Generalising Frailty Assumptions in Survival Analysis: a Geometric Approach

Vahed Maroufy and Paul Marriott

Abstract This paper uses Information Geometry in a practical and important applied statistical context: Cox regression in survival analysis. We explore the geometry of the corresponding model space including its potentially complex boundary. The exact manner that frailty terms in Cox's hazard model are specified has important implications for modelling. For example, it is very common to assume a gamma frailty for reasons of mathematical tractability and convenience. In this paper, we examine if there is a cost to having the gamma as a default option, without further scientific justification. We take a geometric approach to understanding the effect of precise model specification. We use a new, highly flexible but statistically well-behaved, way of specifying the frailty to calibrate modelling assumptions that are very commonly used in practice. We show that the gamma frailty assumption has the effect of considerably under-estimating standard errors when compared to our more general assumptions and, potentially, introducing bias. We comment on the implications of this. The survival times of adult acute myeloid leukaemia patients in northwest England are analyzed

1 Introduction

One of the key insights of Information Geometry is the duality which links the exponential and mixture affine structures, [9]. Finite dimensional affine structures in the first are exponential families and affine structures in the second determine identification conditions in mixture models, [22]. This paper takes this fundamental

Vahed Maroufy

Department of Biostatistics, University of Texas Health, Houston, TX, USA e-mail: vahed.maroufy@uth.tmc.edu

Paul Marriott

Department of Statistics, University of Waterloo, Waterloo, ON, Canada. e-mail: pmarriot@uwaterloo.ca

insight and explores its implications in a practical and important applied statistical context.

The motivating example of this paper is a study of the factors which affect the survival times of adult acute myeloid leukaemia patients. This is an example of a problem in lifetime data analysis, [18]. One of the most popular tools here is Cox's proportional hazard model, [8]. This models the time dependent hazard function, for subject i with covariates X_i, as a product

$$h_i(t) = h_0(t)\exp\{X_i\beta\},$$

where $h_0(t)$ is the time dependent baseline hazard common to all subjects. This will be treated as a non-parametric nuisance function. The second term is a time independent function of the explanatory variables, which is modelled parametrically. Under the proportional hazard assumption the interest parameters can be estimate using a partial likelihood. Often though there are unmeasured individual level covariates which need to be taken into account in the analysis. This is commonly done by adding subject level terms θ_i in the adapted hazard given by (1). Since this will give a parameter for each subject it is convenient to treat these as unobserved random effects, or *frailties*, and simply estimate parameters associated with the distribution of these terms.

This means that we work by mixing over different hazard functions and the key model choice issue is the specification of the frailty distribution. It is here that Information Geometry plays a role through consideration of the fundamental mixture geometry and its impact on inference of the parameters of interest. In particular we use tools associated with the *local mixture model*, see [22], [3]. Of special geometric interest is the potentially complex boundary that these models can have. For example [20] shows that local mixture models can have a boundaries which are boundaries of polytopes or can be a non-smooth union of a finite number of smooth components.

Frailty models have been studied by many researchers; for example, [5], [14], [17], and [11]. Various hazard models, including Cox's regression model, have been generalised by assuming a random frailty variable. The frailty model is commonly chosen to return a tractable marginal likelihood function; hence, gamma, inverse Gaussian and positive stable distributions with closed-form Laplace transformations are regular choices ([6, p.77]). Among these the gamma is the most popular model choice, ([17]; [24]; [25]).

The goal of this paper is to investigate what might be the cost of specifying a particular parametric form, specifically the gamma, for the frailty distribution when this choice has been made purely for convenience. Model uncertainty is a critical problem in applied statistics, and – as we do here – the analysis can be treated in a geometric way. The use of geometry in the area is not new, for example the paper [7] provides an intriguing solution by proposing the 'double the variance' method for addressing the possibility of undetectably small departures from the model. Much more detail on the geometry of model specification can be found in [1].

We investigate the choice of a gamma frailty by calibrating it to a much more general space of mixtures, which can be used as a bench-marking tool. We use the excellent study of [12] to illustrate these effects in a real context. We show that making the gamma frailty assumption can have the effect of considerably underestimating standard errors and its potential misspecification gives rise to important biases.

In general with mixture models, and specifically with the frailty models considered here, the analyst has to balance flexibility with inferential and computational tractability. In this paper we use the recently defined *discrete mixture of local mixture models*, introduced in [21], to achieve the calibration. This new model generates a finite dimensional and suitably parameterized space as a high quality approximation to a general mixture model with unknown mixing mechanism. The model is always identifiable and estimable, and its geometric and inferential properties allow for fast and efficient estimation algorithms. These properties make it suitable as a calibration tool, and we are currently exploring using it as a frailty model in its own right.

Notation, motivation and the main results of the paper, including our proposed method are presented in Section 2. Section 3 looks at the calibration, with §3.1 showing a simulation study, illustrating that the local mixture method returns similar biases, but larger – often more than double – standard deviations for the estimates compared to the Expected-Maximization method of [17] where a gamma frailty is assumed. This shows the considerable impact that fixing on a particular parametric form has on inference, and clearly illustrates that the reduction in standard error may be just an artefact of modelling choice rather than real. In Section 3.2, the survival time of 1043 adults acute myeloid leukemia patients, recorded between 1982 and 1998 in northwest England, is analyzed, again with important differences in inferential conclusions being found. The paper closes with a short discussion in Section 4.

2 Methodology

Throughout this section, we follow the notation and definitions in [18] and [11]. Let (T_i^0, C_i), for $i = 1, \cdots, n$, be the failure time and censoring time of the ith individual, and also let X be the $n \times p$ design matrix of the covariate vectors. Define $T_i = \min(T_i^0, C_i)$ and $\delta_i = I(T_i^0 < C_i)$, where $I(\cdot)$ is an indicator function. In addition, associated with the ith individual, an unobservable covariate θ_i, the frailty, is assumed, where θ_i's follow some distribution, Q. Adapting the proportional hazard model of [8] for the ith individual, conditional on the frailty θ_i, the hazard function is,

$$h_i(t) = \theta_i h_0(t) \exp\{X_i \beta\}, \tag{1}$$

where $h_0(t)$ is the base hazard function, X_i is the ith row of X and $\beta = (\beta_0, \cdots, \beta_{p-1})^T$ is a p-vector regression coefficients. The cumulative hazard and survival functions are, respectively, defined as $H_i(t) = \int_0^t h_i(u)\, du$, and $S_i(t) = \exp\{-H_i(t)\}$, with the base cumulative hazard function $H_0(t)$. We also assume that the frailty θ is independent of X, and further that, given X and θ, censoring is independent and noninformative for θ and (h_0, β), see [11]. The full likelihood function for the parameter vector (β, H_0) is written as

$$L(\beta, H_0, Q) = \prod_{i=1}^{n} \int \left[\left(\theta\, h_0(T_i)\, e^{X_i \beta} \right)^{\delta_i} \exp\left\{ -\theta\, H_0(T_i)\, e^{X_i \beta} \right\} \right] dQ(\theta). \qquad (2)$$

In a similar way to a general mixture model problem, frailty survival models with an unknown frailty distribution, suffer from identification and estimability issues. Although, when all the covariates variables are continuous with a continuous distribution, [10], show that, given the distribution of the time duration variable, all the three multiplicative factors are, at least formally, identified. This theoretical result does not solve the identifiability issue in the general sense. For instance, when there is a discrete covariate then identifiability requires the corresponding regression coefficient to be limited to a known compact set [13, Ch.2].

2.1 Local and global mixture models

To allow full generality of the mixture structure it is tempting to only restrict Q to be a finite discrete distribution with an unknown number of components., i.e.

$$Q(\theta) = \sum_{i=1}^{N} \rho_i \delta_{\theta_i}(\theta),$$

where δ is the indicator function, and $\sum_{i=1}^{N} \rho_i = 1$ and $\rho_i > 0$. Indeed, as shown by [19], the non-parametric maximum likelihood estimate of Q lies in such a family. In that case the perturbation space would be 'parameterised' by N, the number of components, $(\theta_1, \ldots, \theta_N)$, the components, and $(\rho_1, \ldots, \rho_N)$, the mixing weights. However, this parameterization has many problems in implementation. Specifically, it is poorly identified and has complex boundaries. A key problem is that mixture components may be too close to one another to be resolved with a given set of data and so the order of the finite mixture is essentially not estimable, see [21].

The case where there is a single set of closely grouped components – or the much more general situation where Q is any small-variance distribution – is exactly the case which motivated the design of the local mixture model (LMM), see [22], and [2]. The key idea is, essentially, to replace the unknown number of clustered mixing components with a fixed number of low order moments parameterised in an inferentially 'nice' way. The geometric intuition is that for a local perturbation all the mixing component distributions will lie close to a low dimensional linear

(affine) space. This space is spanned by derivatives of the prior and parameterised with a small number of identified parameters.

Definition 1. For a density function, $f(y; \theta)$, belonging to the exponential family, the local mixture model of order k, centred at ϑ, is defined as

$$g_\vartheta(y; \lambda) = f(y; \vartheta) + \sum_{j=1}^{k} \lambda_j f^{(j)}(y; \vartheta), \tag{3}$$

where $\lambda := (\lambda_1, \cdots, \lambda_k) \in \Lambda_\vartheta$, $f^{(j)}(y; \vartheta) = \frac{\partial^j f(y;\theta)}{\partial \theta^j} |_{\theta=\vartheta}$. The parameter space, Λ_ϑ, is defined by

$$\Lambda_\vartheta := \{\lambda \,|\, g_\vartheta(y; \lambda) \geq 0, \forall y\}$$

is convex with a boundary determined by the non-negativity of (3). The structure of this boundary is the union of smooth manifolds and is studied in [20].

In [21], local mixtures are generalised to discrete mixtures of local mixture models. In Definition 1 the point ϑ is fixed, selected by the analyst. This has the natural generalisation to have a set of possible centres, $\vartheta_1, \cdots \vartheta_L$, selected independently of the data. This gives rise to the following.

Definition 2. The discrete mixture of local mixture models is defined by

$$g(y; \underline{\vartheta}, \underline{\lambda}) = \sum_{l=1}^{L} \rho_l \, g_{\vartheta_l}(y; \lambda^l), \tag{4}$$

where $\underline{\lambda} = (\lambda^1, \cdots, \lambda^L)$, $\underline{\vartheta} = (\vartheta_1, \dots, \vartheta_L)$, $\rho_l \geq 0$ and $\sum_{l=1}^{L} \rho_l = 1$.

The model selection task with such models is to select k, L and $\vartheta_1, \cdots \vartheta_L$ to balance estimability with the quality of approximation to a completely general mixture. We use the results of [21, Section 2.1] which show, for a given value of k, how to select the points $\vartheta_1 < \dots < \vartheta_L$ to have a uniform L^1-bound on the difference between $g(y; \underline{\vartheta}, \underline{\lambda})$ and any mixture with mixing distribution Q whose support lies in a compact region containing $\{\vartheta_1, \dots, \vartheta_L\}$.

We emphasis, before we show some of the details in an example, that in this paper, we are using the LMM structure as a calibration tool, in order to evaluate the effect of selecting a particular parametric frailty model. We are generating a rich , but well behaved, class of mixtures which subsume the gamma assumption.

Example 1. To illustrate the flexibility of this framework, Figure 1 shows some discrete mixtures of local mixture models for the exponential distribution on both the density and hazard scales. Using the methods of [21, Section 2.1] some straightforward numerical analysis shows that selecting points, in the rate parametrisation, at $\vartheta_1 = 0.5, \vartheta_2 = 1, \vartheta_3 = 2$ and $\vartheta_4 = 3.5$ and fixing $k = 4$ gives excellent L^1-approximations for all mixtures Q with support in $[0.3, 5]$. In the figure Model 1 is the unmixed exponential model, Model 2 is a single component local mixture centered at $\vartheta_2 = 1$, Models 3 and 4 are two components mixtures of local mixture centred at $(1, 3.5)$ and $(0.5, 1)$, respectively.

In [21] it was shown that an adaptation of the EM algorithm works well for fitting a finite mixture of local mixtures. We only consider here local mixture models of

order $k = 4$. Increasing this degree – while mathematically possible – only adds a small marginal improvement to the local modelling performance, ([23]), at the cost of extra parameters.

The key, novel, issue in the estimation is to deal with the fact that the parameter space, Λ_ϑ, has a complex boundary. For this paper, for the Cox model, we need to consider the case where $f(y;\theta) = \theta \exp(-\theta y)$ and then the parameter space is characterized as the set of all λ's such that, for all $y > 0$,

$$\vartheta \lambda_4 y^4 - (4\lambda_4 + \vartheta \lambda_3)y^3 + (3\lambda_3 + \vartheta \lambda_2)y^2 - (2\lambda_2 + \vartheta \lambda_1)y + \lambda_1 + \vartheta \geq 0. \quad (5)$$

The boundary is characterised as being the cases where (5) has only repeated positive roots, and it can be shown, by direct calculation, that the parameter space is the union of smooth manifolds. This characterisation allows explicit parameterisations of the component manifolds to be computed and these are exploited in the numerical studies of this paper. See Appendix for details.

3 Application to frailty modelling

In this section, we investigate the size of the effect on inference associated with fixing on a particular frailty distribution by comparing a standard gamma frailty model to our more general local mixture approach. In cases where the only reason for selecting a gamma frailty was convenience we argue differences, such as a reduction in standard error, are artefacts of model choice and not real. In order to investigate the best possible case for the gamma frailty assumption we simulate, in §3.1,
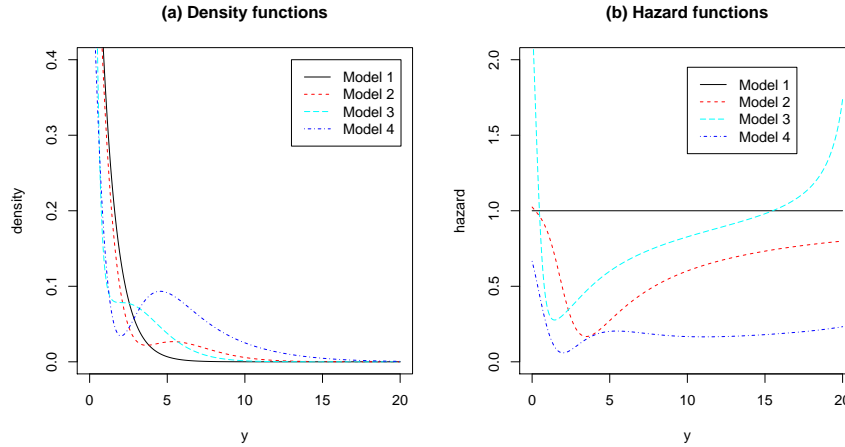

Fig. 1 The flexibility of the model class: (a) four models on a density scale, (b) the same models' hazard functions.

from that frailty distribution, and we only make the comparison to a one component local mixture with $k = 4$. Even with these restrictions we see that standard errors associated with the gamma assumption are considerably smaller than those of the LMM. Furthermore, in the real data example of §3.2, it turns out that the $k = 4$, one component local mixture is the appropriate choice for that data and we find important inferential differences which might be due to either misspecification bias, or artificially reduced standard errors, in the gamma based model.

Intuitively, for any fixed ϑ, the finite dimensional parameter vector λ represents the frailty distribution through its central moments. Local mixing also can be seen as a mechanism that extends a parametric model to a larger and more flexible space of densities which has nice geometric and inferential properties, allowing calibration of a particular parametric assumption. Substituting the local mixture expansion of Equation (3) we obtain

$$l(\beta, H_0, \lambda) = \sum_{i=1}^{n} \left(\delta_i [\log h_0(T_i) + X_i \beta] + \log f(T_i, \beta, \vartheta) \right)$$
$$+ \sum_{i=1}^{n} \log \left(1 + \sum_{j=1}^{k} \lambda_j A_j(\delta_i, y_i) \right), \qquad \lambda \in \Lambda_\vartheta \qquad (6)$$

in which $A_j(\delta_i, y_i) = \frac{f^{(j)}(T_i, \beta, \vartheta)}{f(T_i, \beta, \vartheta)}$, and $y_i = H_0(T_i)e^{X_i \beta}$. We maximize Equation (6) when estimating β, where H_0 and λ are considered as nuisance parameters. Thus, a profile likelihood optimization method is employed. That is, we first maximize for λ over Λ_ϑ to obtain $\hat{\lambda}$ and impute $\hat{H}_0$ for H_0, then maximize $l_p(\beta) = l(\beta, \hat{H}_0, \hat{\lambda})$ to estimate β.

To impute $h_0(t)$ and $H_0(t)$ we use the arguments in [11] to provide a recursive estimate of the cumulative hazard function using the fact that for two consecutive failure times, $T_{(i)}$ and $T_{(i+1)}$, we have $H_0(T_{(i+1)}) = H_0(T_{(i)}) + \Delta H_i$. Substituting this recursive equation into the log-likelihood function in (6), considering the conventions in [4] and taking partial derivative with respect to ΔH_i, we obtain

$$\frac{\partial l}{\partial \Delta H_i} = \frac{1}{\Delta H_i} - \sum_{\ell=i}^{n} e^{X_\ell \beta} + \frac{P'(e^{X_i \beta}[H_0(T_{(i)}) + \Delta H_i])}{P(e^{X_i \beta}[H_0(T_{(i)}) + \Delta H_i])}, \qquad (7)$$

which is a function of just ΔH_i when $\hat{H}_0(T_{(i)})$ is given at time $T_{(i+1)}$, $P(\cdot)$ is a polynomial of degree four with its coefficients linear functions of $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ and $P'(\cdot)$ is its derivative with respect to ΔH_i. When the denominator is not zero, Equation (7) is a polynomial of degree five which can be solved numerically for ΔH_i. Note that when there is no frailty factor – that is $\lambda = (0,0,0,0)$ – then the last term in equation (7) is zero, and the estimate of the cumulative hazard function reduces to the form in [16] which is the estimate in [17] with $\hat{\omega} = 1$.

3.1 Simulation Study

In this section, a simulation study illustrates the effect of making the gamma frailty assumption when compared to the much richer local mixture method. We adapt the method in [17], which assumes a gamma model with mean 1 and variance η for the frailty, and apply the Expectation-Maximization algorithm.

We let $C = 0.01$, $\tau = 4.6$ and we follow a similar set-up as found in [15]. For each individual the event time is $T = [-\log(1-U)\{\theta \exp\{\beta X\}\}^{-1}]^{-1/\tau}C^{-1}$, where $X \sim N(0,1)$, $U \sim \text{uniform}[0,1]$. The censoring distribution is $N(100,15)$, and frailty is assumed to follow a gamma distribution with mean 1 and variance η. Table 1 shows the bias and standard error for the estimates of the regression coefficient using both methods for three different values of η. It is clear that the local mixture method, which does not use any information about the frailty model, returns very similar biases as the Expectation-Maximization method for the gamma frailty. However, the standard deviation for the estimates in the local mixture method are almost twice as large as these for the Expectation-Maximization method.

Table 1 Bias and standard errors of coefficient estimates, when frailty is generated from $\Gamma(\frac{1}{\eta}, \eta)$. LMM is the local mixture method; Gamma the Expectation-Maximization method for the gamma frailty.

			Gamma		LMM	
n	η	β	bias	std	bias	std
200	0.5	$\log 3$	-0.057	0.18	-0.048	0.43
200	0.7	$\log 3$	-0.056	0.21	-0.038	0.40
200	1	$\log 3$	-0.117	0.22	-0.094	0.41

The results in Table 1 convey two important messages. First, without assuming any specific model form for the unobserved frailty, our methodology does equally well in terms of the bias of estimation. Second, our method automatically addresses the possibility of departures from the gamma assumption by doubling the standard errors. We argue that the apparent reduction of standard error, associated with the gamma, is an artefact of model choice which can only be justified if there were good, extra-data, reasons for trusting the gamma assumption.

3.2 Data example

We compare the two methods in the context of a study of the survival time of 1043 adult acute myeloid leukemia patients, recorded between 1982 and 1998 in north-west England ([12]). In this example 16% of the survival times are censored and complete information is available for four covariates; age, sex, white blood cell

count (WBC), and a measure of deprivation for the enumeration district of residence (Dep). [12], studied the data to investigate possible spatial variation in the survival time assuming a gamma marginal frailty with covariance structure for the frailty among the 24 districts. Although they find some indication of spatial variation between the districts, their analysis illustrates that, assuming covariance structure in the frailty variable does not affect the inference on the regression coefficients, and differences in the coefficient estimates are not significantly bigger than one standard deviation. The estimates and standard errors of the regression coefficients, $(\beta_1, \cdots, \beta_4)$, based on the independent gamma frailty are shown in the first two rows of Table 2. The frailty variance is estimated as $\hat{\eta} = 0.772$, indicating the existence of unobserved variation in the patients' survival time.

Table 2 Estimates and standard errors for covariate coefficient are presented. LMM, the local mixture method; Gamma; the estimates in *Henderson et al (2002)* for the gamma frailty. Std are the associated standard errors for the gamma model.

	age	sex	WBC	Dep
Gamma	0.0470	0.0563	0.0057	0.0547
Std	0.0045	0.0505	0.0008	0.0187
LMM	0.0422	0.0156	0.0114	0.0306

Applying the local mixture method we obtain $\hat{\lambda} = (-0.687, 0.017, 0.056, 0.023)$, also reflecting the existence of some unobserved variation. Furthermore, $\hat{\lambda}$ lie in the relative interior of Λ_ϑ, indicating that the one component local mixture is adequate.

The estimates for $(\beta_1, \cdots, \beta_4)$, using the LMM approach, are shown in the third row of Table 2. Comparing the differences between the estimates with the standard errors, we realize that all differences are inside the one standard deviation bound except for β_3, the coefficient for *WBC*, where the ratio is 7.125. This important difference in the estimate is interpreted as the result of a possible misspecification of the frailty model when a gamma distribution is imposed. It could be due to either misspecification bias or an under estimate of the standard error in the gamma frailty model.

4 Discussion

In the context of frailty survival analysis, this paper applies an approximation of a general mixture model – with a completely unspecified mixing mechanism – by a novel family of models, the discrete mixture of local mixture models. These models are built in a geometrically and inferentially convenient way and have many advantages compared to finite mixture models. We do this to generate a calibration tool with which we can evaluate the effect of specifying a particular choice of frailty

model. These geometric properties lead to inferential properties such as concavity of the log-likelihood function, identifiability and orthogonality in parametrization.

Our novel, highly flexible, but inferentially well-behaved, family gives a way of calibrating the effect of making a closed form parametric assumption on the frailty term. We see considerable differences in both the simulation and real data examples. Hence we conclude that making the gamma frailty assumption can have the effect of considerably under-estimating standard errors and its potential misspecification can gives rise to important biases.

Appendix

For the exponential distribution, the density of a single component local mixture has the form

$$\left\{\vartheta\lambda_4 y^4 - (4\lambda_4 + \vartheta\lambda_3)y^3 + (3\lambda_3 + \vartheta\lambda_2)y^2 - (2\lambda_2 + \vartheta\lambda_1)y + \lambda_1 + \vartheta\right\}\exp(-\vartheta y), \tag{8}$$

which is a density when the polynomial is non-negative for all $y > 0$. It will lie on the boundary of the parameter space when all the positive roots of the polynomial are of order 2 or 4 and $\lambda_4 > 0$. Polynomials of this form can be written as

$$\left\{a(y-r)^2(y^2 + 2by + c)\right\}, \tag{9}$$

then the roots of $(y^2 + 2by + c)$ must be repeated when they are positive, and a, b, c and r satisfy

$$\vartheta^5 - \vartheta^4 ar^2 c - (2ar^2 b - 2arc)\vartheta^3 - (2ac + 2ar^2 - 8arb)\vartheta^2 - (12ab - 12ar)\vartheta - 24a = 0. \tag{10}$$

This follows since a density integrates to one. Comparing (8) with (9) and imposing (10) results in an explicit parameterisation of the boundary in terms of b, c and r. The Jacobean of this has full rank as long as $r \neq 0, -b \pm \sqrt{b^2 - c}$. So, under these conditions the boundary is a smooth manifold. In the singular case where $r = -b \pm \sqrt{b^2 - c}$, since roots must have even order, we have $r = -b$ and $b < 0$. This singular set is given by a one dimensional manifold, again with an explicit parameterisation.

References

1. Anaya-Izquierdo, K., Critchley, F., Marriott, P., Vos, P.: The geometry of model sensitivity: an illustration. In: Computational Information Geometry: For Image and Signal Processing. Springer (2016)
2. Anaya-Izquierdo, K., Marriott, P.: Local mixture models of exponential families. Bernoulli **13**, 623–640 (2007)

3. Anaya-Izquierdo, K., Marriott, P.: local mixture of the exponential distribution. Annals of the Insti. Math. Statist. **59**(1), 111–134 (2007)
4. Breslow, N.: Contribution to the discussion on the paper of Cox (1972). Biom **34**, 216–217 (1972)
5. Clayton, D.G.: A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika **65**(1), 141–151 (1978)
6. Cook, R.J., Lawless, J.F.: The statistical analysis of recurrent events (2007)
7. Copas, J., Eguchi, S.: Local model uncertainty and incomplete-data bias (with discussion). Journal of the Royal Statistical Society: Series B (Statistical Methodology) **67**(4), 459–513 (2005)
8. Cox, D.R.: Regression models and life-tables. Journal of the Royal Statistical Society B **34**, 187–200 (1972)
9. Critchley, F., Marriott, P.: Information geometry and its applications: An overview. In: Computational Information Geometry: For Image and Signal Processing, pp. 1–31. Springer (2017)
10. Eleber, C., Ridder, G.: True and spurious duration dependence: The identifiability of the proportional hazard model. Review of Economic Studies **49**, 403–409 (1982)
11. Gorfine, M., Zucher, D.M., Hsu, L.: Prospective survival analysis with general semiparametric shared frialty model: A pseudo full likelihood approach. Biometrika **93**(3), 735–741 (2006)
12. Henderson, R., Shimakura, S., Gorst, D.: Modeling patial variation in leukemia urvival data. Journal of the American Statistical Assiciation **97**(460), 965–972 (2002)
13. Horowitz, J.L.: Semiparametric and Nonparametric Methods in Econometrics. Springer (2010)
14. Hougaard, P.: Survival models for heterogeneus population derived from stable distributions. Biometrika **73**(2), 387–396 (1986)
15. Hsu, L., Chen, L., Gorfine, M., Malone, K.: Semiparametric estimation of marginal hazard function from case-control family studies. Biometrics **60**, 936–944 (2004)
16. Johansen, S.: An extenssion of Cox's regression model. International Statistical Review **51**, 258–262 (1983)
17. Klein, J.P.: Semiparametric estimation of random effects using cox model based on the em algorithm. Biometrics **48**, 795–806 (1992)
18. Lawless, J.F.: Statistical Models and Methods for Lifetime Data. Wiley (1981)
19. Lindsay, B.G.: Mixture Models: Theory, Geometry and Applications. Inst of Mathematical Statistics (1995)
20. Maroufy, V., Marriott, P.: Computing boundaries in local mixture models. Geometric Science of Information, Lecture Notes in Computer Science (9389), 577–585 (2015)
21. Maroufy, V., Marriott, P.: Mixture models: Building a parameter space. Statistics and Computing pp. 1–7 (2016)
22. Marriott, P.: On the local geometry of mixture models. Biometrika **89**, 77–93 (2002)
23. Marriott, P.: Extending local mixture models. AISM **59**, 95–110 (2006)
24. Nielsen, G.G., Gill, R.D., Anderson, P.K., Sorensen, T.I.A.: A counting process approach to maximum likelihood estimation in frailty models. Scandinavian Journal of Statistics **19**(1), 25–43 (1992)
25. Vaupel, J.W., Manton, K.G., Stallard, E.: The impact of heterogeneity in individual frailty on the dynamics of mortality. Demography **16**, 539–454 (1979)