On the Distribution of the Minimum or Maximum of a Random Number of i.i.d. Lifetime Random Variables

Francisco Louzada[1], Estela M. P. Bereta[2], Maria A. P. Franco[2]
[1]Department of Applied Mathematics & Statistics, ICMC, Universidade de São Paulo, São Paulo, Brazil
[2]Department of Statistics, CCET, Universidade Federal de São Carlos, São Carlos, Brazil
Email: louzada@icmc.usp.kr

ABSTRACT

Statisticians are usually concerned with the proposition of new distributions. In this paper we point out that a unified and concise derivation procedure of the distribution of the minimum or maximum of a random number N of independent and identically distributed continuous random variables Y_i, $\{i = 1, 2, \cdots, N\}$ is obtained if one compounds the probability generating function of N with the survival or the distribution function of Y_i. Expressions are then derived in closed form for the density, hazard and quantile functions of the minimum or maximum. The methodology is illustrated with examples of the distributions proposed by Adamidis and Loukas (1998), Kus (2007), Tahmasbi and Rezaei (2008), Barreto-Souza and Cribari-Neto (2009), Cancho, Louzada, and Barriga (2011) and Louzada, Roman and Cancho (2011).

Keywords: Compounding Distributions; Distribution of the Maximum; Distribution of the Minimum; Probability Generating Function

1. Introduction

Several authors have proposed new distributions for the maximum or the minimum as extensions of the exponential distribution, such as [1-7]. In this paper, we obtain an alternative form to the one considered by these authors for obtaining the distribution of the minimum or maximum of N independent and identically distributed (i.i.d.) random variables Y_i, $\{i = 1, 2, \cdots, N\}$, N being also a strictly positive integer random variable with discrete probability function (dpf) and probability generating function (pgf) $G_N(t, \alpha)$ defined throughout the interval $[0, 1]$.

Let $S_0(y, \beta, \gamma)$ and $F_0(y, \beta, \gamma)$ be the survival function and cumulative distribution function of the random variables Y_i, $\{i = 1, 2, \cdots, N\}$. The cumulative distribution of the maximum out of $(Y_1, Y_2, \cdots, Y_N)$ is obtained by composing $G_N(t, \alpha)$ with the cumulative distribution function of Y_i and the survival function of the minimum is obtained by composing $G_N(t, \alpha)$ with the survival function of Y_i.

2. Model Formulation

Let N be a strictly positive random variable with dpf and pgf of $G_N(t, \alpha)$ defined throughout the interval $[0, 1]$. $G_N(t; \alpha)$ is increasing in t and satisfies the

equalities $G_N(0, \alpha) = 0$ and $G_N(0, \alpha) = 1$. Thus $G_N(t, \alpha)$ can be viewed as the value for $t \in [0, 1]$ of a cumulative distribution function.

If $G_N(t, \alpha)$ is an absolutely continuous function, its pdf is denoted by $G_N(t, \alpha)$, and risk function is represented by $h_{G_N}(t, \alpha)$, both supported on $[0, 1]$. Keeping the assumptions made in this section on the random variable N, it follows that the function $G_N^*(t) = G_N(1 - t; \alpha)$ is decreasing function of t being thus the equalities $G_N^*(0) = 1$ and $G_N^*(1) = G_N(0; \alpha) = 0$. So $G_N^*(t)$ is the restriction to the interval $[0, 1]$ of a survival function and if it is absolutely continuous function it has pdf represented by $g_{G_N}^*(t)$ and hazard function denoted $h_{G_N}^*(t)$, both with support in $[0, 1]$.

In this paper α is the vector of parameters of $G_N(t, \alpha)$ and all other Greek letters refer to the parameters of cumulative distribution function of Y, which is represented by $F_0(y, \beta, \gamma)$.

Let $(Y_1, \cdots, Y_N)$ be a sequence of i.i.d. random variables with pdf and N corresponding to the number of random variables Y_i i.i.d. random variables with survival function $S_0(y, \beta, \gamma)$ and the cumulative distribution function $F_0(y, \beta, \gamma)$.

For $N = n$ the survival function of $\min(Y_1, Y_2, \cdots, Y_n)$ is given by $S_{\min}(t) = [S_0(t)]^n$, but when N is a random variable the survival function of the minimum is

given by

$$S_{\min}(t) = \sum_{n=1}^{\infty} \left[S_0(t) \right]^n P(N=n). \qquad (1)$$

Several authors have obtained density functions of the minimum by (2), which requires the calculation of a series, given by

$$f_{\min}(t) = \sum_{n=1}^{\infty} f_{\min}(t|N=n) P(N=n)$$
$$= \sum_{n=1}^{\infty} n \left[S_0(t) \right]^{n-1} f_0(t) P(N=n). \qquad (2)$$

In this paper we show that a more concise way to obtain the functions that determine the distribution of the minimum without the need of the calculation a series by considering the fact that the expression (1) can also be written as,

$$S_{\min}(t) = G_N \left(S_0(t) \right). \qquad (3)$$

Thus, the survival function of the minimum is obtained directly from (3), consequently the pdf of the minimum is obtained by derivation of $G_N \left(S_0(t) \right)$. Similarly, the survival function of maximum $\max(Y_1, Y_2, \cdots, Y_n)$ is obtained from $G_N \left[F_0(t) \right]$, and the cumulative distribution and pdfs of the maximum are obtained by derivation of $G_N \left[F_0(t) \right]$. From (3) follows that the survival function of the minimum and the cumulative distribution function of $\max(Y_1, Y_2, \cdots, Y_n)$ are defined as

$$S_{\min}(t) = G_N \left[S_0(y, \beta, \gamma), \alpha \right],$$
$$F_{\max}(t) = G_N \left[F_0(y, \beta, \gamma), \alpha \right]. \qquad (4)$$

The pdf, hazard and quantile functions of the minimum or maximum of the $(Y_1, Y_2, \cdots, Y_N)$ are defined respectively as

$$f_{\min}(t) = g_N \left[S_0(Y, \beta, \gamma), \alpha \right] f_0(Y, \beta, \gamma),$$
$$f_{\max}(t) = g_N \left[F_0(Y, \beta, \gamma), \alpha \right] f_0(Y, \beta, \gamma). \qquad (5)$$

$$h_{\min}(t) = \frac{g_N \left[S_0(y) \right]}{G_N \left[S_0(y) \right]} f_0(y, \beta, \gamma),$$
$$h_{\max}(t) = \frac{g_N \left[F_0(y) \right]}{1 - G_N \left[F_0(y) \right]} f_0(y, \beta, \gamma). \qquad (6)$$

and

$$Q_{\min}(u) = Q_0 \left[1 - G_N^{-1}(1-u) \right]$$
$$Q_{\max}(u) = Q_0 \left[G_N^{-1}(u) \right]. \qquad (7)$$

where $Q_0(u)$ is the quantile function the of basic distribution of Y_i.

The maximum likelihood estimates (MLEs) of the parameters are obtained by direct maximization of the log-likelihood function, $\log L = \sum_{i=1}^n f_j(t)$, $j = \min$ or $\max$.

The advantage of this procedure is that it runs immediately using existing statistical packages such as R. The EM-algorithm can also be considered as in [6]. Large-sample inference for the parameters can be based on their MLEs and estimated standard errors, or, preferably, on the profile likelihood, the later being invariant under reparametrization and a safer guide in relatively small samples. Different approaches are via the bootstrap or via Bayesian inference.

3. Some Working Examples

Table 1 shows the pgf of N, the survival function and the density function of the minimum or maximum of N i.i.d. random variables for the distributions proposed by [1,3,4,6,7], obtained respectively by considering (4), (5) and (6), assuming $S_0(y) = \exp(-\beta y)$ as the survival function from an exponentiated random variable. However, many new distributions may be obtained by considering a composition of different $G_N(t; \alpha)$ and $S_0(y)$ functions. For instance, assuming $T = \min\{Y_i\}_{i=1}^N$,

$G_N(t; \alpha) = [\alpha t]/[1 - t(1-\alpha)]$ (the geometric pgf) and $S_0(y) = \exp\left[-(\beta y)^\gamma \right]$ (the Weibull survival function), we obtain $S_{\min}(t) = \left[\alpha e^{-(\beta t)^\gamma} \right]/\left[1 - e^{-(\beta t)^\gamma}(1-\alpha) \right]$ and $f_{\min}(t) = \left[\alpha \beta^\gamma \gamma t^{\gamma-1} e^{-(\beta t)^\gamma} \right]/\left[1 - e^{-(\beta t)^\gamma}(1-\alpha) \right]$.

Table 1. The pgf of N and survival function and density function (p.d.f) for $T = \min\{Y_i\}_{i=1}^N$ or $T^* = \max\{Y_i\}_{i=1}^N$.

Authors	$G_N(t; \alpha)$	Variable	Survival function	p.d.f
Adamidis and Loukas (1998)	Geometric $\dfrac{\alpha t}{1 - t(1-\alpha)}$	T	$\dfrac{\alpha e^{-\beta t}}{1 - e^{-\beta t}(1-\alpha)}$	$\dfrac{\alpha \beta e^{-\beta t}}{\left(1 - e^{-\beta t}(1-\alpha) \right)^2}$
Kus (2007)	Poisson $\dfrac{e^{\alpha t} - 1}{e^\alpha - 1}$	T	$\dfrac{1 - e^{\alpha e^{-\beta t}}}{1 - e^\alpha}$	$\dfrac{\alpha \beta e^{-\alpha + \alpha e^{-\beta t} - \beta t}}{1 - e^{-\alpha}}$
Tahmasbi and Rezae (2008)	Logaritmic $\dfrac{\log(1 - \alpha t)}{\log(1 - \alpha)}$	T	$\dfrac{\log(1 - \alpha e^{-\beta t})}{\log(1 - \alpha)}$	$\dfrac{\alpha \beta e^{-\beta t}}{(\alpha e^{-\beta t} - 1)\log(1 - \alpha)}$
Cancho et al. (2011)	$\dfrac{e^{\alpha t} - 1}{e^\alpha - 1}$	T^*	$\dfrac{1 - e^{-\alpha e^{-\beta t}}}{1 - e^{-\alpha}}$	$\dfrac{\alpha \beta e^{-\alpha e^{-\beta t} - \beta t}}{1 - e^{-\alpha}}$
Louzada et al. (2011)	$\dfrac{\alpha t}{1 - t(1-\alpha)}$	T^*	$\dfrac{e^{-\beta t}}{e^{-\beta t}(1-\alpha) + \alpha}$	$\dfrac{\alpha \beta e^{-\beta t}}{\left(e^{-\beta t}(1-\alpha) + \alpha \right)^2}$

We fit the five different distributions presented in **Table 1** in a real data set on the serum-reversal time (days) of 143 children contaminated with HIV by vertical transmission from the University Hospital of the Ribeiro Preto School of Medicine (Hospital das Clnicas da Faculdade de Medicina de Ribeiro Preto) from 1986 to 2001 [8]. Serum-reversal can occur in children born from mothers infected with HIV. In order to compare the distributions we consider the $-\text{LOG} = -\log L(\hat{\alpha}, \hat{\beta})$ values, the Akaike information criterion (AIC) and Bayesian information criterion (BIC). The best distribution corresponds to lower $-\text{LOG}$, AIC and BIC values. The **Table 2** shows the parameter MLEs and their corresponding

standard errors in parentheses, values of the $-\text{LOG}$, AIC and BIC. The values of AIC, BIC and $-\text{LOG}$ provide evidence in favor the CEG distribution. These results are corroborated by the fitted density functions and survival functions of the five distributions superimposed to the histogram and Kaplan-Meier curve. The **Figure 1** presents the fitted density functions on the histogram, and survival function of the EG, EP, EL, PE and CEG distributions superimposed to the data histogram and Kaplan-Meier fit, respectively. The presence of long-term survivals is very common in practice [8]. Our approach should be investigate in the long-term survival context. A possible approach is to consider the mixture model adopted by [9].

Table 2. The parameter MLEs, their corresponding standard errors in parentheses, values of the –LOG, AIC and BIC to the five fitted distributions.

| Distribution	$\hat{\alpha}$	$\hat{\beta}$	–LOG	AIC	BIC
Adamidis and Loukas (1998)	0.9998	0.0019	861.85	1727.69	1733.62
Exponential Geometric(EG)	(0.0030)	(0.0002)			
Kus (2007)	0.8072	0.002	861.85	1727.69	1733.62
Exponential Poisson(EP)	(0.0027)	(0.0001)			
Tahmasbi and Rezaei (2008)	0.3059	0.0020	861.85	1727.69	1733.62
Exponential Logarithmic(EL)	(0.0032)	(0.0001)			
Cancho *et al.* (2011)	7.0350	0.0051	813.38	1630.75	1636.68
Poisson Exponential(PE)	(0.9633)	(0.0003)			
Louzada *et al.* (2011)	0.0190	0.0083	806.37	1616.74	1622.67
Complementary Exponential	(0.0075)	(0.0007)			
Geometric(CEG)					
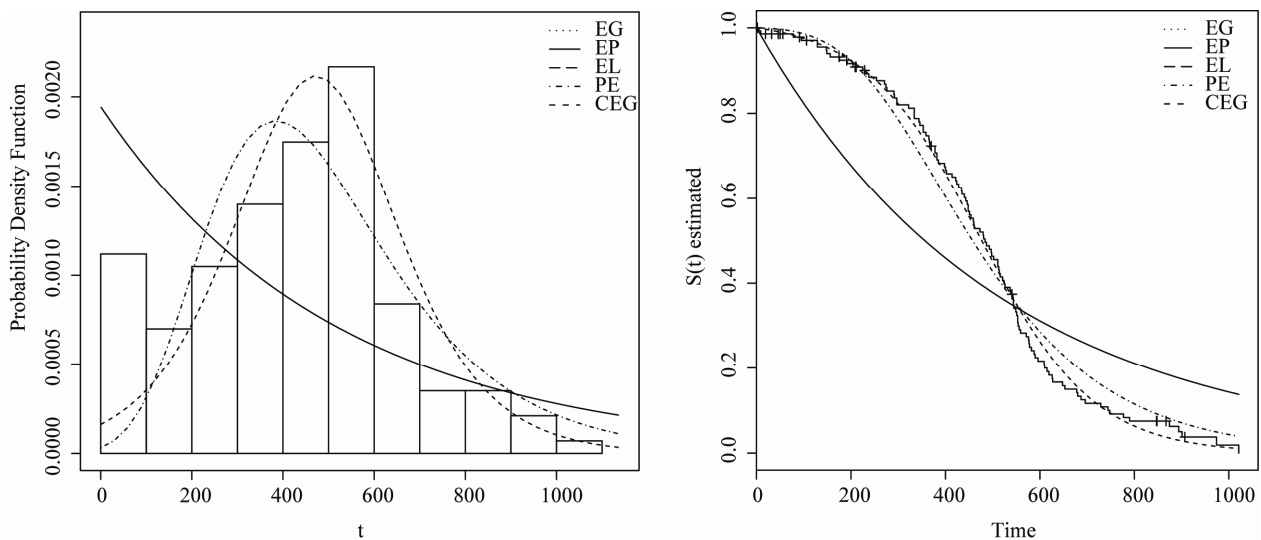
Figure 1. Fitted density functions on the histogram (left panel), and survival function (right panel) of the EP, EG, CEG,PE and El distributions superimposed to the Kaplan-Meier fit.

4. Acknowledgements

REFERENCES

[1] K. Adamidis and S. Loukas, "A Lifetime Distribution with Decreasing Failure Rate," *Statistics and Probability Letters*, Vol. 39, No. 1, 1998, pp. 35-42. doi:10.1016/S0167-7152(98)00012-1

[2] K. Adamidis, T. Dimitrakopoulou and S. Loukas, "On an Extension of the Exponential-Geometric Distribution," *Statistics and Probability Letters*, Vol. 73, No. 3, 2005, pp. 259-269. doi:10.1016/j.spl.2005.03.013

[3] C. Kus, "A New Lifetime Distribution Distributions," *Computational Statistics and Data analysis*, Vol. 11, No. 9, 2007, pp. 4497-4509. doi:10.1016/j.csda.2006.07.017

[4] R. Tahmasbi and S. Rezaei, "A Two-Parameter Life-Time Distribution with Decreasing Failure Rate," *Computational Statistics and Data Analysis*, Vol. 52, No. 8, 2008, pp. 3889-3901. doi:10.1016/j.csda.2007.12.002

[5] W. Barreto-Souza and F. Cribari-Neto, "A Generalization of the Exponential-Poisson Distribution," *Statistics and Probability Letters*, Vol. 79, No. 24, 2009, pp. 2493-2500. doi:10.1016/j.spl.2009.09.003

[6] V. Cancho, F. Louzada and G. Barriga, "The Poisson-Exponential Lifetime Distribution," *Computational Statistics and Data analysis*, Vol. 55, No. , 2011, pp. 677-686. doi:10.1016/j.csda.2010.05.033

[7] F. Louzada, M. Roman and V. Cancho, "The Complementary Exponential Geometric Distribution: Model, Properties, and a Comparison with Its Counterpart," *Computational Statistics and Data analysis*, Vol. 55, No. 8, 2011, pp. 2516-2524. doi:10.1016/j.csda.2011.02.018

[8] M. C. Hsiung, T.-H. Tung2, C.-Y. Chang, Y.-C. Chuang, S.-H. S. K-C. Lee, Y.-P. Chou, H. R, C.-M. Huang, W.-H. Y. C-C. Lin and J. W. M. S. Young, "Long-Term Survival and Prognostic Implications of Chinese Type 2 Diabetic Patients with Coronary Artery Disease after Coronary Artery Bypass Grafting," *Applied Mathematics*, Vol. 1, No. 3, 2009, pp. 139-145.

[9] G. S. C. Perdona and F. Louzada-Neto, "A General Hazard Model for Lifetime Data in the Presence of Cure Rate," *Journal of Applied Statistics*, Vol. 38, No. 7, 2011, pp. 1395-1405. doi:10.1080/02664763.2010.505948