Phase-Type Distributions for Competing Risks

Bo Henry Lindqvist
Norwegian University of Science and Technology, Trondheim, NORWAY
e-mail: bo@math.ntnu.no

Abstract

We extend the phase-type methodology for modeling of lifetime distributions to the case of competing risks. This is done by considering finite state Markov chains in continuous time with more than one absorbing state, letting each absorbing state correspond to a particular risk. We study statistical estimation from (possibly censored) competing risks data modeled by the phase-type approach. Using results from the literature we consider estimation via the EM algorithm as well as Bayesian estimation using Markov chain Monte Carlo methods. Treatment of covariates in competing risks data is also be discussed.

Key Words: censored data, EM algorithm, Markov chain, Markov chain Monte Carlo

1. Introduction

Phase-type distributions represent the time to absorption for a finite state Markov chain in continuous time. The simplest examples are mixtures and convolutions of exponential distributions. The class of phase-type distributions is both flexible and conceptually simple to work with. For a comprehensive introduction we refer to Neuts (1981). A shorter and very useful introduction is given by Aalen (1995). It is interesting to note that it means essentially no loss of generality to work with phase-type distributions, since the class of phase-type distributions (with the number of states taking any finite value) is dense. Hence any lifetime distribution can, at least in principle, be approximated arbitrarily close by a phase-type distribution. Multivariate phase-type distributions can likewise be defined. They possess many of the properties of the univariate phase-type distributions (Assaf et al., 1984).

Phase-type distributions have received much attention in applied probability, in particular queuing theory. Here they generalize the celebrated Erlang distribution. Phase-type distributions are also much used in reliability theory and also in medical statistics. The problem of fitting phase-type distributions to lifetime data has been considered by several authors. One approach is via the EM (expectation-maximization) algorithm, with the possibility of having right-censored or interval-censored observations. There are also Bayesian approaches in the literature, based on Markov chain Monte Carlo methods. We return to these approaches later.

The purpose of the present paper is to extend the phase-type methodology to the case of competing risks. The basic ingredient in a competing risks phase-type model is a finite state Markov chain in continuous time with more than one absorbing state, where each absorbing state corresponds to a particular risk. Expressions for cause specific hazard functions, cumulative incidence functions etc. can now be given in terms of the transition matrix of the underlying Markov chain. Special structures like Coxian models may still be studied in the competing risks framework. Of particular interest is model estimation from (possibly censored) competing risks data. It will in the paper be shown how estimation via the EM algorithm can be performed also in the case of several absorbing states. Likewise, existing methods for Bayesian estimation may be extended. An attempt to treat covariates in competing risks data will also be discussed. In fact, this raises several interesting questions on how to model the influence of covariates in competing risks models in a phase-type setting.

2. Phase-type modeling for competing risks

A phase-type distribution can be described in terms of a Markov process $\{X(t); t \geq 0\}$, say, where the system moves through some or all of K transient states, or phases, before moving to a single absorbing state $K+1$. The time of absorption, T, is then said to have a *phase-type distribution.*

Suppose now instead that there are $m > 1$ absorbing states, named $K+1, K+2, \ldots, K+m$, say. Letting T as above be the time of absorption (in any one of the absorbing states), and let C represent the state where absorption occurs, by defining $C = j$ if $X(T) = K + j$; $j = 1, 2, \ldots, m$.

The pair (T, C) thus can be viewed as an observation from a classical competing risks process, as studied for example by Lawless (2003, Ch. 9). We shall see below how certain functions of common interest for competing risks studies can be derived from the infinitesimal transition matrix $\mathbf{A}$ of the underlying Markov chain. This is a $(K+m) \times (K+m)$ matrix given on block form as

$$\mathbf{A} = \begin{bmatrix} \mathbf{Q} & \mathbf{L} \\ \mathbf{0}_1 & \mathbf{0}_2 \end{bmatrix}. \tag{1}$$

Here $\mathbf{Q}$ is the $K \times K$ matrix corresponding to the transitions between the transient states; $\mathbf{L}$ is the $K \times m$ matrix defining direct transition intensities from the transient states to the absorbing states; while $\mathbf{0}_1$ and $\mathbf{0}_2$ are, respectively, $m \times K$ and $m \times m$ matrices of zeros. Letting $\mathbf{P}(t)$ be the matrix of transition probabilities $P_{ij}(t) = P(X(t) = j | X(0) = i)$ it is well known (e.g. Ross, 1983, Ch. 5) that

$$\mathbf{P}(t) = e^{\mathbf{A}t} = \sum_{i=0}^{\infty} \mathbf{A}^i,$$

and it is then straightforward to show that (1) implies

$$\mathbf{P}(t) = \begin{bmatrix} e^{\mathbf{Q}t} & \mathbf{Q}^{-1}(e^{\mathbf{Q}t} - \mathbf{I})\mathbf{L} \\ \mathbf{0}_1 & \mathbf{I} \end{bmatrix}, \tag{2}$$

where $\mathbf{I}$ is the $K \times K$ identity matrix.

2.1. Functions for competing risks. From (2) we obtain expressions for the subdistribution functions (Lawless, 2003, Ch. 9), given by

$$F_j(t) = P(T \leq t, C = j) = P(X(t) = j) = \mathbf{p}\mathbf{Q}^{-1}(e^{\mathbf{Q}t} - \mathbf{I})\mathbf{L}\mathbf{v}_j \tag{3}$$

for $j = 1, \ldots, m$, (also called cumulative incidence functions in the medical literature). Here $\mathbf{p}$ is the K-vector with entries $p_i = P(X(0) = i)$ for $i = 1, \ldots, K$, which defines the initial distribution of the Markov chain. We will always assume that this distribution has support in the transient states $\{1, \ldots, K\}$, and it is often natural to assume $p_1 = 1$. Finally, $\mathbf{v}_j$ is the m-vector with jth element equal to 1 and the rest equal to 0.

Another function of interest in applications of competing risks is the cause-specific hazard rate, given by

$$h_j(t) = \lim_{\Delta t \to 0} \frac{P(T \le t + \Delta t, C = j | T > t)}{\Delta t} = \frac{F_j'(t)}{P(T > t)} = \frac{\mathbf{p} e^{\mathbf{Q}t} \mathbf{L} \mathbf{v}_j}{\mathbf{p} e^{\mathbf{Q}t} \mathbf{1}_K} \quad (4)$$

(e.g., Braarud, 2012). Here $\mathbf{1}_K$ is a K-vector of all 1s.

2.2. Coxian distributions for competing risks. A so-called Coxian phase-type distribution is obtained when all the transitions from the transient states are either from i to $i+1$ or to the absorbing state $K+1$. It is natural to generalize this property to the competing risks case by allowing transitions to any of the m absorbing states $K+1, \ldots, K+m$ from each of the transient states. The resulting restriction on the matrix $\mathbf{Q}$ is in fact not so strong as it may look, since any upper triangular $\mathbf{Q}$ can be brought on Coxian form by possibly permuting the transient states (see e.g. O'Cinneide, 1989).

3. Statistical inference for ordinary phase-type distributions

Suppose the lifetimes of n observation units are given by $T_1, \ldots, T_n$, where the T_i are i.i.d. from a phase-type distribution as described above, corresponding to an absorbing Markov chain with K transient states.

Asmussen et al. (1996) presented a general approach to estimation of such a phase-type distribution. Their idea was to consider the class of phase-type distributions for a fixed K as a multi-parameter exponential family. Since one then obviously is in the setting of incomplete observations, they suggested to implement the EM algorithm. The extension to right-censored data was treated by Olsson (1996), as a follow-up to the paper by Asmussen et al. (1989).

Bladt et al. (2003) considered Bayesian estimation of phase-type distributions, constructing a Gibbs sampler which draws phase-type parameters from their posterior distribution. They reported as a main advantage of their method, that the uncertainty of estimates of complex functionals of the phase-type distributions could easily be obtained. It is not so clear, on the other hand, how to do this for the Em-algorithm approach.

Faddy et al. (2009) and McGrory et al. (2009) considered, respectively, maximum likelihood estimation and Bayesian estimation for Coxian models which include covariates. The latter authors used a reversible jump MCMC in their analysis, thus including also K as a parameter in the model.

4. Statistical inference for competing risks data without covariates

4.1. Estimation by the EM algorithm. It appears to be rather straightforward to extend the approach of Asmussen et al. (1996) to the case with several absorbing states. The observed data for an observation unit is (t, c), while the full (latent) observation is of the form

$$x = (i_0, i_1, \ldots, i_{\ell-1}, c, s_0, \ldots, s_{\ell-1}),$$

where $t = \sum_{j=0}^{\ell-1} s_j$. Here $i_0, \ldots, i_{\ell-1}$ are the transient states visited before absorption in $K + c$, while $s_0, \ldots, s_{l-1}$ are the corresponding sojourn times. The difference from the approach of Asmussen et al. (1996) is hence that the terms that have to do with the absorption now also should include the identity of the absorbing state (which is actually observed).

Olsson (1996) extended the EM-algorithm of Asmussen et al. (1996) to the case where some observations are right censored or interval censored. This approach can similarly be extended to the case of censored competing risks data. In fact, Olsson's approach has similarities to the modification described above when going from a single to several absorbing states.

4.2. Bayesian estimation. The approach of McGrory et al. (2009), who considered a Coxian model with a single absorbing state, can easily be extended to the competing risks case when there are no covariates. The necessary extension will then consist in replacing their transition rates from the transient states by m sets of transition rates $\mu_{1j}, \ldots, \mu_{Kj}$ for $j = 1, \ldots, m$, one set for each absorbing state ("risk"). Again it is possible to devise a reversible jump Markov Chain Monte Carlo in order to include the number of transient states, K, as a parameter in the model.

5. Competing risks data with covariates

Suppose that for each observation is associated a covariate vector $\mathbf{x}$ to the lifetime T.

McGrory et al. (2009) considered a Coxian model with covariates. Since they modeled the expected value of T, they suggested to multiply the intensity matrix by a factor $\exp\{-\boldsymbol{\beta}'\mathbf{x}\}$. Alternatively, one may in the case of a single absorbing state define a model with covariates by multiplication of the intensity matrix $\mathbf{A}$ by $\exp\{\boldsymbol{\beta}'\mathbf{x}\}$. This approach will be a natural approach of "ordinary" Cox regression type.

In the competing risks case one is interested in investigating how the covariate vector influences either the cause-specific hazard functions or the cumulative incidence functions. As is common in such investigations, we shall consider the case $m = 2$ and assume that the cause C=1, i.e. absorption in state $K + 1$, is of main interest. In this case $\mathbf{L}$ is a $K \times 2$ matrix.

It should be noted that we are not free to model the matrices $\mathbf{Q}$ and $\mathbf{L}$ as general functions of $\mathbf{x}$. This is due to the fact that $\mathbf{A}$, being the infinitesimal transition matrix of the underlying Markov chain, has all row sums equal to 0. The implied relation between $\mathbf{Q}$ and $\mathbf{L}$ is hence seen to be

$$\mathbf{L}\mathbf{1}_m = -\mathbf{Q}\mathbf{1}_K. \tag{5}$$

In the following we suggest a way to parametrize the influence of $\mathbf{x}$ on the joint distribution of (T, C). Let

$$\mathbf{A}_0 = \begin{bmatrix} \mathbf{Q}_0 & \mathbf{L}_0 \\ \mathbf{0}_1 & \mathbf{0}_2 \end{bmatrix} \tag{6}$$

be a basic intensity matrix, similar to the baseline hazard functions of Cox-regression or of the standard modeling of cause-specific hazard rates for competing risks. We consider two different ways that the covariate vector can influence the null model:

1. The speed of the transitions through the set of transient states until absorption may depend on $\mathbf{x}$.
2. The relative tendency of absorption in states $K + 1$ and $K + 2$ may depend on $\mathbf{x}$.

Item 1 is similar to the situation without competing risks, and as already discussed can be accommodated by multiplication of the whole matrix $\mathbf{A}_0$ by a constant of the form $\exp\{\boldsymbol{\beta}'\mathbf{x}\}$. After this multiplication, we can in accordance with Item 2 modify also $\mathbf{L}_0$, but we then need to take into account the relation (5). We shall do this by modifying L_0 in a way that does not change its row sums. More precisely, we consider 2×2 matrices $\mathrm{E}(\mathbf{x})$ with row sums 1, and define

$$\mathbf{L}(\mathbf{x}) = e^{\boldsymbol{\beta}'\mathbf{x}}\mathbf{L}_0\mathrm{E}(\mathbf{x}).$$

One way is to let

$$\mathrm{E}(\mathbf{x}) = \begin{bmatrix} e^{\boldsymbol{\gamma}'\mathbf{x}} & 1 - e^{\boldsymbol{\gamma}'\mathbf{x}} \\ 0 & 1 \end{bmatrix}$$

Then by letting $\mathbf{Q}(\mathbf{x}) = e^{\boldsymbol{\beta}'\mathbf{x}}\mathbf{Q}_0$ it is easily seen that

$$\mathbf{A}(\mathbf{x}) = \begin{bmatrix} \mathbf{Q}(\mathbf{x}) & \mathbf{L}(\mathbf{x}) \\ \mathbf{0}_1 & \mathbf{0}_2 \end{bmatrix}. \qquad (7)$$

satisfies (5). We need, however, to estimate two vectors $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ of coefficients, in addition to the underlying Markov chain.

Note that if $\boldsymbol{\beta} = \mathbf{0}$, then the cause-specific hazard rate for cause 1, see (4), is of the form $\lambda_0(t)e^{\boldsymbol{\gamma}'\mathbf{x}}$, which is a commonly used model in applications of competing risks (see e.g. Lawless, 2003). This follows since $\mathbf{L}(\mathbf{x})\mathbf{v}_j = e^{\boldsymbol{\gamma}'\mathbf{x}}\mathbf{L}_0\mathbf{v}_j$ and by defining

$$\lambda_0(t) = \frac{\mathbf{p}e^{\mathbf{Q}_0 t}\mathbf{L}_0\mathbf{v}_j}{\mathbf{p}e^{\mathbf{Q}_0 t}\mathbf{1}}.$$

In this connection it is, however, a little disappointing to see from (3) that the cumulative incidence functions will depend on the covariates in the same manner as the cause-specific hazard rates. It is indeed a well known fact in competing risks theory that covariates may influence cause-specific hazards and cumulative incidence functions differently. Hence we should in some way retain both the $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ in our model. But then in general, the cause-specific hazard will not be of the simple form mentioned above.

6. Conclusions

It has been demonstrated that the concept of phase-type distributions can be generalized to the competing risks situation in a fairly straightforward manner. Statistical inference may also be performed in a much similar way. It is not quite clear, however, how to find informative and appropriate ways to model the influence of covariates.

References

Aalen, O. O. (1995). Phase Type Distributions in Survival Analysis. Scandinavian Journal of Statistics, 22(4), 447-463

Asmussen, S., Nerman, O., & Olsson, M. (1996). Fitting phase-type distributions via the EM algorithm. Scandinavian Journal of Statistics, 23(4), 419-441.

Assaf, D., Langberg, N. A., Savits, T. H., & Shaked, M. (1984). Multivariate phase-type distributions. Operations Research, 32(3), 688-702.

Bladt, M., Gonzalez, A., & Lauritzen, S. L. (2003). The estimation of Phase-type related functionals using Markov chain Monte Carlo methods. Scandinavian Actuarial Journal, 2003(4), 280-300.

Braarud, I. H. (2012). Phase type and competing risk models in survival analysis. Project work. Department of Mathematical Sciences, Norwegian University of Science and Technology.

Faddy, M.J., Graves, N. & Pettitt, A.N. (2009). Modeling length of stay in hospital and other right skewed data: Comparison of phase-type, gamma and lognormal distributions. Value in Health, 12, 309-314.

Lawless, J. F. (2003). Statistical models and methods for lifetime data. Wiley-Interscience.

McGrory, C. A., Pettitt, A. N., & Faddy, M. J. (2009). A fully Bayesian approach to inference for Coxian phase-type distributions with covariate dependent mean. Computational Statistics & Data Analysis, 53(12), 4311-4321.

Neuts, M. F. (1981). Matrix-geometric solutions in stochastic models. Johns Hopkins University Press, Baltimore, London.

O'Cinneide, C. A. (1989). On non-uniqueness of representations of phase type distributions. Commun. Statist.-Stochastic Models 5, 247-259.

Olsson, M. (1996). Estimation of phase-type distributions from censored data. Scandinavian Journal of Statistics, 23(4), 443-460.

Ross, S. M. (1983). Stochastic Processes. John Wiley and Sons, New York.