Published in final edited form as:

Lifetime Data Anal. 2010 October ; 16(4): 525–546. doi:10.1007/s10985-010-9180-y.

Recurrent events and the exploding Cox model

Håkon K. Gjessing,

Division of Epidemiology, Department of Genes and Environment, Norwegian Institute of Public Health, Oslo, Norway; Department of Public Health and Primary Health Care, University of Bergen, Bergen, Norway

Kjetil Røysland,

Department of Biostatistics, Institute of Basic Medical Sciences, University of Oslo, Oslo, Norway

Edsel A. Pena, and

Department of Statistics, University of South Carolina, Columbia, USA

Odd O. Aalen

Department of Biostatistics, Institute of Basic Medical Sciences, University of Oslo, Oslo, Norway

Abstract

Counting process models have played an important role in survival and event history analysis for more than 30 years. Nevertheless, almost all models that are being used have a very simple structure. Analyzing recurrent events invites the application of more complex models with dynamic covariates. We discuss how to define valid models in such a setting. One has to check carefully that a suggested model is well defined as a stochastic process. We give conditions for this to hold. Some detailed discussion is presented in relation to a Cox type model, where the exponential structure combined with feedback lead to an exploding model. In general, counting process models with dynamic covariates can be formulated to avoid explosions. In particular, models with a linear feedback structure do not explode, making them useful tools in general modeling of recurrent events.

Keywords

Recurrent events; Cox regression; Explosion; Honest process; Birth process; Aalen regression; Stochastic differential equation; Lipschitz condition; Feller criterion; Martingale problem

1 Introduction

Our ability to analyze complex event history data is much more limited than is commonly admitted. There exists an extensive theory for analyzing survival and event history data, but mostly these methods cover relatively simple situations. When we are confronted with complex data where the structure of event histories must be analyzed, like how future developments depend on previous events in the process, existing theory and methods are

limited. The great majority of published analyses are of the standard survival type, possibly with competing risks (although even this is rare in practice). When more complex data are analyzed, this is often done by breaking the problem up into smaller pieces. There are methods for handling recurrent event data (e.g., Cook and Lawless 2007 and Aalen et al. 2008), but not necessarily the more complex information often accompanying such data. There exists methodology for so-called causal analysis of time-dependent confounding, like marginal structural models. But these methods are still in an early phase, they are not necessarily easy to use, and they have limitations related for instance to instability of weighting procedures.

The limited availability of appropriate methods is an increasing problem since the data available for analysis is becoming much more extensive. Studies in epidemiology and social science are increasingly based on cohorts and registries where several events of different types are registered for each individual, together with a lot of other information. There is a mixture of events of various types (e.g., going in and out of jobs, in and out of hospitals, moving between different medical states, etc.) and longitudinal data (e.g., clinical data, psychiatric data, social measures or biomarkers). The same is true for clinical studies: more and more hospitals register all patient information electronically, and only a tiny part would be used in a clinical trial or other clinical research.

Hence, one has joint observation of (possibly) recurrent events of several types and of longitudinal data; the goal should be to give an integrated analysis. There are two problems here: first, how to analyze several types of events, and second, how to combine events and measurements. The tradition in statistics has been to regard event history analysis and longitudinal analysis as two separate fields. In survival analysis, the events always take precedence, and the measurement data collected, i.e., the covariates, are relegated to a secondary role. Covariates are used in a regression model, typically a Cox model, but their internal structure (for instance, causal relations among them) is not analyzed. Internal time-dependent covariates are often seen as a problem because they can create bias when introduced into a Cox model. However, we believe that internal covariates should rather be seen as a challenge to increase our understanding, which could be achieved if the longitudinal and event data were set on an equal footing. See also a discussion of the importance of jointly modeling time-to-event and longitudinal marker data in Tsiatis and Davidian (2004) and Han et al. (2007) which introduced and studied a latent class-based joint model for recurrent events and longitudinal markers.

A proper analysis of this type of data requires that they are treated as stochastic processes; see e.g., Aalen et al. (2008). Time must be properly handled in the model; a fruitful point of view is the French school of process theory, where the emphasis is on local characteristics that explain how the past influences the present and the future. This dynamic approach is also closely related to a causal view. However, if data are viewed as realizations of stochastic processes, one has to take care that the stochastic process models are well defined. In general this is a non-trivial issue, in particular since some common models actually define processes that explode in finite time. The issue of explosions in this context was explored in Aalen et al. (2008, Sect. 8.6.3). In this paper we will give a more comprehensive discussion of conditions that guarantee non-explosion and thus ensure well-defined processes. Our

exposition is informal, with a focus on bringing forward and connecting results that are already known, but not frequently seen in the present context.

2 Recurrent events and covariate processes

Recurrent event models may be generally formulated as counting processes. Consider k individuals, and for a typical individual define a counting process $N(t)$, which counts the number of events experienced by that individual up to time t . The “proneness” to new events for an individual is conveniently described by the *intensity process* $\lambda(t)$, heuristically defined by the relation

$$\lim_{h \rightarrow 0} \frac{1}{h} P \{N(t+h) - N(t) \geq 1 | \mathcal{F}_{t-}\} = \lambda(t). \quad (1)$$

Here, $\mathcal{F}_{t-}$ is the collected history before time t . The intensity process $\lambda(t)$ is thus nonnegative and depends only on the past history of N , i.e., the jumping times of N before t . Since the value of $\lambda(t)$ determines the immediate probabilistic future of $N(t)$, it is natural to base statistical models on λ by letting it depend also on various covariates of interest. For instance, following the Cox tradition (Cox 1972) of exponential dependence on the covariates we may write

$$\lambda(t) = Y(t) \alpha(t) \exp(\beta X(t)), \quad (2)$$

where $\alpha(t)$ is a deterministic function, $X(t) = \{X_1(t), \dots, X_p(t)\}$ is a vector of possibly time-dependent stochastic covariate processes, and $Y(t)$ is an indicator equal to 1 if the process is under observation at time t . This general formulation is often termed the Andersen–Gill model. The difference from the standard survival model is that in the standard model only one event could be observed, so that $Y(t)$ would be zero after the first event. In the Andersen–Gill setting many events may be observed. In the present case $Y(t)$ could function as a censoring indicator, but we will not use this explicitly in the following.

An important issue is to what extent the intensity process is dependent on the past development of the counting process itself. Often this is assumed not to be the case, and the process is then simply a Poisson process. Note that one still may have quite complicated models with the covariate processes depending on various types of external information. A useful formulation of the necessary requirement to get a simple Andersen–Gill model is that the set of processes $\{X_1(t), \dots, X_p(t)\}$ is *locally independent* of $\{N_1(t), \dots, N_k(t)\}$ (Aalen et al. 2008). Surprisingly little attention has been given in the literature to the case where there are more complex dependencies.

It is important to distinguish between external and internal covariates. We call a covariate (process) *external* if it is locally independent of all other processes in the system. An illustration is given in the local dependence plot of Fig. 1, where X_1 is an external covariate because it is not influenced by any other process within the system.

Let us now consider *internal* covariates, which may be of two types: either processes that are locally dependent on other covariate processes, or *dynamic* covariates that for each time t are functions of the past of the counting process. (One may also have combinations of the

two.) The two possibilities are illustrated in the extended local dependence plots in Fig. 2. The left panel in the plot shows an internal covariate process $X_2(t)$. This is internal because the influence of the process $X_1(t)$ is partly going through $X_2(t)$, implying that these two covariates “compete” for attention in the statistical analysis. The right panel shows an internal covariate that is dynamic, that is, $X_2(t)$ is a function of the counting process $N(t)$.

Dynamic covariates, as shown e.g. in the right panel of Fig. 2, indicate a feedback mechanism at work in the counting process. The process is no longer of Poisson type. While relatively little attention has been paid to such more general counting processes, there is still a number of authors who have extended the counting process idea to intensity processes that depend on the past (Aalen et al. (2008, Chapter 8); Peña 2006; Peña et al. 2007; Stocker and Peña 2007; Gandy and Jensen 2004; Miloslavsky et al. 2004; Kalbfleisch and Prentice 2002).

3 Honest processes, feedback and explosions

Mathematically, introducing dependence on the past goes through the intensity process, which should be functionally dependent on past events. This can be formulated as dynamic covariates computed from actual observations. In applications, these covariates may be complex functions of the past. Note that the requirements for a statistical analysis are very different from the requirements needed to understand the probabilistic nature of the model. In the latter case simplicity of the dependence structure is required, but for statistical analysis the important issue is how the intensity process depends on statistical parameters. However, there is one important requirement of a probabilistic kind, namely that the processes must be well defined and actually exist as mathematical objects. That is, they must be “honest” in the terminology of Cox and Miller (1965).

3.1 Explosions and honesty

Let $0 < T_1 < T_2 < \dots$ be a strictly increasing sequence of random variables, and let $N(t) := \sum_{i=1}^{\infty} 1\{t \geq T_i\}$ be the corresponding counting process, i.e., T_i is the time of the i 'th jump of $N(t)$. Define $T := \lim_{i \rightarrow \infty} T_i$. The process $N(t)$ is said to be *honest* if we have $T = \infty$ with probability one; that is, if $N(t)$ is finite with probability one for all t . In contrast, a *dishonest* process will have $P(T < \infty) > 0$, which means that $P(N(t) = \infty) > 0$ for t large enough. In this situation T is called the *explosion time* of the process $N(t)$. In this paper, we will refer interchangeably to honest processes and non-exploding processes. In some instances it may be natural to consider only non-explosion on a finite time interval $0, \tau]$, that is, $N(t)$ is finite a.s. for all $t \leq \tau$, but most of our results relate to the full time interval $[0, \infty)$.

If $N(t)$ is used as a model for counting actual physical events, an explosion would indicate that the system at some point becomes chaotic or uncontrollable. While this could certainly be the case for some real world phenomena, such as an epidemic, most biomedical applications will focus on recurring events such as repeated hospitalizations, repeated infections, etc., that will not exhibit explosive behavior. Consequently, the probabilistic property of non-explosion is important for the overall structure of the model. When fitting an explosive model to non-explosive data by parameter estimation, one may well obtain a reasonably good fit. However, there is a risk that in the fitting process the model parameters

will be forced away from their “true” values simply to avoid parameter regions where the model is explosive and thus cannot provide an acceptable fit. It is clear that inherently explosive models should best be avoided in most applications of event history analysis to repeated events.

Including the past history of N into the model for the intensity process as described in Sect. 2 introduces feedback in the system. This is a natural way of describing individual proneness to events, and is a good alternative to the customary modeling of individual frailty as a random variable directly influencing the intensity. However, feedback also carries an inherent risk of explosion; when events recur more frequently the feedback gets even stronger and the intensity increases even further. The difficulty lies in choosing a model that allows feedback, but in a controlled manner.

3.2 Explosions in the Cox model

Consider, for example, the Andersen-Gill model defined in (2) and assume that one of the covariate processes equals the number of previous jumps in the counting process, i.e., $X_1(t) = N(t-)$, with a positive β coefficient. With the exponential dependence on the covariates in the Cox formulation, it seems clear that even this simple model will produce too much feedback and create an explosion. As an illustration we simulate this process, with an additional time-dependent covariate process $X(t)$, so that the intensity is defined as

$$\lambda(t) = \alpha \exp(\beta N(t-) + bX(t)). \quad (3)$$

The covariate process $X(t)$ is simulated as an Ornstein-Uhlenbeck process, described in more detail below. In Fig. 3 we see that even if $X(t)$ introduces a measure of randomness, all paths inevitably explode due to the feedback mechanism. The time of explosion, and the behavior of the process prior to explosion, will of course depend on the parameters, but explosion is still inevitable. As exemplified in Kalbfleisch and Prentice (2002, chapter 9), this type of feedback modeling seems natural in a Cox regression setting, but may well lead to ill-defined models.

4 The Feller criterion for non-explosion

From the above simulations, it seems clear that the Cox formulation with exponential feedback is guaranteed to explode. Indeed, assume that we can write the intensity process as

$$\lambda(t) = h(N(t-)) \quad (4)$$

for some non-negative function h . That is, N is a *pure birth* process. There is then a simple condition due to Feller (1968), which says that explosion occurs if and only if

$$\sum_{k=0}^{\infty} \frac{1}{h(k)} \text{ converges.} \quad (5)$$

An example thus causing explosion is $h(x) = x^2$, and exponential functions are clearly even worse. In a simple linear case where, for instance, $\lambda(t) = \alpha N(t-) + \beta$ with nonnegative parameters α and β , the sequence in (5) converges, and no-explosions can occur.

Note the intuitive aspect of the Feller condition: when $N(t)$ enters state k the intensity remains constant at $h(k)$ until the next event, and thus the expected time in state k is $1/h(k)$. If (5) holds, it means that the expected total time until explosion is finite. It is intuitively clear that as long as Feller's criterion for non-explosion is satisfied, this is sufficient to guarantee that the process $N(t)$ is otherwise well defined and can be constructed from a series of jump times where the time from one jump to the next is exponentially distributed. For details, see Kallenberg (1997, Theorem 10.19).

4.1 The Jacobsen condition

Feller's condition is very useful in the simple case (4) where the intensity process is a function of $N(t-)$. However, this is clearly a restrictive situation which is of limited modeling value. For instance, if

$$\lambda(t) = \alpha N(t-) + \beta x(t) \quad (6)$$

where $x(t)$ is a deterministic time-dependent covariate, one would still expect no explosions if, for instance, $x(t)$ is bounded by a constant. A condition by Martin Jacobsen (1982, Proposition 1.4.5) relaxes the required condition for non-explosion. Let $N(t)$ be a counting process and let $\lambda(t)$ be a process satisfying some additional regularity conditions. Assume there exists a nonnegative function h such that

$$\lambda(t) \leq h(N(t-)) \text{ and } \sum_{k=0}^{\infty} \frac{1}{h(k)} = \infty. \quad (7)$$

Then there exists a probability measure P on the underlying probability space such that λ is the intensity of N with respect to P . Note that since $N(t)$ is finite for every $t \geq 0$, this ensures the existence of a model (N, λ, P) , where λ is the intensity of N and where no explosions occur. Note that the existence of P is actually a consequence of the condition.

The Jacobsen condition (7) thus extends the Feller condition to a more general sufficient condition for the existence of a non-exploding model. In particular, the linear model (6) is clearly covered by the Jacobsen condition when $x(t)$ is bounded. More generally, many time-inhomogeneous Markov models will be covered, and even possibly non-Markov models.

5 Lipschitz conditions and stochastic differential equations

The heuristic description of an intensity process given in (1) is often formulated more rigorously using martingale theory. In general, given a probability measure P , the intensity process λ is the process that makes $M(t) = N(t) - \int_0^t \lambda(s) ds$ a (local) martingale under P . Or at least this will be true up until a possibly finite explosion time. In differential form this is written $dN(t) = \lambda(t)dt + dM(t)$. Assuming there is feedback as a function of N , we can write $\lambda(t) = h(N(t-), t)$ for a suitable function h . This leads to the stochastic differential equation (SDE)

$$dN(t) = h(N(t-), t) dt + dM(t) \quad (8)$$

for the process N . An SDE like (8) can be interpreted in different ways. One might ask for *weak* or *strong* solutions. Loosely speaking, a weak solution requires finding processes N , M and a probability space such that (8) is satisfied, and M is a martingale on that probability space. A strong solution asks for more: a solution process N must be found for a given martingale M on a given probability space. For statistical modeling, a weak solution should suffice; it is only the distribution of the process that matters, not on what probability space it is constructed. Existence (and uniqueness) of solutions to SDEs is a much-studied topic in the probability literature; a number of conditions for both weak and strong solutions have been derived. Even though strong solutions may be more than what we really need, the standard conditions for existence and uniqueness are very simple and intuitive. We will only give a brief outline here, for details see e.g., Protter (1990), Jacod and Shiryaev (1987), or Applebaum (2004). Note that we derived (8) just by rewriting the definition of $M(t)$. In the Appendix we show in more detail why this is justified.

A function $h(x, t)$ is *Lipschitz* (in x) if

$$|h(x_2, t) - h(x_1, t)| \leq K|x_2 - x_1|, \quad \forall x_1, x_2 \quad (9)$$

for some constant K (independent of t). Clearly, a linear function of the type $h(x, t) = xh_0(t)$ will be Lipschitz if h_0 is bounded. A Lipschitz condition is very similar to a differentiability

condition. Indeed, if $\frac{\partial}{\partial x} h(x, t)$ exists and is bounded, the mean value theorem implies that (9) is satisfied. However, simply asking for differentiability is not enough; the function $h(x) = x^2$ satisfies (9) locally, but h' is not bounded, so h is not Lipschitz. We see that (9) also implies a global linear growth restriction, which appears e.g., as a restriction on $h(x_2, t)$ when setting $x_1 = 0$.

The useful consequence of a Lipschitz condition is that it guarantees the existence and uniqueness of a strong non-exploding solution to (8), see Protter (1990, Theorem V.3.7). In fact, in general it is the growth restriction that gives the existence of a non-exploding solution, the smoothness (“differentiability”) part of (9) is what establishes the uniqueness. Note that it is also common to impose a separate linear growth condition in addition to the Lipschitz condition. As we show below, the Lipschitz condition alone will sometimes suffice. See also Applebaum (2004, Exercise 6.2.2).

The question is now in what way the SDE formulation relates to the Feller and Jacobsen criteria, and what advantages it may have, if any. Note that since the counting process N in our examples is a discrete-valued process, the smoothness dictated by the Lipschitz condition is a moot point; the function $h(x, t)$ needs to be defined only for $x = 0, 1, 2, \dots$, its behavior in between is irrelevant. As a consequence, it is the growth restriction implied by (9) that really matters. To explore this, keep x_1 fixed and assume that $h(x_1, t)$ is bounded as a function of t . This leads to the restriction

$$|h(x_2, t) - h(0, t)| \leq K|x_2|, \quad \forall x_2,$$

which again, since h and x in our applications are non-negative, implies that

$$h(x, t) \leq Kx + K_0, \quad \forall x \geq 0,$$

for some constant K_0 . Thus, with the additional condition that $h(x_1, t)$ is bounded in t for some x_1 we can conclude that the Lipschitz condition is a stronger requirement than the Jacobsen condition, and in particular, if h does not depend on t , it requires more than the Feller condition. Like the Jacobsen condition it is not a necessary condition for non-explosion.

Nevertheless, the SDE formulation offers more than just a reformulation of the problem. Indeed, it extends very naturally to multivariate processes, including covariate processes that are not counting processes. We will return to this in Sect. 7.2. Other extensions that may be useful in applications are to allow h to be random, and models with *functional* coefficients in the SDE, i.e., coefficients that depend on the full path of $N(s)$ just before time t , not only on the value of $N(t-)$. One such example would be $\lambda(t) = \int_0^t b(N(s), s) ds$, where the function $b(x, s)$ itself satisfies a Lipschitz condition. More details can be found in Protter (1990, Chapter V).

6 Examples of deterministic approximations

As an alternative to using the Feller and Jacobsen conditions, and the Lipschitz condition, some feedback models can be checked for explosions by analyzing a corresponding deterministic approximation. In the following we give an intuitive outline of this approach, with some illustrating examples. For this and other examples of deterministic approximations see also Cox and Miller (1965). We start with the simple model defined in (4). If h is bounded, no explosion can occur, so let us for simplicity assume $h(0) > 0$ and $\lim_{x \rightarrow \infty} h(x) = \infty$. If $N(t)$ approaches a point of explosion, $h(N(t-))$ becomes large, and since $dN(t)$ is Poisson with mean $h(N(t-))dt$, its standard deviation would suggest that for large values of $N(t)$, it can be approximated by the solution $f(t)$ to the corresponding deterministic differential equation

$$df(t) = h(f(t)) dt, \quad (10)$$

(possibly depending on the starting value $f(0) = c$). In particular, one would expect explosions of N to be determined by explosions of (10). It should be stressed, however, that the approximation is by no means absolute; it only holds when $N(t)$ is large, so the deterministic approximation and the actual process may be quite different from one another for smaller values. The existence of non-exploding solutions of a deterministic differential equation like (10) is of course also guaranteed by Lipschitz type conditions (Birkhoff and Rota 1989), but the advantage is that sometimes the equations can be solved explicitly. Indeed, since (10) is separable, we see that $f(t)$ is implicitly defined as the solution of

$$\int_c^{f(t)} \frac{1}{h(u)} du = t$$

for a given t . Define the function

$$H(y) = \int_c^y \frac{1}{h(u)} du \quad (11)$$

and note that H is increasing with $H(c) = 0$. The equation for f is thus $H(f) = t$. To guarantee a non-exploding solution f for all $t > 0$ it is clearly necessary and sufficient that

$$H(\infty) = \int_c^\infty \frac{1}{h(u)} du = \infty. \quad (12)$$

If, on the other hand, $H(\infty) = \tau < \infty$, $f(t)$ will explode when t approaches τ . Note that the time of explosion τ will depend on the starting value $c = f(0)$.

We see that condition (12) is in fact equivalent to the Feller-condition (5) under most reasonable circumstances.

6.1 A time-inhomogeneous example

In general, the intensity process might not be a simple function of the counting process as in the pure birth process discussed above. The issue of explosion is then far more complex. To illustrate this complexity, we develop an example of a time-inhomogeneous process where the feedback is exponential. In particular, we define the intensity process by

$$\lambda(t) = \alpha \exp(\beta N(t-)/t), \quad (13)$$

where α and β are (non-random) parameters. This is still a Cox-like model, but explosion is no longer certain. By using the average number of events over time rather than $N(t-)$ itself, it would be reasonable to expect that explosions are less prevalent. It should be noted that neither the Feller/Jacobsen nor the Lipschitz conditions apply directly to this example. However, the approach using a deterministic approximation will still shed some light on the problem. This model was studied in some detail in Aalen et al. (2008, Sect. 8.6.3). We will here summarize and extend some of the results.

The model defined in (13) has, in spite of the averaging factor, a potential for explosions. This possibility, however, depends on the parameters α and β . In essence, there are two options: if $\alpha\beta > e^{-1}$ an explosion will always occur. If $\alpha\beta < e^{-1}$ the situation is more complex: it seems clear that explosions will occur if $N(t)$ has a large starting value, or increases rapidly very early. If no explosions occur, $N(t)$ will approach a homogeneous Poisson process over time. We will discuss these results in more detail below.

To start with an illustration, Fig. 4 shows 100 simulated paths of the process with intensity defined in (13), with parameter values $\alpha = 1$ and $\beta = 1/2.5$. The simulations are based on a discrete approximation to the counting process, with increments $\{N_j^D, j = 0, \dots, 2500\}$, where N_j^D is Poisson distributed with expectation $\exp(N_{j-1}/(2.5j))$, and the starting value is $N_0 = 0$. Since $\alpha\beta$ is larger than e^{-1} , all paths will eventually explode, although this is not obvious from the picture.

When the parameters are chosen even closer to the borderline value, we see in Fig. 5 that (with $\alpha = 1$ and $\beta = 1/2.7$) there are only two visible explosions (among 100 simulations), and the explosions are those paths that increase very rapidly to begin with. The remaining paths also explode, eventually.

When $\alpha\beta < e^{-1}$ we would expect (see below) that some paths (which quickly obtain a large value) still explode, and that some approach parts of a homogeneous Poisson process. Figure 6, with simulations on the other side of the boundary ($\alpha = 1$ and $\beta = 1/3.1$), shows this, although the convergence of the non-exploding paths is slow. It will be shown below that the limit of $N(t)/t$ is the stationary point 1.77.

Approximating our slightly more complex model (13), the corresponding deterministic differential equation would be

$$df(t) = \alpha \exp(\beta f(t)/t) dt. \quad (14)$$

From the heuristic argument above we would in fact expect the explosive behavior of $N(t)$ to be the same as that of $f(t)$. Define H as in (11) above, with $h(u)$ replaced by $g(u) = \alpha e^{\beta u} - u$, and assume that g has two zeros, $u_1 < u_2$, which happens when $\alpha\beta < e^{-1}$. The implicit equation for $f(t)$, derived from (14), is

$$H\left(\frac{f(t)}{t}\right) = \log(t) \quad (15)$$

(see Aalen et al. (2008) for details). From this it was shown that if $0 < c < u_1$ or $u_1 < c < u_2$, there are no explosions, whereas if $c > u_2$, the solution explodes. Once more, if

$$\tau = \tau(c) := H(\infty) < \infty,$$

the explosion will take place at time $\exp(\tau(c))$. It is also seen that if $c < u_1$, $f(t)/t \rightarrow u_1$ as $t \rightarrow \infty$, and indeed this happens also when $u_1 < c < u_2$. The root u_2 marks the threshold between stability and explosion.

For the stochastic model (13) there will be a lot of initial variation, but it seems reasonable to expect the same asymptotic behavior as (14) with a random initial value. With $\alpha = 1$ and $\beta = 1/3.1$, the function g has the two zeros $u_1 = 1.77$ and $u_2 = 4.97$, and examining Fig. 6, for the non-exploding paths, $N(t)/t$ converges to u_1 , albeit very slowly. The slow convergence should also be evident from the $\log(t)$ term on the right hand side of (15).

6.2 Explosions due to additional random covariate processes

As we have seen, even when $\alpha\beta < e^{-1}$ a rapid initial growth of the process defined by (13) may cause it to exceed u_2 and bring it into the region of explosion. Or similarly, a random initial value will have this effect. By the same token, one would expect that a random covariate process could cause instability even if the process itself starts out at zero. Assume, for instance, that

$$\lambda(t) = \alpha \exp(\beta N(t-) / t + bX(t)), \quad (16)$$

where $X(t)$ is a randomly varying covariate process and b is a regression coefficient. For instance, $X(t)$ could be a Wiener process $\sigma W(t)$, with expectation zero and variance $\sigma^2 t$, or it could be the Ornstein-Uhlenbeck process, defined by the stochastic differential equation

$$dX(t) = l_1(l_0 - X(t))dt + \sigma dW(t), \quad (17)$$

where l_0 and $l_1 > 0$ are parameters. The former exhibits steadily increasing variability over time, whereas the latter is stable around the value l_0 , the so-called mean reverting property (see, for example, Aalen et al. (2008, Appendix A.4.3)). Repeating the above simulations with X defined as the Ornstein-Uhlenbeck process, we see in Fig. 7 how the behavior becomes more unpredictable, although the main pattern is not very different from the model without the covariate process. The covariate process X itself is well defined and behaves in a stationary manner.

However, when introducing a Wiener process, the paths of N become erratic, as seen in Fig. 8. In some cases, the process shows initial signs of approaching explosion, but the variability of the covariate process dampens the effect and eventually defuses it.

7 Honesty for multivariate processes

So far, we have restricted our models to univariate counting processes. We will end this paper with a discussion about the multivariate setting. While univariate counting processes are the most common in standard applications, multivariate processes are relevant for models that keep track of different individuals, several types of outcome at the same time, or for instance combinations of outcomes and dynamic covariates. In the following we will present a summary of results rather than delve too deeply into examples. We use the notation $N(t) = (N_1(t), \dots, N_d(t))$ for an d -dimensional counting process with intensity processes $\lambda(t) = (\lambda_1(t), \dots, \lambda_d(t))$, and we make the common assumption that none of the individual processes have jumps at the same point in time. Let $T_1 < T_2 < \dots$ be the jump times of $N(t)$, i.e., the times where any one of the components of $N(t)$ makes a jump. As before, we say that N explodes if $\lim_{i \rightarrow \infty} T_i < \infty$. It is clear that N explodes if and only if at least one of its components explodes. Note that the intensity of an event in *any* component of the process is $\sum_{i=1}^d \lambda_i(t)$, and that $\sum_{i=1}^d N_i(T_k) = k$.

7.1 The Feller condition for multivariate processes

The Feller criterion (5) has an extension to multivariate Markov chains. Let $N(t)$ be an n -dimensional counting process. Assume, as in (4), that $\lambda(t) = (\lambda_1(t), \dots, \lambda_d(t))$ is a function of $N(t-)$, i.e., $\lambda(t) = h(N(t-)) = (h_1(N(t-)), \dots, h_d(N(t-)))$. Proposition 10.21 of Kallenberg (1997) applies to homogeneous multivariate Markov chains, and states that $N(t)$ does not explode if and only if

$$\sum_{k=1}^{\infty} \frac{1}{\sum_{i=1}^d h_i(N(T_k))} = \infty, \quad (18)$$

where $T_1 < T_2 < \dots$ are the jump times of N . This condition is made somewhat difficult by its dependence on the actual path of $N(t)$. In the simple case where $h_i(N(t-)) = h_i(N_i(t-))$, i.e., h_i is a function of only the i 'th component of N , so that the intensity of each component depends only on its own value, it is clear that non-explosion is the case if and only if

$$\sum_{k=1}^{\infty} \frac{1}{h_i(k)} = \infty$$

for all d . In the more general case, where h_i depends on the full vector $N(t)$ and not only the i 'th component, we can derive a sufficient criterion for non-explosion as

$$\sum_{i=1}^d h_i(n_1, \dots, n_d) \leq K_1 + K_2 \sum_{i=1}^d n_i \quad (19)$$

for all $(n_1, \dots, n_d) \in \mathbb{Z}_+^d$, for some $K_1, K_2 > 0$. That is, we assume the intensity rates are restricted by a linear function of the individual components. If we apply (19) to the condition (18), we get

$$\sum_{k=1}^{\infty} \frac{1}{\sum_{i=1}^d h_i(N(T_k))} \geq \sum_{k=1}^{\infty} \frac{1}{K_1 + K_2 \sum_{i=1}^d N_i(T_k)} = \sum_{k=1}^{\infty} \frac{1}{K_1 + K_2 k} = \infty,$$

and non-explosion follows.

7.2 The Lipschitz condition for multivariate processes

In Sect. 5 we saw how the Lipschitz condition applies to the SDE formulation of the explosion problem. One advantage with the Lipschitz condition (9) is that it extends naturally to multivariate processes. In the example in Sect. 6.2 we showed that an exponential feedback mechanism combined with a noisy covariate process X could cause nearly chaotic explosion patterns. However, if we still define X as an Ornstein-Uhlenbeck process, but assume a linear feedback structure for N , the corresponding SDE would be

$$\begin{aligned} dX(t) &= l_l(l_0 - X(t)) dt + \sigma dW(t), \\ dN(t) &= (\beta N(t-) + bX(t)) dt + dM(t). \end{aligned}$$

We can then define the two-dimensional process $Z(t) = [X(t), N(t)]^T$ and the two-dimensional martingale $M(t) = [\sigma W(t), M(t)]^T$, and the SDE can be written as

$$dZ(t) = \left\{ \begin{bmatrix} l_l l_0 \\ 0 \end{bmatrix} + \begin{bmatrix} -l_l & 0 \\ b & \beta \end{bmatrix} Z(t-) \right\} dt + d\tilde{M}(t).$$

This is again a linear equation in $Z(t)$, and the coefficients thus satisfy a Lipschitz condition. This again guarantees existence and uniqueness of a non-exploding solution. In this illustration we thus see that even when both X and N have feedback, non-explosion is

guaranteed by the linear structure. The Lipschitz condition covers not only the counting process part but also the diffusion part in the covariate.

7.3 Non-explosion and the martingale problem

In this section we approach the problem of non-explosion from yet another angle. We have seen that the Feller and Jacobsen criteria provide easily verifiable conditions for non-explosion, and extend at least partially to multivariate processes. Similarly, the closely related Lipschitz conditions are very natural when the problem is put in an SDE form. All these approaches relate to the explicit construction of a solution process, either as a Markov chain with a specified transition intensity, or as a solution to an SDE.

A somewhat more abstract approach is to view the existence as a *martingale problem*. The intuitive understanding of the intensity process is as an instantaneous transition probability, as defined in (1). However, as mentioned in Sect. 5 we can characterize the intensity of N with respect to the probability measure P more formally within the framework of martingales. It is the unique nonnegative process λ such that, in addition to some regularity conditions, each random variable $\lambda(t)$ only depends on the previous history of N , and such that

$$M(t) := N(t) - \int_0^t \lambda(s) ds \quad (20)$$

is a (local) martingale with respect to P and the history generated by N . The problem of honesty of counting process models can be formulated as follows within the martingale framework: Given a counting process N such that $N(t)$ is finite for every $t \geq 0$ and a suitable nonnegative process λ , does there exist a probability measure P on the underlying space such that the process (20) is a (local) martingale with respect to P ? The problem of finding a probability measure that makes a given stochastic process into a martingale is called a martingale problem. Note that the Lipschitz conditions given in Sect. 5 and 7.2 guarantee the existence of a *strong* solution to the corresponding SDE, whereas solutions to the martingale problem will typically correspond to *weak* solutions, i.e., where the probability space and martingale are not given before-hand. In general, the martingale problem is a deep mathematical topic that has been central to probability theory, see Stroock and Varadhan (2006) and Jacod and Shiryaev (1987). We will explore solutions to the martingale problem in the following section and see what conditions for non-explosion they may provide.

We start with a d -variate Poisson process model. This means a d -dimensional counting process N , such that the components $N_1(t), \dots, N_d(t)$ are finite for every t , and a probability measure Q on the underlying space. The components $N_1, \dots, N_d$ are independent with respect to Q and have intensity 1 with respect to the previous history generated by all the counting processes. Moreover, we consider positive processes $\lambda_1, \dots, \lambda_d$ with some additional regularity and such that every random variable $\lambda_i(t)$ only depends on the history of the process N that is observable before t .

The question is whether there exists a probability measure P such that the processes

$$N_1(t) - \int_0^t \lambda_1(s) ds, \dots, N_d(t) - \int_0^t \lambda_d(s) ds \quad (21)$$

are (local) martingales with respect to P . In other words, such that $\lambda_1, \dots, \lambda_d$ define the intensities of $N_1, \dots, N_d$ with respect to P . Since the random variables $N_1(t), \dots, N_d(t)$ are finite for every $t \geq 0$, we would then have an honest model.

We now consider the process

$$R(t) := \prod_{i=1}^d \exp\left(-\int_0^t (\lambda_i(s) - 1) ds\right) \prod_{s \leq t} (\lambda_i(s))^{\Delta N_i(s)}. \quad (22)$$

This is an example of a Doléans–Dade exponential, see Protter (1990, Theorem II 37). It is a consequence of Girsanov’s theorem, Protter (1990, Theorem II 40), that if T is a positive number then

$$E_Q[R(T)] = 1, \quad (23)$$

if and only if there exists a probability measure P such that on the interval $[0, T]$,

1. The random variable $R(T)$ is the likelihood ratio of P and Q (Radon–Nikodym derivative),
2. The processes (21) are (local) martingales with respect to P , i.e., $\lambda_1, \dots, \lambda_d$ are the intensities of $N_1, \dots, N_d$ with respect to P .

Since the random variables $N_1(T), \dots, N_d(T)$ are finite, the condition (23) would now guarantee an honest multivariate counting process model with the given intensities on the interval $[0, T]$. Note that $R(t)$ is the “change of measure” from the standard Poisson measure Q that makes N a standard Poisson process with intensity 1, into the measure P that makes λ the intensity process of N . It is closely related to partial likelihood and estimation in event history models, see Andersen et al. (1993), and van der Vaart (1998).

Remark 1 As an aside, let us briefly consider the space of possibly infinite counting processes on the interval $[0, T]$. One could, without condition (23), consider a probability measure $\tilde{P}$ allowing proper extensions of the processes $\tilde{\lambda}_1, \dots, \tilde{\lambda}_d$ as intensities of the corresponding counting processes. The existence of $\tilde{P}$ would only require regularity conditions; however, such a model could be explosive. The likelihood ratio of $\tilde{P}$ and Q on the space of finite counting processes would equal $R(T)$. This suggests that $E_Q[R(T)]$ is the probability that an outcome in the possibly dishonest model will not explode on the interval $[0, T]$.

The condition (23) is difficult to verify directly. However, note that Q is a Poisson measure, so explicit computations may be within reach in special cases. It turns out that (23) is satisfied if and only if the process R is a martingale on the interval $[0, T]$. The problem of deciding when a Doléans–Dade exponential is a martingale has been studied in several places in the literature. A condition for this to be true was given in Lépingle and Mémin

(1978). A very common approach, however, is through the Novikov condition, Protter and Shimbo (2008, Theorem 6). In our case, the Novikov condition translates into

$$E_Q \left[\exp \left(\sum_{i=1}^d \int_0^T (\lambda_i(s) - 1)^2 ds \right) \right] < \infty. \quad (24)$$

Condition (24) implies condition (23) and so the corresponding model is honest.

While the Novikov condition is often useful in theory, it still requires computing an expected value that may not be easily obtainable from a practical problem formulation. A simpler and more explicit condition is given in Brémaud (1981) VIII Theorem T11. It is shown that if there exist nonnegative K_1 and K_2 and an $\alpha > 1$ such that:

$$\sum_{i=1}^d (\lambda_i(t))^\alpha \leq K_1 + K_2 \sum_{i=1}^d N_i(t) \quad (25)$$

for every t in the interval $[0, T]$, then the Doléans-Dade exponential $R(t)$ is a martingale. Again, this means that (23) is satisfied and we would have an honest multivariate counting process model with the given intensities on the interval $[0, T]$. Note that for the homogeneous Markov formulation used in Sect. 7.1, the condition (25) requires more than (19), since α must be greater than one.

8 Discussion

As we have shown in this paper, the issue of the actual existence of counting process models is a complex one. However, we have brought forth a number of suitable conditions for existence. Clearly, also in a statistical setting one has to pay attention to whether the models one proposes actually exist as mathematical objects. The warnings issued in the present paper are especially relevant for the Cox model since the covariates come into the hazard rate through exponential functions. Aalen's additive model is in this respect much less vulnerable, and is more closely connected to standard Markov processes, such as linear birth processes.

From a statistical point of view it is important to use models that describe how the past may influence the present. Even when the process is mathematically well defined, however, there is a further caveat: one might be tempted to interpret such models as saying something about causality. As always, one has to show caution when causality is involved, and there are at least two issues to consider. Firstly, association between earlier and later events could easily be due to confounding. For instance, if an individual with many previous events has an increased likelihood of future events, then this could be because there is some property of this individual that increases the general likelihood of events occurring. This is discussed as a "frailty effect" in Sect. 8.6.2 of Aalen et al. (2008). Even in the absence of such confounders one has to be careful in sorting out the effects of various dynamic covariates, because they are usually related among themselves. A useful tool might be dynamic path analysis as described in Sect. 8.4 of Aalen et al. (2008), see also Fosen et al. (2006).

The results of this paper raise the issue of whether in practice, when real recurrent event data are available, one could fit recurrent event models with dynamic components or exponential link for the covariate processes, such as the Andersen-Gill model or similar models, e.g., those in Peña et al. (2007). This is an important question requiring further examination. The real data may certainly be used to estimate the model parameters, but then the resulting parameter estimates may reside in the region of the parameter space where explosion occurs with positive probability. How then should the resulting fitted model be interpreted? Is such a fitted model still going to be useful in practice? How should the estimated model parameter values be interpreted? It is possible that an alteration of the model, such as imposing e.g., a Lipschitz condition on the link function for the dynamic component, or possibly by incorporating a component in the “at-risk process”, would guarantee non-explosion. However, these model alterations require more detailed studies in practical applications since they also demand changes in the statistical inference methods and the interpretations of results.

We close by observing that the theory of counting processes has been of considerable use in survival and event history modeling. A general formulation has been given by the Andersen-Gill model. It is now time to move the Andersen-Gill model into a more realistic setting and start looking at which general counting process models are mathematically feasible, in particular when a dynamic covariate structure is brought fully into the modeling perspective.

Acknowledgments

We thank an anonymous referee for useful suggestions.

Appendix

In Sect. 5 we showed how the counting process problem could be formulated as a stochastic differential equation, so that the relevant Lipschitz conditions could be applied. We will here give a more formal argument for Eq. 8.

Let $N(t, dx)$ be a Poisson random measure with intensity measure ν , such that

$N(t, B) = \int_B N(t, dx)$ is a Poisson process with intensity $\nu(B)$ and compensator $t\nu(B)$. To be specific, we assume $N(t, dx)$ is generated as the counting measure of a stable Lévy subordinator, so that $\nu(dx) = \beta x^{1-\beta} dx$ is the corresponding Lévy measure, with $\beta \in (0, 1)$. Let B be the interval $[y^{-1/\beta}, \infty)$ for $y > 0$ and note that

$$\nu(B) = \int_{y^{-1/\beta}}^{\infty} \nu(dx) = y.$$

Thus, $N^y(t) := N(t, B)$ is a Poisson process with intensity y , and we can write

$$N^y(t) = \int_0^{\infty} I(y, x) N(t, dx)$$

where $I(y, x) := 1\{x \leq y^{-1/\beta}\}$ is an indicator function. The compensator of $N^y(t)$ is $t\nu(B) = ty$. For the increments of N^y we can write

$$\begin{aligned} dN^y(t) &= \int_0^\infty I(y, x) N(dt, dx) \\ &= \int_0^\infty I(y, x) \nu(x) dx dt + \int_0^\infty I(y, x) [N(dt, dx) - \nu(x) dx dt] \\ &= y dt + \int_0^\infty I(y, x) [N(dt, dx) - \nu(x) dx dt]. \end{aligned} \quad (26)$$

Recall the counting process $N(t)$ in Sect. 5, having intensity process $\lambda(t) = h(N(t-), t)$. By substituting $\lambda(t)$ for y in (26) we see that, conditional on $\mathcal{F}_{t-}$, the increments $dN^y(t)$ and $dN(t)$ have the same distribution. Thus, $N(t)$ is a solution the stochastic differential equation

$$dN(t) = h(N(t-), t) dt + \int_0^\infty F(N(t-), t, x) [N(dt, dx) - \nu(x) dx dt], \quad (27)$$

where

$$F(n, t, x) := I(h(n, t), x).$$

We can now apply Theorem 6.2.9 of Applebaum (2004), which states that (27) has a strong solution if h and F satisfy the necessary Lipschitz (and possibly growth) conditions. For F the required Lipschitz condition is

$$\int_0^\infty |F(n_2, t, x) - F(n_1, t, x)|^2 \nu(dx) \leq K_1(t) |n_2 - n_1|.$$

Since (assuming, say, that $h(n_2, t) \geq h(n_1, t)$),

$$\begin{aligned} F(n_2, t, x) - F(n_1, t, x) &= I(h(n_2, t), x) - I(h(n_1, t), x) \\ &= 1\{h(n_2, t) \leq x \leq h(n_1, t)\} \end{aligned}$$

we have

$$\int_0^\infty |F(n_2, t, x) - F(n_1, t, x)|^2 \nu(dx) = |h(n_2, t) - h(n_1, t)|,$$

which thus simply turns out as a Lipschitz condition on h .

We have thus shown that the stochastic differential Eq. 8 can be justified by generating the martingale noise part explicitly using a Poisson process associated with the stable Lévy process. Since the driving Poisson random measure $N(dt, dx)$ is given in advance, the solution process $N(t)$ is a strong solution. More details can be found in Applebaum (2004). Protter (1990) approaches the problem using more general semimartingales to generate the noise in the stochastic differential equation.

References

- Aalen, OO.; Borgan, Ø.; Gjessing, HK. Survival and event history analysis: a process point of view. Springer; New York: 2008.
- Andersen, PK.; Borgan, Ø.; Gill, RD.; Keiding, N. Statistical models based on counting processes. Springer-Verlag; New York: 1993.
- Applebaum, D. Lévy processes and stochastic calculus. Cambridge University Press; Cambridge: 2004.
- Birkhoff, G.; Rota, G-C. Ordinary differential equations. Wiley; New York: 1989.
- Brémaud, P. Point processes and queues. Springer-Verlag; New York: 1981.
- Cook, RJ.; Lawless, JF. The statistical analysis of recurrent events. Springer; New York: 2007.
- Cox DR. Regression models and life-tables (with discussion). J R Stat Soc: Ser B (Stat Methodol). 1972; 34:187–220.
- Cox, DR.; Miller, HD. The theory of stochastic processes. Methuen; London: 1965.
- Feller, W. An introduction to probability theory and its applications. 3rd edn. Vol. 1. Wiley; New York: 1968.
- Fosen J, Borgan Ø, Weedon-Fekjær H, Aalen OO. Dynamic analysis of recurrent event data using the additive hazard model. Biomet J. 2006; 48:381–398.
- Gandy A, Jensen U. A nonparametric approach to software reliability. Appl Stoch Model Bus Ind. 2004; 20:3–15.
- Han J, Slate E, Peña E. Parametric latent class joint model for longitudinal biomarkers and recurrent events. Stat Med. 2007; 26:5285–5302. [PubMed: 17542002]
- Jacobsen, M. Statistical analysis of counting processes. Lecture Notes in Statistics. Vol. 12. Springer-Verlag; New York, Heidelberg, Berlin: 1982.
- Jacod, J.; Shiryaev, AN. Limit theorems for stochastic processes. Springer-Verlag; Berlin: 1987.
- Kalbfleisch, JD.; Prentice, RL. The statistical analysis of failure time data. 2nd edn.. Wiley; Hoboken: 2002.
- Kallenberg, O. Foundations of modern probability. Springer-Verlag; Berlin: 1997.
- Lépine D, Mémis J. Sur l'intégrabilité uniforme des martingales exponentielles. Z. Wahrsch. Verw. Gebiete. 1978; 42(3):175–203.
- Miloslavsky M, Keles S, Laan MJ, van der Butler S. Recurrent events analysis in the presence of time-dependent covariates and dependent censoring. J R Stat Soc: Ser B. 2004; 66:239–257.
- Peña EA. Dynamic modeling and statistical analysis of event times. Stat Sci. 2006; 21:487–500.
- Peña EA, Slate EH, González JR. Semiparametric inference for a general class of models for recurrent events. J Stat Plan Inference. 2007; 137:1727–1747. [PubMed: 19823592]
- Protter, P. Springer-Verlag; Berlin: 1990. Stochastic integration and differential equations: a new approach.
- Protter P, Shimbo K. No arbitrage and general semimartingales. In: Markov processes and related topics a Festschrift for Thomas G. Kurtz. IMS Collections. 2008; 4
- Stocker R, Peña E. A general class of parametric models for recurrent event data. Technometrics. 2007; 49:210–220.
- Stroock, D.; Varadhan, SRS. Multidimensional diffusion processes, classics in mathematics. Springer-Verlag; Berlin: 2006.
- Tsiatis A, Davidian M. Joint modelling of longitudinal and time-to-event data: an overview. Stat Sinica. 2004; 14:793–818.
- van der Vaart, AW. Asymptotic Statistics. Cambridge University Press; Cambridge: 1998.

Fig. 1.

Local dependence graph showing the influence of covariate processes on a counting process N . The covariate process X_1 is an external covariate process

Fig. 2.
Extended local dependence graphs for dynamic covariates and for time-dependent internal covariates

Fig. 3.
 $N(t)$, where $N(t)$ is a Cox counting process defined by (3), with $\alpha = 1/1000$, $\beta = 1/2$, $b = 1/2$, with a time-dependent covariate process $X(t)$ defined as in (17), with $l_0 = 0$, $l_1 = 0.01$ and $\sigma = 0$.

Fig. 4.
 $N(t)/t$, where $N(t)$ is the Cox-type counting process defined by (13), with $\alpha = 1$ and $\beta = 1/2.5$

Fig. 5.
 $N(t)/t$, where $N(t)$ is the Cox-type counting process defined by (13), with $\alpha = 1$ and $\beta = 1/2.7$

Fig. 6.
 $N(t)/t$, where $N(t)$ is the Cox-type counting process defined by (13), with $\alpha = 1$ and $\beta = 1/3.1$.
 Initial values are chosen as Poisson distributed with mean equal to 5

Fig. 7.

Left panel $N(t)/t$, where $N(t)$ is the Cox-type counting process defined by (16), with $\alpha = 1$, $\beta = 1/3$, $b = 1/2$. *Right panel* simulations of the corresponding covariate process $X(t)$ defined as in (17), with $l_0 = 0$, $l_1 = 0.01$ and $\sigma = 0.1$

Fig. 8.

Left panel $N(t)/t$, where $N(t)$ is the Cox-type counting process defined by (16), with $\alpha = 1$, $\beta = 1/3$, $b = 1/2$. *Right panel* simulations of the corresponding covariate process $X(t)$ defined as a Wiener process with drift zero and $\sigma = 0.1$. Note that the individual covariate process paths are drawn only up until the point T_∞ where $N(t)$ explodes