On a generalization of Clayton-Oakes model by R. L. Prentice

Hideatsu Tsukahara*

November 18, 2022

Abstract

When we have two failure times T_1 and T_2 and our primary interest is in the association between them, we need families of bivariate survivor functions for statistical modeling. Postulating some desirable properties on certain conditional hazard functions, Clayton (1978, Biometrika) derived such a family now called Clayton-Oakes model. Prentice (2016, Biometrika) proposed a generalization of the Clayton-Oakes model, but it is not clear that his generalized model gives a proper multidimensional survivor function. In this expository note, we begin with the review of the original Clayton-Oakes model and the key concept of cross ratio, and discuss the validity of Prentice's generalization.

1 Introduction

Cox and Oakes [2] list a number of desirable properties for these families:

(i) The association between T_1 and T_2 is governed by a single parameter θ which has a simple physical interpretation.

(ii) The marginal survivor functions can be specified arbitrarily and, if desired, parameterized separately from θ.

(iii) Either negative or positive association should be permissible, and the special cases of independence and the Fréchet-Hoeffding bounds are achievable within the family.

(iv) Reasonably simple parametric and semiparametric procedures are available for estimating θ, even in the presence of censoring in either or both components.

The resulting model is called the Clayton-Oakes model, which is the first semiparametric bivariate model for a pair of survival times, and it has a great advantage of clear interpretation with constant cross ratio function.

Recently Prentice [13] (also Prentice and Zhao [14]) has proposed a generalization of the Clayton-Oakes model, but it is not clear that his generalized model gives a proper multidimensional survivor function. In this expository note, we begin with the review of the original Clayton-Oakes model and the key concept of cross ratio, and discuss the validity of Prentice's generalization.

*Faculty of Economics, Seijo University, 6–1–20 Seijo, Setagaya-ku, Tokyo, 157-8511, Japan, E-mail: tsukahar@seijo.ac.jp

2 Review of Clayton-Oakes Model

Let T_1 and T_2 be two failure times with continuous joint survivor function

$$S(t_1, t_2) = \mathrm{P}(T_1 > t_1, T_2 > t_2), \quad t_1, t_2 \in \mathbb{R}_+.$$

We denote the conditional hazard function of T_1 given $T_2 = t_2$ by $\lambda_1(t_1 \,|\, T_2 = t_2)$, and the conditional hazard function of T_1 given $T_2 \geq t_2$ by $\lambda_1(t_1 \,|\, T_2 \geq t_2)$.

Clayton [1] postulates that they are proportional:

$$\lambda_1(t_1 \,|\, T_2 = t_2) = (1+\theta)\lambda_1(t_1 \,|\, T_2 \geq t_2), \quad t_1, t_2 \in \mathbb{R}_+ \tag{2.1}$$

It is straightforward to see that

$$\lambda_1(t_1 \,|\, T_2 = t_2) = -\frac{\partial}{\partial t_1} \log\left[-\frac{\partial S(t_1, t_2)}{\partial t_2}\right],$$

and

$$\lambda_1(t_1 \,|\, T_2 \geq t_2) = -\frac{\partial}{\partial t_1} \log S(t_1, t_2).$$

Plugging these into (2.1) and integrating with respect to t_1 yield

$$\log\left[-\frac{\partial S(t_1, t_2)}{\partial t_2}\right] - \log\left[-\frac{\partial S_2(t_2)}{\partial t_2}\right] = (1+\theta)[\log S(t_1, t_2) - \log S_2(t_2)].$$

Integrating with respect to t_2 after exponentiation gives

$$S(t_1, t_2)^{-\theta} - S_1(t_1)^{-\theta} = S_2(t_2)^{-\theta} - 1.$$

Thus we obtain

$$S(t_1, t_2) = \left[S_1(t_1)^{-\theta} + S_2(t_2)^{-\theta} - 1\right]^{-1/\theta}.$$

This is a model Clayton [1] first derived, and Oakes [10, 11] studied the model carefully and developed a semiparametric method of inference. Hence this model is called the *Clayton-Oakes model*.

In terms of copula model, this corresponds to a so-called (2-dimensional) Clayton copula:

$$C(u_1, u_2) = \left(u^{-\theta} + v^{-\theta} - 1\right)^{-1/\theta} \vee 0, \quad \theta \in [-1, \infty) \setminus \{0\}. \tag{2.2}$$

This family is an example of *Archimedean copulas*.

2.1 Archimedean copulas

A copula is a distribution function on $[0,1]^d$ with uniform marginals, and joins a multivariate survivor function to their one-dimensional marginals, which indicates that we can separately model univariate marginals and dependence structure. This flexibility of copula models is a great advantage in modeling multivariate event history data. Copulas are also useful in defining various concepts and measures of dependence. For generalities on copulas, consult Joe [6, 7] and Nelsen [9]. Hougaard [5] contains much broader topics in multivariate survival analysis.

Definition 2.1

(i) A nonincreasing and continuous function $\psi : [0, \infty) \to [0, 1]$ which satisfies the conditions $\psi(0) = 1$ and $\lim_{x\to\infty} \psi(x) = 0$ and is strictly decreasing on $[0, x_0)$ where $x_0 := \{x \colon \psi(x) = 0\}$ is called an *Archimedean generator*.

(ii) A d-dimensional copula C is called *Archimedean* if it has the representation

$$C(u_1,\dots,u_d) = \psi\big(\psi^{-1}(u_1)+\cdots+\psi^{-1}(u_d)\big), \quad u_1,\dots,u_d \in [0,1] \tag{2.3}$$

for some Archimedean generator ψ and its inverse $\psi^{-1}:(0,1]\to[0,\infty)$ where, by convention, $\psi^{-1}(0)=x_0$.

For C in (2.3) to be a genuine copula, the following conditions, separately for $d=2$, a fixed $d\geq 3$, all $d\geq 2$, are known to be necessary and sufficient.

Theorem 2.2 *Suppose that ψ is an Archimedean generator.*

*(i) The function $\psi\big(\psi^{-1}(u_1)+\psi^{-1}(u_2)\big)$ is a copula iff ψ is **convex**.*

*(ii) The function $\psi\big(\psi^{-1}(u_1)+\cdots+\psi^{-1}(u_d)\big)$ is a copula for a fixed d iff ψ is **d-monotone** on $[0,\infty)$*

*(iii) The function $\psi\big(\psi^{-1}(u_1)+\cdots+\psi^{-1}(u_d)\big)$ is a copula for all integer $d\geq 2$ iff ψ is **completely monotone** on $[0,\infty)$*

The following are examples of Archimedean generator.

- Clayton: $\psi(x)=(1+x)^{-1/\theta}$, $\psi^{-1}(u)=u^{-\theta}-1$
- Gumbel-Hougaard: $\psi^{-1}(u)=(-\log u)^\theta$
- Frank: $\psi^{-1}(u) = -\log\dfrac{\mathrm{e}^{\theta u}-1}{\mathrm{e}^{\theta}-1}$
- Independence: $\psi^{-1}(t)=-\log t$
- Countermonotone: $\psi^{-1}(t)=1-t$

Archimedean copulas may be related to the frailty model in survival analysis. Let W be a *frailty* (latent variable), and suppose that $T_1,\dots,T_d$ are conditionally independent given W, and that the conditional survivor function of T_i given W is given by

$$S_i(t_i|W) = \big(H_i(x_i)\big)^W \quad \text{(Lehmann alternative)},$$

where H_i is a baseline survivor function. Let $\psi(t) := \mathrm{E}[\mathrm{e}^{-tW}]$ be the Laplace transform of W. Then the marginal survivor function of T_i is given by

$$S_i(t_i) = \psi(-\log H_i(t_i)).$$

Denoting the joint survivor function $T_1,\dots,T_d$ by $S(t_1,\dots,t_d)$, we have

$$\begin{aligned} S(t_1,\dots,t_d) &= \mathrm{E}\left[\big(H_1(t_1)\cdots H_d(t_d)\big)^W\right] \\ &= \psi\left[\psi^{-1}\big(S_1(t_1)\big)+\cdots+\psi^{-1}\big(S_d(t_d)\big)\right] \end{aligned}$$

The corresponding copula is clearly Archimedean and its generator is the Laplace transform of W.

Example 2.3

- If W has a gamma distribution, then the resulting copula is Clayton.
- If W has a positive stable distribution, then the resulting copola is Gumbel-Hougaard.

2.2 Cross ratio function

Suppose that a bivariate survivor function $S(t_1, t_2)$ is absolutely continuous, and put

$$\partial_{12}S(t_1,t_2) = \frac{\partial^2 S}{\partial t_1 \partial t_2}(t_1,t_2), \quad \partial_i S(t_1,t_2) = \frac{\partial S}{\partial t_i}(t_1,t_2),\ i = 1,2.$$

Recall from (2.1) in Clayton-Oakes model, it is postulated that two hazard functions $\lambda_1(t_1 \,|\, T_2 = t_2)$ and $\lambda_1(t_1 \,|\, T_2 \geq t_2)$ are proportional. Since

$$\lambda_1(t_1 \mid T_2 = t_2) = \frac{-\partial_{12}S(t_1,t_2)}{-\partial_2 S(t_1,t_2)}, \quad \lambda_1(t_1 \mid T_2 \geq t_2) = \frac{\partial_1 S(t_1,t_2)}{S(t_1,t_2)},$$

we have

$$\frac{\lambda_1(t_1 \,|\, T_2 = t_2)}{\lambda_1(t_1 \,|\, T_2 \geq t_2)} = \frac{S(t_1,t_2)\partial_{12}S(t_1,t_2)}{\partial_1 S(t_1,t_2)\partial_2 S(t_1,t_2)}.$$

This important function of (t_1, t_2) has its own name.

Definition 2.4 For an absolutely continuous bivariate survivor function $S(t_1, t_2)$, its *cross ratio function* is defined by

$$\theta^*(t_1,t_2) := \frac{\lambda_1(t_1 \,|\, T_2 = t_2)}{\lambda_1(t_1 \,|\, T_2 > t_2)} = \frac{S(t_1,t_2)\partial_{12}S(t_1,t_2)}{\partial_1 S(t_1,t_2)\partial_2 S(t_1,t_2)}$$

Thus the Clayton-Oakes model amounts to assuming that the cross ratio function is constant (independent of (t_1, t_2)).

Furthermore, a bivariate survivor function whose associated copula is Archimedean can be characterized by a property of the cross ratio function.

Theorem 2.5 (Oakes [12]) *There exists a function θ for which $\theta^*(t_1,t_2) = \theta\big(S(t_1,t_2)\big)$ holds iff the survival copula of (T_1, T_2) is Archimedean.*

3 Prentice's Extension

3.1 Trivariate case

In Section 2 of Prentice [13], it is stated that "A trivariate generalization with unrestricted marginal and pairwise marginal survivor functions is

$$\begin{aligned} S(t_1,t_2,t_3) = \big\{&S(t_1,t_2,0)^{-\theta} + S(t_1,0,t_3)^{-\theta} + S(0,t_2,t_3)^{-\theta} \\ &- S(t_1,0,0)^{-\theta} - S(0,t_2,0)^{-\theta} - S(0,0,t_3)^{-\theta} + 1\big\}^{-1/\theta} \vee 0, \end{aligned}$$

where $-1 \leq \theta < \infty$." This is a functional equation in $S(t_1,t_2,t_3)$, and it is not clear that there exists a 3-dimensional **survivor** function $S(t_1,t_2,t_3)$ satisfying this equation. So the problem should be correctly posed in the following way.

Problem Given three bivariate sf's $S_{12}(t_1,t_2)$, $S_{13}(t_1,t_3)$, $S_{23}(t_2,t_3)$ satisfying the compatiblility conditions

$$\begin{aligned} S_{12}(t_1,0) &= S_{13}(t_1,0) =: S_1(t_1), \\ S_{12}(0,t_2) &= S_{23}(t_2,0) =: S_2(t_2), \\ S_{13}(0,t_3) &= S_{23}(0,t_3) =: S_3(t_3), \end{aligned}$$

is the function G defined by

$$G(t_1,t_2,t_3)=\{S_{12}(t_1,t_2)^{-\theta}+S_{13}(t_1,t_3)^{-\theta}+S_{23}(t_2,t_3)^{-\theta}\\ -S_1(t_1)^{-\theta}-S_2(t_2)^{-\theta}-S_3(t_3)^{-\theta}+1\}^{-1/\theta}\vee 0$$

a proper survivor function?

The problem can be stated in terms of copulas as well. Let $C_{12}(u_1,u_2)$, $C_{13}(u_1,u_3)$, $C_{23}(u_2,u_3)$ be the copula associated with $S_{12}(t_1,t_2)$, $S_{13}(t_1,t_3)$, $S_{23}(t_2,t_3)$; namely

$$\begin{aligned} S_{12}(t_1,t_2)&=C_{12}(S_1(t_1),S_2(t_2)),\\ S_{13}(t_1,t_3)&=C_{13}(S_1(t_1),S_3(t_3)),\\ S_{23}(t_2,t_3)&=C_{23}(S_2(t_2),S_3(t_3)). \end{aligned}$$

The problem is then reduced to whether the function

$$\widehat{G}(u_1,u_2,u_3)=\{C_{12}(u_1,u_2)^{-\theta}+C_{13}(u_1,u_3)^{-\theta}+C_{23}(u_2,u_3)^{-\theta}\\ -u_1^{-\theta}-u_2^{-\theta}-u_3^{-\theta}+1\}^{-1/\theta}\vee 0$$

is a (proper) copula for any given copulas C_{12}, C_{13} and C_{23}.

3.2 Counterexample

For simplicity, consider the case $\theta>0$. We have

$$\begin{aligned} \widehat{G}(u_1,u_2,1)&=\{C_{12}(u_1,u_2)^{-\theta}+C_{13}(u_1,1)^{-\theta}+C_{23}(u_2,1)^{-\theta}-u_1^{-\theta}-u_2^{-\theta}\}^{-1/\theta}\\ &=C_{12}(u_1,u_2) \end{aligned}$$

Similarly one can check that $\widehat{G}(u_1,1,u_3)=C_{13}(u_1,u_3)$ and $\widehat{G}(1,u_2,u_3)=C_{23}(u_2,u_3)$.

Let C_{12}, C_{13} and C_{23} be the Gaussian copulas with correlation coefficient $2/3$, $-1/2$ and $1/2$ respectively; S_1, S_2 and S_3 be the standard normal distribution function. For these choices, if

$$G(t_1,t_2,t_3)=\{S_{12}(t_1,t_2)^{-\theta}+S_{13}(t_1,t_3)^{-\theta}+S_{23}(t_2,t_3)^{-\theta}\\ -S_1(t_1)^{-\theta}-S_2(t_2)^{-\theta}-S_3(t_3)^{-\theta}+1\}^{-1/\theta}$$

is a proper survival function, then the corresponding correlation matrix must be

$$\begin{pmatrix} 1 & 2/3 & -1/2\\ 2/3 & 1 & 1/2\\ -1/2 & 1/2 & 1 \end{pmatrix},$$

which is not nonnegative definite. Hence $G(t_1,t_2,t_3)$ cannot be a proper survival function.

Unsolved Problem

Under what condition(s), G is a proper survival function, or $\widehat{G}$ is a proper copula?

Note For the reference, we record the following analytical conditions for a function to be a copula.

Theorem 3.1 *A function $C\colon [0,1]^d \to [0,1]$ is a copula if and only if it satisfies the following three conditions:*

- *C is **grounded**: $C(\boldsymbol{u}) = 0$ whenever at least one of the u_j's equals 0.*
- *C has **uniform marginals**: $C(1,\dots,1,u_j,1,\dots,1) = u_j$ for all $j = 1,\dots,d$.*
- *C is **d-increasing**: For all $\boldsymbol{u},\boldsymbol{v} \in [0,1]^d$ with $u_j < v_j$ for all $j = 1,\dots,d$, we have*

$$\sum (-1)^{\#\{j\colon w_j = u_j\}} C(\boldsymbol{w}) \geq 0$$

  where the sum is taken over all $\boldsymbol{w}$ such that $w_j = u_j$ or v_j for all $j = 1,\dots,d$.

It is typically hard to check whether a given function is copula using these analytical conditions.

4 Remarks

Some other future research problems are the following.

- Is it possible to extend this method of construction to general Archimedean copulas (cf. McNeil and Nešlehová [8])?
- Rigorous theoretical development including general survivor function (not necessarily absolutely continuous). See Dabrowska [3, 4].
- Estimation method for θ
- Can we incorporate covariates in the model?
- Is there any connection with nested/hierarchical Archimedean copulas (Joe [6])?

References

[1] D. G. Clayton. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, Vol. 65, pp. 141–151, 1978.

[2] D. R. Cox and D. Oakes. *Analysis of Survival Data.* Chapman & Hall, London, 1984.

[3] D. M. Dabrowska. Kaplan-meier estimate on the plane. *Annals of Statistics*, Vol. 16, pp. 1475–1489, 1988.

[4] D. M. Dabrowska. Weak convergence of a product integral dependence measure. *Scandinavian Journal of Statistics*, Vol. 23, pp. 551–580, 1996.

[5] P. Hougaard. *Analysis of Multivariate Survival Data.* Springer-Verlag, 2000.

[6] H. Joe. *Multivariate Models and Dependence Concepts.* Chapman and Hall, London, 1997.

[7] H. Joe. *Dependence Modeling with Copulas.* CRC Press, Boca Raton, 2015.

[8] A. J. McNeil and J. Nešlehová. Multivariate archimedean copulas, d-monotone functions and l_1-norm symmetric distributions. *Annals of Statistics*, Vol. 37, pp. 3059–3097, 2009.

[9] R. B. Nelsen. *An Introduction to Copulas*. Springer-Verlag, New York, second edition, 2006.

[10] D. Oakes. A model for association in bivariate survival data. *Journal of the Royal Statistical Society, Series B*, Vol. 44, pp. 414–422, 1982.

[11] D. Oakes. Semiparametric inference in a model for association in bivariate survival data. *Biometrika*, Vol. 73, pp. 353–361, 1986.

[12] D. Oakes. Bivariate survival models induced by frailties. *Journal of the American Statistical Association*, Vol. 84, pp. 487–493, 1989.

[13] R. L. Prentice. Higher dimensional clayton–oakes models for multivariate failure time data. *Biometrika*, Vol. 103, pp. 231–236, 2016.

[14] R. L. Prentice and S. Zhao. *The Statistical Analysis of Multivariate Failure Time Data*. CRC Press, 2019.