

ORDER STATISTICS OF MULTIVARIATE ERLANG MIXTURES WITH APPLICATION IN MULTIPLE LIFETIME THEORY

GUI Wen-yong¹, ZHANG Xiang-lei², DU Xing¹, SANG Li-xin¹

(1. College of Computer Science and Artificial Intelligence, Wenzhou University, Wenzhou 325000, China)

(2. School of Mechanical and Electrical Engineering, Wenzhou University, Wenzhou 325000, China)

Abstract: In this paper, we study the order statistics of a set of dependent variables from a multivariate Erlang mixture and then apply the result to multiple lifetime theory. We derive the analytic density functions of the order statistics using the known mathematical induction method and show that any order statistic still has the form of a univariate Erlang mixture. Several important quantities in life insurance actuarial field also have explicit expressions. The result extends the result of independent random variables to dependent case.

Keywords: order statistics; multivariate Erlang mixture; multiple life theory; dependency

2010 MR Subject Classification: 62H10; 62P05

Document code: A

Article ID: 0255-7797(2022)03-0193-12

1 Introduction

[1] introduced a class of multivariate mixtures of Erlang distributions or multivariate Erlang mixtures and showed its good application in insurance. A multivariate Erlang mixture is defined as a random vector $\mathbf{X} = (X_1, X_2, \dots, X_n)$ with probability density function (pdf)

$$h(\mathbf{x}|\beta) = \sum_{m_1=1}^{\infty} \cdots \sum_{m_n=1}^{\infty} \alpha_{\mathbf{m}} \left\{ \prod_{j=1}^n \frac{\beta(\beta x_j)^{m_j-1} e^{-\beta x_j}}{(m_j-1)!} \right\}, \quad (1.1)$$

where $\mathbf{x} = (x_1, \dots, x_n)$, $\mathbf{m} = (m_1, \dots, m_n)$, $\alpha_{\mathbf{m}}$ are the mixing weights with each $\alpha_{\mathbf{m}} \geq 0$ and $\sum_{m_1=1}^{\infty} \cdots \sum_{m_n=1}^{\infty} \alpha_{\mathbf{m}} = 1$ and β is the common rate parameter ($\theta = \frac{1}{\beta}$ is called the scale parameter). The mixing weights $\alpha_{\mathbf{m}}$ can be viewed as a joint probability function of a multivariate counting random vector $\mathbf{N} = (N_1, N_2, \dots, N_n)$, that is,

$$\alpha_{\mathbf{m}} = P(N_1 = m_1, N_2 = m_2, \dots, N_n = m_n).$$

The expectation-maximum (EM) algorithm is always used to estimate the parameters of a mixture model. A standard EM algorithm and some more modified versions for parameter

* Received date: 2020-07-23

Accepted date: 2020-09-21

Foundation item: Supported by Zhejiang Provincial Natural Science Foundation (LQ17E050003).

Biography: Gui Wenyong(1988–), male, born at Yingtian, Jiangxi, lecture, major in non-life actuarial and machine learning. E-mail: 20180171@wzu.edu.cn.

estimation of Erlang mixtures can be seen in [1–3]. The class of Erlang mixtures are widely used in insurance, reliability theory and many other fields, see [4–8] and the references therein.

Let $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ be the order statistics. The use of order statistics is an extremely important subject in a wide range of statistical applications. Some primary results about the order statistics are given under the assumption that the random variables are identically distributed or independent. The studies can be seen in [9–11] and the references therein. The study on order statistics of Erlang mixtures also can be found in recent years. For example, [12] showed that the order statistics of an independent set of mixed Erlang random variables belong to the same class of Erlang mixtures. More studies can be seen in [13–15].

In this paper, we consider a set of dependent and non-identically distributed random variables with the joint distribution being a multivariate Erlang mixture. [15] derived the distributions of the minimum $X_{1:n}$ and the maximum $X_{n:n}$ and showed that both distributions belong to the class of univariate Erlang mixtures. The purpose of this paper is to generalize the results in [15] and derive the distributions of all order statistics. Furthermore, we show that the distribution of any r th ($r = 1, 2, \dots, n$) order statistic has the form of univariate Erlang mixtures.

We apply the class of multivariate Erlang mixtures to multiple lifetime area. Traditional actuarial theory of multiple life insurance often assumes independence among the future lifetimes, see, for example [16]. However, extensive research over the past years suggests otherwise, see [17, 18] and the references therein. The class of multivariate Erlang mixtures has been showed to flexibly capture the dependency among the variables, making it a reasonable choice. Another common tool used to describe the dependency in multivariate context is the copula method such as in [19]. Compared with the copula method, a multivariate Erlang mixture is more easier to deal with high dimensional data. The results in this paper show that we can get explicit expressions for some important quantities which can improve the accuracy.

This paper is organized as follows. In Section 2, we derive the density functions of the order statistics of a set of variables from a multivariate Erlang mixture and show that the order statistics are still of the form of Erlang mixtures. In Section 3, we apply the multivariate Erlang mixtures to the multiple lifetime theory and explicit results are given for some common quantities. In Section 4, a conclusion is made and some details about the method proposed in this paper are discussed.

2 Order Statistics

For notational simplicity, we denote an Erlang density with shape parameter m and rate parameter β as

$$f(x|m, \beta) = \frac{\beta(\beta x)^{m-1} e^{-\beta x}}{(m-1)!}. \quad (2.1)$$

An Erlang distribution is in fact a gamma distribution with a positive integer shape parameter. The distribution function (df) is given by

$$F(x|m, \beta) = 1 - e^{-\beta x} \sum_{n=0}^{m-1} \frac{x^n}{n!}, \quad (2.2)$$

and the survival function $\overline{F}(x|m, \beta) = 1 - F(x|m, \beta)$.

Let $f_{r:n}(x)$ and $F_{r:n}(x)$ be the density function and the distribution function of the r th order statistic in a sample of size n . It is clear that for $r = 1, \dots, n-1$, we have

$$\begin{aligned} F_{r:n}(x) - F_{r+1:n}(x) &= P(\text{exactly } r \text{ of } X\text{'s such that } X_i \leq x) \\ &= \sum_{S_r} P(X_{s_1} \leq x, \dots, X_{s_r} \leq x, X_{s_{r+1}} > x, \dots, X_{s_n} > x), \end{aligned} \quad (2.3)$$

where \sum_{S_r} denotes the sum over permutations $\{s_1, \dots, s_n\}$ of $\{1, \dots, n\}$ with $s_1 < \dots < s_r$ and $s_{r+1} < \dots < s_n$. For notational convenience, let

$$F_{[r]:n}(x) = \sum_{S_r} P(X_{s_1} \leq x, \dots, X_{s_r} \leq x, X_{s_{r+1}} > x, \dots, X_{s_n} > x),$$

and

$$f_{[r]:n}(x) = \frac{d}{dx} F_{[r]:n}(x).$$

Then we can express the density of the r th order statistic as

$$f_{r:n}(x) = f_{r+1:n}(x) + f_{[r]:n}(x). \quad (2.4)$$

In this section, we will prove that any r th ($1 \leq r \leq n$) order statistic has the form of univariate Erlang mixtures. The derivation will be a little complex. We present our main results in this section and the proof can be seen in the appendix part.

Lemma 2.1 Suppose an n -variate random vector $\mathbf{X} = (X_1, X_2, \dots, X_n)$ has joint probability density function of form (1.1), the density function $f_{[r]:n}(x)$, $r = 1, \dots, n$ can be expressed as

$$f_{[r]:n}(x) = \frac{1}{\beta^{n-1}} \sum_{m_1=1}^{\infty} \cdots \sum_{m_n=1}^{\infty} \tilde{\alpha}_{\mathbf{m}}([r], n) \prod_{i=1}^n f(x|m_i, \beta), \quad (2.5)$$

where

$$\tilde{\alpha}_{\mathbf{m}}([r], n) = \sum_{S_r} \sum_{j=1}^n \sum_{\substack{u_j=m_{s_j}, u_l < m_{s_1}, u_k \geq m_{s_k} \\ l=1, \dots, j-1, k=j+1, \dots, n}} \alpha_{\mathbf{u}} I_{\{j \leq r\}},$$

and \sum_{S_r} denotes the sum over permutations $\{s_1, \dots, s_n\}$ of $\{1, \dots, n\}$ with $s_1 < \dots < s_r, s_{r+1} < \dots < s_n$, the notation $I_{\{j \leq r\}}$ is an indicator function with a value of 1 when $j \leq r$ and otherwise 0.

Remark The density function in Lemma 2.1 also has form of univariate Erlang mixtures. However, this density function is a combination of Erlang distributions rather than a

mixture of Erlang distributions because the coefficients $\tilde{\alpha}_{\mathbf{m}}([r], n)$, $\mathbf{m} = (m_1, m_2, \dots, m_n)$, $r = 1, \dots, n$ are not all positive. The density function can be rewritten as

$$f_{[r]:n}(x) = \sum_{m=1}^{\infty} \alpha_m([r], n) f(x|m, n\beta), \quad (2.6)$$

where

$$\alpha_m([r], n) = n^{-m} \sum_{A_{m,n}} \binom{m-1}{m_1-1, m_2-1, \dots, m_n-1} \tilde{\alpha}_{\mathbf{m}}([r], n)$$

and

$$A_{m,n} = \{(m_1, \dots, m_n) | \sum_{i=1}^n m_i = m + n - 1\}.$$

Theorem 2.2 Suppose an n -variate random vector $\mathbf{X} = (X_1, X_2, \dots, X_n)$ has joint probability density function of form (1.1), the density function of the r th ($r = 1, \dots, n$) order statistic is given by

$$f_{r:n}(x) = \frac{1}{\beta^{n-1}} \sum_{m_1=1}^{\infty} \cdots \sum_{m_n=1}^{\infty} \tilde{\alpha}_{\mathbf{m}}(r, n) \prod_{i=1}^n f(x|m_i, \beta), \quad (2.7)$$

where

$$\tilde{\alpha}_{\mathbf{m}}(r, n) = \sum_{S_r} \sum_{j=1}^r \sum_{\substack{u_j=m_{s_j}, u_l < m_{s_l}, u_k \geq m_{s_k} \\ l=1, \dots, j-1, k=j+1, \dots, n}} \alpha_{\mathbf{u}}.$$

Theorem 2.3 Suppose an n -variate random vector $\mathbf{X} = (X_1, X_2, \dots, X_n)$ has joint probability density function of form (1.1), the distribution of the r th ($r = 1, \dots, n$) order statistic is a univariate Erlang mixture and the density function can be rewritten as

$$f_{r:n}(x) = \sum_{m=1}^{\infty} \alpha_m(r, n) f(x|m, n\beta), \quad (2.8)$$

where

$$\alpha_m(r, n) = n^{-m} \sum_{A_{m,n}} \binom{m-1}{m_1-1, m_2-1, \dots, m_n-1} \tilde{\alpha}_{\mathbf{m}}(r, n)$$

and

$$A_{m,n} = \{(m_1, \dots, m_n) | \sum_{i=1}^n m_i = m + n - 1\}.$$

Now we assume that the marginal random variables X_1, \dots, X_n are mutually independent. According to Corollary 2.3 in [1], the counting random variables N_1, \dots, N_n are also mutually independent. Hence, considering the relationship between the mixing weights in

Erlang mixtures and the corresponding counting random variables, the coefficients in Theorem 2.2 in the independent case can be written as

$$\begin{aligned}\tilde{\alpha}_{\mathbf{m}}(r, n) &= \sum_{S_r} \sum_{j=1}^r P(N_{s_j} = m_j, \bigcap_{k=1, k \neq j}^{r-1} N_{s_k} \leq m_k - 1, \bigcap_{k=r+1}^n N_{s_k} \geq m_k) \\ &= \sum_{S_r} \sum_{j=1}^r P(N_{s_j} = m_j) \prod_{k=1, k \neq j}^{r-1} P(N_{s_k} \leq m_k - 1) \prod_{k=r+1}^n P(N_{s_k} \geq m_k).\end{aligned}$$

[12] studied the order statistics of independent Erlang mixtures and our result is consistent with their result.

Example 1 Consider a trivariate Erlang mixture with joint density function given by

$$\begin{aligned}f(x, y, z) &= 0.2f(x|2, 1)f(y|5, 1)f(z|10, 1) + 0.3f(x|4, 1)f(y|8, 1)f(z|2, 1) \\ &\quad + 0.5f(x|1, 1)f(y|3, 1)f(z|5, 1).\end{aligned}\tag{2.9}$$

In this example, the positive mixing weights are $\alpha_{(2,5,10)} = 0.2, \alpha_{(4,8,2)} = 0.3, \alpha_{(1,3,5)} = 0.5$. The coefficients $\tilde{\alpha}_{\mathbf{m}}(r, n)$ in (2.7) are much simpler in form than the mixing weights $\alpha_{\mathbf{m}}(r, n)$ in (2.8), hence we first obtain the density functions of form (2.7) and then transform to form of univariate Erlang mixtures. Take the first order statistic for example and the parameters are given in Table 1.

Table 1 Parameters of the density function of form (2.7)

m_1	m_2	m_3	coefficient	product term	$m_1 + m_2 + m_3$
1	1	1	0.5	$f(x 1, 1)f(x 1, 1)f(x 1, 1)$	3
1	1	2	0.5	$f(x 1, 1)f(x 1, 1)f(x 2, 1)$	4
1	1	3	0.5	$f(x 1, 1)f(x 1, 1)f(x 3, 1)$	5
1	1	4	0.5	$f(x 1, 1)f(x 1, 1)f(x 4, 1)$	6
1	1	5	0.5	$f(x 1, 1)f(x 1, 1)f(x 5, 1)$	7
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
2	1	1	0.2	$f(x 1, 1)f(x 1, 1)f(x 2, 1)$	4
2	1	2	0.2	$f(x 1, 1)f(x 1, 1)f(x 2, 1)$	5
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots

From the results in Table 1, we obtain the density function of the first order statistic in the form of Erlang mixtures according to Theorem 2.3. Similarly, the parameters of the second order statistic and the third statistic are given in Table 2. The survival curves for the order statistics are shown in Figure 1.

Table 2 Parameters of density functions of all three order statistics

shape	mixing weights			rate
	first order	second order	third order	
1	0.1667	0	0	3
2	0.1667	0	0	
3	0.1667	0	0	
4	0.1395	0.0247	0	
5	0.1136	0.0473	0	
6	0.0885	0.0693	0	
7	0.0638	0.0942	0	
8	0.0379	0.1127	0	
9	0.0248	0.1078	0.0128	0.0263
10	0.0153	0.0971	0.0263	
11	0.0090	0.0840	0.0376	
12	0.0049	0.0717	0.0450	
13	0.0015	0.0625	0.0480	
14	0.0008	0.0512	0.0504	
15	0.0004	0.0418	0.0510	
\vdots	\vdots	\vdots	\vdots	

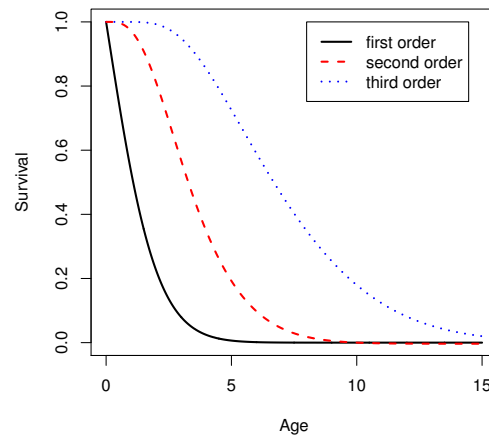


Figure 1: Survival curves for the order statistics

3 Application: Multiple Lifetime Theory

In this section, we consider the payment of an insurance benefit occurs at the moment of death. The theory for analysis of financial benefit based on the death of a single life is

well developed and the theory can be extended to the case involving several lives, see [20], [21] and the references therein. Order statistics are particularly relevant in various contexts including those involving multiple lives in life contingency.

In this section, we consider an insurance contract consisting of n dependent and non-identically distributed lives (X_1, X_2, \dots, X_n) . Denote $(x_1), (x_2), \dots, (x_n)$ be the ages of the members at the start of the contract and let $T(x_i) = (X_i - x_i | X_i > x_i), i = 1, 2, \dots, n$ be the future lifetime of (x_i) .

First, we recall some important definitions in multiple life theory. The r -survivor status of n lives $(x_1), (x_2), \dots, (x_n)$ exists at least r of the n lives survivor, denoted by $(\frac{r:n}{\mathbf{x}})$. In other words, the r -survivor status of n lives fails upon the $(n - r + 1)$ th death among the n lives. When $r = n$, the n -survivor status is also called joint-life status and $r = 1$, the 1-survivor status called the last-survivor status. Obviously, we can note that the future lifetime of r -survivor status is exactly the $(n - r + 1)$ th order statistic of the n future lifetimes $T(x_1), T(x_2), \dots, T(x_n)$. For purpose of analysis, we define another status called the $[r]$ -deferred survivor status which exists while exactly r of the n lives survive. We denote the status by $(\frac{[r]:n}{\mathbf{x}})$.

It should be noted that if n lives has pdf of form (1.1), the joint distribution of the future lifetimes $\{T(x_1), T(x_2), \dots, T(x_n)\}$ is still a multivariate Erlang mixture, see [1]. Hence, the joint pdf of future lifetimes $\{T(x_1), T(x_2), \dots, T(x_n)\}$ still has the density function of form (1.1).

The notation ${}_t p_{\frac{r:n}{\mathbf{x}}}$ represents the survival function of the future lifetime of r -survivor status. Then according to Theorem 2.3, we have

$${}_t p_{\frac{r:n}{\mathbf{x}}} = \sum_{m=1}^{\infty} \alpha_m(n-r+1, n) \int_t^{\infty} \frac{n\beta(n\beta)^{m-1}e^{-n\beta t}}{(m-1)!} dt = \sum_{m=1}^{\infty} \sum_{u=0}^{m-1} \alpha_m(n-r+1, n) \frac{(n\beta t)^u e^{-n\beta t}}{u!}. \quad (3.1)$$

Consider an insurance policy which pays a unit on the $(n - r + 1)$ th death among the n lives, the actuarial present value denoted by $\bar{A}_{\frac{r:n}{\mathbf{x}}}$ will be

$$\begin{aligned} \bar{A}_{\frac{r:n}{\mathbf{x}}} &= \int_0^{\infty} v^t {}_t p_{\frac{r:n}{\mathbf{x}}} \mu_{\frac{r:n}{\mathbf{x}}} dt = \int_0^{\infty} v^t f_{n-r+1:n}(t) dt \\ &= \int_0^{\infty} v^t \sum_{m=1}^{\infty} \alpha_m(n-r+1, n) \frac{n\beta(n\beta)^{m-1}e^{-n\beta t}}{(m-1)!} dt \\ &= \sum_{m=1}^{\infty} \alpha_m(n-r+1, n) \frac{(n\beta)^m}{(\delta + n\beta)^m}. \end{aligned} \quad (3.2)$$

where $\mu_{\frac{r:n}{\mathbf{x}}} = (-\frac{d}{dt} {}_t p_{\frac{r:n}{\mathbf{x}}}) / {}_t p_{\frac{r:n}{\mathbf{x}}}$, and δ is a constant force of interest, we have $v^t = e^{-\delta t}$. The symbols δ and v represent the same meaning in the following part of this section.

For a continuous annuity of 1 payable annually as long as at least r of the n lives survive,

the actuarial present value denoted by $\bar{a}_{\frac{r:n}{\mathbf{x}}}$ is

$$\begin{aligned}\bar{a}_{\frac{r:n}{\mathbf{x}}} &= \int_0^\infty v^t {}_t p_{\frac{r:n}{\mathbf{x}}} dt \\ &= \int_0^\infty v^t \sum_{m=1}^\infty \sum_{u=0}^{m-1} \alpha_m(n-r+1, n) \frac{(n\beta t)^u e^{-n\beta t}}{u!} dt \\ &= \sum_{m=1}^\infty \sum_{u=0}^{m-1} \alpha_m(n-r+1, n) \frac{(n\beta)^u}{(\delta + n\beta)^{u+1}}.\end{aligned}\quad (3.3)$$

Similarly, the notation ${}_t p_{\frac{[r]:n}{\mathbf{x}}}$ represents the survival function of the future lifetime of $[r]$ -deferred survivor status. According to the definition of $[r]$ -deferred survivor status, the density of the status can be seen in Lemma 2.1. Using the form of (2.6), we have

$${}_t p_{\frac{[r]:n}{\mathbf{x}}} = \sum_{m=1}^\infty \alpha_m([r], n) \int_t^\infty \frac{n\beta(n\beta)^{m-1} e^{-n\beta t}}{(m-1)!} dt = \sum_{m=1}^\infty \sum_{u=0}^{m-1} \alpha_m([r], n) \frac{(n\beta t)^u e^{-n\beta t}}{u!}. \quad (3.4)$$

Consider a continuous annuity which is payable annually as long as any of $(x_1), (x_2), \dots, (x_n)$ are alive. The annual payment is c_r when there are exactly r of the n lives survive. Then the actuarial present value is $\sum_{r=1}^n c_r \cdot \bar{a}_{\frac{[r]:n}{\mathbf{x}}}$ and the term $\bar{a}_{\frac{[r]:n}{\mathbf{x}}}$ can be calculated as

$$\begin{aligned}\bar{a}_{\frac{[r]:n}{\mathbf{x}}} &= \int_0^\infty v^t {}_t p_{\frac{[r]:n}{\mathbf{x}}} dt \\ &= \int_0^\infty v^t \sum_{m=1}^\infty \sum_{u=0}^{m-1} \alpha_m([r], n) \frac{(n\beta t)^u e^{-n\beta t}}{u!} dt \\ &= \sum_{m=1}^\infty \sum_{u=0}^{m-1} \alpha_m([r], n) \frac{(n\beta)^u}{(\delta + n\beta)^{u+1}}.\end{aligned}\quad (3.5)$$

The actuarial present value can also be calculated by the formula $\bar{A}_{\frac{[r]:n}{\mathbf{x}}} + \delta \bar{a}_{\frac{[r]:n}{\mathbf{x}}} = 1$, where

$$\begin{aligned}\bar{A}_{\frac{[r]:n}{\mathbf{x}}} &= \int_0^\infty v^t {}_t p_{\frac{[r]:n}{\mathbf{x}}} \mu_{\frac{[r]:n}{\mathbf{x}}} dt \\ &= \int_0^\infty v^t \sum_{m=1}^\infty \alpha_m([r], n) \frac{n\beta(n\beta)^{m-1} e^{-n\beta t}}{(m-1)!} dt \\ &= \sum_{m=1}^\infty \alpha_m([r], n) \frac{(n\beta)^m}{(\delta + n\beta)^m}.\end{aligned}\quad (3.6)$$

Example 2 (Example 1 continued): Suppose the future lifetimes of 3 lives $(x_1), (x_2), (x_3)$ in an insurance policy have the joint density function (2.9) shown in Example 1. We summarise some important quantities mentioned in this section in Table 3. The values in the last three columns are obtained by setting $t = 5$, $\delta = 5\%$ and the mixing weights can be seen in Table 2.

Table 3 Summarize for important quantities

quantity	formula	$r = 1$	$r = 2$	$r = 3$
$t p_{\frac{r:n}{\mathbf{x}}}$	$\sum_{m=1}^{\infty} \sum_{u=0}^{m-1} \alpha_m(n-r+1, n) \frac{(n\beta t)^u e^{-n\beta t}}{u!}$	0.7247	0.1959	0.0064
$\bar{A}_{\frac{r:n}{\mathbf{x}}}$	$\sum_{m=1}^{\infty} \alpha_m(n-r+1, n) \frac{(n\beta)^m}{(\delta + n\beta)^m}$	0.7063	0.8422	0.9371
$\bar{a}_{\frac{r:n}{\mathbf{x}}}$	$\sum_{m=1}^{\infty} \sum_{u=0}^{m-1} \alpha_m(n-r+1, n) \frac{(n\beta)^u}{(\delta + n\beta)^{u+1}}$	5.8661	3.2245	1.2572
$t p_{\frac{[r]:n}{\mathbf{x}}}$	$\sum_{m=1}^{\infty} \sum_{u=0}^{m-1} \alpha_m([r], n) \frac{(n\beta t)^m e^{-n\beta t}}{m!}$	0.5288	0.1895	0.0064
$\bar{A}_{\frac{[r]:n}{\mathbf{x}}}$	$\sum_{m=1}^{\infty} \alpha_m([r], n) \frac{(n\beta)^m}{(\delta + n\beta)^m}$	0.8621	0.9071	0.9371
$\bar{a}_{\frac{[r]:n}{\mathbf{x}}}$	$\sum_{m=1}^{\infty} \sum_{u=0}^{m-1} \alpha_m([r], n) \frac{(n\beta)^u}{(\delta + n\beta)^{u+1}}$	2.8583	1.8587	1.2572

4 Conclusion

In this paper, we have studied the order statistics for the class of multivariate Erlang mixtures. We derive the distribution of any order statistic of some dependent variables coming from multivariate Erlang mixtures without the assumption of independence. Furthermore, we have shown that the order statistics are still of the form of univariate Erlang mixtures. This desirable property enables us to deal with multivariable issues more efficiently. For this purpose, we apply the multivariate Erlang mixtures to multiple lifetime theory. One of the advantages is that we can get explicit expressions for some important quantities while numerical methods may be used to calculate the quantities for a general distribution.

Appendix

A.1 Proof for Lemma 2.1

Proof Without loss of generality, we set $S_r = \{1, 2, \dots, r, r+1, \dots, n\}$, then

$$\begin{aligned}
 F_{S_r}(x) &\triangleq P(X_1 \leq x, \dots, X_r \leq x, X_{r+1} > x, \dots, X_n \geq x) \\
 &= \sum_{m_1=1}^{\infty} \dots \sum_{m_n=1}^{\infty} \alpha_{\mathbf{m}} \prod_{i=1}^r \sum_{u_i=m_i}^{\infty} \frac{(\beta x)^{u_i} e^{-\beta x}}{u_i!} \prod_{i=r+1}^n \sum_{u_i=0}^{m_i-1} \frac{(\beta x)^{u_i} e^{-\beta x}}{u_i!} \\
 &= e^{-n\beta x} \sum_{u_1=1}^{\infty} \dots \sum_{u_r=1}^{\infty} \sum_{u_{r+1}=0}^{\infty} \dots \sum_{u_n=0}^{\infty} H_{S_r}(\mathbf{u}+1) \prod_{i=1}^n \frac{(\beta x)^{u_i}}{u_i!} \\
 &= e^{-n\beta x} \sum_{u_1=2}^{\infty} \dots \sum_{u_r=2}^{\infty} \sum_{u_{r+1}=1}^{\infty} \dots \sum_{u_n=1}^{\infty} H_{S_r}(\mathbf{u}) \prod_{i=1}^n \frac{(\beta x)^{u_i-1}}{(u_i-1)!}
 \end{aligned}$$

$$= \sum_{u_1=1}^{\infty} \cdots \sum_{u_n=1}^{\infty} H_{S_r}(\mathbf{u}) \prod_{i=1}^n \frac{f(x|u_i, \beta)}{\beta},$$

the last equation holds due to the fact $P(N_i = 0) = 0$ and

$$H_{S_r}(\mathbf{m}) = P(N_1 \leq m_1 - 1, \dots, N_r \leq m_r - 1, N_{r+1} \geq m_{r+1}, N_n \geq m_n).$$

Then take the derivative with respect to x , we have

$$f_{S_r}(x) = \sum_{m_1=1}^{\infty} \cdots \sum_{m_n=1}^{\infty} H_{S_r}(\mathbf{m}) \sum_{j=1}^n \{f(x|m_j - 1, \beta) - f(x|m_j, \beta)\} \prod_{i=1, i \neq j}^n \frac{f(x|m_i, \beta)}{\beta}.$$

Obviously, we have,

$$\begin{aligned} & \sum_{m_j=1}^{\infty} H_{S_r}(\mathbf{m}) (f(x|m_j - 1, \beta) - f(x|m_j, \beta)) \\ &= \sum_{m_j=0}^{\infty} H_{S_r}(\mathbf{m} + e_j) f(x|m_j, \beta) - \sum_{m_j=1}^{\infty} H_{S_r}(\mathbf{m}) f(x|m_j, \beta) \\ &= \sum_{m_j=1}^{\infty} [H_{S_r}(\mathbf{m} + e_j) - H_{S_r}(\mathbf{m})] f(x|m_j, \beta), \end{aligned}$$

where the notation e_j represents an n -length vector with the j th entry equals 1 and others 0.

We repeat the procedure for all summations over $S_r = \{s_1, s_2, \dots, s_n\}$ of $\{1, 2, \dots, n\}$ with conditions $s_1 < \dots < s_r$ and $s_{r+1} < \dots < s_n$ and the result holds.

A.2 Proof for Theorem 2.2

Proof We apply the mathematical deduction method.

- (1) Let $r = n$, according to the results in [15], it's obviously true.
- (2) Assume that the result holds for any $(r + 1)$ th order statistic, namely, we have

$$f_{r+1:n}(x) = \sum_{j=1}^{r+1} \sum_{m_1=1}^{\infty} \cdots \sum_{m_n=1}^{\infty} \beta [H_{r+1}(\mathbf{m} + e_j) - H_{r+1}(\mathbf{m})] \prod_{i=1}^n \frac{f(x|m_i, \beta)}{\beta},$$

where

$$H_r(\mathbf{m}) = \sum_{S_r} Pr(N_{s_1} \leq m_1 - 1, \dots, N_{s_r} \leq m_r - 1, N_{s_{r+1}} \geq m_{r+1}, N_{s_n} \geq m_n).$$

- (3) To prove that the result also holds for r th order statistic, from (2.4), we need to prove

$$\sum_{j=1}^r [H_r(\mathbf{m} + e_j) - H_r(\mathbf{m})] = \sum_{j=1}^{r+1} [H_{r+1}(\mathbf{m} + e_j) - H_{r+1}(\mathbf{m})] + \sum_{j=1}^n [H_r(\mathbf{m} + e_j) - H_r(\mathbf{m})].$$

Compare the left hand with the second term on the right hand, we further simplify the problem by proving

$$\sum_{j=r+1}^n [H_r(\mathbf{m} + e_j) - H_r(\mathbf{m})] = - \sum_{j=1}^{r+1} [H_{r+1}(\mathbf{m} + e_j) - H_{r+1}(\mathbf{m})].$$

According the definition of notation $H_r(\mathbf{m})$, we have

$$\begin{aligned} & \sum_{j=r+1}^n [H_r(\mathbf{m} + e_j) - H_r(\mathbf{m})] \\ &= - \sum_{j=r+1}^n \sum_{S_r} P(N_{s_j} = m_j, \bigcap_{i=1}^r N_{s_i} \leq m_i - 1, \bigcap_{l=r+1, l \neq j}^n N_{s_l} \geq m_l) \\ &= - \sum_{j=1}^{r+1} \sum_{S_{r+1}} P(N_{s_j} = m_j, \bigcap_{i=1, i \neq j}^{r+1} N_{s_i} \leq m_i - 1, \bigcap_{l=r+1}^n N_{s_l} \geq m_l) \\ &= - \sum_{j=1}^{r+1} [H_{r+1}(\mathbf{m} + e_j) - H_{r+1}(\mathbf{m})]. \end{aligned}$$

It means the conclusion also holds for r th order statistic, then we finish the proof.

References

- [1] Lee S C, Lin X S. Modeling dependent risks with multivariate Erlang mixtures[J]. ASTIN Bulletin: The Journal of the IAA, 2012, 42(1): 153–180.
- [2] Verbelen R, Antonio K, Claeskens G. Multivariate mixtures of Erlangs for density estimation under censoring[J]. Lifetime Data Analysis, 2016, 22(3): 429–455.
- [3] Gui Wenyong, Huang Rongtan, Lin X S. Fitting the Erlang mixture model to data via a GEM-CMM algorithm[J]. Journal of Computational and Applied Mathematics, 2018, 343: 189–205.
- [4] Yin Cuihong, Lin X S. Efficient estimation of Erlang mixtures using iSCAD penalty with insurance application[J]. ASTIN Bulletin: The Journal of the IAA, 2016, 46(3): 779–799.
- [5] Gui Wenyong, Huang Rongtan, Lin Jianhua, Lin X S. Conditional Residual Lifetimes of $(n - k + 1)$ -out-of- n Systems with Mixed Erlang Components[J]. Journal of Mathematical Study, 2017, 50(1): 1–16.
- [6] Fung T C, Badescu A L, Lin X S. Multivariate Cox Hidden Markov models with an application to operational risk[J]. Scandinavian Actuarial Journal, 2019, 8: 686–710.
- [7] Blostein M, Miljkovic T. On modeling left-truncated loss data using mixtures of distributions[J]. Insurance: Mathematics and Economics, 2019, 85: 35–46.
- [8] Virchenko Y P, Novoseltsev A D. Probability distributions unimodality of finite sample extremes of independent erlang random variables[J]. Journal of Physics Conference Series, 2020, DOI:10.1088/1742-6596/1479/1/012104.
- [9] Saran J, Pushkarna N, Tiwari R. Moment properties of generalized order statistics from lindley distribution[J]. Journal of Statistics Applications & Probability, 2015, 4: 429–434.
- [10] Malik M R, Kumar D. Relations for moments of progressively type-ii right censored order statistics from erlang-truncated exponential distribution[J]. Statistics, 2017, 18(4): 651–668.

- [11] Rashwan N I. Moments of order statistics from nonidentically Distributed Lomax, exponential Lomax and exponential Pareto Variables[J]. Journal of Advances in Mathematics, 2018, 14(1): 7431–7438.
- [12] Landriault D, Moutanabbir K, Willmo GE. A note on order statistics in the mixed Erlang case[J]. Statistics & Probability Letters, 2015, 106: 13–15.
- [13] Barakat H M, Abdelkader Y H. Computing the moments of order statistics from nonidentical random variables[J]. Statistical Methods and Applications, 2004, 13(1): 15–26.
- [14] Hlynka M, Brill PH, Horn W. A method for obtaining Laplace transforms of order statistics of Erlang random variables[J]. Statistics & Probability Letters, 2010, 80(1): 9–18.
- [15] Willmot G E, Woo J K. On some properties of a class of multivariate Erlang mixtures with insurance applications[J]. ASTIN Bulletin: The Journal of the IAA, 2015, 45(1): 151–173.
- [16] Denuit M, Frostig E. Life insurance mathematics with random life tables[J]. North American Actuarial Journal, 2009, 13(3): 339–355.
- [17] Denuit M, Cornet A. Multilife premium calculation with dependent future lifetimes[J]. Journal of Actuarial Practice, 1999, 7: 147–180.
- [18] Dufresne F, Hashorva E, Ratovomirija G, Toukourou Y. On age difference in joint lifetime modelling with life insurance annuity applications[J]. Annals of Actuarial Science, 2018, 12(2): 350–371.
- [19] Pak A, Dashti H. Distributional results for dependent Type-II Hybrid censored order statistics[J]. Pakistan Journal of Statistics and Operation Research, 2017, 13(1): 201–210.
- [20] Denuit M, Frostig E. Association and heterogeneity of insured lifetimes in the Lee-Carter framework[J]. Scandinavian Actuarial Journal, 2007, 1(1): 1–19.
- [21] Gerber H U. Life insurance mathematics[M]. New York: Springer Science & Business Media, 2013.

多维Erlang混合的次序统计量及其在多生命理论中的应用

桂文永¹, 张祥雷², 杜 兴¹, 桑立鑫¹

(1.温州大学计算机与人工智能学院, 浙江 温州 325000)

(2.温州大学机电工程学院, 浙江 温州 325000)

摘要: 本文研究来自多维Erlang混合分布的具有相关性的随机变量的次序统计量并将相关结论应用到多生命理论中. 利用熟知的数学归纳法推导出次序统计量分布密度函数的解析表达式并证明其分布为一维Erlang混合分布, 寿险精算理论中常见的重要统计量可以算出精确解. 这一结论将随机变量相互独立情形下的结果推广到变量存在相关性的情形.

关键词: 次序统计量; 多维Erlang混合; 多生命理论; 相关性

MR(2010)主题分类号: 62H10; 62P05 中图分类号: O212.1