Chapter 2
Stochastic processes in event history analysis

Event histories unfold in time. Therefore, one would expect that tools from the theory of stochastic processes would be of considerable use in event history analysis. This is indeed the case, and in the present chapter we will review some basic concepts and results for stochastic processes that will be used in later chapters of the book.

Event histories consist of discrete events occurring over time in a number of individuals. One can think of events as being counted as they happen. Therefore, as indicated in Section 1.4, counting processes constitute a natural framework for analyzing survival and event history data. We shall in this chapter develop this idea further, and in particular elaborate the fundamental martingale concept that makes counting processes such an elegant tool. In this book the focus is on models in continuous time. However, as some concepts and results for martingales and other stochastic processes are more easily understood in discrete time, we first, in Section 2.1, consider the time-discrete case. Then, in Section 2.2, we discuss how the concepts and results carry over to continuous time. To keep the presentation fairly simple, we restrict attention to univariate counting processes and martingales in this chapter. Extensions to the multivariate case are summarized in Appendix B.

With processes unfolding over time, one will also naturally come across processes that do not consist of discrete jumps of unit size, like counting processes do, where each jump corresponds to the occurrence of an event. For instance, one may imagine that an event is really just a manifestation of some underlying process that could be continuous; for example, a heart attack may occur when a blood clot grows beyond a certain size. In general there may be the idea of some continuous underlying process crossing a threshold and producing an event (cf. Chapter 10). This way of thinking is very natural and useful, and we shall apply some continuous stochastic processes. The most basic continuous stochastic processes is the Wiener process (or Brownian motion), which has independent increments that are normally distributed with mean zero and variance proportional to the length of the time interval. A review of the basic properties of the Wiener process is provided in Section 2.3.1, while a more extensive review and discussion of the more general diffusion processes is provided in Appendix A.4.

O.O. Aalen et al., *Survival and Event History Analysis: A Process Point of View*, Statistics for Biology and Health, DOI 10.1007/978-0-387-68560-1_2,

Transformations of Wiener processes arise as limits of martingales associated with counting processes when the number of individuals increases. Such approximations, which are given by the martingale central limit theorem, play an essential role in the statistical analysis of models based on counting processes. In Sections 2.3.2 and 2.3.3 conditions under which the martingale central limit theorem hold are discussed and formally stated.

The idea of independent increments is fundamental in stochastic process theory. From statistical inference one is well acquainted with the independent identically distributed random variables that form the basis of many statistical models. In stochastic process theory we have the Poisson and Wiener processes with their stationary and independent increments (cf. Sections 2.2.4 and 2.3.1). More general processes with stationary and independent increments also exist and are denoted Lévy processes. These have nice and important properties that are particularly useful in the theory of frailty and generalized frailty (cf. Chapters 6 and 11). An introduction to Lévy processes is given in Appendix A.5.

When processes do not have independent increments, the fundamental view taken in this book is that of the "French school" of stochastic processes. One then seeks to explain the future development of a process by means of what has happened previously. This is a dynamic point of view, connecting the past, present and future. This differs fundamentally from the theory of stationary processes that plays such a large role in time series analysis. The connection between the past and the future is given by means of *local characteristics* that describe how the past influences the changes taking place. The intensity process of a counting process is an example of a local characteristic. The mathematical foundation underlying this theory is given by a theorem called the *Doob-Meyer decomposition*. Although this is in its general form a quite heavy mathematical result, the intuitive content is simple and not hard to grasp (cf. Section 2.2.3).

The stochastic process concepts and results we present in this chapter and in Appendices A and B are based on quite heavy mathematics if one wants to go into every detail, and a number of regularity assumptions must be made to be mathematically precise. We shall not state these assumptions, but refer to Andersen et al. (1993) for the theory of counting processes with associated martingales and stochastic integrals and to the references provided in Appendix A for the theory of Wiener processes, diffusions, and Lévy processes. [For example, we will not state integrability conditions in this book, and thus we will not worry about whether a stochastic process is a (local) martingale or a (local) square integrable martingale.] It should be pointed out, however, that the basic ideas and results for stochastic processes are mostly relatively simple, and that they can be understood at an intuitive "working technical" level without going into mathematical details. This is the level of presentation we aim at in this book. Very often, the intuitive content in stochastic processes tends to drown in complex mathematical presentations, which is probably the reason this material is not so much used in applied statistics. We want to contribute to a "demystification" of martingales, stochastic integrals, and other stochastic process concepts.

2.1 Stochastic processes in discrete time

Although in this book we operate in continuous time, where events can occur at any time, it will be useful in this section to consider time-discrete processes. Mathematically, such processes are much simpler than the time-continuous ones, which may often be derived as limits of the time-discrete processes. So if we understand the basic ideas in a discrete context, it will also give us the required insight in the continuous setting. The results for time-discrete processes are also of interest in their own right, for example, for studying longitudinal data in discrete time (Borgan et al., 2007; Diggle et al., 2007).

2.1.1 Martingales in discrete time

Let $M = \{M_0, M_1, M_2, \ldots\}$ be a stochastic process in discrete time. The process M is a *martingale* if

$$\mathrm{E}(M_n \mid M_0, M_1, \ldots, M_{n-1}) = M_{n-1} \tag{2.1}$$

for each $n \geq 1$. Hence, the martingale property simply consists in asserting that the conditional expectation of a random variable in the process given the past equals the previous value. This innocent-looking assumption has a far greater depth than would appear at first look. The martingale property may be seen as a requirement to a fair game. If M_{n-1} is the collected gain after $n-1$ games, then the expected gain after the next game should not change.

For the applications we have in mind, it will always be the case that $M_0 = 0$. As some of the following formulas become simpler when this is the case, we will tacitly assume that $M_0 = 0$ throughout.

In Chapter 1 we talked about the past in a rather unspecified fashion. The past could be just what is generated by the observed process as typified in formula (2.1). However, the past could also be defined as a *wider* amount of information. Often the past includes some external information (e.g., covariates) in addition to the previous values of the process itself. It is important to note that a process may be a martingale for some definitions of the past and not for others. In stochastic process theory the past is usually formulated as a σ-algebra of events. We will not give a formal definition of this σ-algebra. For our purpose it suffices to see it as the family of events that can be decided to have happened or not happened by observing the past. Such a σ-algebra is often termed $\mathscr{F}_n$ and is a formal way of representing what is known at time n. We will denote $\mathscr{F}_n$ as the *history at time n*, so that the entire history is represented by the increasing family of σ-algebras $\{\mathscr{F}_n\}$. The family has to increase, since our past knowledge will increase as time passes.

We shall now give a more general formulation of (2.1). Assume that $\mathscr{F}_n$, for each n, is generated by $M_1, \ldots, M_n$ plus possibly some external information. A technical formulation would be that the process $M = \{M_0, M_1, M_2, \ldots\}$ is *adapted* to the history $\{\mathscr{F}_n\}$. This means that, for each n, the random variables $M_1, \ldots, M_n$

are measurable with respect to the σ-algebra $\mathscr{F}_n$. A practical implication is that $M_1, \ldots, M_n$ may be considered as constants given the history $\mathscr{F}_n$; in particular:

$$\mathrm{E}(M_m \mid \mathscr{F}_n) = M_m \qquad \text{for all } m \leq n. \tag{2.2}$$

The process $M = \{M_0, M_1, M_2, \ldots\}$ is a martingale with respect to the history $\{\mathscr{F}_n\}$ if

$$\mathrm{E}(M_n \mid \mathscr{F}_{n-1}) = M_{n-1} \qquad \text{for all } n \geq 1, \tag{2.3}$$

cf. (2.1). This is equivalent to the more general statement

$$\mathrm{E}(M_n \mid \mathscr{F}_m) = M_m \qquad \text{for all } n > m \tag{2.4}$$

(Exercise 2.1). Note the difference between (2.2) and (2.4). While (2.2) states that we know the past and present of the process M, (2.4) states that the expected value of the process in the future equals its present value.

As a consequence of (2.4) we get, using double expectations and the assumption $M_0 = 0$,

$$\mathrm{E}(M_n) = \mathrm{E}\{\mathrm{E}(M_n \mid \mathscr{F}_0)\} = \mathrm{E}(M_0) = 0. \tag{2.5}$$

Since the martingale has mean zero for all n, we say that it is a *mean zero martingale*. By a similar argument one may show that

$$\mathrm{Cov}(M_m, M_n - M_m) = 0 \qquad \text{for all } n > m$$

(Exercise 2.2), that is, the martingale has *uncorrelated increments*.

By (2.2), a reformulation of (2.3) is as follows:

$$\mathrm{E}(M_n - M_{n-1} \mid \mathscr{F}_{n-1}) = 0 \qquad \text{for all } n \geq 1. \tag{2.6}$$

Here $\triangle M_n = M_n - M_{n-1}$, $n = 1, 2, \ldots$, are denoted martingale differences. Notice that (2.6) would also hold if the process had independent zero-mean increments. In this sense the concept of martingale differences is a weakening of the independent increment concept. We could also say that any sum of independent zero-mean random variables is a martingale (Exercise 2.3). The assumption of independence pervades statistics, but in many cases a martingale-type assumption would be sufficient for demonstrating unbiasedness, asymptotic normality, and so on.

2.1.2 Variation processes

Two processes describe the variation of a martingale $M = \{M_0, M_1, \ldots\}$. The *predictable variation process* is denoted $\langle M \rangle$ and for $n \geq 1$ is defined as the sum of conditional variances of the martingale differences:

$$\langle M \rangle_n = \sum_{i=1}^{n} \mathrm{E}\{(M_i - M_{i-1})^2 \mid \mathscr{F}_{i-1}\} = \sum_{i=1}^{n} \mathrm{Var}(\triangle M_i \mid \mathscr{F}_{i-1}), \tag{2.7}$$

while $\langle M \rangle_0 = 0$. The *optional variation process* $[M]$ is defined by

$$[M]_n = \sum_{i=1}^{n} (M_i - M_{i-1})^2 = \sum_{i=1}^{n} (\triangle M_i)^2 \tag{2.8}$$

for $n \geq 1$, and $[M]_0 = 0$. The following statements can be proved by simple calculations:

$$M^2 - \langle M \rangle \text{ is a mean zero martingale,} \tag{2.9}$$

$$M^2 - [M] \text{ is a mean zero martingale.} \tag{2.10}$$

We shall prove the second of these statements and leave (2.9) as an exercise for the reader (Exercise 2.4). To prove (2.10), we first note that $M_0^2 - [M]_0 = 0$. Then writing

$$M_n^2 = (M_{n-1} + M_n - M_{n-1})^2, \quad \text{and} \quad [M]_n = [M]_{n-1} + (M_n - M_{n-1})^2,$$

we get

$$\begin{aligned}
&\mathrm{E}(M_n^2 - [M]_n \mid \mathscr{F}_{n-1}) \\
&= \mathrm{E}(M_{n-1}^2 + 2M_{n-1}(M_n - M_{n-1}) - [M]_{n-1} \mid \mathscr{F}_{n-1}) \\
&= M_{n-1}^2 - [M]_{n-1} + 2M_{n-1}\mathrm{E}(M_n - M_{n-1} \mid \mathscr{F}_{n-1}) \\
&= M_{n-1}^2 - [M]_{n-1},
\end{aligned}$$

which gives exactly the martingale property.

2.1.3 Stopping times and transformations

One major advantage of the martingale assumption is that one can make certain manipulations of the process without destroying the martingale property. The independence property, on the other hand, would not survive these manipulations.

Our first example of this is the concept of an *optional stopping time* (or just stopping time). An example could be the first time M passes above a certain limit. We denote such a time by T and the value of the process by M_T. In general, a time T is called an optional stopping time if the event $\{T = t\}$ is only dependent on what has been observed up to and including time t. If T is defined as a first passage time, then it is an optional stopping time. It is easy to construct times that are not optional stopping times; for instance, let T be equal to the last time the process passes above a certain limit. The optional stopping time property can be decided from the past and present observations, while in the last example we would have to look ahead.

Stopping a fair game at an optional stopping time T preserves the fairness of the game. This is connected to a preservation of the martingale property under optional stopping. For a martingale M, define *M stopped at T* as follows:

$$M_n^T = M_{n \wedge T}. \tag{2.11}$$

(Here $n \wedge T$ denotes the minimum of n and T.) It can be proved that M^T is a martingale (Exercise 2.6). The idea of *fairness of a game* is strongly connected to the idea of *unbiasedness* in statistics, and this is a major reason for the statistical usefulness of martingales. Stopping will commonly occur in event history data in the form of censoring, and preserving the martingale property ensures that estimates and tests remain essentially unbiased.

A more general formulation of (2.11), which is of great use in our context, comes through defining a *transformation* of a process as follows. Let $X = \{X_0, X_1, X_2, \ldots\}$ be some general process with a history $\{\mathscr{F}_n\}$, and let $H = \{H_0, H_1, H_2, \ldots\}$ be a *predictable* process, that is, a sequence of random variables where each H_n is measurable with respect to $\mathscr{F}_{n-1}$ and hence is known one step ahead of time. The process Z defined by

$$Z_n = H_0 X_0 + H_1(X_1 - X_0) + \cdots + H_n(X_n - X_{n-1}) \tag{2.12}$$

is denoted the *transformation* of X by H and written $Z = H \bullet X$.

Using (2.6) and the predictability of H, the following simple calculation shows that if M is a martingale, then so is $Z = H \bullet M$:

$$\begin{aligned} \mathrm{E}(Z_n - Z_{n-1} \mid \mathscr{F}_{n-1}) &= \mathrm{E}(H_n(M_n - M_{n-1}) \mid \mathscr{F}_{n-1}) \\ &= H_n \mathrm{E}(M_n - M_{n-1} \mid \mathscr{F}_{n-1}) \\ &= 0. \end{aligned}$$

Hence *a transformation preserves the martingale property*. Moreover, since $Z_0 = H_0 M_0 = 0$, the transformation is a mean zero martingale. Considering games, we see again that the fairness of the game is preserved; one often says that there is no betting system that can beat a fair game.

Note that for $n \geq 1$ we may write

$$Z_n = (H \bullet M)_n = \sum_{s=1}^{n} H_s \triangle M_s, \tag{2.13}$$

where $\triangle M_s = M_s - M_{s-1}$. In Section 2.2.2 we shall meet transformations under the more sophisticated disguise of stochastic integrals, and the formulation there will be seen to be very similar to that of (2.13). Most of the properties we will need for stochastic integrals are easily derived for transformations.

The variation processes obey the following rules under transformation:

$$\langle H \bullet M \rangle = H^2 \bullet \langle M \rangle \quad \text{and} \quad [H \bullet M] = H^2 \bullet [M],$$

or formulated as sums:

$$\langle H \bullet M \rangle_n = \sum_{s=1}^{n} H_s^2 \triangle \langle M \rangle_s, \tag{2.14}$$

$$[H \bullet M]_n = \sum_{s=1}^{n} H_s^2 \triangle [M]_s. \tag{2.15}$$

We will prove the first of these statements. By (2.13) and (2.7) we have:

$$\triangle (H \bullet M)_s = H_s \triangle M_s,$$
$$\triangle \langle M \rangle_s = \mathrm{Var}(\triangle M_s \mid \mathscr{F}_{s-1}).$$

Then using (2.7) and the predictability of H, the predictable variation process of the transformation $H \bullet M$ becomes

$$\begin{aligned}\langle H \bullet M \rangle_n &= \sum_{s=1}^{n} \mathrm{Var}(H_s \triangle M_s \mid \mathscr{F}_{s-1}) \\ &= \sum_{s=1}^{n} H_s^2 \, \mathrm{Var}(\triangle M_s \mid \mathscr{F}_{s-1}) \\ &= \sum_{s=1}^{n} H_s^2 \triangle \langle M \rangle_s.\end{aligned}$$

This proves (2.14). The proof of (2.15) is similar and is left as an exercise (Exercise 2.7).

2.1.4 The Doob decomposition

Martingales arise naturally whenever we try to explain the developments in a stochastic process as a function of its previous development and other observations of the past. It is possible to decompose an arbitrary stochastic process into a sequence of martingale differences and a predictable process. Let $X = \{X_0, X_1, X_2, \ldots\}$ be some general process, with $X_0 = 0$, with respect to a history $\{\mathscr{F}_n\}$, and define a process $M = \{M_0, M_1, M_2, \ldots\}$ by

$$M_0 = X_0, \quad M_n - M_{n-1} = X_n - E(X_n \mid \mathscr{F}_{n-1}).$$

It is immediately clear that the $\triangle M_n = M_n - M_{n-1}$ are martingale differences, since the expectation given the past $\mathscr{F}_{n-1}$ is zero. We can therefore write

$$X_n = E(X_n \mid \mathscr{F}_{n-1}) + \triangle M_n; \tag{2.16}$$

this is the Doob decomposition. The quantity $E(X_n \mid \mathscr{F}_{n-1})$ is a function of the past only, and hence the process taking these values is predictable. The martingale differences $\triangle M_n$ are often termed *innovations*, since they represent what is new and unexpected compared to past experience. Hence, formula (2.16) decomposes a process into what can be predicted from the past and what is new and "surprising," the innovations.

A time-continuous generalization of the Doob decomposition, called the Doob-Meyer decomposition, is in fact the key to the counting process approach in this book; cf. Section 2.2.3.

2.2 Processes in continuous time

We will now discuss stochastic processes in continuous time. We first consider time-continuous martingales and stochastic integrals and indicate how the results of the previous section carry over to the time-continuous case. We also discuss the Doob-Meyer decomposition for time-continuous stochastic processes, generalizing the Doob decomposition of Section 2.1.4. Then we briefly discuss the well-known Poisson process and show how this in a natural way gives rise to a time-continuous martingale. Finally, we consider counting processes and martingales derived from counting processes, and we review a number of results that will be of great use in later chapters of the book. To keep the presentation fairly simple, we do not here consider vector-valued counting processes, martingales, and stochastic integrals. The relevant results for such processes, which are multivariate extensions of the results presented in this chapter, are collected in Appendix B.

In practical applications, stochastic processes will be observed over a finite time interval. Unless otherwise stated, we will assume throughout the book that the time-continuous stochastic processes we consider are defined on the finite interval $[0,\tau]$.

Formally, we say that a stochastic process $X = \{X(t);\, t \in [0,\tau]\}$ is *adapted* to a history $\{\mathscr{F}_t\}$ (an increasing family of σ-algebras) if $X(t)$ is $\mathscr{F}_t$-measurable for each t. This means that at time t we know the value of $X(s)$ for all $s \leq t$ (possibly apart from unknown parameters). A realization of X is a function of t and is called a *sample path*. If the sample paths of a stochastic process are right-continuous and have left-hand limits, we say that the process is *cadlag* (continue à droite, limité à gauche). Unless otherwise stated, all time-continuous processes we encounter in the book are assumed to be cadlag.

Also in continuous time will we need the concept of a *stopping time*. We say that T is a stopping time if the event $\{T \leq t\}$ is $\mathscr{F}_t$-measurable for each t. This means that at time t we know whether $T \leq t$ or $T > t$.

2.2.1 *Martingales in continuous time*

A stochastic process $M = \{M(t); t \in [0,\tau]\}$ is a martingale relative to the history $\{\mathscr{F}_t\}$ if it is *adapted* to the history and satisfies the *martingale property:*

$$\mathrm{E}(M(t) \mid \mathscr{F}_s) = M(s) \quad \text{for all } t > s. \tag{2.17}$$

Note that the martingale property (2.17) corresponds to (2.4) for the time-discrete case. A heuristic way of formulating the martingale property, corresponding to (2.6), is to say that M is a martingale provided that

$$\mathrm{E}(dM(t) \mid \mathscr{F}_{t-}) = 0.$$

Here $dM(t)$ is the increment of M over the small time interval $[t, t+dt)$, and $\mathscr{F}_{t-}$ means the history until *just before* time t.

As for the time-discrete case, we will tacitly assume throughout that $M(0) = 0$. This will cover all our applications. Then, by the argument used in (2.5), we have $\mathrm{E}M(t) = 0$ for all t, that is, M is a *mean zero martingale*. Similarily we may show that a martingale has *uncorrelated increments*, that is,

$$\mathrm{Cov}(M(t) - M(s), M(v) - M(u)) = 0 \tag{2.18}$$

for all $0 \le s < t < u < v \le \tau$ (cf. Exercise 2.2).

We also for a time-continuous martingale M introduce the *predictable variation process* $\langle M \rangle$ and the *optional variation process* $[M]$. These are defined as the appropriate limits (in probability) of their time-discrete counterparts:

$$\langle M \rangle (t) = \lim_{n \to \infty} \sum_{k=1}^{n} \mathrm{Var}(\triangle M_k | \mathscr{F}_{(k-1)t/n}), \tag{2.19}$$

and

$$[M](t) = \lim_{n \to \infty} \sum_{k=1}^{n} (\triangle M_k)^2, \tag{2.20}$$

where the time interval $[0, t]$ is partitioned into n subintervals each of length t/n, and $\triangle M_k = M(kt/n) - M((k-1)t/n)$ is the increment of the martingale over the kth of these subintervals. Informally, we have from (2.19) that

$$d\langle M \rangle(t) = \mathrm{Var}(dM(t) \mid \mathscr{F}_{t-}), \tag{2.21}$$

that is, the increment $d\langle M \rangle(t)$ of the predictable variation process over the small time interval $[t, t+dt)$ is the conditional variance of the increment of the martingale.

In a similar manner as for a discrete-time martingale, the following results hold [cf. (2.9) and (2.10)]:

$$M^2 - \langle M \rangle \text{ is a mean zero martingale,} \tag{2.22}$$

$$M^2 - [M] \text{ is a mean zero martingale.} \tag{2.23}$$

By (2.22) and (2.23), $M^2(t) - \langle M \rangle (t)$ and $M^2(t) - [M](t)$ have mean zero for all t. Therefore, since $M(t)$ has mean zero:

$$\mathrm{Var}(M(t)) = \mathrm{E}\left(M(t)^2\right) = \mathrm{E}\langle M \rangle(t) = \mathrm{E}[M](t). \tag{2.24}$$

This shows how the variation processes of a martingale are closely linked to its variance, a fact that will be useful in later chapters when deriving estimators for the variances of statistical estimators and test statistics.

We will often encounter situations with several martingales, and it is then fruitful to define covariation processes for pairs of martingales M_1 and M_2. Corresponding to (2.19) and (2.20) we may define the *predictable covariation process* $\langle M_1, M_2 \rangle$ as the limit (in probability) of the sum of conditional covariances $\mathrm{Cov}(\triangle M_{1k}, \triangle M_{2k} | \mathscr{F}_{(k-1)t/n})$ and the *optional covariation process* $[M_1, M_2]$ as the limit of the sum of the products $\triangle M_{1k} \triangle M_{2k}$. Informally we may write

$$d\langle M_1, M_2 \rangle(t) = \mathrm{Cov}(dM_1(t), dM_2(t) \,|\, \mathscr{F}_{t-}).$$

Note that by the preceding definitions, $\langle M, M \rangle = \langle M \rangle$ and $[M, M] = [M]$. In a similar manner as (2.22) and (2.23) we have that:

$$M_1 M_2 - \langle M_1, M_2 \rangle \text{ is a mean zero martingale,} \tag{2.25}$$

$$M_1 M_2 - [M_1, M_2] \text{ is a mean zero martingale.} \tag{2.26}$$

As a consequence of these results

$$\mathrm{Cov}\,(M_1(t), M_2(t)) = \mathrm{E}\,(M_1(t) M_2(t)) = \mathrm{E}\langle M_1, M_2 \rangle(t) = \mathrm{E}[M_1, M_2](t) \tag{2.27}$$

for all t; cf. (2.24).

The rules for evaluating the (co)variation processes of linear combinations of martingales are similar to the rules for evaluating (co)variances of linear combinations of ordinary random variables. As an example, the predictable variation processes of a sum of martingales may be written

$$\langle M_1 + M_2 \rangle = \langle M_1 \rangle + \langle M_2 \rangle + 2 \langle M_1, M_2 \rangle, \tag{2.28}$$

and a similar relation holds for the optional variation processes.

2.2.2 Stochastic integrals

We will now introduce the stochastic integral as an analog to the transformation for discrete-time martingales. Let $H = \{H(t); t \in [0, \tau]\}$ be a stochastic process that is *predictable*. Intuitively this means that for any time t, the value of $H(t)$ is *known just before t* (possibly apart from unknown parameters). A formal definition of predictability in continuous time is a bit intricate, and we will not go into details about this. But we note that sufficient conditions for H to be predictable are:

- H is *adapted* to the history $\{\mathscr{F}_t\}$.
- The sample paths of H are *left-continuous*.

Predictability may sound like an uninteresting technical assumption. It turns out, however, to be very important in practical calculations when defining test statistics and estimators.

We can now introduce the stochastic integral:

$$I(t) = \int_0^t H(s)\,dM(s).$$

This is a general concept valid for martingales from a much broader setting than counting processes. However, in our context most martingales will arise from counting processes (Section 2.2.5). Stochastic integration is the exact analogue of transformation of discrete-time martingales as defined in (2.12). Analogously to formula (2.13), the stochastic integral can be defined as a limit of such a transformation in the following sense:

$$I(t) = \lim_{n\to\infty} \sum_{k=1}^{n} H_k \triangle M_k,$$

where we have partitioned the time interval $[0,t]$ into n subintervals of length t/n, and we let $H_k = H((k-1)t/n)$ and $\triangle M_k = M(kt/n) - M((k-1)t/n)$. (In the general theory of stochastic integrals, which include integrals with respect to Wiener process martingales, this limiting definition is not valid and one has to introduce the concept of an Itô integral. In this book, Itô integrals are only used briefly in connection with stochastic differential equations in Section 10.4 and Appendix A.4.)

As for a transformation, the major interesting fact about a stochastic integral is that $I(t)$ is a mean zero martingale with respect to $\{\mathscr{F}_t\}$. Hence, *the martingale property is preserved under stochastic integration*. This follows from the corresponding fact for transformations (Section 2.1.3), since the stochastic integral can be seen as a limit of discrete-time versions.

In analogy with formulas (2.14) and (2.15), the following rules hold for evaluating the variation processes of a stochastic integral:

$$\left\langle \int H\,dM \right\rangle = \int H^2\,d\langle M\rangle, \tag{2.29}$$

$$\left[\int H\,dM\right] = \int H^2\,d\,[M], \tag{2.30}$$

while the following rules hold for the covariation processes:

$$\left\langle \int H_1\,dM_1, \int H_2\,dM_2 \right\rangle = \int H_1 H_2\,d\langle M_1, M_2\rangle, \tag{2.31}$$

$$\left[\int H_1\,dM_1, \int H_2\,dM_2\right] = \int H_1 H_2\,d\,[M_1, M_2]; \tag{2.32}$$

cf. Exercise 2.8.

2.2.3 The Doob-Meyer decomposition

In Section 2.1.4 we saw how a discrete-time stochastic process can be decomposed into a predictable process and a sequence of martingale differences. A similar result holds for processes in continuous time and is know as the *Doob-Meyer decomposition*.

To state the content of this decomposition, we first need to consider a specific class of processes. An adapted process $X = \{X(t); t \in [0,\tau]\}$ is called a *submartingale* if it satisfies

$$\mathrm{E}(X(t) \,|\, \mathscr{F}_s) \geq X(s) \quad \text{for all } t > s. \tag{2.33}$$

[Note the similarity with the martingale property (2.17).] Thus a submartingale is a process that tends to increase as time passes. In particular, any nondecreasing process, like a counting process, is a submartingale. The Doob-Meyer decomposition states that any submartingale X can be decomposed *uniquely* as

$$X = X^* + M, \tag{2.34}$$

where X^* is a nondecreasing predictable process, often denoted the *compensator* of X, and M is a mean zero martingale. Heuristically we have that

$$dX^*(t) = \mathrm{E}(dX(t) \,|\, \mathscr{F}_{t-})$$

and

$$dM(t) = dX(t) - \mathrm{E}(dX(t) \,|\, \mathscr{F}_{t-}).$$

Like the Doob decomposition in discrete time (Section 2.1.4), the Doob-Meyer decomposition (2.34) therefore tells us what can be predicted from the past, $dX^*(t)$, and what is the innovation, or surprising element, $dM(t)$.

In (2.22) we noted that if M is a martingale, so is $M^2 - \langle M \rangle$. Here M^2 is a submartingale (by Jensen's inequality) and $\langle M \rangle$ is a nondecreasing predictable process (by construction). This shows that the predictable variation process $\langle M \rangle$ is the compensator of M^2, and this offers an alternative definition of the predictable variation process.

The Doob-Meyer decomposition (2.34) extends immediately to a *special semimartingale*, that is, a process X that is a difference of two submartingales. But then the compensator X^* is no longer a nondecreasing predictable process, but a finite variation predictable process (i.e., a difference of two nondecreasing predictable processes).

2.2.4 The Poisson process

A homogeneous Poisson process describes the distribution of events that occur entirely independently of one another. One imagines a basic rate of occurrence, or intensity, denoted λ, such that the probability of an event occurring in the time interval $[t, t+dt)$ is $\lambda\, dt$. A homogeneous Poisson process has a number of well known properties:

- The time between events is exponentially distributed with probability density $\lambda e^{-\lambda t}$.
- The expected value and the variance of the number of events in a time interval of length h are both equal to λh.
- The number of events in a time interval of length h is Poisson distributed; the probability of exactly k events occurring is $(\lambda h)^k e^{-\lambda h}/k!$.
- The process has independent increments, that is, the number of events in nonoverlapping intervals are independent.

We let $N(t)$ be the number of events in $[0,t]$ and introduce the process

$$M(t) = N(t) - \lambda t, \tag{2.35}$$

obtained by centering the Poisson process (by subtracting its mean). Further we denote by $\mathscr{F}_t$ the information about all events that happen in the time interval $[0,t]$. Due to the independent increments of a Poisson process, we have for all $t > s$:

$$\mathrm{E}\{M(t) - M(s) \,|\, \mathscr{F}_s\} = \mathrm{E}\{M(t) - M(s)\} = \mathrm{E}\{N(t) - N(s)\} - \lambda(t-s) = 0,$$

which yields

$$\mathrm{E}\{M(t) \,|\, \mathscr{F}_s\} = M(s). \tag{2.36}$$

This is the martingale property (2.17), and hence the process (2.35) is a martingale. It follows that λt is the compensator of the Poisson process $N(t)$ (cf. Section 2.2.3). By a similar argument:

$$\mathrm{E}\{M^2(t) - \lambda t \,|\, \mathscr{F}_s\} = M^2(s) - \lambda s \tag{2.37}$$

(Exercise 2.9), which shows that the process $M^2(t) - \lambda t$ is a martingale. Thus λt is also the compensator of $M^2(t)$, and it follows by the comment at the end of Section 2.2.3 that the martingale (2.35) has a predictable variation process

$$\langle M \rangle(t) = \lambda t. \tag{2.38}$$

In the next subsection we will see that relations similar to (2.35), (2.36), and (2.38) are in general valid for counting processes.

2.2.5 Counting processes

As introduced in Section 1.4, a counting process $N = \{N(t); t \in [0, \tau]\}$ is a right-continuous process with jumps of size 1 at event times and constant in between. We assume that the counting process is adapted to the history $\{\mathscr{F}_t\}$, which is just a technical way of saying that the history is generated by N and possibly some external information as well.

The intensity process $\lambda(t)$ of a counting process (w.r.t. the history $\{\mathscr{F}_t\}$) is heuristically defined by

$$\lambda(t)dt = \mathrm{P}(dN(t) = 1 \mid \mathscr{F}_{t-}) = \mathrm{E}(dN(t) \mid \mathscr{F}_{t-}), \tag{2.39}$$

cf. (1.11) and (1.12). To give a precise mathematical definition of an intensity process, first note that since the counting process is nondecreasing, it is a submartingale (Section 2.2.3). Hence by the Doob-Meyer decomposition (2.34), there exist a unique predictable process $\Lambda(t)$, called the *cumulative intensity process*, such that $M(t) = N(t) - \Lambda(t)$ is a mean zero martingale.

Throughout the book we will consider the case where the cumulative intensity process is absolutely continuous. Then there exists a *predictable process* $\lambda(t)$ such that

$$\Lambda(t) = \int_0^t \lambda(s)ds, \tag{2.40}$$

and this gives a formal definition of the intensity process $\lambda(t)$ of the counting process. Further for the absolute continuous case, we have that

$$M(t) = N(t) - \int_0^t \lambda(s)ds \tag{2.41}$$

is a mean zero martingale. As indicated in Section 1.4, this is a key relation that we will use over and over again.

The predictable and optional variation processes of M are defined as the limits in (2.19) and (2.20), respectively. We first look at the latter of the two. To this end, note that the martingale (2.41) has jumps of size 1 at the jump times of N, and that it is continuous between the jump times. When the limit is approached in (2.20), only the jumps will remain. Thus the optional variation process becomes

$$[M](t) = N(t). \tag{2.42}$$

As for the predictable variation process, we will be content with a heuristic argument. By (2.21) and (2.41) we have that

$$\begin{aligned} d\langle M\rangle(t) &= \mathrm{Var}(dM(t) \mid \mathscr{F}_{t-}) \\ &= \mathrm{Var}(dN(t) - \lambda(t)dt \mid \mathscr{F}_{t-}) \\ &= \mathrm{Var}(dN(t) \mid \mathscr{F}_{t-}), \end{aligned}$$

since $\lambda(t)$ is predictable, and hence a fixed quantity given $\mathscr{F}_{t-}$. Now $dN(t)$ may only take the value 0 or 1, and it follows using (2.39) that

$$d\langle M\rangle(t) \approx \lambda(t)dt\,\{1 - \lambda(t)dt\} \approx \lambda(t)dt.$$

This motivates the relation

$$\langle M \rangle (t) = \int_0^t \lambda(s) ds, \tag{2.43}$$

which is another key result that will be used a number of times in later chapters.

Note that (2.41) and (2.43) imply that the cumulative intensity process (2.40) is the compensator of both the counting process, $N(t)$, and the square of the martingale, $M^2(t)$. Note also that (2.41) and (2.43) are similar to the relations (2.35) and (2.38) we derived for the Poisson process in Section 2.2.4. Thus one may say that a counting process has the same "local behavior" as a Poisson process. It is this "local Poisson-ness" of counting processes that is the source of their nice properties.

So far we have considered just a single counting process. In practice we will often have several of them, for instance, corresponding to different individuals or to different groups we want to compare. We assume that the counting processes are adapted to the same history $\{\mathscr{F}_t\}$, so the history is generated by all the counting processes and possibly some external information as well.

Generally, we shall require that no two counting processes in continuous time can jump simultaneously. Consider a pair N_1 and N_2 of counting processes, with corresponding martingales M_1 and M_2. Since the counting processes do not jump simultaneously, the same applies for the martingales M_1 and M_2. From this it follows that

$$\langle M_1, M_2 \rangle (t) = 0 \quad \text{for all } t \tag{2.44}$$

$$[M_1, M_2](t) = 0 \quad \text{for all } t \tag{2.45}$$

(Exercise 2.10). We say that the martingales are *orthogonal*. By (2.25) and (2.26) the orthogonality of M_1 and M_2 is equivalent to the fact that the product $M_1 M_2$ is a martingale.

2.2.6 Stochastic integrals for counting process martingales

In the context of counting processes the stochastic integral

$$I(t) = \int_0^t H(s)\, dM(s)$$

is simple to understand. Using formula (2.41) one simply splits the integral in two as follows:

$$I(t) = \int_0^t H(s)\, dN(s) - \int_0^t H(s)\, \lambda(s)\, ds.$$

For given sample paths of the processes, the last integral is simply an ordinary (Riemann) integral. The first integral, however, is to be understood as a sum of the values of H at every jump time of the counting process. Thus

$$\int_0^t H(s)\,dN(s) = \sum_{T_j \le t} H(T_j),$$

where $T_1 < T_2 < \cdots$ are the ordered jump times of N.

Using (2.29), (2.30), (2.42), and (2.43), we get the following expressions for the predictable and optional variation processes of a stochastic integral of a counting process martingale:

$$\left\langle \int H\,dM \right\rangle(t) = \int_0^t H^2(s)\lambda(s)ds, \tag{2.46}$$

$$\left[\int H\,dM\right](t) = \int_0^t H^2(s)\,dN(s). \tag{2.47}$$

Consider counting processes $N_1, N_2, \ldots, N_k$ with no simultaneous jumps and with intensity processes $\lambda_1, \lambda_2, \ldots, \lambda_k$ (w.r.t. the same history). Then the corresponding martingales $M_1, M_2, \ldots, M_k$ are orthogonal [cf. (2.44)], and (2.31) implies that $\langle \int H_j\,dM_j, \int H_l\,dM_l\rangle(t) = 0$ for all t when $j \neq l$. Using (2.28), we then get the important relation:

$$\left\langle \sum_{j=1}^k \int H_j\,dM_j \right\rangle(t) = \sum_{j=1}^k \int_0^t H_j^2(s)\lambda_j(s)ds. \tag{2.48}$$

In a similar manner, the following result holds for the optional variation process:

$$\left[\sum_{j=1}^k \int H_j\,dM_j\right](t) = \sum_{j=1}^k \int_0^t H_j^2(s)dN_j(s). \tag{2.49}$$

2.2.7 The innovation theorem

The intensity process of a counting process N *relative to a history* $\{\mathscr{F}_t\}$ is given informally by

$$\lambda^{\mathscr{F}}(t)\,dt = \mathrm{E}(dN(t)\,|\,\mathscr{F}_{t-}), \tag{2.50}$$

cf. (2.39). We here make the dependence on the history $\{\mathscr{F}_t\}$ explicit in the notation to point out that the intensity process depends on the history, and that if the history is changed, the intensity process *may* change as well. We will now have a closer look at this.

Consider a counting process N, and let $\{\mathscr{N}_t\}$ be the history (or filtration) generated by the counting process (denoted the *self-exiting filtration*). In Sections 2.2.5 and 2.2.6 it is a key assumption that a counting process is adapted to the history. This means that $\{\mathscr{N}_t\}$ is the smallest history we may consider if the results of Sections 2.2.5 and 2.2.6 are to hold true.

Usually, however, we will consider histories that are not only generated by N, but are generated by N as well as by other counting processes, censoring processes, covariates, etc. that are observed in parallel with N. Consider two such histories, $\{\mathscr{F}_t\}$ and $\{\mathscr{G}_t\}$, and assume that they are *nested*, that is, that $\mathscr{F}_t \subseteq \mathscr{G}_t$ for all t. Thus, at any time t, all information contained in $\mathscr{F}_t$ is also contained in $\mathscr{G}_t$, but $\mathscr{G}_t$ may contain information that is not contained in $\mathscr{F}_t$. Using double expectations, we then have

$$\mathrm{E}(dN(t) \mid \mathscr{F}_{t-}) = \mathrm{E}\{\mathrm{E}(dN(t) \mid \mathscr{G}_{t-}) \mid \mathscr{F}_{t-}\}.$$

By this and (2.50) it follows that the intensity processes of N with respect to the two histories are related as follows:

$$\lambda^{\mathscr{F}}(t) = \mathrm{E}(\lambda^{\mathscr{G}}(t) \mid \mathscr{F}_{t-}). \tag{2.51}$$

This result is called the *innovation theorem*.

It is important to note that the innovation theorem applies only to histories that are nested. Further, in order for N to be adapted to the histories, both of these need to contain $\{\mathscr{N}_t\}$. Thus the innovation theorem holds provided that $\mathscr{N}_t \subseteq \mathscr{F}_t \subseteq \mathscr{G}_t$.

When considering more than one history, it is important to use notation for intensity processes that points out their dependence on the history, as we have done in this subsection. In later chapters, however, it will usually be clear from the context how the history is defined. Then we will just write $\lambda(t)$ for the intensity process without explicitly mentioning the history.

2.2.8 Independent censoring

In Sections 1.4.2 and 1.4.3 we gave an informal discussion of the concept of *independent censoring*; the main point being that independent censoring preserves the form of the intensity processes of the counting processes at hand. We will now discuss more formally the concept of independent censoring.

In order to do that we have to operate with three different models:

(i) a model for the (hypothetical) situation *without censoring*, that is, where all occurrences of the event of interest are observed

(ii) a *joint model* for the (hypothetical) situation where all occurrences of the event of interest *as well as* the censoring processes are observed

(iii) a model for the situation *with censoring*, that is, for the occurrences of the event actually observed

The parameters of interest are defined according to model (i), the concept of independent censoring is defined by means of model (ii), while model (iii) is the one used in the actual statistical inference; cf. Figure 2.1.

We start by considering model (i), that is, the (hypothetical) situation where all occurrences of the event of interest are observed. Let $N_1^c(t), \ldots, N_n^c(t)$ be the counting processes registering these occurrences for each of n individuals, *assuming*

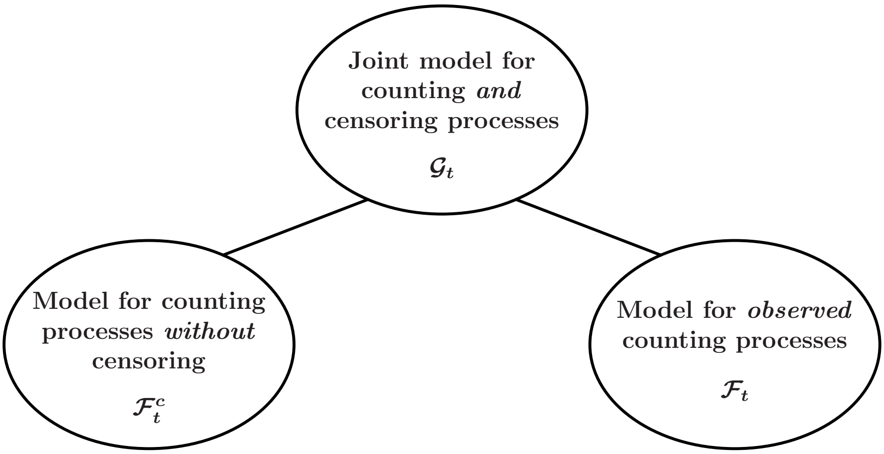

Fig. 2.1 *The three models that are involved in the definition of independent censoring. The model parameters are defined for the model without censoring, the conditions on the censoring are formulated for the joint model, while the statistical methods are derived and studied for the model for the observed counting processes.*

complete observation, and denote by $\mathscr{F}_t^c$ the information that would then have been available to the researcher by time t. The history $(\mathscr{F}_t^c)$ is generated by *all* the counting processes $N_i^c(t)$ and possibly also by covariate processes that run in parallel with the counting processes. We assume that the $(\mathscr{F}_t^c)$-intensity processes of the counting processes take the form

$$\lambda_i^{\mathscr{F}^c}(t) = Y_i^c(t)\alpha_i(t); \qquad i = 1, \dots, n. \tag{2.52}$$

Here $Y_i^c(t)$ is a left-continuous $(\mathscr{F}_t^c)$-adapted indicator process that takes the value 1 if individual i *may* experience an event at time t, and is equal to 0 otherwise. (For example, we will have $Y_i^c(t) = 0$ if individual i has died by time t or if, in a multistate model, the individual is in a state from which the event of interest cannot take place.) The $\alpha_i(t)$ in (2.52) are our key model parameters and, as we will see in later chapters, a main aim of a statistical analysis is to infer how these (and derived quantities) depend on covariates and vary over time. It is important to note that the $\alpha_i(t)$ may be random and depend on covariates as well as previous occurrences of the event (through *the past* $\mathscr{F}_{t-}^c$).

The study of the $\alpha_i(t)$ is complicated by incomplete observation of the counting processes. To handle this, for each $i = 1, \dots, n$ we introduce a left-continuous binary censoring process $Y_i^o(t)$ that takes the value 1 if individual i is under observation "just before" time t, and the value 0 otherwise. For the special case of right-censoring, $Y_i^o(t) = I\{t \le C_i\}$ for a right-censoring time C_i. The observed counting processes are then given by

$$N_i(t) = \int_0^t Y_i^o(u)\,dN_i^c(u); \qquad i = 1, \ldots, n. \tag{2.53}$$

The censoring processes will often create some extra randomness, causing the observed counting processes $N_i(t)$ not to be adapted to the history $(\mathscr{F}_t^c)$. Then we can *not* define the intensity processes for the observed counting processes relative to the complete history $(\mathscr{F}_t^c)$.

To handle this problem, we have to consider the larger model (ii), that is, the *joint model* for the completely observed counting processes $N_i^c(t)$ *and* the censoring processes $Y_i^o(t)$. To this end we consider the larger history $(\mathscr{G}_t)$ generated by the complete history $(\mathscr{F}_t^c)$ *as well as* by the censoring processes. This history corresponds to the (hypothetical) situation where all occurrences of the event of interest are observed, and we *in addition* observe the censoring processes. We may then consider the $(\mathscr{G}_t)$-intensity processes $\lambda_i^{\mathscr{G}}(t)$ of the completely observed counting processes $N_i^c(t)$. If the censoring processes carry information on the likelihood of occurrence of the event, implying that individuals under observation have a different risk of experiencing the event than similar individuals that are not observed, the $\lambda_i^{\mathscr{G}}(t)$ will differ from the $(\mathscr{F}_t^c)$-intensity processes (2.52). We will, however, assume that this is not the case, so that the intensity processes relative to the two histories are the same:

$$\lambda_i^{\mathscr{G}}(t) = \lambda_i^{\mathscr{F}^c}(t); \qquad i = 1, \ldots, n. \tag{2.54}$$

When (2.54) is fulfilled, we say that *censoring is independent*.

Before we discuss the consequences of the independent censoring assumption (2.54), we will take a closer look at the assumption itself. To this end we concentrate on the situation with right-censoring, where n individuals are followed until observation stops at censoring or death (or, more generally, at the entry into an absorbing state). For this situation we have $Y_i^o(t) = I\{t \le C_i\}$ for (potential) right-censoring times C_i; $i = 1, \ldots, n$. A number of different right-censoring schemes are possible. For example:

- Censoring at fixed times that may differ between the individuals, that is, $C_i = c_i$ for given constants c_i (type I censoring).
- Censoring of all individuals at the time T when a specified number of occurrences have taken place, that is, $C_i = T$ for $i = 1, \ldots, n$ (type II censoring).
- Censoring of all individuals when the event has not occurred in a certain time interval.
- Censoring at random times C_i that may differ between the individuals and that are independent of the completely observed counting processes $N_i^c(t)$ (random censoring).

For the first three censoring schemes, the censoring times C_i are *stopping times* relative to the complete history $(\mathscr{F}_t^c)$. Therefore no additional randomness is introduced by the censoring, and there is no need to enlarge the history. Thus the histories $(\mathscr{G}_t)$ and $(\mathscr{F}_t^c)$ are the same, and the independent censoring assumption (2.54) is automatically fulfilled. For the last censoring scheme mentioned, additional randomness is introduced by the censoring. But as the (potential) censoring times C_i are assumed to be independent of the completely observed counting processes $N_i^c(t)$, the independent censoring assumption (2.54) holds for this censoring scheme as well.

We then consider the data actually observed, and we denote the history corresponding to the observed data by $(\mathscr{F}_t)$. The counting processes $N_i(t)$ given by (2.53) are observed, so these are adapted to $(\mathscr{F}_t)$. However, we do not necessarily observe the censoring processes $Y_i^o(t)$; e.g. we do not observe censoring after death. What is observed for each $i = 1, \ldots, n$ is the left-continuous process

$$Y_i(t) = Y_i^c(t)\, Y_i^o(t), \tag{2.55}$$

taking the value 1 if individual i is at risk for the event of interest "just before" time t and the value 0 otherwise. The processes $Y_i(t)$ are therefore adapted to $(\mathscr{F}_t)$ and, due to their left-continuity, they are in fact $(\mathscr{F}_t)$-predictable. We may then adopt an argument similar to the one used to derive the innovation theorem in the previous subsection, to find the $(\mathscr{F}_t)$-intensity processes $\lambda_i^{\mathscr{F}}(t)$ of the observed counting processes. Using (2.52), (2.53), (2.54) and (2.55) we obtain

$$\begin{aligned}
\lambda_i^{\mathscr{F}}(t)\,dt &= \mathrm{E}(\,dN_i(t)\,|\,\mathscr{F}_{t-}) = \mathrm{E}(\,Y_i^o(t)\,dN_i^c(t)\,|\,\mathscr{F}_{t-})\\
&= \mathrm{E}\{\mathrm{E}(\,Y_i^o(t)\,dN_i^c(t)\,|\,\mathscr{G}_{t-})\,|\,\mathscr{F}_{t-}\} = \mathrm{E}\{Y_i^o(t)\,\mathrm{E}(\,dN_i^c(t)\,|\,\mathscr{G}_{t-})\,|\,\mathscr{F}_{t-}\}\\
&= \mathrm{E}\{Y_i^o(t)\,\lambda_i^{\mathscr{G}}(t)\,dt\,|\,\mathscr{F}_{t-}\} = \mathrm{E}\{Y_i^o(t) Y_i^c(t)\,\alpha_i(t)\,dt\,|\,\mathscr{F}_{t-}\}\\
&= Y_i(t)\,\mathrm{E}\{\alpha_i(t)\,|\,\mathscr{F}_{t-}\}\,dt
\end{aligned}$$

In order to progress further, an assumption on the observation of covariates is needed. Typically, the complete history $(\mathscr{F}_t^c)$ will be generated by the fully observed counting processes $N_i^c(t)$ as well as by covariate processes running in parallel with the counting processes, and the $\alpha_i(t)$ of (2.52) may depend on these covariates. For statistical inference on the $\alpha_i(t)$ to be feasible, we must assume that the covariates that enter into the specification of the $\alpha_i(t)$ are available to the researcher. A technical way of formulating this assumption is to assume that the $\alpha_i(t)$ are $(\mathscr{F}_t)$-predictable. Then the intensity processes of the observed counting processes take the form

$$\lambda_i^{\mathscr{F}}(t) = Y_i(t)\alpha_i(t)\,; \qquad i = 1, \ldots, n. \tag{2.56}$$

Comparing this with (2.52), we see that *the form of the intensity processes is preserved under independent censoring*.

There is a close connection between drop-outs in longitudinal data and censoring for survival and event history data. In fact, *independent censoring* in survival and event history analysis is essentially the same as *sequential missingness at random* in longitudinal data analysis (e.g., Hogan et al., 2004).

2.3 Processes with continuous sample paths

In Section 2.1 we considered processes in discrete time, while the counting processes that were the focus of Section 2.2 have discrete state space. We will also consider processes where both time and state space are continuous. Examples of such processes are Wiener processes and Gaussian martingales. These have applications as models for underlying, unobserved processes and as limiting processes of stochastic integrals of counting process martingales.

2.3.1 The Wiener process and Gaussian martingales

The Wiener process, also called Brownian motion, has a similar fundamental character as the Poisson process. While the latter is the model of completely random events, the Wiener process is the model of completely random noise. In fact, the so-called white noise is a kind of derivative of the Wiener process. (This goes beyond the ordinary derivative, which is not valid here.)

Let $W(t)$ denote the value of the Wiener process at time t, and consider a time interval $(s,t]$. Then the increment $W(t) - W(s)$ over this interval is normally distributed with

$$\mathrm{E}\{W(t) - W(s)\} = 0 \quad \text{and} \quad \mathrm{Var}\{W(t) - W(s)\} = t - s.$$

Further, the Wiener process has continuous sample paths, and the increment of the Wiener process over a time interval is independent of its increments over all nonoverlapping intervals. Figure 2.2 shows one realization of the Wiener process.

The Wiener process may be modified in a number of ways; a drift may be introduced such that the process preferentially moves in a certain direction. Further extension of the Wiener process yields the diffusion processes. These extensions of the Wiener process are reviewed in Appendix A.4.

One role of the Wiener process and its allies is as models for underlying processes. By this we mean that an observed event reflects something that occurs on a deeper level. A divorce does not just happen; it is the end result of a long process of deterioration of a marriage. A myocardial infarction is the result of some disease process. Such underlying processes may be seen as diffusions. One would not usually observe them directly but could still use them as the basis for statistical models, which we shall show in Chapters 10 and 11.

Another role of the Wiener process is that time-transformations of the Wiener process arise as limits in a number of applications. In particular, this will be the case for stochastic integrals of counting process martingales, and this provides the basis for the necessary asymptotic theory of estimators and test statistics.

Let $V(t)$ be a strictly increasing continuous function with $V(0) = 0$, and consider the process $U(t) = W(V(t))$. The process U inherits the following properties from the Wiener process:

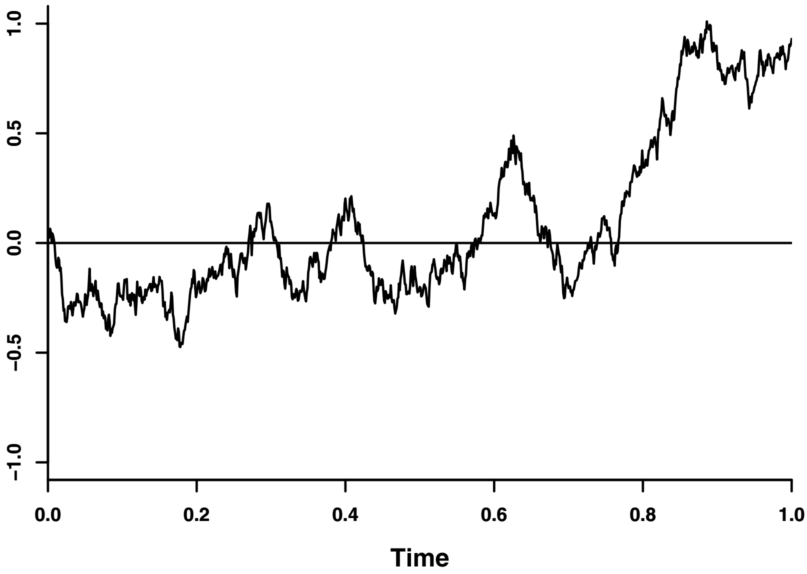

Fig. 2.2 *Simulation of a sample path of the Wiener process.*

- The sample paths are continuous.
- The increments over nonoverlapping intervals are independent.
- The increment over an interval $(s,t]$ is normally distributed with mean zero and variance $V(t)-V(s)$.

From these properties one may show that U is a mean zero martingale with predictable variation process $\langle U \rangle(t) = V(t)$ (Exercise 2.12). It is common to denote U a *Gaussian martingale*.

2.3.2 Asymptotic theory for martingales: intuitive discussion

We have pointed out a number of times that martingales are to be considered as processes of noise, or error, containing random deviations from the expected. In other parts of statistics, one is used to errors being approximately normally distributed. This holds for martingales as well. In fact there are central limit theorems for martingales that are closely analogous to those known for sums of independent random variables.

If we have a sequence of counting processes, where the number of jumps increases and gets more and more dense, the properly normalized associated martingales (or stochastic integrals with respect to these martingales) will converge to a limiting martingale with continuous sample path. The limiting martingale is closely connected to a Wiener process, or Brownian motion. In fact, if the predictable variation process of the limiting martingale equals a deterministic function $V(t)$, then the limiting martingale is exactly the Gaussian martingale discussed at the end of the previous subsection. This follows from the nice fact that a martingale with continuous sample paths is uniquely determined by its variation process.

So there are two things to be taken care of to ensure that a sequence of martingales converges to a Gaussian martingale:

(i) The predictable variation processes of the martingales shall converge to a deterministic function.
(ii) The sizes of the jumps of the martingales shall go to zero.

It is important to note that these two assumptions concern entirely different aspects of the sequence of processes. The first assumption implies a stabilization on the sample space of processes, while the second one is a requirement on the sample paths of the processes.

Figure 2.3 illustrates the convergence of a sequence of normalized counting process martingales to a Gaussian martingale. More specifically the figure shows $n^{-1/2}M(t)$ for $n = 10, 50, 250$, and 1250, where $M(t) = N(t) - \Lambda(t)$ is derived from censored Weibull survival data as described in Example 1.18 (for $n = 10$). In fact, the upper left-most panel of Figure 2.3 shows a normalized version of the martingale in the right-hand panel of Figure 1.14.

2.3.3 Asymptotic theory for martingales: mathematical formulation

There exist several versions of the central limit theorem for martingales that formalize requirements (i) and (ii) of the previous subsection. A very general and elegant theorem was formulated by Rebolledo (1980). The version of Rebolledo's theorem we present here is taken from Andersen et al. (1993, section II.5), where more details, and other versions of the conditions, can be found. Helland (1982) showed how Rebolledo's theorem in continuous time can be deduced from the simpler central limit theorem for discrete time martingales.

Let $\widetilde{M}^{(n)}$, $n \geq 1$, be a sequence of mean zero martingales defined on $[0, \tau]$, and let $\widetilde{M}_\varepsilon^{(n)}$ be the martingale containing all the jumps of $\widetilde{M}^{(n)}$ larger than a given $\varepsilon > 0$. Let $\xrightarrow{P}$ denote convergence in probability, and consider the conditions:

(i) $\langle \widetilde{M}^{(n)} \rangle(t) \xrightarrow{P} V(t)$ for all $t \in [0, \tau]$ as $n \to \infty$, where V is a strictly increasing continuous function with $V(0) = 0$.

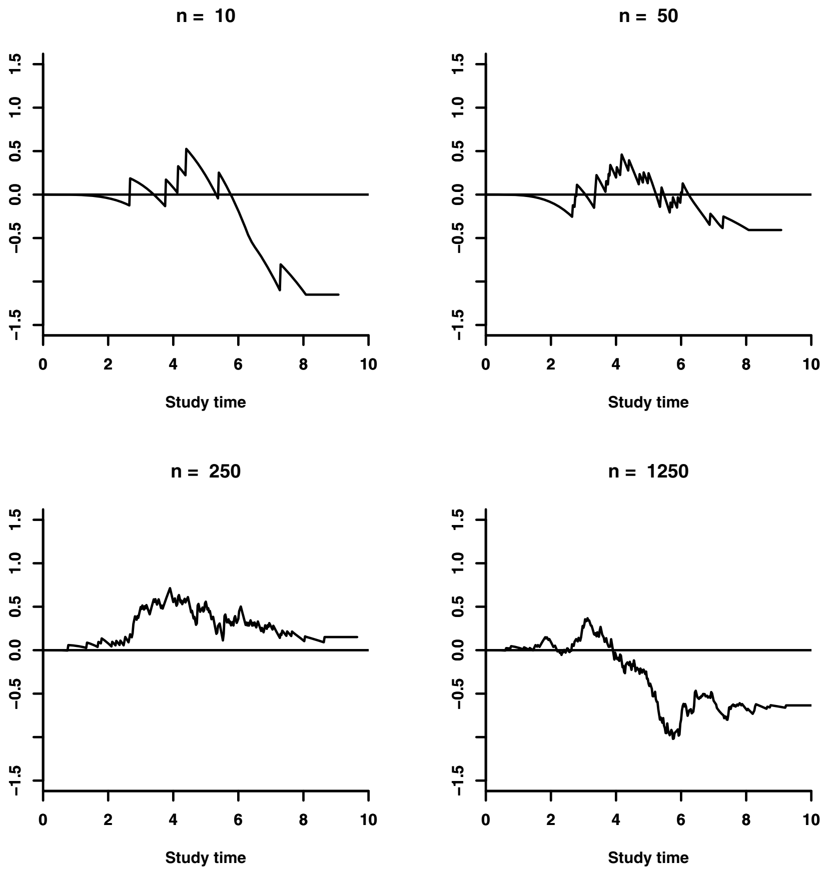

Fig. 2.3 *Illustration that a sequence of normalized counting process martingales converges to a Gaussian martingale (see text for details).*

(ii) $\langle \widetilde{M}_{\varepsilon}^{(n)} \rangle(t) \xrightarrow{P} 0 \quad$ for all $t \in [0, \tau]$ and all $\varepsilon > 0$ as $n \to \infty$.

Then, as $n \to \infty$, the sequence of martingales $\widetilde{M}^{(n)}$ converges in distribution to the mean zero Gaussian martingale U given by $U(t) = W(V(t))$.

Hence, under quite general assumptions there will be convergence in distribution to a limiting Gaussian martingale. Since the relevant statistics may also be functionals of the processes, many different probability distributions may arise from the theory. Note in particular that $\widetilde{M}^{(n)}(t)$ converges in distribution to a normally distributed random variable with mean zero and variance $V(t)$ for any given value of t.

We will use the martingale central limit theorem to derive the limiting behavior of (sequences of) stochastic integrals of the form $\int_0^t H^{(n)}(s)\, dM^{(n)}(s)$, where $H^{(n)}(t)$

is a predictable process and $M^{(n)}(t) = N^{(n)}(t) - \int_0^t \lambda^{(n)}(s)ds$ is a counting process martingale. More generally, we will consider sums of stochastic integrals

$$\sum_{j=1}^{k} \int_0^t H_j^{(n)}(s)\, dM_j^{(n)}(s), \tag{2.57}$$

where $H_j^{(n)}(t)$ is a predictable process for each n and

$$M_j^{(n)}(t) = N_j^{(n)}(t) - \int_0^t \lambda_j^{(n)}(s)ds$$

is a counting process martingale; $j = 1,\ldots,k$. For this situation, conditions (i) and (ii) take the form

$$\sum_{j=1}^{k} \int_0^t (H_j^{(n)}(s))^2\, \lambda_j^{(n)}(s)\, ds \xrightarrow{P} V(t) \text{ for all } t \in [0,\tau], \tag{2.58}$$

$$\sum_{j=1}^{k} \int_0^t (H_j^{(n)}(s))^2\, I\{|H_j^{(n)}(s)| > \varepsilon\}\, \lambda_j^{(n)}(s)\, ds \xrightarrow{P} 0 \text{ for all } t \in [0,\tau]. \tag{2.59}$$

When we are going to use the martingale central limit theorem to show that an estimator or a test statistic converges in distribution, we have to check that conditions (2.58) and (2.59) hold. As we do not focus on regularity conditions in this book, we will not check the two conditions in detail in later chapters.

Assume that we may write $V(t) = \int_0^t v(s)ds$. Then, apart from some regularity conditions, a sufficient condition for (2.58) is that

$$\sum_{j=1}^{k} (H_j^{(n)}(s))^2 \lambda_j^{(n)}(s) \xrightarrow{P} v(s) > 0 \text{ for all } s \in [0,\tau], \text{ as } n \to \infty. \tag{2.60}$$

Furthermore, if k is fixed, a sufficient condition for (2.59) is that

$$H_j^{(n)}(s) \xrightarrow{P} 0 \text{ for all } j = 1,\ldots,k \text{ and } s \in [0,\tau], \text{ as } n \to \infty. \tag{2.61}$$

When $k = k_n$ is increasing with n (as is the case when $k = n$), it may be more involved to check condition (2.59), and we will not go into details on how this can be done.

To summarize: if (2.58) and (2.59) hold, then the sequence of sums of stochastic integrals $\sum_{j=1}^{k} \int_0^t H_j^{(n)}(s)\, dM_j^{(n)}(s)$, $n \geq 1$, converges in distribution to a mean zero Gaussian martingale with variance function $V(t) = \int_0^t v(s)ds$. In particular, for any given value of t, it converges in distribution to a normally distributed random variable with mean zero and variance $V(t)$. Under some regularity conditions, a sufficient condition for (2.58) is (2.60), and when k is fixed, a sufficient condition for (2.59) is (2.61).

In some applications in later chapters, we will need a multivariate version of the martingale central limit theorem. This is given in Appendix B.3.

2.4 Exercises

2.1 Show that (2.3) and (2.4) are equivalent. [*Hint:* By a general result for conditional expectations we have that $\mathrm{E}(M_n \mid \mathscr{F}_{m_1}) = \mathrm{E}\{\mathrm{E}(M_n \mid \mathscr{F}_{m_2}) \mid \mathscr{F}_{m_1}\}$ for all $0 \leq m_1 < m_2 < n$.]

2.2 Let $M = \{M_0, M_1, M_2, \ldots\}$ be a martingale. Show that $\mathrm{Cov}(M_m, M_n - M_m) = 0$ for all $n > m$. [*Hint:* Use the rule of double expectation to show that $\mathrm{Cov}(M_m, M_n - M_m) = \mathrm{E}\{\mathrm{E}(M_m\,(M_n - M_m) \mid \mathscr{F}_m)\}$.]

2.3 Let $M_n = \sum_{k=0}^{n} X_k$, where the $X_1, X_2, \ldots$ are independent random variables with zero mean and variance σ^2, and $X_0 = 0$.

a) Show that $M = \{M_0, M_1, M_2, \ldots\}$ is a martingale (w.r.t. the history generated by the process itself).
b) Use (2.7) to find the predictable variation process $\langle M \rangle$.
c) Use (2.8) to find the optional variation process $[M]$.

2.4 Prove (2.9), that is, that $M_0^2 - \langle M \rangle_0 = 0$ and that $\mathrm{E}(M_n^2 - \langle M \rangle_n \mid \mathscr{F}_{n-1}) = M_{n-1}^2 - \langle M \rangle_{n-1}$.

2.5 Assume that the processes $M_1 = \{M_{10}, M_{11}, \ldots\}$ and $M_2 = \{M_{20}, M_{21}, \ldots\}$, with $M_{10} = M_{20} = 0$, are martingales with respect to the history $\{\mathscr{F}_n\}$. The *predictable covariation process* $\langle M_1, M_2 \rangle$ is for $n \geq 1$ defined by

$$\langle M_1, M_2 \rangle_n = \sum_{i=1}^{n} \mathrm{E}\{\triangle M_{1i} \triangle M_{2i} \mid \mathscr{F}_{i-1}\} \tag{2.62}$$
$$= \sum_{i=1}^{n} \mathrm{Cov}(\triangle M_{1i}, \triangle M_{2i} \mid \mathscr{F}_{i-1}),$$

while $\langle M_1, M_2 \rangle_0 = 0$. The *optional covariation process* $[M_1, M_2]$ is defined by

$$[M_1, M_2]_n = \sum_{i=1}^{n} \triangle M_{1i} \triangle M_{2i} \tag{2.63}$$

for $n \geq 1$ and $[M_1, M_2]_0 = 0$.

a) Show that $M_1 M_2 - \langle M_1, M_2 \rangle$ and $M_1 M_2 - [M_1, M_2]_n$ are mean zero martingales.
b) Show that $\mathrm{Cov}(M_{1n}, M_{2n}) = \mathrm{E}\,\langle M_1, M_2 \rangle_n = \mathrm{E}\,[M_1, M_2]_n$ for all n.

2.6 Show that the process M^T defined by (2.11) is a martingale. [*Hint:* Use the martingale preservation property of the transformation $Z = H \bullet M$ with an appropriate choice of the predictable process H.]

2.7 Prove the statement (2.15). [*Hint:* Use the definition (2.8).]

2.8 Assume that the processes $M_1 = \{M_{10}, M_{11}, \ldots\}$ and $M_2 = \{M_{20}, M_{21}, \ldots\}$, with $M_{10} = M_{20} = 0$, are martingales with respect to the history $\{\mathscr{F}_n\}$, and let $H_1 = \{H_{10}, H_{11}, \ldots\}$ and $H_2 = \{H_{20}, H_{21}, \ldots\}$ be predictable processes. Show that

$$\langle H_1 \bullet M_1, H_2 \bullet M_2 \rangle_n = \sum_{s=1}^{n} H_{1s} H_{2s} \triangle \langle M_1, M_2 \rangle_s$$

$$[H_1 \bullet M_1, H_2 \bullet M_2]_n = \sum_{s=1}^{n} H_{1s} H_{2s} \triangle [M_1, M_2]_s.$$

[*Hint:* Use the definitions (2.62) and (2.63) in Exercise 2.5.]

2.9 Prove (2.37). [*Hint:* Use that $M^2(t) = N^2(t) - 2\lambda t N(t) + (\lambda t)^2$ and the properties of the Poisson process $N(t)$.]

2.10 Assume that the counting processes N_1 and N_2 do not jump simultaneously. Prove that the covariation processes $\langle M_1, M_2 \rangle$ and $[M_1, M_2]$ of the corresponding martingales are both identically equal to zero. [*Hint:* $N_1 + N_2$ is a counting process with $[M_1 + M_2] = N_1 + N_2$. The result follows from the analog to (2.28).]

2.11 Let $N(t)$ be an inhomogeneous Poisson process with intensity $\lambda(t)$. Then the number of events $N(t) - N(s)$ in the time interval $(s,t]$ is Poisson distributed with parameter $\int_s^t \lambda(u)\,du$, and the number of events in disjoint time intervals are independent. Let $\mathscr{F}_t$ be generated by $N(s)$ for $s \leq t$, and let $M(t) = N(t) - \int_0^t \lambda(u)\,du$.

a) Prove that $\mathrm{E}(M(t) \mid \mathscr{F}_s) = M(s)$ for all $s \leq t$, that is, that $M(t)$ is a martingale.
b) Prove that $\mathrm{E}\left(M^2(t) - \int_0^t \lambda(u)\,du \mid \mathscr{F}_s\right) = M^2(s) - \int_0^s \lambda(u)\,du$, that is, that $M^2(t) - \int_0^t \lambda(u)\,du$ is a martingale. Note that this shows that $\langle M \rangle(t) = \int_0^t \lambda(u)du$.

2.12 Let $W(t)$ be the Wiener process. Then the increment $W(t) - W(s)$ over the time interval $(s,t]$ is normally distributed with mean zero and variance $t - s$, and the increments over disjoint time intervals are independent. Let $V(t)$ be a strictly increasing continuous function with $V(0) = 0$, and introduce the stochastic process $U(t) = W(V(t))$. Finally let $\mathscr{F}_t$ be generated by $U(s)$ for $s \leq t$.

a) Prove that $\mathrm{E}(U(t) \mid \mathscr{F}_s) = U(s)$ for all $s \leq t$, that is, that $U(t)$ is a martingale.
b) Prove that $\mathrm{E}\left(U^2(t) - V(t) \mid \mathscr{F}_s\right) = U^2(s) - V(s)$, that is, that $U^2(t) - V(t)$ is a martingale. Note that this shows that $\langle U \rangle(t) = V(t)$.

Springer

http://www.springer.com/978-0-387-20287-7

Survival and Event History Analysis
A Process Point of View
Aalen, O.; Borgan, O.; Gjessing, H.
2008, XVIII, 540 p., Hardcover
ISBN: 978-0-387-20287-7