On Simulation Some Life Data Distributions

Prof. Emeritus Ion Văduva
Faculty of Mathematics and Computer Science,
University of Bucharest, Romania
vaduva@fmi.unibuc.ro; oriv@clicknet.ro

Abstract
First, the paper presents in short algorithms for simulating some distributions which are used in the paper. Then, some probability distributions are introduced, which are mixtures between distributions of a minim and a maxim of sequences of life data having Lomax (Pareto) or Weibull distributions (called target *distributions), mixed up with truncated geometric or Poisson distributions. These distributions could be used to calculate the reliability of multicomponent serial or parralel systems and particularly for simulating such sistems. Apart from composition methods for simulating these distributions, some other simulation algorithms,based on inverse or rejection methods, are presented.*
Keywords: *distribution functions*
ACM/AMS Classification: 62G30

1. INTRODUCTION

In [1,2,3] were introduced some mixtures of life data distributions, deriving from some known distributions, which involve mixtures with Poisson distribution. Here we continue to introduce similar distributions which involve further mixtures with Poisson distributions or *geometric distributions*. More precisely, we consider a life time L having a *target* cummulative distribution function (*cdf*) $\phi(x)$ and denote $\varphi(x)$ its probability density function (pdf). Hence $\varphi(x) = \phi'(x)$.

If $L_1, L_2, ..., L_n$ are indepemdent and identicaly distributed (*iid*) random variables, then, let us consider the *extremum ranking* variables, namely,

$$V = \min_{1\leq i\leq n} L_i;\ W = \max_{1\leq i\leq n} L_i. \tag{1.1}$$

Note that the random variables V and W have the following $cdf's$ respectively

$$\phi_V(x) = (1-(1-\phi(x))^n;\ \phi_W(x) = (\phi(x))^n \tag{1.2}$$

and the correspondinjg $pdf's$ are

$$\varphi_V(x) = n\varphi(x)(1-\phi(x))^{n-1};\ \varphi_W(x) = n\varphi(x)(\phi(x))^{n-1}. \tag{1.3}$$

In this paper we consider as target distributions the $Lomax(a, \theta)$ and the *standard* $Weibull(\nu)$ distributions. The sample size n is assummed to be a discrete random variable N^* which takes values $N^* = 1, 2, ...$ In the following N^* is either a truncated (on $[1, \infty)$) $Poissaon(\lambda)$ random variable, or a truncated $Geometric(p), 0 < p < 1$ random variable. In the $Poisson(\lambda)$ case, the truncated distribution is

$$P(N^* = k) = \frac{1}{e^{\lambda - 1}} \frac{\lambda^k}{k!}, \ \lambda > 0, \ k = 1, 2, ... \tag{1.4}$$

and in the $Geometric(p)$ case, the truncated distribution is

$$P(N^* = k) = pq^{k-1}, \ k = 1, 2..., \ 0 < p < 1, \ q = 1 - p. \tag{1.5}$$

The mixed distribution will be obtained as distributions of V or W with random $n = N^*$ (i.e. truncated $Poisson(\lambda)$ or $Geometric(p)$).

1.1 Simulation of discrete distributions involved

- The $Poisson(\lambda)$ distribution of the discrete random variable N is

$$P(N = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \ k = 0, 1, 2, ... \tag{1.6}$$

There are several known ways of simulating this distribution. One convenient way is based on the following property:

If $n \in \mathcal{N}^+$ and $p, 0 < p < 1$ are selected such as $n \to \infty$ (i.e. n is large) and $\lambda = n.p$ then $Poisson(\lambda)$ distribution can be approximated with the Binomial $Bin(n, p)$ distribution.

The $Bin(n, p)$ distrinution is in the forrm

$$P(N = k) = C_n^k p^k q^{n-k}, \ q = 1 - p, \ k = 0, 1, ..., n \tag{1.7}$$

As n is large it is also known that

The random variable

$$Z = \frac{N - np}{\sqrt{npq}} \tag{1.8}$$

is approximately distributed as normal $N(0, 1)$

As the normal $N(0, 1)$ random variable can be easily simulated (see for instance [4],p.68-69 and & 3.5.4), the random variable N can be simulated as

Generate $Z \to N(0, 1)$;

Take $N = \{np + Z\sqrt{npq}\}$.

(The notation $\{x\}$, $\forall x \in R$ means the closest integer from x).

- The Geometric $Geom(p)$ $(0 < p < 1)$ random variable N has the following distribution

$$P(N = k) = pq^k,\ k = 0, 1, ..., q = 1 - p \tag{1.9}$$

Note that the *cummulative* geometric distribution function is

$$F(x) = P(N < x) = \sum_{i=1}^{x-1} pq^i = 1 - q^x,\ x - 0, 1, 2, ... \tag{1.10}$$

therefore, the simulationn of N in this case can be done by *the inverse* method as follows

Generate U an uniform $(0,1)$ random number;

Take $N = \left[\frac{\log(U)}{\log(q)}\right]$.

(The $[x], \forall x \in R$, denotes *the integer part of* x).

The described algorithms simulate discrete random variables N having $Poisson(\lambda)$ or Geometric $Geom(p)$ distributions. For the purpose of this paper, we need to simulate *truncated* random variables $N^* = N, N > 0$. Any of these random variables are easily genereted by the following *rejection* algorithm.

repeat

Generate $N \to Poisson(\lambda)$ *(or* $Geom(p)$*);*

until $N > 0$;

Take $N^* = N$.

1.2 Simulation of discrete distributions involved

As concerns *Weibull distribution* it is enough to consider only the *standard* $Weibull(0, 1, \nu)$ *distribution* having the *pdf* in the form

$$f(x, \nu) = \begin{cases} 0 & \text{if } x \leq 0 \\ \nu x^{\nu-1} e^{-x^\nu} & \text{if } x > 0. \end{cases} \tag{1.11}$$

The corresponding *cdf* is

$$F(x, \nu) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 - e^{-x^\nu} & \text{if } x > 0, \end{cases} \tag{1.12}$$

and its *inverse* is the solution of the equation $F(x) = U$ i.e.

$$F^{-1}(x) = (-\log U)^{\frac{1}{\nu}},\ \text{for } 0 < U < 1. \tag{1.12'}$$

The *Lomax*(a,θ) *distribution* has the *pdf* in the form

$$f(x,a,\theta)=\begin{cases}0 & \text{if } x\le 0\\ \frac{a\theta}{(1+\theta x)^{a+1}} & \text{if } x>0.\end{cases} \tag{1.13}$$

The *Lomax cdf* is

$$F(x,a,\theta)=\begin{cases}0 & \text{if } x\le 0\\ 1-\frac{1}{(1+\theta x)^{a}} & \text{if } x>0\end{cases} \tag{1.14}$$

and its inverse (the solution of equation $F(x)=U$) is

$$F^{-1}(U)=\frac{1}{\theta}\left((\frac{1}{U})^{\frac{1}{a}}-1\right),\ for\ 0<U<1. \tag{1.14'}$$

If U is an uniform random number over $(0,1)$, then the target distributions are simply simulated by the *inverse method* as

$$X=F^{-1}(U),$$

where F^{-1} is either given by (1.12') or by (1.14').

2. RESULTS: MIXED DISTRIBUTIONS OBTAINED

The $pdf's$ of mixed distributions are in the form

$$f_V(x)=\sum_{k=1}^{\infty}P(N^*=k)k\varphi(x)(1-\phi(x))^{k-1}, \tag{2.1}$$

$$f_W(x)=\sum_{k=1}^{\infty}P(N^*=k)k\varphi(x)(\phi(x))^{k-1}. \tag{2.2}$$

2.1 The case of truncated $Poisson(\lambda)$ distribution

For the Poisson distribution Kus [2] derived a new distribution (of W), when the target distribution is *extreme value distribution.*In [1], Lupu derived distributions on V and W when the target distribution is exponential $Exp(\lambda)$.

Here, we assume first that the target distribution is $Weibull(0,1,\nu)$, $\nu>0$, i.e a Weibull standard which has the *pdf*

$$\varphi(x)=\nu x^{\nu-1}e^{-x^{\nu}},x>0. \tag{2.3}$$

Therefore, for the variable V we have

$$f_V^{(PW)}(x) = \sum_{k=1}^{\infty} \frac{1}{e^\lambda - 1} \frac{\lambda^k}{k!} k\nu x^{\nu-1} e^{-x^\nu} e^{-(k-1)x^\nu} =$$

$$= \frac{\lambda}{e^\lambda - 1} \nu x^{\nu-1} e^{-x^\nu} \sum_{k=1}^{\infty} \frac{(\lambda e^{-x^\nu})^{k-1}}{(k-1)!}$$

which, after some calculations, gives finally

$$f_V^{(PW)}(x) = \frac{\lambda}{e^\lambda - 1} \nu x^{\nu-1} e^{-x^\nu} e^{\lambda e^{-x^\nu}}. \tag{2.4}$$

(The upper script (PW) means $Poisson - Weibull$ and the lower script V refers to the random variable in (1.1)).

One can see that the corresponding *cdf* is

$$F_V^{(PW)}(x) = \frac{1}{e^\lambda - 1} (e^\lambda - e^{\lambda e^{-x^\nu}}). \tag{2.5}$$

For the random variable W we have

$$f_W^{(PW)}(x) = \sum_{i=1}^{\infty} \frac{1}{e^\lambda - 1} \frac{\lambda^k}{k!} k\nu x^{\nu-1} e^{-x^\nu} (1 - e^{-x^\nu})^{k-1}$$

which after some calculation gives

$$f_W^{(PW)}(x) = \frac{\lambda}{e^\lambda - 1} \nu x^{\nu-1} e^{-x^\nu} \sum_{i=1}^{\infty} \frac{\lambda(1 - e^{-x^\nu})^{k-1}}{(k-1)!}$$

i.e.

$$f_W^{(PW)}(x) = \frac{\lambda}{e^\lambda - 1} \nu x^{\nu-1} e^{-x^\nu} e^{\lambda(1 - e^{-x^\nu})}. \tag{2.6}$$

(Here again, the upper script (PW) refers to $Poisson - Weibull$ mixture and the lower script W refers to the random variable W in (1.1)).

In this case the *cdf* is

$$F_W^{(PW)}(x) = \frac{e^{\lambda e^{-x^\nu}} - 1}{e^\lambda - 1}. \tag{2.7}$$

2.2 **The case of a truncated** *Geometric*(p) **distribution**

The calculations bellow use the following elementary result
Lemma 1 *If* $0 < \rho < 1$ *then*

$$\sum_{k=1}^{\infty} k\rho^{k-1} = \frac{1}{(1-\rho)^2}.$$

In this case we deal only with $Weibull(0,1,\nu)$ target distribution and with $Pareto(a,\theta)(Lomax(a,\theta))$ target distribution.

- In the case of *Geometric*(p) distribution, the random variable V in (1.1) when the target distribution is Weibull has the *pdf*

$$f_V^{(GW)}(x) = \sum_{i=1}^{\infty} pq^{k-1}k\nu x^{\nu-1}e^{-x^{\nu}}(e^{-x^{\nu}})^{k-1}.$$

Using Lemma 1, after some calculation one obtain

$$f_V^{(GW)}(x) = p\nu x^{\nu-1}\frac{e^{-x^{\nu}}}{(1-qe^{-x^{\nu}})^2}. \tag{2.8}$$

(Here the meaning of upper script (GW) is obvious).

The corresponding *cdf* is

$$F_V^{(GW)}(x) = \frac{1}{q} - \frac{p}{q}.\frac{1}{1-e^{-x^{\nu}}}. \tag{2.8$'$}$$

The mixture of geometric distribution with random variable W when the target distribution is Weibull has the *pdf*

$$f_W^{(GW)}(x) = p\sum_{i=1}^{\infty} k\nu x^{\nu-1}e^{-x^{\nu}}q^{k-1}(1-e^{-x^{\nu}})^{k-1}$$

which after some calculations based on Lemma 1, gives

$$f_W^{(GW)}(x) = p\nu x^{\nu-1}e^{-x^{\nu}}\frac{1}{(p+qe^{-x^{\nu}})^2}. \tag{2.9}$$

After some calculations the corresponding *cdf* is

$$F_W^{(GW)}(x) = \frac{p}{q}\frac{1}{p+qe^{-x^{\nu}}} - \frac{p}{q}. \tag{2.9$'$}$$

- Let us consider now the *Lomax*(a,θ) target distribution which has the *pdf*

$$f(x)=\frac{a\theta}{(1+\theta x)^{a+1}},\ x>0\,a>0,\ \theta>0. \tag{2.10}$$

The mixture distribution of the *Geometric*(p) with the random variable V, when the target distribution is *Lomax*(a,θ) has the *pdf*

$$f_V^{(GL)}(x)=\sum_{i=1}^{\infty}pq^{k-1}.\frac{a\theta}{(1+\theta x)^{a+1}}\frac{k}{((1+\theta x)^a)^{k-1}}.$$

Using Lemma 1, after saome calculation we obtain

$$f_V^{(GL)}(x)=\frac{pa\theta(1+\theta x)^{a-1}}{[(1+\theta x)^a-q]^2},x>0. \tag{2.11}$$

The *cdf* of $f_V^{(GL)}(x)$ is (after some calculations!)

$$F_V^{(GL)}(x)=1-\frac{p}{(1+\theta x)^a-q},\ x>0. \tag{2.12}$$

The mixture distribution of the *Geometric*(p) with the random variable W, when the target distribution is *Lomax*(a,θ) has the *pdf*

$$f_W^{(GL)}(x)=\sum_{i=1}^{\infty}pq^{k-1}\frac{ka\theta}{(1+\theta x)^{a+1}}\left(1-\frac{1}{(1+\theta x)^a}\right)^{k-1},\ x>0.$$

Using again Lemma 1, after some calculations, we obtain the final form

$$f_W^{GL)}(x)=pa\theta\frac{(1+\theta x)^{a-1}}{(p(1+\theta x)^a+q)^2},\ x>0. \tag{2.13}$$

The corresponding *cdf* of $f_W^{(GL)}(x)$ is

$$F_W^{(GL)}(x)=1-\frac{1}{p(1+\theta x)^a+q},\ x>0. \tag{2.14}$$

3. SIMULATION OF INTRODUCED DISTRIBUTIONS

3.1 Direct simulation as composition algorithms

Since all distributions are *mixture* distributions, they can be simulated by *composition algorithms* in the form

begin
Simulate n *a samplimg value of* N^**;*
Simulate $L_1, L_2, ..., L_n$ *having the target distribution;*
Calculate V *(or* W*) as in (1.1);*
Take $X = V$ (or $X = W$)
end.

The randon variable X produced by the algorithm has the corresponding mixture distribution. Implementation of all algorithms does not imply any difficulty.

3.2 Simulation based on the inverse method

For all introduced distributions the *cdf's* were specified.Therefore simulation of any random variable X which has the *cdf* $F(x)$ is done according to the following *inverse* algorithm

begin
Simulate a random number U *uniform over (0,1);*
Calculate $X = F^{-1}(U)$*;*
end.

(In the algorithm, F^{-1} is the inverse function of F).
In the following we list-up the inverse F^{-1} of *cdf's* introduced in section 2.

1. The inverse of the *cdf* $F = F_V^{(PW)}$ is

$$F^{-1}(U) = \{-\log[\log[e^\lambda - U(e^\lambda - 1)]^{\frac{1}{\lambda}}]\}^{\frac{1}{\nu}}. \tag{3.1}$$

2. The inverse of the *cdf* $F = F_W^{(PW)}$ is

$$F^{-1}(U) = \{-\log[\log[1 + U(e^\lambda - 1)]^{\frac{1}{\lambda}}]\}^{\frac{1}{\nu}}. \tag{3.2}$$

3. The inverse of the *cdf* $F = F_V^{(GW)}$ is

$$F^{-1}(U) = [-\log\left(\frac{q(1-U)}{1-qU}\right)]^{\frac{1}{\nu}}. \tag{3.3}$$

4. The inverse of the *cdf* $F = F_W^{(GW)}$ is

$$F^{-1}(U) = \left[-\log(\frac{q(1-U)}{p+qU}\right]^{\frac{1}{\nu}}. \tag{3.4}$$

5. The inverse of the *cdf* $F = F_V^{(GL)}$ is

$$F^{-1}(U) = \frac{1}{\theta}\left[\left(\frac{1-qU}{1-u}\right)^{\frac{1}{a}} - 1\right]. \tag{3.5}$$

6. The inverse of the *cdf* $F = F_W^{(GL)}$ is

$$F^{-1}(U) = \frac{1}{\theta}\left[\left(\frac{1-qU}{1-U}\right)^{\frac{1}{a}} - 1\right]. \tag{3.6}$$

For implementation of these simulation procedures, it must take care to avoid calculation of the log function when its *argument* is close to zero (i.e. < 0.0000001!).

3.3 **Simulation based on rejection method**

In this subsection we use the following known result [4,5]

Theorem 1 *Assume that X is a random variable to be simulated and its pdf is $f(x)$. Assume that Y is another random variable for wich a simulation method is known and its pdf is $h(x)$, such that functions f and g have the same support. Assume also that exists a finite constant $\alpha > 0$, such as*

$$\frac{f(x)}{h(x)} \leq \alpha. \tag{3.7}$$

Let U be an uniform (0,1) variate stochastically independent from Y. Then, the conditional pdf of Y given that

$$0 < U \leq \frac{f(Y)}{\alpha h(y)}, \tag{3.8}$$

is f. (Therefore X is simulated by Y which satisfies condition (3.8))

From this Theorem results the following *general rejection algorithm* for simulating the random variable X :

Algorithm R

repeat
Generate U an uniform random number over (01);
Generate Y having pdf h, such as Y and U are stochastically independent;
until $U \leq \frac{f(Y)}{\alpha h(Y)}$;*(This is the condition C).*
Take $X = Y$.

It is shown that the probability to come out from the cycle **repeat... until** (i.e the *acceptance probability*), is

$$p_a = \frac{1}{\alpha}, \tag{3.9}$$

therefore $1 < \alpha < \infty$ and the algorithm is fast if p_a is close to one.

In the following we apply this theorem for simulating, by the rejection method, the distributions introduced in section 2. In fact will be presented only important features necessary to implement **Algorithm R** for each distribution, namely:

- The enveloping density $h(x)$;
- The constant α;
- The acceptance probability p_a;
- The simpliffied expression of the condition C in the cycle **repeat... until C.**

- **A note** regarding selection of the *enveloping density*:we will try (and it will work!) as *enveloping densities* the densities corresponding to *target distributions.*

  a. **The distribution** $F_V^{(PW)}$. The enveloping density selected is the standard $Weibull(0,1,\nu)$ density

  $$h(x) = \nu x^{\nu-1} e^{-x^\nu}, x > 0, \nu > 0. \tag{3.10}$$

  The *enveloping constant* is

  $$\alpha = \frac{\lambda e^\lambda}{e^\lambda - 1} \tag{3.10'}$$

  and the *acceptance probability* is

  $$p_a = \frac{e^\lambda - 1}{\lambda e^\lambda}. \tag{3.10"}$$

  The simplified form of the predicate C is

  $$C = true, iff, U \le e^{-\lambda + \lambda e^{-Y^\nu}}. \tag{3.10'''}$$

  b. **The distribution** $F_W^{(PW)}$. The enveloping density is the same as in **a.**; the constant α and the acceptance probability are also the same. The final form of the predicate C is

  $$C = true, iff, U \le e^{-\lambda + \lambda(1 - e^{-Y^\nu})}. \tag{3.11}$$

  Even if the probability p_a is the same for cases **a.** and **b.**, the rejection algorithm in the case **a.** is faster because the condition C in this case requires a smaller time-complexity.

c. **The distribution** $F_V^{(GL)}$. The enveloping density $h(x)$ is also given by (3.10); the constant α and the probability p_a are

$$\alpha = \frac{1}{p}, \quad p_a = p. \tag{3.12}$$

The final form of the predicate C is

$$C = true, \; iff \, U \leq \frac{p^2}{(1 - qe^{-Y^\nu})^2}. \tag{3.12$'$}$$

d. **The distribution** $F_W^{(GW)}$. The enveloping density $h(x)$ is also (3.10); The constants α and p_a are the same as in case **c.** and the predicate C is

$$C = true, \; iff \, U \leq \frac{p^2}{(p + qe^{-Y^\nu})^2}. \qquad (3.13).$$

Note that complexity ot the two rejection algorithms (cases **c.** and **d.**) differs only by the time necessary to evaluate predicates C.

e. **The distribution** $F_V^{(GL)}$. The enveloping density selected is the $Lomax(a, \theta)$ density, i.e.

$$h(x) = \frac{a\theta}{(1 + \theta x)^{a+1}}, x > 0, a, \theta > 0. \tag{3.14}$$

The enveloping constant α and the acceptance probability are

$$\alpha = \frac{1}{p}, \; p_a = p, \tag{3.14$'$}$$

and the final form of the predicate C is

$$C = true, \; iff \, U \leq \frac{p^2 a\theta(1 + \theta Y)^{2a}}{[(1 + \theta Y)^a - q]^2}. \tag{3.14$''$}$$

f. **he distribution** $F_W^{(GL)}$..The enveloping density is also $Lomax(a, \theta)$ and the constants α, p_a are the same as in the case **e.**, i.e. are in the form (3.14'). The final form of the predicate C is

$$C = true, \; iff \, U \leq \frac{p^2 a\theta(1 + \theta Y)^{2a}}{[p(1 + \theta Y)^a + q]^2}. \tag{3.15}$$

If the Weibull and Lomax random variates are simulated by the inverse methods, then the final form of predicates could be further simplified, thus reducing execution time.

- Note that the **moments** of mixed distributions introduced exist and they could be simply calculated.

References

1. KUŞ, KOSKUN (2007). *A New Lifetime Distribution* Computational Statisyics and Data Analysis, 51(2007), p.4497-4509.
2. LUPU, ELENA (2008). *On Simulation of Some Lifetime Distributions*, Analele Universităţii Bucureşti,Matematică-Informatică, Anul LVII.
3. LUIS ANTONIO PEREZ GONZALEZ, ION VĂDUVA (2011). *Simulation of Some Mixed Lifetime Distributions*, Analele Universităţii Bucureşti, Informatică,Anul LX, p.17-32.
4. VĂDUVA, Ion (1977). *Modele de simulare cu calculatorul*, Bucureşti, Ed. Tehnică.
5. VĂDUVA, ION (2004). *Modele de simulare*, Ed. Univ. Bucureşti