small interval [t; t + u] of time (measured in hours) will then be expressed by

$$p_k(t \quad u) = e^{-\lambda(t).u} \frac{\{\lambda(t).u\}^k}{k!},$$

in which $\lambda(t) = \lambda_1/24$ if $[t; t + h] \subset [0;4] \vee [20,24]$,

$$\lambda(t) = \lambda_1/24 + \lambda_2/16 \text{ if } [t; t + h] \subset [4;7] \cup [18;20]$$

and

$$\lambda(t) = \lambda_1/24 + \lambda_2/16 + \lambda_3/11 \text{ if } [t; t + h] \subset [7;18].$$

---

2) To be precise: the ships for which the railway bridge over the watering at Dordrecht (Holland) has to be opened.

(We do not consider the case in which the instants 0, 4, 7, 18, 20, 24 are an inner point of [t; t + h]).

In practise it turned out that the values of λ_i never exceeded 10. Take e.g. $\lambda_1 = 4.8$, $\lambda_2 = 8.0$ and $\lambda_3 = 4.4$. This means, for instan that at 8 a.m. we can expect an arrival on an average within 1/1.1 hour. But what do we know for instance about the expected inter-arrival time at, 3 a.m. or 7 p.m. ?

If we generate inter-arrival times X_k by means of a negative exponential distribution with parameter $\lambda(t)$ and $t = A_k$ (= k^{th} instant of arrival then we simulate sytematically too few or too many arrivals. Examples of this phenomenon are given in the next section. In the third section two methods of getting rid of these difficulties (but involving some new difficulties, however) are presented.

2. SOME EXAMPLES

In this section we present the results of four different simulations.

Example 1: 20 times, starting at instant 0 of the first day, with subsequent inter-arrival times during 100 [3] successive days are simulated by means of a negative exponential distribution function with parameter $(\lambda_1 + \lambda_2 + \lambda_3)/24 = 17.2/24$. That is to say: we stop if $\sum_i X_i \geq 100 \times 24 = 2400$ hours and we list the 20 numbers k (one of each simulation run) so that

$$\sum_{i=1}^{k} X_i \leq 2400 \text{ and } \sum_{i=1}^{k+1} X_i > 2400, \text{ where } X_i \text{ is the } i^{th}$$

generated inter-arrival time. The resulting numbers of arrivals are reproduced in table 2.

Example 2: 20 times subsequent inter-arrival times, again during 100 successive days are simulated starting at instant 0 of the first day. Once more we make use of a negative exponential distribution, but now with an instant-dependent

3) The numbers 20 and 100 are chosen arbitrarily.

parameter as given by table 1:

Instant element of	parameter
[0; 4) or [20; 24)	0.2
[4; 7) or [18; 20)	0.7
[7; 18)	1.1

In order to generate an inter-arrival time we use the parameter belonging to the instant of the previous arrival; the instant A_k of the k^{th} arrival is given by

$$A_k = \sum_{i=1}^{k} X_i - 24.n_k,$$

where

$$n_k = \left[(\sum_{i=1}^{k} X_i)/24 \right] \quad 4)$$

The results are presented in table 2.

Example 3: Now the parameter λ_k used as parameter of a negative exponential distribution will be:

$$\lambda_k = \frac{A_k + 24n_k}{2400} \times 20.2/24 + \frac{2400-(A_k + 24n_k)}{2400} \times 14.2/24$$

(see figure 1)

Each simulation run applies again to 100 days; 20 runs are made. The results are shown in table 2.

Example 4: The same as example 3, but now with:

$$\lambda_k = \frac{A_k + 24n_k}{2400} \times 14.2/24 + \frac{2400 -(A_k + 24n_k)}{2400} \times 20.2/24$$

The results can be found in table 2.

(see figure 2)

According to simulation (1) the number of arrivals is Poisson-distributed with parameter 17.2/24 x 2400 = 1720. The sample mean is in agreement with this. The sample variance (2703), however, differs substantially from its theoretical value 1720; later simulations have shown that this was accidental.

4) $[\mu] = j$ if j is the integer with $j \leq \mu < j + 1$.

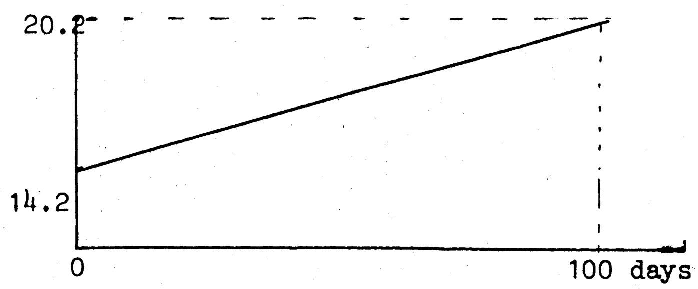


figure 1

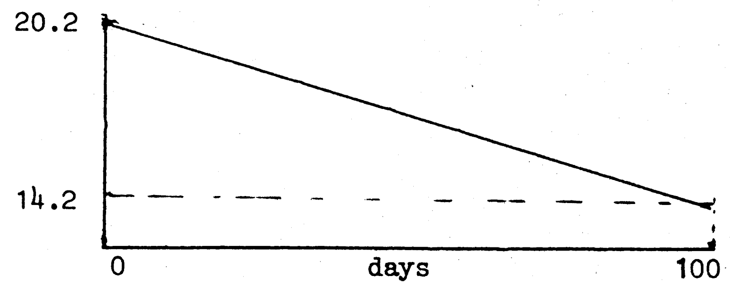


figure 2

Table 2 SIMULATED NUMBERS OF ARRIVALS PER RUN, ACCORDING TO THE FOUR EXAMPLES OF SECTION 2.

RUN	EXAMPLE (1)	(2)	(3)	(4)
1	1711	1381	1709	1712
2	1762	1530	1763	1761
3	1761	1731	1758	1761
4	1669	1467	1672	1670
5	1673	1407	1672	1673
6	1800	1390	1798	1805
7	1707	1531	1706	1703
8	1711	1426	1713	1710
9	1723	1575	1720	1723
10	1739	1444	1743	1744
11	1746	1441	1741	1743
12	1611	1447	1613	1619
13	1760	1546	1762	1756
14	1832	1311	1830	1827
15	1711	1547	1714	1716
16	1630	1576	1629	1629
17	1700	1491	1701	1704
18	1734	1430	1734	1734
19	1727	1472	1725	1729
20	1698	1475	1698	1695
MEAN	1720.3	1465.8	1720.0	1720.7
S.D.	52.0	69.7	51.3	50.9

In simulations (3) and (4) the parameter is systematically too low and too high, respectively. Nevertheless the sample mean does not differ significantly from 1720. We believe that this will always be the case because the parameter changes slowly with regard to the generated inter-arrival times; further simulations confirmed this idea.

Simulation (2), however. shows great differences just like numerous similar simulations that we carried out. This may be the reason why at 3.9.h, for instance an inter-arrival time of, say, 9 hours (with parameter 0.2)is generated so that a considerable part of the daytime is "passed over" this means that from 3.9 h to 12.9 h of that day no arrivals take place. The generated inter-arrival times are no longer small compared with changes of the parameter (as exemplified by the pattern in figure 3).

One may expect that this systematic underestimation of the number of arrivals due to violation of the Poisson-postulates (Feller (1967), pp. 446ff) is considerable.

In the next two sectionswe shall discuss two simple methods of avoiding this underestimation. We are aware of the fact that there are also more difficult (and therefore more general) solutions to the problem (see e.g. Luchak (1956)), but we were not looking for such solutions.

3. STRAIGHTFORWARD SIMULATION OF THE ALTERED POSTULATES

The altered system of postulates is

(3.1) $$P_1(t,u) = \lambda(t).u + O(u);$$

(3.2) $$P_n(t,u) = O(u);$$

(3.3) the numbers of arrivals in non-overlapping time interval are stochastically independent.

$P_n(t,u) = Pr\{n \text{ arrivals in interval } [t, t + u]\}$

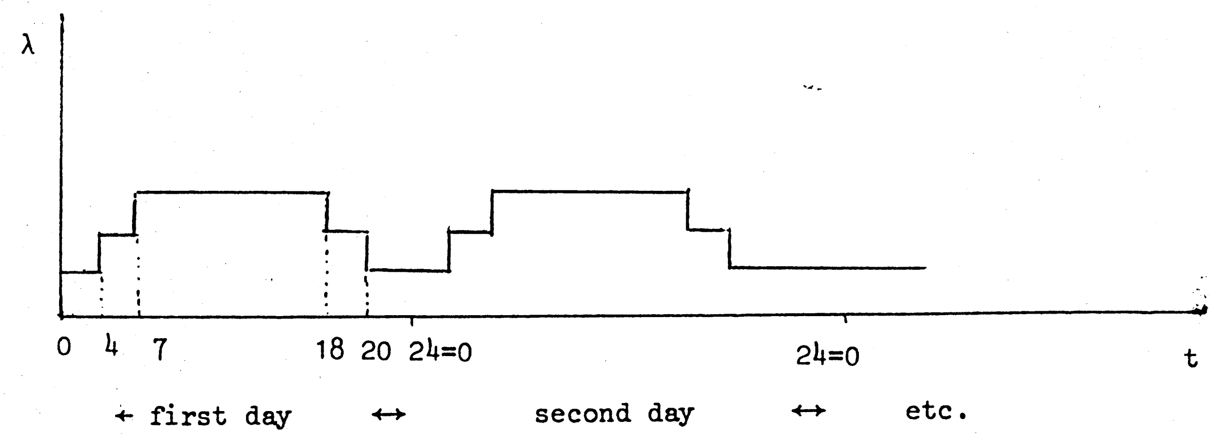


figure 3

With the same data as in simulation (2) of the previous section we simulate the arrivals in 600 successive days, as follows: starting with the first minute of the first day we decide for every minute whether or not there is an arrival. For every minute we draw a random number R uniformly from [0,1] (and independently of other drawings as far as possible with pseudo-random numbers).If $R \leq \lambda(t)/60$ there is an arrival in that minute and otherwise there is none; $\lambda(t)$ is taken from table 1. The results are reproduced in table 3.

Table 3.

Total number of arrivals per time-interval

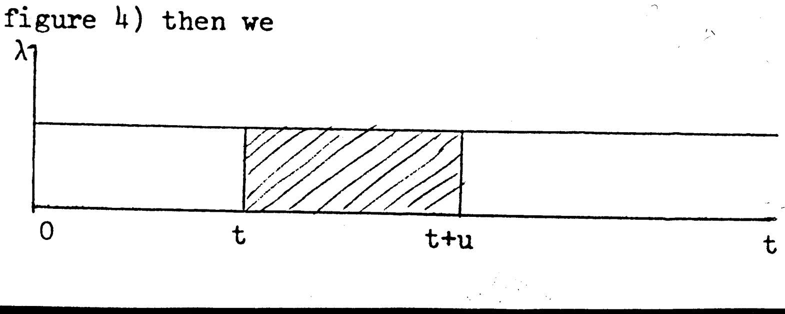
	time-interval				
	[0;4)	[4;7)	[7;18)	[18;20)	[20;24]
Simulated number of arrivals	461	1303	7163	829	472
theoretical number	480	1260	7260	840	480

For a description of the random generation process see the appendix.

The method is extremely simple and requires, simple arithmetic operations only, apart from the subrontime of random-generation. Nevertheless the computer time is considerable: the computation of the arrivals according to table 3 with a well-designed Fortran IV program took almost 45 minutes on the Econometric Institute's 2.2 μ IBM 1130 computer. This makes application of this method inadvisable especially in view of the arbitrariness of the choice of the time intervals (in this case minutes) except perhaps in small simulation runs. Furthermore, we also have the following method.

4. A GENERALIZATION OF THE POISSON-PROCESS

If the parameter   of a Poisson-process is instant-independent (see figure 4) then we



can generate an inter -arrival time distribution as follows: draw a random number R uniformly from [0;1]; compute $u = -\ln R/\lambda$; if the previous arrival instant is t then the next will be t + u. The "surface" of the hatched area is.

$$(-\ln R/\lambda).\ \lambda = -\ln R.$$

This suggested to us the idea of the following generalization.

Suppose our process will follow the postulates mentioned in the section 3, with an instant-dependent parameter $\lambda(t)$, (see figure 5), $\lambda(t) > 0$.

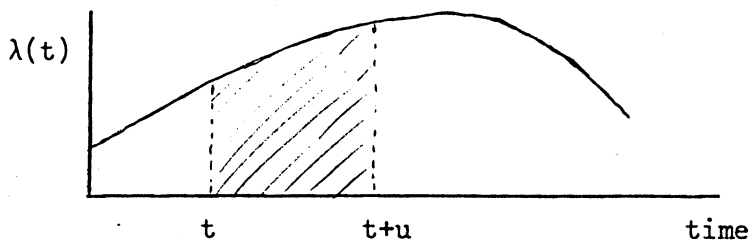


figure 5

If t is again the instant of the previous arrival and if R is a random draw from [0;1] then u will be such that

$$\int_t^{t+u} \lambda(v)dv = -\ln R$$

We wrote a computer program for generating inter-arrival times on the basis of this principle in the case where λ is a function of t. not necessarily continuous, graphically approximated by (parts of) straight lines. The program turned out to be about ten times faster than the "minute-method" of the previous section. The results of a first simulation run over 600 days can be found in table 4.

Total numbers of arrivals during 600 days per time interval

	time-interval				
	[0;4)	[4;7)	[7;18)	[18;20)	[20;24]
Simulated numbers	479	1309	7353	808	426
theoretical numbers	480	1260	7260	840	480

The results of 20 simulation runs of 100 days. With this method as described in section 2 (table 2) are reproduced in table 5.

Table 5. Numbers of arrivals during 100 days (20 simulations)

1720	1718	1719	1704	1756
1759	1678	1684	1790	1713
1710	1710	1751	1720	1650
1761	1817	1695	1648	1698

mean 1720.05
variance 1817.41
s.d. 42.63

Finally, we derive some theoretical properties of the inter-arrival time.

U(t) is the random variable such that

$$\int_t^{t+U(t)} \lambda(v)dv = - \ell n\ R;\ R \text{ is uniformly distributed over } [0;1]$$

Hereafter we shall write U for U(t)

$$F(u|t) = Pr\{U \leq u|t\}$$

$$= Pr\{- \ell n\ R \leq \int_t^{t+u} \lambda(v)dv\}$$

$$= Pr\{R \geq exp(- \int_t^{t+u} \lambda(v)dv)\}$$

$$= 1 - exp(- \int_t^{t+u} \lambda(v)dv).$$

Consequently, the density function will be

$$f(u|t) = \lambda(t+u)exp(- \int_t^{t+u} \lambda(v)dv)$$

If $\lambda(u)$ is constant over time, F en f are the probability distribution and density function of the negative exponential distribution respectively. If λ has a primitive function Λ than

$$F(u|t) = 1 - e^{-\{\Lambda(t+u) - \Lambda(t)\}}$$

and

$$f(u|t) = \lambda(t+u)e^{-\{\Lambda(t+u) - \Lambda(t)\}}$$

$P_n(t,u) = \Pr\{n \text{ arrivals in interval } [t,t+u]\}$

$$(4.1) \qquad P_o(t,u) = \Pr\{U > u|t\} =$$

$$= 1 - F(u|t)$$

$$= \exp\left(- \int_t^{t+u} \lambda(v)dv\right)$$

$$= \exp(-\{\Lambda(t+u) - \Lambda(t)\}$$

Now we shall prove by complete induction that for n = 0, 1, 2, ...

$$(4.2) \qquad P_n(t,u) = \frac{\{\Lambda(t+u)-\Lambda(t)\}^n}{n!} e^{-\{\Lambda(t+u)-\Lambda(t)\}}$$

Proof: From (4.1) we conclude that (4.2) is right for n = 0. Suppose (4.2) is right for any n = k. Then

$$P_{k+1}(t,u) = \int_0^u f(x|t).P_k(t + x, u - x)\, dx$$

$$= \int_0^u \lambda(t + x)e^{-\{\Lambda(t+x)-\Lambda(t)\}}\frac{\{\Lambda(t+x+u-x)-\Lambda(t+x)\}^k}{k!}. e^{-\{\Lambda(t+x+u-x)-\Lambda(t+x)\}}$$

$$= \int_0^u \lambda(t+x)e^{-\{\Lambda(t+u)-\Lambda(t)\}}\frac{\{\Lambda(t+u)-\Lambda(t+x)\}^k}{k!}\, dx$$

$$= e^{-\{\Lambda(t+u)-\Lambda(t)\}}\left[-\frac{\{\Lambda(t+u)-\Lambda(t+x)\}^{k+1}}{(k+1)!}\right]_0^u$$

$$= e^{-\{\Lambda(t+u)-\Lambda(t)\}}\frac{\{\Lambda(t+u)-\Lambda(t)\}^{k+1}}{(k+1)!} .$$

If λ is a constant, $\Lambda(v) = \lambda v$. In that case (4.2) becomes the regular Poisson distribution.

From the postulates mentioned in section 3 we can deduce the following difference-differential equation

$$\frac{d}{dt} P_n(t,u) = - \lambda(t+u)P_n(t,u) + \lambda(t+u)P_{n-1}(t,u).$$

One can easily verify that (4.2)satisfies this equation, viz. as a particular solution.

REFERENCES

[1] W.Feller(1967), An Introduction to Probability Theory and its Applications I, Wiley, New-York.

[2] Th.H. Naylor, J.L. Balintfy, D.S. Burduck, K. Chu, (1966), Computer Simulation Techniques, Wiley, New-York.

[3] G. Luchak (1956),"The Solution of the Single-channel quening Equations characterized by a Time-dependent Poisson-distributed arrival Rate and a general Class of Holding Times", Opns. Res. 4 (1956) pp. 711-732.

APPENDIX

The random-number generating procedure

The uniformly distributed (pseudo-)random numbers used in this report are computed by means of the well-known multiplicative congruential relation:

$$r_{n+1} \equiv a.\ r_n \pmod{b},$$

with $b = 2^{31}$ $a = 2^{16} + 3$ and r_o an arbitrarily chosen odd natural number. The computer algorithm is such that two words of 16 bits each are combined to one (pseudo-random-)number. Hence $b = 2^{31}$.
The choice of $a = 2^{16} + 3$ has been made for three reasons:

1. To obtain the maximum cycle-length, viz. 2^{29} (the above-mentioned computer requires about 40 hours to compute all the numbers of one cycle).

2. To obtain minimal serial correlation between subsequent pseudo-random numbers; for that reason one has to choose a as near as possible to $b^{\frac{1}{2}}$ in view of the other critiria.
3. Computational convenience (a = 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 if written in digitals).