Lifetime Data Anal (2009) 15:147–176
DOI 10.1007/s10985-008-9108-y

Review and implementation of cure models based on first hitting times for Wiener processes

Jeremy Balka · Anthony F. Desmond · Paul D. McNicholas

Received: 2 March 2008 / Accepted: 19 November 2008 / Published online: 4 January 2009

Abstract The development of models and methods for cure rate estimation has recently burgeoned into an important subfield of survival analysis. Much of the literature focuses on the standard mixture model. Recently, process-based models have been suggested. We focus on several models based on first passage times for Wiener processes. Whitmore and others have studied these models in a variety of contexts. Lee and Whitmore (Stat Sci 21(4):501–513, 2006) give a comprehensive review of a variety of first hitting time models and briefly discuss their potential as cure rate models. In this paper, we study the Wiener process with negative drift as a possible cure rate model but the resulting defective inverse Gaussian model is found to provide a poor fit in some cases. Several possible modifications are then suggested, which improve the defective inverse Gaussian. These modifications include: the inverse Gaussian cure rate mixture model; a mixture of two inverse Gaussian models; incorporation of heterogeneity in the drift parameter; and the addition of a second absorbing barrier to the Wiener process, representing an immunity threshold. This class of process-based models is a useful alternative to the standard model and provides an improved fit compared to the standard model when applied to many of the datasets that we have studied. Implementation of this class of models is facilitated using expectation-maximization (EM) algorithms and variants thereof, including the gradient EM algorithm. Parameter estimates for each of these EM algorithms are given and the proposed models are applied to both real and simulated data, where they perform well.

J. Balka · A. F. Desmond (✉) · P. D. McNicholas
Department of Mathematics & Statistics, University of Guelph, Guelph, ON, Canada N1G2W1
e-mail: tdesmond@uoguelph.ca

J. Balka
e-mail: jbalka@uoguelph.ca

P. D. McNicholas
e-mail: pmcnicho@uoguelph.ca

Keywords Cure rate · EM algorithm · First hitting time · Gradient EM algorithm · Inverse Gaussian · Mixture models · Survival analysis · Wiener process

1 Background

1.1 The cure rate

This section is an overview of the commonly used methods in modeling populations that contain immune individuals. Let T be a random variable representing survival time to some event of interest. Let $S(t)$ represent the survivor function for the population, where $S(t) = P(T > t)$. The cure rate p is then defined as $p = \lim_{t\to\infty} S(t)$.

Note that, throughout this paper, we assume independent random censoring as described by Lawless (2003, p. 54).

1.2 Common cure rate models

1.2.1 The Kaplan-Meier estimator

A nonparametric estimator of the cure rate is the minimum value of the Kaplan-Meier curve; $\hat{p} = \min\{\hat{S}(t)\}$. This estimator of p is sometimes referred to as the Kaplan-Meier cure rate estimator (KMCRE). The KMCRE is well known to give misleading estimates of the cure rate (details in Maller and Zhou 1996).

1.2.2 Parametric cure rate mixture models

In the analysis of survival data with a possible immune proportion, parametric cure rate mixture models are a natural method of analysis. Boag (1949), Berkson and Gage (1952), Farewell (1982) and many others have extensively modeled populations with a suspected immune proportion. The standard parametric cure rate mixture model takes the form

$$S_{\text{pop}}(t) = p + (1 - p)S_0(t),$$

where $S_0(t)$ is a proper parametric survival distribution and $S_{\text{pop}}(t)$ is the population survival function. Many choices for $S_0(t)$ are possible; common choices that have been used in the literature are the Weibull distribution, the log-logistic distribution, and the lognormal distribution. Peng et al. (1998) use a generalized F-distribution in a cure rate mixture model. The generalized F-distribution is very flexible and includes many of the commonly used survival time distributions as special cases. However, it is quite complicated and its application is computationally intensive.

Parametric cure rate mixture models are not without their faults however. Farewell (1982) discussed some of the problems involved and suggested that mixture models not be used, unless there is strong scientific evidence of an immune proportion. When the two populations (immune and susceptible individuals) are assumed to exist, the inferences will be based on this assumption whether it is valid or not. Farewell (1982)

also advocates careful inspection of the likelihood function, since in many applications it can be quite flat as the cure rate p changes. It can be difficult to distinguish between a proportion of cured individuals and a long right tail of the distribution of failure times for the susceptible proportion.

1.3 Alternatives to the standard cure rate mixture model

1.3.1 Yakovlev's model

Yakovlev et al. (1993) discuss an alternative to the standard cure rate model that has a strong appeal for its natural biological interpretation. Chen et al. (1999) discuss a formulation of this model in a Bayesian framework. The model is often phrased in terms of times to relapse of cancer, since it has a clear biological interpretation in that context. In the formulation of Chen et al. (1999), an individual is assumed to have N carcinogenic cells, where N has a Poisson distribution with mean θ.

Let Z_i be the time that the ith carcinogenic cell produces an observable cancer mass. The Z_i are assumed to be independent and identically distributed with some distribution function $F(t)$. The time to relapse of the cancer is given by $T = \min\{Z_i : 1 \leq i \leq N\}$. The survival function for the population is

$$S_{\text{pop}}(t) = P(\text{no detectable tumour by time } t) = e^{-\theta} + \sum_{k=1}^{\infty} S(t)^k \frac{\theta^k}{k!} e^{-\theta} = e^{-\theta F(t)}.$$

Note that $S_{\text{pop}}(\infty) = e^{-\theta} > 0$, which implies there are cured individuals present in the population.

1.3.2 Proportional hazards cure models

The standard proportional hazards model (Cox 1972) as commonly used, does not usually allow for a cured proportion. However, as a referee has pointed out, it is possible to have $S_0(\infty) > 0$ in the baseline survival function. More commonly, in the standard cure rate mixture model, a cured proportion is incorporated into the survival function as follows:

$$S(t \mid \mathbf{x}) = p(\mathbf{x}) + (1 - p(\mathbf{x}))S_0(t)^{\exp(\boldsymbol{\beta}'\mathbf{x})},$$

where $p(\mathbf{x})$ typically has a logistic-linear link with the covariates and $S_0(\infty) = 0$. Estimation of the model parameters can be more computationally difficult than in the standard cure rate mixture model. Kuk and Chen (1992) used a Monte Carlo simulation method to estimate the model parameters. Taylor (1995) implemented the expectation-maximization algorithm (described in Sect. 3.1) on a version of the model with no covariates. Sy and Taylor (2000) and Peng and Dear (2000) extended the method to allow for inclusion of covariate information.

1.3.3 Frailty models

In standard frailty models, a non-negative frailty random variable Y is introduced, acting multiplicatively on the hazard rate (cf. Duchateau and Janssen 2008). The frailty models can be adapted to allow for a cured proportion by choosing a distribution for Y that has a point mass at 0.

Aalen (1992) considered a family of compound Poisson distributions. In this model, the form of the frailty random variable Y is

$$Y = \begin{cases} X_1 + X_2 + \ldots + X_N & \text{if } N > 0 \\ 0 & \text{if } N = 0, \end{cases}$$

where N has a Poisson distribution and $X_1, X_2, \ldots, X_N$ are independent, each having a gamma distribution. Longini and Halloran (1996) also extend the standard frailty model to allow for a cured proportion by using a frailty mixture model in which the frailty is 0 with probability p, and has a continuous distribution with probability $1 - p$.

1.3.4 Defective distributions

Defective distributions integrate to less than one for certain values of the parameters, without explicitly including the parameter p. There are several choices of such distributions that lead to interesting possibilities for cure rate models. Haybittle (1959) fits a Gompertz model to survival times for breast cancer. Cantor and Shuster (1992) also used the Gompertz distribution to model times to death for patients with pediatric cancer and in this model, the modified Gompertz hazard function is given by $\lambda(t) = \alpha e^{\beta t}, \alpha > 0, \beta < 0$. This yields a survival function for the population of $S_{pop}(t) = \exp\left\{\beta^{-1}\alpha\left(1 - e^{\beta t}\right)\right\}$, and a cure rate of $p = e^{\beta^{-1}\alpha}$. Gieser et al. (1998) extended this model to include covariate information.

Note that the cure rate p is a function of the parameters of the failure time distribution of the susceptibles. This could be a drawback to the use of defective distributions as cure rate models. Of course, if the assumed relationship between the cure rate and failure time distribution is correct, the assumption will aid in accurate estimation of the parameters. However, if this assumption is not correct, then the estimates of the cure rate and failure time distribution for the susceptibles will be biased.

Another choice of defective distribution that has been used in the literature is the inverse Gaussian distribution; see, for example, Whitmore (1979). The inverse Gaussian cure rate model and several related models will be discussed in Sect. 2.

2 Cure rate models based on the first passage time of a Wiener process

2.1 Introduction

In many situations, it is natural to view the survival event of interest as the end result of an underlying process. For example, in cancer studies, death is the end result of the progression of cancer in the body. Attempting to model the underlying process may

aid in a more complete understanding of the nature of the event. Cox (1999) outlines four different types of relations between such a process and failure; our models fall under the first of those types. Aalen and Gjessing (2001) suggested that more attention should be paid to survival models viewed from a process point of view. This section views survival times as the first passage time of an underlying Wiener process. For applications of first hitting times to survival analysis, see Lee and Whitmore (2004), Whitmore et al. (1998) and Lee and Whitmore (2006). The inverse Gaussian distribution arises as the first passage time to an absorbing barrier in a Wiener process. Cure rate models based on the inverse Gaussian distribution have seen limited use in the literature; see, however, Whitmore (1979) and Lee and Whitmore (2006).

2.2 The inverse Gaussian distribution

The inverse Gaussian distribution was first derived by Schrödinger (1915), who discovered it as the probability distribution of the first passage time in one dimensional Brownian motion. Tweedie (1945) studied this distribution further and first proposed the name inverse Gaussian distribution. Let $Y(t)$ be a Wiener process with drift δ and volatility ν starting at the origin. That is, let $Y(t)$ have the following properties:

1. For any $t_1 < t_2 < t_3 < t_4$, $Y(t_2) - Y(t_1)$ and $Y(t_4) - Y(t_3)$ are independent.
2. $Y(t_2) - Y(t_1)$ is normally distributed with mean $\delta(t_2 - t_1)$ and variance $\nu(t_2 - t_1)$.

The distribution of the first passage time to an absorbing barrier $u > 0$ units from the origin is given by

$$f(t) = \frac{u}{\sqrt{2\nu\pi t^3}} e^{-(u-\delta t)^2/2\nu t}, \tag{1}$$

where $\delta > 0$ and $\nu > 0$. There are three parameters, but the density depends statistically on only two.

The popularity of the inverse Gaussian distribution as a lifetime model has also increased since the review paper of Chhikara and Folks (1978). In more recent years, it has been used for extensive application; examples include modeling labour turnover (Whitmore 1979), strike duration (Lancaster 1972) and times to task completion (Desmond and Chapman 1993). Its origin as the first passage time in Brownian motion is very appealing for a lifetime model. Lee and Whitmore (2006) give a review of general first hitting time models and suggest their potential for use as cure rate models. The inverse Gaussian distribution is also flexible, with shapes ranging from highly skewed to symmetrical, depending on the value of the shape parameter $\phi = \delta/\nu$.

2.3 The defective inverse Gaussian distribution

If $\delta > 0$, then the underlying Wiener process will reach the barrier at $Y = u$ with certainty. The inverse Gaussian distribution has a property that is useful when modeling situations where immune individuals may be present; for values of $\delta < 0$, the distribution is defective, that is $S(\infty) > 0$. This is easy to visualize when viewed

through the underlying process. If the mean drift is away from the absorbing barrier, not all individuals will reach the barrier. Whitmore (1979) called the inverse Gaussian distribution allowing for negative values of δ the 'defective' inverse Gaussian (DIG) distribution.

For $\delta < 0$, there is a point mass at infinity of $1 - e^{2\delta/\nu}$. When used as a cure rate model, the proportion of immunes in the population is

$$p = S(\infty) = \begin{cases} 1 - e^{2\delta/\nu} & \text{for } \delta < 0 \\ 0 & \text{for } \delta \geq 0. \end{cases}$$

Another property of the DIG distribution easily yields the failure time distribution of the susceptibles in the population. The distribution of first passage times, conditional on reaching the barrier, is a proper inverse Gaussian distribution with drift parameter $|\delta|$.

The DIG model has the appealing property of a simple interpretation of the underlying process, but there are some disadvantages that may restrict its use as an effective cure rate model. Both the cure rate, and the shape of the failure time distribution of the susceptible proportion depend only on $\phi = \delta/\nu$. If the actual underlying process is a Wiener process with negative drift, this relationship will aid in the estimation of the cure rate. The DIG model will be the best model in this situation, and more information can be gained about p from the observed failure times. If the underlying process is a Wiener process with positive drift ($\delta > 0$) and the true value of $p = 1 - e^{-2\delta/\nu}$, then it is also the same situation. Since it will generally be impossible to know whether the true process has negative drift, or has positive drift with $p = 1 - e^{-2\delta/\nu}$, this may be too restrictive an assumption in many practical cases.

Farewell (1986) suggests that one should be quite sure of the presence of a cured proportion in the population before they are used, since these mixture models will attempt to fit this proportion, whether or not it is truly different from zero. This may not be as much of a problem with the DIG model, since the estimation procedure will not necessarily seek to fit the model with $\delta < 0$.

2.4 The inverse Gaussian cure rate mixture model

For values of $\delta > 0$, the inverse Gaussian is a proper distribution, and an ordinary cure rate mixture model can be fitted using standard techniques. The likelihood function is

$$\mathcal{L}(p^*, \delta, \nu \mid \mathbf{t}) = \prod_{i \in D} (1 - p^*) f(t_i) \prod_{i \in C} [p^* + (1 - p^*) S(t_i)],$$

where $f(t)$ is defined as in (1) and

$$S(t) = 1 - \left[\Phi\left(\frac{-1 + \delta t}{\sqrt{\nu t}} \right) + e^{2\delta/\nu} \Phi\left(-\frac{1 + \delta t}{\sqrt{\nu t}} \right) \right],$$

for $\delta > 0$, $\nu > 0$. The parameter p^* represents the proportion of individuals not subject to the Wiener process, or subject to a different process that will never reach the

absorbing barrier. If δ is restricted to positive values, $p^* = p$, which is the cure rate. A modification to this standard cure rate model involves allowing values of $\delta < 0$, as in the DIG, but still including a parameter p^*. In this case, the cure rate is

$$p = \begin{cases} p^* + (1 - p^*)(1 - e^{2\delta/\nu}) & \text{for } \delta < 0 \\ p^* & \text{for } \delta \geq 0. \end{cases}$$

In the absence of covariate information, allowing $\delta < 0$ in the inverse Gaussian cure rate mixture is redundant, as the proportion of immune individuals accounted for by negative drift can simply be absorbed into the parameter p^*. For models involving covariates, there may be differences between the two choices of mixture model. Allowing $\delta < 0$ values in the inverse Gaussian cure rate mixture is somewhat of a hybrid between the DIG and the standard inverse Gaussian cure rate mixture. It results in a more difficult interpretation of the parameter estimates, but a more flexible model. The inverse Gaussian cure rate mixture model with unrestricted δ, will be referred to (somewhat loosely) as the inverse Gaussian cure rate mixture (IGCRM) herein and, in the absence of covariate information, δ will be restricted to positive values. The added parameter makes the IGCRM model a more flexible model than the DIG model. Maximum likelihood estimates of the parameters in the IGCRM can be found through maximization techniques applied directly to the likelihood function, or *via* the expectation-maximization (EM) algorithm and its variants, discussed in Sect. 3.

2.5 A Wiener process with two absorbing barriers

The addition of a second absorbing barrier to the underlying Wiener process results in another cure rate model. Consider a Wiener process with drift δ and variance ν, with absorbing barriers at u, and $-l$, where $u, l > 0$. The process terminates when it reaches either barrier. The distribution of the first passage time to one of the barriers, conditional on not first reaching the other barrier, can be found in many texts on stochastic processes; for example, see Feller (1986).

This distribution has strong potential for use as a cure rate model, with a simple, natural interpretation. Let the upper barrier represent the survival event of interest, and the lower barrier $-l$ represent an immunity threshold. Figure 1 illustrates two sample paths of a Wiener process, with absorbing barriers at u and $-l$.

In Fig. 1, Sample 2 is absorbed into the upper barrier at time t_2 and since the upper barrier represents the event of interest, t_2 will be an observed failure time. Sample 1 is absorbed into the lower barrier at time t_1. The individual corresponding to Sample 1 achieves immunity at time t_1, and will therefore never reach the upper barrier. This individual will never experience the event of interest and will eventually result in a censored observation.

An uncensored event time corresponds to the first passage time to the upper barrier, conditional on having never reached the lower barrier. The density function of the first passage time to the upper barrier depends on four parameters: δ, ν, u, and l, only three of which are free, in a statistical sense. For the sake of consistency with the other models discussed herein, the upper barrier u is set at 1. This is equivalent to

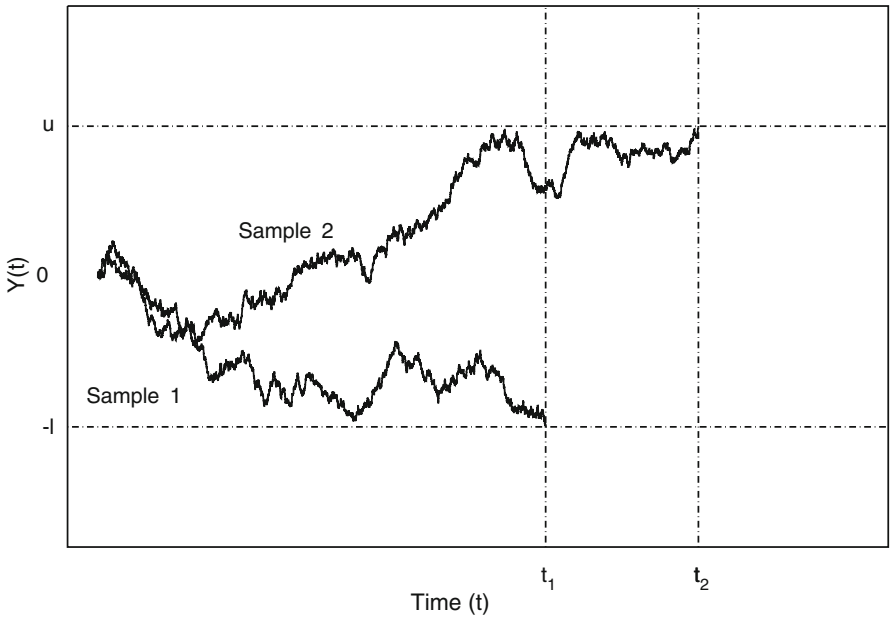

Fig. 1 Two samples of a Wiener process with two absorbing barriers

starting with a two barrier Wiener process with parameters δ', ν', u', and l', and letting $\delta = \delta'/u'$, $\nu = \nu'/(u')^2$, $u = u'/u' = 1$, $l = l'/u'$. Under this parameterization, the density function of the time to reach the upper barrier before reaching the lower barrier is given by

$$f(t \mid \delta, \nu, l) = e^{\delta/\nu} \frac{\nu}{(1+l)^2} \sum_{k=1}^{\infty} e^{z_k t} k\pi \sin\left(\frac{k\pi}{(1+l)}\right),$$

where

$$z_k = -\frac{1}{2}\left(\frac{\delta^2}{\nu} + \frac{k^2\pi^2\nu}{(1+l)^2}\right).$$

This expression is well known in the theory of Brownian motion; for example, see Knight (1981). Note that this distribution is defective, since not all individuals will reach the upper barrier. The probability of not being absorbed in the upper barrier by time t is given by

$$S(t \mid \delta, \nu, l) = 1 - e^{\delta/\nu}\left[\frac{\sinh(\frac{\delta l}{\nu})}{\sinh\left(\frac{\delta(1+l)}{\nu}\right)} - \frac{\nu}{(1+l)} \sum_{k=1}^{\infty} \frac{e^{-\lambda_k t}}{\lambda_k} k\pi \sin\left(\frac{k\pi}{1+l}\right)\right],$$

where

$$\lambda_k = \frac{1}{2}\left[\frac{\delta^2}{\nu} + \frac{k^2\pi^2\nu}{(1+l)^2}\right]$$

(Pelsser 2000). The proportion of individuals that will never be absorbed by the upper barrier (the cure rate) is given by

$$p = S(\infty \mid \delta, \nu, l) = 1 - \frac{e^{\delta/\nu}\sinh\left(\frac{\delta l}{\nu}\right)}{\sinh\left(\frac{\delta(1+l)}{\nu}\right)}.$$

Right censored times imply that the process has not been absorbed in the upper barrier by time t. Thus right censored times contribute a likelihood factor of $S(t)$. The likelihood function is the standard likelihood for right censored survival times; $\mathcal{L}(\delta, \nu, l \mid \mathbf{t}) = \prod_{i\in D} f(t_i)\prod_{i\in C} S(t_i)$. Both the density function and survivor function involve an infinite sum. Maximum likelihood procedures involve truncating the sum when the contribution of the terms becomes negligible.

Covariates can be included in a similar fashion as the previously discussed models, letting the drift parameter have a linear relationship with the covariates. The value of the lower barrier may also be allowed to depend on the covariates. Note that the two barrier model reduces to the DIG model for $l = \infty$ and so the two barrier model is a generalization of the DIG model, yet still preserves the interpretation of the underlying process. The two barrier model has an advantage over the inverse Gaussian cure rate mixture model in that the cure rate is determined completely by the process and is not modelled as a separate quantity. The two barrier Wiener process cure rate model is a natural extension of the inverse Gaussian distribution to incorporate a cure rate. The major drawback of the two barrier model is that the density and survival functions are much more complicated, resulting in the analysis being significantly more computationally intensive.

2.6 Mixture of two inverse Gaussian distributions

An alternative Wiener process model arises if the population is a mixture of individuals subject to one of two different Wiener processes. Whitmore and Su (2007) considered such models in a study of fetal birth weight. Let π represent the proportion of individuals that are subject to a process with positive drift, while the remaining proportion $1-\pi$ are subject to a Wiener process with possibly negative drift. Individuals subject to positive drift will experience the event of interest with certainty and only those individuals that experience negative drift are potentially immune. For this model, the density function of time to failure for the population is $f_{\text{pop}}(t) = \pi f_1(t) + (1-\pi) f_2(t)$, where $f_1(t)$ is the inverse Gaussian density with parameters δ_1, $\nu_1 > 0$, and $f_2(t)$ is the DIG density with parameters $-\infty < \delta_2 < \infty$, and $\nu_2 > 0$. This model reduces to the DIG model for $\pi = 0$. The cure rate is given by

$$p = \begin{cases} (1-\pi)(1-e^{2\delta_2/\nu_2}) & \text{for } \delta_2 < 0 \\ 0 & \text{for } \delta_2 \geq 0. \end{cases}$$

This five parameter model has much more flexibility than the two parameter DIG or three parameter IGCRM and can provide a substantially better fit to some data sets. The added parameters can result in difficulty fitting the model, as widely varying parameter values can result in similar values of the likelihood function. The likelihood for this model is given by

$$\mathcal{L}(\delta_1, \delta_2, \nu_1, \nu_2, \pi \mid \mathbf{t}) = \prod_{i \in D} [\pi f_1(t) + (1-\pi) f_2(t)] \prod_{i \in C} [\pi(1 - F_1(t)) + (1-\pi)(1 - F_2(t))],$$

where $F_j(t)$ is the cumulative distribution function of the inverse Gaussian distribution with parameters δ_j and ν_j, for $j = 1, 2$. Parameter estimates can be found by maximizing the likelihood function. A method based on the EM algorithm is discussed in Sect. 3. This defective inverse Gaussian-inverse Gaussian mixture model will be called the DIGIGMIX model hereafter.

2.7 Heterogeneity in the drift parameter

In many practical cases, it may be reasonable to think that some individuals have higher drift than others. This may be accounted for by measured covariates, but there may also be unexplained individual heterogeneity in the population. Whitmore (1986) extended the defective inverse Gaussian model by deriving a mixture model allowing for heterogeneity in both δ and ν. Whitmore also derived a restricted version of this model, allowing only for heterogeneity in δ. In the restricted model, the drift parameter was assumed to have a normal distribution. This section will discuss this latter model, allowing for heterogeneity in δ only. Desmond and Chapman (1993) modelled times to task completion at a large automobile plant using this mixture distribution. However, neither Whitmore (1986) nor Desmond and Chapman (1993) acknowledge explicitly the defective nature of this model. Aalen and Gjessing (2001) point out the defective nature of this model and hence its potential as a cure rate model; they also point out the connection with frailty.

Consider (1) and assume that the drift parameter is normally distributed with mean d and variance $\nu\tau$. Then the marginal distribution of the time to failure is,

$$f(t \mid d, \nu, \tau) = \frac{1}{\sqrt{2\pi\nu t^3(\tau t + 1)}} \exp\left\{-\frac{(td-1)^2}{2\nu t(\tau t + 1)}\right\}.$$

As δ can take on negative values in the conditional normal distribution, the marginal distribution of T is defective. Note that for $\tau = 0$, there is no heterogeneity in δ and this distribution reduces to the DIG distribution. Score tests for $H_0 : \tau = 0$ were studied by Desmond and Yang (2006). Likelihood ratio tests and Wald tests were studied by Desmond and Chapman (1992).

To construct the likelihood function and estimate the model parameters, the survivor function is needed. The survivor function is given by

$$S(t \mid d, \nu, \tau) = \Phi\left(\frac{1 - dt}{\sqrt{\nu(t^2\tau + t)}}\right) - e^{2(d+\tau)/\nu}\Phi\left(-\frac{1 + t(2\tau + d)}{\sqrt{\nu(t^2\tau + t)}}\right).$$

The cure rate p is given by the limiting value of the survivor function:

$$p = \Phi\left(-\frac{d}{\sqrt{\nu\tau}}\right) - e^{2(d+\tau)/\nu}\Phi\left(-\frac{2\tau + d}{\sqrt{\nu\tau}}\right).$$

This inverse Gaussian-normal mixture model will be called the IGNMIX for the remainder of this work.

3 Parameter estimation in cure rate mixture models

3.1 The expectation-maximization algorithm

The analysis of cure rate data often involves finding maximum likelihood estimates of the model parameters. The likelihood function for the various models rarely has closed form solutions for the maximum likelihood estimators, necessitating the use of numerical techniques. The expectation-maximization (EM) algorithm (Dempster et al. 1977) is an iterative technique used to obtain maximum likelihood estimates for a likelihood function $\mathcal{L}(\boldsymbol{\theta} \mid \mathbf{t})$ where data is either incomplete or is treated as incomplete. This latent data, denoted $\mathbf{z}$, together with the observed data $\mathbf{t}$, is called the complete-data $(\mathbf{t}, \mathbf{z})$ and it is the complete-data likelihood function $\mathcal{L}_c(\boldsymbol{\theta} \mid \mathbf{t}, z)$ that is maximized. For simplicity of notation, these likelihoods will be written as $\mathcal{L}(\boldsymbol{\theta})$ and $\mathcal{L}_c(\boldsymbol{\theta})$ herein.

The algorithm alternates between an expectation (E) step and a maximization (M) step. In its most general setting, the E-step involves finding the expected value of the complete-data log-likelihood, conditional on the observed data, denoted $Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(k)})$, where $\boldsymbol{\theta}^{(k)}$ is the estimated value of $\boldsymbol{\theta}$ at iteration k. In the M-step, the value of $\boldsymbol{\theta}$ that maximizes this expected complete-data log-likelihood is computed. Applications of the EM algorithm to cure rate models can be found in Kuo and Peng (1995) and Peng and Dear (2000). Herein, EM algorithms are used to estimate parameters of the DIG, IGCRM and DIGIGMIX cure rate models.

3.2 Generalized EM algorithms

In some situations, even with the addition of the latent data, the solution to the M-step is not available in closed form. In these situations, a generalized EM (GEM) algorithm is often used. A GEM algorithm increases the expected complete-data log-likelihood at each iteration. Dempster et al. (1977) show that this condition is sufficient to ensure

that $\mathcal{L}(\boldsymbol{\theta}^{(k+1)}) \geq \mathcal{L}(\boldsymbol{\theta}^{(k)})$. Further properties of GEM algorithms and an excellent overview of EM algorithms and their variants is given by McLachlan and Krishnan (2008).

In order to implement a GEM algorithm, a method of updating parameter estimates that results in an increase in the incomplete-data likelihood is required. One simple method is to perform one step of a Newton-Raphson iteration scheme at each maximization step. Lange (1995a) considers this method and calls it the gradient EM algorithm. The use of a full Newton step does not guarantee an increase in the incomplete-data likelihood, and there may be boundary crossing problems for the parameter estimates.

More generally, Lange (1995a) and Rai and Matthews (1995) consider estimators of the form

$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} + s^{(k)} S_0(\boldsymbol{\theta}^{(k)}) I_0^{-1}(\boldsymbol{\theta}^{(k)}), \tag{2}$$

where

$$S_0(\boldsymbol{\theta}^{(k)}) = \left[\frac{\partial \log \mathcal{L}_c(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(k)}}, \quad I_0(\boldsymbol{\theta}^{(k)}) = \left[-\frac{\partial^2 \log \mathcal{L}_c(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2}\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(k)}},$$

and $0 < s^{(k)} \leq 1$. Since the maximization step is replaced by one step of a maximization procedure, Rai and Matthews (1995) call this the EM1 algorithm. When $s^{(k)} = 1$ the EM1 algorithm reduces to the gradient EM algorithm. The step fraction $s^{(k)}$ is chosen such that the incomplete-data likelihood function increases at each iteration.

One drawback of the EM algorithm is that the rate of convergence can be quite slow, especially when the fraction of missing information is large. Lange (1995b) suggests a quasi-Newton alternative to speed the rate of convergence of the algorithm; replacing (2) with

$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} + s^{(k)} S_0(\boldsymbol{\theta}^{(k)}) \{I_0(\boldsymbol{\theta}^{(k)}) + B^{(k)}\}^{-1}, \tag{3}$$

where $B^{(k)}$ is chosen so that $\{I_0(\boldsymbol{\theta}^{(k)}) + B^{(k)}\}$ is closer to the Hessian of the incomplete-data likelihood than $I_0(\boldsymbol{\theta}^{(k)})$. Lange (1995b) suggests a choice of $B^{(k)}$ that can provide a significant improvement in the rate of convergence.

The EM and EM1 algorithms will be applied to the IGCRM model in Sect. 3.4, the EM1 algorithm and the quasi-Newton algorithm (Lange 1995b) will be applied to the DIG model in Sect. 3.5 and the EM1 algorithm will be applied to the DIGIGMIX model in Sect. 3.6.

3.3 The EM algorithm applied to cure rate mixture models

The likelihood function of the standard cure rate mixture model can be written as

$$\mathcal{L}(\boldsymbol{\theta} \mid \mathbf{t}) = \prod_{i \in D} (1-p) f(t_i \mid \boldsymbol{\theta}) \prod_{i \in C} \left[p + (1-p) S(t_i^* \mid \boldsymbol{\theta})\right],$$

where t_i represents the uncensored failure time for the ith individual, and t_i^* represents a right censoring time for the ith individual. The vector $\mathbf{t}$ contains the failure and censored times. In a more general setting, a logistic link is often used for p and covariates are often included in some form. Even in this simplest case, however, the survival function $S(t)$ is often not in closed form and typically there will not be closed form solutions for the maximum likelihood estimators of $\boldsymbol{\theta}$.

Kuo and Peng (1995) considered simplifying the likelihood by introducing two latent variables: one variable representing whether an individual is immune or susceptible, the other representing unknown failure times for censored susceptible individuals. Achcar and de Araujo Pereira (1999) used two latent variables in a Gibbs sampler of a (non-cure rate) mixture model. Peng and Dear (2000) used a simplified version, introducing a single latent variable representing whether an individual is immune or susceptible.

To implement the algorithm, let $\gamma = 1$ for susceptible individuals and $\gamma = 0$ for cured individuals. An uncensored failure time for the ith individual implies that $\gamma_i = 1$. The value of γ is unobserved for right censored individuals. The second latent variable represents the failure times. Let t_i be the (unknown) failure time for the right-censored ith individual conditional on that individual being susceptible. For uncensored individuals, t_i is the observed failure time. Let $\mathbf{t}_c$ be a vector containing the values of t_i for the observed and censored times. Incorporating these latent variables, the complete-data likelihood is

$$\mathcal{L}_c(\boldsymbol{\theta} \mid \boldsymbol{\gamma}, \mathbf{t}_c) = \prod_{i=1}^{n} [(1-p) f(t_i)]^{\gamma_i} \prod_{i=1}^{n} p^{1-\gamma_i}.$$

This yields a straightforward application of the EM algorithm. The complete-data log-likelihood is

$$\log \mathcal{L}_c(\boldsymbol{\theta} \mid \boldsymbol{\gamma}, \mathbf{t}_c) = \sum_{i=1}^{n} \gamma_i \log(1-p) + \sum_{i=1}^{n} (1-\gamma_i) \log(p) + \sum_{i=1}^{n} \gamma_i \log(f(t_i \mid \boldsymbol{\theta})).$$

Then,

$$Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(k)}) = \log(1-p) \sum_{i=1}^{n} E^{(k)}(\gamma_i) + \log(p) \sum_{i=1}^{n} \left[1 - E^{(k)}(\gamma_i)\right] + \sum_{i=1}^{n} E^{(k)}(\gamma_i) E^{(k)}(\log f(t_i)),$$

where $\gamma_i = 1$ for observed failure times and

$$E^{(k)}(\gamma_i) = E(\gamma_i \mid \boldsymbol{\theta}^{(k)}, t_i^*) = P(\gamma_i = 1 \mid \boldsymbol{\theta}^{(k)}, t_i^*) = \frac{(1-p^{(k)}) S(t_i^* \mid \boldsymbol{\theta}^{(k)})}{p^{(k)} + (1-p^{(k)}) S(t_i^* \mid \boldsymbol{\theta}^{(k)})}$$

for censored values. The maximization step maximizes $Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(k)})$. For any standard cure rate mixture model, where p is not a parameter in the failure time distribution of the susceptible proportion, the value of p that maximizes the expected complete-data likelihood at the next iteration is

$$p^{(k+1)} = 1 - \frac{\sum_{i=1}^{n} E^{(k)}(\gamma_i)}{n}.$$

Since $E^{(k)}(\gamma_i)$ represents the estimated probability that the ith individual is a susceptible at the kth iteration, this estimator of p is a natural estimator of the cure rate at the $(k+1)$st iteration. Note that since $0 \leq E^{(k)}(\gamma_i) \leq 1$, $0 \leq p^{(k+1)} \leq 1$.

3.4 The EM algorithm applied to the inverse Gaussian cure rate mixture model

Assume a standard cure rate mixture model with the inverse Gaussian distribution as the distribution of the failure times of the susceptibles:

$$f(t \mid \delta, \nu) = \frac{1}{\sqrt{2\nu\pi t^3}} e^{-(1-\delta t)^2/2t},$$

for $\delta, \nu > 0$. Consider first the case where there are no covariates present, as the algorithm is considerably simpler in this case. Given the parameter estimates at the kth iteration, $\boldsymbol{\theta}^{(k)} = (\delta, \nu, p)^{(k)}$, the expectation step involves finding

$$\begin{aligned} Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(k)}) = &\sum_{i=1}^{n} E^{(k)}(\gamma_i) \log(1-p) + \sum_{i=1}^{n} \left[1 - E^{(k)}(\gamma_i)\right] \log(p) \\ &+ \sum_{i=1}^{n} E^{(k)}(\gamma_i) E^{(k)}(\log f(t_i)). \end{aligned}$$

By solving $\partial Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(k)})/\partial \boldsymbol{\theta} = 0$ for p, δ, and ν, then at the next iteration,

$$\begin{aligned} p^{(k+1)} &= 1 - \frac{\sum_{i=1}^{n} E^{(k)}(\gamma_i)}{n}, \\ \delta^{(k+1)} &= \frac{\sum_{i=1}^{n} E^{(k)}(\gamma_i)}{\sum_{i=1}^{n} E^{(k)}(\gamma_i) E^{(k)}(t_i)}, \\ \nu^{(k+1)} &= \frac{\sum_{i=1}^{n} E^{(k)}(\gamma_i)\, E^{(k)}(1/t_i) - \frac{\left(\sum_{i=1}^{n} E^{(k)}(\gamma_i)\right)^2}{\sum_{i=1}^{n} E^{(k)}(\gamma_i) E^{(k)}(t_i)}}{\sum_{i=1}^{n} E^{(k)}(\gamma_i)}. \end{aligned}$$

For the expectation step, $E^{(k)}(\gamma_i)$, $E^{(k)}(t_i)$, and $E^{(k)}(1/t_i)$ are required. For right censored times,

$$E^{(k)}(\gamma_i) = E(\gamma_i \mid \boldsymbol{\theta}^{(k)}, t_i^*) = \frac{(1-p^{(k)}) S(t_i^* \mid \boldsymbol{\theta}^{(k)})}{p^{(k)} + (1-p^{(k)}) S(t_i^* \mid \boldsymbol{\theta}^{(k)})}.$$

Whitmore (1979) derived the following results:

$$E^{(k)}[t_i] = E(t_i \mid \boldsymbol{\theta}^{(k)}, t_i^*) = \frac{1 - S(1/t_i^*(\delta^{(k)})^2)}{\delta^{(k)} S(t_i^*)}, \tag{4}$$

$$E^{(k)}[1/t_i] = E(1/t_i \mid \boldsymbol{\theta}^{(k)}, t_i^*) = \nu^{(k)} + (\delta^{(k)})^2 E(t_i \mid \boldsymbol{\theta}^{(k)}, t_i^*) - \frac{2t_i^* \nu^{(k)} f(t_i^* \mid \boldsymbol{\theta}^{(k)})}{S(t_i^* \mid \boldsymbol{\theta}^{(k)})}. \tag{5}$$

To illustrate the speed of convergence, the EM algorithm was used to find the maximum likelihood estimates for the inverse Gaussian cure rate mixture model, using a data set of 100 observations simulated from this distribution. The assumed parameter values for the simulated data set are $\delta = 0.005$, $\nu = 0.05$ and $p = 0.3$. The individuals were subject to exponential censoring, resulting in 43 censored values, and 57 uncensored values. Starting values for δ and ν in the iteration scheme were found by treating the censored values as observed failure times and calculating the maximum likelihood estimates for this uncensored sample.

The starting value for p was chosen to be the proportion of censored values in the sample. Since there are closed form solutions for the updated parameter estimates, there is no need to implement one of the hybrid EM-Newton methods, but for comparison purposes the gradient EM was applied to the same data. Not surprisingly, the EM algorithm achieves near optimal likelihood much faster than the gradient EM algorithm.

In most cure rate applications, one of the major points of interest is to examine the effect of covariates on the cure rate and on the distribution of susceptibles. When covariates are present, the implementation of the EM algorithm is not as straightforward. Recall that the maximization step involves maximizing

$$Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(k)}) = \sum_{i=1}^{n} -E^{(k)}(\gamma_i) \log\left(\frac{p_i}{1 - p_i}\right) + \sum_{i=1}^{n} \log(p_i) + \sum_{i=1}^{n} E^{(k)}(\gamma_i) E^{(k)}(\log f(t_i)).$$

The cure rate p is commonly allowed to have a logistic-linear link with the covariates

$$\log\left(\frac{p_i}{1 - p_i}\right) = \boldsymbol{\tau}' \mathbf{x}_i,$$

where p_i is the probability that the ith individual is immune, $\boldsymbol{\tau}' = (\tau_0, \tau_1, \ldots, \tau_r)$ is a vector of $(r + 1)$ parameters and $\mathbf{x}_i' = (1, x_{i1}, x_{i2}, \ldots, x_{ir})$ is a vector of covariates corresponding to the ith individual. Then,

$$Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(k)}) = \sum_{i=1}^{n} -E^{(k)}(\gamma_i)\boldsymbol{\tau}'\mathbf{x}_i + \sum_{i=1}^{n} \log\left(\frac{e^{\boldsymbol{\tau}'\mathbf{x}_i}}{1+e^{\boldsymbol{\tau}'\mathbf{x}_i}}\right) + \sum_{i=1}^{n} E^{(k)}(\gamma_i)E^{(k)}(\log f(t_i)). \tag{6}$$

And $\partial Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(k)})/\partial\tau_j = 0$ implies that

$$\sum_{i=1}^{n}[1 - E^{(k)}(\gamma_i)]x_{ij} = \sum_{i=1}^{n} \frac{e^{\tau_j x_{ij}}}{1+e^{\tau_j x_{ij}}} x_{ij},$$

and there are no closed form solutions for the elements of $\boldsymbol{\tau}$. Note that the third summation term in (6) involves only parameters associated with the assumed failure time distribution, and not $\boldsymbol{\tau}$.

In the maximization step, estimation of the parameters of the failure time distribution can proceed separately from the estimation of $\boldsymbol{\tau}$. Depending on the assumed distribution for the susceptible proportion, there may be closed form solutions for the parameters of the failure time distribution in the M-step. Numerical techniques may be needed only to update the elements of $\boldsymbol{\tau}$. But since at each iteration of the EM algorithm a full M-step would require an iterative technique, it may be more efficient to use a hybrid (EM-Newton) method.

Consider the IGCRM with covariate information, where the drift parameter for the ith individual has a linear relationship with the covariates $\delta_i = \boldsymbol{\beta}'\mathbf{x}_i$, where $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_r)'$ is a vector of $(r+1)$ parameters. The IGCRM defined in Sect. 2.4 does not restrict δ to positive values for models involving covariates. In the case of no covariates, δ is restricted to positive values for the sake of simplicity, as any proportion of immunes accounted for by negative δ can simply be absorbed into the parameter p^*. However, when covariates are included, imposing the restriction $\delta > 0$ can be unnecessarily restrictive. A small disadvantage to allowing negative values of δ in the IGCRM is that the application of an EM-Newton algorithm is slightly more complicated. The application to the IGCRM of two EM-Newton algorithms based on different latent variables is discussed below.

The incomplete-data likelihood for the IGCRM is given by

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\tau}, \nu \mid \mathbf{t}) = \prod_{i \in D}[(1-p_i^*)f(t_i \mid \delta_i, \nu)] \prod_{i \in C}[p_i^* + (1-p_i^*)(1 - F(t_i^* \mid \delta_i, \nu))],$$

where $p_i^* = e^{\boldsymbol{\tau}'\mathbf{x}_j}/(1+e^{\boldsymbol{\tau}'\mathbf{x}_j})$ is the probability that the ith individual is not subject to the Wiener process. For $\delta_i > 0$, $p_i^* = p_i$ is the probability that the ith individual is immune.

With the addition of the latent variables γ and t defined above, the complete-data likelihood function is given by

$$\mathcal{L}_c(\boldsymbol{\beta}, \boldsymbol{\tau}, \nu \mid \boldsymbol{\gamma}, \mathbf{t}_c) = \prod_{i=1}^{n} [(1 - p_i^*) f(t_i \mid \delta_i, \nu)]^{\gamma_i} [p_i^* + I(\delta_i < 0)(1 - p_i^*)(1 - e^{2\delta_i/\nu})]^{1-\gamma_i},$$

where the indicator function $I(\delta_i < 0)$ takes on the values

$$I(\delta_i < 0) = \begin{cases} 1 & \text{for } \delta_i < 0 \\ 0 & \text{for } \delta_i \geq 0. \end{cases}$$

The expected complete-data log-likelihood is given by

$$Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(k)}) = \sum_{i=1}^{n} E^{(k)}(\gamma_i) E^{(k)} \left(\log \left\{ (1 - p_i^*) f(t_i \mid \delta_i, \nu) \right\} \right) + \sum_{i=1}^{n} E^{(k)}(1 - \gamma_i) \log[p_i^* + I(\delta_i < 0)(1 - p_i^*)(1 - e^{2\delta_i/\nu})].$$

For uncensored failure times, $\gamma_i = 1$. For censored times,

$$E^{(k)}(\gamma_i) = E\left(\gamma_i \mid \nu^{(k)}, \delta_i^{(k)}, p_i^{*(k)}, t_i^*\right) = 1 - \frac{p_i^{*(k)} + I(\delta_i^{(k)} < 0)(1 - p_i^{*(k)})(1 - e^{2\delta_i^{(k)}/\nu^{(k)}})}{1 - (1 - p_i^{*(k)}) F(t_i^* \mid \delta_i, \nu)}.$$

Calculation of $E(\log f(t_i))$ requires $E(t_i)$ and $E(1/t_i)$, and proceeds as above. The presence of the $\log[p_i^* + I(\delta_i < 0)(1 - p_i^*)(1 - e^{2\delta_i/\nu})]$ term in the $Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(k)})$ function results in somewhat lengthy derivative terms, and they are not presented here.

Alternatively, a further simplification of the likelihood can be achieved with an additional latent variable α. Let $\alpha_i = 0$ if individual i is subject to the Wiener process, and $\alpha_i = 1$ if individual i is not subject to the process. Then the complete-data likelihood simplifies to

$$\mathcal{L}_c(\boldsymbol{\beta}, \boldsymbol{\tau}, \nu \mid \boldsymbol{\gamma}, \boldsymbol{\alpha}, \mathbf{t}_c) = \prod_{i=1}^{n} [(1 - p_i^*) f(t_i)]^{\gamma_i(1-\alpha_i)} p_i^{*(1-\gamma_i)(\alpha_i)} \times [(1 - p_i^*)(1 - e^{2\delta_i/\nu})]^{(1-\gamma_i)(1-\alpha_i)}.$$

The expected complete-data likelihood is given by

$$Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(k)}) = \sum_{i=1}^{n} E^{(k)}(\gamma_i) E^{(k)}(1 - \alpha_i \mid \gamma_i = 1) E^{(k)}\{\log[(1 - p_i^*) f(t_i \mid \delta_i, \nu)]\}$$

$$+\sum_{i=1}^{n} E^{(k)}(1-\gamma_i)E^{(k)}(\alpha_i \mid \gamma_i = 0)\log p_i^*$$
$$+\sum_{i=1}^{n} E^{(k)}(1-\gamma_i)E^{(k)}(1-\alpha_i \mid \gamma_i = 0)\log[(1-p_i^*)(1-e^{2\delta_i/\nu})].$$

The resulting first and second derivatives that are required for implementation of a hybrid EM-Newton algorithm are much simpler in this case than for the method including only γ and t as the latent variables. Calculation of $E^{(k)}(\gamma_i)$, $E^{(k)}(t_i)$, and $E^{(k)}(1/t_i)$ can proceed as above. All susceptible individuals are subject to the process, implying $E^{(k)}(\alpha_i \mid \gamma_i = 1) = 0$, which simplifies the first summation term in $Q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(k)})$ above. For cured individuals:

$$E^{(k)}(\alpha_i \mid \gamma_i = 0) = E\left(\alpha_i \mid \nu^{(k)}, \delta_i^{(k)}, p_i^{*(k)}, \gamma_i = 0\right)$$
$$= \begin{cases} 1 & \text{for } \delta_i^{(k)} > 0 \\ p_i^{*(k)} / \left[p_i^{*(k)} + (1-p_i^{*(k)})(1-e^{2\delta_i/\nu})\right] & \text{for } \delta_i^{(k)} < 0. \end{cases}$$

The algorithms for the two hybrid EM-Newton methods for the IGCRM model with covariates will proceed identically if $\delta_i^{(k)} > 0$ for all i, k. In most practical cases, there will be little difference between the methods, as the drift parameter often remains positive when the parameter p^* is included in the model. To compare the speed of convergence for these two methods for cases with some $\delta_i < 0$, the EM1 algorithm was applied to a simulated data set. The simulated data results from an assumed IGCRM model with one covariate, measured at three levels. The slope β_1 was assumed to be negative, with resulting δ values of $(0.101, 0.001, -0.099)$. The values of p^* and ν were assumed to be 0.2 and 2, respectively, for all levels. We observed that optimal likelihood is achieved more quickly when the latent variable α is not included.

3.5 Hybrid methods for the DIG model

In this section, the hybrid maximization methods reviewed in Sect. 3.2 are applied to the DIG model. For the DIG model, the complete-data likelihood is given by

$$\mathcal{L}_c(\boldsymbol{\theta}|\boldsymbol{\gamma}, \mathbf{t}_c) = \prod_{i=1}^{n} \left[\frac{1}{\sqrt{2\pi\nu t_i^3}} e^{-(1-\delta_i t_i)^2/2\nu t_i}\right]^{\gamma_i} \left[1 - e^{2\delta_i/\nu}\right]^{1-\gamma_i}.$$

This complete-data likelihood has no closed form solutions for the maximum likelihood estimators of δ and ν, even in the simple case of no covariates. Consider a defective inverse Gaussian regression model where the drift parameter has a linear relationship with the covariates, $\delta_i = \boldsymbol{\beta}'\mathbf{x}_i$. To implement either the Newton or quasi-Newton hybrid EM algorithms to find the maximum likelihood estimates of $\boldsymbol{\theta} = (\delta, \nu)$, expressions for

$$\frac{\partial \log \mathcal{L}_c(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \frac{\partial^2 \log \mathcal{L}_c(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2},$$

are required. To save space, the necessary expressions are not reproduced herein. In the iteration scheme, γ_i and t_i corresponding to censored values are replaced by their conditional expected values. For censored values,

$$\begin{aligned} E^{(k)}(\gamma_i) &= E(\gamma_i \mid \nu^{(k)}, \delta_i^{(k)}) \\ &= \begin{cases} 1 & \text{for } \delta_i^{(k)} > 0 \\ \left[1 - F(t_i^* \mid \nu^{(k)}, \delta_i^{(k)}) - p_i^{(k)}\right] / \left[1 - F(t_i^* \mid \nu^{(k)}, \delta_i^{(k)})\right] & \text{for } \delta_i^{(k)} < 0 \end{cases} \end{aligned}$$

where $p_i^{(k)} = 1 - e^{2\delta_i^{(k)}/\nu^{(k)}}$. For right censored values, the conditional expectations of t_i and $1/t_i$, given the individual is a susceptible, are simply Eqs. 4, and 5, with $\delta_i^{(k)}$ replaced by $|\delta_i^{(k)}|$:

$$\begin{aligned} E^{(k)}\left(t_i \mid \delta_i^{(k)}, \nu_i^{(k)}, t_i^*\right) &= \frac{1 - S(1/t_i^*(\delta_i^{(k)})^2 \mid \nu^{(k)}, |\delta_i^{(k)}|)}{|\delta_i^{(k)}| S(t_i^* \mid \nu^{(k)}, |\delta_i^{(k)}|)} \\ E^{(k)}\left(1/t_i \mid \delta_i^{(k)}, \nu_i^{(k)}, t_i^*\right) &= \nu^{(k)} + (\delta_i^{(k)})^2 E(t_i \mid \delta_i^{(k)}, \nu_i^{(k)}, t_i^*) \\ &\quad - \frac{2t_i^* \nu^{(k)} f(t_i^* \mid \nu^{(k)}, |\delta_i^{(k)}|)}{S(t_i^* \mid \nu^{(k)}, |\delta_i^{(k)}|)}. \end{aligned}$$

To illustrate the speed of convergence for the different methods, the gradient EM and the quasi-Newton algorithm were applied to the DIG model, using the cancer data that was analyzed by Boag (1949). Figure 2 compares the observed data log-likelihoods at each step of the iteration scheme.

Note that for iterations 0 and 1, the methods are identical, as the first step of the quasi-Newton method is a one-step Newton. The results for this data set illustrate that the quasi-Newton approach can yield a faster rate of convergence.

3.6 A hybrid method for the DIGIGMIX model

This section uses an EM-type algorithm to find the maximum likelihood estimates for the DIGIGMIX model proposed in Sect. 2.6. The incomplete-data likelihood for the DIGIGMIX model is given by

$$\begin{aligned} \mathcal{L}(\pi, \delta_1, \delta_2, \nu_1, \nu_2 \mid \mathbf{t}) = \prod_{i \in D} [\pi f_1(t_i) + (1-\pi) f_2(t_i)] \prod_{i \in C} \big[\pi(1 - F_1(t_i^*)) \\ + (1-\pi)(1 - F_2(t_i^*))\big], \end{aligned}$$

where $f_1(t)$ is the density function of a proper inverse Gaussian ($\delta_1 > 0$) and $f_2(t)$ is the density function of a defective inverse Gaussian distribution ($-\infty < \delta_2 < \infty$).

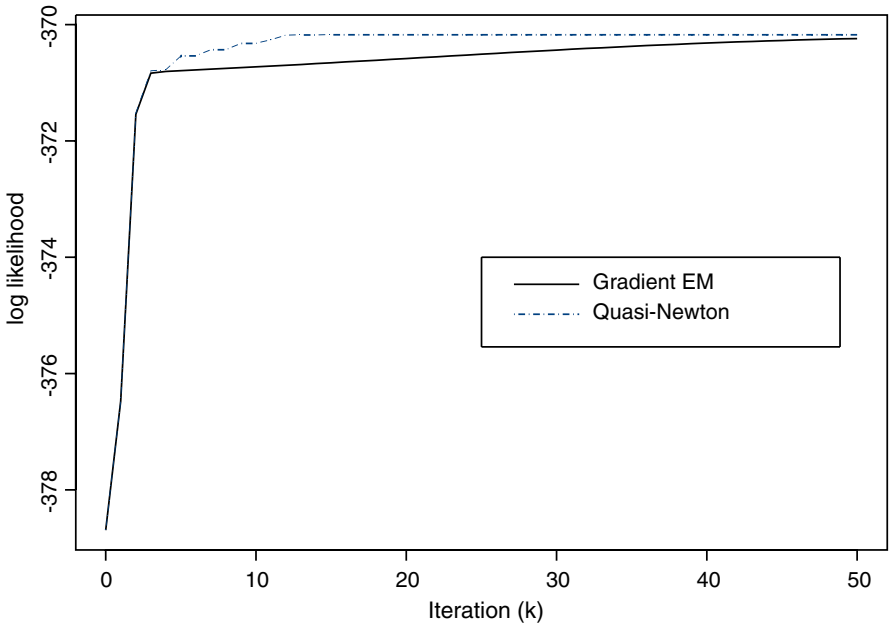

Fig. 2 Plot of $\log(\mathcal{L}(\boldsymbol{\theta}^{(k)}))$ versus k for the gradient EM and quasi-Newton algorithms applied to the DIG model

Their respective cumulative distribution functions are represented by $F_1(t)$ and $F_2(t)$. The form of the likelihood, as well as the non-closed form solutions for $F_1(t)$ and $F_2(t)$, make this a complicated likelihood requiring numerical techniques to maximize. In a similar approach as for the DIG and IGCRM models, latent variables are added to simplify the likelihood. Let the latent variable α represent whether an individual is in group one or two ($\alpha = 1$ for group one, $\alpha = 0$ for group two). Note that α is unobservable for all individuals. Let t_{1i} represent the failure time for the ith individual, conditional on that individual being in group one ($\alpha = 1$). Let t_{2i} represent the failure time for the ith individual, conditional on being in group two ($\alpha = 0$). The latent variables t_{1i} and t_{2i} are unobserved for censored times. For uncensored failure times, $t_{1i} = t_{2i} = t_i$. As above, let the latent variable γ_i represent the immunity of individual i ($\gamma_i = 1$ for susceptibles,$\gamma_i = 0$ for immunes). The complete-data likelihood is given by

$$\mathcal{L}_c(\boldsymbol{\theta} \mid \mathbf{t}) = \prod_{i=1}^{n} [\pi f_1(t_{1i})]^{\alpha_i} [(1-\pi) f_2(t_{2i})]^{(1-\alpha_i)\gamma_i} [(1-\pi)(1-e^{2\delta_2/\nu_2})]^{(1-\alpha_i)(1-\gamma_i)}.$$

The expected complete-data log-likelihood can be written as

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) = \sum_{i=1}^{n} E^{(k)}(\alpha_i) \log(\pi) + \sum_{i=1}^{n} (1 - E^{(k)}(\alpha_i)) \log(1-\pi)$$

$$+\sum_{i=1}^{n} E^{(k)}(\alpha_i)E^{(k)}(\log(f_1(t_{1i}))|\alpha = 1)$$
$$+\sum_{i=1}^{n} E^{(k)}(1-\alpha_i)E(\gamma_i|\alpha = 0)E^{(k)}(\log f_2(t_{2i})|\alpha = 0, \gamma = 1)$$
$$+\sum_{i=1}^{n}(1-e^{2\delta_2/\nu_2})E^{(k)}((1-\alpha_i))E^{(k)}(1-\gamma_i|\alpha = 0),$$

where $E^{(k)}(X)$ is the expected-value of X, conditional on the observed data, evaluated at the parameter estimates at the kth iteration. The required conditional expectations can be calculated as in the Appendix.

This section, and the supporting mathematical details in the Appendix, illustrate that the EM algorithm and the hybrid EM-Newton extensions can be used to effectively find maximum likelihood estimates in cure rate mixture models.

4 Data analysis

4.1 Introduction

This Section analyzes several data sets, with an emphasis on comparing the fit of the different Wiener process cure rate models discussed in Sect. 2. The fit of these models will also be compared to the more commonly used Weibull and lognormal cure rate mixture models. Goodness-of-fit will be assessed using well-established methods. The analysis will show that the Wiener process models can provide an improved fit when compared to the Weibull and lognormal mixture models for some data sets.

4.2 Dataset I: time to infection for burn patients

4.2.1 Background

Ichida et al. (1993) investigated the effect of a change in disinfectant practices for the time to infection for burn patients. *Staphylococcus aureus* infections are common in burn patients and can contribute to increased length of stay in hospital and even death. Not all patients experience such an infection and any survival analysis should incorporate the possibility of a cured proportion in the population. In the study, the time to occurrence of a *staphylococcus aureus* infection was recorded for 154 burn patients. One of the major interests of the study was to investigate the different rates of infection for two disinfectant bathing practices. Let X_1 be the disinfectant practice ($X_1 = 0$ for a 10% povidone-iodine solution, $X_1 = 1$ for a 4% chlorhexidine gluconate solution). Another important variable that affects staphylococcus infections is the percentage of body area that is burned. Let X_2 represent the percentage of body surface area burned.

For the cure rate mixture models, the cure rate is assumed to have a logistic-linear link with the covariates:

$$\log\left(\frac{p_i}{1-p_i}\right) = \gamma_0 + \gamma_1 x_{1i} + \gamma_2 x_{2i},$$

where $p_i = P(i\text{th individual is immune})$, x_{1i} is the bathing method for the ith individual and x_{2i} is the percentage of body area burned for the ith individual. For the Weibull mixture, the scale parameter λ was assumed to be a log-linear function of the covariates: $\log\lambda_i = \tau_0 + \tau_1 x_{1i} + \tau_2 x_{2i}$. For the lognormal model (with median lifetime e^{μ}), μ was assumed to be a linear function of the parameters: $\mu_i = \theta_0 + \theta_1 x_{1i} + \theta_2 x_{2i}$. For the Wiener process models, the drift parameter δ was assumed to be a linear function of the covariates: $\delta_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$. For the two barrier model, the absolute value of the lower barrier l was assumed to be a log-linear function of the covariates: $l_i = \exp\{\gamma_0 + \gamma_1 x_{1i} + \gamma_2 x_{2i}\}$.

4.2.2 Comparison of the different models

Table 1 illustrates the log-likelihood and the Akaike Information Criterion (AIC) for the full models. As a group, the Wiener process models provide a good fit to the data, under the full models. Judging by the AIC, they provide a better fit than either the Weibull or lognormal cure rate mixtures. The two barrier model has the largest log-likelihood of all the models, providing a better fit than the inverse Gaussian cure rate mixture. The improvements in the fit of the IGCRM and two barrier models over the DIG are fairly small, and the DIG results in the lowest value of the AIC.

Cox-Snell residual plots exhibit some curvature indicating imperfect fits for all of the models. Plots of deviance residuals show no obvious trends and no major problems with the fit of the models. There is one slight outlier, corresponding to an individual with a low percentage surface area that experienced an infection earlier than expected.

Figure 3 compares the survivor functions for the different Wiener process models, for the two different bathing methods. For each plot, the burn percentage is fixed at its mean (24.69%). The curves for the Wiener process models are similar over the range of the data.

Table 2 compares the estimates of the cure rate at different values of the covariates for the models. The defective inverse Gaussian model results in an estimated cure

Table 1 Log-likelihood and Akaike information criteria for the time to infection for burn patients

Model	Log-likelihood	# of Parameters	AIC
Weibull mixture	−245.4131	7	504.8262
Lognormal mixture	−242.1064	7	498.2129
IGCRM	−241.1666	7	496.3332
DIG	−241.4241	4	490.8481
Two barrier	−239.8341	7	493.6868

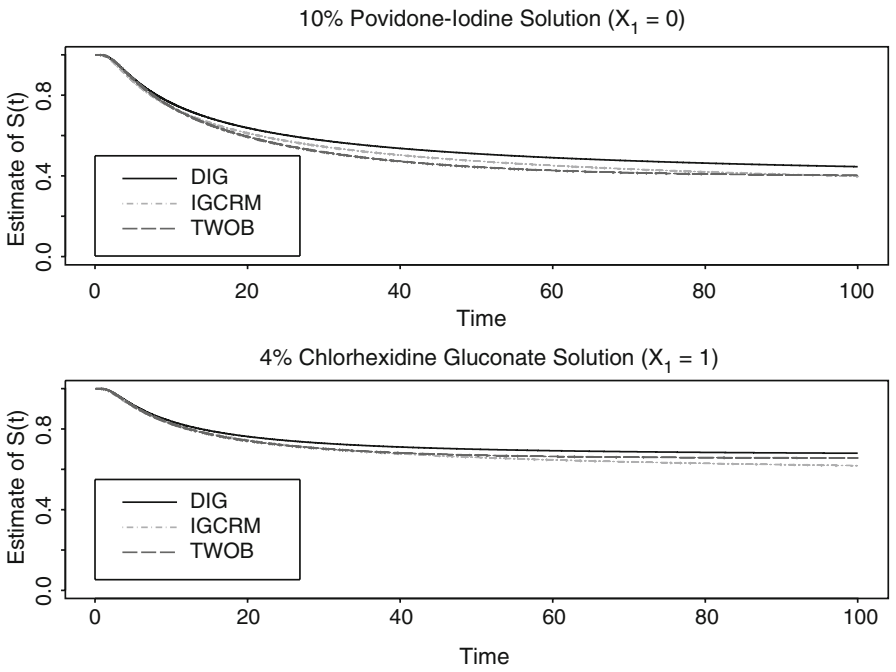

Fig. 3 Estimated survivor functions for the Wiener process models for the different bathing methods (burn percentage fixed at the mean)

Table 2 Comparison of estimated cure rates

	$X_1 = 0$		$X_1 = 1$	
	$X_2 = 10$	$X_2 = 70$	$X_2 = 10$	$X_2 = 70$
Weibull	0.577	0.191	0.767	0.362
Lognormal	0.547	0.198	0.736	0.363
DIG	0.494	0	0.734	0.401
IGCRM	0.470	0.105	0.686	0.549
TWOB	0.504	0.177	0.642	0.536

rate of 0 for patients with 70% of their body burned, under the 10% povidone-iodine disinfectant method ($X_1 = 0$). The estimated cure rate for the defective inverse Gaussian model can be very different from the other models and a point estimate of 0 is more common than for the other cure rate models.

As indicated by the AIC, the Wiener process models provide a good fit to this time to infection data, compared to the Weibull and lognormal mixtures. The Wiener process models provide similar fits over the range of the data, but they can differ a great deal in the estimated cure rate.

4.3 Data set II: a two barrier simulation

This section analyzes a data set simulated from a two barrier Wiener process. A simple method is used which approximates a random sample from this distribution. The

simulated data results from a Wiener process with $\delta = 0.005$ and $\nu = 0.05$, with an upper absorbing barrier at $u = 1$ and a lower absorbing barrier at $-l = -2$. The true cure rate for this population under these assumptions is $p = 0.269$. To approximate the process, each unit of time was split into one hundred segments, and the change in the process in each segment was generated from a normal distribution with mean 0.01δ and variance 0.01ν. The time at which the process first reached the upper barrier was recorded for 1000 individuals.

Lesosky and Horrocks (2004) found that for certain values of the parameters and time increments, this type of simulation method was not adequate to properly simulate from the two barrier distribution. However, the values used here should provide a sample that roughly approximates one from the two barrier distribution. For this simulation, Type I right censoring was used. If the process first reached the lower barrier, or had not reached either barrier by $t = 100$, the time was considered to be right censored at 100. Of the 1000 individuals, 315 were censored at 100.

4.3.1 Comparison of the models

Table 3 gives the values of the log-likelihoods and AIC. The IGNMIX and IGCRM models seem to provide the best overall fit. Surprisingly, the fit of these two models is marginally better than the two barrier model for this simulation. All of the Wiener process models yield similar fits over the range of the data. The Weibull and lognormal mixture models provide a poor fit to the data, judging by the AIC. The Weibull model yields an especially poor fit. This is evidenced by both the AIC and the systematic curvature in the goodness-of-fit plot. The plots for the other models show very straight lines.

The estimates of the cure rates for the different models are compared in Table 4. The estimate of the drift parameter for the DIG model is positive, resulting in an estimated

Table 3 Log-likelihood and AIC values for the two barrier simulation data

Model	Log-likelihood	No. of Parameters	AIC
Weibull mixture	−3549.477	3	7104.954
Lognormal mixture	−3511.889	3	7029.778
DIG	−3505.619	2	7015.238
IGCRM	−3503.000	3	7011.999
IGNMIX	−3502.989	3	7011.979
Two barrier	−3503.705	3	7013.411
DIGIGMIX	−3502.987	5	7015.973

Table 4 Estimates of the cure rate for the two barrier simulation

Model	$\hat{\mathbf{p}}$	Model	$\hat{\mathbf{p}}$
Weibull mixture	0.3031	IGNMIX	0.1592
Lognormal mixture	0.2668	Two barrier	0.2727
DIG	0	DIGIGMIX	0.1576
IGCRM	0.2227	True p	0.2693

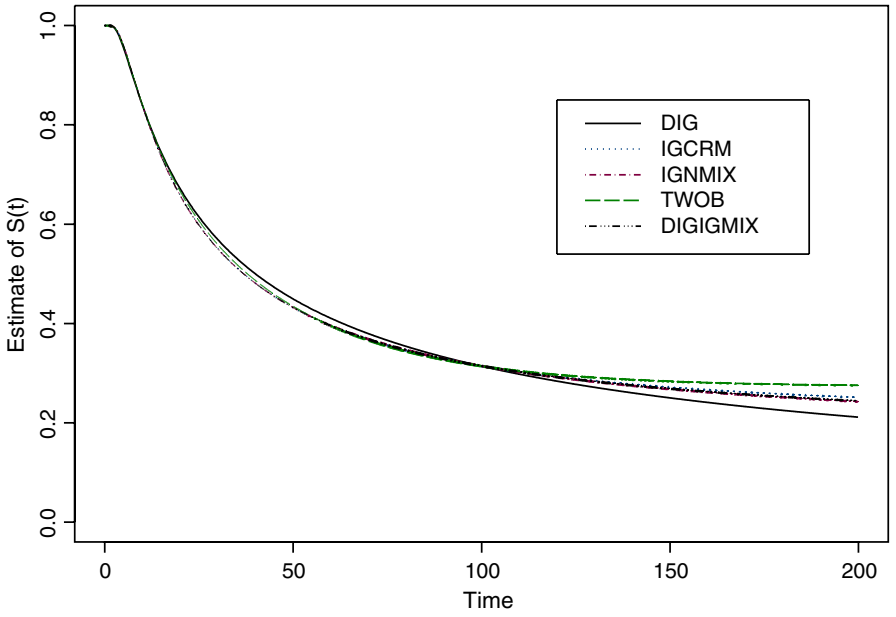

Fig. 4 Estimated survivor curves for the Wiener process models

cure rate of 0. When heterogeneity in the drift parameter is allowed (the IGNMIX model), the estimated cure rate is 0.1592, closer to the other models, but still relatively low. Not surprisingly, the two barrier model provided the best estimate of the cure rate.

Figures 4 and 5 illustrate the estimated survival and hazard functions for the Wiener process models. The estimated survival curves are similar over the range of the observed data, but differ in the right tail. Beyond the right extreme of the observed data ($t = 100$), the DIG survivor curve drops more quickly than the other models, reaching 0 in the limit.

All the Wiener process models fit this simulated data from a two barrier process reasonably well. The fit of Weibull and lognormal mixtures is considerably worse. The defective inverse Gaussian estimate of the cure rate is 0, quite different from the other models.

5 Conclusions

5.1 Comments on the fit of the models to the data sets

The defective inverse Gaussian model does have some initial appeal as a cure rate model, but it is somewhat lacking when it comes to actually estimating the cure rate. The model often provides a similar fit over the range of the data as the other Wiener process models, but it tends to have more area in the far right tail of the estimated

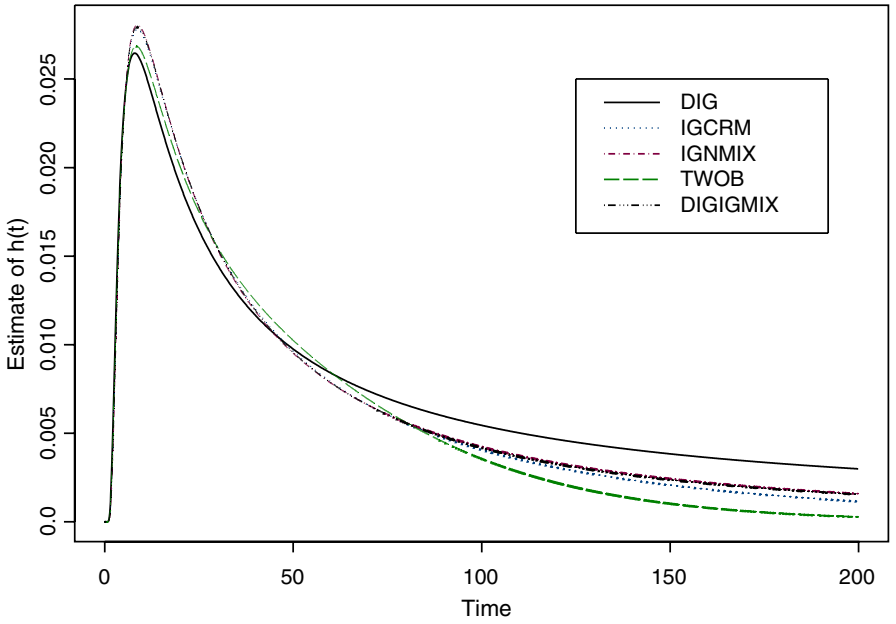

Fig. 5 Estimated hazard functions for the Wiener process models

density function for the population. This often leads to a lower estimate of p than the other models. Frequently, the estimated cure rate is 0.

The inverse Gaussian cure rate mixture model is a useful, more flexible alternative to the defective inverse Gaussian. One drawback to this model is that the interpretation of the survival times and the cure rate as the result of an underlying Wiener process is lost to a certain extent.

Allowing for heterogeneity in the drift parameter (the IGNMIX model) can sometimes improve the fit over the DIG, but often results in very similar fits to the DIG model. Many of the comments regarding the DIG also apply to the IGNMIX model. The estimated cure rate for this distribution is often much lower than the other models.

The two barrier model is a potentially useful generalization of the defective inverse Gaussian model. For the data sets analyzed above, the fit over the range of the data was often similar to the defective inverse Gaussian, but the estimated cure rate was often closer to the standard cure rate mixtures. This is also true of the inverse Gaussian cure rate mixture model, but the two barrier model has the advantage of preserving the interpretation of the survival times and cure rate as the realization of an underlying Wiener process.

The DIGIGMIX model is another interesting alternative to the DIG model. It is much more flexible, and its nature as a mixture of distributions allows it to provide a reasonable fit to some unusual data. The Wiener process models provided the best fit to the burn data. The fit of these models was better than the more commonly used Weibull and lognormal mixture models. These models have seen very limited

use as cure rate models, but should be considered as possible alternatives in many situations.

5.2 Extensions

Extensions of the models considered here to allow for frailty or heterogeneity between individuals are of considerable interest. Many hazard-based frailty models have been studied, generally *via* the introduction of a frailty random variable acting multiplicatively on the hazard function. Alternative models of frailty regarded as influencing an underlying process may frequently be more biologically plausible as persuasively argued by Aalen and Gjessing (2001). Specifically, these authors study the Wiener process with randomized drift and demonstrate its utility in incorporating covariates both internal and external (Kalbfleisch and Prentice, 2002). In addition, this model is a defective survival distribution allowing for a cured fraction. Aalen (1994) appears to be the first to recognize its potential as a frailty model in survival analysis. Here we have only considered the simple model with randomized drift denoted by IGNMIX in order to compare it with our models. However, all of the models discussed herein have potential to be extended to allow for unobserved heterogeneity.

Acknowledgements This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC). Specifically, in the form of a scholarship to the first author and respective Discovery Grants to the second and third authors. We are grateful to an associate editor and a referee for comments which have led to improvements in this work.

Appendix

Results for hybrid methods for the DIGIGMIX model

The required conditional expectations can be calculated using the following formulae.

$$E(\alpha_i)=P(\text{individual } i \text{ is in group 1})=\begin{cases}\dfrac{\pi f_1(t_i)}{\pi f_1(t_i)+(1-\pi) f_2(t_i)} & \text{for uncensored times.}\\[2ex] \dfrac{\pi S_1(t_i^*)}{\pi S_1(t_i^*)+(1-\pi)S_2(t_i^*)} & \text{for right censored times.}\end{cases}$$

For uncensored failure times $E(\gamma_i) = P(\text{individual } i \text{ is a susceptible}) = 1$. For censored times,

$$\begin{aligned}E^{(k)}(\gamma_i \mid \alpha = 0) &= E(\gamma_i \mid \alpha_i = 0, \delta_2^{(k)}, \nu^{(k)}, t_i^*)\\ &= \begin{cases}1 & \text{for } \delta_2^{(k)} > 0\\ \dfrac{1-F\left(t_i^* \mid \delta_2^{(k)}, \nu^{(k)}\right)-p_i^{(k)}}{1-F\left(t_i^* \mid \delta_2^{(k)}, \nu^{(k)}\right)} & \text{for } \delta_2^{(k)} < 0,\end{cases}\end{aligned}$$

where $p_i^{(k)} = 1 - e^{2\delta_2^{(k)}/\nu^{(k)}}$. Also,

$$E^{(k)}(\log f(t) \mid \alpha = 1) = -\frac{1}{2}\log(2\pi\nu_1^{(k)}) - \frac{1}{2}E^{(k)}(\log t^3 \mid \alpha = 0) - E^{(k)}\left(\frac{\left(1-\delta_1^{(k)}t\right)^2}{2\nu_1^{(k)}t}\right),$$

$$E^{(k)}(\log f(t) \mid \alpha = 0, \gamma = 1) = -\frac{1}{2}\log(2\pi\nu_2^{(k)}) - \frac{1}{2}E^{(k)}(\log t^3 \mid \alpha = 0) - E^{(k)}\left(\frac{\left(1-|\delta_2^{(k)}|t\right)^2}{2\nu_2^{(k)}t}\right).$$

For censored values, $E(t_{1i} \mid t_{1i} > t_i^*)$ and $E(1/t_{1i} \mid t_{1i} > t_i^*)$ are given by (4) and (5), with $\delta^{(k)}$ and $\nu^{(k)}$ replaced by $\delta_1^{(k)}$ and $\nu_1^{(k)}$, respectively. $E(t_{2i} \mid t_{2i} > t_i^*)$ and $E(1/t_{2i} \mid t_{2i} > t_i^*)$ are given by (4) and (5) with $\delta^{(k)}$ and $\nu^{(k)}$ replaced by $|\delta_2^{(k)}|$ and $\nu_2^{(k)}$, respectively. In a simple case involving no covariates, closed form solutions exist for the updated estimates of π, δ_1, and ν_1;

$$\pi^{(k+1)} = \frac{\sum_{i=1}^n E^{(k)}(\alpha_i)}{n},$$

$$\delta_1^{(k+1)} = \frac{\sum_{i=1}^n E^{(k)}(\alpha_i)}{\sum_{i=1}^n E^{(k)}(\alpha_i)E^{(k)}(t_{1i})},$$

$$\nu_1^{(k+1)} = \frac{\sum_{i=1}^n E^{(k)}(\alpha_i)E^{(k)}\left(\frac{1}{t_{1i}}\right) - \frac{\left(\sum_{i=1}^n E^{(k)}(\alpha_i)\right)^2}{\sum_{i=1}^n E^{(k)}(\alpha_i)E^{(k)}(t_{1i})}}{\sum_{i=1}^n E^{(k)}(\alpha_i)}.$$

No closed form solutions exist for the next iteration of δ_2 and ν_2. As in the DIG model, these estimates can be updated at each iteration using one step of a Newton iteration.

This hybrid EM-EM1 algorithm can be used to find the maximum likelihood estimates for the DIGIGMIX model, where π^k, δ_1^k, and ν_1^k can be found using their closed form estimators, and δ_2^k and ν_2^k can be updated using a one-step Newton procedure.

References

Aalen OO (1992) Modelling heterogeneity in survival analysis by the compound Poisson distribution. Ann Appl Probab 2:951–972

Aalen OO (1994) Effects of frailty in survival analysis. Stat Methods Med Res 3:227–243

Aalen OO, Gjessing HK (2001) Understanding the shape of the hazard rate: a process point of view. Stat Sci 16:1–22

Achcar J, de Araujo Pereira G (1999) Mixture models for type II censored survival data in the presence of covariates. Comput Stat 14:233–250
Berkson J, Gage RP (1952) Survival curve for cancer patients following treatment. J Am Stat Assoc 47: 501–515
Boag JW (1949) Maximum likelihood estimates of the proportion of patients cured by cancer therapy. J R Stat Soc B 11:15–34
Cantor AB, Shuster JJ (1992) Parametric versus non-parametric methods for estimating cure rates based on censored survival data. Stat Med 11:931–937
Chen M-H, Ibrahim JG, Sinha D (1999) A new Bayesian model for survival data with a surviving fraction. J Am Stat Assoc 94:909–919
Chhikara RS, Folks JL (1978) The inverse Gaussian distribution and its statistical application-a review. J R Stat Soc B 40:263–289
Cox DR (1972) Regression models and life-tables (with discussion). J R Stat Soc B 34:187–220
Cox DR (1999) Some remarks on failure-times, surrogate markers, degradation, wear, and the quality of life. Lifetime Data Anal 5:307–314
Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm (with discussion). J R Stat Soc B 39:1–38
Desmond AF, Chapman GR (1992) Likelihood inference for inverse Gaussian mixtures. Technical report, Department of Mathematics and Statistics, University of Guelph, Ontario, Canada
Desmond AF, Chapman GR (1993) Modelling task completion data with inverse Gaussian mixtures. Appl Stat 42(4):603–613
Desmond AF, Yang ZL (2006) Score tests of goodness-of-fit for the inverse Gaussian distribution. Technical report, Department of Mathematics and Statistics, University of Guelph
Duchateau L, Janssen P (2008) The frailty model. Springer, New York
Farewell VT (1982) The use of mixture models for the analysis of survival data with long-term survivors. Biometrics 38:1041–1046
Farewell VT (1986) Mixture models in survival analysis: are they worth the risk. Can J Stat 14:257–262
Feller W (1986) An introduction to probability theory and its applications. Wiley, New York
Gieser PW, Chang MN, Rao PV, Shuster JJ, Pullen J (1998) Modelling cure rates using the Gompertz model with covariate information. Stat Med 17:831–839
Haybittle JL (1959) The estimation of the proportion of patients cured after treatment for cancer of the breast. Br J Radiol 32:725–733
Ichida J, Wassell J, Keller M, Ayers L (1993) Evaluation of protocol change in burn-care management using the Cox proportional hazards model with time-dependent covariates. Stat Med 12:301–310
Kalbfleisch JD, Prentice RL (2002) The statistical analysis of failure time data, 2nd edn. Wiley, New York
Knight F (1981) Essentials of Brownian motion and diffusion, math surveys and monographs no. 18. Amer Math Soc, Providence, Rhode Island
Kuk AYC, Chen CH (1992) A mixture model combining logistic regression with proportional hazards regression. Biometrika 79:531–541
Kuo L, Peng F (1995) A mixture-model approach to the analysis of survival data. Technical report 95–31, Department of Statistics, University of Connecticut
Lancaster T (1972) A stochastic model for the duration of a strike. J R Stat Soc A 135:257–271
Lange K (1995a) A gradient algorithm locally equivalent to the EM algorithm. J R Stat Soc B 57:425–437
Lange K (1995b) A quasi-Newton acceleration of the EM algorithm. Stat Sin 5:1–18
Lawless JF (2003) Statistical models and methods for lifetime data. 2nd edn. Wiley, New York
Lee M-LT, Whitmore GA (2004) First hitting time models for lifetime data. In: Balakrishnan N, Rao CR (eds) Handbook of statistics, vol 23. Elsevier, Amsterdam, pp 537–543
Lee M-LT, Whitmore GA (2006) Threshold regression for survival analysis: modeling event times by a stochastic process reaching a boundary. Stat Sci 21(4):501–513
Lesosky M, Horrocks J (2004) Generating random variables from the inverse Gaussian and first-passage-two-barrier distributions. Technical report, Department of Mathematics and Statistics, University of Guelph
Longini IN, Halloran ME (1996) A frailty mixture model for estimating vaccine efficacy. Appl Stat 45: 165–173
Maller RA, Zhou X (1996) Survival analysis with long-term survivors. Wiley, West Sussex
McLachlan GJ, Krishnan T (2008) The EM algorithm and extensions, 2nd edn. Wiley, New York
Pelsser A (2000) Pricing double barrier options using laplace transforms. Finance and Stochastics 4:95–104

Peng Y, Dear KBG (2000) A nonparametric mixture model for cure rate estimation. Biometrics 56:237–243
Peng Y, Dear KBG, Denham JW (1998) A generalized F mixture model for cure rate estimation. Stat Med 17:813–830
Rai SN, Matthews DE (1995) Improving the EM algorithm. Biometrics 49:587–591
Schrödinger E (1915) Zur theorie der fall-und steigversuche an teilchen mit brownscher bewegung. Phys Ze 16:289–295
Sy JP, Taylor JMG (2000) Estimation in a Cox proportional hazards cure model. Biometrics 56:227–236
Taylor JMG (1995) Semi-parametric estimation in failure time mixture models. Biometrics 51:899–907
Tweedie MCK (1945) Inverse statistical variates. Nature 155:453
Whitmore GA (1979) An inverse Gaussian model for labour turnover. J R Stat Soc A 142:468–478
Whitmore GA (1986) Normal-gamma mixtures of inverse Gaussian distributions. Scand J Stat 13:211–220
Whitmore GA, Su Y (2007) Modeling low birth weights using threshold regression: results for U.S. birth data. Lifetime Data Anal 13:161–190
Whitmore GA, Crowder MJ, Lawless JF (1998) Failure inference from a marker process based on a bivariate Wiener model. Lifetime Data Anal 4:229–251
Yakovlev AY, Asselain B, Bardou VJ, Fourquet A, Hoang R, Rochefediere A, Tsodikov AD (1993) A simple stochastic model of tumor recurrence and its applications to data on premenopausal breast cancer. In: Asselain CDCLJPMB, Boniface M, Tranchefort J (eds) Biometrie et Analyse de Donnees Spatio-Temporelles, pp 66–82