Sankhyā B : The Indian Journal of Statistics
https://doi.org/10.1007/s13571-021-00259-w
© 2021, The Author(s)

The Sibling Distribution for Multivariate Life Time Data

Ingvild M. Helgøy and Hans J. Skaug
University of Bergen, Bergen, Norway

Abstract

A flexible class of multivariate distributions for continuous lifetimes is proposed. The distribution is defined in terms of the age-at-death of m siblings. The expression for the joint density is derived using classical results from mathematical demography. The parameters of the distribution are the age-specific birth and death rates, in addition to a vector of relative death times for the m siblings. For the case of constant birth and death rates we are able to derive an explicit expression for the bivariate sibling density, which is proven to be MTP_2, and hence has positive dependence. Further, we show that a special case of the sibling distribution belongs to the Block-Basu class of multivariate distribution. In the general case, with age-dependent birth and death rates, evaluation of the density involves numerical integration, but is still feasible.

AMS (2000) *subject classification.* Primary 62N99; Secondary 60E05.
Keywords and phrases. Copula, Frailty, Life time distribution, Mathematical demography, Close-Kin Mark-Recapture.

1 Introduction

Classes of multivariate densities for multivariate life time and survival data are well studied in the statistical and demographic literature (Hougaard, 2001; Barreto-Souza and Mayrink, 2019). A common approach for making survival times positively dependent goes via "shared frailties" (Hougaard, 2001, Chpt. 7). A frailty is a latent random variable that proportionally scales the hazard rate in a group of individuals, hence inducing dependence between otherwise independent lifetimes. In the present paper we introduce a new class of latent variable models, named the "sibling distribution", which is defined in terms of the age-at-death for each of m half siblings. There is no information available about their common mother, except that she had m offspring in total, and was alive at a specified point in time, taken to be $t = 0$ for convenience. The two latent variables of the model are the mother's birth and death times. The building blocks of the sibling distribution are

the individual birth rates $\beta(a)$ and death rates $\phi(a)$, where a is the age of an individual. We define the distribution for general functions, $\beta(a)$ and $\phi(a)$, but for most part we shall assume that $\beta(a)$ and $\phi(a)$ are constants as functions of a.

The positive dependence between the sibling's life times comes from conditioning on their (absolute) times of death, in combination with a shared dependence on the mother's life span. The times-of-death become parameters of the distribution. This somewhat implicit construction will be seen to be a mixture distribution, and can be studied using general theory for multivariate dependence (Shaked and Spizzichino, 1998; Khaledi and Kochar, 2001). We prove that the bivariate constant-rate sibling distribution is multivariate totally positive of order 2 (Karlin and Rinott, 1980), which for instance imply that the correlation is positive.

The constant-rate sibling distribution turns out to have marginal distributions that are perturbated exponential distributions. The exponential distribution plays a special role for univariate life times and several multivariate extensions can be found in the literature (Marshall and Olkin, 1967; Freund, 1961; Block and Basu, 1974; Arnold and Strauss, 1988; Gumbel, 1960; Hougaard, 1986; Sarkar, 1987). One of these extensions was introduced by Block and Basu (Block and Basu, 1974). The Block-Basu bivariate lifetime distribution can be derived by omitting the singular part of a bivariate exponential distribution as outlined by Marshall and Olkin (Marshall and Olkin, 1967), but can also be viewed as a reparametrization of Freund's distribution (Freund, 1961). We will see that the constant-rate (birth and death) sibling distribution reduces to a Block-Basu distribution, which will be used to shed light on the sibling distribution.

An alternative route to construction of multivariate life time distributions goes via copulae (Andersen, 2005). The implication also goes in the other direction; our sibling distribution induces a novel symmetric two-parameter copula.

The remaining part of the paper is organized as follows. In Section 2 we introduce the general sibling distribution. Explicit expressions in the bivariate case, along with positive dependence property, are derived in Section 3. In Section 4 we discuss the relationship to the Block-Basu distribution and in Section 5 we address simulation and parameter estimation. Finally, we provide a discussion in Section 6.

2 The Sibling Age Distribution

Consider a female who over her lifespan is known to have had m offspring. Denote by t_j and x_j the time of death and age-at-death, respectively, of the

j'th offspring. The offspring are arbitrarily ordered, not according to the time of birth. We shall view t and x as random variables taking values on the real line. For notational simplicity we let $j = 0$ refer to the mother, and we condition on the fact that the mother was alive at time $t = 0$, i.e. on the event that $t_0 - x_0 \leq 0 \leq t_0$. This assumption should be kept in mind at all times when reading this paper. We define $a_0 = x_0 - t_0$ as the age of the mother at $t = 0$, and we denote the joint density of (a_0, t_0) by $g(a_0, t_0)$. Let $y_j = t_j - x_j$, be the birth time of the *j'th* offspring. Please refer to Fig. 1 for an illustration of key quantities.

We denote random variables by capital letters. Conditionally on $(A_0, T_0) = (a_0, t_0)$, and hence on $X_0 = x_0 = a_0 + t_0$, the density of Y_j is

$$f_Y(y_j|a_0, t_0) = \frac{\beta(y_j - y_0)}{\int_0^{x_0} \beta(a)da}, \qquad y_j \in (y_0, t_0), \tag{2.1}$$

where $\beta(a)$ is the age-specific rate at which the mother produces offspring. We shorten our notation for conditional densities, e.g. we write f_Y rather than the full $f_{Y|A_0,T_0}$. The joint conditional density of (X_j, T_j) is

$$f_{X,T}(x_j, t_j|a_0, t_0) = f_Y(y_j|a_0, t_0)f_X(x_j) = f_Y(t_j - x_j|a_0, t_0)f_X(x_j), \tag{2.2}$$

on the support

$$\{(x_j, t_j) : (t_j - t_0)_+ \leq x_j \leq t_j - y_0\}, \tag{2.3}$$

where $z_+ = \max(z, 0)$. The constraints on x_j express the fact that $x_j \geq 0$ and $y_0 \leq y_j \leq t_0$, i.e. the offspring must be born in the time window when the mother is alive (see Fig. 1). The latter is related to Eq. 2.3 via the algebraic equivalence $y_0 \leq y_j \leq t_0 \;\Leftrightarrow\; t_j - t_0 \leq t_j - y_j \leq t_j - y_0$. Further, the marginal density $f_X(x_j)$ in Eq. 2.2 is defined in terms of the survival function $l(x_j) = \Pr(X > x_j)$ via $f_X(x_j) = -l'(x_j)$. Finally, the marginal density of T_j is obtained from Eq. 2.2 and Eq. 2.3,

$$f_T(t_j|a_0, t_0) = \int_{(t_j - t_0)_+}^{t_j - y_0} f_Y(t_j - x|a_0, t_0)f_X(x)dx, \qquad t_j \geq y_0. \tag{2.4}$$

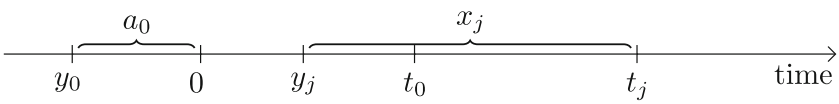

Figure 1: Birth (y_0) and death (t_0) times of mother, and corresponding times (y_j and t_j) for the j'th offspring. Further, a_0 is the age of the mother at the reference point $t = 0$

The births and deaths of the m siblings are assumed to be conditionally independent, given $(A_0, T_0) = (a_0, t_0)$, so the joint density of $X_{1:m} = (X_1, \ldots, X_m)$ and $T_{1:m} = (T_1, \ldots, T_m)$ is $\prod_{j=1}^m f_{X,T}(x_j, t_j|a_0, t_0)$, where the density $f_{X,T}(x_j, t_j|a_0, t_0)$ is given by Eq. 2.2. We are now in position to define the sibling distribution as the conditional distribution of $X_{1:m}$, given $T_{1:m} = t_{1:m}$.

Remark 1. The sibling distribution of $X_{1:m}$ has density

$$f(x_{1:m}|t_{1:m}) = \frac{\int_0^\infty \int_0^\infty \prod_{j=1}^m f_{X,T}(x_j, t_j|a_0, t_0) g(a_0, t_0) da_0 dt_0}{\int_0^\infty \int_0^\infty \prod_{j=1}^m f_T(t_j|a_0, t_0) g(a_0, t_0) da_0 dt_0}, \; x_{1:m} \in R_+^m, \tag{2.5}$$

where $f_{X,T}$ and f_T are given by Eqs. 2.2 and 2.4, respectively, and g is the joint density of (A_0, T_0) for which we will derive the density (2.7) below.

The parameters of the sibling distribution are $t_{1:m} \in R^m$, in addition to whatever parameters are hidden in the functional forms of the functions $\beta(a)$ and $l(a)$. Note that the t_j are not restricted to be positive, i.e. the offspring may have died before $t = 0$. Mostly, we shall parameterize $l(a)$ in terms of the age-specific death rate $\phi(a)$, which is related to the survival function through the well known relationship $l(x) = \exp\left(-\int_0^x \phi(a) da\right)$.

In order to derive an expression for the density $g(a_0, t_0)$, occurring in Eq. 2.5, we use the theory for stable age distributions from mathematical demography (Caswell and Keyfitz, 2005), which we now briefly review. A population in which the age-specific rates $\phi(a)$ and $\beta(a)$ do not change with time will settle into a stable age distribution. Further, the population will grow at a rate r given as the solution to the "characteristic equation"

$$\int_0^\infty \beta(a) l(a) e^{-ra} da = 1.$$

The stable age distribution has density $f_A(a) = l(a)e^{-ra}/\int_0^a l(u)e^{-ru}du$, for $a \geq 0$. Our point of view is that the mother is randomly selected among all females alive at $t = 0$, so that the density of A_0 is given by f_A. This is yet not taking into account the fact that she has m offspring over her life time. The joint density of A_0 and T_0 is

$$f_{A_0,T_0}(a_0, t_0) = f_{T_0|A_0}(t_0|a_0) f_{A_0}(a_0), \quad a_0 \geq 0, t_0 \geq 0, \tag{2.6}$$

where $f_{T_0|A_0}(t_0|a_0) = -l'(a_0 + t_0)/l(a_0)$. Conditionally on A_0 and T_0, and hence on the length of the time period $X_0 = A_0 + T_0$ that she is alive, her total number of offspring M is Poisson distributed with mean $B(x_0) = \int_0^{x_0} \beta(u) du$.

Knowing that the mother had m offspring over her lifetime perturbs the distribution (2.6) as follows

$$g(a_0, t_0) \propto f_{A_0,T_0}(a_0, t_0) B(a_0 + t_0)^m e^{-B(a_0+t_0)}. \tag{2.7}$$

We have now completely specified the sibling distribution via its density (2.5). Instead of providing results about its properties in the general case, we turn to a special case in which explicit results can be found. In Section 6 we briefly return with some discussion of the general case.

3 Constant Birth and Death Rates

We shall refer to the situation

$$\beta(a) = \beta \quad \text{and} \quad \phi(a) = \phi \quad \text{for all } a, \tag{3.1}$$

as the constant-rate sibling distribution. Under this assumption it follows that the conditional densities (2.1) and (2.2) reduce respectively to

$$f_Y(y_j|a_0, t_0) = \frac{1}{a_0 + t_0}, \qquad y_j \in (-a_0, t_0), \tag{3.2}$$

and

$$f_{X,T}(x_j, t_j|a_0, t_0) = \frac{1}{a_0 + t_0} \phi e^{-\phi x_j}, \qquad (t_j - t_0)_+ \leq x_j \leq t_j + a_0. \tag{3.3}$$

Further, the marginal density (2.4) becomes

$$f_T(t_j|a_0, t_0) = \frac{e^{-\phi t_j}}{a_0 + t_0} \begin{cases} e^{\phi t_j} - e^{-\phi a_0}, & -a_0 \leq t_j \leq t_0, \\ e^{\phi t_0} - e^{-\phi a_0}, & t_j > t_0, \end{cases} \tag{3.4}$$

and the joint density (2.7) becomes

$$g(a_0, t_0) \propto (a_0 + t_0)^m e^{-(\phi+\beta)t_0} e^{-2\beta a_0}, \qquad a_0, t_0 \geq 0. \tag{3.5}$$

Using these expressions we are able to find an explicit expression for the sibling distribution (2.5) of order $m = 2$. We shall first assume that $t_1 = t_2$ which simplifies expressions somewhat. Derivations for the case $t_1 \neq t_2$ are very similar.

Consider the distribution of the life times X_1 and X_2, given that $T_1 = T_2 = t$, where t is the common time of death of the two siblings. We have the following expression for the sibling density, which due to symmetry is presented only for $x_1 \leq x_2$.

THEOREM 1. *With constant-rates* (3.1)*, the sibling density* (2.5) *with* $m = 2$ *becomes (for* $t > 0$*):*

$$f(x_1, x_2) = C_1^{-1} \begin{cases} e^{-(\beta+\phi)(t-x_1)-\phi(x_1+x_2)}, & 0 \le x_1 \le x_2 \le t, \\ e^{2\beta(t-x_2)-(\beta+\phi)(t-x_1)-\phi(x_1+x_2)}, & 0 \le x_1 \le t \le x_2, \\ e^{2\beta(t-x_2)-\phi(x_1+x_2)}, & 0 \le t \le x_1 \le x_2, \end{cases} \tag{3.6}$$

and for $t < 0$*:*

$$f(x_1, x_2) = C_2^{-1} e^{-(2\beta+\phi)x_2-\phi x_1}, \quad t < 0 \le x_1 \le x_2, \tag{3.7}$$

where C_1 *and* C_2 *are normalizing constants.*

PROOF. See Appendix A.

By integrating over three branches in Eq. 3.6 we obtain the following expression for the normalizing constant:

$$C_1 = \begin{cases} \frac{e^{-2\phi t}(7\beta+5\phi)}{(2\beta+\phi)(\beta^2-\phi^2)} + \frac{4e^{-(\beta+2\phi)t}}{\phi(2\beta+\phi)} - \frac{2e^{-(\beta+\phi)t}}{\phi(\beta-\phi)}, & \beta \neq \phi, \\ \frac{12\phi e^{\phi t}t-7e^{\phi t}+8}{6\phi^2 e^{3\phi t}}, & \beta = \phi, \end{cases} \tag{3.8}$$

and similarly integration of Eq. 3.7 yields

$$C_2 = \left[(2\beta + \phi)(\beta + \phi)\right]^{-1}. \tag{3.9}$$

Note that $f(x_1, x_2)$ does not depend on t when $t < 0$. When we in addition set $\beta = 0$ (interpreted as a limit), X_1 and X_2 become independent, exponentially distributed. Further interpretation of the case that $t < 0$ is given in Section 4 below.

Like the exponential distribution, the constant-rate sibling distribution is closed under change of scale. If we define $X_j' = cX_j$ for $c > 0$, the parameters of the resulting sibling distribution are $\phi' = c^{-1}\phi$, $\beta' = c^{-1}\beta$ and $t' = ct$. Hence, we may set $\phi = 1$ and reparameterize the distribution in terms of (c, β, t), which for some purposes is useful.

As seen from Eq. 3.6, the density has a piecewise definition. When using symmetry to include also the case $x_1 > x_2$, the definition of the sibling density splits the first quadrant, $x_1, x_2 \ge 0$, into six regions (Fig. 2 with $t_1 = t_2$). We see that $\log\{f(x_1, x_2)\}$ is piecewise linear over these regions, and is continuous (but not differentiable) across the boundaries of the regions. The density is unimodal, with the mode at $(x_1, x_2) = (0, 0)$ when $\beta < \phi$, and while $\beta > \phi$ the mode is at $(x_1, x_2) = (t, t)$. Figure 3 shows $f(x_1, x_2)$ for three different parameter.

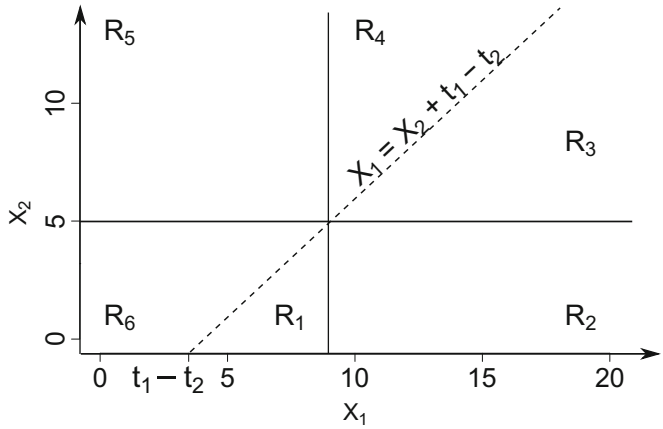

Figure 2: The six different regions of the sibling density when $t_1 = 9$ and $t_2 = 5$. The dashed line indicates ages where the siblings are born at the same time ($y_1 = y_2$)

In order to present the sibling distribution for the case $t_1 \neq t_2$, it is advantageous to introduce a general piecewise log-linear density f over the regions $R_1, \ldots, R_6$ in Fig. 2:

$$f(x_1, x_2) = C^{-1} e^{b_k + c_k x_1 + d_k x_2}, \qquad (x_1, x_2) \in R_k, \quad k = 1, \ldots, 6. \tag{3.10}$$

Here, $b_{1:6} = (b_1, \ldots, b_6)$, $c_{1:6} = (c_1, \ldots, c_6)$ and $d_{1:6} = (d_1, \ldots, d_6)$ are constants satisfying the constraints $c_2, c_3, c_4, d_3, d_4, d_5 < 0$, which are needed for

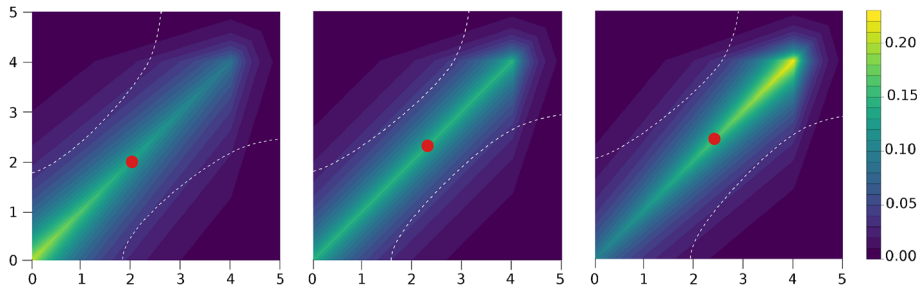

Figure 3: Bivariate sibling density $f(x_1, x_2)$ with parameters $\phi = 1$, $\beta = 0.8, 1.0, 1.2$ (left to right) and $(t_1, t_2) = (4, 4)$. The red dots show expected value (μ, μ). The dashed white curve is the contour $c = 1$ of $c(x_1, x_2)$ given by Eq. 3.15

f to be a proper density. Straightforward, but tedious, integration yields the normalization constant:

$$\begin{aligned} C = &-\frac{\mathrm{e}^{b_1}\left(\mathrm{e}^{d_1 t_2} c_1\left(\mathrm{e}^{-d_1 t_1}-\mathrm{e}^{c_1 t_1}\right)+\mathrm{e}^{c_1 t_1}\left(c_1+d_1\right)-c_1-d_1\right)}{c_1 d_1\left(c_1+d_1\right)} \\ &+\frac{\mathrm{e}^{c_2 t_1+b_2}\left(1-\mathrm{e}^{d_2 t_2}\right)}{c_2 d_2}+\frac{\mathrm{e}^{c_3 t_1+d_3 t_2+b_3}}{c_3\left(c_3+d_3\right)}+\frac{\mathrm{e}^{c_4 t_1+d_4 t_2+b_4}}{d_4\left(c_4+d_4\right)} \\ &-\frac{\mathrm{e}^{d_5 t_2+b_5}\left(\mathrm{e}^{c_5 t_1}-1\right)}{c_5 d_5}+\frac{\mathrm{e}^{d_6 t_2+b_6}\left(c_6\left(\mathrm{e}^{-d_6 t_1}-1\right)+d_6\left(\mathrm{e}^{c_6 t_1}-1\right)\right)}{c_6 d_6\left(c_6+d_6\right)}, \end{aligned} \tag{3.11}$$

where $c_1+d_1 \neq 0$ and $c_6+d_6 \neq 0$ in addition to the constraint $c_2, c_3, c_4, d_3, d_4, d_5 < 0$. The constants C_1 and C_2 in Theorem 1 are special cases of this.

It is easy to derive the moment generating function of Eq. 3.10:

$$M(s_1, s_2) = E\left(e^{s_1 X_1+s_2, X_2}\right) = \frac{C(b_{1:6}, c_{1:6}+s_1, d_{1:6}+s_2, t_1, t_2)}{C(b_{1:6}, c_{1:6}, d_{1:6}, t_1, t_2)}, \tag{3.12}$$

where $c_{1:6}+s_1 = (c_1+s_1, \ldots, c_6+s_1)$ and $d_{1:6}+s_2 = (d_1+s_2, \ldots, d_6+s_2)$. Moments of X_1 and X_2 of various orders can be obtained by repeated differentiation of $M(s_1, s_2)$ at $s_1 = s_2 = 0$. The resulting expressions are complex, and not well suited for interpretation, but are nevertheless useful for numerical evaluation.

THEOREM 2. *With constant rates* (3.1)*, the sibling density* (2.5) *with* $m = 2$ *and* $t_1 \neq t_2$ *has density given by* Eq. 3.10 *with coefficients as specified in* Table 1.

PROOF. The proof is very similar to that of Theorem 1 and is omitted.

3.1. Positive Association Intuitively, X_1 and X_2 are positively associated under the sibling distribution, due to their dependence of the lifespan

Table 1: Choice of coefficients in Eq. 3.10 yielding the order 2 sibling density when $t_1 \neq t_2$

j	b_j	c_j	d_j
1	$-(\beta+\phi)t_2$	$-\phi$	β
2	$2\beta t_1-(\beta+\phi)t_2$	$-(2\beta+\phi)$	β
3	$2\beta t_1$	$-(2\beta+\phi)$	$-\phi$
4	$2\beta t_2$	$-\phi$	$-(2\beta+\phi)$
5	$2\beta t_2-(\beta+\phi)t_1$	β	$-(2\beta+\phi)$
6	$-(\beta+\phi)t_1$	β	$-\phi$

$(-A_0, T_0)$ of their shared mother. Further, for each marginal ($j = 1, 2$) we must have that X_j is stochastically increasing in the parameter t_j. The informal argument for the latter is that since the mother is known to have been alive at $t = 0$, the larger the death time t_j the older (x_j) the individual is likely to be. We now set out to prove these claims rigorously.

We start out by proving the positive association between X_1 and X_2. An appropriate notion of positive association is the so-called Multivariate Total Positivity of order 2 (MTP_2). A bivariate density $f(x)$, $x \in R^2$, is is said to be MTP_2 if

$$f(x \vee z) f(x \wedge z) \geq f(x) f(z), \tag{3.13}$$

for any $x, z \in R^2$, where $x \vee z = (\max(x_1, z_1), \max(x_2, z_2))$ and $x \wedge z = (\min(x_1, z_1), \min(x_2, z_2))$. See Karlin and Rinott (1980) for a comprehensive overview of properties of MTP_2 distributions.

THEOREM 3. *The sibling densities* (3.6) *and* (3.7) *are MTP*$_2$.

This is proved in Appendix C using the definition Eq. 3.13 of MTP_2 directly. We believe that an alternative proof may be based on the fact that mixture distributions, of which the numerator of Eq. 2.5 is an example, under certain conditions are MTP_2 (Khaledi and Kochar, 2001; Shaked and Spizzichino, 1998). Using this approach it may be possible to prove that more general sibling distributions than (3.6) are MTP_2.

MTP_2 is a strong positive dependence property, which among other things imply that $\mathrm{cov}(X_1, X_2) \geq 0$. Although covariance is (arguably) not the most relevant dependence measure for life times, it nevertheless the most common dependence measure in general, and it is therefore useful to have establish this result.

3.2. Marginal Distribution and Copula The marginal densities in Eq. 3.6 are both given as

$$f(x) = \frac{e^{-\phi(t+x)}}{C_1} \begin{cases} e^{-\beta(t-x)} \left[\frac{\beta+\phi}{\beta\phi} - \frac{e^{-\beta x}}{\beta} - \frac{2\beta e^{\phi(x-t)}}{\phi(2\beta+\phi)} \right], & x \in (0, t], \\ e^{2\beta(t-x)} \left[\frac{\beta+\phi}{\beta\phi} - \frac{e^{-\beta t}}{\beta} - \frac{2\beta e^{\phi(t-x)}}{\phi(2\beta+\phi)} \right], & x \in (t, \infty), \end{cases} \tag{3.14}$$

where C_1 is given by Eq. 3.8. As a local measure of dependency between X_1 and X_2 we introduce

$$c(x_1, x_2) = \frac{f(x_1, x_2)}{f(x_1) f(x_2)}. \tag{3.15}$$

The $c = 1$ contour of $c(x_1, x_2)$ is displayed in Fig. 3. The region in which $c(x_1, x_2) > 1$ is located around the diagonal $x_1 = x_2$. This reflects the positive dependence in the sibling distribution.

We can obtain an analytical expression for the cumulative joint distribution $F(x_1, x_2)$ by integrating Eq. 3.6. Similarly, we get an expression for the cumulative marginal distribution function $G(x)$ by integrating Eq. 3.14. Then we can define a copula (Nelsen, 2007), $F\left(G^{-1}(x_1), G^{-1}(x_2)\right)$, based on the sibling distribution, where G^{-1} denotes the inverse of G. Because the sibling distribution is closed under change of scale, we set $\phi = 1$, and the copula thus has β and t as free parameters. We do not explore this copula further in this paper.

We next prove that X (X_1 or X_2) is stochastically increasing in t, in the sense of the following theorem.

THEOREM 4. *For any $t' > t > 0$ we have*

$$P(X > x|T = t) \leq P(X > x|T = t'). \tag{3.16}$$

It turns out to be easier to prove the more general statement that (X_1, X_2) is multivariate stochastically increasing (Shaked and Shanthikumar, 2007, p. 265), which imply Eq. 3.16. The reason this is simpler is that the key quantity, the ratio $f(x; t')/f(x; t)$, which is involved in Theorem 6.B.8. of Shaked and Shanthikumar (2007, p. 265), takes on a simpler form for the bivariate density (3.6) than for the univariate density (3.14). The details of the proof are given in Appendix D.

Stochastic monotonicity of a random variable X implies that $E(X)$ is an increasing function of t (Shaked and Shanthikumar, 2007, p. 4). This means that, for given ϕ and β, there is a one-to-one correspondence between t and $\mu = E(X)$. This fact will play a crucial role when we later devise an estimator for the parameters ϕ, β and t.

3.3. The Role of β and t Because the family of constant-rate sibling distributions is closed under change of scale we set $\phi = 1$. In this section we will study the effect of varying β and t on two characteristics: the correlation (COR) between X_1 and X_2 and $\mathrm{CV}(X_j) = \sqrt{\mathrm{Var}(X_j)}/\mathrm{E}(X_j)$. Without conditioning on $T_j = t$, we have that X_j is exponentially distributed with rate $\phi = 1$. The process of conditioning on T_j can be expected to deduce $\mathrm{CV}(X_j)$. Rather than trying to prove this formally, we provide numerical evidence.

Figure 4 shows correlation and CV as functions of β (top) and t (bottom), and indeed we see that $\mathrm{CV} \leq 1$ for all β and t. When β and t are both close to zero we have $\mathrm{CV} \approx 1$ and $\mathrm{COR} \approx 0$, which reflects the fact that X_1 and X_2 are then approximately independent and exponentially distributed. For increasing β the correlation increases, but not necessarily monotonically, and approaches an asymptotic level (top-left). From the corresponding plot of

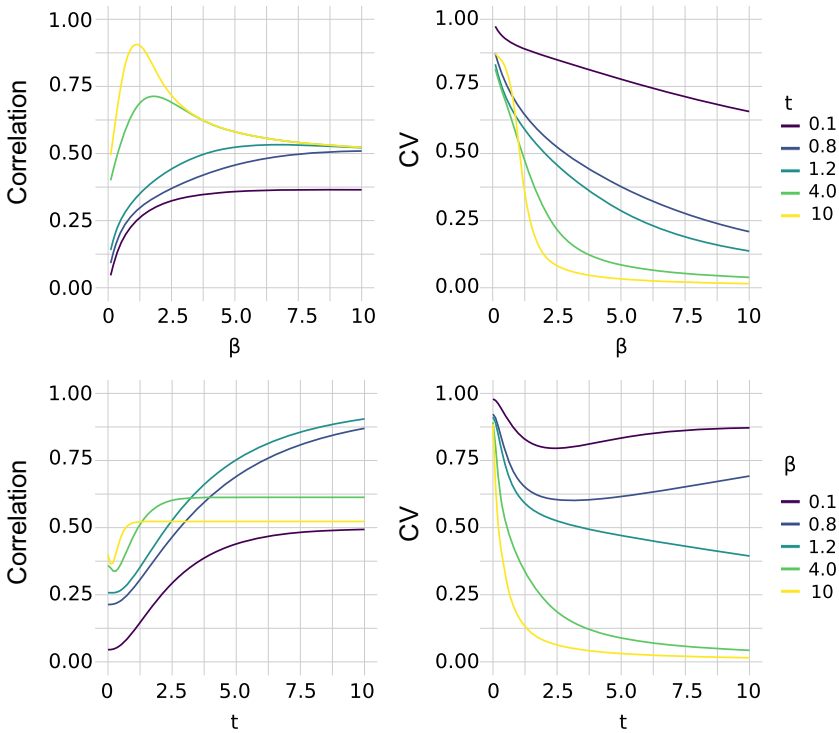

Figure 4: Correlation (left) and CV (right) of the sibling distribution ($\phi = 1$) as a function of its parameters. The parameter ϕ is set to 1 and the plots on the top row are plotted as a function of β for five different values of t. The plots on the bottom are plotted as a function of t for five different values of β

the CV (top-right) we notice that the overall trend is that the CV decreases as the value of β increases. The decrease is steepest for the highest values of t.

Further, we see from the bottom row of Fig. 4 that the correlation increases as a function of t, except for very small t. It can be shown that when $\beta = \phi$ the correlation approaches 1 as $t \to \infty$. When $\beta \neq \phi$, the correlation does not approach 1 as $t \to \infty$, but flattens out at a lower value which depends on the value of β. The CV decrease quickly as a function of t, especially for the higher values of β. When $\beta < \phi$ we see that the CV first decrease, then starts to increase for higher values of t.

4 Relationship to the Block-Basu Distribution

In this section we clarify the relationship between the constant-rate sibling distribution and the Block-Basu distribution (Block and Basu, 1974),

and we shall use this relationship to interpret the sibling distribution. One way of deriving the Block-Basu distribution goes via (Freund, 1961), and we will refer to this as the "Freund interpretation". Let now X_1 and X_2 be the lifetimes of two components assumed to be independently exponentially distributed with rate parameters α_1 and α_2, respectively. When one of the component fails, the rate for the remaining component changes from α_1 to α_1^* or from α_2 to α_2^*, depending on which component fails first. The resulting density (see Freund (1961)) is

$$f(x_1, x_2) = \begin{cases} \alpha_1 e^{-x_1(\alpha_1+\alpha_2)} \alpha_2^* e^{-\alpha_2^*(x_2-x_1)}, & 0 \leq x_1 \leq x_2, \\ \alpha_2 e^{-x_2(\alpha_1+\alpha_2)} \alpha_1^* e^{-\alpha_1^*(x_1-x_2)}. & 0 \leq x_2 \leq x_1. \end{cases} \tag{4.1}$$

When setting

$$\alpha_j = \beta + \phi, \qquad \alpha_j^* = 2\beta + \phi, \qquad j = 1, 2, \tag{4.2}$$

we see that Eq. 4.1 reduces to the symmetric sibling density (3.7) with $t \leq 0$. Since $t = 0$ is the time point at which the mother is known to have been alive, we are effectively considering a sibling distribution where both offspring are dying before their mother.

The Freund interpretation yields that $X^{(1)} = \min(X_1, X_2)$, i.e. the age of the youngest sibling, has an exponential distribution with rate parameter $\alpha_1 + \alpha_2 = 2(\beta + \phi)$. Further, the age difference between the oldest and the youngest, $X^{(2)} = \max(X_1, X_2) - \min(X_1, X_2)$, has an exponential distribution with rate parameter $\alpha^* = 2\beta + \phi$. In the following paragraphs we interpret the rates (4.2) in the context of the sibling distribution.

If we consider the case with $t = 0$, we know that the mother and her two offspring were all alive at (or just prior to) $t = 0$. The Freund interpretation requires us to look backwards in time, starting from $t = 0$. The "failure" of a component corresponds to an offspring being born. We first look at the event $X^{(1)} > x$, which can be broken into three sub events:

(i) The mother was born prior to $t = -x$, i.e. $a_0 > x$. Because the stable age distribution of A_0 is exponential with rate β, we have $P(A_0 > x) = \exp(-\beta x)$.

(ii) Both offspring were born prior to $-x$, and because we are conditioning on there being $m = 2$ siblings in total, this implies that there were no additional births in $(-x, 0)$. The latter has probability $\exp(-\beta x)$.

(iii) Both offspring survived the interval $(-x, 0)$, which has probability $\exp(-2\phi x)$.

When combining the independent events i)–iii) we get the Freund interpretation $P\left(X^{(1)} > x\right) = \exp\left[-2(\beta+\phi)x\right]$.

The event $X^{(2)} > x$ can be interpreted similarly, but we must shift our point of view backwards in time to $t = -x^{(1)}$ when the youngest sibling was born. The mother would have to be born prior to $t = -(x + x^{(1)})$. Using the stable age distribution of A_0, this has conditional probability

$$P\left(A_0 > x + x^{(1)} | A_0 > x^{(1)}\right) = \exp\left(-\beta x\right).$$

Secondly, there couldn't have been any births between $t = -\left(x + x^{(1)}\right)$ and $t = -x^{(1)}$, which has probability $\exp\left(-\beta x\right)$. Finally, the offspring that was born at $t = -x^{(2)}$ survived from $t = -\left(x + x^{(1)}\right)$ until $t = -x^{(1)}$, which has probability $\exp\left(-\phi x\right)$. In total we get $P\left(X^{(2)} > x\right) = \exp\left[-(2\beta+\phi)x\right]$, which is the Freund interpretation of the sibling distribution. Similar arguments applies to the situation $t < 0$.

Finally, we discuss a few additional insights gained from the Freund interpretation (4.2). First, the age difference between the siblings, $X_1 - X_2$ follows a Laplace distribution with rate $2\beta + \phi$. Further, note that $\alpha_j^* \to \alpha_j = \phi$ as $\beta \to 0$. Hence, X_1 and X_2 are independent in the limit $\beta \to 0$, each having an exponential distribution with rate ϕ.

5 Simulation, Estimation and Application to Real Data

We first devise an algorithm for sampling (x_1, x_2) from the density (3.6). Rather than sampling directly from Eq. 3.6, which would be feasible albeit a bit technical, we choose to go back to the definition of the sibling distribution. This involves explicitly sampling (a_0, t_0) for the mother. As a byproduct, our algorithm sheds light on the sibling distribution, through an expression for the conditional density of (A_0, T_0), given (T_1, T_2).

We also construct a hybrid moment/maximum likelihood estimator for the parameter vector $\theta = (\beta, \phi, t)$. The statistical properties of this estimator are investigated on simulated data.

5.1. Simulation The joint density of (A_0, T_0), (X_1, T_1) and (X_2, T_2) is given by

$$g(a_0, t_0) \prod_{j=1}^{2} f(x_j, t_j | a_0, t_0) = f(t_1, t_2) f(a_0, t_0 | t_1, t_2) \prod_{j=1}^{2} f(x_j | t_j, a_0, t_0), \quad (5.1)$$

where for notational simplicity we suppress subscripts on densities and range of variables in this section. The quantities on the left-hand side are given

by Eqs. 3.3 and 3.5, while the right-hand side is a generic refactoring of the joint density in terms of conditional densities. Dividing through by $f(t_1, t_2)$ in Eq. 5.1 we obtain the target distribution, and the right-hand side (5.1) suggests the following algorithm for sampling (X_1, X_2) conditionally on (T_1, T_2):

(i) Sample (A_0, T_0) from $f(a_0, t_0|t_1, t_2)$,

(ii) Using (a_0, t_0) from (i), draw X_j from $f(x_j|t_j, a_0, t_0)$, independently for $j = 1, 2$.

Step (ii) is straight forward, and is seen from Eq. 3.3 to amount to sampling from an exponential distribution with x_j constrained to a certain interval. Step (i) requires more careful consideration. We have

$$f(a_0, t_0|t_1, t_2) \propto g(a_0, t_0) \prod_{j=1}^{2} f(t_j|a_0, t_0),$$

where $f(t_j|a_0, t_0)$ is given by Eq. 3.4 and $g(a_0, t_0)$ by Eq. 3.5. We have found experimentally for $t_1 = t_2$ that the following two-step procedure works well. We start by independently drawing A_0 and T_0 from exponential distributions with rates 2β and $\min(\phi + \beta, t_1^{-1}, t_2^{-1})$, respectively. This is repeated K times to get a pre-sample $\{(A_{0k}, T_{0k}), k = 1 \ldots, K\}$, from which a single pair (A_0, T_0) is drawn with probabilities

$$p_k \propto f(a_{0k}, t_{0k}|t_1, t_2) \exp\left[2\beta a_{0k} + \min(\phi + \beta, t_1^{-1}, t_2^{-1}) t_{0k}\right].$$

5.2. Estimation Consider n observation pairs $\{(x_{1i}, x_{2i});\ i = 1, \ldots, n\}$ from Eq. 3.6. While $f(x_1, x_2)$ is continuous as a function of $\theta = (\beta, \phi, t)$, it is not differentiable in t at $t = x_1$ and $t = x_2$. This implies that the log-likelihood

$$l(\beta, \phi, t) = n^{-1} \sum_{i=1}^{n} \log\left(f(x_{1i}, x_{2i}; \beta, \phi, t)\right)$$

has $2n$ points where the derivative is not differentiable. Standard numerical optimization algorithms typically either do not use derivative information at all, or requires the objective function to be continuously differentiable in all variables. The former types of algorithms are slow and unstable, and the latter type are not directly applicable to our setting. We thus devise a special two-stage estimation algorithm.

Because X_1 and X_2 have the same marginal distribution we define the overall sample mean $\bar{x} = (2n)^{-1} \sum_{i=1}^{n} (x_{1i} + x_{2i})$. We denote by $\mu(\beta, \phi, t)$ the expectation of X_1 and X_2, and impose the constraint $\mu(\beta, \phi, t) = \bar{x}$ on

the parameter estimation problem. An analytical expression for $\mu(\beta, \phi, t)$ is given in Appendix B. The expression is complicated, but nevertheless well suited for numerical evaluation. Recall that we have proven earlier that $\mu(\beta, \phi, t)$ is increasing as a function of t.

Our estimation algorithm iterates between the following two steps:

1. For given $\hat{t}$, let $\hat{\beta}$ and $\hat{\phi}$ be the maximizer of $l(\beta, \phi, \hat{t})$.
2. For given $\hat{\beta}$ and $\hat{\phi}$, let $\hat{t}$ be the solution of the equation $\mu(\hat{\beta}, \hat{\phi}, t) = \bar{x}$.

Both 1) and 2) are solved numerically using the software TMB (Kristensen et al., 2016).

5.3. Simulation Experiment Using the algorithm of Section 5.1, we sampled $n = 1000$ observation pairs (x_1, x_2) to which the estimator of Section 5.2 was applied. This process was repeated 1000 times to assess the statistical properties of the estimator. Table 2 shows the results for 20 different parameter combinations (one row per combination). Moments characterizing the distribution of (X_1, X_2), obtained from the moment generating function (3.12), are also given in the table.

From Table 2 we see that the estimator for the parameter vector $\theta = (\beta, \phi, t)$ is overall stable with respect to the different combinations of the input parameters. More specifically, for the parameter t in the first column, we see that the mean values of the estimates are all very close to the true value of t, but they are slightly worse in the case when $\beta > \phi$. The same trend can be seen in the standard deviations, which are higher in these situations. For the parameter β we see that the mean of the estimates are more accurate for higher values of t, but we also see the trend with better predictions when $\beta > \phi$. The standard deviations are however quite stable for all combinations of the input parameters. The mean values of the estimates for the parameter ϕ are all very similar and they do not seem to be affected by the different combinations of the input parameters. The standard deviations are higher when $t = 2$, but otherwise quite similar. Figure 5 shows the marginal density (3.14) for some of the parameter combinations used in Table 2.

5.4. Application to Real Data We have presented the sibling distribution as a distribution for life times, but it may in fact be applied to any set of non-negative quantities with positive dependence. The constant-rate case is applicable only when the CV is less than one, and when $t_1 = t_2$ the marginals must be the same. We use the "twinData" dataset found in the R-package "OpenMx" (Neale et al., 2016) as an example. These are BMI measurements on twins (around age 18), but nevertheless satisfy the above mentioned restrictions (Table 3).

Table 2: Performance of the estimator $(\hat{\beta}, \hat{\phi}, \hat{t})$ with $n = 1000$ observations drawn using the algorithm in Section 5.2

$\hat{t}$				$\hat{\beta}$			$\hat{\phi}$			Moments		
	True	Mean	SD	True	Mean	SD	True	Mean	SD	E(X)	CV(X)	COR
1	2.00	2.05	0.21	0.60	0.63	0.05	1.00	1.01	0.15	1.20	0.65	0.37
2	2.00	2.03	0.15	0.80	0.83	0.05	1.00	1.01	0.15	1.21	0.61	0.40
3	2.00	2.02	0.08	1.20	1.23	0.06	1.00	1.01	0.13	1.27	0.54	0.46
4	2.00	2.00	0.07	1.40	1.43	0.06	1.00	1.00	0.13	1.30	0.51	0.48
5	4.00	4.06	0.15	0.60	0.64	0.04	1.00	1.01	0.05	1.84	0.66	0.57
6	4.00	4.03	0.10	0.80	0.83	0.04	1.00	1.00	0.05	2.02	0.61	0.62
7	4.00	4.01	0.06	1.20	1.21	0.05	1.00	1.01	0.06	2.43	0.49	0.68
8	4.00	4.00	0.05	1.40	1.41	0.05	1.00	1.00	0.07	2.62	0.43	0.70
9	6.00	6.05	0.18	0.60	0.64	0.04	1.00	1.00	0.04	2.30	0.71	0.69
10	6.00	6.03	0.12	0.80	0.83	0.04	1.00	1.00	0.04	2.73	0.63	0.75
11	6.00	6.01	0.06	1.20	1.21	0.04	1.00	1.01	0.05	3.71	0.45	0.80
12	6.00	6.00	0.05	1.40	1.40	0.05	1.00	1.00	0.06	4.16	0.37	0.81
13	8.00	8.03	0.25	0.60	0.65	0.04	1.00	1.00	0.03	2.59	0.75	0.76
14	8.00	8.02	0.14	0.80	0.83	0.04	1.00	1.00	0.04	3.32	0.66	0.82
15	8.00	8.01	0.06	1.20	1.20	0.04	1.00	1.01	0.05	5.12	0.42	0.87
16	8.00	8.00	0.05	1.40	1.40	0.04	1.00	1.01	0.05	5.87	0.32	0.86
17	10.00	9.98	0.33	0.60	0.65	0.03	1.00	1.00	0.03	2.77	0.79	0.80
18	10.00	10.00	0.15	0.80	0.83	0.03	1.00	1.00	0.03	3.81	0.69	0.87
19	10.00	10.00	0.06	1.20	1.20	0.04	1.00	1.01	0.04	6.63	0.39	0.90
20	10.00	9.99	0.05	1.40	1.39	0.05	1.00	1.00	0.05	7.69	0.28	0.89

The column "True" shows the values used in the simulations, while "Mean" and "SD" are, respectively, the average and standard deviation of $(\hat{\beta}, \hat{\phi}, \hat{t})$ across 1000 repetitions. The three rightmost columns show respectively $E(X)$, $CV(X) = \sqrt{\mathrm{Var}(X)}/E(X)$, and $\mathrm{COV}(X_1, X_2)$, all calculated using the moment generating function for the true parameter values

The dataset consists of BMI measurements for 3569 (male/female, monozygotic/dizygotic) twin pairs. The fitted sibling distribution is unable to accomodate the light left-hand tail of data (Fig. 6). The fitted density is huge perturbation of the unconditional (on T) distribution of X, which for the constant-rate case is an exponential distribution. This illustrates the flexibility of the distribution. Table 3 shows the parameters of the fitted distribution. The lack of fit is reflected in estimated moments not fitting the empirical moments very well. If we look at the estimated parameter values in Fig. 6 we see that these are "outside in the normal range", in the following sense. The expected life length of an individual is $1/\hat{\phi} = 1/0.94 = 1.12$, but

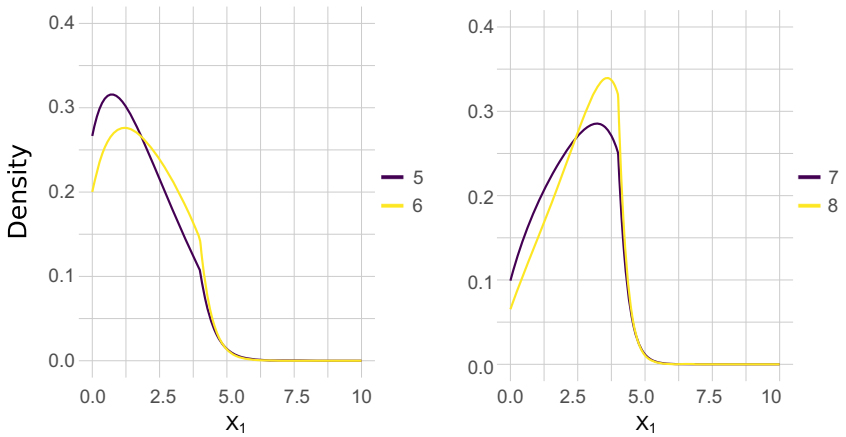

Figure 5: Marginal density (3.14) when $\beta < \phi$ (left) and when $\beta > \phi$ (right). The legend refers to the leftmost column of Table 2, which gives the parameter setting used for the different density curves

the mother-offspring duo spanned (mother's birth to offspring's death) at least $\hat{t} = 21.77$ time units.

6 Discussion

The idea of a sibling distribution was conceived during our work with the recently invented Close-Kin Mark-Recapture method (Bravington et al., 2016), in which the joint age distribution of half-siblings plays a crucial role. Its usefulness as a distribution for multivariate life time data in general remains to be explored. The fact that it is a mixture (over A_0 and T_0) of independent life times makes it amenable to analysis in the framework of Shaked and Spizzichino (1998). However, due to the conditioning on

Table 3: Fitted sibling distribution to BMI data

	Estimated	Empirical
Mean	21.79	21.77
SD	1.82	0.94
CV	0.08	0.04
COR	0.87	0.53
β	1.52	-
ϕ	0.94	-
t	23.51	-

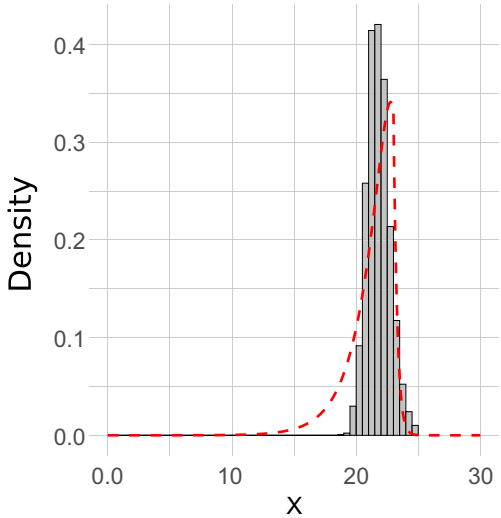

Figure 6: Marginal distribution of BMI data, with fitted sibling distribution. The dashed red curve shows the sibling density (3.6)

T_j its structure, and in particular the moments, is complicated. Moreover, the non-differentiability of the likelihood with respect to the parameter t prevents straight forward application of maximum likelihood estimation.

Our numerical experiments indicate that the constant-rate (3.1) distribution has CV ≤ 1. This is clearly limiting for a general purpose life time distribution, but this restriction can be removed by choosing a non-constant $\phi(a)$. We have not been able to obtain explicit expressions for the sibling density under more general conditions. In general, the sibling density (2.5) can be evaluated numerically. Both the numerator and denominator involves two-dimensional integrals (with respect to a_0 and t_0). The integrand of the denominator is a product of m functions on form (2.4), which each involves a one dimensional integral. This is by no means computationally prohibitive, but specially tailored numerical integration schemes would have to be devised in order for the general distribution to be practically useful.

We have shown that the constant-rate sibling distribution with $t_1 = t_2 = 0$ coincides with the exchangeable Block-Basu distribution ($\alpha_1 = \alpha_2$ and $\alpha_1^* = \alpha_2^*$). The non-exchangeable case is not a sibling distribution, as the sibling framework requires that ϕ and β are the same for both siblings. Conversely, the sibling distribution with $t_1 \neq t_2$, or $t_1 = t_2 > 0$, is not a Block-Basu distribution. Also, when allowing age-specific rates, $\phi(a)$ and $\beta(a)$, the sibling distribution is no longer a Block-Basu distribution. Kundu

and Gupta (2010) extended the Block-Basu distribution by deriving it from components that were Weibull distributed instead of exponential. The additional shape parameter makes the extended Block-Basu distribution more flexible at the cost of being less computational tractable.

Finally, we have proven that the constant-rate sibling distribution is MTP_2 and stochastically increasing in t. We conjecture, based on literature for mixture distributions (Shaked and Spizzichino, 1998, p. 273; Shaked and Shanthikumar, 2007) that these properties hold for a wider class of sibling distributions, but possibly not for all.

Acknowledgments. Parts of this work have been done in the context of CEDAS (Center for Data Science, University of Bergen, Norway).

Funding. Open access funding provided by University of Bergen (incl Haukeland University Hospital).

Open Access. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

ANDERSEN, EW (2005). Two-stage estimation in copula models used in family studies. *Lifetime Data Anal.* **11**, 333–350.

ARNOLD, BC and STRAUSS, D (1988). Bivariate distributions with exponential conditionals. *J. Am. Stat. Assoc.* **83**, 522–527.

BARRETO-SOUZA, W and MAYRINK, VD (2019). Semiparametric generalized exponential frailty model for clustered survival data. *Ann. Inst. Stat. Math.* **71**, 679–701.

BLOCK, HW and BASU, A (1974). A continuous, bivariate exponential extension. *J. Am. Stat. Assoc.* **69**, 1031–1037.

BRAVINGTON, MV, SKAUG, HJ, ANDERSON, EC et al. (2016). Close-kin mark-recapture. *Stat. Sci.* **31**, 259–274.

CASWELL, H and KEYFITZ, N (2005). *Applied mathematical demography*, 3rd edn. Springer, New York.

FREUND, JE (1961). A bivariate extension of the exponential distribution. *J. Am. Stat. Assoc.* **56**, 971–977.

GUMBEL, EJ (1960). Bivariate exponential distributions. *J. Am. Stat. Assoc.* **55**, 698–707.

HOUGAARD, P (1986). A class of multivariateate failure time distributions. *Biometrika* **73**, 671–678.

HOUGAARD, P (2001). *Analysis of multivariate survival data*. Springer, New York.

KARLIN, S and RINOTT, Y (1980). Classes of orderings of measures and related correlation inequalities. I. Multivariate totally positive distributions. *J. Multivar. Anal.* **10**, 467–498.

KHALEDI, BE and KOCHAR, S (2001). Dependence properties of multivariate mixture distributions and their applications. *Ann. Inst. Stat. Math.* **53**, 620–630.

KRISTENSEN, K, NIELSEN, A, BERG, CW, SKAUG, HJ and BELL, B (2016). TMB: Automatic differentiation and Laplace approximation. *J. Stat. Softw.* **70**, 1–21.

KUNDU, D and GUPTA, RD (2010). A class of absolutely continuous bivariate distributions. *Statistical Methodology* **7**, 464–477.

MARSHALL, AW and OLKIN, I (1967). A multivariate exponential distribution. *J. Am. Stat. Assoc.* **62**, 30–44.

NEALE, M, HUNTER, M, PRITIKIN, J, ZAHERY, M, BRICK, T, KIRKPATRICK, R, ESTABROOK, R, BATES, T, MAES, H and BOKER, S (2016). OpenMx 2.0: Extended structural equation and statistical modeling. *Psychometrika* **81**, 535–549.

NELSEN, RB (2007). An introduction to copulas. Springer Science & Business Media.

SARKAR, SK (1987). A continuous bivariate exponential distribution. *J. Am. Stat. Assoc.* **82**, 667–675.

SHAKED, M and SHANTHIKUMAR, JG (2007). Stochastic orders. Springer Science & Business Media.

SHAKED, M and SPIZZICHINO, F (1998). Positive dependence properties of conditionally independent random lifetimes. *Math. Oper. Res.* **23**, 944–959.

Publisher's Note. Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Appendix A. Proof of Theorem 1

The quantities involved in the expression (2.5) for the sibling density are given by Eqs. 3.3, 3.4 and 3.5. The evaluation of the integrals over (a_0, t_0) in the numerator and denominator of Eq. 2.5 is made difficult by the constraints (2.3). Below we show how these constraints split the first quadrant of the (x_1, x_2) plane into six disjoint regions $R_1, \ldots, R_6$ (see Fig. 2). In each of these the integrand is just a simple exponential function. Because the density is exchangeable when $t_1 = t_2$ it is sufficient to specify the expression only over the regions R_1, R_2 and R_3.

Recall that $y_j = t_j - x_j$ denotes the birth time ($j = 0, 1, 2$). The integrand of the numerator in Eq. 2.5 is

$$\begin{aligned} &f_{X,T}(x_1, t|a_0, t_0) f_{X,T}(x_2, t|a_0, t_0) g(a_0, t_0) \\ &\propto e^{-2\beta a_0} e^{-(\beta+\phi)t_0} \times \exp\left\{-\phi(x_1 + x_2)\right\}, \end{aligned} \tag{A.1}$$

for values of a_0 and t_0 such that the constraints (2.3) are satisfied for $j = 1, 2$, and zero otherwise. The term after $\times$ does not depend on a_0 and t_0, but is included for later reference. Because we assume $x_1, x_2 \geq 0$, we can replace the inequality $(t_j - t_0)_+ \leq x_j$ in Eq. 2.3 by $t - t_0 \leq x_j$, which again is equivalent to $t_0 \geq y_j$. Similarly, the last inequality $x_j \leq t_j - y_0$ in Eq. 2.3 can be re-expressed as $a_0 \geq -y_j$. Together with the basic constraints $a_0, t_0 \geq 0$ we get $a_0 \geq \max(0, -y_1, -y_2)$ and $t_0 \geq \max(0, y_1, y_2)$. Depending on the relative values of y_1 and y_2, the lower bounds of the integrals over a_0 and t_0 will be qualitatively different. There are 6 different cases, corresponding to the partition $R_1, \ldots, R_6$ of the (x_1, x_2) plane shown in Fig. 2. When integrating (A.1) with respect to a_0 and t_0, and replacing y_j by $t - x_j$ ($j = 1, 2$), we get

$$f_{X,T}(x_1, t) f_{X,T}(x_2, t) \propto e^{-\phi(x_1+x_2)} \times \begin{cases} e^{-(\beta+\phi)(t-x_1)}, & (x_1, x_2) \in R_1, \\ e^{2\beta(t-x_2)} e^{-(\beta+\phi)(t-x_1)}, & (x_1, x_2) \in R_2, \\ e^{2\beta(t-x_2)}, & (x_1, x_2) \in R_3, \end{cases}$$

which is Eq. 3.6.

The proof when $t < 0$ follows in the same vein, where we start out with the integrand of the numerator in Eq. 2.5 given by Eq. A.1. We must find values of a_0 and t_0 such that the constraints (2.3) still hold. The inequality $(t_j - t_0)_+ \leq x_j$ in Eq. 2.3 which is equivalent to $t_0 \geq y_j$ can be replaced with $t_0 \geq 0$ since we have $y_1, y_2 < 0$ in combination with the constraint $t_0 \geq 0$. For the last inequality in Eq. 2.3, the above arguments still apply such that this term is replace, with by $a_0 \geq -y_j$. In total we get $a_0 \geq \max(0, -y_1, -y_2)$ and $t_0 \geq 0$.

Appendix B. Analytical expressions

The expectation under the density (3.14) is

$$\begin{aligned} \mu(\beta, \phi, t) &= \mathrm{E}(X) \\ &= 6\beta\left(\phi - \beta/3\right)\left(\beta + \phi\right)^2\left(2\beta + \phi\right)^2 \mathrm{e}^{(-\beta+\phi)t} \\ &\quad + 4\left(\phi^2\left(t\beta - 1\right) + 2\beta^2\phi t + 2\beta^2\right)\left(-\beta + \phi\right)^2\left(\beta + \phi\right)^2 \mathrm{e}^{-t\beta} \\ &\quad + \left(4 - 10t\beta\right)\phi^6 - \left(34\beta^2 t + 4\beta\right)\phi^5 - \left(18\beta^3 t + 51\beta^2\right)\phi^4 + \left(34\beta^4 t - 7\beta^3\right)\phi^3 \\ &\quad + \left(28\beta^5 t - 23\beta^4\right)\phi^2/8\beta\left(2\beta + \phi\right)\left(-\beta + \phi\right)\phi\left(\beta + \phi\right) \\ &\quad \left(1/2\left(2\beta + \phi\right)\left(\beta + \phi\right)\mathrm{e}^{(-\beta+\phi)t} + \left(-\beta^2 + \phi^2\right)\mathrm{e}^{-t\beta} - 5/4\phi\left(\phi + 7/5\beta\right)\right). \end{aligned}$$

This expression is obtained using computer algebra system Maple.

Appendix C. Proof of Theorem 3

We shall work with $g(x) = \log f(x)$, $x \in R^2$, which for the sibling density that Eq. 3.6 is a piecewise linear function. The MTP_2 property of f is equivalent supermodularity of g, i.e.

$$g(x \vee z) + g(x \wedge z) \geq g(x) + g(z). \tag{C.1}$$

It is trivial to check that a linear function g is supermodular, so it follows directly that Eq. 3.7 is MTP_2.

The density (3.6) requires more careful attention due to its piecewise definition. It suffices to consider x and z such that the four points x, z, $(x \vee z)$ and $(x \wedge z)$ form a non-degenerate rectangle with sides parallel to the axes. This happens when either $x_1 < z_1$ and $x_2 > z_2$ or when $x_1 > z_1$ and $x_2 < z_2$. We will consider only the former, where x is the upper-left corner (and z the lower-right corner) of the rectangle. The other case is handled in the same way. When the rectangle is degenerate (a line or a point) it can be checked that Eq. C.1 holds for any function g. Under these constraints, if x and z lie in the same region $(R_1, \ldots, R_6)$ of Fig. 7 all four corners of the rectangle lies lies in same region, and by linearity of g within each region we have that Eq. C.1 is satisfied.

The sum of two supermodular functions is again supermodular, so we may add $\phi(x_1 + x_2)$ to all three branches of the logarithm of Eq. 3.6, so that we may work with

$$g(x_1, x_2) = \begin{cases} -(\beta + \phi)(t - x_1), & 0 \leq x_1 \leq x_2 \leq t, \\ 2\beta(t - x_2) - (\beta + \phi)(t - x_1), & 0 \leq x_1 \leq t \leq x_2, \\ 2\beta(t - x_2), & 0 \leq t \leq x_1 \leq x_2. \end{cases} \tag{C.2}$$

The extension of g to all six regions of Fig. 7 (with $t_1 = t_2 = t$) is $g(x_1, x_2) = g(x_2, x_2)$ when $x_1 > x_2$. The fact that $g(x_1, x_2)$ does not depend on x_2 in regions R_1 and R_4, and not on x_2 in R_3 and R_6 are visualized via the a level curve (green dashed line) of g in Fig. 7. It is seen that g is unimodal, with the mode at $(x_1, x_2) = (t, t)$.

We start out by restricting ourselves to the case $x_2 < t$. Under the facts established above, and the assumed restrictions on x and z, there are only three qualitatively different cases that must be considered. Using the red part of Fig. 7 as a reference, these are: i) $x = A$ and $z = C'$, ii) $x = A$ and $z = D'$ and iii) $x = C$ and $z = D'$. With our geometric approach, supermodularity is something that is checked for rectangles. It has the property that we can split a rectangle (A, C, C', A') in two, (A, B, B', A')

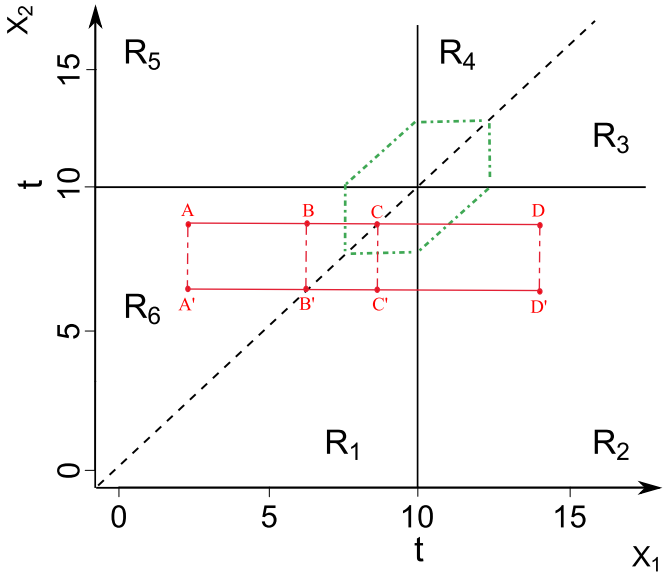

Figure 7: Part of the proof of Theorem C. The green dashed line is an example of a level curve (C.2). The red rectangles are used to prove supermodularity

and (B, C, C', B'), and it is sufficient to check Eq. C.1 for the two parts. Hence, to check all of i)–iii) it suffices to check all the red sub-rectangles in Fig. 7, which we will now do. First, (A, B, B', A') lies in a single region (R_6) so Eq. C.1 holds. For (B, C, C', B') we find by looking at the level curves of g that $g(C) > g(B)$ and $g(C') = g(B')$, which imply that Eq. C.1 holds. Finally, the (solid black) vertical line $x_1 = t$ splits (C, D, D', C') in two parts which each line entirely in R_1 and R_2, respectively. This completes the proof for $x_2 < t$.

The situation that $z_2 > t$, i.e. the red part of the figure is moved above the (solid black) horizontal line $x_2 = t$, follows by symmetry. The remaining case, $x_2 > t > z_1$, can be handled by splitting in two the rectangle horizontally at the x-axis, for each of which we know Eq. C.1 holds. Since supermodularity is also additive when splitting a rectangle horizontally, we have completed the proof.

Appendix D. Proof of Theorem 4

We prove that $f(x_1, x_2; t)$ is multivariate stochastically increasing (Shaked and Shanthikumar, 2007, Definition (6.B.1)) in the parameter t, where

$f(x_1, x_2; t)$ is given by Eq. 3.6. This implies Eq. 3.16, i.e. the marginals are also stochastically increasing (Shaked and Shanthikumar, 2007, Theorem 6.B.16 (c)).

We prove that the conditions in Theorem 6.B.8 in (Shaked and Shanthikumar, 2007) are satisfied. First, it follows from the MTP_2 property that (X_1, X_2) is "associated" in the sense of the theorem (Karlin and Rinott, 1980, Eq. (1.7)). Define the function $h(x_1, x_2) = \log\left[f(x_1, x_2; t')/f(x_1, x_2; t)\right]$. The main condition of Theorem 6.B.8 is that $h(x_1, x_2)$ is increasing in (x_1, x_2) for $t' > t$. To verify this we will check that

$$\frac{\partial h}{\partial x_1} \geq 0 \quad \text{and} \quad \frac{\partial h}{\partial x_2} \geq 0, \tag{D.1}$$

which together with the continuity of h is sufficient.

We build on the proof of Theorem 3, and denote the six regions of Fig. 7 associated with t' by $R_1', \ldots, R_6'$. We need to verify Eq. D.1 when $(x_1, x_2) \in R_j \cap R_k'$ for different values of j and k. When $j = k$ it follows that $h(x_1, x_2) = 0$ which implies Eq. D.1. Next, due to the fact that $t \leq t'$ many combinations of j and k cannot occur, and we are left with the following list to check:

(x_1, x_2)	$h(x_1, x_2)$
$R_4 \cap R_5'$	$(\phi + \beta)x_1$
$R_4 \cap R_6'$	$(\phi + \beta)x_1 + 2\beta x_2$
$R_5 \cap R_6'$	$2\beta x_2$
$R_3 \cap R_2'$	$(\phi + \beta)x_2$
$R_3 \cap R_1'$	$(\phi + \beta)x_2 + 2\beta x_1$
$R_2 \cap R_1'$	$2\beta x_1$

For all of these combinations (D.1) holds. Note that we have skipped additive terms in h that does not depend on x_1 or x_2. This completes the proof.

Ingvild M. Helgøy
Hans J. Skaug
Department of Mathematics,
University of Bergen, P.O. Box 7803,
N-5020 Bergen, Norway
E-mail: Ingvild.Helgoy@uib.no
Hans.Skaug@uib.no

Paper received: 17 September 2020; accepted 28 May 2021.