Sampling Algorithm of Order Statistics for Conditional Lifetime Distributions

G. R. Pasha
Department of Statistics, Bahauddin Zakariya University Multan, Pakistan
Email: drpasha@bzu.edu.pk

Muhammad Aslam
Department of Statistics, Bahauddin Zakariya University Multan, Pakistan
Email: aslamasadi@bzu.edu.pk

Atif Akbar
Department of Statistics, Bahauddin Zakariya University Multan, Pakistan
Email: atifakber@bzu.edu.pk

Abstract

We construct a sampling algorithm to generate ordered statistics from conditional lifetime distributions by expressing the conditional distribution of ordered statistics in terms of cumulative hazard function. We use uniform spacing algorithm to generate uniform (0,1) ordered statistics in the proposed algorithm.

Keywords: Cumulative hazard function; Reliability function; Conditional lifetime distribution; Order Statistics; Sampling algorithm.

Introduction

A lifetime distribution represents the length of the life of a material, a human being or a device and its applications range from investigations of manufactured items to life expectancies of livings in different diseases. The theoretical population models used to describe unit lifetimes in which the population under consideration is constituted by the entire possible unit lifetimes for all of the units that could be manufactured based or biomedical based on a particular design and choice of material and manufacturing process. While a random sample from the population is the collection of failure times observed for randomly selected units. Thus, statistical analysis of lifetime, survival time or failure time data have developed into an important topic for researchers in many areas especially in the engineering and biomedical sciences. Such analysis requires some generated data for simulation purposes that result in intrigue of random number generation from the desired distributions. Simulation is fundamental to the investigation of complex stochastic systems (see Fishman, 1996; Evans and Swatz, 2000). A lot of effort has been spent on the construction of sampling algorithms of random number generation, especially, to help the statisticians who are mainly interested in generating independent sequences of random variables (e.g., see Devroye, 1986).

Order Statistics that describe random variables arranged in order of magnitude, are widely used in statistical models and inference and play a key role in lifetime-distribution study. The conditional distributions of order statistics can be expressed in terms of the cumulative hazard function. Newby (1979) uses these results to construct an algorithm to simulate samples of order statistics from lifetime distributions. Many lifetime distribution models e.g. Exponential, Gamma, Weibull, Lognormal etc. having one or more parameters, are used to describe lifetime of the individuals. Normal, Exponential, Rayleigh, Beta, Gamma, and *t* distributed random variates can be obtained by manipulation of the order statistics defined by samples of IID uniform random variates in (0,1). Box and Muller (1958) derive normal random variates in this manner. Newby (1979) also uses uniform (0,1) random variates in the formulation of algorithm for order statistics from lifetime distributions.

In this paper, we extend Newby's algorithm by using uniform (0,1) order statistics rather than simple uniform (0,1) random variates as proposed by Newby for the generation of more pragmatic samples of order statistics from conditional lifetime distributions. Devroye (1986) has discussed three main algorithms to generate uniform (0,1) order statistics;

A. Sorting
   (i) Generate IID uniform (0, 1) random variates $U_1, U_2, \ldots, U_n$.
   (ii) Obtain $U_{(1)}, U_{(2)}, \ldots, U_{(n)}$ by sorting the U_i's .

B. Via exponential spacing
   (i) Generate IID exponential random variates $E_1, E_2, \ldots, E_{n+1}$ and compute their sum K.
   (ii) Set $U_{(0)} = 0$
   $U_{(i)} = U_{(i-1)} + E_i/K;$ $\quad i = 1, 2, \ldots, n$

C. Via uniform spacing
   (i) Generate a uniform (0, 1) random variate U
   (ii) Set $U_{(n+1)} = 1$
   $U_{(i)} = \mathrm{U}^{1/i}\, U_{(i+1)};$ $\quad i = n, n\text{-}1, \ldots, 1$

Algorithms B and C were developed in a series of papers by Lurie and Hartley (1972), Schucany (1972) and Lurie and Manson (1973). Algorithm A is the naïve approach and least efficient since additional storage proportional to sample size *n*. Devroye (1986) finds only the method C to be one-pass method. So following this idea we use this method in the algorithm given by Newby (1979).

In Section 2, on the lines of Newby (1979), we define hazard function, cumulative hazard function and reliability or survival function for a random variable from any lifetime distribution and derive the conditional distribution of the order statistics. In Section 3, we construct a new sampling algorithm to generate order statistics from conditional lifetime distributions by using uniform spacing algorithm in the Newby's algorithm. In Section 4, we present an example of Rayleigh distribution to illustrate the proposed algorithms. In Section 5, we give some applications of the proposed sampling algorithm and finally the Section 6 concludes.

Conditional Distribution of Order Statistics from Lifetime Distributions

Let $t_1, t_2, \ldots, t_n$ be a random sample of size n from a lifetime distribution having density $f(t)$ and distribution function $F(t)$ defined on $[0, \infty)$. The hazard function of the r.v. T is

$$h(t) = f(t)/\{1 - F(t)\}, \tag{2.1}$$

and the cumulative hazard function is

$$H(t) = \int_0^t h(u)\,du. \tag{2.2}$$

Following Gendenko et al. (1969), Newby (1979) shows that the distribution F and reliability R may be expressed as

$$F(t) = 1 - exp\{- H(t)\} \tag{2.3}$$

$$R(t) = 1 - F(t) = exp\{- H(t)\} \tag{2.4}$$

Consider the experiment of putting n identical items on test at time zero and recording the failure times, $t_1, t_2, \ldots, t_n$. This generates an ordered sample of size n, $t_{(1)}, t_{(2)}, \ldots, t_{(n)}$, and there is probability zero of two failure times coinciding. Now at time, when j items have failed so that the times $t_{(1)}<t_{(2)}< \ldots <t_{(j)}$ have been noted, the conditional reliability of the remaining $(n - j)$ items is

$$R(t_{(j+1)} \mid t_{(j)}) = [exp\{-H(t_{(j+1)})\}/exp\{-H(t_{(j)}\}]^{(n - j)}$$
$$= exp[- (n - j)\{H(t_{(j+1)})-H(t_{(j)}\}] \tag{2.5}$$

Since order statistics have the Markov's property (Pyke, 1965, Feller, 1966), Newby (1979) derives the conditional distribution of $t_{(j+1)}$ given $t_{(j)}, t_{(j-1)}, \ldots, t_{(1)}$ using (2.3), (2.4), and (2.5) as

$$F(t_{(j+1)} \mid t_{(j)}, t_{(j-1)}, \ldots, t_{(1)}) = F(t_{(j+1)}) \mid t_{(j)})$$
$$= 1 - exp[- (n - j)\{H(t_{(j+1)})-H(t_{(j)}\}] \tag{2.6}$$

Sampling Algorithm for Ordered Statistics

By taking logarithm and rearranging (2.6), it is easy to show that

$$t_{(j+1)} = H^{-1}[H(t_{(j)} - ln\{1 - F(t_{(j+1)} \mid t_{(j)})\}/(n - j)] \quad (3.1)$$

Since the observed value of any distribution function F is a uniform r.v. on (0,1), the observed value of $F(t_{(j+1)} \mid t_{(j)})$ will be uniform (0,1) for each j and whatever the value of $t_{(j)}$.

Thus, Newby (1979) suggest the following algorithms to generate sample of order statistics;

$$h_0 = 0,$$
$$h_{j+1} = h_j - ln(u_j)/(n - j); \qquad j = 0, 1, \ldots, n-1,$$
$$t_{(j+1)} = H^{-1}(h_{j+1}),$$

where u_j is an observation from the uniform (0,1) distribution.

Since, in the sense of Newby (1979), the sequence $t_{(j)}$ is increasing for H is an increasing function and the sequence h_j is also increasing, so it will be more coherent to use ordered uniform random numbers in (0,1) rather than simple uniform (0,1) values as used by Newby (1979). Therefore, we use algorithm C to generate ordered uniform (0,1) statistics via uniform spacing due to one-pass method and storage efficiency as described in Section 1(see Devroye, 1986 for details), in the Newby's algorithm and propose a new sampling algorithm to generate ordered statistics from conditional lifetime distributions as

(i) $u_{(n+1)} = 1$
$u_{(i)} = u^{1/i}\, u_{(i+1)};$ $\qquad i = n, n-1, \ldots, 1$ and $u \sim U(0,1)$

(ii) $h_0 = 0,$
$h_{j+1} = h_j - ln(u_{j+1})/(n - j);$ $\qquad j = 0, 1, \ldots, n-1,$
$t_{(j+1)} = H^{-1}(h_{j+1}).$

Example

Consider the example of Rayleigh distribution that is widely used in the lifetime study of electronic components with the following density having parameter θ, for illustration.

$$f(t) = (t/\theta^2)\, exp[-t^2/(2\theta^2)]; \qquad t > 0, \quad \theta > 0.$$

The distribution function and cumulative hazard functions are

$$F(t) = 1 - exp[-t^2 / (2\theta^2)],$$
$$H(t) = t^2 / (2\theta^2),$$

and the inverse of the cumulative hazard function is

$$G(u) = (2u\theta^2)^{1/2}$$

The sequence of ordered statistics from the Rayleigh distribution can be generated by

(i) $u_{(n+1)} = 1$
$u_{(i)} = u^{1/i} u_{(i+1)};$ $\quad i = n, n\text{-}1, \ldots, 1$ and $u \sim U(0,1)$

(ii) $h_0 = 0,$
$h_{j+1} = h_j - ln(u_{j+1})/(n - j);$ $\quad j = 0, 1, \ldots, n\text{-}1,$
$t_{(j+1)} = (2\, h_{j+1}\theta^2)^{1/2}$.

For numerical application, we generate an ordered sample of size 25 from Rayleigh distribution with θ =5. We use both algorithms for comparison. Since for generation of said numbers using both algorithms, one has to take one random number from Uniform (0,1). Our randomly selected number is 0.3450. In Newby's algorithm, we use this number to generate the values of hazard function and then order statistics are generated. But in our proposed algorithm, we first generate ordered Uniform numbers using this randomly selected number 0.3450 and then find the values of the hazard function to generate required ordered numbers.

The summary statistics of the samples from both algorithms is as follows:

Summary Statistics

	Newby's Algorithm	New Proposed Algorithm
Mean	6.55	7.09
Standard Error	0.66	0.35
Kurtosis	-0.17	-0.04
Skewness	0.55	-0.84
Count	25	25

Applications

The measure of equipment's reliability or the survival time of a diseased person is the infrequency with which failures occur in time and the simulations based on the above algorithm may be helpful for the statistical analysis and solution of reliability or survival problem providing further insight to develop, understand and apply the method. This algorithm also simulates the situations where censoring truncates the samples before all *n* failure times have been observed. For instance, if a life test on *n* items is to be terminated at time T then sampling stops at the *j*th observation where $t_{j-1} < T \leq t_j$.

Conclusions

We extend the Newby's (1979) algorithm to generate ordered statistics from conditional lifetime distributions by expressing the conditional distribution of ordered statistics in terms of cumulative hazard function. We use uniform spacing algorithm to generate uniform (0,1) ordered statistics for the generation of the required samples. So it becomes more pragmatic to use uniform (0,1) order statistics rather than simple uniform (0,1) for the generation of samples of order statistics from conditional lifetime distributions.

References

1. Box, G. E. P. and Muller, M. E., 1958. A Note on the Generation of Random Normal Deviates. Annals Math. Stat, V. 29, pp. 610-61.
2. Devroye, L., 1986. Non-Uniform Random Variate Generation. Springer-Verlag, New York.
3. Evans, M. J. and Swatz, T. B., 2000. Simulation from Non-Standard Distributions using Envelope Methods. Proc. Winter Simulation Conference.
4. Feller, W., 1966. An Introduction to Probability Theory and Its Applications. Vol. II. New York Wiley.
5. Fishman, G. S., 1996. Monte Carlo: Concepts, Algorithms, and Applications. Springer-Verlag, New York.
6. Gnedenko, B. V., Belyayev, Y. K. and Solovyev, A. D., 1969. Mathematical Methods of Reliability. New York: Academics Press.
7. Lurie, D. and Hartley, H. O., 1972. Machine-generation of Order Statistics for Monte Carlo Computations. Am. Statist. 26, 26-27.
8. Lurie, D. and Mason, R. L., 1973. Empirical Investigation of Several Techniques for Computer Generation of Order Statistics. Commun. Stat. 2, 363-371.
9. Newby, M. J., 1979. The Simulation of Order Statistics from Life Distributions. Appl. Statist. Vol. 28, No. 30, pp. 298-300.
10. Pyke, R., 1965. Spacing. J. R. Statist. Soc. B, Vol. 27, pp 395-468.
11. Schucany, W. R., 1972. Order Statistics in Simulation. J. Stat. Camp. Sim. 1,281-286.