

The Cox-Aalen Models as Framework for Construction of Bivariate Probability Distributions, Universal Representation

Jerzy K. Filus

Dept. of Mathematics and Computer Science, Oakton Community College
jkfilus98@gmail.com

Lidia Z. Filus

Dept. of Mathematics, Northeastern Illinois University
L-Filus@neiu.edu

Starting with the Aalen (1989) version of Cox (1972) 'regression model' we show the method for construction of "any" joint survival function given marginal survival functions. Basically, however, we restrict ourselves to model positive stochastic dependences only with the general assumption that the underlying two marginal random variables are centered on the set of nonnegative real values. With only these assumptions we obtain nice general characterization of bivariate probability distributions that may play similar role as the copula methodology. Examples of reliability and biomedical applications are given.

Key words: Cox model, Aalen additive hazards model, construction of bivariate probability distributions given marginal distributions, "joiner" as dependence function "connecting" the marginals, general characterization of bivariate distributions, similarity to the copula methodology, reliability and biomedical applications

Introduction

Let Y be a random variable of the main interest and random variables X_1, \dots, X_k be its explanatory variables. One may (between others) interpret Y as the random life time of a technical or biological object, and the random variables X_1, \dots, X_k as stresses the object may endure.

Let, for example, Y be the residual life time of a patient who was diagnosed with a kind of cancer. If this patient was not a smoker nor systematically drinking alcohol, her/his residual life time Y_0 is considered to be the baseline, and the corresponding survival function

$S_0(y) = P(Y_0 \geq y)$ is known to be given by the expression:
 $\exp\left[-\int_0^y \lambda_0(t, \theta) dt\right]$, where $\lambda_0(t, \theta)$ is the baseline hazard rate.

In the opposite case we will assume that there are (random) measures, say X_1, \dots, X_k , of how much the person was smoking and/or drinking and/or etc ... (These might include the time length of the consumption multiplied by the corresponding approximated amount of a given substance used daily per one kilo of weight of the patient.)

Here we assume that the probability distributions of the random variables X_1, \dots, X_k are known, and if dependent we know their joint probability distribution.

Realize that in the case of one single patient one should know the given (as measured) realization (x_1, \dots, x_k) of the random vector (X_1, \dots, X_k) .

A realization would be of interest for doctors while the whole random vector distribution is more of interest for statisticians who deal with actuary investigations. The main problem is how (much) the stresses realizations x_1, \dots, x_k influence the hazard rate of the given life time Y .

It is assumed that when the stresses are involved, the baseline hazard rate $\lambda_0(y, \theta)$ will be turned into some other (similar) function, say, $\lambda(y,)$ of time y . This stochastic phenomenon of the stresses' impact may be described in two different ways:

Firstly, the stresses may influence the hazard rate $\lambda_0(y, \theta)$ "from inside" by transforming its parameter(s) θ into other value(s), say, $\theta^* = \theta(x_1, \dots, x_k)$ so that the parameter "becomes" a piecewise continuous function of the set of stresses' realizations x_1, \dots, x_k .

The resulting hazard rate $\lambda(,)$ becomes the conditional one:

$$\lambda(y, \theta(x_1, \dots, x_k)) = \lambda(y | x_1, \dots, x_k).$$

The latter hazard rate corresponds to the conditional probability density of Y , given the values x_1, \dots, x_k .

This way of influencing the hazard rate "from inside" is called the 'parameter dependence method' and was discussed, between others, in Filus and Filus (2012) and (2013).

Secondly, the description of the influence on the hazard rate by the same stresses "from outside" is very well known as the classical Cox model (1972).

In that model the baseline hazard rate $\lambda_0(y)$ preserves its parameter(s) but instead is multiplied by a factor which depends on the values x_1, \dots, x_k .

In further development of theory this factor may also become dependent on the time y .

With this modification, the Cox (1972) model for the transformed hazard rate function $\lambda(y)$ of the life time Y , given the stresses x_1, x_2, \dots, x_m , takes the very well-known form:

$$\lambda(y | x_1, x_2, \dots, x_m) = \Phi(y; x_1, x_2, \dots, x_m) \times \lambda_0(y), \quad (1)$$

where $\lambda_0(y)$ is the baseline hazard rate.

In this paper, (while preserving the physical interpretation that the random quantities X_1, X_2, \dots, X_m affect hazard rate [so the probability distribution] of the quantity Y as explanatory variables) instead of the Cox model (1) we will consider its modification i.e., the Aalen (1989) "additive hazards model" given by:

$$\lambda(y | z_1, z_2, \dots, z_k) = \lambda_0(y) + \psi(y; z_1, z_2, \dots, z_k), \quad (2)$$

The reason, that we choose the Aalen additive model over the multiplicative by Cox is a gain of the possibility to extend the dependence problem to cases when two sets of random variables, here Y and Z_1, \dots, Z_k , are mutually explanatory to each other.

In the case $k=1$ (when the random vector (Z_1, \dots, Z_k) reduces to one single random variable $Z_1 = Z$) it turns out to be easy to model the mutual dependence by construction of bivariate survival functions of the random vector (Y, Z) .

That construction results, as a very special case, with the Gumbel (1960) first exponential bivariate survival function (distribution).

Moreover, this yields a nice and universal characterization, together with the construction, of (say, *almost*)

all continuous bivariate probability distributions when both marginal distributions of Y and of Z are explicitly given in advance.

The Aalen Model Further Development

Consider the Aalen's "conditional hazard rate", as given by (2).

(In particular, $\psi (y; z_1, z_2, \dots, z_k)$ in (2) may be considered a linear function in z_1, z_2, \dots, z_k whose coefficients may be functions of time y.)

Make the following notations for the corresponding baseline and for the affected survival functions:

$$\Pr(Y_0 \geq y) = S_0(y) = \exp [- \int_0^y \lambda_0(t) dt] , \quad (3)$$

and

$$\Pr(Y \geq y) = S(y | z_1, z_2, \dots, z_k) = \exp [- \int_0^y \lambda(t | z_1, z_2, \dots, z_k) dt] , \quad (4)$$

where the hazard rates $\lambda_0(t)$ and $\lambda(t | z_1, z_2, \dots, z_k)$ are related by formula (2). (It always will be assumed that $P(Y \geq 0) = 1$)

From the Aalen additive model we have relation:

$$\begin{aligned} \lambda_0(t) &\rightarrow \lambda(t | z_1, z_2, \dots, z_k) \text{ equivalent to} \\ S_0(y) &\rightarrow S(y | z_1, z_2, \dots, z_k). \end{aligned}$$

Note that the Aalen additive version (2) of the Cox model (1) is always equivalent to multiplying the baseline probability (survival function) $S_0(y)$ by the factor

$$\exp [- \int_0^y \int_0^{z_1, \dots, z_k} \psi (t; u_1, u_2, \dots, u_k) du_1 \dots du_k dt] .$$

This factor, in a quite good sense, may be regarded as a "probability".

Pursuing in this direction one obtains the Aalen additive model in the form of the following product of probabilities:

$$\begin{aligned} S(y | z_1, z_2, \dots, z_k) &= \exp [- \int_0^y \lambda_0 (t) dt] * \exp [- \int_0^y \int_0^{z_1, \dots, z_k} \psi (t; u_1, u_2, \dots, u_k) du_1 \dots du_k dt] \\ &= S_0(y) * \exp [- \int_0^y \int_0^{z_1, \dots, z_k} \psi (t; u_1, u_2, \dots, u_k) du_1 \dots du_k dt] . \end{aligned} \quad (5)$$

As it follows from above, formula (5) represents the conditional survival function of Y, given the events

$$Z_1 \geq z_1, \dots, Z_k \geq z_k :$$

$$S(y) = S(y | z_1, z_2, \dots, z_k) = S_0(y) * \exp [- \int_0^y \int_0^{z_1, \dots, z_k} \psi(t; u_1, u_2, \dots, u_k) du_1 \dots du_k dt] \quad (5^*)$$

The factor

$$\exp [- \int_0^y \int_0^{z_1, \dots, z_k} \psi(t; u_1, u_2, \dots, u_k) du_1 \dots du_k dt]$$

we propose to call "Aalen factor" as, apparently, the latter formula (5*) is the full version of the operation:

$$S_0(y) \rightarrow S(y) \text{ as determined by the Aalen's model (2).}$$

It appears that the latter "Aalen's addition" and the equivalent Aalen's multiplication (5*) of the "probabilities" well reflects an underlying 'physical' impact of the stresses Z_1, Z_2, \dots, Z_k on the baseline life time Y_0 . Meanwhile, realize that the original "Cox multiplication":

$$\lambda(y | x_1, x_2, \dots, x_m) = \Phi(y; x_1, x_2, \dots, x_m) \times \lambda_0(y) ,$$

does not produce as nice formula for the $S_0(y) \rightarrow S(y)$ operation.

Instead of the “probabilities” multiplication (5*) the original Cox model produces the exponential formula

$$S(y | x_1, x_2, \dots, x_k) = S_0(y)^{\Phi(x_1, x_2, \dots, x_m)}.$$

In this exponential case a “*physical action*” of the quantities X_1, X_2, \dots, X_k on the quantity Y is not as tangible as with the Aalen model. So, for the original Cox model, the *physical interpretation* in terms of the probabilities is not so direct.

The Application of the Aalen’s Paradigm to Bivariate Probability Distributions Construction

The specific, (Aalen’s) multiplicative character of the conditional survival function (5*) gives one the possibility to handle mutual dependence model, when the random variable Y and the random vector (Z_1, Z_2, \dots, Z_k) are explanatory to each other.

Notice: To keep it simpler we now replace the random vector (Z_1, Z_2, \dots, Z_k) by a single random variable Z (the $k = 1$ case) that has some (arbitrary) known survival function

$$S_2(z) = \exp\left[-\int_0^z \mu_0(t) dt\right].$$

Also, for clarity we will use the notation $S_1(y)$ and $S_2(z)$ for the probabilities $P(Y \geq y)$ and $P(Z \geq z)$ respectively.

We then obtain in place of (5*) the following distribution conditioned on an event $Z \geq z$:

$$S(y) = S(y | z) = S_1(y) * \exp\left[-\int_0^y \int_0^z \psi(t; u) du dt\right] \quad (5^{**})$$

Now, multiplying (5**) by $S_2(z) = \Pr(Z \geq z)$ one obtains the joint survival function of the random vector (Y, Z) in the form:

$$P(Y \geq y, Z \geq z) = S_1(y) S_2(z) \exp\left[-\int_0^y \int_0^z \psi(t; u) du dt\right] \quad (6)$$

Express the latter equality by:

$$P(Y \geq y, Z \geq z) = \exp\left[-\int_0^y \lambda_0(t) dt - \alpha(y, z) - \int_0^z \mu_0(t) dt\right], \quad (6^*)$$

where the expression $\alpha(y, z) = \int_0^y \int_0^z \psi(t; u) du dt$ analytically describes the mutual stochastic dependence between the random variables Y, Z .

When $\alpha(y, z) \rightarrow 0$, the variables Y, Z “become” independent.

We (mostly) will assume that $\psi(t; u) \geq 0$ for all t, u (the positive dependence).

Example: As an underlying example illustrating the stochastic dependence model (6*) one may consider the reliability of two interacting components in a series system whose dependent life times are these Y, Z .

At first, we are given the independent (baseline) life times Y_0, Z_0 , of the components estimated separately in laboratory conditions, so that the corresponding baseline components failure (hazard) rates are $\lambda_0(y), \mu_0(z)$. When put into the system, the components “start” to interact physically affecting each other. So the side effects of the operating process of each working component impacts changes in the inner physical structure of the other component.

Here the

“(micro-shocks \rightarrow micro-damages) \rightarrow (micro-hazard rates changes) and their cumulation” pattern may be employed, see for example Filus and Filus (2008). These (micro)-changes may, for example, accelerate

degradation (or adaptation) or other physical (chemical) processes resulting in the affected component ‘statistically earlier’ failure. Mathematically this means that the presence of the other component at a given instantaneous time epoch t causes the component’s failure (hazard) rate to be higher (or lower, but this case needs a separate treatment) than the original, by an “infinitesimal value”, say $\psi(t; u) dt du$.

Compare with the ‘Aalen scheme’.

Now, in particular, the components “in system” life times Y, Z may be jointly distributed according to (6*) i.e.,

$$P(Y \geq y, Z \geq z) = \exp[-\int_0^y \lambda_0(t) dt - \alpha(y, z) - \int_0^z \mu_0(t) dt],$$

with

$$\alpha(y, z) = \int_0^y \int_0^z \psi(t; u) du dt.$$

Other examples of tandems of two, mutually affecting random quantities Y, Z may be taken from a bio-medical investigations. Y, Z , may be considered as levels of some chemicals

(hormones, in particular) in a human or an animal body.

Also, an example may provide the pair ‘pulse rate’ Y and the accompanying ‘blood pressure’ Z .

On *Universal Representation of Bivariate Survival Functions*

The Representation

Now we will depart slightly from the Aalen model toward more general paradigms. Thus we may drop the assumption on existence of the hazard rates $\lambda_0(y), \mu_0(z)$, and the function $\alpha(y, z)$ need not anymore be expressed by the integral $\int_0^y \int_0^z \psi(t; u) du dt$.

Instead of formulas (6) and (6*) for the joint survival function of (Y, Z) we may use the following, more general, formula:

$$S(y, z) = S_1(y) S_2(z) \exp[-\alpha(y, z)], \tag{7}$$

where $\alpha(y, z)$ is only assumed to be a nonnegative measurable real function, defined for $y \geq 0, z \geq 0$.

However, for simplicity, we specify the subclass of the functions $\alpha(y, z)$ as satisfying the following three conditions. So we consider any $\alpha(y, z)$ to be:

- (1) continuous with respect to y and with respect to z ,
- (2) nondecreasing with respect to both variables,
- (3) $\alpha(0, z) = \alpha(y, 0) = 0$.

From the last condition it directly follows:

Property 1. If the above condition (3) is satisfied then both the marginal probability distributions of the joint distribution given by formula (7) are preserved in the sense that they are the same as the baseline distributions $S_1(y), S_2(z)$ originally given. \square

If the marginal and the joint densities of the considered throughout random variables Y, Z exist, the concern is on nonnegativity of them. While the marginal densities, say, $f(y), g(z)$ are nonnegative due to nonnegativity of the corresponding hazard rates $\lambda_0(y), \mu_0(z)$ we need some special condition to assure nonnegativity of the joint density $h(y, z)$.

Thus, first of all, if the joint density $h(y, z)$ exists the function

$\alpha(y, z)$ must have continuous partial derivatives of first order $\alpha_y(y, z), \alpha_z(y, z)$ and the continuous second

order mixed partial derivative $\alpha_{y,z}(y, z) = \psi(y, z)$.

In this case, the above conditions (1), (2), (3) for the $\alpha(y, z)$ are not always satisfactory in order to “connect” [to the joint distribution] two survival functions $S_1(y)$, $S_2(z)$ that possess the hazard rates $\lambda_0(y)$, $\mu_0(z)$, respectively.

The additional condition, that must be satisfied too, has the form of the following inequality:

$$[\lambda_0(y) + \alpha_y(y, z)] \times [(\mu_0(z) + \alpha_z(y, z))] \geq \alpha_{y,z}(y, z). \quad (8)$$

This follows from the form of the joint density:

$$h(y, z) = \partial^2 / \partial y \partial z S(y, z) = \\ \{ [\lambda_0(y) + \alpha_y(y, z)] \times [(\mu_0(z) + \alpha_z(y, z)) - \alpha_{y,z}(y, z) \} \exp[- \Lambda_1(y) - \Lambda_2(z) - \alpha(y, z)]$$

which must be nonnegative.

Here, $d / dy \Lambda_1(y) = \lambda_0(y)$, and $d / dz \Lambda_2(z) = \mu_0(z)$ and $\alpha_{y,z}(y, z) = \psi(y, z)$.

A simpler condition than (8) for the nonnegativity of $h(y, z)$, that is satisfactory too (but not necessary) for the existence of the joint survival function (6*) is the following obtained from (8):

$$\lambda_0(y) \mu_0(z) \geq \alpha_{y,z}(y, z). \quad (8^*)$$

The condition (8) or (8*) together with the conditions (1), (2), (3) are satisfactory for connecting two survival functions $S_1(y)$, $S_2(z)$ by a given joiner $\alpha(y, z)$ into the bivariate model $S(y, z)$. As it was pointed out above, there is a huge number of such models.

Particular Cases of the Bivariate Models

A. Obviously, when $\alpha(y, z)$ reduces to zero for all y, z then model (7) describes independent random variables.

B. If the baseline hazard rates exist and are constant, say λ_0, μ_0 , we call this model “exponential”.

In this case we may choose as the ‘dependence function’ $\alpha(y, z) = ayz$ ($a > 0$) to obtain the following special case of model (7):

$$S(y, z) = \exp [- \lambda_0 y - ayz - \mu_0 z] \quad (9) \\ (a \leq \lambda_0 \mu_0),$$

which, apparently, represents the first bivariate exponential Gumbel (1960) probability distribution.

C. One also obtains the following ‘Weibullian version’ of the above bivariate Gumbel:

$$S(y, z) = \exp [- \lambda_0 y^{\gamma_1} - ay^{\gamma_1} z^{\gamma_2} - \mu_0 z^{\gamma_2}], \quad (10)$$

where γ_1 and γ_2 are positive reals.

The models (9) and (10) are special cases of the representation (7).

Remark 2. As for above Weibullian version (10) of the Gumbel exponential model (where γ_1 and γ_2 are positive reals and $a \leq \lambda_0 \mu_0$) it can be extended to the following one:

$$S(y, z) = \exp [- \lambda_0 y^{\gamma_1} - ay^{\delta_1} z^{\delta_2} - \mu_0 z^{\gamma_2}] \quad (11)$$

However, in order to satisfy the requirement (8), we must restrict it by imposing the conditions:

$$0 < \delta_i \leq \gamma_i \text{ (for } i = 1, 2 \text{)}.$$

Remark 3. The form of model (7) and its Property 1 allows us to construct a huge number of bivariate probability distributions from (7). For this realize that:

1. For any given fixed ‘dependence function’ $\alpha(y, z)$, ‘any’ pair of two probability distributions [not necessarily from the same class], determined by $S_1(y)$, $S_2(z)$ (so that inequality (8) or (8*) is satisfied), may be “connected” by formula (7) to “become” the bivariate survival

(distribution) function in which they remain the marginals.

2. For any fixed pair of probability distributions, given by $S_1(y)$, $S_2(z)$, there is a wide class of “possible” dependence functions $\alpha(y, z)$ [only determined by the conditions (1), (2), (3) together with inequality (8) or (8*) from section 4.1], so that the two given distributions can be “connected” into the bivariate model (7) in as many ways as there are possible proper functions: $\alpha(y, z)$. \square

We propose to call the dependence function $\alpha(y, z)$ “joiner”.

Similarity to the Copula Concept

Recall we restrict our analysis to the probability distributions centered on nonnegative real values only. At this point, notice that the possibility to construct a wide class of joint survival functions, given any single pair $S_1(y)$, $S_2(z)$ of marginal survival functions resembles the copula methodology for bivariate probability distributions construction see (1959). This is, however, a “different story” as neither (6*) nor (7) representations are copula representations. To obtain a proposition that is somehow equivalent to the Sklar’s theorem (1959) for the copula representations, one may characterize the class of (all “reasonable”) bivariate survival functions $S(y, z)$ as given by representations (7) , i.e., by the class of all proper joiners $\alpha(y, z)$ and corresponding marginals so that either (8) or (8*) holds.

Restating somewhat Remark 2 we obtain the following two propositions:

Proposition 1. Suppose we are given any continuous joint survival function, say, $S(y, z)$ with its continuous marginals $S_1(y)$, $S_2(z)$, then the corresponding unique “joiner representation” (7) for $S(y, z)$ exists and is given by a unique continuous function $\alpha(y, z)$. \square

The next proposition is somewhat inverse to Proposition 1. Although it is not necessary, but for better clarity, in the following proposition we will assume that for all the corresponding survival functions the probability densities exist.

Proposition 2. For any pair of survival functions $S_1(y)$, $S_2(z)$ and any function $\alpha(y, z)$, satisfying the conditions (1)-(3) from section 4.1 (and satisfying the inequality (8) or (8*) with respect to the given survival functions $S_1(y)$, $S_2(z)$), there is unique joint survival function $S(y, z)$ with $S_1(y)$, $S_2(z)$ as marginals, and which can be written in the form (7). \square

For the proof of Proposition 1 just set

$$\alpha(y, z) = -\log \{ S(y, z) / S_1(y)S_2(z) \}$$

The latter short proof exhibits the arbitrariness of the joint distribution $S(y, z)$ to have representation (7). Proposition 2 is obvious. \square

The advantage of the above joiner representation (7) over the copula representation mainly lies in the fact that the joiner $\alpha(y, z)$ (especially when expressed in terms of the hazard rates (6*) within the “Aalen paradigm”) usually has (or may have) a direct physical or any other ‘real life’ interpretation, while most of the copulas are given by rather purely mathematical relationships without such references to modeled realities. Thus, finding a copula for a given problem in applications is usually not easier than finding the joint

distribution itself.

The important fact is that for the joiner representations, when given by (6*), there exists a straight relation between this universal representation of bivariate probability distributions and

the Aalen's additive formula (2) (with $k = 1$ and $z_1 = z$) for the underlying stochastic dependence pattern. Recall, the Aalen model is the modification of the model by Cox (1972).

Remark 4. It seems quite possible also to "connect" any two known, or newly constructed, stochastic processes, say Y_t and Z_t by a "joiner function" $\alpha_t(y_t, z_t)$ [satisfying inequality (8) for each t] which should be a (continuous in the continuous time case) function of time t . In particular, a constant function might be a proper one in many cases. In such a way new bivariate stochastic processes, say (Y_t, Z_t) could be constructed. \square

Extension of the above ideas for higher dimensions should not be difficult.

References

- Aalen, O. O., A Linear Regression Model for the Analysis of the Life Times, *Statistics in Medicine*, 8, 907-925, 1989.
- Cox, D. R., Regression Models and Life Tables (with discussion), *Journal of the Royal Statistical Society, B*, 74, 187-220, 1972.
- Filus J. K, Filus L. Z. , A method for multivariate probability distributions construction via parameter dependence, *Communications in Statistics: Theory and Methods*, Vol. 42, Num. 4, 716-721, 2013.
- Filus J. K, Filus L. Z. , Multivariate "pseudodistributions" as natural extension of the multivariate normal density pattern – Theory, *AIP Conference Proceedings*, Vol. 1479 (1) American Institute of Physics – Sep 26, 2012.
- Filus J. K, Filus L. Z. , On Multi-Component System Reliability with 'Microshocks – Microdamages' Type of Components Interaction, *Proceedings of the International Multi Conference of Engineers and Computer Scientists*, 2008, Vol. 2 IMECS, 19 – 21, March, 2008, Hong Kong.
- Gumbel E. J., Bivariate Exponential Distributions, *Journal of the American Statistical Association*, 55, 698- 707, 1960.
- Sklar A., Fonctions de repartition a n dimensions et leurs marges, *Publications de l'Institut de Statistique de l'Universite de Paris*, 8: 229-231, 1959.