STK4080 SURVIVAL AND EVENT HISTORY ANALYSIS

Slides 5: Counting processes and martingales

Bo Lindqvist
Department of Mathematical Sciences
Norwegian University of Science and Technology
Trondheim

https://www.ntnu.no/ansatte/bo.lindqvist
bo.lindqvist@ntnu.no

University of Oslo, Autumn 2019

A stochastic process is a collection of random variables, $\{X(t) : t \in \mathcal{T}\}$
Index set $\mathcal{T}$ is usually time, e.g.

- $\mathcal{T} = [0, \infty)$ ("continuous time") or
- $\mathcal{T} = \{0, 1, \ldots\} \equiv \mathcal{N}$ ("discrete time")

COUNTING PROCESSES $\{N(t) : t \geq 0\}$:

- $N(t) - N(s) = \#$ events in interval $(s, t]$ for $s < t$
- $N(t) \in \{0, 1, \ldots\}$
- $N(t) \uparrow t$ in steps of height 1

EXAMPLES

DISCRETE TIME
Random walk X_n, e.g. $X_n =$ better's gain after n games

CONTINUOUS TIME
Poisson process $N(t)$
Wiener process $W(t)$ (Brownian motion)

IMPORTANT IN SURVIVAL ANALYSIS

The dynamic aspects of a process,
i.e. given the past - what is most likely to happen in the future? What is the probability of various outcomes in the future?
Dynamic representations of processes are important for:

- Modeling purposes
- Deriving likelihood functions for statistical inference
- Studying behaviour of estimators

CONDITIONAL DISTRIBUTIONS GIVEN PAST

DISCRETE TIME:
Distribution of $X_n, X_{n+1}, \ldots$ **given** $X_1, X_2, \ldots, X_{n-1}$ (or "given $\mathcal{F}_{n-1}$")

CONTINUOUS TIME
Distribution of $X(s)$ for $s \geq t$ **given** $X(u)$ for $0 \leq u < t$ (or "given $\mathcal{F}_{t-}$")

JOINT DISTRIBUTIONS

$X_1, X_2, \ldots, X_n$ random variables

$$f(x_1, x_2, \ldots, x_n) = P(X_1 = x_1, \ldots, X_n = x_n) \text{ (discrete)}$$

$$f(x_1, x_2, \ldots, x_n) = \lim_{\Delta_i \to 0} \frac{P(x_1 \le X_1 \le x_1 + \Delta_1, \ldots, x_n \le X_n \le x_n + \Delta_n)}{\Delta_1 \cdots \Delta_n} \text{ (contin.)}$$

MARGINAL DISTRIBUTIONS

For $m < n$,

$$f(x_1, x_2, \ldots, x_m) = \begin{cases} \sum_{x_{m+1}, \ldots, x_n} f(x_1, x_2, \ldots, x_m, x_{m+1}, \ldots, x_n) \\ \int \cdots \int f(x_1, x_2, \ldots, x_m, x_{m+1}, \ldots, x_n) dx_{m+1} \cdots dx_n \end{cases}$$

CONDITIONAL DISTRIBUTIONS

$$f(x_{m+1}, \ldots, x_n | x_1, x_2, \ldots, x_m) = \frac{f(x_1, x_2, \ldots, x_m, x_{m+1}, \ldots, x_n)}{f(x_1, x_2, \ldots, x_m)}$$

THE RULE OF DOUBLE EXPECTATION

$E(X_n|x_1, x_2, \ldots, x_{n-1}) \equiv g(x_1, x_2, \ldots, x_{n-1})$ is a number;

$E(X_n|X_1, X_2, \ldots, X_{n-1}) \equiv g(X_1, X_2, \ldots, X_{n-1})$ is a random variable; given as a function of $X_1, X_2, \ldots, X_{n-1}$

As such it has an expected value. What is this value?

$$
\begin{aligned}
E\left[E(X_n|X_1, X_2, \ldots, X_{n-1})\right] &= E[g(X_1, X_2, \ldots, X_{n-1})] \\
&= \sum_{x_1, x_2, \ldots, x_{n-1}} g(x_1, x_2, \ldots, x_{n-1}) f(x_1, x_2, \ldots, x_{n-1}) \\
&= \sum_{x_1, x_2, \ldots, x_{n-1}} E(X_n|x_1, x_2, \ldots, x_{n-1}) f(x_1, x_2, \ldots, x_{n-1}) \\
&= \sum_{x_1, x_2, \ldots, x_{n-1}} \frac{\sum_{x_n} x_n f(x_1, x_2, \ldots, x_n)}{f(x_1, x_2, \ldots, x_{n-1})} f(x_1, x_2, \ldots, x_{n-1} \\
&= \sum_{x_1, x_2, \ldots, x_{n-1}} \sum_{x_n} x_n f(x_1, x_2, \ldots, x_n) \\
&= \sum_{x_n} x_n \sum_{x_1, x_2, \ldots, x_{n-1}} f(x_1, x_2, \ldots, x_n) \\
&= \sum x_n f(x_n) = E(X_n)
\end{aligned}
$$

INTERPRETATION OF "THE RULE OF DOUBLE EXPECTATION"

$$E\left[E(X_n|X_1, X_2, \ldots, X_{n-1})\right] = E(X_n)$$

"If I observe $X_1, X_2, \ldots, X_{n-1}$ many times, and each time compute the expected value $E(X_n|X_1, X_2, \ldots, X_{n-1})$, then on average I will get the value $E(X_n)$".

EXERCISE 1

Throw a die several times. Let Y_i be result in ith throw, and let $X_i = Y_1 + \ldots + Y_i$ be the sum of the i first throws.

(a) Find an expression for $f(x_1, x_2)$.

(b) Find an expression for $f(x_2|x_1)$.

(c) Find an expression for $E(X_2|x_1)$ as a function of x_1.

(d) Find the expectation of the random variable $E(X_2|X_1)$. Find $E(X_2)$ from this. Find also $E(X_2)$ without using the rule of double expectation.

(e) Find also $E(X_3|X_2)$, $E(X_3|X_1, X_2)$ and $E(X_3|X_1)$.

A CONDITIONAL DOUBLE EXPECTATION RESULT

$$E(X_3|X_1) = E[E(X_3|X_1, X_2)|X_1]$$

which if we suppress the conditioning on X_1 throughout,
is merely the rule of double expectation:

$$E(X_3) = E[E(X_3|X_2)]$$

THE HISTORY OF A PROCESS

Notation:

$$\mathcal{F}_n = \text{“information contained in } X_1, X_2, \ldots, X_n\text{”}$$

or “the history at time n”, or the 'past' at time n; meaning the set of all events that can be described in terms of $X_1, X_2, \ldots, X_n$, (these form a so called σ-algebra $\mathcal{F}_n$).

Remark 1: $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3 \subset \cdots$

EXAMPLE: The event $\{X_2 > 2X_4\}$ is in $\mathcal{F}_4$ and $\mathcal{F}_{17}$ but not in $\mathcal{F}_2$.

Remark 2: In applications it may be convenient to let the $\mathcal{F}_n$ contain also auxiliary (random) information, provided $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3 \subset \cdots$ still holds.

Unless otherwise stated, we assume that $\mathcal{F}_n$ is generated from $X_1, X_2, \ldots, X_n$ alone.

- Can now write $E(X_n|X_1, X_2, \ldots, X_{n-1}) = E(X_n|\mathcal{F}_{n-1})$
- More generally we can consider $E(Y|\mathcal{F}_{n-1})$ or $E(Y|\mathcal{F}_n)$ for any random variable Y which is a function of the full process $X_1, X_2, \ldots$, i.e. $Y \in \mathcal{F}$, where $\mathcal{F}$ is the information in all of $X_1, X_2, \ldots$, so $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots \subset \mathcal{F}$

CONDITIONAL EXPECTATION GIVEN THE HISTORY

Let $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3 \subset \cdots \subset \mathcal{F}$ and let $Y \in \mathcal{F}$. Then the

conditional expectation

$$E(Y|\mathcal{F}_n) \; (= E(Y|X_1, X_2, \ldots, X_n))$$

is interpreted as "what can be said about Y after having observed (only) $X_1, X_2, \ldots, X_n$, i.e. knowing only the history $\mathcal{F}_n$"

It is uniquely characterized by

1. $E(Y|\mathcal{F}_n) \in \mathcal{F}_n$, i.e. is a function of $X_1, X_2, \ldots, X_n$ only.
2. For any $Z \in \mathcal{F}_n$ we have

$$E[ZE(Y|\mathcal{F}_n)] = E[ZY]$$

For intuition:

- Special case of "2" is rule of double expectation:

$$E[E(Y|X_1, X_2, \ldots, X_n)] = E[Y] \quad (\text{put } Z \equiv 1)$$

- For $Z \in \mathcal{F}_n$ it holds that $E(ZY|\mathcal{F}_n) = ZE(Y|\mathcal{F}_n)$, since Z is a 'constant' with respect to $\mathcal{F}_n$. Take expectations to get "2"

PROPERTIES OF CONDITIONAL EXPECTATIONS

Recall: For $Y \in \mathcal{F}$, i.e. a function of all the $X_1, X_2, \ldots$

1. $E(Y|\mathcal{F}_n) \in \mathcal{F}_n$, i.e. is a function of $X_1, X_2, \ldots, X_n$ only.
2. For any $Z \in \mathcal{F}_n$ we have

$$E[ZE(Y|\mathcal{F}_n)] = E[ZY]$$

From these we can derive

(i) $E[E(Y|\mathcal{F}_n)] = E(Y)$ (rule of double expectation)

(ii) $E(aY_1 + bY_2|\mathcal{F}_n) = aE(Y_1|\mathcal{F}_n) + bE(Y_2|\mathcal{F}_n)$ (linearity)

(iii) If $Y \in \mathcal{F}_n$ then $E(Y|\mathcal{F}_n) = Y$

(iv) If Y is independent of $\mathcal{F}_n$ i.e. of $X_1, X_2, \ldots, X_n$, then $E(Y|\mathcal{F}_n) = E(Y)$

(v) If $Z \in \mathcal{F}_n$ then $E(ZY|\mathcal{F}_n) = ZE(Y|\mathcal{F}_n)$

Example: $E(X_nX_{n+1}|\mathcal{F}_n) = X_nE(X_{n+1}|\mathcal{F}_n)$

EXERCISE 2

Let $X_1, X_2, \ldots$ be independent with $E(X_n) = \mu$ for all n.

Let $S_n = X_1 + \ldots + X_n$.

Find $E(S_4|\mathcal{F}_2)$ and in general $E(S_n|\mathcal{F}_m)$ when $m \leq n$

GENERALIZATION OF $E(X_3|X_1) = E[E(X_3|X_1, X_2)|X_1]$

We have:

$$E(X_n|X_1, \ldots, X_{m_1}) = E[E(X_n|X_1, \ldots, X_{m_2})|X_1, \ldots, X_{m_1}] \tag{1}$$

for $1 \leq m_1 < m_2 < n$

Verbally:

"If I want to predict, at time m_1, what can be said about X_n at time $m_2 > m_1$, then this is the same as what can be said about X_n at time m_1".

Now (1) can be written

$$E(X_n|\mathcal{F}_{m_1}) = E[E(X_n|\mathcal{F}_{m_2})|\mathcal{F}_{m_1}] \text{ for } 1 \leq m_1 < m_2 < n$$

More generally we have, for any $Y \in \mathcal{F}$,

$$E(Y|\mathcal{F}_{m_1}) = E[E(Y|\mathcal{F}_{m_2})|\mathcal{F}_{m_1}] \text{ for } 1 \leq m_1 < m_2$$

which can be proved from the two characterizing properties of conditional expectations.

A stochastic process $M = \{M_0, M_1, M_2, \ldots\}$ is called a *martingale* if

$$E(M_n | M_1, \ldots, M_{n-1}) = M_{n-1} \text{ for } n = 1, 2, \ldots$$

or, more compactly,

$$E(M_n | \mathcal{F}_{n-1}) = M_{n-1} \text{ for } n = 1, 2, \ldots \tag{2}$$

where $\mathcal{F}_k =$ information contained in $M_0, M_1, \ldots, M_k$.

Thus - given the history to time $n-1$, the expected value at the next step equals the current state.

This can be taken as the definition of a *fair game:*

If the gambler's gain after $n-1$ games is M_{n-1}, then the expected gain after the next game is the same as the current.

MORE GENERAL DEFINITION OF MARTINGALE

As mentioned earlier, we may let the $\mathcal{F}_n$ contain extra information, provided

- $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \cdots$
- $M_n \in \mathcal{F}_n$ for each n
  ("the process M is *adapted* to the history $\{\mathcal{F}_n\}$")
- $E|M_n| < \infty$ for each n
- $E(M_n|\mathcal{F}_{n-1}) = M_{n-1}$ for $n = 1, 2, \ldots$

EXPECTED VALUE OF MARTINGALE IS CONSTANT

Recall definition:

$$E(M_n|\mathcal{F}_{n-1}) = M_{n-1}$$

Taking the expected value on each side, the rule of double expectation gives

$$E(M_n) = E(M_{n-1})$$

Hence

$$E(M_0) = E(M_1) = \cdots = \text{(usually) } 0$$

("mean zero martingale")

Often it is assumed that $M_0 = 0$.

EXERCISE 3

Let $X_1, X_2, \ldots$ be independent with

$$P(X_i = 1) = P(X_i = -1) = 1/2$$

Think of X_i as result of a game where one flips a coin and wins 1 unit if "heads" and loses 1 unit if "tails".

Let $M_n = X_1 + \ldots + X_n$ be the gain after n games, $n = 1, 2, \ldots$

Show that M_n is a (mean zero) martingale.

EXERCISE 4

(a) Let $X_1, X_2, \ldots$ be independent with $E(X_n) = 0$ for all n.
Let $M_n = X_1 + \ldots + X_n$.
Show that M_n is a (mean zero) martingale.

(b) Let $X_1, X_2, \ldots$ be independent with $E(X_n) = \mu$ for all n.
Let $S_n = X_1 + \ldots + X_n$.
Compute $E(S_n|\mathcal{F}_{n-1})$ (see Exercise 2). Is $\{S_n\}$ a martingale?
Can you find a simple transformation of S_n which is a martingale?

(c) Let $X_1, X_2, \ldots$ be independent with $E(X_n) = 1$ for all n.
Let $M_n = X_1 \cdot X_2 \cdot \ldots \cdot X_n$.
Show that M_n is a martingale. What is $E(M_n)$?

EXERCISE 5

Show that for a martingale $\{M_n\}$ is

(a) $E(M_n|\mathcal{F}_m) = M_m$ for all $m < n$.
This is essentially Exercise 2.1 in book.

(b) $E(M_m|\mathcal{F}_n) = M_m$ for all $m < n$

(c) Give verbal (intuitive) interpretations of each of (a) and (b).

EXERCISE 6

Define the *martingale differences* by

$$\Delta M_n = M_n - M_{n-1}$$

(a) Show that the definition of martingale, $E(M_n|\mathcal{F}_{n-1}) = M_{n-1}$, is equivalent to

$$E(M_n - M_{n-1}|\mathcal{F}_{n-1}) = 0, \text{ i.e. } E(\Delta M_n|\mathcal{F}_{n-1}) = 0 \qquad (3)$$

(*Hint: Why is* $E(M_{n-1}|\mathcal{F}_{n-1}) = M_{n-1}$*?*)

(b) Show that for a martingale we have

$$\mathrm{Cov}(M_{n-1} - M_{n-2}, M_n - M_{n-1}) = 0 \text{ for all } n \qquad (4)$$

i.e. $\mathrm{Cov}(\Delta M_{n-1}, \Delta M_n) = 0$. Explain in words what this means.

(c) Show that (3) and (4) automatically hold when $M_n = X_1 + \ldots + X_n$ for independent $X_1, X_2, \ldots$.

Note that in this case the differences $\Delta M_n = M_n - M_{n-1} = X_n$ are independent.

Thus (4) shows that martingale differences correspond to a weakening of the independent increments property

STOPPING TIMES

Let $M = \{M_0, M_1, \ldots\}$ be a process adapted to $\{\mathcal{F}_n\}$.

Let T be a positve integer valued random variable.

T is a *stopping time* if for all n the event $\{T = n\}$ is in $\mathcal{F}_n$, i.e. at any time n we can decide from $M_0, M_1, \ldots, M_n$ whether we should stop or not!

Examples of stopping times:

- Let k be a fixed integer and let $T = k$
- Let A be a set and let T be the first time the process hits A. Example: If M_n is gain after n games (Exercise 3), then we may stop when $M_n \geq 10$.
- In survival analysis: T is usually a *censoring* time, i.e. unit is not observed beyond T even if still alive.

STOPPED PROCESS

The *stopped process* is denoted M^T and is defined by

$$M_n^T = M_{n \wedge T} \equiv \begin{cases} M_n \text{ if } n \leq T \\ M_T \text{ if } n > T \end{cases}$$

(so we "freeze" the process at time T).

It will be shown later, by using a more general result on transformations of martingales (Exercise 2.6 in book), that the stopped process M^T is a martingale.

TRANSFORMATION OF A MARTINGALE

Consider again EXERCISE 3: Let $X_1, X_2, \ldots$ be independent with $P(X_i = 1) = P(X_i = -1) = 1/2$. X_i is result of a game where one flips a coin and wins 1 unit if "heads" and loses 1 unit if "tails".

$M_n = X_1 + \ldots + X_n$ is the gain after n games, and is a martingale (Exercise 3).

Suppose now that in the nth game we make the bet H_n, where in order to determine H_n we use information from the first $n-1$ games (i.e. $\mathcal{F}_{n-1}$). Now we either lose or win an amount H_n in the nth game. The gain after n games is therefore

$$\begin{aligned} Z_n &= H_1 X_1 + H_2 X_2 + \ldots + H_n X_n \\ &= H_1(M_1 - M_0) + H_2(M_2 - M_1) + \ldots + H_n(M_n - M_{n-1}) \end{aligned}$$

where $M_0 = 0$. This is a special case of what is called *the transformation of M by H* and is written $Z = H \bullet M$.

The interesting and useful property is that $Z = H \bullet M$ is a martingale whenever M is. The clue is that $H_n \in \mathcal{F}_{n-1}$ so that H is a so-called *predictable* process.

Main result: The transformation of a martingale by a predictable process is again a martingale.

PROOF OF MARTINGALE PROPERTY

Let

$$Z_n = H_1(M_1 - M_0) + H_2(M_2 - M_1) + \ldots + H_n(M_n - M_{n-1})$$

Then

$$\begin{aligned} E(Z_n - Z_{n-1}|\mathcal{F}_{n-1}) &= E(H_n(M_n - M_{n-1})|\mathcal{F}_{n-1}) \\ &= H_n\, E(M_n - M_{n-1}|\mathcal{F}_{n-1}) \\ &= 0 \end{aligned}$$

where we use that $H_n \in \mathcal{F}_{n-1}$.

Conclusion: "You cannot beat a fair game". But - see next slide...

ON MARTINGALES IN WIKIPEDIA

"*Originally, martingale referred to a class of betting strategies that was popular in 18th century France.* The simplest of these strategies was designed for a game in which the gambler wins his stake if a coin comes up heads and loses it if the coin comes up tails. The strategy had the gambler double his bet after every loss, so that the first win would recover all previous losses plus win a profit equal to the original stake. Since as a gambler's wealth and available time jointly approach infinity his probability of eventually flipping heads approaches 1, the martingale betting strategy was seen as a sure thing by those who practiced it. Of course in reality the exponential growth of the bets would eventually bankrupt those foolish enough to use the martingale for a long time.

The concept of martingale in probability theory was introduced by Paul Pierre Lévy, and much of the original development of the theory was done by Joseph Leo Doob. Part of the motivation for that work was to show the impossibility of successful betting strategies."

MARTINGALE STRATEGY - FIRST WIN IS IN nth GAME

Game #	Stake (H)	Result (Lose/Win)	Gain (Z)
1	1	L	-1
2	2	L	-3
3	4	L	-7
4	8	L	-15
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$n-1$	2^{n-2}	L	$-(2^{n-1}-1)$
n	2^{n-1}	W	$2^{n-1}-(2^{n-1}-1)=1$
$n+1$	0	-	1
$n+2$	0	-	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$

This process, stopped at first win ("heads"), is a transformation of the simple betting process with $X_i = \pm 1$, with

$$\begin{aligned} H_n &= 2^{n-1} \text{ if } X_1 = X_2 = \ldots = X_{n-1} = -1 \\ H_n &= 0 \text{ if } X_i = 1 \text{ for some } i = 1, 2, \ldots, n-1 \end{aligned}$$

ROULETTE: PLAY RED/BLACK OR ODD/EVEN

Requires unbounded capital, and requires that bank sets no upper bound for bets - in order to have a sure win!

EXERCISE 7

Suppose that you use the martingale betting strategy, but that you in any case must stop after n games.

Calculate the expected gain. Can you "beat the game"?

(Hint: You either win 1 or lose $2^n - 1$ *units.)*

EXERCISE 8

Do Exercise 2.6 in book:

Show that the stopped process M^T is a martingale. *(Hint: Find a predictable process H such that $M^T = H \bullet M$).*

This suggests that $E(M_T) \equiv E(M_T^T) = E(M_0^T) = E(M_0) = 0$.

But this does obviously not hold for the martingale strategy case (where indeed the process is a martingale, stopped at a stopping time).

What is the clue here?

state that under certain conditions, the expected value of a martingale at a stopping time, is equal to its initial value, i.e. $E(M_T) = E(M_0)$.

Here are some conditions which, together with the condition $P(T < \infty) = 1$, are each sufficient (but none of which are satisfied in the martingale stratgy case):

- there is a constant d such that $|M_n^T| \leq d$ for all n
- there is a constant τ such that $P(T \leq \tau) = 1$
- $E(T) < \infty$ and there is a constant c such that $|M_n - M_{n-1}| \leq c$ for all n

THE DOOB DECOMPOSITION – EXERCISE 9

Let $X = \{X_0, X_1, X_2, \ldots\}$ be a process with $X_0 = 0$, adapted to history $\{\mathcal{F}_n\}$.

Define process $M = \{M_0, M_1, M_2, \ldots\}$ by $M_0 = 0$ and

$$\Delta M_n \equiv M_n - M_{n-1} = X_n - E(X_n|\mathcal{F}_{n-1})$$

To see that M is a martingale, consider

$$\begin{aligned} E(M_n - M_{n-1}|\mathcal{F}_{n-1}) &= E(X_n - E(X_n|\mathcal{F}_{n-1})|\mathcal{F}_{n-1}) \\ &= E(X_n|\mathcal{F}_{n-1}) - E(X_n|\mathcal{F}_{n-1}) = 0 \end{aligned}$$

The *Doob decomposition* is now

$$X_n = E(X_n|\mathcal{F}_{n-1}) + \Delta M_n$$

X is thus decomposed in a predictable part $\{E(X_n|\mathcal{F}_{n-1})\}$ and a so-called *innovation* part ΔM consisting of martingale differences.

EXERCISE 9

Suppose $X_n = U_1 + \ldots + U_n$ where $U_1, \ldots, U_n$ are independent with $E(U_i) = \mu$.

Find the Doob decomposition of the process X and identify the predictable part and the innovation part.

Key elements:

- $X = \{X(t) : t \in [0, \tau]\}$
- History (information) at time t is given by $\mathcal{F}_t$, with $\mathcal{F}_s \subset \mathcal{F}_t$ when $s < t$. Typically, $\mathcal{F}_t$ corresponds to observation of $X(u)$ for $0 \leq u \leq t$
- $\{\mathcal{F}_t\}$ is called a *filtration*
- The process X is said to be *adapted* to $\{\mathcal{F}_t\}$ if $X(t) \in \mathcal{F}_t$ for all t.
- The process is called *cadlag* if its *paths* (trajectories on a graph) are right continuous with left hand limits
- A time T is a *stopping time* if $\{T \leq t\} \in \mathcal{F}_t$ for each t. This means that at time t we know whether $T \leq t$ or $T > t$

IN CONTINUOUS TIME	IN DISCRETE TIME
$\mathcal{F}_s \subset \mathcal{F}_t$ whenever $s < t$ $M = \{M(t) : t \in [0, \tau]\}$	$\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3 \subset \cdots$ $M = \{M_n : n = 0, 1, \ldots\}$
Definition: $E(M(t)\|\mathcal{F}_s) = M(s)$ for all $t > s$	*Definition:* $E(M_n\|\mathcal{F}_m) = M_m$ for all $n > m$
Property: $\mathrm{Cov}(M(t) - M(s), M(v) - M(u)) = 0$ for $0 \le s < t < u < v \le \tau$	*Property:* $\mathrm{Cov}(M_k - M_j, M_n - M_m) = 0$ for $0 \le j < k \le m < n$

EQUIVALENT DEFINITIONS OF MARTINGALES

- Discrete time:

$$E(\Delta M_n|\mathcal{F}_{n-1}) = E(M_n - M_{n-1}|\mathcal{F}_{n-1}) = 0$$

- Continuous time:

$$E(dM(t)|\mathcal{F}_{t-}) \equiv E(M((t+dt)-) - M(t-)|\mathcal{F}_{t-}) = 0$$

Here we think of $\mathcal{F}_{t-}$ as the history up to, but not including, time t.

Thus $\mathcal{F}_{t-}$ contains the information in $\{M(u) : 0 \leq u < t\}$

Note that $dM(t)$ is defined as the increment of $M(t)$ in the time interval $[t, t+dt)$.

Recall for discrete time: *The transformation of M by H*, $Z = H \bullet M$, is defined for martingale M and *predictable process* H by

$$Z_n = H_1(M_1 - M_0) + H_2(M_2 - M_1) + \ldots + H_n(M_n - M_{n-1}) = \sum_{i=1}^{n} H_i \Delta M_i$$

Continuous time: Need to define *predictable process.* The process $H = \{H(t)\}$ is *predictable* if, informally, the value of $H(t)$ is known immediately before t. A sufficient condition for predictability of H is that it is adapted to $\mathcal{F}_t$ and have *left continuous* sample paths.

The *stochastic integral* of a predictable process H with respect to a martingale M is now,

$$I(t) = \int_0^t H(s)dM(s) =_{\text{def}} \lim_{n\to\infty} \sum_{i=1}^{n} H_i \Delta M_i$$

where $[0, t]$ is partitioned into n parts of length t/n and

- $H_i = H((i-1)t/n), \qquad \Delta M_i = M(it/n) - M((i-1)t/n)$

As for transformations in discrete time: **The $I(t)$ are (mean zero) martingales.**

THE DOOB-MEYER DECOMPOSITION

The process $X = \{X(t) : t \in [0, \tau]\}$ is a *submartingale* if

$$E(X(t)|\mathcal{F}_s) \geq X(s) \text{ for all } t > s$$

Every non-decreasing process (e.g. counting process) must be a submartingale, since $E\{X(t) - X(s)|\mathcal{F}_s\} \geq 0$ for all $t > s$]

Doob-Meyer decomposition for submartingales:

$$X(t) = X^*(t) + M(t) \quad \text{(uniquely)}$$

- X^* is a nondecreasing *predictable* process, called the **compensator** of X
- M is a mean zero martingale

Heuristic proof, suggested by discrete time Doob: Define M and X^* by

$$dX(t) = E(dX(t)|\mathcal{F}_{t-}) + dM(t) = dX^*(t) + dM(t)$$

Then $M(t)$ is a martingale since

$$E(dM(t)|\mathcal{F}_{t-}) = E\{dX(t) - E(dX(t)|\mathcal{F}_{t-})|\mathcal{F}_{t-}\} = 0$$

and $X^*(t)$ is nondecreasing since by the submartingale-property

$$dX^*(t) \equiv E(dX(t)|\mathcal{F}_{t-}) = E(X((t+dt)-) - X(t-)|\mathcal{F}_t-) \geq 0$$

$N(t) = \#$ events in $[0, t]$

Characterizing properties:

- $N(t) - N(s)$ is Poisson-distributed with parameter $\lambda(t - s)$ for $s < t$
- $N(t)$ has independent increments, i.e. $N(t) - N(s)$ is independent of $\mathcal{F}_s$ for $s < t$.

EXERCISE 11

Show that

$$M(t) = N(t) - \lambda t$$

is a martingale. Identify the compensator of $N(t)$.

Hint:

For $t > s$ is $E\{N(t)|\mathcal{F}_s\} = N(s) + \lambda(t - s)$

So far we have mostly looked at expected values. In applications we will also be interested in the variation of the processes:

BACK TO DISCRETE TIME ....

The *predictable* variation process:

$$\langle M \rangle_n = \sum_{i=1}^{n} E\{(M_i - M_{i-1})^2 | \mathcal{F}_{i-1}\} = \sum_{i=1}^{n} \mathrm{Var}(\Delta M_i | \mathcal{F}_{i-1})$$

The *optional* variation process:

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

EXERCISE 12: This is Exercise 2.3 in book, plus a new questions (d). (a) was done as our Exercise 3(a)

Let $X_1, X_2, \ldots$ be independent with $E(X_n) = 0$ and $\text{Var}(X_n) = \sigma^2$ for all n. Let $M_n = X_0 + X_1 + \ldots + X_n$ where $X_0 = 0$.

(a) Show that M_n is a (mean zero) martingale.

(b) Compute $\langle M \rangle_n$

(c) Compute $[M]_n$

(d) Compute $E \langle M \rangle_n$, $E [M]_n$ and $\text{Var}(M_n)$.

EXERCISE 13: Consider a general discrete time martingale M. Show that:

(a) $M_n^2 - \langle M \rangle_n$ is a mean zero martingale - Exercise 2.4 in book

(b) $M_n^2 - [M]_n$ is a mean zero martingale - See book p. 45

(c) Use (a) and (b) to prove that

$$Var(M_n) = E \langle M \rangle_n = E [M]_n$$

What is the intuitive essence of this result?

THE PREDICTABLE VARIATION OF A TRANSFORMATION

Recall:

$$
\begin{aligned}
(H \bullet M)_n &= H_1(M_1 - M_0) + H_2(M_2 - M_1) + \ldots + H_n(M_n - M_{n-1}) \\
\Delta(H \bullet M)_n &= H_n(M_n - M_{n-1}) = H_n \Delta M_n \\
\langle M \rangle_n &= \sum_{i=1}^{n} E\{(M_i - M_{i-1})^2 | \mathcal{F}_{i-1}\} = \sum_{i=1}^{n} \mathrm{Var}(\Delta M_i | \mathcal{F}_{i-1})
\end{aligned}
$$

From this:

$$
\begin{aligned}
\langle H \bullet M \rangle_n &= \sum_{i=1}^{n} \mathrm{Var}(\Delta(H \bullet M)_i | \mathcal{F}_{i-1}) \\
&= \sum_{i=1}^{n} \mathrm{Var}(H_i \Delta M_i | \mathcal{F}_{i-1}) \\
&= \sum_{i=1}^{n} H_i^2 \mathrm{Var}(\Delta M_i | \mathcal{F}_{i-1}) \\
&= \sum_{i=1}^{n} H_i^2 \Delta \langle M \rangle_i \\
&= (H^2 \bullet \langle M \rangle)_n
\end{aligned}
$$

THE OPTIONAL VARIATION OF A TRANSFORMATION

Recall:

$$
\begin{aligned}
(H \bullet M)_n &= H_1(M_1 - M_0) + H_2(M_2 - M_1) + \ldots + H_n(M_n - M_{n-1}) \\
\Delta(H \bullet M)_n &= H_n(M_n - M_{n-1}) = H_n \Delta M_n \\
[M]_n &= \sum_{i=1}^{n} (M_i - M_{i-1})^2 = \sum_{i=1}^{n} (\Delta M_i)^2
\end{aligned}
$$

From this:

$$
\begin{aligned}
[H \bullet M]_n &= \sum_{i=1}^{n} (\Delta(H \bullet M)_i)^2 \\
&= \sum_{i=1}^{n} (H_i \Delta M_i)^2 \\
&= \sum_{i=1}^{n} H_i^2 (\Delta M_i)^2 \\
&= \sum_{i=1}^{n} H_i^2 \Delta [M]_i \\
&= (H^2 \bullet [M])_n
\end{aligned}
$$

RESULTING FORMULAS

$$\begin{aligned} \langle H \bullet M \rangle &= H^2 \bullet \langle M \rangle \\ [H \bullet M] &= H^2 \bullet [M] \end{aligned}$$

with similarities to the elementary formula

$$\mathrm{Var}(aX) = a^2 \mathrm{Var}(X)$$

VARIATION PROCESSES IN CONTINUOUS TIME

The discrete *predictable* variation process

$$\langle M \rangle_n = \sum_{i=1}^{n} \mathrm{Var}(\Delta M_i | \mathcal{F}_{i-1}) \qquad \text{where } \Delta M_i = M_i - M_{i-1}$$

becomes in continuous time

$$\langle M \rangle (t) = \lim_{n \to \infty} \sum_{i=1}^{n} \mathrm{Var}(\Delta M_i | \mathcal{F}_{(i-1)t/n})$$

where

- $[0, t]$ is partitioned into n parts of length t/n
- $\Delta M_i = M(it/n) - M((i-1)t/n)$

Informally:

$d \langle M \rangle (t) = \mathrm{Var}(dM(t) | \mathcal{F}_{t-})$

VARIATION PROCESSES IN CONTINUOUS TIME

The discrete *optional* variation process

$$[M]_n = \sum_{i=1}^{n} (\Delta M_i)^2 \quad \text{where } \Delta M_i = M_i - M_{i-1}$$

becomes

$$[M](t) = \lim_{n\to\infty} \sum_{i=1}^{n} (\Delta M_i)^2$$

where

- $[0, t]$ is partitioned into n parts of length t/n
- $\Delta M_i = M(it/n) - M((i-1)t/n)$

For processes of finite variation (as in our applications), we have

$$[M](t) = \sum_{s\le t} (M(s) - M(s-))^2 \text{ , i.e. sum of squares of all jumps of } M(t)$$

FROM DISCRETE TO CONTINUOUS TIME

IN CONTINUOUS TIME	IN DISCRETE TIME
$M^2(t) - \langle M \rangle (t)$ is a mean zero martingale	$M_n^2 - \langle M \rangle_n$ is a mean zero martingal
$M^2(t) - [M](t)$ is a mean zero martingale	$M_n^2 - [M]_n$ is a mean zero martingal
$Var(M(t)) = E \langle M \rangle (t) = E [M](t)$	$Var(M_n) = E \langle M \rangle_n = E [M]_n$

CHARACTERIZATION OF $\langle M \rangle$ IN CONTINUOUS CASE

First, M^2 is a submartingale because

$$E(M^2(t)|\mathcal{F}_s) \geq (E(M(t)|\mathcal{F}_s))^2 = M(s)^2$$

by Jensen's inequality.

Since $M^2 - \langle M \rangle$ is a martingale, i.e. $M^2 = \langle M \rangle +$ a martingale, it follows by uniqueness of Doob-Meyer decomposition that

- $\langle M \rangle$ is the compensator of M^2

But we also have that $M^2 - [M]$ is a martingale. Why doesn't it follow from this that also $[M]$ is the compensator of M^2?

Recall that $I(t) = \int_0^t H(s)dM(s)$ is a mean zero martingale whenever M is, and is similar to the transformations $H \bullet M$.

Recall for discrete time:

$$\begin{aligned} \langle H \bullet M \rangle &= H^2 \bullet \langle M \rangle \\ [H \bullet M] &= H^2 \bullet [M] \end{aligned}$$

Corresponding formulas for continuous time:

$$\begin{aligned} \left\langle \int HdM \right\rangle &= \int H^2 d\langle M \rangle \\ \left[\int HdM \right] &= \int H^2 d[M] \end{aligned}$$

EXERCISE 14

Consider again the Poisson process, where we have shown that

$$M(t) = N(t) - \lambda t$$

is a martingale.

Prove now that:

$$M^2(t) - \lambda t$$

is a martingale (Exercise 2.9 in book)

Then prove that $\langle M \rangle (t) = \lambda t$

What is $[M](t)$?

Hint:

For $t > s$ is $E(N(t)|\mathcal{F}_s) = N(s) + \lambda(t-s)$

For $t > s$ is $E(N^2(t)|\mathcal{F}_s) = N(s)^2 + 2N(s)\lambda(t-s) + \lambda(t-s) + (\lambda(t-s))^2$

GENERAL COUNTING PROCESS

Recall that $N(t)$ is right continuous, is nondecreasing with jumps of size 1, and is adapted to $\mathcal{F}_t$

Doob-Meyer decomposition:

$$N(t) = \Lambda(t) + M(t) \tag{5}$$

where the compensator Λ is a (unique) nondecreasing, predictable process. ($\Lambda(t) = \lambda t$ for the Poisson process, i.e. a determinisitic function)

Assume we can represent $\Lambda(t) = \int_0^t \lambda(s)ds$.

Then $\lambda(t)$ is called the *intensity process*, while $\Lambda(t)$ is the *cumulative* intensity process.

Classical formulation of counting process martingale:

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

From

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

we have

$$dM(t) = dN(t) - \lambda(t)dt$$

which implies since $\lambda(t)$ is predictable and M is a martingale,

$$0 = E(dM(t)|\mathcal{F}_{t-}) = E(dN(t)|\mathcal{F}_{t-}) - \lambda(t)dt$$

so that we have the important connection

$$\lambda(t)dt = E(dN(t)|\mathcal{F}_{t-})$$

Further, since steps are of size 1,

$$\lambda(t)dt = E(dN(t)|\mathcal{F}_{t-}) = P(dN(t) = 1|\mathcal{F}_{t-}) = P(\text{event in } [t, t + dt)|\mathcal{F}_{t-})$$

$$[M](t) = N(t) = \text{ process itself}$$

since we have seen that for processes of finite variation, we have

$$[M](t) = \sum_{s \le t} (M(s) - M(s-))^2 \text{ , i.e. sum of squared jumps of } M(t)$$

Furthermore,

$$\begin{aligned} d\langle M\rangle(t) &= \mathrm{Var}(dM(t)|\mathcal{F}_{t-}) \\ &= \mathrm{Var}(dN(t) - \lambda(t)|\mathcal{F}_{t-}) \\ &= \mathrm{Var}(dN(t)|\mathcal{F}_{t-}) \\ &= \lambda(t)dt(1 - \lambda(t)dt) \\ &\approx \lambda(t)dt \end{aligned}$$

Here we used that $dN(t)$ conditional on $\mathcal{F}_{t-}$ is a $0-1$-variable with expected value $\lambda(t)dt$.

Hence

$$\langle M\rangle(t) = \int_0^t \lambda(s)ds = \Lambda(t)$$

STOCHASTIC INTEGRAL OF GENERAL COUNTING PROCESS

Consider a general counting process $N(t)$ and recall the decompositiom

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

Let further $T_1 < T_2 < \cdots$ be the jump times of $N(t)$.

Then

$$\begin{aligned} I(t) &= \int_0^t H(s)dM(s) \\ &= \int_0^t H(s)dN(s) - \int_0^t H(s)\lambda(s)ds \\ &= \sum_{T_j \le t} H(T_j) - \int_0^t H(s)\lambda(s)ds \end{aligned}$$

VARIATION PROCESSES OF STOCHASTIC INTEGRALS OF COUNTING PROCESS MARTINGALES

Recall first general facts:

$$\begin{aligned}\left\langle \int HdM \right\rangle &= \int H^2 d\langle M\rangle \\ \left[\int HdM\right] &= \int H^2 d[M]\end{aligned}$$

and the special facts for counting processes

$$\langle M\rangle(t) = \int_0^t \lambda(s)ds$$

$$[M](t) = N(t)$$

From this we get the formulas:

$$\begin{aligned}\left\langle \int HdM \right\rangle(t) &= \int_0^t H^2(s)\,\lambda(s)ds \\ \left[\int HdM\right](t) &= \int_0^t H^2(s)\,dN(s)\end{aligned}$$

EXAMPLE - THE MULTIPLICATIVE INTENSITY MODEL

Counting process $N(t)$ has intensity function

$$\lambda(t) = \alpha(t)Y(t)$$

where $\alpha(t)$ is a deterministic function representing the risk for each unit under study, and $Y(t)$ is the predictable process counting the number of units which are present immediately before t and hence may fail at time t.

We have

$$N(t) = \int_0^t \lambda(s)ds + M(t)$$

so we can write

$$dN(s) = \lambda(s)ds + dM(s) = \alpha(s)Y(s) + dM(s)$$

We want to estimate $\alpha(t)$, but it is easier to estimate its integral $A(t) = \int_0^t \alpha(s)ds$, which we will do.

Assume for simplicity that we know that $Y(t) > 0$ for all t.

Then - dividing through the equation by $Y(t)$ we get

$$\frac{1}{Y(s)}dN(s) = \alpha(s)ds + \frac{1}{Y(s)}dM(s)$$

THE MULTIPLICATIVE INTENSITY MODEL (cont.)

$$\frac{1}{Y(s)}dN(s) = \alpha(s)ds + \frac{1}{Y(s)}dM(s)$$

Integrating we get that

$$\int_0^t \frac{1}{Y(s)}dN(s) = A(t) + \int_0^t \frac{1}{Y(s)}dM(s)$$

The rightmost expression is a mean zero integral (since $Y(s)$ is a predictable process), so this suggests the (Nelson-Aalen) estimator

$$\hat{A}(t) = \int_0^t \frac{1}{Y(s)}dN(s) = \sum_{T_j \le t} \frac{1}{Y(T_j)}$$

PROPERTIES OF THE ESTIMATOR

We have

$$\hat{A}(t) - A(t) = \int_0^t \frac{1}{Y(s)} dM(s)$$

and hence $\hat{A}(t)$ is an unbiased estimator (since the right hand side is a mean zero martingale). Further (see page 54),

$$\left\langle \hat{A} - A \right\rangle (t) = \int_0^t \frac{1}{Y^2(s)} \alpha(s) Y(s) ds = \int_0^t \frac{1}{Y(s)} \alpha(s) ds$$

$$\left[\hat{A} - A \right] (t) = \int_0^t \frac{1}{Y^2(s)} dN(s) = \sum_{T_j \leq t} \frac{1}{Y^2(T_j)} \tag{6}$$

We also know from our earlier formulas that

$$Var(\hat{A}(t)) = \mathrm{Var}(\hat{A}(t) - A(t)) = E\left[\hat{A}(t) - A(t)\right]$$

so that (6) is an unbiased estimator of $Var(\hat{A}(t))$

ASYMPTOTIC DISTRIBUTION OF NELSON-AALEN ESTIMATOR

Suppose that our data were based on observation of a large number n units. We shall let n tend to infinity in the relation

$$\sqrt{n}(\hat{A}(t) - A(t)) = \int_0^t \sqrt{n} \frac{1}{Y(s)} dM(s)$$

and hence we need to look at the limiting behaviour of the mean zero martingale on the right hand side.

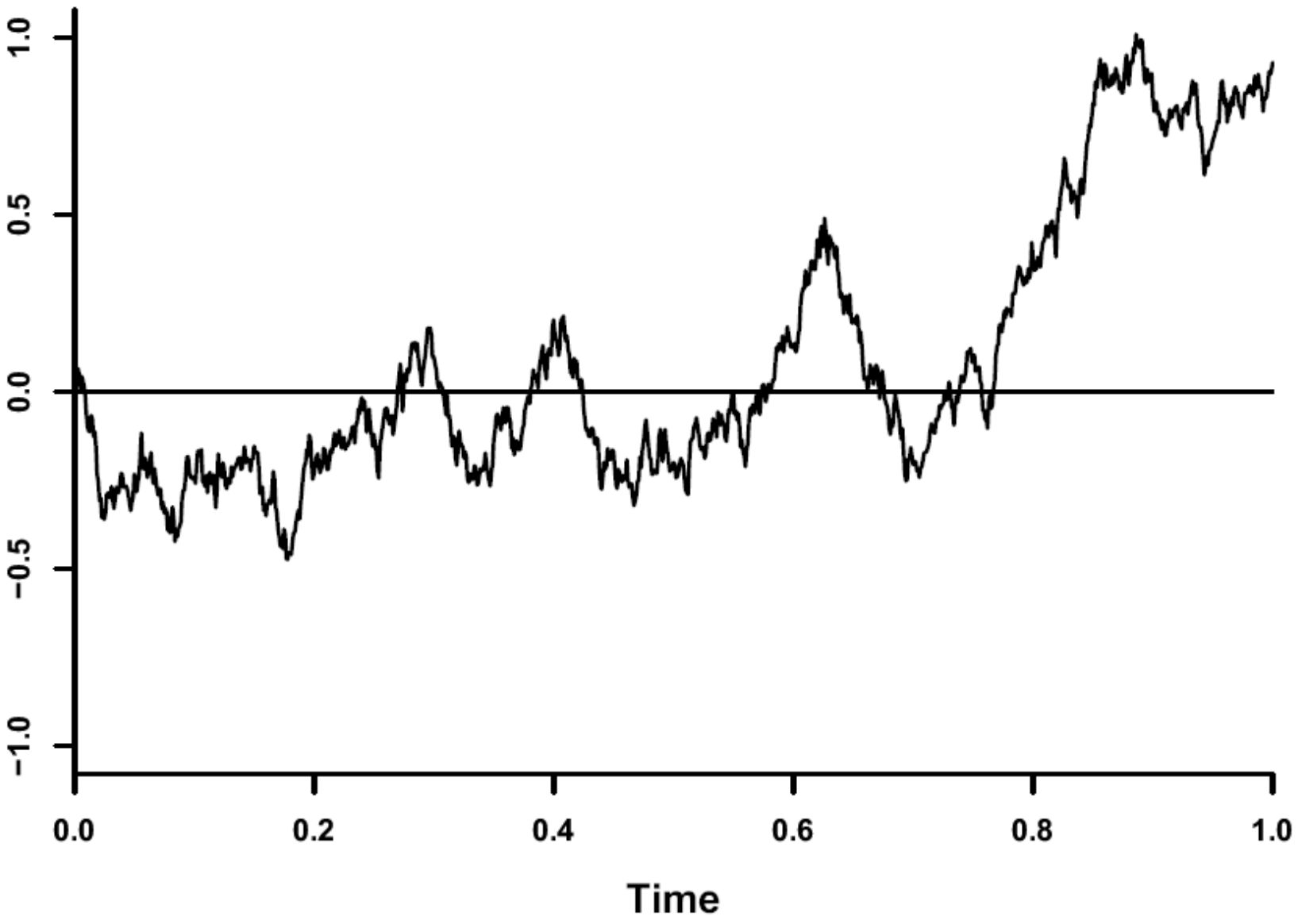
This brings us to the need for an asymptotic theory for martingales.

WIENER PROCESS

A Wiener process (Brownian motion) with variance parameter 1 is a stochastic process $W(t)$ with values in the real numbers satisfying

1. $W(0) = 0$
2. For any $s_1 \leq t_1 \leq s_2 \leq t_2 \leq \cdots \leq s_n \leq t_n$, the random variables $W(t_1) - W(s_1), \ldots, W(t_n) - W(s_n)$ are independent.
3. For any $s < t$, the random variables $W(t) - W(s)$ are normal with expected value 0 and variance $(t - s)$
4. The paths are continuous.

WIENER PROCESS



Let $V(t)$ be a strictly increasing continuous function with $V(0) = 0$ and consider the stochastic process

$$U(t) = W(V(t))$$

(i.e. a time transformation of $W(t)$.)

The process $U(t)$

- is a mean zero martingale
- has predictable variation process $\langle U \rangle (t) = V(t)$

(exercise 2.12 in book).

U is called a Gaussian martingale.

REBOLLEDO'S MARTINGALE CONVERGENCE THEOREM

Let $\tilde{M}^{(n)}(t)$ be a sequence of mean zero martingales for $t \in [0, \tau]$.

Define

$$\tilde{M}_{\epsilon}^{(n)}(t) = \sum_{s \le t \,:\, |\tilde{M}^{(n)}(s) - \tilde{M}^{(n)}(s-)| \ge \epsilon} |\tilde{M}^{(n)}(s) - \tilde{M}^{(n)}(s-)|$$

to be the sum of the jumps larger than ϵ of the martingale $\tilde{M}^{(n)}(t)$.

Then under the conditions:

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

$\left\langle \tilde{M}_{\epsilon}^{(n)}(t) \right\rangle \to 0$ in probability for all $t \in [0, \tau]$ and all $\epsilon > 0$ as $n \to \infty$,

the sequence $\tilde{M}^{(n)}(t)$ converges in distribution to the mean zero Gaussian martingale U given by $U(t) = W(V(t))$.

As seen earlier, we consider stochastic integrals with predictable variation prosesses

$$\left\langle \int HdM \right\rangle (t) = \int_0^t H^2(s)\ \lambda(s)ds$$

Rebolledo's conditions can then be written

$$\int_0^t (H^{(n)}(s))^2 \lambda^{(n)}(s)ds \to V(t) \text{ for all } t \in [0,\tau]$$

$$\int_0^t (H^{(n)}(s))^2 I\{|H^{(n)}(s)| > \epsilon\}\lambda^{(n)}(s)ds \to 0 \text{ for all } t \in [0,\tau] \text{ and all } \epsilon > 0$$

Sufficent conditions for these are, with $V(t) = \int_0^t v(s)ds$,

$$(H^{(n)}(s))^2 \lambda^{(n)}(s) \to v(s) > 0 \text{ for all } s \in [0,\tau]$$

$$H^{(n)}(s) \to 0 \text{ for all } s \in [0,\tau]$$

Recall that

$$\sqrt{n}(\hat{A}(t) - A(t)) = \int_0^t \frac{\sqrt{n}}{Y(s)} dM(s)$$

so that we have

$$H^{(n)}(t) = \frac{\sqrt{n}}{Y(t)}$$

Assume that there is a deterministic positive function $y(t)$ such that $Y(t)/n \to y(t)$ in probability. Then the two sufficient conditions for Rebolledo's theorem are satisfied:

$$(H^{(n)}(s))^2 \lambda^{(n)}(s) = \frac{n}{Y^2(s)} \cdot \alpha(s) Y(s) = \frac{\alpha(s)}{Y(s)/n} \to \frac{\alpha(s)}{y(s)} \equiv v(s)$$

$$H^{(n)}(s) = \frac{\sqrt{n}}{Y(s)} = \frac{1/\sqrt{n}}{Y(s)/n} \to 0$$

Conclusion:

$$\sqrt{n}(\hat{A}(t) - A(t)) = \int_0^t \frac{\sqrt{n}}{Y(s)} dM(s)$$

converges in distribution to a mean zero Gaussian martingale with predictable variation process

$$V(t) = \int_0^t v(s)ds = \int_0^t \frac{\alpha(s)}{y(s)} ds$$

In particular, for each t is $\hat{A}(t)$ asymptotically normally distributed, and

$$\mathrm{Var}(\hat{A}(t)) \approx \frac{1}{n} \cdot \int_0^t \frac{\alpha(s)}{Y(s)/n} ds = \int_0^t \frac{\alpha(s)}{Y(s)} ds (= \left\langle \hat{A} - A \right\rangle (t))$$

THE INNOVATION THEOREM (Section 2.2.7)

We have defined the intensity of a counting process N with respect to a filtration $\mathcal{F}_t$ by

$$\lambda^{\mathcal{F}}(t) = E(dN(t)|\mathcal{F}_{t-})$$

We have earlier tacitly assumed that the underlying filtration is given.

For example, we have often assumed the so-called *self-exiting* filtration $\mathcal{N}_t$ generated by the history of $N(t)$ alone.

There may also be other histories which contain more information, and with respect to which we get a different intensity.

Recall the general result, for any $Y \in \mathcal{F}$,

$$E(Y|\mathcal{F}_{m_1}) = E[E(Y|\mathcal{F}_{m_2})|\mathcal{F}_{m_1}] \text{ for } 1 \leq m_1 < m_2 < n$$

Using this we have, if $\mathcal{F}_t \subseteq \mathcal{G}_t$ (i.e. $\mathcal{G}$ contains more information than $\mathcal{F}$), that

$$E(dN(t)|\mathcal{F}_{t-}) = E[E(dN(t)|\mathcal{G}_{t-})|\mathcal{F}_{t-}]$$

so it follows that

$$\lambda^{\mathcal{F}}(t) = E(\lambda^{\mathcal{G}}(t)|\mathcal{F}_{t-})$$

This is the *innovation* theorem.

INDEPENDENT CENSORING (Section 2.2.8 in book)

Let $N_i^c(t)$ be counting process for individual i, when censoring is not present ($i = 1, 2, \ldots, n$), with intensity process

$$\lambda_i^c(t) = Y_i^c(t)\alpha_i(t)$$

These processes are adapted to the history $\{\mathcal{F}_t^c\}$, say.

Let *censoring* be represented as $0-1$-processes $Y_i^0(t)$ such that the process N_i^c is observed only when $Y_i^0(t) = 1$ (e.g. $Y_i^0(t) = I(t \leq C_i)$ for censoring times C_i).

Observed process is then $N_i(t) = \int_0^t Y_i^0(s)dN_i^c(s)$ which jumps when $N_i^c(t)$ jumps, but only if $Y_i^0(t) = 1$ at the jump.

Let $\mathcal{G}_t$ be the history $\mathcal{F}_t^c$ augmented by the information in all the $Y_i^0(s)$ for $s \leq t$.

Independent censoring means by definition that the intensity of the counting process $N_i^c(t)$ is the same for the two histories $\mathcal{G}$ and $\mathcal{F}^c$, i.e.

$$\lambda_i^{\mathcal{G}}(t) = \lambda_i^{\mathcal{F}^c}(t) \text{ for all } t$$

INDEPENDENT CENSORING (cont.)

What is the intensity of the observed process $N_i(t)$, with respect to its filtration $\mathcal{F}_t$? (Which is smaller than both $\mathcal{F}^c$ and $\mathcal{G}$)

Note that $Y_i(t) = Y_i^0(t)Y_i^c(t)$ is adapted to $\mathcal{F}$ (why?), which will be used below:

$$
\begin{aligned}
\lambda_i^{\mathcal{F}}(t)dt &= E\{dN_i(t)|\mathcal{F}_{t-}\} \\
&= E\{Y_i^0(t)dN_i^c(t)|\mathcal{F}_{t-}\} \\
&= E\{Y_i^0(t)E(dN_i^c(t)|\mathcal{G}_{t-})|\mathcal{F}_{t-}\} \\
&= E\{Y_i^0(t)\lambda_i^{\mathcal{G}}(t)dt|\mathcal{F}_{t-}\} \\
&= \textit{assuming independent censoring} \\
&= E\{Y_i^0(t)\lambda_i^{\mathcal{F}^c}(t)dt|\mathcal{F}_{t-}\} \\
&= E\{Y_i^0(t)Y_i^c(t)\alpha_i(t)dt|\mathcal{F}_{t-}\} \\
&= Y_i(t)E\{\alpha_i(t)|\mathcal{F}_{t-}\}dt \\
&= Y_i(t)\alpha_i(t)dt
\end{aligned}
$$

where we assume that $\alpha_i(t)$ is adapted to $\mathcal{F}_{t-}$.

This shows that the form of the intensity is preserved under independent

PREDICTABLE COVARIATION PROCESSES

Recall predictable variation process:

$$d\langle M\rangle(t) = \mathrm{Var}(dM(t)|\mathcal{F}_{t-})$$

where $dM(t) = M((t+dt)-) - M(t-)$

Similarly we can define the predictable covariation process

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

From this

$$\begin{aligned} d\langle M_1 + M_2\rangle(t) &= \mathrm{Var}(dM_1(t) + dM_2(t)|\mathcal{F}_{t-}) \\ &= \mathrm{Var}(dM_1(t)|\mathcal{F}_{t-}) + \mathrm{Var}(dM_2(t)|\mathcal{F}_{t-}) \\ &+ 2\mathrm{Cov}(dM_1(t), dM_2(t)|\mathcal{F}_{t-}) \\ &= d\langle M_1\rangle(t) + d\langle M_2\rangle(t) + 2d\langle M_1, M_2\rangle(t) \end{aligned}$$

Hence

$$\langle M_1 + M_2\rangle = \langle M_1\rangle + \langle M_2\rangle + 2\langle M_1, M_2\rangle$$

OPTIONAL COVARIATION PROCESSES

Recall optional variation process:

$$[M](t) = \lim_{n\to\infty} \sum_{i=1}^{n} (\Delta M_i)^2$$

where $\Delta M_i = M(it/n) - M((i-1)t/n)$

Similarly we can define the optional covariation process

$$[M_1, M_2](t) = \lim_{n\to\infty} \sum_{i=1}^{n} \Delta M_{1i} \Delta M_{2i}$$

Also:

$$[M_1 + M_2] = [M_1] + [M_2] + 2[M_1, M_2]$$

Exercise 2.5 is to show:

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

$$M_1 M_2 - [M_1, M_2 \text{ is a mean zero martingale}$$

TWO COUNTING PROCESSES

Let $N_1(t)$ and $N_2(t)$ be counting processes.

Assume that they cannot jump simultaneously.

EXERCISE 15

Do exercise 2.10 in book: Show that

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

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

i.e. M_1 and M_2 are *orthogonal* martingales.

Hint: Argue that $N_1 + N_2$ is a counting process with

$$\langle M_1 + M_2 \rangle = \langle M_1 \rangle + \langle M_2 \rangle$$

$$[M_1 + M_2] = [M_1] + [M_2]$$