Introduction to stochastic analysis

Dario Gasbarra

2010-02-03

1 Introduction

Let x_t and h_t measurable functions $\mathbb{R}^+ \mapsto \mathbb{R}$, where x_t has finite variation and h_t is bounded on every compact interval.

A function of finite variation has a representation

$$x_t = x_0 + x_t^{\oplus} - x_t^{\ominus},$$

where $x_t^{\oplus}, x_t^{\ominus}$ are non-decreasing functions with $x_0^{\oplus} = x_0^{\ominus} = 0$. We can always choose a representation where the corresponding measures $x^{\oplus}(dt), x^{\ominus}(dt)$ are mutually singular. Then, the variation of the function x over the interval $[0, t]$ is defined as

$$v_t(x) := x_t^{\oplus} + x_t^{\ominus}$$

For example when x_t has almost everywhere a derivative $\dot{x}_t$,

$$x_t^{\oplus} = \int_0^t (\dot{x}_s)^+ ds, \; x_t^{\ominus} = \int_0^t (\dot{x}_s)^- ds \;\text{ and }\; v_t(x) = \int_0^t |\dot{x}_s| ds$$

where $x^{\pm} := \max(\pm x, 0)$.

We have learned from the Probability Theory or Real Analysis courses that in such case the integral

$$I_t = \int_0^t h_s dx_s$$

is well defined as a Lebesgue Stieltjes integral. When the integrand h is piecewise continuous or it has finite variation this is a Riemann Stieltjes integral defined as limit of Riemann sums.

$$I_t = \lim_{\Delta(\Pi) \to 0} \sum_i h_{s_i} (x_{t_{i+1}} - x_{t_i})$$

where $\Pi = \big\{ 0 = t_0 \le s_0 \le t_1 \le s_1 \le t_2 \le \dots \le t_{n-1} \le s_n \le t_n = t \big\}$ is a partition of $[0, t]$ and $\Delta(\Pi) := \max_{i \le n}(t_i - t_{i-1})$

This Riemann-Stieltjes integral does not depend on the sequence of partitions and the choice of the middle point.

In 1900, Luis Bachelier in his Ph.D. thesis *Theorie de la speculation* invented a new probabilistic model to descibe the behaviour of the stock exchange in Paris. This is a stochastic process $(B_t(\omega))_{t \in \mathbb{R}^+}$, defined in continuous time as follows:

Definition 1. *1. For $0 \leq s \leq t$, the increments $(B_t(\omega) - B_s(\omega))$ are stochatically independent over disjoint intervals, and have gaussian distribution with* 0 *mean and variance* $(t-s)$.

2. for (P-almost) all ω the trajectory $t \mapsto B_t(\omega)$ is continuous.

In 1905 Einstein introduced independently the very same mathematical model and results to explain the thermal motion of pollen particles suspended in a liquid, which haad been observed by the botanist Brown.

Unfortunately, the importance of the work of Bachelier was not recognized at his times, so that B_t is called *Brownian motion* or *Wiener process*, after Norbert Wiener who started the theory of stochastic integration. In textbooks it is also denoted by W_t. In honour of Bachelier we like to use the B_t notation.

In fact, although Kolmogorov (1933) showed that the paths $B_t(\omega)$ are almosty surely Hölder continuous that is the random quantity

$$\sup\left\{ \frac{|B_t(\omega) - B_s(\omega)|}{|t-s|^\alpha} : 0 \leq s, t, \leq T,\ s \neq t \right\} < \infty \quad P - \text{almost surely}$$

for all $0 < \alpha < 1/2$ in every compact $[0.T]$, with probability 1 the paths are nowhere differentiable and have infinite variation.

For integrand paths $h_s(\omega)$ of finite variation using the integration by parts formula we define for every ω

$$\int_0^t h_s(\omega) dB_t(\omega) := B_t(\omega) h_t(\omega) - h_0(\omega) B_0(\omega) - \int_0^t B_s(\omega) dh_s(\omega)$$

This trick does not work for the integral

$$\int_0^t B_s(\omega) dB_s(\omega)$$

It was in 1944 that Kyoshi Ito extended Wiener integral to the class of *non-anticipative* integrand processes. This was the beginning of modern stochastic analysis.

For the history, in 1940 the german-french mathematician Wolfgang Doeblin fighting on the french side was surrounded by the nazis and, before commiting suicide, sent to the french academy of sciences a letter to be opened 60 years later. This letter, published in year 2000, contained many of the ideas on stochastic differential equations that Ito was developing.

In Ito calculus we have the change of variable formula

Theorem 1. *for $f \in C^2$*

$$f(B_t) = f(B_0) + \int_0^t f'(B_s) dB_s + \frac{1}{2} \int_0^t f''(B_s) ds$$

We give a sketch of the proof (later we will generalize it by using martingale theory) by using telescopic sum and Taylor expansion.

For $0 = t_0 < t_1 < \cdots < t_n = t$

$$f(B_t) = f(B_0) + \sum_i^n \big(f(B_{t_i}) - f(B_{t_{i-1}}) \big)$$

$$= f(B_0) + \sum_i^n f'(B_{t_{i-1}})(B_{t_i} - B_{t_{i-1}}) + \frac{1}{2} \sum_i^n f''(B_{S_{i-1}})(B_{t_i} - B_{t_{i-1}})^2 =$$

$$f(B_0) + \sum_i^n f'(B_{t_{i-1}})(B_{t_i} - B_{t_{i-1}}) + \frac{1}{2} \sum_i^n f''(B_{t_{i-1}})(B_{t_i} - B_{t_{i-1}})^2 + R_n$$

where $S_{i-1}(\omega) \in [t_{i-1}, t_i]$ is a random time depending on the trajectory $B_t(\omega)$ and the random remainder is

$$R_n(\omega) = \frac{1}{2} \sum_i^n (f''(B_{S_{i-1}}) - f''(B_{t_{i-1}}))(B_{t_i} - B_{t_{i-1}})^2$$

For almost all ω, the composition of continuous functions $s \mapsto f''(B_s(\omega))$ is continuous, and it is uniformly continuous on compact intervals $[0, t]$. Therefore for all $\varepsilon > 0$ there is a random $\delta(\omega) > 0$ such that

$$|f''(B_s(\omega)) - f''(B_r(\omega))| < \varepsilon$$

for $s, r \in [0, t]$ with $|s - r| < \delta(\omega)$. In particular,

$$|R_n(\omega)| \le \varepsilon \sum_i^n (B_{t_i} - B_{t_{i-1}})^2$$

when $\Delta(\Pi) < \delta(\omega)$. We use the following result:

Lemma 1. *As the step-size of the partition $\Delta \to 0$,*

$$\left(\sum_i^n (B_{t_i} - B_{t_{i-1}})^2 \right) \stackrel{L^2(P)}{\rightarrow} t := \int_0^t dB_s dB_s \ , \qquad (1)$$

which implies convergence in probability. This limit is called quadratic variation and denoted by the square bracket $[B, B]_t = [B]_t$

It follows that $|R_n(\omega)| \stackrel{P}{\rightarrow} 0$ as $\Delta(\Pi) \to 0$. We have also

$$\frac{1}{2} \sum_i^n f''(B_{t_{i-1}})(B_{t_i} - B_{t_{i-1}})^2 \stackrel{P}{\rightarrow} \int_0^t f''(B_s) ds$$

where the integral on the right hand side is defined for almost every ω as a Riemann-Stieltjes integral. This follows directly when $f''(x)$ is piecewise constant and otherwise we can approximate $f(x)$ by a sequence of functions $f^{(n)}(x)$ with piecewise constant second derivatives converging uniformly on compacts.

This means that sequence of Riemann sums converges in probability to the Ito's stochastic integral

$$\lim_{\Delta(\Pi) \to 0} \sum_i f'(B_{t_i})(B_{t_{i+1}} - B_{t_i}) \stackrel{P}{\rightarrow} \int_0^t f'(B_s) dB_s = f(B_t) - f(B_0) - \frac{1}{2} \int_0^t f''(B_s) ds \qquad (2)$$

Note that for a continuous path x_t with finite variation $v_t(x)$ on the interval $[0,t]$ we have

$$\sum_i^n (x_{t_i} - x_{t_{i-1}})^2 \le v_t(x) \max_{i=1,\dots,n} |x_{t_i} - x_{t_{i-1}}| \to 0 \quad \text{as} \quad \Delta(\Pi) \to 0$$

since x_s is uniformly continuous on $[0,t]$ so that the second order term in Ito's formula vanishes, giving

$$f(x_t) = f(x_0) + \int_0^t f'(x_s)dx_s$$

Proof of Lemma 1 By definition of Brownian motion the increments $\Delta B_i = (B_{t_i} - B_{t_{i-1}})$ are independent Gaussian with zero mean and variance $\Delta t = (t_i - t_{i-1})$. Since from gaussianity it follows $E\big((\Delta B_i)^4\big) = 3(\Delta t_i)^2$, using independence,

$$E\bigg(\Big\{t - \sum_i (\Delta B_i)^2\Big\}^2\bigg) E\bigg(\Big\{\sum_i \Big((\Delta B_i)^2 - \Delta t_i\Big)\Big\}^2\bigg) =$$
$$\sum_i E\bigg(\Big\{(\Delta B_i)^2 \Delta t_i\Big\}^2\bigg) + 2\sum_{j<i} E\bigg(\Big\{(\Delta B_i)^2 - \Delta t_i\Big\}\Big\{(\Delta B_j^2 - \Delta t_j\Big\}\bigg) =$$
$$\sum_i E\Big((\Delta B_i)^4 + (\Delta t_i)^2 - 2(\Delta B_i)^2 \Delta t_i\Big) + 2\sum_{j<i} E\Big((\Delta B_i)^2 - \Delta t_i\Big) E\Big((\Delta B_j^2 - \Delta t_j\Big) =$$
$$\sum_i \Big(3(\Delta t_i)^2 + (\Delta t_i)^2 - 2\Delta t_i\Big) = 2\sum_i (\Delta t_i)^2 \le \max_i \Delta t_i \sum_i \Delta t_i = t\Delta(\Pi) \to 0$$

From this convergence in quadratic mean follows convergence in probability by using Chebychev inequality.

If the sequence of partitions is refining, meaning that $\Pi_n \supseteq \Pi_{n-1}$, we can use the martingale convergence theorem to replace convergence in probability with the stronger almost sure convergence in the definitions of the quadratic variation and Ito integral $\left(\int_0^t f(B_s)dB_s\right)$. We give a simple proof for the special case of dyadic partitions.

Lemma 2. *Consider the dyadic partitions $D_n := (t_i^{(n)} = i2^{-n}t : i = 0, 1, \dots, 2^n)$. For the sequence $\Pi_n = D_n$ the limiting relation 1 holds also for P-almost sure convergence.*

Proof since $\Delta t_i^{(n)} = t_i^{(n)} - t_{i-1}^{(n)} = 2^{-n}$,

$$E\bigg(\Big\{t - \sum_{i \in D_n} (\Delta B_i^{(n)})^2\Big\}^2\bigg) = 2 \cdot 2^{-n} t$$

Let $\varepsilon > 0$ and

$$A_{n,\varepsilon} = \bigg\{\omega : \Big|t - \sum_{i \in D_n} (\Delta B_i^{(n)})^2\Big| > \varepsilon\bigg\}$$

Since from Chebychev's inequality

$$P(A_{n,\varepsilon}) \leq 2t2^{-n}\varepsilon^{-2}$$

we have

$$\sum_n P(A_{n,\varepsilon}) \leq \varepsilon^{-2} 4t < \infty$$

From Borel Cantelli lemma it follows $\forall \varepsilon > 0$

$$P\big(\limsup_n A_{n,\varepsilon}\big) = 0$$

This means almost sure convergence, since taking countable union over $\varepsilon = m^{-1}$ and complement, we obtain

$$P\bigg(\bigcup_{m\in\mathbb{N}} \bigcap_{k\in\mathbb{N}} \bigcup_{n\geq k} A_n\bigg) = 0 \Longleftrightarrow 1 = P\bigg(\bigcap_{m\in\mathbb{N}} \bigcup_{k\in\mathbb{N}} \bigcap_{n\geq k} A_n^c\bigg) = P\bigg(\lim_{n\to\infty} \sum_{i\in D_n} (\Delta B_i^{(n)})^2 = t\bigg) \square$$

Remark It is not wrong to say that the Brownian path $(B_t(\omega) : t \in [0,1])$ corresponds to an infinite dimensional random vector uniformly distributed on the surface of the infinite dimensional unit sphere.

2 Kolmogorov's extension

We prove first Daniell-Kolmogorov extension theorem which tells when a stochastic process (X_t) indexed by a time parameter $t \in T$ exists as collection of random variables.

Whether this collection of random variables can be combined together into a random path with some continuity properties with respect to the parameter, is the content of Kolmogorov's continuity theorem.

Definition 2. *Let $(\Omega, \mathcal{F}, P)$ be a probability triple. A stochastic process is a collection of random variables $(X_t(\omega))_{t\in T}$ with values in $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)$ with parameter set T.*

In these lectures we will consider $T = \mathbb{N}, \mathbb{Z}, \mathbb{R}, \mathbb{R}^+, \mathbb{Q}$ but some other index sets may appear.

Definition 3. *Let $X = (X_t(\omega))_{t\in T}$ and $X' = (X'_t(\omega))_{t\in T}$ $\mathbb{R}$-valued stochastic processes on the respective probability spaces $(\Omega, \mathcal{F}, P)$ and $(\Omega', \mathcal{F}', P')$. We say that X and X' are versions the same process if their finite dimensional laws coincide: $\forall k \in \mathbb{N}, t_1 \dots t_k \in T$ $B_1, \dots B_k \in \mathcal{B}(\mathbb{R}^d)$*

$$P\Big(X_{t_1} \in B_1, \dots, X_{t_k} \in B_k\Big) = P'\Big(X'_{t_1} \in B_1, \dots, X'_{t_k} \in B_k\Big)$$

Definition 4. *Let $X = (X_t(\omega))_{t\in T}$ and $Y = (Y_t(\omega))_{t\in T}$ $\mathbb{R}$-valued stochastic processes on the same probability space $(\Omega, \mathcal{F}, P)$ We say that X and Y are modifications of each other if $\forall t \in T$*

$$P\big(X_t = Y_t\big) = 1$$

Definition 5. *Let $X = (X_t(\omega))_{t\in T}$ and $Y = (Y_t(\omega))_{t\in T}$ $\mathbb{R}$-valued stochastic processes on the same probability space $(\Omega, \mathcal{F}, P)$ We say that X and Y are indistinguishable when*

$$P\big(\omega : X_t(\omega) = Y_t(\omega)\ \forall t \in T\big) = 1$$

Exercise 1. *When X and Y are indistinguishable, they are modification of each other. When X and Y are each others' modifications, they share the same finite dimensional laws. Show a simple example of a X, Y which are modfication of each other but not indistinguishable.*

Definition 6. *We say that the family of finite dimensional distributions*

$$P_{t_1,\dots,t_n} : \mathcal{B}(\mathbb{R}^n) \to [0,1], \qquad \text{with } n \in \mathbb{N}, t_1, \dots, t_n \in T$$

is consistent , when

- $$P_{t_1,\dots,t_n}(A_1 \times \dots \times A_n) = P_{t_{\pi(1)},\dots t_{\pi(n)}}\big(A_{t_{\pi(1)}} \dots \times A_{t_{\pi(n)}}\big)$$
  $$\forall n \in \mathbb{N}, A_1, \dots A_n \in \mathcal{B}(\mathbb{R}), t_1, \dots, t_n \in T, \quad \forall \text{ permutation } \pi$$

- $$P_{t_1,\dots,t_n}(A_1 \times \cdots \times A_n) = P_{t_1,\dots,t_n,t_{n+1}}(A_1 \times \cdots \times A_n, \mathbb{R})$$

Theorem 2. *(Daniell-Kolmogorov,1933) Let*

$$\left(P_{\mathbf{t}} : \mathbf{t} \in \bigcup_{n=1}^{\infty} T^n \right)$$

a consistent family of finite dimensional probability distributions with arbitrary index set T.

There exist a unique probability measure $\mathbf{P}$ on the product space $\Omega = \mathbb{R}^T$ equipped with the cylinder σ-algebra generated by the product topology, such that $\forall n \in \mathbb{N}, t_1, \dots, t_n \in \mathbb{N}, B_n \in \mathcal{B}(\mathbb{R}^n)$,

$$\mathbf{P}\left(\omega \in \mathbb{R}^T : \ (\omega_{t_1}, \dots, \omega_{t_n}) \in B_n \right) = P_{t_1,\dots,t_n}(B_n) \tag{3}$$

Proof

The elements of $\Omega = \mathbb{R}^T$ are functions $t \mapsto \omega_t$. $\sigma(\mathcal{C})$ coincides with the smallest σ-algebra on $\Omega = R^T$ which makes the canonical evalutions $\omega \mapsto X_t(\omega) = \omega_t$ measurable for all $t \in T$.

We define the cylinders' algebra $\mathcal{C}$ with typical elements

$$C = \left\{ \omega \in \mathbb{R}^T : (\omega_{t_1}, \dots, \omega_{t_n}) \in B_n \right\}$$

where $n \in \mathbb{N}, t_1, \dots, t_n \in \mathbb{N}, B_n \in \mathcal{B}(\mathbb{R}^n)$.

We take (3) as a definition of the map $\mathbf{P} : \mathcal{C} \to [0, 1]$.

By using the consistency assumption you can check that $\mathbf{P}(C)$ does not depend on the particular representation of a cylinder $C \in \mathcal{C}$.

Since every finite number of cylinders can be represented on a common index set, since the finite dimensional distributions are proabilities, it is also not difficult to check that $\mathbf{P}$ is finitely additive on $\mathcal{C}$.

The next step is to use Charatheodory's theorem to extend $\mathbf{P}$ to a σ-additive probability measure defined on the σ-algebra $\sigma(C)$.

All we need to show is that $\mathbf{P}$ is σ-additive on the algebra $\mathcal{C}$, that is

If $\{C_n : n \in \mathbb{N}\} \subseteq \mathcal{C}$ is a sequence of cylinders such that

$$C_n \supseteq C_{n+1} \forall n, \ \text{and} \ \bigcap_{n \in \mathbb{N}} C_n = \emptyset,$$

necessarily $\lim_{n \to \infty} \mathbf{P}(C_n) = 0$.

We proceed by contradiction, assuming $\mathbf{P}(C_n) \geq \varepsilon > 0 \ \forall n$ and showing that $\bigcap\limits_{n \in \mathbb{N}} C_n \neq \emptyset$.

By choosing the representations and eventually repeating the cylinders in the sequence, we always find a sequence $(t_n) \subseteq T$ and a sequence of cylinders $\{D_n : n \in \mathbb{N}\}$ with representations

$$D_n = \left\{ \omega \in \mathbb{R}^T : (\omega_{t_1}, \dots, \omega_{t_n}) \in A_n \right\}$$

where $A_n \in \mathcal{B}(\mathbb{R}^n)$, such that $D_n \supseteq D_{n+1} \forall n$, and for all $m \in \mathbb{N}$ there is some n such that $D_n = C_m$.

It follows that $\mathbf{P}(D_n) \geq \varepsilon > 0 \ \forall n$ and $\bigcap_{n\in\mathbb{N}} C_n = \bigcap_{n\in\mathbb{N}} D_n$.

Now since $P_{t_1,\dots,t_n}$ is a probability measure on $\mathbb{R}^n$, and A_n is Borel measurable, there is a closed set $F_n \subseteq A_n$ with $P_{t_1,\dots,t_n}(A_n \subseteq F_n) < \varepsilon 2^{-n}$. By σ-additivity, intersecting F_n with a ball large enough centered around the origin we find also a compact $K_n \subseteq A_n$ with

$$P_{t_1,\dots,t_n}(A_n \setminus K_n) < \varepsilon 2^{-n}$$

Consider the cylinders

$$F_n = \left\{ \omega \in \mathbb{R}^T : (\omega_{t_1}, \dots, \omega_{t_n}) \in K_n \right\}$$

Since these are not necessarily included into each other we take the intersections

$$F'_n = \bigcap_{m=1}^{n} F_k = \left\{ \omega \in \mathbb{R}^T : (\omega_{t_1}, \dots, \omega_{t_n}) \in K'_n \right\}$$

where $K'_n \subseteq K_n$ are compacts. We have

$$\begin{aligned} &P_{t_1,\dots,t_n}(K'_n) = \mathbf{P}(F'_n) = \mathbf{P}(D_n) - \mathbf{P}(D_n \setminus F'_n) = \\ &P_{t_1,\dots,t_n}(A_n) - P_{t_1,\dots,t_n}\left(\bigcup_{m=1}^{n} (A_n \setminus K_m) \right) \\ &\geq P_{t_1,\dots,t_n}(A_n) - P_{t_1,\dots,t_n}\left(\bigcup_{m=1}^{n} (A_m \setminus K_m) \right) \\ &\geq \mathbf{P}(D_n) - \sum_{m=1}^{n} \mathbf{P}(D_m \setminus F'_m) \geq \varepsilon - \sum_{m=1}^{n} \varepsilon 2^{-m} > 0 \end{aligned}$$

Therefore for each n, $\exists (x_1^{(n)} \dots, x_n^{(n)}) \in K'_n \neq \emptyset$.

Since the sequence F'_n is non-increasing, necessarily the sequence $(x_1^{(n)}) \subseteq K'_1$. By compactness, there is a convergent subsequence $x_1^{(n_l)} \to x_1^* \in K'_1$.

The subsequence $(x_1^{(n_l)}, x_2^{(n_l)}) \subseteq K'_2$, and there is a convergent subsequence with limit $(x_1^*, x_2^*) \in K'_2$.

By induction, we find a sequence (x_n^*) with $(x_1^*, \dots, x_n^*) \in K'_n \ \forall n$. The set

$$D^* = \left\{ \omega \in \mathbb{R}^T : \omega_{t_n} = x_n^* \quad \forall n \right\} \subseteq F'_n \subseteq D_n \quad \forall n \in \mathbb{N}$$

is nonempty, contradicting the hypothesis □

Definition 7. *A Borel space* $(S, \mathcal{S})$ *is a measurable space which can be mapped by a one-to-one measurable map* f *with measurable inverse to a Borel subset of the unit interval* $([0,1], \mathcal{B}([0,1]))$.

Lemma 3. *In a Borel space, the* σ*-algebra* $\mathcal{S}$ *is countably generated.*

Corollary 1. *Kolmogorov extensions theorem applies to processes* $(X_t(\omega))_{t\in T}$ *taking vaues in a Borel space* $(S,\mathcal{S})$*, (for example* $\mathbb{R}^d$*), without restrictions on the parameter set* T*.*

Proof By using a measurable bijection $f: S \leftrightarrow B \in \mathcal{B}([0,1])$, we define first a stochastic process $(Y_t(\omega))$ with values in $[0,1]$ and obtain $X_t(\omega) = f^{-1}(Y_t(\omega))$ with values in S.

Exercise 2. *A separable metric space* (S,d) *equipped with the Borel* σ*-algebra generated by the open sets is a Borel space.*

Hint: there is countable set $\{x_n\}_{n\in\mathbb{N}}$ which is dense in S. $\forall x \in S$ there is a subsequence $\{x_{n_k}\}_{k\in\mathbb{N}}$ such that $d(x_{n_k}, x) \to 0$.

Solution: We construct such subsequence explicitely as follows: let

$$n_k = \arg \min_{1\le m\le 2^k} \{d(x_m, x)\}$$

where we use lexicographic order in case of ambiguity.

Since $n_k \le 2^k$ it has a binary expansion

$$n_k = \sum_{m=0}^{k-1} x_m^{(k)} 2^m, \quad x_m^{(k)} \in \{0,1\}$$

so we can code n_k by the word $(x_0^{(k)}, \dots, x_{k-1}^{(k)}) \in \{0,1\}^k$, By concatenating these words we obtain the binary expansion of some $u \in [0,1]$. This map is one-to-one, from u we can recover the subsequence and (x_{n_k}) and the limiting point x_0. Although this map does not need to be continuous, it is measurable with measurable inverse: you can check that the image of a ball centered around some x_n is a Borel set in $[0,1]$, and the inverse image of $(k2^{-n}, (k+1)2^{-n}]$ $0 \le k \le 2^{-n}$ is a Borel set in S.

Warning: Working with random processes taking values in non-separable spaces can be tricky, since Kolmogorov theorem does not apply directly. During this lecture course we will stay on the safe side.

3 Continuity

So far we have constructed the probability measure $\mathbf{P}$ on $(\Omega = \mathbb{R}^T, \sigma(\mathcal{C}))$ such that the canonical process $X_t(\omega) = \omega_t$ follows the specified family of finite dimensional distribution. Suppose T is a topological space which is not countable, for example $T = \mathbb{R}$. In such case, the set

$$A = \{\omega : t \mapsto \omega_t \text{ is continuous at all } t \in T \ \}$$

does not belong to $\sigma(\mathcal{C})$ simply because to check continuity in an uncountable set we need uncountably many evaluations of the function $t \mapsto \omega_t$. In other words, $\mathbf{1}_A(\omega)$ is not a random variable.

Theorem 3. *(Kolmogorov's continuity criterium)*

We denote the dyadic subsets of $[0,1]^d$ *by*

$$D = \bigcup_{m\in\mathbb{N}} D_m \quad \textit{where} \quad D_m := \{2^{-m}(k_1,\dots,k_d) : 0 \le k_i \le 2^m\},\ m \in \mathbb{N}.$$

Note that D is countable and dense in $[0,1]^d$.

On a probability space $(\Omega, \mathcal{F}, P)$, let $(X_t : t \in T = [0,1]^d)$ a stochastic process with values in a normed vector space $(E, \| \cdot \|_E)$ (for example $E = \mathbb{R}^m$) When for $p, r > 0$

$$E\bigg(\| X_t - X_s \|_E^p \bigg) \le c|t-s|^{d+r}$$

for all $t, s \in T$, then for all $0 < \alpha < r/p$

$$\| X_t(\omega) - X_s(\omega) \|_E \le K_\alpha(\omega)|t-s|^\alpha \quad \forall s, t \in D$$

with $K_\alpha \in L^p(\Omega)$, in particular $K_\alpha(\omega) < \infty$ P-almost surely.

Proof

Let $N_m = \big\{(s,t) \in D_m : |s-t| = 2^{-m}\big\}$, the set of nearest neighbors pairs at level m.

Since $\#N_m = \frac{1}{2} \sum_{s \in D_m} \#\big\{\text{neighbors of } s\big\} \le 2^{-1} 2^{d(m+1)} 2d$

$$E\bigg(\sup_{(s,t) \in N_m} \| X_t - X_s \|^p \bigg) \le \sum_{(s,t) \in N_m} E\big(\| X_t - X_s \|^p \big) \le (2^{d(m+1)} d)(c 2^{-m(d+r)}) = 2^d dc 2^{-mr} \quad (4)$$

For $t \in D$ let t_m the nearest element in D_m.

Either $t_{m+1} = t_m$ or $|t_{m+1} - t_m| = 2^{-(m+1)}$, that is $(t_m, t_{m+1}) \in N_{m+1}$. Define analogously (s_m) for $s \in D$. Since $t, s \in D$ implies $t, s \in D_k$ for some k large enough, by using telescopic sums

$$X_t - X_s = (X_{t_m} - X_{s_m}) + \sum_{k=m}^{\infty} (X_{t_{k+1}} - X_{t_k}) - \sum_{k=m}^{\infty} (X_{s_{k+1}} - X_{s_k})$$

where we sum over finitely many non-zero terms. Note that if $t, s \in D$, $t \ne s$, necessarily $2^{-(m+1)} < |t-s| \le 2^{-m}$ for some $m \in \mathbb{N}$. In such case, $(t_m - s_m) = 2^m$ that is t_m and s_m are neighbors in D_m By starting the telescoping sum from such m,

$$\| X_t - X_s \| \le \| t_m - s_m \| + \sum_{k=m}^{\infty} \| X_{t_{k+1}} - X_{t_k} \| + \sum_{k=m}^{\infty} \| X_{s_{k+1}} - X_{s_k} \|$$

which gives

$$\sup\big\{ \| X_t - X_s \|^p : t, s \in D, 2^{-(m+1)} < |t-s| \le 2^{-m} \big\} \le 3 \sum_{k=m}^{\infty} \sup_{(t,s) \in N_m} \| X_{t_{k+1}} - X_{t_k} \|^p$$

By the triangle inequality in $L^p(\Omega, P, E)$ and (4)

$$E\bigg(\sup_{s,t \in D : |s-t| < 2^{-m}} \| X_t - X_s \|^p \bigg)^{1/p} \le 3 \sum_{k=m}^{\infty} E_P\bigg(\sup_{(t,s) \in N_k} \| X_t - X_s \|^p \bigg)^{1/p}$$

$$\le \bar{c} \sum_{k=m}^{\infty} 2^{-kr/p} = \bar{c} 2^{-mr/p}$$

Fix $\alpha < (r/p)$. By taking union over disjoint sets

$$E\bigg(\sup_{(s,t) \in D : s \ne t} \bigg\{ \frac{\| X_t - X_s \|}{|t-s|^\alpha} \bigg\}^p \bigg)^{1/p} \le \bar{c} \sum_{m=0}^{\infty} 2^{m\alpha} 2^{-mr/p} < \infty$$

which implies

$$K_\alpha(\omega) := \sup_{(s,t)\in D: s\neq t} \frac{\| X_t(\omega) - X_s(\omega) \|}{|t-s|^\alpha} < \infty \quad P\text{-almost surely} \tag{5}$$

Note that $\omega \mapsto K_\alpha(\omega)$ is measurable and $K_\alpha \in L^p(\Omega)$. By taking countable intersections of these events with $\alpha_n = \frac{r}{p}\left(\frac{n}{n+1}\right)$, almost surely (5) holds simultaneously for all $\alpha < r/p$ □

Corollary 2. *Under the assumptions of Theorem 3, when $(E, \|\cdot\|)$ is complete, there is a modification $\widetilde{X}_t(\omega)$ of the process $X_t(\omega)$ with α-Hölder continuous trajectories for all $0 < \alpha < r/p$.*

Proof It follows outside a measurable set $\mathcal{N}$ with $P(\mathcal{N}) = 0$, the paths $t \mapsto X_t(\omega)$ are uniformly continuous on the compact D.

Therefore for each $t \in [0,1]$

$$\widetilde{X}_t(\omega) := \begin{cases} \lim_{s\to t, s\in D} X_s(\omega) & \omega \in \mathcal{N}^c \\ x_0 & \omega \in \mathcal{N} \end{cases}$$

is well defined and measurable ($x_0 \in E$ is chosen arbitrarily).

This follows since, for $\omega \in \mathcal{N}^c$, if $s_n, s'_n \in D_n$ are dyadic sequences with $s_n \to t$ and $s'_n \to t$, $\forall \varepsilon > 0\ \exists n_\varepsilon(\omega)$ such that $\forall m, n > n_\varepsilon(\omega)$

$$\max\Big\{ \| X_{s_n}(\omega) - X_{s'_n}(\omega) \|,\ \| X_{s_m}(\omega) - X_{s_n}(\omega) \|,\ \| X_{s'_m}(\omega) - X_{s'_n}(\omega) \| \Big\} < \varepsilon$$

Therefore for $\omega \in \mathcal{N}^c$ $X_{s_n}(\omega)$ and $X_{s'_n}(\omega)$ are Cauchy sequences in the complete space E with a common limit.

Note that $\widetilde{X}_s(\omega) = X_s(\omega)$ for $s \in D$, and since $(X_s(\omega))_{s\in D}$ is α-Hölder continuous when $\omega \in N^c$, $0 < \alpha < 2/p$ by contruction $(\widetilde{X}_s(\omega))_{s\in[0,1]^d}$ is α-Hölder continuous $\forall\omega$ and all $0 < \alpha < r/p$.

From the hypothesis on increments' moments, by Chebychev inequality we get for fixed $t \in [0,1]^d$

$$X_s \overset{P}{\to} X_t \text{ as } s \to t,\ s \in T$$

in probability. By starting with a dyadic sequence, we find a subsequence $(s_k) \subseteq D$ such that $s_k \to t$ and P-almost surely

$$\lim_k X_{s_k}(\omega) = X_t(\omega)$$

Since $X_s(\omega) = \widetilde{X}_s(\omega)$ $\forall s \in D$, it follows that $\forall t \in [0,1]^d$

$$P(\{\omega : X_t(\omega) = \widetilde{X}_t(\omega)\}) = 1$$

that is $\widetilde{X}_t(\omega)$ is a continuous modification of $X_t(\omega)$.

In particular $\widetilde{X}_t$ and X_t have the same finite dimensional distributions □

Note that this continuous modification is unique up to indistinguishability. If $\hat{X}_t(\omega)$ is another continuous modification of $X_t(\omega)$, necessarily

$$P(\hat{X}_s(\omega) = X_s(\omega) = \tilde{X}_s(\omega) \quad \forall s \in D) = 1$$
$$\implies P(\hat{X}_t(\omega) = \tilde{X}_t(\omega) \quad \forall t \in [0,1]^d) = 1$$

Corollary 3. *On the probability space $(\Omega = (\mathbb{R})^{\mathbb{R}}, \sigma(\mathcal{C}))$, there is a probability measure $\mathbf{P}_W$ (the Wiener measure) and a stochastic process $B_t(\omega)$ which satisfies definition 1. Morover there is a modification which has locally α-Hölder continuous paths $t \mapsto B_t(\omega)$ $\forall \omega \in \Omega$ for any $0 < \alpha < 1/2$.*

Locally means that α-Hölder continuity holds on compacts.

Note by taking images, the Wiener measure $\mathbf{P}_W$ is also defined on the spaces $C(\mathbb{R}^+;\mathbb{R}), C^\alpha(\mathbb{R}^+;\mathbb{R})$ of continuous and locally α-Hölder continuous functions, for $0 < \alpha < 1/2$. Under the Wiener measure, in these function spaces the canonical process is a Brownian motion.

Proof We first take $T = [0,1]$ $\Omega = \mathbb{R}^{[0,1]}$ Definition 1 determines consistently the family of finite dimensional distributions of Brownian motion. By Kolmogorov extension theorem, there a probability measure $\mathbf{P}_W$ on $(\Omega, \sigma(\mathcal{C}))$ consistent with the finite dimensional distributions' specification. In particular the canonical process $X_t(\omega) = \omega_t$ has gaussian increments $(X_t(\omega) - X_s(\omega)) \sim N(0, t-s)$.

The gaussian distribution has the following property: if $G(\omega)$ is a gaussian random variable with $E(G) = 0$, then $E(G^{2n+1}) = 0$ $\forall n$, and there are constants (c_n) such that

$$E(G^{2n}) = c_n \{E(G^2)\}^n$$

By the continuity theorem with $d = 1$ and $p = 2n, n \in \mathbb{N}$ we get

$$E(|X_t - X_s|^{2n}) = c_n |t-s|^n = c_n |t-s|^{1+(n-1)} \quad \forall n \in \mathbb{N}$$

from which it follows that $(X_t(\omega))$ has a modification $(B_t(\omega))$ which is α-Hölder continuous for all α with

$$\alpha < \sup_{n \in N} \frac{(n-1)}{2n} = 1/2$$

Let $(B_t^{(n)})_{t \in [0,1]}$ a sequence of independent copies of the Brownian motion defined on the canonical space of continuous function $\Omega_n = C([0,1], \mathbb{R})$ equipped with the Wiener measure. Note that since $C([0,1], \mathbb{R})$ is separable there is not problem to apply Kolomogorov theorem to define the product measure on the infinite product space.

By concatenating these independent copies into a single continuous path we obtain a Brownian motion indexed by $T = [0, +\infty)$, or $T = \mathbb{R}$.

4 Exercises

1. When $f \in C^2$, from 1 we get the *semimartingale decomposition* of the process $f(B_t(\omega))$ as an Ito integral plus a process with finite variation of on compacts. Write the semimartingale decomposition in the following cases

- $f(x) = x^n$; $f(x) = \sin(x)$; $f(x) = \exp(x)$.
- $f(x) = h(g(x))$, $f(x) = h(x)g(x)$ where h and g are some functions above.

2. Let $(B_t(\omega))_{t\geq 0}$ and $(W_t(\omega))_{t\geq 0}$ two indepenent Brownian motions defined on the same probability space. Adapt the proof of lemma 1 to show that *quadratic covariation*

$$[W, B]_t = \lim_{\Delta(\Pi)\to 0} \sum_i (W_{t_{i+1}} - W_{t_i})(B_{t_{i+1}} - B_{t_i}) \xrightarrow{P} 0 \tag{6}$$

where we take the limit over partitions $\Pi = (0 = t_0 \leq t_1 \leq \cdots \leq t_n = t)$, $n \in \mathbb{N}$ as $\Delta(\Pi) \to 0$ Hint: take the limit in $L^2(P)$ and use independence.

The process $[W, B]_t$ is called *quadratic covariation* .

Adapt the proof of lemma 2 that for the dyadic sequence of partitions $\Pi_n = D_n$ we have also almost sure convergence in (6).

3. Let $(S, \mathcal{S})$ a Borel space, and $K(x, dy)$ a probability kernel, that is a map $K : S \times \mathcal{S} :\to [0, 1]$, such that

   (a) $\forall x \in S$ the map $A \mapsto K(x, A)$ is a probability measure

   (b) $\forall A \in \mathcal{S}$, the map $x \mapsto K(x, A)$ is measurable

   - Let $x \in S$. Use Kolmogorov's extension theorem to show that there exist a probability measure $\mathbf{P}_x$ on the space of sequences $\Omega = S^{\mathbb{N}}$ such that the collection of canonical random variables $(X_t(\omega) = \omega_t,\ t \in \mathbb{N})$ satisfies for all $n,\ A_1, \dots, A_n \in \mathcal{S}$,

$$\mathbf{P}_x\Big(X_0(\omega) \in A_0, X_1(\omega) \in A_1, \dots, X_n(\omega) \in A_n \Big) =$$
$$\mathbf{1}_{A_0}(x) \int_{A_1\times\cdots\times A_{n-1}\times A_n} K(x_{n-1}, dx_n)K(x_{n-2}, dx_{n-1}) \dots K(x, dx_1)$$

   - Let

$$K^n(x, dy) := \mathbf{P}_x(X_n \in dy), n \in \mathbb{N} \tag{7}$$

   Show that $K^n(x, dy)$ $n \in \mathbb{N}$ is a probability kernel.

   - Show that the Chapman-Kolmogorov equation holds: for all $m, n \in N$

$$K^{n+m}(x, dy) = \int_S K^n(x, dz)P^m(z, dy) \tag{8}$$

   - Let $\pi(dx)$ a probability measure on $(S, \mathcal{S})$. Show that there exist a probability measure $\mathbf{P}_\pi$ on the space of sequences $\Omega = S^{\mathbb{N}}$ such that the collection of canonical random variables $(X_t(\omega) = \omega_t,\ t \in \mathbb{N})$ satisfies for all $n,\ A_0, A_1, \dots, A_n \in \mathcal{S}$,

$$\mathbf{P}_\pi\Big(X_0(\omega) \in A_0, X_1(\omega) \in A_1, \dots, X_n(\omega) \in A_n \Big)$$
$$= \int_{A_0\times A_1\times\cdots\times A_{n-1}\times A_n} K(x_{n-1}, dx_n)K(x_{n-2}, dx_{n-1}) \dots K(x_0, dx_1)\pi(dx_0)$$

The process $(X_t(\omega) : t \in \mathbb{N})$ is called the time homogeneous Markov-process with initial distribution $\pi(dx)$ and transition kernel $K(x, dy)$.

- Let $(S, \mathcal{S}, \lambda)$ a measurable space, where we assume that λ is a σ-finite positive measure with no atoms, that is $\lambda(\{x\}) = 0$ for all $x \in S$.

4. - Use Kolmogorov consistency theorem to show that there is real valued random process $(W_A(\omega) : A \in \mathcal{S}$ with $\lambda(A) < \infty)$ indexed by sets, such that:
$\forall A \in \mathcal{S}$ with $\lambda(A) < \infty$, $W_A(\omega)$ is a gaussian random variable with 0 mean and variance $\lambda(A)$; $W_A \perp\!\!\!\perp W_B$ when $A, B \in \mathcal{S}$ and $A \cap B = \emptyset$.

   Moreover we ask that $A \to W_A(\omega)$ is finitely additive for all ω, that is for $A \cap B = \emptyset$, with $\lambda(A \cup B) < \infty$, $W_{A\cup B}(\omega) = W_A(\omega) + W_B(\omega)$
$W_A(\omega)$ is called Wiener noise or white noise driven by λ.

   Hint: to show consistency use the following property: if $W_i \sim \mathcal{N}(\mu_i, \sigma_i^2),\ i = 1, 2$ are independent gaussian random variables, then their convolution is gaussian:

$$(W_1 + W_2) \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

- On the settings of the previous exercise, use Kolmogorov consistency theorem to show that there is a $\mathbb{N}$-valued random process $(N_A(\omega) : A \in \mathcal{S})$ such that:

   $\forall A \in \mathcal{S}$ with $\lambda(A) < \infty$, $N_A(\omega)$ is a Poisson random variable with parameter $\lambda(A)$; $N_A \perp\!\!\!\perp N_B$ when $A, B \in \mathcal{S}$ and $A \cap B = \emptyset$.

   Moreover we ask that $A \to N_A(\omega)$ is finitely additive for all ω, that is for $A \cap B = \emptyset$, with $\lambda(A \cup B) < \infty$, $N_{A\cup B}(\omega) = N_A(\omega) + N_B(\omega)$
The centered process $\widetilde{N}_A(\omega) := (N_A(\omega) - \lambda(A))$ is called Poisson noise driven by λ.

   Hint: to show consistency use the following property: if $N_i \sim \text{Poisson}(\lambda_i),\ i = 1, 2$ are independent Poisson random variables, then their convolution is Poisson:

$$(N_1 + N_2) \sim \text{Poisson}(\lambda_1 + \lambda_2)$$

5 Conditional expectation

Let $(\Omega, \mathcal{F}, P)$ a probability space and $\mathcal{G} \subseteq \mathcal{F}$ a sub σ-algebra. Let $X(\omega) \geq 0$ be a random variable $\mathcal{F} \geq 0$. A $\mathcal{G}$-measurable random variable $Y(\omega)$ is a version of the conditional expectation $E_P(X|\mathcal{G})(\omega)$ if $\forall G \in \mathcal{G}$

$$E_P(X\mathbf{1}_G) = E_P(Y\mathbf{1}_G)$$

More in general when $X(\omega) = X^+(\omega) - X^-(\omega)$ with $X^{\pm}(\omega) \geq 0$, we take define

$$E_P(X|\mathcal{G})(\omega) = E_P(X^+|\mathcal{G})(\omega) - E_P(X^-|\mathcal{G})(\omega)$$

the right hand side is well defined. Otherwise the conditional expectation does not exists.

Altough in most of the textbooks it is assumed $E_P(|X|) < \infty$, our extended definition makes sense and could be useful.

For example, let $Z(\omega) = \lfloor X(\omega) \rfloor \in \mathbb{Z}$, the integer part of the random variable X, and let $\mathcal{G} = \sigma(Z)$.

Then the random variable

$$Y(\omega) := \sum_{z\in\mathbb{Z}} \frac{\int_{[z,z+1)} x P_X(dx)}{P_X([z, z+1))} \mathbf{1}(Z(\omega) = z)$$

with the convention that $\frac{0}{0} = 0$, satisfies the definition of $E_P(X|\mathcal{G})(\omega)$ even when X in not integrable (in such case Y is also not integrable).

Lemma 4. $X(\omega) \geq 0$ P $a.s \Longrightarrow E_P(X|\mathcal{G})(\omega) \geq 0$.

Proof By contradiction, assume that $Y(\omega) = E_P(X|\mathcal{G})(\omega) < 0$ with positive probability. Then $\exists n$ such that $P(G) > 0$, where

$$G = \big\{\omega : Y(\omega) < -1/n\big\}$$

is $\mathcal{G}$-measurable since Y is. Then by the definition of conditional expectation

$$0 \leq E_P(X\mathbf{1}_G) = E_P(Y\mathbf{1}_G) \leq -\frac{1}{n}P(G) < 0$$

which gives a contradiction since the last inequality is strict.

Proposition 1. *These properties follow directly from the definition of conditional expectation and positivity, when the conditional expectations do exist.*

1. *Linearity*

2. *Monotone convergence: if* $0 \leq X_n(\omega) \uparrow X(\omega)$ P *a.s.* $\Longrightarrow E_P(X_n|\mathcal{G})(\omega) \uparrow E_P(X|\mathcal{G})(\omega)$ P *a.s.*

3. *Fatou lemma:* $0 \leq X_n(\omega) \Longrightarrow E_P(\liminf X_n|\mathcal{G})(\omega) \leq \liminf_n E_P(X_n|\mathcal{G})(\omega)$ P *a.s.*

4. *Dominated convergence: if* $|X_n(\omega)| \leq Y(\omega)$ *where* $Y(\omega)$ *is* $\mathcal{G}$ *measurable and* $X_n(\omega) \to X(\omega)$ P *almost surely, then* $E_P(X_n|\mathcal{G})(\omega) \to E_P(X|\mathcal{G})(\omega)$ P*-almost surely.*

5. *if* Y *is* $\mathcal{G}$ *measurable,*

$$E_P(XY|\mathcal{G})(\omega) = Y(\omega)E_P(X|\mathcal{G})$$

6. *when* $\mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{G}$ *are nested* σ*-algebrae*

$$E_P(X|\mathcal{H}) = E_P\big(E_P(X|\mathcal{G})\big|\mathcal{H}\big)$$

7. *When* $\mathcal{H}$ *is independent from the* σ*-algebra* $\sigma(X) \vee \mathcal{G}$,

$$E_P(X|\mathcal{G} \vee \mathcal{H}) = E_P(X|\mathcal{H})$$

Hint: it is enough to use independence checking the definition of conditional expectation for the sets $\{G \cap H : H \in \mathcal{H}, G \in \mathcal{G}\}$ *which generate the* σ*-algebra* $\mathcal{G} \vee \mathcal{H}$.

8. *Jensen inequality: if $f(x)$ is a convex function (for example $f(x) = |x|^p$ for $p \geq 1$),*

$$f\big(E_P(X|\mathcal{G})\big) \leq E_P(f(X)|\mathcal{G})$$

Theorem 4. *When $X \in L^2(\Omega, \mathcal{F}, P)$, then the conditional expectation $Y = E_P(X|\mathcal{G})$ exists as the orthogonal projection of X to the closed subspace $L^2(\omega, \mathcal{G}, P)$.*

Hint. By using completeness one shows the orthogonal projection is well defined as the element of $L^2(\omega, \mathcal{G}, P)$ minimizing

$$E_P((X - Z)^2)$$

among all $Z \in L^2(\omega, \mathcal{G}, P)$. Since $(Y + tZ) \in L^2(\omega, \mathcal{G}, P)$ for every $t \in \mathbb{R}$,

$$E_P((X - Y - tZ)^2) \geq E_P((X - Y)^2) \Longleftrightarrow t^2 E_P(Z^2) - 2t E_P((X - Y)Z) \geq 0$$

for all t. Letting $t \to 0$ we see that necessarily $E_P((X - Y)Z) = 0$, so that $Y = E_P(X|\mathcal{G})$ according to the definition.

Corollary 4. *When $X \in L^1(\Omega, \mathcal{F}, P)$ the conditional expectation $Y = E_P(X|\mathcal{G})$ exists in $L^1(\Omega, \mathcal{G}, P)$*

Proof When $X(\omega) \geq 0$ take $X^{(n)}(\omega) = (X(\omega) \wedge n) \in L^2$. By the previous theorem and positivity exists $0 \leq Y^{(n)} = E_P(X^{(n)}|\mathcal{G}) \uparrow Y(\omega)$, with $\mathcal{G}$-measurable limit. By using the monotone convergence theorem we then check that $Y(\omega)$ satisfies the definition of conditional expectation. More in general by decomposíng $X(\omega) = (X^+(\omega) - X^-(\omega))$ with $X^{\pm} = (\pm X, 0)$ the result follows from linearity.

Definition 8. *Regular versions*

6 Martingales

Definition 9. *Let $(\Omega, \mathcal{F})$ a probability space. A filtration is an increasing collection of σ-algebrae $(\mathcal{F}_t : t \in T)$ where $T = \mathbb{N}, \mathbb{R}^+, \mathbb{Z}, \mathbb{R}$ such that for all $s \leq t$ $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$*

Definition 10. *A stochastic process $(X_t : t \in T)$ is adapted to the filtration $(\mathcal{F}_t : t \in T)$, if X_t is $\mathcal{F}_t$-meaasurable for all $t \in T$.*

Definition 11. *A random variable $\tau(\omega) \in T = \mathbb{R}^+, \mathbb{N}$ is a $(\mathcal{F}_t)$-stopping time if*

$$\{\omega : \tau(\omega) \leq t\} \in \mathcal{F}_t \quad \forall t \in T$$

Equivalently the counting process $N_t(\omega) := \mathbf{1}(\tau(\omega) \leq t)$ is adapted to the filtation.

Definition 12. *Let $\tau(\omega)$ an $(\mathcal{F}_t)$-stopping time Define*

$$\mathcal{F}_\tau := \big\{A \in \mathcal{F} : A \cap \{\tau \leq t\} \in \mathcal{F}_t \quad \forall t \in T\big\}.$$

as the stopped σ-algebra.

Exercise 3.

- *Check that $\mathcal{F}_\tau$ is a σ-algebra.*
- *If $\sigma \le \tau$ are $(\mathcal{F}_t)$-stopping times then $\mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$*

Definition 13. *A (sub,super)-martingale with respect to the filtration $(\mathcal{F}_t)_{t\in T}$ is an adapted and integrable process $(X_t : t \in T) \subseteq L^1(P)$ which satisfies the martingale property: for $s \le t$*

$$E_P(M_t|\mathcal{F}_s) = M_s$$

(respectively $\ge, \le$)

Note the martingale property depends both on the probability measure and on the filtration.

Exercise 4. *Let $(X_t : t \in \mathbb{N}) \subseteq L^1(P)$ independent random variables with $E(X_t) = 0$, and $\mathcal{F}_t = \sigma(X_1, X_2, \dots, X_t)$ Then $M_t = (X_1 + \dots + X_t)$ is a $(\mathcal{F}_t)$-martingale*

Exercise 5. *Let $(X_t : t \in \mathbb{N}) \subseteq L^1(P)$ independent random variables with $E(X_t) = 1$, and $\mathcal{F}_t = \sigma(X_1, X_2, \dots, X_t)$ Then $M_t = (X_1 \times \dots \times X_t)$ is a $(\mathcal{F}_t)$-martingale*

Exercise 6. *Let $X_n(\omega) \in \mathbb{R}^d$ a discrete time Markov chain with initial distribution π and transition kernel K*

Define the operator $(Kf)(x) = \int_{\mathbb{R}^d} f(y)K(y, dx) = E_x(f(X_1))$

Check that this is a martingale

$$M_t(f) = \sum_{s=1}^{t}(f(X_s) - (Kf)(X_{s-1}))$$

Taking telescopic sums

$$f(X_t) = f(X_0) + \sum_{s=1}^{t}(f(X_s) - f(X_{s-1}) =$$

$$f(X_0) + \sum_{s=1}^{t}(f(X_s) - Kf(X_{s-1}) + \sum_{s=1}^{t}((Kf)(X_{s-1}) - f(X_{s-1}))$$
$$= f(X_0) + M_t(f) + A_t(f)$$

(decomposition into martingale and predictable part)

Definition 14. *A process $(Y_t(\omega) : t \in \mathbb{N})$ is predictable with respect to the discrete-time filtration $(\mathcal{F}_t : t \in \mathbb{N})$, if Y_t is $\mathcal{F}_t$-measurable for all $t \in T$.*

Proposition 2. *Let (X_t) be a martingale and (Y_t) a predictable process in the discrete-time filtration $(\mathcal{F}_t : t \in \mathbb{N})$. Define the martingale transform*

$$M_t(\omega) = \sum_{s=1}^{t} Y_s(M_s - M_{s-1})$$

When $E(|Y_s \Delta M_s|) < \infty$ $\forall s \in T$, (M_t) is a martingale.

Proof From the definition we see that M_t is adapted and integrability follows from triangle inequality. We check the martingale property:

$$E_P(M_t - M_{t-1}|\mathcal{F}_{t-1}) = E_P(Y_t(X_t - X_{t-1})|\mathcal{F}_{t-1}) = Y_t E_P(X_t - X_{t-1}|\mathcal{F}_{t-1}) = 0$$

where we use predictability of Y_t together with the definition of conditional expectation.

In order to check integrability it is enough to use Hölder inequality,

$$E(|Y_s \Delta M_s|) \leq \| Y_s \|_{L_p} \| \Delta M_s \|_{L_q}$$

for conjugate exponents $p, q \in [1, +\infty]$, $p^{-1} + q^{-1} = 1$.

6.1 Convergence of forward martingales

Theorem 5. *(Doob)*

Let $(X_t : t \in \mathbb{N})$ a supermartingale with $E_P(X_t^-) < \infty$,
(here $x^- = \max(-x, 0)$)).
Then $X_\infty(\omega) \in L^1(\Omega)$ and

$$\lim_{t\to\infty} X_t(\omega) = X_\infty(\omega) \quad P\text{-almost surely}$$

with $X_\infty(\omega) \in L^1(\Omega)$

Note : although $X_\infty(\omega) \in L^1(\Omega)$ we don't have necessarily convergence in $L^1(P)$ sense.

Proof Note first that by the supermartingale propery, $\forall t \in \mathbb{N}$

$$E(X_t^+) \leq E(X_0) + E(X_t^-)$$

so that

$$\sup_t E(X_t^+) \leq E(X_0) + \sup_t E(X_t^-)$$

where $E(|X_0|) < \infty$, so that the sequence $(X_t)_{t\in\mathbb{N}}$ is bounded in $L^1(P)$.

Given $a < b$, we define a sequence of stopping times

$$\sigma_0(\omega) = \inf\{s \in \mathbb{N} : X_s(\omega) < a\}$$
$$\tau_i(\omega) = \inf\{s > \sigma_i(\omega) : X_s(\omega) \geq b\}, \sigma_i(\omega) = \inf\{s > \tau_{i-1}(\omega) : X_s(\omega) < a\},\ i \geq 1$$

We have $0 \leq \sigma_i < \tau_i < \sigma_{i+1} < \dots$, To check that these are stopping times, note that for each $t \in \mathbb{N}$ the events

$$\{\omega : \sigma_i(\omega) \leq t\} \quad \text{and} \quad \{\omega : \tau_i(\omega) \leq t\}\mathcal{T}$$

are $\mathcal{F}_t$ since they depend on the trajectory of the $(\mathcal{F}_t)$-adapted process X_t up to time t.

Define the investement strategy

$$C_t(\omega) = \begin{cases} 1 & t \in (\sigma_i, \tau_i] \text{ for some } i \in \mathbb{N} \\ 0 & t \in (\tau_i, \sigma_{i+1}] \end{cases}$$

Note that since τ_i and σ_i are stopping times, for all $t \in N$

$$\{C_t = 1\} = \bigcup_{i \in \mathbb{N}} \{t \in (\sigma_i, \tau_i]\} = \bigcup_{i \in \mathbb{N}} \{\sigma_i \leq (t-1)\} \cap \{\tau_i \leq (t-1)\}^c \in \mathcal{F}_{t-1}$$

Since $C_t(\omega) \in \{0,1\}$ is a non-negative and bounded predictable process, it follows that the martingale transform

$$Y_t(\omega) = \sum_{s=1}^{t} C_s(\omega) \Delta X_s$$

has the supermartingale property.

Note that

$$Y_t \geq (b-a) U_{[a,b]}([0,t]) - (X_t - a)^-$$

By taking expectation, since $E(Y_t) \leq E(Y_0) = 0$ from the supermartingale property, we obtain *Doob upcrossing inequality*

$$E_P\big(U_{[a,b]}([0,t])\big) \geq \frac{1}{(b-a)} E_P\big((X_t - a)^-\big)$$

Now since $U_{[a,b]}([0,t])$ is non-decreasing, for every ω exists

$$U_{[a,b]}([0,\infty),\omega) := \lim_{t\to\infty} U_{[a,b]}([0,t]) \in \mathbb{N} \cup \{+\infty\}$$

and by monotone convergence theorem, since

$$(X_t - a)^- = \max(a - X_t, 0) \leq |a| + X_t^-$$

we obtain

$$E_P\big(U_{[a,b]}([0,\infty),\omega)\big) = \lim_{t\to\infty} E_P\big(U_{[a,b]}([0,t])\big) \leq \frac{1}{(b-a)} \left(|a| + \sup_{t\in\mathbb{N}} E_P(X_t^-) \right) < \infty$$

In particular $U_{[a,b]}([0,\infty),\omega) < \infty$ P-almost surely.

Now let

$$\begin{aligned} N &= \big\{\omega : \liminf_{t\to\infty} X_t(\omega) \lneq \limsup_{t\to\infty} X_t(\omega)\big\} \\ &= \bigcup_{a<b\in Q} \big\{\omega : \liminf_{t\to\infty} X_t(\omega) \leq a < b \leq \limsup_{t\to\infty} X_t(\omega)\big\} \\ &= \bigcup_{a<b\in\mathbb{Q}} \big\{U_{[a,b]}([0,\infty),\omega) = \infty\big\} \end{aligned}$$

so that $P(N) = 0$ since is the countable union of null sets.

This means that P-almost surely $(X_t(\omega))_{t\in\mathbb{N}}$ is a converging sequence with limit $X_\infty(\omega) := \limsup_{t\to\infty} X_t(\omega), \quad \forall \omega \in \Omega$.

Note that a priori $X_\infty(\omega) \in [-\infty, \infty]$.

By using Fatou lemma

$$E(|X_\infty|) = E(\liminf_t |X_t|) \leq \liminf_t E(|X_t|) \leq \sup_t E(|X_t|) < \infty$$

In particular $|X_\infty(\omega)| < \infty$ P-almost surely $\square$.

Corollary 5. *A non-negative supermartingale X_t has almost surely an integrable limit X_∞ with $E_P(X_\infty) \le E_P(X_t)$ for $t < \infty$.*

Proof For all $t \in \mathbb{N}$

$$E_P(|X_t|) \le E_P(X_t) = E_P\big(E_P(X_t|\mathcal{F}_0)\big) \le E_P(X_0) = E_P(|X_0|)$$

so that L^1 boundedness follows for free and Doob convergence theorem applies
□

Corollary 6. *Let $(X_t : t \in \mathbb{N})$ a submartingale with $E_P(X_t^+) < \infty$. Then for P almost all ω $\exists \lim_{t\to\infty} X_t(\omega) = X_\infty(\omega) \in L^1(P)$.*

Proof Apply the theorem to the supermartingale $(-X_t)$

6.2 Uniform integrability

Let M_t a martingale bounded in $L^1(P)$, with limit $M_\infty \in L^1(P)$ Does the martingale property holds at infinity ?, that is

$$M_s = E(M_\infty|\mathcal{F}_s), \quad s \ge 0$$

Definition 15. *A collection of random variables $\{X_t : t \in T\}$ is uniformly integrable (UI) if*

$$\lim_{k\to\infty} \sup_{t\in T} E_P\big(|X_t|\mathbf{1}(|X_t| > k)\Big) = 0$$

Note that a single random variable $X \in L^1(P)$ since by monotone convergence theorem $E(|X| \wedge k) \uparrow E(|X|) < \infty$,

$$E(|X| - |X| \wedge k) = E(|X|\mathbf{1}(|X| > k)) \downarrow 0 \quad \text{as} k \uparrow \infty$$

Proposition 3. *The collection of random variables $\{X_t : t \in T\} \subseteq L^1(P)$ is UI if and only if*

for all $\varepsilon > 0$ $\exists\delta$ such that

$$\sup_{t\in T} E_P\big(|X_t|\mathbf{1}_A\big) < \varepsilon$$

when $P(A) \le \delta$.

We have the following characterization convergence in $L^1(P)$-norm:

Proposition 4. *Let $(X_n : n \in \mathbb{N}) \subseteq L^1(P)$ and $X \in L^1(P)$ $E(|X_n - X|) \to 0$ if and only if*

- *$X_n \xrightarrow{P} X$, and*
- *The collection $\{X_n : n \in \mathbb{N}\}$ is UI*

UI is a compactness condition in $L^1(P)$ when we replace the norm topology by the weak topology:

Proposition 5. *(Dunford Pettis) The collection of random variables $\mathcal{C}\{X_t : t \in T\} \subseteq L^1(P)$ is UI if and only if it is weakly compact in $L^1(P)$ that is for every sequence $(t_n) \subseteq T$ there is a subsequence (t_{n_k}) and a random variable $X \in L^1(P)$ such that $\forall A \in \mathcal{F}$*

$$E_P\big((X_{n_k} - X)\mathbf{1}_A\big) \to 0$$

Proof of $\Longrightarrow$ It is enough to consider the case when $X_t(\omega) \geq 0 \ \forall t$, since weak compacteness of $\mathcal{C}$ will follow from weak compactness of $(X_t^+ : t \in T)$ and $(X_t^- : t \in T)$. Let $(X_n : n \in \mathbb{N}) \subseteq \mathcal{C}$ and for $M \in \mathbb{N}$ consider the truncated random variables $X_n^(M) = X_n(\omega) \wedge M$. For fixed M, the sequence $(X_n^{(M)} : n \in \mathbb{N})$ is bounded in $L^2(P)$.

Banach-Alaoglu's theorem says that closed balls in the dual space of a Banach space are compact under the weak-star topology. Since $L^2(P)$ is the dual of itself and $\mathbf{1}_A \in L^2(P)$, it follows that for every $M \in \mathbb{N}$ there is a subsequence (n_k) (which at first depends on M) and a r.v. $X^{(M)} \in L^2(P)$ such that $\forall A \in \mathcal{F}$

$$E_P\left((X_{n_k}^{(M)} - X^{(M)})\mathbf{1}_A\right) \to 0 \text{ as } M \to \infty$$

which means $X_{n_k}^{(M)} \to X^{(M)}$ weakly in $L^1(P)$ (the dual of $L^1(P)$ is $L^\infty(P)$ the space of essentially bounded random variables, by a monotone class argument it is enough to check convergence using indicators). By taking further subsequences and using a diagonal argument we find a further subsequence (n_k) such that the convergence 6.2) holds simultaneously for all $M \in \mathbb{N}$. For $M, N \in \mathbb{N}$, by Fatou lemma for

$$\begin{aligned}
&E\big(|X^{(M+N)} - X^{(M)}|\big) \leq \liminf_k E\big(|X_{n_k}^{(M+N)} - X_{n_k}^{(M)}|\big) \\
&\leq \sup_{t\in T} E\big(|X_t - M|\mathbf{1}(|X_t| > M)\big) \\
&\leq 2\sup_{t\in T} E\big(|X_t|\mathbf{1}(|X_t| > M)\big) \to 0 \text{ as } M \to \infty
\end{aligned}$$

because of the UI assumption.

Therefore $(X^{(M)} : M \in \mathbb{N})$ is a Cauchy sequence in the complete space $L^1(P)$ and it converges in $L^1(P)$ norm to a limit $X \in L^1(P)$

For $A \in \mathcal{F}$,

$$\begin{aligned}
&\Big|E_P\big((X_{n_k} - X)\mathbf{1}_A\big)\Big| \\
&= \Big|E_P\big((X_{n_k} - X_{n_k}^{(M)})\mathbf{1}_A\big) + E_P\big((X_{n_k}^{(M)} - X^{(M)})\mathbf{1}_A\big) + E_P\big((X^{(M)} - X)\mathbf{1}_A\big)\Big| \\
&\leq 2E_P\big(X_{n_k}\mathbf{1}(X_{n_k} > M)\big) + E_P\big((X_{n_k}^{(M)} - X^{(M)})\mathbf{1}_A\big) + E_P\big(|X^{(M)} - X|\big)
\end{aligned}$$

where we choose first M large enough to make $E_P\big(|X^M - X|\big)$ small, and for such fixed M the first to terms are arbitrarily small for k large enough.

6.3 UI martingales

Lemma 5. *Let $X \in L^1(P)$. Then the family*

$$\Big\{Y = E_P(X|\mathcal{G}) : \mathcal{G} \subseteq \mathcal{F} \text{ sub-}\sigma\text{-algebra}\Big\}$$

is uniformly integrable.

Proof Let $\varepsilon > 0$ and δ such that $E_P\big(|X|\mathbf{1}_A\big) < \varepsilon$ when $P(A) \le \delta$.
Choose $k > \delta E(|X|)^{-1}$.
Let $Y = E_P(X|\mathcal{G})$ with $\mathcal{G} \subseteq \mathcal{F}$ sub-σ-algebra.
From Jensen inequality

$$|Y| \le E\big(|X||\mathcal{G}\big)$$

so that $E(|Y|) \le E(|X|)$ and by Chebychev inequality

$$P(|Y| > K) \le K^{-1}E(|Y|) \le K^{-1}E(|X|) < \delta$$

Since $\{\omega : |Y(\omega)| > K\} \in \mathcal{G}$, by the Jensen inequality for conditional expectations

$$E\big(|Y|\mathbf{1}(|Y| > K)\big) \le E\big(|X|\mathbf{1}(|Y| > K)\big) < \varepsilon$$

Proposition 6.

- *Let $(M_t : t \in \mathbb{N})$ an UI martingale. Then*

$$M_t \stackrel{L^1(P)}{\to} M_\infty \text{ and } M_t = E_P(M_\infty|\mathcal{F}_t)$$

- *Let $M_\infty \in L^1(P)$ and define $M_t = E_P(M_\infty|\mathcal{F}_t)$. Then $(M_t : t \in [0, +\infty])$ is an UI martingale.*

Proof

- From the UI property follows that for any $K \ge 0$

$$\sup_{t\in\mathbb{N}} E_P(|M_t|) \le K + \sup_{t\in T} E_P\big(|M_t|\mathbf{1}(|M_t| > K)\big) < \infty$$

  so that Doob martingale convergence theorem applies, there exists $M_\infty \in L^1(P)$ such that $M_t(\omega) \to M_\infty(\omega)$ P a.s. By the UI assuption, using the characterization of $L^1(P)$ convergence we have $E_P\big(|M_t - M_\infty|\big) \to 0$.

  To show the martingale property,let's fix $t \ge 0$ and $A \in \mathcal{F}_t$.

  The sequence $M_T(\omega)\mathbf{1}_A(\omega) \to M_{-\infty}(\omega)\mathbf{1}_A(\omega)$ as $T \to \infty$ and it is obviously an UI family, so that by the martingale property and charaterization of $L^1(P)$ convergence, for $T \ge t$,

$$E_P(M_t\mathbf{1}_A) = E_P(M_T\mathbf{1}_A) \to E_P(M_\infty\mathbf{1}_A) \quad \Box$$

- When $M_\infty \in L^1(P)$ From the properties of the conditional expectation it follows that $M_t = E_P(M_\infty|\mathcal{F}_t)$ is integrable, adapted and satisfies the martingale property.

  Uniform integrability follows from lemma (**??**)□.

6.4 Convergence of backward martingales

Definition 16. *A backward filtration is an increasing family of σ-algebrae $(\mathcal{F}_t : (-t) \in T)$ where $T = \mathbb{N}, \mathbb{R}$, For $0 \ge t \ge u$*

$$\mathcal{F} \supseteq \mathcal{F}_t \supseteq \mathcal{F}_u \supseteq \mathcal{F}_{-\infty} = \bigcap_{(-t)\in T} \mathcal{F}_t$$

where $\mathcal{F}_{-\infty}$ is the tail σ-algebra . The interpretation is that information in $\mathcal{F}_t$ decreases as $t \downarrow -\infty$.

Definition 17. *A backward (sub,super)-martingale with respect to the backward filtration* $(\mathcal{F}_t)_{(-t)\in T}$ *is an adapted and integrable process* $(X_t : -t \in T) \subseteq L^1(P)$ *which satisfies the martingale property: for* $0 \geq t \geq u$

$$E_P(X_t|\mathcal{F}_u) = X_u$$

(respectively $\geq, \leq$*)*

Theorem 6. *(Doob backward martingale convergence) Let* $(X_t : -t \in \mathbb{N})$ *a backward submartingale.*

1. *P-almost surely, exists the limit*

$$X_{-\infty}(\omega) = \lim_{t\to-\infty} X_t(\omega) \in [-\infty, \infty)$$

2. *Under the assumption*

$$\sup_{-t\in\mathbb{N}} E(X_t^-) < +\infty$$

   $X_{-\infty}(\omega) \in L^1(P)$ *and is P-a.s. finite.*

3. *When* (X_t) *is a backward martingale the assumption 3 always holds,* $X_t = E(X_0|\mathcal{F}_t)$ *for* $t \leq 0$ *is uniformly integrable and*

$$X_{-\infty}(\omega) = E(X_0|\mathcal{F}_{-\infty})(\omega)$$

   that is the martingale property hold in the extended time index set $(-\mathbb{N}) \cup \{-\infty\}$.

Proof By copying the proof of the forward convergence theorem for $U_{(a,b)}([t,0])$ the number of upcrossing the supermartigale $(-X_t)$ in the interval $[t,0]$, with $(-t) \in \mathbb{N}$, $a < b \in \mathbb{R}$, we get

$$E_P(U_{[a,b]}([t,0])) \leq \frac{E_P((a+X_0)^-)}{(b-a)} \leq \frac{(|a| + E_P(|X_0|)}{(b-a)}$$

which implies as in the forward case

$$X_{-\infty}(\omega) := \limsup_{t\to-\infty} X_t(\omega) = \lim \inf_{t\to-\infty} X_t(\omega) \quad P\text{-almost surely}$$

When X_t is a backward martingale by Fatou lemma

$$E(|X_\infty|) = E(\liminf_t |X_t|) \leq \liminf_t E(|X_t|) = \liminf_t E\big(|E(X_0|\mathcal{F}_t)|\big)$$
$$\leq \liminf_t E\big(E(|X_0||\mathcal{F}_t)|\big) \leq E(|X_0|) < \infty$$

where we used Jensen inequality. In the submartingale case, since $X_t \leq E(X_0|\mathcal{F}_t)$ when $t < 0$, we have only

$$X_t^+ \leq E(X_0|\mathcal{F}_t)^+ \leq E(X_0^+|\mathcal{F}_t)$$
$$X_t^- \geq E(X_0|\mathcal{F}_t)^-$$

In this case to complete the Fatou lemma argument we need the upper bound (3).

Finally let $A \in \mathcal{F}_{-\infty} \subseteq \mathcal{F}_{-t}\ \forall t \le 0$.' Since $X_t = E_P(X_0|\mathcal{F}_t)$ is uniformly integrable, when we use the definition of conditional expectation we can take the limit inside the exepctation getting

$$E_P(X_0\mathbf{1}_A) = E_P(X_t\mathbf{1}_A) \to E_P(X_\infty\mathbf{1}_A)$$

which means $X_{-\infty} = E_P(X_t|\mathcal{F}_{-\infty})$.

Exercise 7. *(martingale proof of the law of large numbers)*

Let $(X_n(\omega) : n \in \mathbb{N})$ independent and identically distributed random variables with $X_1 \in L^1(P)$, and let

$$S_n(\omega) = X_1(\omega) + \cdots + X_n(\omega)$$

We introduce the backward filtration $(\mathcal{T}_{-n} : n \ge 1)$ with $\mathcal{T}_{-n} = \sigma(S_n, S_{n+1}, \dots)$, and the backward martingale $M_{-n} = E_P(X_1|\mathcal{T}_{-n})$.

Note that the information in $\mathcal{T}_{-n}$ is decreasing as $n \to \infty$.

Since by symmetry the joint laws of the pairs (S_n, X_k) coincide for $k = 1, \dots, n$

$$E_P(X_k|\mathcal{T}_{-n}) = E_P(X_k|\sigma(S_n)) = E_P(X_1|\sigma(S_n)) \text{ and}$$
$$S_n = E_P(X_1 + \cdots + X_n|\sigma(S_n)) = \sum_{k=1}^{n} E_P(X_k|\sigma(S_n)) = nE_P(X_1|\sigma(S_n))$$

which gives $M_{-n}(\omega) = n^{-1}S_n(\omega)$. By martingale backward convergence theorem exists $M_{-\infty}(\omega) = \lim_{n\to\infty} n^{-1}S_n(\omega)$. Note that $M_{-\infty}(\omega)$ is $\mathcal{T}_{-n}$-measurable for all n, which means it belongs to the tail σ-algebra

$$\mathcal{T}_{-\infty} = \bigcap_{n\in\mathbb{N}} \mathcal{T}_{-n}$$

By Kolmogorov 0-1 law $M_{-\infty}(\omega) \equiv c$ is deterministic, and necessarily $c = E(X_1)$.

6.5 Exchangeability and De Finetti's theorem

Definition 18. *The random sequence $(X_t)_{t\in\mathbb{N}}$ with values in $(S, \mathcal{S})$ is infinitely exchangeable if for all n, $t_1, \dots, t_n \in \mathbb{N}$ and π permutation of $\{1, \dots, n\}$, we have $(X_{t_1}, \dots, X_{t_n})$ and $(X_{t_{\pi(1)}}, \dots, X_{t_{\pi(n)}})$ have the same distribution under P.*

Note that for infinitely exchangeable real valued random variables $(X_n)_{n\in\mathbb{N}}$, it follows as in the i.i.d. case that

$$M_{-n}(\omega) = n^{-1}S_n(\omega) := E(X_1|\mathcal{T}_{-n})$$

is a backward martingale with almost sure limit

$$M_{-\infty}(\omega) = E(X_1|\mathcal{T}_{-\infty})(\omega) \text{ as } n \to \infty$$

However without assuming independence, the tail σ-algebra $\mathcal{T}$ is non-trivial, and $M_{-\infty}(\omega)$ is truly random. This observation leads us to an important result: De Finetti's theorem.

Definition 19. *A random sequence $(X_k(\omega))$ with values in $(S, \mathcal{S})$ is conditionally independent and identically distributed given the σ-algebra $\mathcal{G}$ if for every n, $k_1, \dots, k_n$, $A_1 \dots A_n \in \mathcal{S}$.*

$$P(X_{k_1} \in A_1, \dots, X_{k_n} | \mathcal{G})(\omega) = \prod_{i=1}^{n} P(X_1 \in A_i | \mathcal{G})(\omega) \quad P \ a.s$$

By taking expectation we see that a conditionally i.i.d. sequence is infinitely exchangeable.

Theorem 7. *(De Finetti) When the sequence (X_k) takes value in a Borel space $(S, \mathcal{S})$, the opposite implication holds conditionally on a tail σ-algebra $\mathcal{T}_{-infty}$ to be defined below.*

Proof We introduce the empirical random measure

$$\mu_n(dx; \omega) = n^{-1} \sum_{i=1}^{n} \mathbf{1}(X_i(\omega) \in dx)$$

which generates the σ-algebra $\sigma(\mu_n) = \sigma\{\mu_n(A) : A \in \mathcal{S}\} \subseteq \mathcal{F}$.

Note that $\sigma(\mu_n) \subseteq \sigma(X_1, \dots, X_n)$ but for $n > 1$ it is strictly smaller since it forgets the time-order of the random variables.

We introduce the decreasing sequence of σ-algebrae

$$\mathcal{T}_{-n} := \bigvee_{k \geq n} \sigma(\mu_k) \downarrow \mathcal{T}_{-\infty} = \bigcap_{n \in \mathbb{N}} \mathcal{T}_{-n} \text{ , the tail } \sigma\text{-algebra .}$$

Fix $k \in \mathbb{N}$ and a bounded measurable test function $f(x_1, \dots, x_k) \in \mathbb{R}$. For $n \geq k$ we use symmetry to compute $E_P(f(X_1, \dots, X_k) | \mathcal{T}_{-n})(\omega)$.

Define the random probability measure $\mu_n^{\circ} : \mathcal{S}^{\otimes k} \to [0, 1]$ as the regular version of the conditional probability of $(X_1, \dots, X_k)$ given $\sigma(\mu_n)$ (which exists since S is a Borel space). By symmetry,

$$\mu^{\circ k}(f; \omega) := \int_{S^k} f(x) \mu^{\circ k}(dx; \omega) = \frac{1}{n!} \sum_{\pi} f(X_{\pi(1)}(\omega), \dots, X_{\pi(k)}(\omega))$$

$$= \frac{(n-k)!}{n!} \sum_{1 \leq i_1, \dots, i_k \leq n \text{ distinct}} f(X_{i_1}, X_{i_2}, \dots, X_{i_k})$$

where the sum is taken over all permuations π of $\{1, \dots, n\}$. Note that $\mu^{\circ k}(dx; \omega)$ is $\sigma(\mu_n)$-measurable, since it depends only on the values taken by $\{X_1(\omega), \dots, X_n(\omega)\}$ and not by their order. Note also that $\mu_n^{\circ k}(dx)$ is not a product measure since by taking permutations repeated indexes are excluded.

By using exchangeability we see that $(X_1, \dots, X_k, \mu_n)$ and $(X_{\pi(1)}, \dots, X_{\pi(k)}, \mu_n)$ have the same law for every permutation π of $\{1, \dots, n\}$ which implies

$$E_P(f(X_1, \dots, X_k | \sigma(\mu_n))(\omega) = E_P(f(X_{\pi(1)}, \dots, X_{\pi(k)}) | \sigma(\mu_n))(\omega)$$

By summing over π and dividing by the number of permutations it follows that

$$\mu_n^{\circ k}(f) = E_P(f(X_1, \dots, X_k) | \sigma(\mu_n)) = E_P(f(X_1, \dots, X_k) | \sigma(\mathcal{T}_{-n}))$$

Note first that $\mathcal{T}_{-n} = \sigma(\mu_n, X_{n+1}, X_{n+2}, \dots)$ since find $X_{n+1}(\omega)$ by comparing μ_n and μ_{n+1}.

Then from infinite exchangeability it follows that in law, for $m \in \mathbb{N}$

$$(X_1, X_2, \dots, X_n, X_{n+1}, X_{n+2}, \dots X_{n+m}) \stackrel{\mathcal{L}}{=} (X_{\pi(1)}, X_{\pi(2)}, \dots, X_{\pi(n)}, X_{n+1}, X_{n+2}, \dots X_{n+m})$$

for any permutation π of $\{1, \dots, n\}$.

Exercise 8. $(X_1, \dots, X_n)$ *and* $(X_{n+1}, X_{n+2}, \dots)$ *are conditionally independent given* $\sigma(\mu_n)$*,*

Solution *Let* $Y(\omega) \in \sigma(\mu_n)$ *with* $|Y(\omega)| \le c$*. Necessarily* $Y = f(X_1(\omega), \dots, X_n(\omega))$ *for some bounded symmetric and measurable* $f(x_1, \dots, x_n)$*.*

Let $g(x_1, \dots, x_n)$ *and* $h(y_1, \dots, y_m)$ *bounded measurable functions, not necessarily symmetric. Using the definition of conditional expectation together with exchangeability,*

$$\begin{aligned}
&E_P\Big(Y E_P\big(g(X_1, \dots, X_n)h(X_{n+1}, \dots, X_{n+m})|\sigma(\mu_n)\big)\Big) = \\
&E_P\Big(Y g(X_1, \dots, X_n)h(X_{n+1}, \dots, X_{n+m})\Big) = \\
&E_P\Big(Y g(X_{\pi(1)}, \dots, X_{\pi(n)})h(X_{n+1}, \dots, X_{n+m})\Big) = \\
&E_P\Big(Y \frac{1}{n!}\sum_{\pi} g(X_{\pi(1)}, \dots, X_{\pi(n)})h(X_{n+1}, \dots, X_{n+m})\Big) = \\
&E_P\Big(Y E_P(g(X_1, \dots, X_n|\sigma(\mu_n))h(X_{n+1}, \dots, X_{n+m})\Big) = \\
&E_P\Big(Y E_P(g(X_1, \dots, X_n|\sigma(\mu_n))E_P(h(X_{n+1}, \dots, X_{n+m})|\sigma(\mu_n))\Big)
\end{aligned}$$

where the sum is taken over the permutations of $\{1, \dots, n\}$*. Since* Y *is an arbitrary bounded and* $\sigma(\mu_n)$ *measurable,*

$$\begin{aligned}
&E_P\big(g(X_1, \dots, X_n)h(X_{n+1}, \dots, X_{n+m})|\sigma(\mu_n)\big) \\
&= E_P(g(X_1, \dots, X_n|\sigma(\mu_n))E_P(h(X_{n+1}, \dots, X_{n+m})|\sigma(\mu_n))
\end{aligned}$$

for all bounded measurable g *and* h*, which corresponds to conditional independence given* $\sigma(\mu_n)$

In other words, $\mathcal{T}_{-n}$ does not contain any information about the ordering of the first n-variables.

Therefore $M_{-n}(f) := \mu^{\circ k}(f)$ is a backward $(\mathcal{T}_{-n})$-martingale with P-a.s. limit $M_{-\infty}(f)$.

The σ-algebra $\mathcal{S}$ of the Borel space S is countably generated, this relation holds simultaneously for all bounded measurable functions f outside a P-null set.

Since $(X_1, \dots, X_k)$ takes value in a Borel space, the conditional probability has a regular version, there is a $\mathcal{T}_{-\infty}$-measurable probability kernel $\mu^{(k)}_{-\infty}(dx;\omega)$

on $S_1 \times \cdots \times S_k$ such that P almost surely for all bounded measutable f

$$M_{-\infty}(f,\omega) = E_P(f(X_1,\dots,X_k)|\sigma(\mathcal{T}_{-\infty}))(\omega) \tag{9}$$
$$= \int_{S_1,\dots,S_k} f(x_1,\dots,x_k)\mu^{(k)}_{-\infty}(dx_1,\dots dx_k;\omega) \tag{10}$$

For P almost all ω the family of finite dimensional distributions

$$\left\{\mu^{(k)}_{-\infty}(dx_1,\dots dx_k;\omega) : k \in \mathbb{N}\right\}$$

is consistent (check this) and by Kolmogorov construction they define a random measure $\mu_\infty(\cdot;\omega)$ on the space of sequences $(x_k : k \in \mathbb{N}) \subseteq S$.

We show that necessarily for almost all ω $\mu_\infty(\cdot,\omega)$ is the infinite product of identical measures.

Consider again $f(x_1,\dots,x_k)$, and let $\mu_n^{\otimes k}$ denote the k-fold product measure. Since

$$\mu_n^{\circ k}(f) = n^{-k} \sum_{1\le i_1,\dots,i_k\le n} f(X_{i_1},\dots,X_{i_k})$$

which includes also terms with repeated indexes, we have

$$\mu_n^{\circ k}(f) - \mu_n^{\otimes k}(f) =$$
$$\mu_n^{\circ k}(f)\left(1 - \frac{n!}{n^k(n-k)!}\right) + n^{-k} \sum_{i_1,\dots i_k : i_l = i_m \text{ for some } l\neq m} f(X_{i_1},\dots,X_{i_k})$$

where in the first term we collect the terms without repetead variables and in the second term at least one variable is repeated.

Therefore for fixed k,$\forall\omega$,

$$|\mu_n^{\circ k}(f) - \mu_n^{\otimes k}(f)|$$
$$\le \| f \|_\infty \left(1 - \prod_{l=0}^{k-1} \frac{(n-l)}{n} + n^{-k}\binom{k}{2} n^{k-1}\right) \to 0 \text{ as } n\to\infty$$

where $\| f \|_\infty = \sup_{x\in S}|f(x)|$.

Since for P-almost all ω, for every $A_1, A_2 \dots \in \mathcal{S}$

$$\mu_n^{\circ 1}(A_1) = \mu_n^{\otimes 1}(A_1) \to \mu^{(1)}_{-\infty}(A_1)$$

and

$$\mu_{-n}^{\circ k}(A_1\times A_2\times\cdots\times A_k) \to \mu^{(k)}_{-\infty}(A_1\times A_2\times\cdots\times A_k)$$

as $n\to-\infty$, it follows also that

$$\mu_{-n}^{\otimes k}(A_1\times A_2\times\cdots\times A_k) = \prod_{i=1}^k \mu^{(1)}_{-n}(A_i) \to \prod_{i=1}^k \mu^{(1)}_{-\infty}(A_i)$$

with the same limit, so that

$$\mu^{(k)}_{-\infty} = (\mu^{(1)}_{-\infty})^{\otimes k}$$

is a product measure over $S_1 \times \cdots \times S_k$ and $\mu_{-\infty}$ is a product measure over S^∞. For bounded measurable functions $g_1, \ldots, g_k$

$$E_P(g_1(X_1)\ldots g_k(X_k)|\mathcal{T}_{-\infty})(\omega) = \prod_{i=1}^k \biggl\{ \int_S g_i(x) \mu_{-\infty}^{(1)}(dx, \omega) \biggr\}$$

By taking expectation we obtain,

$$E_P(g_1(X_1)\ldots g_k(X_k)) = \int_{\mathcal{M}(S)} \biggl\{ \prod_{i=1}^k \int_S g_i(x)\mu(dx) \biggr\} Q(d\mu)$$

where Q is the probability distribution of $\mu_{-\infty}(\omega)$ in $\mathcal{M}(S)$, the space of probability measures on $(S, \mathcal{S})$.

In other words, an infinitely exchangeable random sequence with values in a Borel space is a mixture of i.i.d. sequences □

Remark

When S is a separable metric space equipped with the Borel σ-algebra, since continuous functions are measurable, it follows directly that P almost surely $\mu_{-n}^{\otimes k} \stackrel{w}{\to} \mu_{-\infty}^{(k)}$ in the sense of weak convergence of probabilities. Note we did not need to check tightness because the limiting measure was defined as a regular conditional probability.

Exercise 9. *De Finetti proved his theorem first in the simplest case when $S = \{0, 1\}$. In this case $\mathcal{M}(S) = [0, 1]$, and $S_n = (X_1 + \cdots + X_n)$ is a sufficient statistics For an infinitely exchangeable sequence of coin tosses, the limit $\mu(\omega) := \lim_{n\to\infty} n^{-1} S_n(\omega) \in [0, 1]$ exists almost surely, with distribution $Q(d\mu)$. Conditionally on $\sigma(\mu)$ the coin tosses are conditionally independent Bernoulli random variables with common random parameter $\mu(\omega)$. The measure $Q(d\mu)$ is the prior probability for the parameter μ. This theorem is the key to understand the Bayesian approach in statistical inference.*

7 Change of measure and Radon-Nikodym theorem

Definition 20. *Let μ and ν positive measures on the probability space $(\Omega, \mathcal{F})$.*

We say that ν is absolutely continuous with respect to μ, (also μ dominates ν) if for all $A \in \mathcal{F}$ $\mu(A) = 0 \Longrightarrow \nu(A) = 0$. In this case we use the notation $\nu \ll \mu$.

Sometimes we need absolute continuity with respect to some sub-σ-algebra $\mathcal{G} \subseteq \mathcal{F}$. We say that μ dominates ν on $\mathcal{G}$ and denote $\nu \stackrel{\mathcal{G}}{\ll} \mu$.

When both $\mu \ll \nu$ and $\nu \ll \mu$ we say that the measures are equivalent (that is they have the same null sets) and denote $\mu \sim \nu$.

Lemma 6. *Let $Q \ll P$ be probability measures on the space $(\Omega, \mathcal{F})$. Then for all $\varepsilon > 0$ there is $\delta > 0$ such that for $A \in \mathcal{F}$ $P(A) < \delta \Longrightarrow Q(A) < \varepsilon$*

Proof Otherwise there is $\varepsilon > 0$ and a sequence $(A_n : n \in \mathbb{N}) \subseteq \mathcal{F}$ with $P(A_n) \le 2^{-n}$ and $Q(A_n) \ge \varepsilon > 0$ By Borel Cantelli lemma $P(\limsup A_n) = 0$, while by reverse Fatou lemma

$$Q(\limsup A_n) \ge \limsup Q(A_n) \ge \varepsilon > 0$$

which is in contradiction with the assumption $Q \ll P$ □

Theorem 8. *(Radon-Nikodym) Letμ and ν σ-finite positive measures on the measurable space $(\Omega, \mathcal{F})$. When $\nu \ll \mu$, there is a measurable function $Z : (\Omega, \mathcal{F}) \to (\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$, such that the change of measure formula holds*

$$\nu(A) = \int_\Omega Z(\omega)\mathbf{1}_A(\omega)\mu(d\omega) \quad \forall A \in \mathcal{F}$$

Proof Since both μ and ν are σ-finite, there is a countable partition $\Omega = \bigcup_{n\in\mathbb{N}} \Omega_n$ of disjoint measurable sets, such that both $\mu(\Omega)_n < \infty$ and $\nu(\Omega)_n < \infty$. By taking $P_n(d\omega) = \mu(d\omega)/\mu(\Omega_n)$ and $Q_n(d\omega) = \nu(d\omega)/\nu(\Omega_n)$ on each Ω_n, we see that it is enough to prove the theorem for probability measures $Q \ll P$.

We assume first that $\mathcal{F}$ is *countably generated* (we say also *separable*) $\mathcal{F} = \sigma(F_n : n \in N)$ where $\{F_n\}_{n\in\mathbb{N}} \subseteq \mathbb{N}$. This is the case when $(\Omega, \mathcal{F})$ is a Borel space. We will drop this assumption later.

Consider the filtration $\{\mathcal{F}_n\}$ where $\mathcal{F}_n = \sigma(F_1, \dots, F_n)$, with $\mathcal{F} = \bigvee_{n\in\mathbb{N}} \mathcal{F}_n$.

For each n, by taking intesections of $F_1, \dots F_n$, we find a $\mathcal{F}_n$-measurable partition of Ω $\{A_1^{(n)}, \dots, A_{m_n}^{(n)}\}$ with $\mathcal{F}_n = \sigma(A_k^{(n)} : k = 1, \dots, m_n)$.

We define the $\mathcal{F}_n$ measurable random variable

$$Z_n(\omega) = \sum_{k=1}^{m_n} \frac{Q(A_k^{(n)})}{P(A_k^{(n)})}\mathbf{1}(\omega \in A_k^{(n)})$$

with the convention that $0/0 = 0$ (or if you like $0/0 = 1$, it does not matter).

Note that by absolute continuity, $Q(A_k^{(n)}) = 0$ when $P(A_k^{(n)}) = 0$ so that $Z_n(\omega)$ takes values in $[0, +\infty)$.

It follows that $Q(A) = E_P(Z_n\mathbf{1}_A)$ $\forall A \in \mathcal{F}_n$.

On fact it is enough to check this property for some $A = A_k^{(n)}$ $k \in \{1, \dots, m_n\}$, since these sets generate the σ-algebra $\mathcal{F}_n$. But this follows directly from the definition.

Note that for every $\mathcal{F}_n$-measurable random variable $X(\omega)$ (which is necessarily a simple r.v.) it follows directly that

$$E_Q(X) = E_P(XZ_n)$$

Note also that $E_P(Z_n) = Q(\Omega) = 1$.

The process $(Z_n(\omega))_{n\in\mathbb{N}}$ is a $(P, \{\mathcal{F}_n\})$-martingale. We have seen that (Z_n) is adapted and it is P-integrable since it takes finitely many finite values.

For all $A \in \mathcal{F}_n$ also $A \in \mathcal{F}_{n+1}$, so that

$$E_P(Z_n\mathbf{1}_A) = Q(A) = E_P(Z_n\mathbf{1}_A)$$

which by definition of conditional expectation means

$$E_P(Z_{n+1}|\mathcal{F}_n)(\omega) = Z_n(\omega).$$

Since $(Z_n(\omega))$ is a non-negative martingale, in particular it is a supermartingale bounded from below, and by Doob forward martingale convergence theorem it follows that P almost surely exists

$$Z_\infty(\omega) = \lim_{n\to\infty} Z_n(\omega)$$

and $Z_\infty \in L^1(\Omega, \mathcal{F}, P)$. In order to define $Z(\omega)$ for all ω we take the lim sup.

In order to show that $Q(A) = E_P(Z_\infty \mathbf{1}_A)\ \forall A \in \mathcal{F}$, since the sets F_n generate the σ-algebra, it is enough to show that $Q(F_n) = E_P(Z_\infty \mathbf{1}_{F_n})\ \forall n$.

Since $Q(F_n) = E_P(Z_m F_n)$ for all $m \geq n$, in order to show that

$$E_P(Z_\infty F_n) = \lim_{m\to\infty} E_P(Z_m F_n) = Q(F_n)$$

we need to check uniform P-integrability for the martingale (Z_n).

Since $Q \ll P$, by lemma 6 for given $\varepsilon > 0$ we can find $\delta > 0$ such that for $A \in \mathcal{F}$ and $P(A) < \delta$ follows $Q(A) < \varepsilon$.

By Chebychev inequality

$$P(Z_n > K) < K^{-1} E_P(Z_n) = K^{-1} \quad \forall n$$

Choose $K > \delta^{-1}$. Since $\{\omega : Z_n(\omega) > K\} \in \mathcal{F}_n$, by the change of measure formula

$$\sup_n E_P(Z_n \mathbf{1}(Z_n > K)) = \sup_n Q(Z_n > K) < \varepsilon$$

which is the UI-condition:

$$\lim_{K\to\infty} \sup_n E_P(Z_n \mathbf{1}(Z_n > K)) = 0$$

So far we have proved the R-N theorem for countably generated σ-algebrae. We extend the proof by using convergence of generalized sequences.

We recall this concept from topology:

Definition 21. *In a topological space $(E, \mathcal{T})$ a net is a generalized sequence $(x_\alpha : \alpha \in \mathcal{I})$ indexed by a directed set, that is a partially ordered set $(\mathcal{I}, \leq)$ such that for every two elements $\alpha, \beta \in \mathcal{I}$ there is an element $\alpha \vee \beta$*

$$\alpha \vee \beta \geq \alpha, \alpha \vee \beta \geq \beta. \gamma \geq \alpha \text{ and } \alpha \geq \beta \Longrightarrow \gamma \geq \alpha \vee \beta$$

We say that $x_\alpha \to x \in E$ when for every open set $U \ni x$ there is an element $\bar{\alpha}$ such that $x_\alpha \in U$ for all $\alpha \geq \bar{\alpha}$.

We consider now the partially order set

$$\mathbb{G} := \left\{ \mathcal{G} \subseteq \mathcal{F} : \mathcal{G} \text{ is a countably generated } \sigma\text{-algebra} \right\}$$

where $\mathcal{F}$ is not assumed to be separable. Here the ordering relation is the inclusion $\subseteq$. Note that $\mathcal{G}' \vee \mathcal{G}'' := \sigma(\mathcal{G}', \mathcal{G}'')$ is a separable sub σ-algebra.

For each $\mathcal{G} \in \mathbb{G}$ we have shown that there is a random variable $0 \leq Z_{\mathcal{G}}(\omega) \in L^1(\Omega, \mathcal{G}, P)$ such that the change of variable formula holds in $\mathcal{G}$:

$$Q(A) = E_P(Z_{\mathcal{G}} \mathbf{1}_A) \quad \forall A \in \mathcal{G}$$

We show that $(Z_{\mathcal{G}} : \mathcal{G} \in \mathbb{G})$ is a Cauchy net in $L^1(\Omega, \mathcal{F}, P)$, and by completeness it has a limit $Z \in L^1(\Omega, \mathcal{F}, P)$.

By Cauchy net we mean the following: for all $\varepsilon > 0$ there is a $\overline{\mathcal{G}} \in \mathbb{G}$ such that if $\mathcal{G}' \subseteq \overline{\mathcal{G}}$, $\mathcal{G}'' \subseteq \overline{\mathcal{G}}$, $\mathcal{G}', \mathcal{G}'' \in \mathbb{G}$, then

$$E_P\big(|Z_{\mathcal{G}'} - Z_{\mathcal{G}''}|\big) < \varepsilon$$

By the triangle inequality this it is equivalent to

$$E_P\big(|Z_{\bar{\mathcal{G}}} - Z_{\mathcal{G}'}|\big) < \varepsilon$$

If $(Z_{\mathcal{G}})$ was not a Cauchy net we would find some $\varepsilon > 0$ and a sequence $(\mathcal{G}_n : n \in \mathbb{N}) \subseteq \mathbb{G}$ such that $\mathcal{G}_n \subseteq \mathcal{G}_{n+1}$ and

$$E_P\big(|Z_{\mathcal{G}_n} - Z_{\mathcal{G}_{n+1}}|\big) \geq \varepsilon > 0$$

Let $\mathcal{G}_\infty = \bigvee_{n\in\mathbb{N}} \mathcal{G}_n$. $\mathcal{G}_\infty \in \mathbb{G}$ and by the previous argument $(Z_{\mathcal{G}_n} : n)$ would be an uniformly integrable martingale in the filtration $\{\mathcal{G}_n\}$, which necessarly is convergent in $L^1(P)$, giving a contradiction.

In a complete metric space (E, d) every Cauchy net $(x_\alpha : \alpha \in \mathcal{I})$ is convergent, that is there is an element $x^* \in E$ such that forall ε $\exists\overline{\alpha}$ with $d(x^*, x_\alpha) \leq \varepsilon$ $\forall\alpha \geq \overline{\alpha}$.

We sketch the proof: for every n let $\overline{\alpha}_n$ such that $d(x, x_\alpha) \leq n^{-1}$ $\forall\alpha \geq \overline{\alpha}_n$, and we can choose $\alpha_n \geq \alpha_{n-1}$.

Therefore $\widetilde{x}_n = x_{\alpha_n}$ is a Cauchy sequence and it as a limit $x^* \in E$, which by definition it is also the limit of the net (x_α).

Therefore the generalized Cauchy sequence $(Z_{\mathcal{G}} : \mathcal{G} \in \mathbb{G})$ has necessarily a limit $Z_\infty(\omega) \in L^1(\Omega, \mathcal{F}, P)$.

We next check the change of measure formula.

Let $A \in \mathcal{F}$ and $\mathcal{G} \in \mathbb{G}$ such that

$$E_P(|Z_\infty - Z_{\mathcal{G}'}|) < \varepsilon$$

for all $\mathcal{G}' \supseteq \mathcal{G}$, $\mathcal{G}' \in \mathbb{G}$.

Let $\widetilde{\mathcal{G}} := \sigma(\mathcal{G} \vee F) \in \mathbb{G}$.

Since

$$Q(A) = E_P(Z_{\widetilde{\mathcal{G}}} \mathbf{1}_A)$$

we have

$$\Big| E_P(Z_\infty \mathbf{1}_A) - Q(A) \Big| \leq E_P\Big(\big|Z_\infty - Z_{\widetilde{\mathcal{G}}}\big| \Big) < \varepsilon$$

where $\varepsilon > 0$ is arbitrarily small $\square$

8 The Likelihood ratio process

9 Doob optional sampling and optional stopping theorems

Lemma 7. *Let $(X_t : t \in \mathbb{N})$ a supermartingale and $0 \leq \tau(\omega) \leq k$ a bounded stopping time.*

Then $E(X_k|\mathcal{F}_\tau)(\omega) \leq X_\tau$.

Proof For $A \in \mathcal{F}_\tau$ by definition $A \cap \{\tau = t\} \in \mathcal{F}_t$. By using the supermartingale property

$$E_P(X_k \mathbf{1}_A) = \sum_{t=0}^{k} E_P\big(X_k \mathbf{1}(A \cap \{\tau = t\})\big) \le \sum_{t=0}^{k} E_P\big(X_t \mathbf{1}(A \cap \{\tau = t\})\big) = E_P(X_\tau \mathbf{1}_A)$$

Theorem 9. *Let $(M_t : t \in \mathbb{N})$ an UI martingale, and τ a stopping time. Then*

$$E_P(M_\infty | \mathcal{F}_\tau)(\omega) = M_\tau(\omega)$$

Proof Since $\mathcal{F}_{\tau \wedge k} \subseteq \mathcal{F}_k$, $k \in \mathbb{N}$ and (M_t) is an UI-martingale

$$E_P(M_\infty | \mathcal{F}_{\tau\wedge k}) = E_P(E_P(M_\infty|\mathcal{F}_k)|\mathcal{F}_{\tau\wedge k}) = E_P(M_k|\mathcal{F}_{\tau\wedge k})$$

Let's assume that $M_\infty(\omega) \ge 0$, otherwise we work with M_∞^+, M_∞^- separately.
For $A \in \mathcal{F}_\tau$,

$$E_P\big(M_\infty \mathbf{1}_{A \cap \{\tau \le k\}}\big) = E_P\big(M_k \mathbf{1}_{A\cap\{\tau\le k\}}\big)$$

by the martingale property, since $A \cap \{\tau \le k\}$ is $\mathcal{F}_k$-measurable,

$$= E_P(M_{\tau\wedge k} \mathbf{1}_{A\cap\{\tau\le k\}}\big) = E_P(M_\tau \mathbf{1}_{A\cap\{\tau\le k\}}\big) =$$

where we used lemma 7 for the bounded stopping time $(\tau \wedge k) \le k$ together with the fact that $A \cap \{\tau \le k\}$ is also $\mathcal{F}_{(\tau\wedge k)}$-measurable. To check this, for all $t \in \mathbb{N}$ we have

$$A \cap \{\tau \le k\} \cap \{\tau \wedge k \le t\} = A \cap \{\tau \le k \wedge t\} \in \mathcal{F}_{(t\wedge k)} \subseteq \mathcal{F}_t$$

Since $\mathbf{1}(\tau(\omega) \le k) \uparrow \mathbf{1}(\tau(\omega) < \infty)$ as $k \uparrow \infty$, by the monotone convergence theorem it follows

$$E_P(M_\infty \mathbf{1}_A \mathbf{1}(\tau < \infty)) = E_P(M_\tau \mathbf{1}_A \mathbf{1}(\tau < \infty))$$

and since $M_\tau \mathbf{1}(\tau < \infty)$ is $\mathcal{F}_\tau$-measurable this means

$$E(M_\infty|\mathcal{F}_\tau)(\omega)\mathbf{1}(\tau(\omega) < \infty) = M_\tau(\omega)\mathbf{1}(\tau(\omega) < \infty)$$

The result follows since

$$M_\infty(\omega)\mathbf{1}(\tau(\omega) = \infty) = M_\tau(\omega)\mathbf{1}(\tau(\omega) = \infty) \quad \square$$

Corollary 7. *Let $\tau(\omega) \ge \sigma(\omega)$ stopping times. and $(M_t : t \in \mathbb{N})$ an UI martingale.*

Then $\mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$ and

$$E_P(M_\tau|\mathcal{F}_\sigma) = M_\sigma \tag{11}$$

and by taking expectation $E_P(M_\tau) = E_P(M_0)$ for all stopping times τ.

Proof*: If $\mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$*

$$M_\sigma = E_P(M_\infty|\mathcal{F}_\sigma) = E_P(E_P(M_\infty|\mathcal{F}_\tau)|\mathcal{F}_\sigma) = E_P(M_\tau|\mathcal{F}_\sigma) \tag{12}$$

Corollary 8. *When M_t is an UI martingale, the stopped process M_t^τ is also an UI martingale in the filtration $(\mathcal{F}_t)$.*

Proof We have seen that if (M_t) is a martingale, the stopped process M_t^τ is a martingale, since it is the martingale transform of a bounded integrand.

Next note that since $(\tau(\omega) \wedge t) \uparrow \tau$ as $t \uparrow \infty$, $E(M_\infty|\mathcal{F}_\tau) = M_\tau$, and $\mathcal{F}_{\tau\wedge t} \subseteq \mathcal{F}_\tau$, it follows

$$E_P(M_\tau|\mathcal{F}_{\tau\wedge t}) = E_P(M_\infty|\mathcal{F}_{\tau\wedge t}) = M_{\tau\wedge t} = \big(M_t \mathbf{1}(\tau > t) + M_\tau \mathbf{1}(\tau \le t)\big) \in \mathcal{F}_t$$

Now since $\mathcal{F}_{\tau\wedge t} \subseteq \mathcal{F}_t$, we have also

$$E_P(M_\tau|\mathcal{F}_t) = E(M_\tau \mathbf{1}(\tau \le t)|\mathcal{F}_t) + E_P(M_\tau \mathbf{1}(\tau > t)|\mathcal{F}_t)$$

where on the right hand side $M_\tau \mathbf{1}(\tau \le t) \in \mathcal{F}_t$ since $M_\tau \in \mathcal{F}_\tau$.

Since $\mathcal{F}_{(\tau\vee t)} \supseteq \mathcal{F}_t$, we have

$$E_P(M_\tau \mathbf{1}(\tau > t)|\mathcal{F}_t) = E_P(M_{(\tau\vee t)} \mathbf{1}(\tau > t)|\mathcal{F}_t) =$$
$$E_P(M_{(\tau\vee t)}|\mathcal{F}_t)\mathbf{1}(\tau > t) = M_t \mathbf{1}(\tau > t)$$

So that

$$E_P(M_\tau|\mathcal{F}_t) = M_\tau \mathbf{1}(\tau \le t) + M_t \mathbf{1}(\tau > t) = M_{t\wedge\tau} = E_P(M_\tau|\mathcal{F}_{t\wedge t})$$

Exercise 10. *Since the stopped process can represented as a martingale transform of a bounded predictable integrand one would hope that martingale transforms with respect to a bounded predictable integrand preserves uniform integrability, but this is not true.*

In fact convergence in $L^1(P)$ sense of martingales is tricky. Cherny has constructed an uniformly integrable martingale $(X_t : t \in \mathbb{N})$ and a **bounded**-*predictable integrand $(H_t : t \in \mathbb{N})$, (that is $|H_t(\omega)| \le c$ for some constant), such that the martingale transform $(H \cdot X)_t$ is a martingale which is not bounded in $L^1(P)$ and therefore it is not uniformly integrable*

Let X_n a sequence of independent random variables, $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$ and $A_n \in \mathcal{F}_n$ defined as follows:

$$a_n = 2n, \quad b_n = \frac{2n}{2n^2 - n + 1}, \quad p_n = \frac{n-1}{2n^2} \quad n \in \mathbb{N} X_0 = X_1 = 1, A_1 = \Omega,$$
$$A_{n+1} = \{\omega : X_{n+1} = a_1 \cdot \dots \cdot a_{n+1}\}$$
$$P(X_{n+1} = a_2 \cdot \dots a_n a_{n+1}|A_n) = p_{n+1}$$
$$P(X_{n+1} = a_2 \cdot \dots a_n b_{n+1}|A_n) = 1 - p_{n+1}$$
$$P(X_{n+1} = X_n|A_n^c) = 1$$

Note that the process X_n stops the first time the event A_n^c appears, and X_n is a martingale since

$$E(X_{n+1}|\mathcal{F}_n) = X_n\Big(\mathbf{1}_{A_n^c} + \mathbf{1}_{A_n}\big\{a_{n+1}p_{n+1} + b_{n+1}(1 - p_{n+1})\big\}\Big) = X_n$$

For $n < m$

$$E(|X_m - X_n|) = E(|X_m - X_n|\mathbf{1}_{A_n}) = E(|X_m - X_n|\mathbf{1}_{A_{n+1}}) + E(|X_m - X_n|\mathbf{1}_{A_n}\mathbf{1}_{A_{n+1}^c}) =$$

One can check by induction that $Y_{m,n} := (X_m - X_n)\mathbf{1}_{A_{n+1}} > 0$ for $m > n$.

$$
\begin{aligned}
&Y_{n+1,n} = (X_{n+1} - X_n)\mathbf{1}_{A_{n+1}} = a_1 \dots a_n(a_{n+1} - 1)\mathbf{1}_{A_{n+1}} \geq 0,\\
&(X_m - X_n)\mathbf{1}_{A_{n+1}} = (X_m - X_{m-1} + X_{m-1} - X_n)\mathbf{1}_{A_{n+1}} =\\
&Y_{m-1,n} + (X_m - X_{m-1})\mathbf{1}_{A_{m-1}} =\\
&Y_{m-1,n} + a_2 \dots a_{m-1}\Big(\mathbf{1}_{A_m}(a_m - 1) + \mathbf{1}_{A_{m-1}}\mathbf{1}_{A_m^c}(b_m - 1)\Big)
\end{aligned}
$$

Now when $\omega \in A_{m-1}^c$ the second term is zero and the first term is non-negative by induction. When $\omega \in A_{m-1}$ this gives

$$
= a_2 \dots a_{m-1}\Big(1 + \mathbf{1}(A_m)(a_m - 1) + \mathbf{1}_{A_m^c}(b_m - 1)\Big) \geq 0
$$

Using the positivity property of $Y_{m,n}$,

$$
E(|X_m - X_n|\mathbf{1}_{A_{n+1}}) = E((X_m - X_n)\mathbf{1}_{A_{n+1}}) = E((X_{n+1} - X_n)\mathbf{1}_{A_{n+1}}) = E(|X_{n+1} - X_n|\mathbf{1}_{A_{n+1}})
$$

so that

$$
\begin{aligned}
&E(|X_m - X_n|) = E(|X_m - X_n|\mathbf{1}_{A_{n+1}}) + E(|X_m - X_n|\mathbf{1}_{A_n}\mathbf{1}_{A_{n+1}^c}) =\\
&E((X_m - X_n)\mathbf{1}_{A_{n+1}}) + E(|X_{n+1} - X_n|\mathbf{1}_{A_n}\mathbf{1}_{A_{n+1}^c})\\
&= E((X_{n+1} - X_n)\mathbf{1}_{A_{n+1}}) + E(|X_{n+1} - X_n|\mathbf{1}_{A_n}\mathbf{1}_{A_{n+1}^c}) \text{ by the martingale property,}\\
&= E(|X_{n+1} - X_n|\mathbf{1}_{A_n}\mathbf{1}_{A_{n+1}}) + E(|X_{n+1} - X_n|\mathbf{1}_{A_n}\mathbf{1}_{A_{n+1}^c}) =\\
&E(|X_{n+1} - X_n|\mathbf{1}_{A_n}) = a_2 \dots a_n \times p_2 \dots p_n \times \big((a_{n+1} - 1)p_{n+1} + (1 - b_{n+1})(1 - p_{n+1})\big) =\\
&a_2 \dots a_n p_2 \dots p_n \times \big(1 - b_{n+1} + (a_{n+1} + b_{n+1} - 2)p_{n+1}\big)\\
&\leq a_2 \dots a_n p_2 \dots p_n(a_{n+1}p_{n+1} + 1) = \frac{1}{n}\left(\frac{n}{n+1} + 1\right) \leq 2/n
\end{aligned}
$$

therefore X_n is a Cauchy sequence and it converges in $L^1(P)$, which means that it is an UI martingale.

Consider now the martingale transform $(H \cdot X)_t$ of the bounded deterministic integrand

$$
H_n = \mathbf{1}(n \text{ is even })
$$

For $m > n$,

$$
\begin{aligned}
&E\left(\left|\mathbf{1}_{A_{2n}}\mathbf{1}_{A_{2n+1}^c}(H \cdot X)_{2m}\right|\right) = E\left(\mathbf{1}_{A_{2n}}\mathbf{1}_{A_{2n+1}^c}\sum_{k=1}^{n}(X_{2k} - X_{2k-1})\right)\\
&\geq E\left(\mathbf{1}_{A_{2n}}\mathbf{1}_{A_{2n+1}^c}(X_{2n} - X_{2n-1})\right),
\end{aligned}
$$

since the remaining terms are non-negative on the event $\mathbf{1}_{A_{2n}}\mathbf{1}_{A_{2n+1}^c}$,

$$
= p_2 \dots p_{2n}(1 - p_{2n+1})a_2 \dots a_{2n-1}(a_{2n} - 1) \geq \frac{1}{4}p_2 \dots p_{2n}a_2 \dots a_{2n} = \frac{1}{8n}
$$

We have

$$
\Omega = A_1^c \cup (A_1 \cap A_2^c) \cup \dots \cup (A_{2m} \cap A_{2m+1}^c) \cup A_{2m+1}
$$

where the union is taken over disjoint sets,

$$E_P\left(\left|(H\cdot X)_{2m}\right|\right) \ge \sum_{n=1}^{m} E_P\left(\mathbf{1}_{A_{2n}}\mathbf{1}_{A^c_{2n+1}}\left|(H\cdot X)_{2m}\right|\right) \ge \sum_{n=1}^{m}\frac{1}{8n} \to \infty$$

as $m\to\infty$, the martingale $(H\cdot X)_n$ is not bounded in $L^1(P)$.

Corollary 9. *Let $(X_t : t \in \mathbb{N})$ an UI submartingale with Doob decomposition*

$$X_t = X_0 + M_t + A_t$$

where M_t is a martingale and A_t is a predictable non-decreasing process with $M_0 = A_0 = 0$.

Then

1. *(M_t) is an UI-martingale and $E_P(A_\infty) < \infty$.*
2. *For every stopping time τ*

$$E(X_\infty|\mathcal{F}_\tau)(\omega) \ge X_\tau(\omega)$$

Proof By Doob forward martingale convergence theorem

$$\exists X_\infty = \lim_{t\to\infty} X_t(\omega)$$

P-almost surely and in $L^1(P)$ sense. By monotonicity $A_t(\omega)\to A_\infty(\omega)$ P-a.s. Since $E(M_t) = 0 \forall t\in\mathbb{N}$, by the bounded convergence theorem

$$E(A_\infty) = \lim_{n\to\infty} E(X_t - X_0) = E_P(X_\infty) - E_P(X_0)$$

which means that $A_t \to A_\infty$ also in $L^1(P)$ sense.

Therefore

$$M_t \to M_\infty = X_\infty - X_0 - A_\infty$$

P-a.s. and in $L^1(P)$.

For a stopping time τ, we have since M_t is an UI-martingale

$$E_P(X_\infty|\mathcal{F}_\tau) = X_0 + E_P(M_\infty|\mathcal{F}_\tau) + E_P(A_\infty|\mathcal{F}_\tau) = X_0 + M_\tau + A_\tau + E_P(A_\infty - A_\tau|\mathcal{F}_\tau)$$

where the last term on the right hand side is non-negative □

Lemma 8. *Let $(X_t(\omega) : t\in\mathbb{N})$ be a non-negative martingale. Since it is non-negative is automatically bounded in L^1 and by Doob convergence theorem exists $\lim_{t\to\infty} X_t(\omega) = X_\infty(\omega)$ P almost surely with $X_\infty\in L^1(P)$. Then X_t is uniformly integrable if and only if $E(X_\infty) = E(X_0)$*

Proof

Necessity follows from the characterization of L^1-convergence. For sufficiency, by Fatou lemma for $A\in\mathcal{F}_t$

$$E_P(X_\infty\mathbf{1}_A) \le \liminf_{T\to\infty} E(X_T\mathbf{1}_A) = E(X_t\mathbf{1}_A)$$

which gives the supermartingale property at $T=\infty$:

$$E_P(X_\infty|\mathcal{F}_t) \le X_t$$

Now by assumption

$$0 = E_P(X_t - X_\infty) = E_P\big(X_t - E_P(X_\infty|\mathcal{F}_t)\big)$$

which means $X_t = E_P(X_\infty|\mathcal{F}_t)$ P almost surely □

10 Martingale maximal inequalities

For a process $(X_t : t \in T)$, $T = \mathbb{R}$ or $\mathbb{N}$ we define the running maximum

$$X_t^* = \max_{0 \le s \le t} X_s(\omega)$$

Theorem 10. *Let $(X_s : s = 1, \dots, T)$ a $\{\mathcal{F}_t\}$-submartingale with $X_t(\omega) \ge 0$ P a.s. $\forall t$.Then for $c > 0$*

$$cP\big(X_T^* \ge c\big) \le E_P\big(X_T \mathbf{1}(X_T^* > c)\big) \le E_P\big(X_T\big)$$

Proof *Let $A := \{\omega : X_T^*(\omega) \ge c\}$ and*

$$A_t := \big\{\omega : X_1(\omega) < c, \dots, X_{t-1}(\omega) < c, X_t(\omega) \ge c\big\}$$

$A = \bigcup_{t=1}^T A_t$ *with $A_t \cap A_s = \emptyset$ for $s \neq t$.*

By the submartingale property

$$E_P(X_T \mathbf{1}_A) = \sum_{s=1}^T E_P(X_T \mathbf{1}_{A_s}) \ge$$
$$\sum_{s=1}^T E_P(X_s \mathbf{1}_{A_s}) \ge c \sum_{s=1}^T P(A_s) = cP(A)$$

Lemma 9. *Let $X(\omega) \ge 0, Y(\omega) \ge 0$ random variables with $Y \in L^p(\Omega, \mathcal{F}, P)$, $p > 1$ for which*

$$cP(X > c) \le E_P(Y \mathbf{1}(X > c)), \quad c > 0$$

then

$$\parallel X \parallel_p \le q \parallel Y \parallel_p \qquad \text{with } \frac{1}{p} + \frac{1}{q} = 1$$

Proof Assume first that $X \in L^p$.By using Fubini's theorem

$$E_P(X^p) = \int_\Omega \bigg(\int_0^{X(\omega)} pt^{p-1} dt \bigg) P(d\omega) = \int_0^\infty P(X \ge t) pt^{p-1} dt \le$$
$$\frac{p}{p-1} \int_0^\infty tP(X \ge t)(p-1)t^{p-2} dt \le q \int_0^\infty E_P\big(Y \mathbf{1}(X \ge t)\big)(p-1)t^{p-2} dt \le$$
$$qE_P\bigg(Y \int_0^{X(\omega)} (p-1)t^{p-2} dt\bigg) = qE_P(YX^{p-1})$$
$$(\text{ Hölder }) \ \le qE_P(Y^p)^{1/p} E_P(X^{q(p-1)})^{1/q} = q \parallel Y \parallel_p \parallel X \parallel_p$$

which gives

$$\parallel X \parallel_p^{p(1-1/q)} = \parallel X \parallel_p \le q \parallel Y \parallel_p$$

Without assuming that $X \in L^p$, take the truncated r.v.

$$X^{(n)}(\omega) := X(\omega) \wedge n \uparrow X(\omega) \text{ as } n \uparrow \infty$$

Note that $\{\omega : X(\omega) \wedge n \ge c\} = \emptyset$ for $n < c$,

and for $n \ge c$, $\{\omega : X(\omega) \wedge n \ge c\} = \{\omega : X(\omega) \ge c\}$ and the assumption of this lemma holds for $X^{(n)}(\omega)$. The result follows from the monotone convergence theorem as $n \uparrow \infty$ $\square$

Theorem 11. *(Doob's L^p maximal inequality) Let $(M_t : t = 1 \ldots, T)$ a martingale with $M_t \in L^p$ $\forall t$. Then for $1 < p < \infty$*

$$\| M_T^* \|_p \leq q \| M_T \|_p$$

Proof $|M_t|$ is a submartingale, by the maximal inequality

$$cP\big(|M_T^*| > c\big) \leq E_P\big(|M_T| \mathbf{1}(|M_T^*| > c)\big)$$

and we to apply the previous result with $X = |M_T^*|$ and $Y = |M_T|$.

Corollary 10. *When $(M_t : t \in \mathbb{N})$ is a martingale in $L^2(P)$, we obtain*

$$E_P\big((M_T^*)^2\big) \leq 4E_P\big(M_T^2\big) = 4\bigg\{ E_P(M_0^2) + E_P\big(\langle M, M\rangle\big)\bigg\}$$

Theorem 12. *(Kakutani) On a probability space $(\Omega, \mathcal{F}, P)$ let $(X_t : t \in \mathbb{N})$ P-independent random variables with $X_t(\omega) \geq 0$ and $E_P(X_t) = 1$.*

Let $\mathcal{F}_t = \sigma(X_1, \ldots, X_t)$ and

$$M_t = X_1 X_2 \ldots X_t, \quad a_t = \big\{E(\sqrt{X_t})\big\} \in (0, 1]$$

M_t is a non-negative $(\mathcal{F}_t)$-martingale with $E(M_t) = 1$ and by Doob forward convergence theorem it has P almost surely a limit $M_\infty(\omega)$ as $t \to \infty$, with $M_\infty \in L^1(P)$.

The following statements are equivalent:

1. *M_t is uniformly integrable*
2. *$E_P(M_\infty) = 1$*
3. *$\prod_{t=1}^{\infty} a_t > 0$*
4. *$\sum_{t=1}^{\infty}(1 - a_t) < \infty$*

Otherwise $M_\infty(\omega) = 0$ P a.s.

Proof 1) $\Longrightarrow$ 2) by the characterization of $L^1(P)$ convergence.

2) $\Longrightarrow$ 1): since $M_t \geq 0$ we can use Fatou's lemma: $\forall A \in \mathcal{F}_s$

$$E_P\big(M_\infty \mathbf{1}_A\big) = E_P\big(\liminf_{t\to\infty} M_t \mathbf{1}_A\big)$$
$$\leq \liminf_{t\to\infty} E_P\big(M_t \mathbf{1}_A\big) = E_P\big(M_s \mathbf{1}_A\big)$$

where we used the martingale property. This is the supermartingale property at $t = \infty$:

$$M_s(\omega) \geq E_P(M_\infty | \mathcal{F}_s)(\omega) \quad P \text{ a.s.}$$

By assumption

$$E_P\bigg(M_s - E_P(M_\infty|\mathcal{F}_s)\bigg) = E_P(M_s) - E_P(M_\infty) = 0$$

which implies that (M_s) is an UI martingale:

$$M_s(\omega) = E_P(M_\infty|\mathcal{F}_s)(\omega) \quad P \text{ a.s.}$$

1) $\Longrightarrow$ 2): Define

$$N_t(\omega) = \frac{\sqrt{M_t(\omega)}}{a_1 a_2 \dots a_t}$$

(N_t) is a martingale in $L^2(P)$.

By Doob L^p martingale inequality with $p = 2$,

$$E_P\left(\sup_{s \le t} M_s\right) \le E_P\left(\sup_{s \le t} N_s^2\right) \le 4E(N_t^2) = \frac{4}{a_1^2 \dots a_t^2}$$

and by the monotone convergence theorem

$$E_P\left(\sup_{s \in \mathbb{N}} M_s\right) = \lim_{t \to \infty} E_P\left(\sup_{s \le t} M_s\right) \le 4 \prod_{t \in \mathbb{N}} a_t^{-2}$$

Now if $\prod_{t \in \mathbb{N}} a_t > 0$, this gives a finite upper bound, and necessary (M_t) is an UI martingale since it is dominated by $\left(\sup_{s \in \mathbb{N}} M_s\right) \in L^1(P)$.

In case $\prod_{t \in \mathbb{N}} a_t = 0$, by Fatou lemma

$$E_P\left(\sqrt{M_\infty}\right) = E_P\left(\liminf_t \sqrt{M_t}\right) \le \liminf_t E_P\left(\sqrt{M_t}\right) = \lim_t a_1 a_2 \dots a_t = 0$$

which implies $M_\infty = 0$ P a.s.

3) $\Longrightarrow$ 4): On another probability space, take a sequence $(Y_n : n \in \mathbb{N})$ of independent Bernoulli random variables with

$$P(Y_n = 1) = 1 - P(Y_n = 0) = a_n \in (0, 1]$$

Let $B_n = \{\omega : Y_n(\omega) = 1\}$, and $B = \bigcap_{n \in \mathbb{N}} B_n$.

Using σ-additivity,

$$P(B) = \prod_{n \in \mathbb{N}} P(B_n) = \prod_{n \in \mathbb{N}} a_n$$

Note that since $P(B_n) = a_n > 0 \ \forall n$,

$$P(B) = 0 \Longleftrightarrow P(\liminf_n B_n) = 0 \Longleftrightarrow P(\limsup_n B_n^c) = 1$$

By the first and second Borel Cantelli lemma for independent events this is equivalent to

$$\infty = \sum_{n=1}^{\infty} P(B_n^c) = \sum_{n=1}^{\infty} (1 - a_n) \quad \square$$

As an application we appply Kakutani theorem to the likelihood ratio process.

On a probability space $(\Omega, \mathcal{F})$ consider a sequence of random variables $(X_n(\omega) : n \in \mathbb{N})$ which generate the filtration $(\mathcal{F}_n)$, $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$.

We consider two probability measures P and Q such that the random variables $(X_n(\omega))$ form an independent sequence under both measures P and Q.

$Q \overset{loc}{\ll} P$ (P dominates Q locally), which means that for all n and for all $A_n \in \mathcal{F}_n$, $P(A_n) = 0 \Longrightarrow Q(A_n) = 0$.

By the Radon-Nikodym theorem, for each $n \in \mathbb{N}$ there is an $\mathcal{F}_n$-measurable Radon-Nikodym derivative

$$0 \le Z_n(\omega) = \frac{dQ_n}{dP_n}(\omega) \text{ such that } Q(A) = E_P\big(Z_n \mathbf{1}_{A_n}\big) \quad \forall A \in \mathcal{F}_n$$

where Q_n and P_n are the restrictions of Q and P on the σ-algebra $\mathcal{F}_n$.

Now $Z_n(\omega)$ is a martigale, since if $A \in \mathcal{F}_m$ then $A \in \mathcal{F}_n$ $\forall m \ge n$ and by using twice the change of measure formula

$$E_P\big(Z_m \mathbf{1}_A\big) = Q(A) = E_P\big(Z_n \mathbf{1}_A\big)$$

Let's assume that $X_n(\omega) \in \mathbb{R}^d$ with densities $Q(X_n \in dx) = g_n(x)dx$ and $P(X_n \in dx) = f_n(x)dx$.

By assumption outside a set of Lebesgue measure 0, $g_n(x) = 0$ when $f_n(x) = 0$. In particular the function

$$z_n(x) = \frac{g_n(x)}{f_n(x)}$$

is well defined outside a set of Lebesgue measure 0.

It follows that

$$Z_n(\omega) = z_1(X_1(\omega)) z_2(X_2(\omega)) \dots z_n(X_n(\omega))$$

By Kakutani's theorem Z_n is UI martingale if and only if

$$\prod_{n=1}^{\infty} E_P\big(\sqrt{z_n(X_n)}\big) > 0$$
$$\Longleftrightarrow \sum_{n=1}^{\infty} \bigg(1 - E_P\big(\sqrt{z_n(X_n)}\big)\bigg) < \infty$$

Exercise 11. *Let X_n i.i.d. standard gaussian with $E_P(X_n) = 0$ and $E_P(X_n^2) = 1$ under the measure P and let $X_n \sim \mathcal{N}(\mu_n, 1)$ and independent under the measure Q.*

In this case

$$z_n(x) = \frac{(2\pi)^{-1/2} \exp\bigg(-\frac{1}{2}(x - \mu_n)^2\bigg)}{(2\pi)^{-1/2} \exp\bigg(-\frac{1}{2}x^2\bigg)} = \exp\bigg(x\mu_n - \frac{1}{2}\mu_n^2\bigg)$$

Then $P \sim Q$ on the σ-algebra $\mathcal{F}_\infty$ if and only if

$$0 < \prod_{n=1}^{\infty} E_P\Big(\sqrt{\exp\Big(x\mu_n - \frac{1}{2}\mu_n^2\Big)}\Big) = \prod_{n=1}^{\infty} E_P\Big(\exp\Big(\frac{1}{2}x\mu_n - \frac{1}{4}\mu_n^2\Big)\Big)$$
$$= \prod_{n=1}^{\infty} \exp(-\frac{1}{8}\mu_n^2) = \exp\Big(-\frac{1}{8}\sum_{n=1}^{\infty}\mu_n^2\Big)$$

which is equivalent to

$$\sum_{n=1}^{\infty} \mu_n^2 < \infty$$

In fact, if $\mu_n = \mu \neq 0\ \forall \mu$, then P and Q are singular on $\mathcal{F}_\infty$.
For example by the law of large numbers the set

$$A = \big\{\omega : \lim_{n\to\infty} n^{-1}(X_1(\omega) + \cdots + X_n(\omega)) = \mu\big\}$$

has $Q(A) = 1$ and $P(A) = 0$

Exercise 12. *Suppose now that under P the random variables (X_n) are i.i.d. Poisson(1) distributed, while under Q (X_n) are independent with respective distributions Poisson(λ_n) with $\lambda_n > 0$.*
In this case

$$z_n(x) = \Big(\exp(-\lambda_n)\lambda_n^x/n!\Big)\Big/\Big(\exp(-1)/n!\Big) = \exp\big(x\log(\lambda_n) + 1 - \lambda_n\big),$$
$$E_P\big(\sqrt{z_n(X_n)}\big) = \exp\Big(\frac{1}{2}(1-\lambda_n)\Big)E_P\Big(\sqrt{\lambda_n}^{X_n}\Big) =$$
$$\exp\Big(\sqrt{\lambda_n} - 1 + \frac{1}{2}(1-\lambda_n)\Big) = \exp\Big(-\frac{1}{2}(\sqrt{\lambda_n}-1)^2\Big)$$

since for a Poisson(1) distributed random variable X, $E_P\big(\theta^X\big) = \exp(\theta - 1)$.
Therefore $Q \sim P$ on $\mathcal{F}_\infty$ if and only if

$$0 < \prod_{n=1}^{\infty} \exp\Big(-\frac{1}{2}(\sqrt{\lambda_n}-1)^2\Big) = \exp\Big(-\frac{1}{2}\sum_{n=1}^{\infty}(\sqrt{\lambda_n}-1)^2\Big)$$
$$\iff \sum_{n=1}^{\infty}(\sqrt{\lambda}_n - 1)^2 < \infty$$

11 Continuous time

Moving from discrete time to continuous part, we need some technical assumptions.

We will work with the filtration $(\mathcal{F}_t : t \in \mathbb{R}^+)$ on the probability space $(\Omega, \mathcal{F}, P)$.

We say that the filtration $(\mathcal{F}_t)$ satisfies the *usual conditions* if

1. The filtration is completed by the P-null sets

$$\mathcal{F}_0 \supseteq \mathcal{N}^P := \{A \subseteq \Omega : P(A) = 0\}$$

2. The filtration is right-continuous

$$\forall t \geq 0 \quad \mathcal{F}_t = \mathcal{F}_{t+} := \bigcap_{u>t} \mathcal{F}_u$$

Next we discuss why these usual assumptions are needed.

Lemma 10. *Let $\tau(\omega) \geq 0$ be a random time and $(\mathcal{F}_t : t \geq 0)$ a filtration which in general is smaller than the filtration $(\mathcal{F}_{t+} : t \geq 0)$.*

1. *$\tau(\omega)$ is a stopping time with respect to the filtration $(\mathcal{F}_t+)$ if and only if $\{\tau < t\} \in \mathcal{F}_t \ \forall t \geq 0$.*
2. *When the filtration is right continuous τ is also a $(\mathcal{F}_t)$-stopping time.*

Proof When τ is a $(\mathcal{F}_t+)$-stopping time

$$\{\omega : \tau(\omega) < t\} = \bigcup_{n\in\mathbb{N}} \{\omega : \tau(\omega) \leq t - n^{-1}\} \in \mathcal{F}_t$$

since $\{\tau(\omega) \leq t - n^{-1}\} \in \mathcal{F}_{t-1/n}$ by definition of stopping time.
On the other hand, from the assumption

$$\{\omega : \tau(\omega) \leq t\} = \bigcap_{n\in\mathbb{N}} \{\omega : \tau(\omega) < t + n^{-1}\} \in \mathcal{F}_{t+} \quad \square$$

Exercise 13. *We show a filtration which is not right-continuous, generated by a continuous process. Consider the probability space of continuous functions started at zero*

$$\Omega = \big\{\omega \in C(\mathbb{R}^+, \mathbb{R}) : \omega_0 = 0\big\}$$

equipped with the Borel σ-algebra, where the canonical process is $X_t(\omega) = \omega_t$, Let $(\mathcal{F}^0_t)$ be the "raw" filtraton generated by X, with $\mathcal{F}^0_t = \sigma(\omega_s : s \leq t)$.
Note that $A \in \mathcal{F}^0_t$ if and only if for all $\omega, \widehat{\omega} \in \Omega$, with $\omega_s = \widehat{\omega}_s \quad \forall s \in [0, t]$,

$$\omega \in A \Longleftrightarrow \widehat{\omega} \in A$$

meaning that A depends only on the path ω restricted to the interval $[0, t]$.
For $a > 0$, consider first the random time

$$\tau(\omega) = \inf\big\{t > 0 : \omega_t \geq a\big\}$$

Now $\forall t > 0$,

$$\{\omega : \tau(\omega) \leq t\} = \{\omega : \inf_{q\leq t, q\in\mathbb{Q}^+} (a - \omega_q)^+ = 0\}$$

now since $(a - \omega_q)^+$ is $\mathcal{F}^0_q$ measurable by taking the infimum over the countable set $[0, t] \cap \mathbb{Q}$, we see that this event is $\mathcal{F}^0_t$ measurable.

Next we construct a random time which is a $(\mathcal{F}^0_{t+})$-stopping time but not a $(\mathcal{F}^0_t)$-stopping time. This shows that the raw filtration $(\mathcal{F}^0_t)$ is not right continuous, even if it is generated by a continuous process. Let

$$\widetilde{\tau}(\omega) = \inf\big\{t > 0 : \omega_t > a\big\}$$

For each $t > 0$,

$$\{\omega : \widetilde{\tau}(\omega) < t\} = \bigcup_{q \in \mathbb{Q}^+, q < t} \{\omega : \omega_q > a\} \in \mathcal{F}_t$$

meaning that $\widetilde{\tau}$ is a $(\mathcal{F}^0_{t+})$ stopping time.

However $\widetilde{\tau}$ is not a $(\mathcal{F}^0_t)$-stopping time. For fixed t, consider a set of paths which are crossing the level a for the first time at time t:

$$\begin{aligned} A_t &= \{\omega : \widetilde{\tau}(\omega) = t\} \\ &= \big\{\omega : \omega_q < a; \forall q < t, \ \ \omega_t = a, \ \ \exists N : \omega_{t+1/n} > a \quad \forall n > N\big\} \end{aligned}$$

For $\omega \in A_t$, consider the reflected path $\widehat{\omega}$

$$\widehat{\omega}_s = \begin{cases} \omega_s & s \in [0,t] \\ 2a - \omega_s & s > t \end{cases}$$

Now by construction when $\omega \in A_t$, $\tau(\widehat{\omega}) > \tau(\omega) = t$, since by construction $\widehat{\omega}$ attains the local maxima a at time t, and may cross the level a only later.

Which means, the event $\{\widetilde{\tau} \le t\}$ is $\mathcal{F}^0_{t+}$ measurable but not $\mathcal{F}^0_t$ measurable: by observing the paths on the interval $[0,t]$ we cannot distinguish between $\omega \in A_t$ and the corresponding $\widehat{\omega}$. For that we need to observe a little bit of the future, that is the extra information contained in $\mathcal{F}^0_{t+}$

Things may change when we complete the filtration with respect to a probability measure: Let P^W the Brownian measure on Ω, such that the canonical process $X_t(\omega) = \omega_t$ is a Brownian motion, and let $(\mathcal{F}_t)$ the filtration completed by the P^W-null events.

In the previous example it is not difficult to show that for each fixed $t > 0$ $P^W(A_t) = 0$, meaning that the probability that the Brownian motion will cross the level a for the first time at the pre-specified time t is zero, and by reflection this is equal to the probability that the Brownian motion attains the local maximum a at time t. Therefore

$$\{\widetilde{\tau} \le t\} = \{\widetilde{\tau} < t\} \cap \{\widetilde{\tau} = t\} \in \sigma(\mathcal{F}^0_t, \mathcal{N}^P) = \mathcal{F}_t$$

$\widetilde{\tau}$ is a stopping time with respect to the P^W-completed filtration $(\mathcal{F}_t)$.

We have seen that continuous process can generate filtrations which are not right continuous. On the other hand, the discontinuous raw filtration generated by a process with jumps may become continuous after completing with the P-null sets.

Proposition 7. *The completed filtration generated by a time-homogeneous process with independent increments is right continuous.*

Proof We give for the case of Brownian motion, but you can check that it goes through also for the Poisson process, (the same proof works for Lévy processes which we have not introduced yet).

Let $\mathcal{F}_t$ the completed Brownian filtration.

Fix $m, n \in \mathbb{N}$, $0 = u_0 < u_1 < \cdots < u_m = s_0 \le t < s_1 < \cdots < s_n$ and $\eta_h, \theta_k \in \mathbb{R}$. We compute the conditional characteristic functions. Let

$$G(\omega) = \exp\Big(i\eta_1(B_{u_1} - B_{u_0}) + \cdots + i\eta_m(B_{u_m} - B_{u_{m-1}}) + i\theta_1(B_{s_1} - B_{s_0}) + \cdots + \theta_n(B_{s_n} - B_{s_{n-1}}) \Big),$$

where $i = \sqrt{-1}$. By using the independence of the increments,

$$M_t := E_P(G|\mathcal{F}_t) =$$
$$= \exp\Big(i\eta_1(B_{u_1} - B_{u_0}) + \dots + i\eta_m(B_{u_m} - B_{u_{m-1}}) + i\theta_1(B_t - B_{u_m}) \Big) \times$$
$$\times \exp\Big(-\frac{1}{2}\Big\{ \theta_1^2(s_1 - t) + \sum_{k=2}^{n} \theta_k^2(s_k - s_{k-1}) \Big) $$

We see that the map $t \mapsto M_t(\omega)$ is right-continuous since B_t has right-continuous paths. By the martingale backward convergence theorem, P-almost surely

$$E(G|\mathcal{F}_{t+}) = M_{t+} = \lim_{u \downarrow t} M_u = M_t = E(G|\mathcal{F}_t)$$

By using the regular version of the conditional probabilty, since the characteristic function characterizes the conditional probability we see that

$$E(G|\mathcal{F}_{t+}) = E(G|\mathcal{F}_t)$$

for all bounded random variables $G(\omega)$. in particular taking $G = \mathbf{1}_A$ with $A \in \mathcal{F}_{t+}$ it follows there is an $\mathcal{F}_t$-measurable set A' such that A and A' differ at most by a set of measure zero. Since $\mathcal{F}_t$ contains the null sets, A is $\mathcal{F}_t$-measurable □

Lemma 11. *Let $D^+ = \{k2^{-n} : k, n \in \mathbb{N}\}$ be the dyadic set.*

Let $(M_u)_{u \in D^+}$ be a right-continuous martingale in the filtration $(\mathcal{F}_u)_{u \in D^+}$ satisfying the usual conditions.

For $t \in \mathbb{R}^+$ define

$$M_t(\omega) := \limsup_{u \downarrow t, u \in D^+} M_u(\omega) \text{ and } \mathcal{F}_t = \bigcap_{u > t, u \in D^+} \mathcal{F}_u$$

Then $(M_t)_{t \in \mathbb{R}^+}$ be a right-continuous martingale in the filtration $(\mathcal{F}_t)_{t \in \mathbb{R}^+}$ which satisfies the usual conditions.

Proof Let $t \in \mathbb{R}^+$ That $(\mathcal{F}_t)_{t \in \mathbb{R}^+}$ is right continuous follows from the definition. Also we see that $\limsup_{u \downarrow t, u \in D} M_u$ is $\mathcal{F}_t$-measurable.

Let $u_n \in D^+$ with $u_n \downarrow t$, and consider the time-discrete backward filtration $\widehat{\mathcal{F}}_{-n} = \mathcal{F}_{u_n}$. By definition

$$\mathcal{F}_t = \widehat{\mathcal{F}}_{-\infty} = \bigcap_n \mathcal{F}_{u_n}$$

The process $(M_{u_n} : n \in \mathbb{N})$ is a backward martingale, and by Doob backward convergence theorem (6)

$$M_t(\omega) = \lim_{n \to \infty} M_{u_n} \qquad P\text{-almost surely and in } L^1(P)$$

In particular it follows that $M_t \in L^1(P)$.

To check the martingale property, let $s, t \in \mathbb{R}$ with $s \le t$, $A \in \mathcal{F}_s$, and let $r_n \in D^+$ with $r_n \downarrow s$ and $u_n \in D^+$ with $u_n \downarrow t$. Since $s \le t$ we can choose sequences such that $r_n \le u_n$. Note that $A \in \mathcal{F}_{r_n} \supseteq \mathcal{F}_s\ \forall n$.

Since $M_{u_n}(\omega) \to M_t(\omega)$ and $M_{r_n}(\omega) \to M_s(\omega)$ P-almost surely and in $L^1(P)$

$$E_P\big(M_t \mathbf{1}_A\big) = \lim_{n\to\infty} E_P(M_{u_n}\mathbf{1}_A) = \lim_{n\to\infty} E_P(M_{r_n}\mathbf{1}_A) = E_P\big(M_s \mathbf{1}_A\big)$$

where we used the martingale property of $(M_u)_{u\in D^+}$; $\square$.

12 Localization

Definition 22. *We say that a property holds locally with respect to the filtration $(\mathcal{F}_t)$ for the process $(X_t(\omega))$, if there is a localizing sequence of $(\mathcal{F}_t)$-stopping times $\tau_n(\omega) \uparrow \infty$ such that for each n the stopped process $X_t^{\tau_n}(\omega) := X_{t\wedge\tau_n}(\omega)$ satisfyies that property.*

For example every $(\mathcal{F}_t)$-adapted process $(X_t : t \in \mathbb{R}^+)$ with continuous paths is locally bounded: as localizing sequence we can take

$$\tau_n(\omega) := \inf\big\{t : |X_t(\omega)| > n\big\}$$

which gives $|X_{t\wedge\tau_n}(\omega)| \le n$.

13 Doob decomposition in continuous time

We recall that the (total) variation of a function $s \mapsto x(s)$ in the interval $[0,t]$ is given by

$$V_{[0,t]}(x) := \sup_{\Pi} \sum_{t_i\in\Pi} |x(t_i) - x(t_{i-1})|$$

where the supremum is taken over partitions $\Pi = (0 = t_0 \le t_1 \le \dots, \le t_n = t)$ of the interval $[0,t]$. It follows that $x(s)$ has finite first variation if and only if $x(s) = x(0) + x^{\oplus}(s) - x^{\ominus}(s)$ with $x^{\oplus}, x^{\ominus}$ non-decreasing functions.

Lemma 12. *A continuous local martingale (M_t) with almost surely finite (total) variation is necessarly constant.*

Proof Without loss of generality we assume that $M_0(\omega) = 0$. Let $\tau_n(\omega) \uparrow \infty$ a localizing sequence of stopping times such that for each n the stopped process $M_{t\wedge\tau_n}$ is a martingale. We define stopping times

$$\sigma_n = \tau_n \wedge \inf\{t : V_{[0,t]}(X(\omega)) > n\} \le \tau_n$$

By Doob optional sampling theorem, the stopped process $M_t^{\sigma_n}(\omega)$ is a martingale with

$$|M_t^{\sigma_n}| \le V_{[0,t]}(M^{\sigma_n}) \le n \quad \forall t \ge 0$$

Since $\sigma_n(\omega) \to \infty$, it is a localizing sequence. In order to simplify the notation, let's fix n and assume that $M_t(\omega) := M_t^{\sigma_n}(\omega)$ is a true martingale, which has bounded first variation. By the discrete integration by parts formula, for a sequence $(0 = t_0 \le t_1 \le t_2 \le \dots)$, with $t_n \to \infty$. We have

$$M(t)^2 = 2\sum_{i=1}^{\infty} M(t_{i-1}\wedge t)(M(t_i\wedge t) - M(t_{i-1}\wedge t)) + \sum_{i=1}^{\infty}(M(t_i\wedge t) - M(t_{i-1}\wedge t))^2$$

Since $s \mapsto M_s(\omega)$ is uniformly continuous on $[0,t]$, there is a random $\delta(\omega)$ such that

$$\sum_i (M_{t_i\wedge t}-M_{t_{i-1}\wedge t})^2 \le \sup_i |M_{t_i\wedge t}-M_{t_{i-1}\wedge t}| \sum_i |M_{t_i\wedge t}-M_{t_{i-1}\wedge t}| \le \varepsilon V_{[0,t]}(M) \le \varepsilon n$$

when $\Delta(\Pi) = \sup_i\big\{(t_i \wedge t) - (t_{i-1}\wedge t)\big\} < \delta(\omega)$. This means

$$\sum_i (M_{t_i\wedge t} - M_{t_{i-1}\wedge t})^2 \to 0 \qquad P\text{-almost surely}$$

as $\Delta(\Pi) \to 0$, and we have

$$M_t^2 = \lim_{\Delta(\Pi)\to 0} 2\sum_{i=1}^{\infty} M_{t_{i-1}\wedge t}(M_{t_i\wedge t}-M_{t_{i-1}\wedge t}) := 2\int_0^t M_s dM_s \qquad P\text{-almost surely}$$

where for almost every ω the limit of Riemann-sums is a Riemann-Stieltjes integral. Note also that the sum containes a fixed number of nonzero terms. By taking expectation,

$$\begin{aligned} E_P(M_t^2) &= 2E_P\left(\lim_{\Delta(\Pi)\to 0}\sum_{i=1}^{\infty} M_{t_{i-1}\wedge t}(M_{t_i\wedge t} - M_{t_{i-1}\wedge t})\right) \\ &= 2\lim_{\Delta(\Pi)\to 0} 2E_P\left(\sum_{i=1}^{\infty} M_{t_{i-1}\wedge t}(M_{t_i\wedge t} - M_{t_{i-1}\wedge t})\right) = \\ & \lim_{\Delta(\Pi)\to 0} 2\sum_{i=1}^{\infty} E_P\left(M_{t_{i-1}\wedge t} E_P(M_{t_i\wedge t} - M_{t_{i-1}\wedge t}|\mathcal{F}_{t_{i-1}\wedge t})\right) = 0 \end{aligned}$$

where we used the martingale property, which gives $M_t(\omega) = M_0(\omega) = 0 \ \forall t$. The interchange of limit and expectation is justified by the bounded convergence theorem, since $M_t(\omega)$ has bounded variation.

$$\left|\sum_{i=1}^{\infty} M_{t_{i-1}\wedge t}(M_{t_i\wedge t} - M_{t_{i-1}\wedge t})\right| \le V_{[0,t]}(M(\omega))^2 \le n^2 \qquad P\text{-almost surely} .$$

Coming back to the local martingale, we have

$$M_t(\omega) = \lim_{n\to\infty} M_{t\wedge\sigma_n}(\omega) = 0 \qquad P\text{-almost surely} \ \square$$

The next two technical lemma are not very intuitive but useful:

Lemma 13. *Suppose $(A_n : n \in \mathbb{N})$ is a $(\mathcal{F}_n)$-predictable and non-decreasing process with $A_0 = 0$, such that*

$$E_P(A_\infty - A_n|\mathcal{F}_n)(\omega) \le C \quad \forall n$$

Then $E_P(A_\infty^2) \le 2C^2$.

Proof

$$(A_n)^2 = \sum_{k=1}^{n}\sum_{h=1}^{n} \Delta A_k \Delta A_k = 2\sum_{k=1}^{n}\sum_{h=k}^{n} \Delta A_h \Delta A_k - \sum_{k=1}^{n}(\Delta A_k)^2$$
$$= 2\sum_{k=1}^{n}(A_n - A_{k-1})\Delta A_k - \sum_{k=1}^{n}(\Delta A_k)^2$$

where $\Delta A_k = (A_k - A_{k-1})$, and since the terms $(A_n)^2$ and $\sum_{k=1}^{n}(\Delta A_k)^2$ are non-negative and non-decreasing, the monotone convergence theorem applies

$$E_P(A_\infty^2) = 2E\bigg(\sum_{k=0}^{\infty}(A_n - A_{k-1})\Delta A_k\bigg) - E_P\bigg(\sum_{k=1}^{\infty}(\Delta A_k)^2\bigg)$$

where we can exchange the order of summation and integration. By taking conditional expectation inside and using predictability,

$$E_P(A_\infty^2) \le 2\sum_{k=0}^{\infty} E_P\Big(E_P\big((A_\infty - A_{k-1})\Delta A_k \big| \mathcal{F}_{k-1}\big)\Big)$$
$$= 2\sum_{k=0}^{\infty} E_P\Big(E(A_\infty - A_{k-1}|\mathcal{F}_{k-1})\Delta A_k\Big) \le 2CE_P\bigg(\sum_{k=1}^{\infty}\Delta A_k\bigg) = 2CE_P(A_\infty) \le 2C^2$$

Lemma 14. *Suppose $A_n^{(1)}$ and $A_n^{(2)}$ are two predictable processes satisfying the hypothesis of lemma 13 and $B_n = (A_n^{(1)} - A_n^{(2)})$. Suppose that there is a r.v. $Y(\omega) \ge 0$ with $E_P(Y^2) < \infty$ and*

$$\big|E_P(B_\infty - B_n|\mathcal{F}_n)(\omega)\big| \le N_n(\omega) := E_P(Y|\mathcal{F}_n)(\omega) \quad \forall n\ .$$

Then there exists c such that

$$E_P\Big(\sup_{n\in N} B_n^2\Big) \le c\Big(E_P(Y^2) + CE(Y^2)^{1/2}\Big)$$

Proof We shall need the following estimate: since

$$|\Delta B_k| = |\Delta A_k^{(1)} - \Delta A_k^{(2)}| \le \Delta A_k^{(1)} + \Delta A_k^{(2)},$$

it follows

$$E_P(B_\infty^2) = 2E\bigg(\sum_{k=0}^{\infty} E(B_n - B_{k-1}|\mathcal{F}_k)\Delta B_k\bigg) - E_P\bigg(\sum_{k=1}^{\infty}(\Delta B_k)^2\bigg) \le 2E_P\big((A_\infty^{(1)} + A_\infty^{(2)})Y\big)$$
$$\le 2E_P\big(Y^2\big)^{1/2}\Big(E_P\big(\{A_\infty^{(1)}\}^2\big)^{1/2} + E_P\big(\{A_\infty^{(2)}\}^2\big)^{1/2}\Big) \le 2^{5/2}CE_P(Y^2)$$

where we used Hölder and Cauchy-Schwartz inequalities together with lemma 13.

Let $M_n := E_P(B_\infty|\mathcal{F}_n)$, $X_n := M_n - B_n$, and $N_n = E_P(Y|\mathcal{F}_n)$ from the triangle inequality

$$E_P\Big(\sup_n B_n^2\Big)^{1/2} \le E_P\Big(\sup_n X_n^2\Big)^{1/2} + E_P\Big(\sup_n M_n^2\Big)^{1/2}$$

Note that

$$|X_n| = \big|E_P(B_\infty - B_n|\mathcal{F}_n)\big| \le E(Y|\mathcal{F}_n) = N_n$$

Since (M_n) and (N_n) are martingales bounded in $L^2(P)$ we can use Doob L^p-martingale inequality

$$E_P\biggl(\sup_n B_n^2\biggr)^{1/2} \le E_P\biggl(\sup_n N_n^2\biggr)^{1/2} + 2E_P\biggl(M_\infty^2\biggr)^{1/2}$$
$$\le 2E_P(Y^2)^{1/2} + 2E_P(B_\infty^2) \le 2E_P(Y^2)^{1/2} + 2^{7/2}CE_P(Y^2) \quad \square$$

Theorem 13. *Suppose $(X_t : t \in \mathbb{R}^+)$ is a $(\mathcal{F}_t)$-supermartingale with continuous paths. Then we have the Doob-Meyer decomposition*

$$X_t(\omega) = X_0(\omega) + M_t(\omega) - A_t(\omega)$$

*where $M_0(\omega) = A_0(\omega) = 0$, M_t is a continuous $(\mathcal{F}_t)$-***local** *martingale and A_t is continuous and non-decreasing. Moreover (M_t) and (A_t) are uniquely determined up to indistinguishable processes.*

Remark : this result can be extended to processes with jumps.

Proof, Uniqueness: Suppose that we have two Doob-Meyer decompositions

$$X_t - X_0 = M_t - A_t = \widetilde{M}_t - \widetilde{A}_t$$

It follows that

$$S_t := (M_t - \widetilde{M}_t) = (A_t - \widetilde{A}_t)$$

is a continuous local martingale with paths of finite variation, and by lemma 12 necessarily $S_t(\omega) = S_0(\omega) = 0\ \forall t$ P-almost surely.

Existence : by considering the stopped process $X_t^{\tau_C} = X_{t\wedge\tau_C}$, where

$$\tau_C(\omega) = \inf\bigl\{s : |X_s(\omega)| > C \text{ or } s > C\bigr\}$$

we reduce first the problem to the case where X is a bounded and uniformly continuous process, which is constant on the interval $[C,\infty)$. Without loss of generality we assume that $X_0(\omega) = 0$.

Fix k and $m \in \mathbb{N}$, and consider $\mathcal{F}_k^m = \mathcal{F}_{k2^{-m}}$, $k \in \mathbb{N}$.

Construct for each $m \in \mathbb{N}$ the discrete time Doob decomposition

$$X_{k2^{-m}}(\omega) = M_k^{(m)} + A_k^{(m)}$$

Define for each m the continuous time filtration

$$\overline{\mathcal{F}}_t^{(m)}(\omega) = \mathcal{F}_{k2^{-m}}(\omega) \quad \text{when } (k-1)2^{-m} < t \le k2^{-m}$$

and the continuous time process

$$\overline{A}_t^{(m)}(\omega) = A_k^{(m)}(\omega) \quad \text{when } (k-1)2^{-m} < t \le k2^{-m}\ .$$

Note that for each m, $\overline{A}_t^{(m)}$ is $(\mathcal{F}_t)$-adapted, since in the time-discrete Doob decomposition $A_k^{(m)}(\omega)$ is $\mathcal{F}_{(k-1)2^{-m}}$-measurable.

Consider the *modulus of continuity*

$$W(\delta,\omega) := \sup_{s\le K, |s-t|\le\delta} |X_t(\omega) - X_s(\omega)|$$

$W(\delta)$ is a bounded random variable since $X_t(\omega)$ is bounded, and because $X_t(\omega)$ has uniformly continuous paths $W(\delta) \to 0$ P-almost surely as $\delta \to 0$. By the bounded convergence theorem $W(\delta) \to 0$ in $L^2(P)$ sense.

We show that $\overline{A}_t^{(m)}$ converges in $L^2(P)$ uniformly in t as $m \to \infty$.

For $m > n$, $\overline{A}_t^{(m)}$ and $\overline{A}_t^{(n)}$ are constant on the intervals $\big((k-1)2^{-m}, k2^{-m}\big]$, we have

$$\sup_t |\overline{A}_t^{(m)} - \overline{A}_t^{(n)}| = \sup_{k\in\mathbb{N}} |\overline{A}_{k2^{-m}}^{(m)} - \overline{A}_{k2^{-m}}^{(n)}|$$

Fix $t = k2^{-m}$ for some k. and let $(l-1)2^{-n} < t \le l2^{-n}$. Denote $u = l2^{-n}$. By the discrete time Doob decomposition

$$E_P(\overline{A}_\infty^{(m)} - \overline{A}_t^{(m)} | \overline{\mathcal{F}}_t^{(m)})(\omega) = E_P(A_\infty^{(m)} - A_k^{(m)} | \mathcal{F}_{k2^{-m}})(\omega) = E_P(X_t - X_\infty | \mathcal{F}_{k2^{-m}})(\omega)$$

On the other hand

$$E_P(\overline{A}_\infty^{(n)} - \overline{A}_t^{(n)} | \overline{\mathcal{F}}_t^{(m)})(\omega) = E_P(A_\infty^{(n)} - A_l^{(n)} | \mathcal{F}_t)(\omega) = E_P\left(E_P(A_\infty^{(n)} - A_l^{(n)} | \mathcal{F}_u) \middle| \mathcal{F}_t \right)(\omega) =$$
$$E_P\left(E_P(X_u - X_\infty | \mathcal{F}_u) \middle| \mathcal{F}_t \right)(\omega) = E_P\big(X_u - X_\infty \big| \mathcal{F}_t\big)(\omega)$$

Then the difference of conditional expectations is bounded:

$$\left| E_P(\overline{A}_\infty^{(m)} - \overline{A}_t^{(m)} | \mathcal{F}_t) - E_P(\overline{A}_\infty^{(n)} - \overline{A}_t^{(n)} | \mathcal{F}_t) \right|$$
$$\le E_P\Big(|X_t - X_u| \big| \mathcal{F}_t\big) \le E_P\big(W(2^{-n}) \big| \mathcal{F}_t\big)$$

The assumptions of lemma 14 are satisfied, giving

$$E_P\left(\sup_t (\overline{A}_t^{(m)} - \overline{A}_t^{(n)})^2 \right) \le c \left\{ E_P\big(W(2^{-n})^2\big) + 2C E_P\big(W(2^{-n})^2\big)^{1/2} \right\} \to 0 \quad \text{as } n, m \to \infty,\ m > n$$

It can be shown that the space of processes

$$S_2 := \left\{ Z(t,\omega)\ (\mathcal{F}_t)\text{-adapted process with } \| Z \|_{\mathcal{S}_2}^2 := E_P\left(\sup_t Z_t^2 \right) < \infty \right\}$$

is complete under the $\| \cdot \|_{\mathcal{S}_2}$ norm. There is a process $A_t(\omega) \in S_2$ such that

$$E_P\left(\sup_t \big\{ A_t^{(n)} - A_t \big\}^2 \right) \to 0$$

From convergence in quadratic mean it follows that there is a subsequence (n_i) such that

$$\sup_t |A_t^{(n_i)}(\omega) - A_t(\omega)| \to 0 \quad P\text{-almost surely} .$$

Next we show that $A_t(\omega)$ is continuous.

$$\Delta\overline{A}_t^n = E_P\left(X_{(k-1)2^n} - X_{(k)2^n} \middle| \mathcal{F}_{(k-1)2^{-n}} \right) \le E_P\big(W(2^{-n}) \big| \mathcal{F}_{(k-1)2^{-n}} \big)$$

where on the right hand side we have an uniformly integrable martingale. We have

$$E_P\left(\sup_t (\Delta\overline{A}_t^n)^2 \right) \le E_P\left(\sup_k E_P\big(W(2^{-n}) \big| \mathcal{F}_{(k-1)2^{-n}} \big)^2 \right) \le 4E_P\left(W(2^{-n})^2 \right) \to 0 \quad \text{ as } n \to \infty$$

by Doob L^p-martingale inequality. In particular there is a further subsequence (n_j) such that

$$\sup_t \Delta\overline{A}_t^{n_j}(\omega) \to 0 \quad P\text{- almost surely as } j \to \infty$$

Almost sure continuity follows:

$$\begin{aligned} &\sup_t |\Delta A_t(\omega)| \le \sup_t |\Delta A_t(\omega) - \Delta A_t^{(n_j)}(\omega)| + \sup_t |\Delta A_t^{(n_j)}(\omega)| \\ &\le 2\sup_t |A_t(\omega) - A_t^{(n_j)}(\omega)| + \sup_t |\Delta A_t^{(n_j)}(\omega)| \end{aligned}$$

which for almost all ω is arbitrary small for j large enough.

We show that $M_t := (X_t + A_t)$ is a $(\mathcal{F}_t)$-martingale. Since M_t is continuous and square integrable since $X_t(\omega)$ and $A_t(\omega)$ are.

By using lemma 11 it is enough to show the martingale property for $s < t$ with $s, t \in D_n = \{k2^{-n} : k \in \mathbb{Z}\}$, $n \in \mathbb{N}$, and $B \in \mathcal{F}_s$:

$$\begin{aligned} &E_P\big(M_t \mathbf{1}_B\big) = E_P\big(M_s \mathbf{1}_B\big) \\ &\iff E_P\big((X_t + A_t)\mathbf{1}_B\big) = E_P\big((X_s + A_s)\mathbf{1}_B\big) \\ &\iff E_P\big((X_t + \overline{A}_t^{(n)})\mathbf{1}_B\big) - E_P\big((X_s + \overline{A}_s^{(n)})\mathbf{1}_B\big) = \\ &= E_P\big((\overline{A}_t^{(n)} - A_t)\mathbf{1}_B\big) - E_P\big((\overline{A}_s^{(n)} - A_s)\mathbf{1}_B\big) \end{aligned}$$

where we can choose n and in the last equation the left hand side is zero by the discrete time martingale property

By the Cauchy Schwartz inequality,

$$\left| E_P\big((\overline{A}_t^{(n)} - A_t)\mathbf{1}_B\big) \right| \le E_P\left(\sup_t (\overline{A}_t^{(n)} - A_t)^2 \right)^{1/2} \sqrt{P(B)} \longrightarrow 0.$$

For the general case, by using the localization

$$X_t = \lim_{C\to\infty} X_t^{\tau_C}(\omega) = X_0 + \lim_{C\to\infty} M_t^{(C)}(\omega) - \lim_{C\to\infty} A_t^{(C)}(\omega) = X_0 + M_t - A_t$$

where $M_t^{(C)}$ are continuous true martingales and $A_t^{(C)}$ are continuous increasing processes with $M_0^{(C)}(\omega) = A_0^{(C)}(\omega) = 0$ and

$$M_t^{(C)}(\omega) = M_t^{(C+1)}(\omega) \text{ and } A_t^{(C)}(\omega) = A_t^{(C+1)}(\omega) \text{ on } [0, \tau_C]$$

This implies that the limits $M_t(\omega)$ and $A_t(\omega)$ exist with $M_t^{(C)} = M_{t\wedge\tau_C}$ and $A_t^{(C)} = A_{t\wedge\tau_C}$. Therefore A_t is continuous and non-decreasing and M_t is a *local* martingale with localizing sequence $(\tau_C : C \in \mathbb{N})$.

Note that without additional assumptions, it is not possible to show that M_t is a true martingale, since for $t > s$ and $B \in \mathcal{F}_s$, we would like to have

$$. \quad E_P\big((M_t - M_s)\mathbf{1}_B\big) = E_P\big(\lim_{C\to\infty}(M_{t\wedge\tau_C} - M_{s\wedge\tau_C})\mathbf{1}_B\big) \tag{13}$$

$$\overset{?}{=} \lim_{C\to\infty} E_P\big((M_{t\wedge\tau_C} - M_{s\wedge\tau_C})\mathbf{1}_B\big) = 0 \tag{14}$$

however the interchange of limit and expectation is not always justified □

Definition 23. 1. *We say that a right continuous $(\mathcal{F}_t)$-adapted process $(X_t(\omega))$ is in the class DL if for each $t > 0$ the family of random variables*

$$\Big\{ X_\tau(\omega) : \tau \text{ is a stopping time with } \tau(\omega) \le t \text{ a.s.} \Big\}$$

is uniformly integrable,

2. *the right continuous adapted process $(X_t(\omega))$ is in the class D is the family of random variables*

$$\Big\{ X_\tau(\omega) : \tau \text{ is a stopping time} \Big\}$$

is uniformly integrable.

Exercise 14. 1. *A local martingale M_t of class DL is a true martingale.*

2. *A local martingale M_t of class D is an uniformly integrable martingale.*

Proof

1. Let (τ_n) be a localizing sequence. For $0 \le s \le t$, $B \in \mathcal{F}_s$ we have

$$E_P\big((M_t - M_s)\mathbf{1}_B\big) = E_P\big(\lim_{n\to\infty}(M_{t\wedge\tau_n} - M_{s\wedge\tau_n})\mathbf{1}_B\big) = \lim_{n\to\infty} E_P\big((M_{t\wedge\tau_n} - M_{s\wedge\tau_n})\mathbf{1}_B\big) = 0$$

where the last step is justified since the family $\{|M_{t\wedge\tau_n} - M_{s\wedge\tau_n}| : n \in \mathbb{N}\}$ is uniformly integrable by assumption.

2. M_t is a martingale by the previous step, and it is clear that M_t is uniformly integrable since determistic times are stopping times.

Corollary 11. *A continuous $(\mathcal{F}_t)$-supermartingale of class DL has unique Doob-Meyer decomposition*

$$X_t(\omega) = X_0(\omega) + M_t(\omega) - A_t(\omega)$$

where $M_0(\omega) = A_0(\omega) = 0$, M_t is a continuous true $(\mathcal{F}_t)$-martingale and A_t is continuous and non-decreasing.

Moreover if X_t is of class D, the martingale M_t is uniformly integrable and A_t is integrable.

Proof When X_t is of class DL, for t and $B \in \mathcal{F}_t$, by the characterization of convergence in $L^1(P)$ we have

$$E_P\big(|X_t - X_{t\wedge\tau_C}|\big) \to 0 \text{ as } C \to \infty$$

Since A is non-decreasing by the monotone convergence theorem

$$E_P\big(A_t - A_{t\wedge\tau_C}\big) \to 0 \text{ as } C \to \infty$$

Therefore

$$\| M_t - M_{t\wedge\tau_C} \|_{L^1(P)} \le \| X_t - X_{t\wedge\tau_C} \|_{L^1(P)} + \| A_t - A_{t\wedge\tau_C} \|_{L^1(P)} \to 0$$

which justifies the interchange of limit and expectation in equation 13.

When X_t is of class D it is uniformly integrable, therefore $X_t \to X_\infty$ almost surely and in $L^1(P)$ by the Doob martingale convergence theorem, and by the martingale property

$$E_P(A_\infty) = \lim_{t\uparrow\infty} E_P(A_t) = \lim_{t\uparrow\infty} E_P(X_t - X_0) = E_P(X_\infty - X_0) < \infty,$$

which means that

$$M_t = (X_t - X_0 - A_t) \to M_\infty = (X_\infty - X_0 - A_\infty)$$

P-almost surely and in $L^1(P)$ sense. In particular M_t is uniformly integrable. $\square$.

14 Quadratic and predictable variation of a continuous local martingale

Let M_t be a continuous local martingale in the $(\mathcal{F}_t)$-filtration, and (τ_n) a localizing sequence. Note that we can choose (τ_n) such that $|M_t^{\tau_n}(\omega)| \le n$.

By Jensen inequality, the stopped process $(M_t^{\tau_n})^2$ is a $(\mathcal{F}_t)$-submartingale, with Doob decomposition

$$(M_t^{\tau_n})^2 = M_0^2 + N_t^{(n)} + \langle M^{\tau_n}\rangle_t$$

where $\langle M^{\tau_n}\rangle_t$ is a continuous non-decreasing process and $N_t^{(n)}$ is a local martingale.

Since $\tau_n \le \tau_{n+1}$ and the Doob-Meyer decomposition is unique it follows that

$$N_t^{(n)}\mathbf{1}(\tau_n > t) = N_t^{(n+1)}\mathbf{1}(\tau_n > t) = N_t\mathbf{1}(\tau_n > t) \quad \text{and}$$
$$\langle M^{\tau_n}\rangle_t\mathbf{1}(\tau_n > t) = \langle M^{\tau_{n+1}}\rangle_t\mathbf{1}(\tau_n > t) = \langle M\rangle_t\mathbf{1}(\tau_n > t)$$

where $N_t := \lim_{n\uparrow\infty} N_t^{(n)}$ is a local martingale and $\langle M\rangle_t = \lim_{n\uparrow\infty}\langle M^{\tau_n}\rangle_t$ is a continuous increasing process, which give the Doob-Meyer decomposition

$$M_t^2 = M_0^2 + N_t + \langle M\rangle_t$$

The process $\langle M\rangle_t$ is the *predictable variation* of the local martingale M_t. Note that

$$M_t - M_s = 0 \quad P\text{-almost surely} \implies \langle M\rangle_t = \langle M\rangle_s \quad P\text{-almost surely}$$

Definition 24. *Let M_t, $\widetilde{M}_t$ $(\mathcal{F}_t)$-local martingales. We define by polarization the predictable covariation as*

$$\langle M, \widetilde{M}\rangle_t := \frac{1}{4}\big(\langle M + \widetilde{M}\rangle_t - \langle M - \widetilde{M}\rangle_t\big) = \frac{1}{2}\big(\langle M + \widetilde{M}\rangle_t - \langle M\rangle_t - \langle \widetilde{M}\rangle_t\big)$$

Note that $\langle M, M\rangle_t = \langle M\rangle_t$.

Proposition 8. *$\langle M, \widetilde{M}\rangle_t$ is the unique continuous process of finite (total) variation such that $\langle M, \widetilde{M}\rangle_0 = 0$ and*

$$M_t\widetilde{M}_t = M_0\widetilde{M}_0 + \widehat{N}_t + \langle M, \widetilde{M}\rangle_t \tag{15}$$

where $\widehat{N}_t$ is a local martingale with $\widehat{N}_t = 0$.

Proof Since $(M_t \pm \widetilde{M}_t)$ are local martingales with Doob-Meyer decompositions

$$(M_t \pm \widetilde{M}_t)^2 = (M_0 \pm \widetilde{M}_0)^2 + N_t^{(\pm)} + \langle M \pm \widetilde{M}\rangle_t$$

we use the polarization identity

$$M_t\widetilde{M}_t = \frac{1}{4}\Big\{(M_t + \widetilde{M}_t)^2 - (M_t - \widetilde{M}_t)^2\Big\}$$

to obtain the semimartingale decomposition (15) with $\widehat{N}_t = \big(N_t^{(+)} - N_t^{(-)}\big)/4$ □

Exercise 15. *Let $(B_t, \widetilde{B}_t)_{t\geq 0}$ a pair of independent Brownian motion, and consider the filtration $\mathcal{F}_t = \sigma(B_s, \widetilde{B}_s : s \leq t) \vee \mathcal{N}^P$ completed by the sets of measure zero.*

B_t and B_t are square integrable martingales.

$$\begin{aligned}
&E_P\big(B_t\widetilde{B}_t - B_s\widetilde{B}_s\big|\mathcal{F}_s\big)\\
&= B_sE_P\big(\widetilde{B}_t - \widetilde{B}_s||\mathcal{F}_s\big) + \widetilde{B}_sE_P\big(B_t - B_s||\mathcal{F}_s\big) + E_P\big((B_t - B_s)(\widetilde{B}_t - \widetilde{B}_s)\big|\mathcal{F}_s\big) =\\
&B_sE_P\big(\widetilde{B}_t - \widetilde{B}_s\big) + \widetilde{B}_sE_P\big(B_t - B_s\big) + E_P\big((B_t - B_s)E_P\big(\widetilde{B}_t - \widetilde{B}_s\big) = 0
\end{aligned}$$

therefore the product $(B_t\widetilde{B}_t)$ is a martingale and from the uniqueness of the Doob-Meyer decomposition it follows that $\langle B, \widetilde{B}\rangle_t = 0$.

For $\alpha \in [0, 1]$, consider the process

$$W_t = \sqrt{\alpha}B_t + \sqrt{(1-\alpha)}\widetilde{B}_t$$

It follows that (W_t) is a Brownian motion adapted to the filtration $\mathcal{F}_t$. We have

$$\begin{aligned}
&E_P\big(B_tW_t - B_sW_s\big|\mathcal{F}_s\big)\\
&= B_sE_P\big(W_t - W_s||\mathcal{F}_s\big) + \widetilde{W}_sE_P\big(W_t - W_s||\mathcal{F}_s\big) + E_P\big((B_t - B_s)(W_t - W_s)\big|\mathcal{F}_s\big)\\
&= 0 + \sqrt{\alpha}E_P\big((B_t - B_s)^2\big|\mathcal{F}_s\big) + \sqrt{(1-\alpha)}E_P\big((B_t - B_s)(\widetilde{B}_t - \widetilde{B}_s)\big|\mathcal{F}_s\big)\\
&= \sqrt{\alpha}(\langle B\rangle_t - \langle B\rangle_s) = \sqrt{\alpha}(t - s)
\end{aligned}$$

It follows that $\rangle B, W\rangle_t = \sqrt{\alpha}\langle B\rangle_t = \sqrt{\alpha}t$

Theorem 14. *Let M be a continuous martingale with $|M_t(\omega)| \le C < \infty \ \forall t > 0$. Then*

$$[M]_t = \lim_{|\Delta|\to 0} \sum_{k=1}^{\infty} (M_{t\wedge t_k} - M_{t\wedge t_{k-1}})^2$$

where the limit exists in $L^2(P)$ sense uniformly on compacts, with

$$\Delta = (0 \le t_0 < t_1 < t_n \dots), \quad |\Delta| := \sup_i (t_i - t_{i-1})$$

$[M]_t$ is continuous and non-decreasing and satisfies:

$$M_t^2 = M_0^2 + [M]_t + N_t$$

where N_t is a true martingale. In other words $[M]_t = \langle M \rangle_t$.

Proof Without loss of generality we assume $M_0 = 0$, otherwise consider $M_t = (M_t - M_0)$. Lets denote

$$T_t^{\Delta}(M) := \sum_{k=1}^{\infty} (M_{t\wedge t_k} - M_{t\wedge t_{k-1}})^2 \tag{16}$$

It follows that $(M_t^2 - T_t^{\Delta}(M))$ is a martingale since:

$$(M_t - M_s)^2 = M_t^2 - M_s^2 + 2M_s(M_t - M_s)$$

and by the martingale property

$$E\big((M_t - M_s)^2 | \mathcal{F}_s\big) = E\big(M_t^2 - M_s^2 | \mathcal{F}_s\big) \tag{17}$$

For $s = s_0 < s_1 < \dots < s_n = t$

$$\sum_{k=1}^{n} E\big(M_{s_k}^2 - M_{s_{k-1}}^2 | \mathcal{F}_s\big) = \sum_{k=1}^{n} E\big(\{M_{s_k} - M_{s_{k-1}}\}^2 | \mathcal{F}_s\big) = E(T_t^{\Delta}(M) - T_s^{\Delta}(M) \big| \mathcal{F}_s\big)$$

In particular for fixed partitions Δ, Δ'

$$X_t := T_t^{\Delta}(M) - T_t^{\Delta'}(M)$$

is a martingale. We will show that $X_t = X_t^{\Delta,\Delta'} \to 0$ in $L^2(P)$ as $|\Delta|, |\Delta'| \to 0$.

Denote $\Delta\Delta' = \Delta \cup \Delta'$, the coarsest partition of $\mathbb{R}^+$ containing both Δ and Δ'. Note that for fixed Δ, Δ', X_t is bounded on compact intervals, since is the sum of finitely many squared differences of the bounded process M.

Consider the process $T_t^{\Delta\Delta'}(X)$, which is defined as in 16 replacing the martingale M_t with the martingale X_t.

From 17 we see that

$$(X_t^2 - T_t^{\Delta\Delta'}(X))$$

is also a martingale. Since $(a-b)^2 \le 2(a^2 + b^2)$, we have

$$E(X_t^2) = E(T_t^{\Delta\Delta'}(X)) \le 2E_P\Big(T_t^{\Delta\Delta'}(T^{\Delta}(M)) + T_t^{\Delta\Delta'}(T^{\Delta'}(M))\Big)$$

We show that $E_P\left(T_t^{\Delta\Delta'}(T^\Delta(M))\right) \to 0$.

Let $s_k \in \Delta\Delta'$, $t_l \in \Delta$ such that $t_l \le s_k < s_{k+1} \le t_{l+1}$.

$$\begin{aligned}
&T^\Delta_{s_{k+1}}(M) - T^\Delta_{s_k}(M) = (M_{s_{k+1}} - M_{t_l})^2 - (M_{s_k} - M_{t_l})^2 \\
&= (M_{s_{k+1}} - M_{s_k})^2 + 2(M_{s_{k+1}} - M_{s_k})(M_{s_k} - M_{t_l}) = (M_{s_{k+1}} + M_{s_k} - 2M_{t_k})(M_{s_{k+1}} - M_{s_k})
\end{aligned}$$

and assuming that $t = s_n \in \Delta\Delta'$

$$\begin{aligned}
&T_t^{\Delta\Delta'}(T^\Delta(M)) = \sum_{k=0}^{n-1}\big(T^\Delta_{s_{k+1}}(M) - T^\Delta_{s_k}(M)\big)^2 \\
&\le \sup_{k\le n}(M_{s_{k+1}} + M_{s_k} - 2M_{t_k})^2 \sum_{k=0}^{n-1}(M_{s_{k+1}} - M_{s_k})^2 \\
&= \sup_{k\le n}(M_{s_{k+1}} + M_{s_k} - 2M_{t_k})^2 T_t^{\Delta\Delta'}(M)
\end{aligned}$$

By taking expectation and using Cauchy-Schwartz inequality

$$E_P\left(T_t^{\Delta\Delta'}(T^\Delta(M))\right) \le E_P\left(\sup_{k\le n}(M_{s_{k+1}} + M_{s_k} - 2M_{t_k})^4\right)^{1/2} E_P\big(\big\{T_t^{\Delta\Delta'}(M)\big\}^2\big)^{1/2}$$

Since for P-almost all ω $M_s(\omega)$ is a continuous martingale, it is uniformly continuous on the compact $[0,t]$,

$$\sup_{k\le n}|M_{s_{k+1}} + M_{s_k} - 2M_{t_k}| \to 0$$

P-a.s. as $|\Delta|, |\Delta'| \to 0$. Since $|M_t(\omega)| \le C$, covergence in $L^p(\Omega)$ follows as well.

In order to complete the proof we show that

$$E_P\big(\big\{T_t^\Delta(M)\big\}^2\big)$$

remains bounded as $|\Delta| \to 0$.

Assuming that $t = t_n \in \Delta$, denoting $\Delta M_k = (M_{t_k} - M_{t_{k-1}})$

$$\begin{aligned}
&\big\{T_t^\Delta(M)\big\}^2 = \sum_{k=1}^n(\Delta M_k)^4 + 2\sum_{k=1}^n\left(\sum_{j>k}^n(\Delta M_j)^2\right), E_P\left(\big\{T_t^\Delta(M)\big\}^2\right) \\
&\le E_P\left(T_t^\Delta(M)\sup_{k\le n}(\Delta M_k)^2\right) + 2\sum_{k=1}^n E_P\left((M_t - M_{t_k})^2(\Delta M_k)^2\right)
\end{aligned}$$

where in the last term we have taken conditional expectation with respect to $\mathcal{F}_{t_k}$ and used the martingale property

$$E_P\big(M_{t_n}^2 - M_{t_k}^2\big|\mathcal{F}_{t_k}\big) = E_P\big((M_t - M_{t_k})^2\big|\mathcal{F}_{t_k}\big)$$

We get

$$\begin{aligned}
&E_P\left(\big\{T_t^\Delta(M)\big\}^2\right) \le E_P\left(T_t^\Delta(M)\sup_{k\le n}\left\{(\Delta M_k)^2 + 2(M_t - M_{t_k})^2\right\}\right) \\
&\le E_P(T_t^\Delta(M))12C^2 = E_P(M_t^2)12C^2 \le 12C^4
\end{aligned}$$

This shows that for each t and every sequence of partitions Δ_n with $|\Delta_n| \to 0$,

$T_t^{\Delta_n}(M)$ is a Cauchy sequence in $L^2(\Omega)$.

Since for fixed k, n $(T_t^{\Delta_n}(M) - T_t^{\Delta_k}(M))$ is a martingale, by the Doob L^p-martingale inequality

$$E_P\left(\sup_{s\le t}\big(T_s^{\Delta_n}(M) - T_s^{\Delta_k}(M)\big)^2\right) \le 4E_P\left(\big(T_t^{\Delta_n}(M) - T_t^{\Delta_k}(M)\big)^2\right)$$

which means that $T_s^{\Delta}(M)$ is a Cauchy sequence in $L^2(\Omega)$ uniformly on each compact $[0,t]$.

Therefore there exists a limiting process $[M_t]$ such that

$$E_P\left(\sup_{s\le t}\big([M]_s - T_s^{\Delta}(M)\big)^2\right) \to 0$$

as $|\Delta_n| \to 0$, which does not depend on the choice of the sequence (Δ_n). In particular there is a subsequence n_j such that

$$\sup_{s\le t}\big|[M]_s - T_s^{\Delta}(M)\big| \to 0 \qquad P\text{-almost surely}.$$

It follows that $[M]_s$ is non-decreasing since $T_s^{\Delta}(M)$ is non-decreasing.

Since the approximating processes $T_s^{\Delta}(M)$ are continuous and converging P-almost surely uniformly on compacts, by the Ascoli-Arzela equicontinuity criterium it follows that the limiting process $[M]_t$ is almost surely continuous.

Next we check the martingale property: for $s \le t$, $A \in \mathcal{F}_s$

$$E_P\left((M_t^2 - M_s^2)\mathbf{1}_A\right) = E_P\left((T_t^{\Delta}(M) - T_s^{\Delta}(M))\mathbf{1}_A\right) \to E_P\left(([M]_t - [M]_s)\mathbf{1}_A\right)$$

since $T_t^{\Delta}(M) \stackrel{L^2}{\to} [M]_t$. Therefore $(M_t^2 - [M]_t)$ is a true martingale and by the uniqueness of the Doob-Meyer decomposition $[M]_t = \langle M\rangle_t$. (This does not hold for processes with jumps!) $\square$.

Corollary 12. *Let M_t be a continuous local martingale. Then the process*

$$[M]_t = \lim_{|\Delta|\to 0}\sum_{k=1}^{\infty}\big(M_{t\wedge t_k} - M_{t\wedge t_{k-1}}\big)^2$$

exists as a limit in probability, it is non-decreasing and we have $[M]_t = \langle M\rangle_t$ in the Doob-Meyer decomposition

$$M_t^2 = M_0^2 + [M]_t + N_t$$

where N_t is a local martingale with $N_0 = 0$.

By polarizarion we obtain also the quadratic covariation of two **continuous** *local martingales M_t and $\widetilde{M}_t$,*

$$[M]_t = \lim_{|\Delta|\to 0}\sum_{k=1}^{\infty}\big(M_{t\wedge t_k} - M_{t\wedge t_{k-1}}\big)\big(\widetilde{M}_{t\wedge t_k} - \widetilde{M}_{t\wedge t_{k-1}}\big)$$

which coincides with the predictable covariation $\langle M, \widetilde{M}\rangle_t$.

Proof Without loss of generality we assume that $M_0 = 0$. There is a localizing sequence $\tau_n \uparrow \infty$ of stopping times such that and $M_t^{\tau_n}$ is a true martingale with $|M_t^{\tau_n}| \le n$.

$N_t^{(n)} = \big(M^2_{t\wedge\tau_n} - [M^{\tau_n}]_t\big)$ is a true martingale which is constant on the interval $[\tau_n, \infty)$.

Since $N_t^{(n+1)} = \big(M^2_{t\wedge\tau_{n+1}} - [M^{\tau_{n+1}}]_t\big)$ is also a true martingale, by the uniqueness of the Doob-Meyer decomposition it follows that

$$[M^{\tau_{n+1}}]_t \mathbf{1}(\tau_n > t) = [M^{\tau_n}]_t \mathbf{1}(\tau_n > t)$$

Define

$$[M]_t(\omega) = \sum_{n=1}^{\infty} \mathbf{1}(\tau_{n-1} < t \le \tau_n)[M^{\tau_n}]_t$$

with $\tau_{n-1} \equiv 0$. Note that this sum for each ω contains finitely many nonzero terms.

It follows that $(M_t^2 - [M]_t)$ is a local martingale with localizing sequence τ_n.

Next we show convergence in probability of $T_t^{\Delta}(M)$ to $[M]_t$ for fixed t:

$$P\bigg(\big|[M]_t - T_t^{\Delta}(M)\big| > \varepsilon \bigg) =$$
$$P\bigg(\{\tau_n \le t\} \bigcap \Big\{ \big|[M]_t - T_t^{\Delta}(M)\big| > \varepsilon \Big\} \bigg) + P\bigg(\{\tau_n > t\} \bigcap \Big\{ \big|[M]_{t\wedge\tau_n} - T^{\Delta}_{t\wedge\tau_n}(M)\big| > \varepsilon \Big\} \bigg)$$
$$P\big(\tau_n \le t\big) + P\bigg(\big|[M^{\tau_n}]_t - T_t^{\Delta}(M^{\tau_n})\big| > \varepsilon \bigg)$$

where for n large enough the first term is is arbitrarily small since $\mathbf{1}(\tau_n \le t) \to 0$ P-a.s, and for such fixed n we let $|\Delta| \to 0$ to make the second term small $\square$.

15 Ito-isometry and stochastic integral

Proposition 9. *Let $\mathcal{M}^2$ be the space of continuous martingales $M_t(\omega)$, bounded in $L^2(\Omega)$, with norm*

$$\| M \|^2_{\mathcal{M}^2} := E_P\big(M^2_\infty\big) = E_P\big(\langle M \rangle_\infty\big)$$

$\mathcal{M}^2$ is complete and it is an Hilbert space with scalar product

$$(M,N)_{\mathcal{M}^2} := E_P\big(M_\infty N_\infty\big) = E_P\big(\langle M, N\rangle_\infty\big)$$
$$E_P\bigg(\sup_{t\ge 0} M_t^2\bigg)^{1/2} \le 2 \| M \|_{\mathcal{M}^2} = E_P\bigg(\sup_{t\ge 0} M_t^2\bigg)^{1/2}$$

by Doob's L^p martingale inequality

Proof Since $\sup_{t\ge 0} E_P(M_t^2) < \infty$, $M_t \to M_\infty$ P-almost surely and in $L^2(P)$.

To show that $\mathcal{M}^2$ is complete, if $(M^{(n)})_{n\in\mathbb{N}} \subseteq \mathcal{M}^2$ is a Cauchy sequence, then $(M^{(n)}_\infty)_{n\in\mathbb{N}}$ is a Cauchy sequence in the complete space $L^2(\Omega)$, therefore $\exists M_\infty \in L^2(\Omega)$ such that $E_P\big((M^{(n)}_\infty - M_\infty)^2\big) \to 0$.

Define $M_t(\omega) := E_P(M_\infty|\mathcal{F}_t)(\omega)$, it follows that $M^{(n)} \to M \in \mathcal{M}^2$, which implies

$$E_P\left(\sup_{t\geq 0}(M_t - M_t^{(n)})^2\right) \to 0$$

In particular there is a subsequence (n_j) such that for P-almost all ω

$$\sup_{t\geq 0}\left|M_t^{(n_j)}(\omega) - M_t(\omega)\right| \to 0$$

which implies that $t \mapsto M_t(\omega)$ is continuous. $\square$.

Definition 25. *We say that the process $Y(s,\omega)$ is a simple predictable with respect to the filtration $(\mathcal{F}_t)$, if it is adapted and left-continuous taking finitely many values, that is*

$$Y_s(\omega) := \sum_{i=1}^n \mathbf{1}_{(a_i,b_i]}(s)\eta_i(\omega), \quad n \in \mathbb{N}$$

with $0 \leq a_1 < b_1 \leq a_2 < b_2 \leq \dots < b_{n-1} \leq a_n < b_n < \infty$ *and* $\eta_i(\omega)$ *is* $\mathcal{F}_{a_i}$*-measurable.*

Definition 26. *Given the filtration $(\mathcal{F}_t)_{t\geq 0}$ we define on the measurable space $\Omega\times\mathbb{R}^+$ the predictable σ-algebra $\mathcal{P}$ generated by the left continuous $(\mathcal{F}_t)$-adapted processes. When $(\omega,t) \mapsto Y_t(\omega)$ is $\mathcal{P}$-measurable, we say that the processY is $(\mathcal{F}_t)$-predictable.*

Lemma 15. *Let $(M_t) \in \mathcal{M}^2$ a continuous martingale, and $Y_t \in \mathcal{S}$ a bounded simple predictable process with representation 18. We define the Ito integral as*

$$(Y\cdot M)_t := \int_0^t Y_s dM_s := \sum_{i=1}^n \eta_i(M_{b_i\wedge t} - M_{a_i\wedge t})$$

For $Y \in \mathcal{S}$, the map $Y \mapsto \int_0^\infty Y_s dM_s$ is an isometry between $L_a^2\big(\Omega\times\mathbb{R}^+, P(d\omega)\otimes \langle M\rangle(\omega,dt)\big)$ and $\mathcal{M}^2$, with

$$E_P\left(\left\{\int_0^\infty Y_s dM_s\right\}^2\right) = E_P\left(\int_0^\infty Y_s^2 d\langle M\rangle_s\right)$$

We have the property: for all $(N_t) \in \mathcal{M}^2$,

$$\langle (Y\cdot M), N\rangle_t := \int_0^\infty Y_s d\langle M,N\rangle_s := \sum_{i=1}^n \eta_i\big(\rangle M_{b_i\wedge t}, N\rangle - \langle M_{a_i\wedge t}, N\rangle\big)$$

Proof By taking conditional expectation and using the martingale property

$$E_P\left(\left\{\int_0^\infty Y_s dM_s\right\}^2\right) =$$

$$\sum_{i=1}^n E_P\Big((\eta_i^2(M_{b_i}-M_{a_i})^2\Big) + 2\sum_{i=1}^n\sum_{1\leq j<n} E_P\Big(\eta_i\eta_j(M_{b_i}-M_{a_i})(M_{b_j}-M_{a_j})\Big) =$$

$$\sum_{i=1}^n E_P\Big((\eta_i^2 E_P\big((M_{b_i}-M_{a_i})^2|\mathcal{F}_{a_i}\Big) + 2\sum_{i=1}^n\sum_{1\leq j<n} E_P\Big(\eta_i\eta_j(M_{b_j}-M_{a_j})E_P(M_{b_i}-M_{a_i}|\mathcal{F}_{a_i}\Big) =$$

$$\sum_{i=1}^n E_P\Big((\eta_i^2(\langle M\rangle_{b_i} - \langle M\rangle_{a_i})\Big) = E_P\left(\int_0^\infty Y_s^2 d\langle M\rangle_s\right)$$

Theorem 15. *(Kunita-Watanabe inequality) Let $(N_t),(M_t)\in\mathcal{M}^2$ and $(Y_s),(U_s)$ jointly measurable processes. Then P-almost surely for $t\in[0,+\infty]$*

$$\int_0^t |Y_s||U_s|d|\langle M,N\rangle_s| \le \left(\int_0^t Y_s^2 d\langle M\rangle_s\right)^{1/2}\left(\int_0^t U_s^2 d\langle N\rangle_s\right)^{1/2}$$

By Hölder inequality, we have also for $p,q>1,\ p^{-1}+q^{-1}=1$

$$E_P\left(\int_0^t |Y_s||U_s|d|\langle M,N\rangle_s|\right) \le E_P\left(\left\{\int_0^t Y_s^2 d\langle M\rangle_s\right\}^{p/2}\right)^{1/p}\left(\int_0^t U_s^2 d\langle N\rangle_s\right\}^{q/2}\right)^{1/q}$$

Note that we need joint measurability since we want that the maps $t\mapsto Y(t,\omega)$ $t\mapsto U(t,\omega)$ are $\mathcal{B}(\mathbb{R}^+)$-measurable for all $\omega\in\Omega$, in order to use the Lebesgue-Stieltjes integral.

The integral on the left hand side is a Lebesgue-Stieljes integral taken ω-wise with respect to the total variation of the process $\langle M,N\rangle_t(\omega)$

Proof Note that $\forall r\in\mathbb{R}$ $(M_t+rN_t)\in\mathcal{M}^2$ and

$$0\le\langle M+rN\rangle_t=\langle M\rangle_t+r^2\langle N\rangle_t+2r\langle N,M\rangle_t$$

Since this quadratic equation in the unknown r has at most one real solution, we get the inequality for the discriminant

$$\langle N,M\rangle_t^2-\langle M\rangle_t\langle N\rangle_t\le 0 \iff \big|\langle N,M\rangle_t\big|\le\sqrt{\langle M\rangle_t}\sqrt{\langle N\rangle_t}$$

The same inequality hold for increments.

By taking

$$Y_s'=|Y_s|,\quad U_s'=|U_s|\frac{d\langle M,N\rangle}{d|\langle M,N\rangle|}(s)$$

where the last term on the right hand side is the Radon-Nikodym derivative of $\langle M,N\rangle$ with respect to its total variation, it is enough to show that

$$\left|\int_0^t Y_sU_sd\langle M,N\rangle_s\right|\le\left(\int_0^t Y_s^2 d\langle M\rangle_s\right)^{1/2}\left(\int_0^t U_s^2 d\langle N\rangle_s\right)^{1/2}$$

Assume that there is a finite Borel-measurable partition of $[0,t]=\cup_{i=1}^n B_i$ and random variables $\widetilde{Y}_i(\omega),\widetilde{U}_i(\omega)$ such that

$$Y_s(\omega)=\sum_{i=1}^n\widetilde{Y}_i(\omega)\mathbf{1}_{B_i}(s)\quad U_s(\omega)=\sum_{i=1}^n\widetilde{U}_i(\omega)\mathbf{1}_{B_i}(s)$$

Denote

$$\Delta V_i=\int_{B_i}dV_s$$

when $V_s = \langle M,N\rangle_s, \langle M\rangle_s, \langle N\rangle_s$, is a process of finite variation.

$$\begin{aligned}
&\left|\int_0^t Y_s U_s d\langle M,N\rangle_s\right| = \left|\sum_{i=0}^n \widetilde{Y}_i\widetilde{U}_i \Delta\langle M,N\rangle_i\right| \\
&\le \sum_{i=0}^n |\widetilde{Y}_i||\widetilde{U}_i| \sqrt{\Delta\langle M\rangle_i}\sqrt{\Delta\langle N\rangle_i} \\
&\le \left(\sum_{i=0}^n \widetilde{Y}_i^2 \Delta\langle M\rangle_i\right)^{1/2}\left(\sum_{i=0}^n \widetilde{U}_i^2 \Delta\langle N\rangle_i\right)^{1/2} \\
&= \left(\int_0^t Y_s^2 d\langle M\rangle_s\right)^{1/2}\left(\int_0^t U_s^2 d\langle N\rangle_s\right)^{1/2}
\end{aligned}$$

where we used the Cauchy Schwartz inequality for sums. Since the sets B_i are Borel-measurable but not necessarily intervals the integrals are Lebesgue-Stieltjes integrals.

The result follows for example by the monotone convergence theorem for the Lebesgue-Stieltjes integrals splitting first the integrands into positive and negative part anb then approximating from below by simple $\mathcal{F}\otimes\mathcal{B}(\mathbb{R}^+)$-measurable processes $\square$.

Lemma 16. . *An $(\mathcal{F}_t)$-adapted process (M_t) is a martingale if and only for all* **bounded** *$(\mathcal{F}_t)$-stopping times τ, the random variable $M_\tau(\omega) \in L^1(P)$ and*

$$E_P\big(M_\tau\big) = E_P(M_0)$$

Proof The necessity follows from Doob's optional stopping theorem.
Sufficiency: let $s \le t$ and $A \in \mathcal{F}_s$. Define the random time

$$\tau(\omega) := s\,\mathbf{1}_A(\omega) + t\,\mathbf{1}_{A^c}(\omega)$$

Note that $\forall u \ge 0$

$$\{\tau(\omega) \le u\} = \begin{cases} \Omega & t \le u \\ A & s < u \le t \\ \emptyset & 0 \le s \le u \end{cases}$$

which is $\mathcal{F}_u$ measurable in all cases, therefore τ is a bounded stopping time.

$$\begin{aligned}
&E_P(M_0) = E_P\big(M_\tau\big) = E_P(\mathbf{1}_A M_s + \mathbf{1}_{A^c} M_t) = \\
&E_P\big(M_t\big) + E_P\big(\mathbf{1}_A(M_s - M_t)\big) = E_P(\big(M_0\big) - E_P\big(\mathbf{1}_A(M_t - M_s)\big) \\
&\Longrightarrow E_P\big(\mathbf{1}_A(M_t - M_s)\big) = 0
\end{aligned}$$

which gives the martingale property.

Theorem 16. *Let $(M_t) \in \mathcal{M}^2$ and $Y(s,\omega)$ a progressively measurable process with*

$$E_P\left(\int_0^\infty Y_s^2 d\langle M\rangle_s\right) < \infty$$

1. *There exists an unique martingale in $\mathcal{M}^2$ which will be denoted by $(Y \cdot M)_t = \int_0^t Y_s dM_s$ such that $\forall\ (N_t) \in \mathcal{M}^2$,*

$$E_P\Big((Y \cdot M)_\infty N_\infty\Big) = E_P\Big(\langle Y \cdot M, N\rangle_\infty\Big) = E_P\Big(\int_0^\infty Y_s d\langle M, N\rangle_s\Big) \quad (18)$$

2. $(Y \cdot M)_0 = 0$ *and for all $(N_t) \in \mathcal{M}^2$*

$$(Y \cdot M)_t N_t - \int_0^t Y_s d\langle M, N\rangle_s,$$

is a true martingale.

3. *Ito isometry holds:*

$$\langle Y \cdot M\rangle_t = \int_0^t Y_s^2 d\langle M, M\rangle_s \quad \forall t \in [0, +\infty]. \quad (19)$$

From the uniqueness result it follows that for simple predictable integrands this definition of Ito integral coincides with the Riemann sums definition given in (15).

Proof: (Uniqueness) Let $(L_t) \in \mathcal{M}_2$ with the same property. Then taking $N_t = \{(Y \cdot M)_t - L_t\} \in \mathcal{M}^2$ we have

$$E_P\Big(\sup_{t\geq 0}\{(Y \cdot M)_\infty - L_\infty\}^2\Big) \leq 4E_P\Big(\{(Y \cdot M)_\infty - L_\infty\}^2\Big) = 0$$

which means that $\{L_t\}$ and $\{Y \cdot M)_t\}$ are indistinguishable processes.

(Existence): The map

$$N_\infty \mapsto \varphi(N) := E_P\Big(\int_0^\infty Y_s d\langle M, N\rangle_s\Big)$$

is linear since the predictable covaration is bilinear. It is also continuous in $\mathcal{M}^2$ norm: by Kunita-Watanabe and Cauchy-Schwartz inequalities

$$|\varphi(N)| = \Big|E_P\Big(\int_0^\infty Y_s d\langle M, N\rangle_s\Big)\Big| \leq E_P\Big(\int_0^\infty Y_s^2 d\langle M\rangle_s\Big)^{1/2} E_P\Big(\langle N\rangle_\infty\Big)^{1/2} =$$
$$E_P\Big(\int_0^\infty Y_s^2 d\langle M\rangle_s\Big)^{1/2} \parallel N \parallel_{\mathcal{M}^2}$$

under the assumption that

$$E_P\Big(\int_0^\infty Y_s^2 d\langle M\rangle_s\Big)^{1/2}$$

By the Riesz representation theorem in the Hilbert space $\mathcal{M}^2$ there exists a continuous martingale $\{(Y \cdot M)_t\} \in \mathcal{M}^2$ such that

$$E_P\Big(\int_0^\infty Y_s d\langle M, N\rangle_s\Big) = \varphi(N) = \big((Y \cdot M), N\big)_{\mathcal{M}^2} =$$
$$E_P\Big((Y \cdot M)_\infty N_\infty\Big) = E_P\Big(\langle Y \cdot M, N\rangle_\infty\Big)$$

Note: if you think about it, up to now we did not need predictability or progressive measurability of (Y_s), since by Kunita Watanabe inequality we just need joint measurability.

Next we show that

$$X_t := N_t \int_0^t Y_s dM_s - \int_0^t Y_s d\langle M, N\rangle_s$$

is a martingale for all $N \in \mathcal{M}^2$ and by definition of predictable covariation

$$\langle Y \cdot M, N\rangle_t = \int_0^t Y_s d\langle M, N\rangle_s$$

which implies 19.

Let τ be a $(\mathcal{F}_t)$-stopping time. By Cauchy Schwartz and Kunita Watanabe inequalities it follows that $X_\tau \in L^1(P)$.

Since $(Y \cdot M)_t$ and (N_t) are uniformly integrable martingales (since they are bounded in $L^2(\Omega, \mathcal{F}, P)$), we write

$$E_P\big((Y \cdot M)_\tau N_\tau\big) = E_P\bigg(E_P\big((Y \cdot M)_\infty | \mathcal{F}_\tau\big) N_\tau\bigg) = E_P\bigg((Y \cdot M)_\infty N_\tau\bigg) =$$

$$E_P\bigg((Y \cdot M)_\infty N^\tau_\infty\bigg) = E_P\bigg(\langle (Y \cdot M), N^\tau\rangle_\infty\bigg) = \text{ by the defining property (20)}$$

$$= E_P\bigg(Y \cdot \langle M, N^\tau\rangle_\infty\bigg) = E_P\bigg(Y \cdot \langle M, N\rangle_\tau\bigg)$$

and the result follows by lemma 16. Note that here we need the assumption that $Y_s(\omega)$ is progressively measurable, since in order to apply the lemma we need to show

$$\int_0^t Y_s d\langle M, N\rangle_s$$

and X_t are $(\mathcal{F}_t)$-adapted.

To show that $(Y \cdot M)_0 = 0$, take $N_t \equiv N_0 \; \forall t \in \mathcal{M}_0$, a constant $\mathcal{F}_0$-measurable process.

Note that in this case by Kunita-Watanabe inequality $|\langle M, N\rangle_t| \le \sqrt{\langle M\rangle_t}\sqrt{\langle N\rangle_t} = 0$ since $[N, N]_t = \langle N, N\rangle_t = 0$ by the definition of quadratic variation.

Then

$$0 = E_P\bigg(\int_0^t Y_s d\langle M, N\rangle_s\bigg) = E_P\bigg((Y \cdot M)_t N_t\bigg) = E_P\bigg((Y \cdot M)_t N_0\bigg) = E_P\bigg((Y \cdot M)_0 N_0\bigg)$$

which implies $(Y \cdot M)_0 = 0$ since $N_0 \in L^2(\Omega, \mathcal{F}_0, P)$ is arbitrary.

By taking $N_t = M_t$, we also obtain

$$\langle M, (Y \cdot M)\rangle_t = \int_0^t Y_s d\langle M, M\rangle_s$$

and by taking $N_t = (Y \cdot M)_t$,

$$\langle (Y \cdot M), (Y \cdot M)\rangle_t = \int_0^t Y_s d\langle M, (Y \cdot M)\rangle_s = \int_0^t Y_s^2 d\langle M, M\rangle_s$$

since in Lebesgue-Stieltjes integrals we have

$$d\langle M,(Y\cdot M)\rangle_s = d_s\left(\int_0^s Y_u d\langle M,M\rangle_u\right) = Y_s d\langle M,M\rangle_s$$

Remark This proof is a bit abstract since we used Riesz representation theorem. A more standard proof for predictable integrands consists in approximating the integrand Y_s by a sequence $(Y_s^{(n)})$ of simple predictable (left-continuous and adapted) integrands in the space $L^2\big(\Omega\times\mathbb{R}^+,\mathcal{P},P(d\omega)\langle M\rangle(\omega,dt)\big)$ obtaining by Ito isometry a Cauchy sequence of Ito integrals in $\mathcal{M}^2$.

A constructive extension of this line of proof to progressively measurable integrands for which the Lebesgue-Stieltjes integral $\int_0^t Y_s d\langle M\rangle_s$ is not necessarily well defined as a Riemann-Stieltjes integral, is a bit technical, since one needs an intermediate approximation step in order to work with Riemann sums (see for example the details in Karatzas and Schreve).

Proposition 10. *Let (M_t) a continuous local martingale and $(Y_t(\omega))$ a progressively measurable process with*

$$\int_0^t Y_s^2 d\langle M\rangle_s < \infty \qquad P \textit{ almost surely } \forall t\in\mathbb{R}^+$$

Then there is a local martingale which we denote by $(Y\cdot M)_t = \int_0^t Y_s dM_s$ such that $(Y\cdot M)_0 = 0$ and

$$\langle (Y\cdot M),N\rangle_t = \int_0^\infty Y_s d\langle M,N\rangle_s \tag{20}$$

for every continuous local martingale N.

Proof Let τ_n a localizing sequence of stopping times $\tau_n\uparrow\infty$ such that both stopped processes M^{τ_n} and N^{τ_n} are martingales in $\mathcal{M}^2$. By 16 there exists a process $(Y\cdot M^{\tau_n})_t$ verifying the theorem in the stochastic interval $\{(\omega,t): t\le \tau_n(\omega)\}$.

Lemma 17. *Let (M_s) a continuous local martingale $(Y_s^{(n)})_{n\in\mathbb{N}}$ a sequence of locally bounded progressively measurable integrands such that for all $s|Y_s^{(n)}(\omega)|\to 0$ P-almost surely and there is a locally bounded process $Y_s(\omega)$ such that $|Y_s^{(n)}(\omega)|\le Y_s(\omega)$,$\forall s,n$, P-almost surely.*

Then for all $t\ge 0$

$$\sup_{s\le t}\left|\int_0^t Y_s^{(n)}dM_s\right|\to 0 \qquad \textit{in probability as } n\to\infty$$

Let $\tau(\omega)$ be a stopping time such that both stopped processes M_s^τ and Y_s^τ are bounded. Then by the bounded convergence theorem

$$E_P(\int_0^\tau (Y_s^{(n)})^2 d\langle M_s\rangle\Big) \to 0 \text{ as } n\to\infty$$

which implies

$$\int_0^\tau Y_s^{(n)} dM_s \to 0 \quad \text{in } L^2(\Omega, \mathcal{F}, P) \text{ and in probability as } n \to \infty$$

To comple the argument we for any fixed t choose the localizing stopping time such that $P(\tau \le t) < \varepsilon$ and conclude as in corollary (**??**).

Proposition 11. *(Extension by localization) Let (M_t) a continous local martingale and Y_s is progressive such that $\forall 0 \le t < \infty$*

$$P\left(\int_0^t Y_s^2 d\langle M\rangle_s < \infty\right) = 1$$

Define the stopping time $(\tau_n) := \inf\left\{t \ge 0 : \int_0^t Y_s^2 \langle M\rangle_s \ge n\right\}$. Then $\tau_n \to \infty$ P-a.s.

The stochastic integral defined as

$$\int_0^t Y_s dM_s = \lim_{n\to\infty} \int_0^{t\wedge\tau_n} Y_s dM_s$$

is a local martingale with localizing sequence τ_n.

Proof

Definition 27. *We say that $X_t = X_0 + M_t + A_t$ is a semimartingale when $M_0 = A_0 = 0$, M_t is a continuous local martingale and A_t is $(\mathcal{F}_t)$-adapted with locally finite variation.*

(X_t) is continuous if and only if A_t is continuous.

For Y_t progressive such that $\forall 0 \le t < \infty$

$$\int_0^t Y_s^2 d\langle M\rangle_s < \infty \quad \text{and} \quad \int_0^t |Y_s||dA|_s < \infty \quad P\text{-almost surely}$$

where the integral on the right side is with respect to the total variation of A, we define

$$\int_0^t Y_s dX_s = \int_0^t Y_s dM_s + \int_0^t Y_s dA_s$$

16 Ito representation theorem

Let $B_t = (B_t^{(1)}, \dots, B_t^{(d)})$ a d-dimensional Brownian motion.

Theorem 17. *Let $Y \in L^2(\Omega, \mathcal{F}_T^B, P)$, $T \in (0, +\infty]$ a real valued random variable. Then there is a progressive process $H_s(\omega) \in \mathbb{R}^d$ with*

$$E_P\left(\int_0^T H_s^2 ds\right) < \infty$$

$$Y(\omega) = E_P(Y) + \int_0^T H_s dB_s$$

$H_s(\omega)$ is unique $P(d\omega) \times ds$ almost surely.

Proof Uniqueness: if $\widetilde{H}_s$ has the same property, then by Ito isometry

$$\int_\Omega \left(\int_0^T \big(H_s(\omega) - \widetilde{H}_s(\omega) \big)^2 ds \right) P(d\omega) = 0$$

Existence:

$$\mathcal{H} = \left\{ \int_0^T H_s dB_s : \ H \text{ is progressive and in } L^2(\Omega \times [0,T], dP \times dt) \right\}$$

is a closed subspace of $L^2(\Omega, \mathcal{F}_T^B, P)$, which follows since the space of progressive integrands in $L^2(\Omega \times [0,T], dP \times dt)$ is complete.

We show that it is total, in the sense that if $Y \in L^2(\Omega, \mathcal{F}_T^B, P)$ such that $E_P\left(Y \int_0^T H_s dB_s \right) = 0$ for all progressive $H \in L^2(\Omega \times [0,T], dP \times dt)$, then $Y(\omega) = E_P(Y)$ is deterministic.

The theorem follows since this implies that the random variable $\big(Y(\omega) - E_P(Y)\big)$ coincides with its orthogonal projection on the closed subspace $\mathcal{H}$.

Without loss of generality assume that $E_P(Y) = 0$, otherwise take $\tilde{Y}(\omega) = (Y(\omega) - E_P(Y))$. For $f \in L^2([0,T], dt)$ consider the square integrable martingale

$$M_t^{(f)} = \exp\left(i \int_0^t f(s) dB_s + \frac{1}{2} \int_0^t f(s)^2 ds \right)$$

By Ito formula

$$M_T^{(f)} - 1 = i \int_0^T M_s^{(f)} f(s) dB_s$$

Since the real and imaginary parts of the right hand side are stochastic integral in $\mathcal{H}$, it follows that

$$0 = E_P\left(Y(M_T^{(f)} - 1) \right) = E_P\left(Y M_T^{(f)} \right) - E_P(Y) = E_P\left(Y M_T^{(f)} \right)$$

When $f(s) = \sum_{i=1}^n \theta_k \mathbf{1}_{[0,t_k]}(s)$ for $\theta_k \in \mathbb{R}^d$ $t_k \in [0,T]$, $k = 1, \dots n, n \in \mathbb{N}$ it follows that

$$\begin{aligned}
0 &= E_P\left(Y \exp\left(i \sum_{k=1}^n \theta_k \cdot B_{t_k} + \frac{1}{2} \sum_{h,k=1}^n \theta_h \theta_k \ (t_h \wedge t_k) \right) \right) \\
&= E_P\left(Y \exp\left(i \sum_{k=1}^n \theta_k \cdot B_{t_k} \right) \right) \exp\left(\frac{1}{2} \sum_{h,k=1}^n \theta_h \theta_k \ (t_h \wedge t_k) \right) \\
&\Longrightarrow E_P\left(Y \exp\left(i \sum_{k=1}^n \theta_k \cdot B_{t_k} \right) \right) = 0
\end{aligned}$$

By the Lévy inversion theorem, which holds not on only for probability measures but also for finite signed measure, a measure is characterized by its characteristic function.

The signed measure on $\mathbb{R}^{k\times d}$ defined as

$$\mu_{t_1,\dots t_n}(A_1 \times A_n) := E_P\bigg(Y\mathbf{1}\big(B_{t_1} \in A_1, \dots, B_{t_n} \in A_n\big)\bigg)$$

for $A_k \in \mathcal{B}(\mathbb{R}^d), k = 1, \dots, n$ is equal to zero.

Since the cylinders generate the σ-algebra $\mathcal{F}_T^B$, it follows by Dynkin extension theorem that

$$E_P(Y\mathbf{1}_F) = 0 \qquad \forall F \in \mathcal{F}_T^B$$

By assumption $Y \in \mathcal{F}_T^B$ measurable, the result follows by taking $F = \mathbf{1}(Y > 0)$, and $F = \mathbf{1}(Y < 0)$.

Corollary 13. *Let (M_t) a martingale in the Brownian filtration bounded in L^2, that is $E_P(M_\infty^2) < \infty$ Then*

$$M_t = E_P(M_\infty|\mathcal{F}_t^B)(\omega) = M_0 + \int_0^t H_s dB_s$$

where the integrand $H \in L^2(\Omega\times\mathbb{R}^+, dP\times dt)$ is progressive and unique $P(d\omega)\times dt$ almost surely. Note that since $\mathcal{F}_0^B$ is P-trivial, $M_0 = E_P(M_0) = E_P(M_\infty)$ is deterministic.

Let $F(\omega) = f(B_T(\omega))$ for some $f(x) \in L^2(\mathbb{R}, \gamma(x)dx)$.

$$\begin{aligned}
&E\big(f(B_T)\big|\mathcal{F}_t\big) = E\big(f(B_t + (B_T - B_t))\big|\mathcal{F}_t\big) \\
&= E\bigg(f(x + G\sqrt{T-t})\bigg)\bigg|_{x=B_t(\omega)} \\
&= \int_{\mathbb{R}} f(B_t(\omega) + y\sqrt{T-t})\gamma(y)dy = \\
&\int_{\mathbb{R}} f(u)\frac{1}{\sqrt{T-t}}\gamma\bigg(\frac{B_t - u}{\sqrt{T-t}}\bigg)dy =
\end{aligned}$$

where $G(\omega) \sim \mathcal{N}(0,1)$ is a standard gaussian random variable with

$$P(G \in dy) = \gamma(y)dy = (2\pi)^{-1/2}\exp\big(-y^2/2\big)dy$$

Next we apply Ito formula and integration by parts to

$$g(B_t, u; t, T) = \frac{1}{\sqrt{T-t}}\gamma\bigg(\frac{B_t - u}{\sqrt{T-t}}\bigg) = \frac{P(B_T \in du|B_t)}{du}$$

We do the calculation in steps:

$$\gamma'(y) = -y\gamma(y),\ \gamma''(y) = \gamma(y)(y^2-1),\ \frac{d}{dt}(T-t)^{-1/2} = \frac{1}{2}(T-t)^{-3/2}$$

and for a continuous semimartingale Y_t

$$d\gamma(Y_t) = \gamma(Y_t)\Big(-Y_t dY_t + \frac{1}{2}(Y_t^2 - 1)d\langle Y\rangle_t\Big)$$

Now for $Y_t = \frac{(B_t - u)}{\sqrt{T-t}}$ we have using integration by parts

$$dY_t = \frac{1}{\sqrt{T-t}}dB_t + \frac{1}{2}\frac{(B_t-u)}{(T-t)^{3/2}}dt, \quad d\langle Y\rangle_t = \frac{1}{(T-t)}dt$$

Therefore

$$d\gamma(Y_t) = \gamma(Y_t)\Big(-\frac{(B_t-u)}{T-t}dB_t - \frac{1}{2}\frac{(B_t-u)^2}{(T-t)^2}dt + \frac{1}{2}\Big(\frac{(B_t-u)^2}{T-t} - 1\Big)\frac{1}{T-t}dt\Big) =$$
$$-\gamma(Y_t)\Big(\frac{B_t-u}{T-t}dB_t + \frac{1}{2(T-t)}dt\Big)$$

Integrating by parts:

$$d\Big(\frac{1}{\sqrt{T-t}}\gamma(Y_t)\Big) = \frac{1}{\sqrt{T-t}}\gamma(Y_t)\Big(-\frac{B_t-u}{T-t}dB_t - \frac{1}{2(T-t)}dt + \frac{1}{2(T-t)}dt\Big)$$
$$= \frac{1}{\sqrt{T-t}}\gamma\Big(\frac{B_t-u}{\sqrt{T-t}}\Big)\Big(\frac{u-B_t}{T-t}\Big)dB_t$$

Therefore we have simply

$$g(B_t,u,t,T) = g(0,u,0,T) + \int_0^t g(B_s,u,s,T)\Big(\frac{u-B_s}{T-s}\Big)dB_s$$

and the stochastic exponential representation

$$g(B_t,u,t,T) = g(0,u,0,T)\mathcal{E}\Big(\int_0^{\cdot}\Big(\frac{u-B_s}{T-s}\Big)dB_s\Big)_t$$
$$= g(0,u,0,T)\exp\Big(\int_0^t\Big(\frac{u-B_s}{T-s}\Big)dB_s - \frac{1}{2}\int_0^t\Big(\frac{u-B_s}{T-s}\Big)^2 ds\Big)$$

Integrating with respect to du we get

$$E_P(f(B_T)|\mathcal{F}_t) =$$
$$\int_{\mathbf{R}} f(u)g(B_t,u,t,T)du = \int_{\mathbf{R}} f(u)g(0,u,0,T)du + \int_{\mathbf{R}}\Big(\int_0^t f(u)\Big(\frac{u-B_s}{T-s}\Big)g(B_s,u,s,T)dB_s\Big)du$$
$$= E_P(f(B_T)) + \int_0^t\Big(\int_{\mathbf{R}} f(u)\Big(\frac{u-B_s}{T-s}\Big)g(B_s,u,s,T)du\Big)dB_s$$
$$= E_P(f(B_T)) + \int_0^t \frac{E_P\big(f(B_T)(B_T-B_s)\big|\mathcal{F}_s\big)}{(T-s)}dB_s$$

where we used a stochastic Fubini theorem (to be explained in the next paragraph) in order to invert the order of integration w.r.t. between du and dB_s. Note that since by assumption $f(B_T) \in L^2(\Omega)$, the term

$$\frac{E_P(f(B_T)(B_T-B_s)|\mathcal{F}_s)}{T-s} = \frac{\mathrm{Cov}(f(B_T), B_T|\mathcal{F}_s)}{\mathrm{Var}(B_T|\mathcal{F}_s)}$$
$$= \frac{E_P((f(B_T)-f(B_s))(B_T-B_s)|\mathcal{F}_s)}{T-s}$$

is the conditional correlation between $f(B_T)$ and B_T given $\mathcal{F}_s$.

Note also that we proved in between that $g(x,u,s,T)$ satisfies the heat equation

$$\frac{\partial}{\partial s} g(x,u,s,T) + \frac{1}{2}\frac{\partial^2}{\partial x^2} g(x,u,s,T) = 0$$

with boundary condition $g(x,u,T,T) = \delta_0(x-u)$ the Dirac delta function in the sense of Schwartz distributions.

Up to now we just assumed that $f \in L^2(\mathbb{R}, d\gamma)$. When $f(x) = f(0) + \int_0^x f'(u)du$ is absolutely continuous with respect to Lebesgue measure we can use the *gaussian integration by parts* formula

$$E(f(B_t)B_t) = tE_P(f'(B_t))$$

which holds when $B_t \sim \mathcal{N}(0,t)$ gaussian.

In this case we write Ito's representation also as

$$E_P\big(f(B_T)\big|\mathcal{F}_t\big) = E_P\big(f(B_T)\big) + \int_0^t E_P\big(f'(B_T)\big|\mathcal{F}_s\big)dB_s$$

Example Let $F(\omega) = f\left(\int_0^T h(s)dB_s\right)$, where $h(s) \in L^2([0,T], ds)$ is deterministic and $E_P\big(f(\| h \|_2 G)^2\big) < \infty$, for $G(\omega)$ standard gaussian r.v.

Then we have the representation

$$F(\omega) = E_P\big(f(\| h \|_2 G)\big) + \int_0^T \frac{E_P\left(f\left(\int_0^T h(s)dB_s\right)\int_t^T h(s)dB_s\middle|\mathcal{F}_s\right)}{\int_t^T h(s)^2 ds} h(t)dB_t$$

Hint: define the determistic time change

$$\tau(u) = \inf\left\{t : \int_0^t h(s)^2 ds \geq u\right\}$$

Then by Lévy characterization theorem $\widetilde{B}_u := \int_0^{\tau(u)} h(s)dB_s$ is a Brownian motion and $\mathcal{F}_u^{\widetilde{B}} = \mathcal{F}_{\tau(u)}^B$.

Letting $\widetilde{T} = \int_0^T h(s)^2 ds$.

In Malliavin calculus these ideas are extended to more general setting where there is not need to use the Markov property.

Theorem 18. *Stochastic Fubini theorem.*

Let $(\Theta, \mathcal{A}, \alpha(d\theta))$ be a measurable space, where $\alpha(d\theta)$ is a finite measure, and $H(s,\omega,\theta)$ a jointly measurable process, such that the map $\theta \mapsto H(s,\omega,\theta)$ is $\mathcal{A}$-measurable for each (s,ω) and the map $(s,\omega) \mapsto H(s,\omega,\theta)$ is $(\mathcal{F}_t)$-progressive for each $\theta \in \Theta$.

Assuming that for all t, P-almost surely

$$\int_{[0,t]\times\Theta} H(s,\omega,\theta)^2 (\alpha \otimes \langle M \rangle)(d\theta \times dt) < \infty$$

which by the classical Fubini theorem does not depend on the order of integration.

Then

$$\int_0^t \Big(\int_\Theta H(s,\omega,\theta)\alpha(d\theta) \Big) dM_s = \int_\Theta \Big(\int_0^t H(s,\omega,\theta) dM_s \Big) \alpha(d\theta)$$

is a local martingale which does not depend on the order of integration.

Proof Without loss of generality assume that $\alpha(d\theta)$ is a probability measure. By the definition of joint measurability is a sequence of simple integrands of the form

$$H^{(n)}(s,\omega,\theta) = \sum_{i=1}^n H_i^{(n)}(s,\omega)\mathbf{1}(\theta \in A_i^{(n)})$$

where $(A_1^{(n)}, \dots A_n^{(n)})$ is a measurable partition of Θ and $H_i^{(n)}(s,\omega)$ are progressive processes, such that

$$\int_{[0,T]\times\Theta} \big\{ H^{(n)}(s,\omega,\theta) - H_s(s,\omega,\theta) \big\}^2 d\langle M \rangle_s \alpha(d\theta) \to 0$$

in probability.

By the linearity of Ito integral the stochastic Fubini theorem holds for the simple integrands $H^{(n)}$. Note also that by Jensen inequality

$$\int_0^T \Big(\int_\Theta \big(H^n(s,\omega,\theta) - H(s,\omega,\theta) \big) \alpha(d\theta) \Big)^2 d\langle M \rangle_s$$
$$\leq \int_{[0^T]\times\Theta} \big(H^n(s,\omega,\theta) - H(s,\omega,\theta) \big)^2 \alpha(d\theta) \otimes d\langle M \rangle_s \xrightarrow{P} 0$$

This implies

$$\int_\Theta \Big(\int_0^T H^{(n)}(s,\theta) dB_s \Big) \alpha(d\theta) =$$
$$\int_0^T \Big(\int_\Theta H^{(n)}(s,\theta)\alpha(d\theta) \Big) dB_s \xrightarrow{P} \int_0^T \Big(\int_\Theta H^{(n)}(s,\theta)\alpha(d\theta) \Big) dB_s$$

and since

$$\int_\Theta \Big(\int_0^T \big(H^{(n)}(\omega,s,\theta) - H(\omega,s,\theta) \big)^2 d\langle M \rangle_s \Big) \alpha(d\theta) \xrightarrow{P} 0$$

It follows that

$$\int_\Theta \Big(\int_0^T (H^{(n)}(s,\theta) dB_s \Big) \alpha(d\theta) \xrightarrow{P} \int_\Theta \Big(\int_0^T (H(s,\theta) dB_s \Big) \alpha(d\theta)$$

Proposition 12. *Gaussian integration by parts formula. If $G(\omega) \sim \mathcal{N}(0,1)$ is centered gaussian and $f(x) = f(0) + \int_0^t f'(y)dy$ is absolutely continuous such that both $(f'(G) - f(G)G)$ and $f(G)$ are in $L^1(P)$. Then*

$$E_P\big(f(G)G\big) = E_P\big(f'(G)\big)$$

'**Proof** We recall that the standard gaussian density $\gamma(x)$, satisfies $\gamma'(x) = -x\gamma(x)$ Integrating by parts, for all $a \le b \in \mathbb{R}$

$$f(b)\gamma(b) - f(a)\gamma(a) = \int_a^b (f'(y) - f(y)y)\gamma(y)dy$$

If $f(x)$ is compactly supported, the left-hand side equals zero for $|a|$ and $|b|$ large. As $a \to -\infty$ and $b \to +\infty$ the left hand side converges to $E_P\big(f'(G) - f(G)G\big)$.

More in general we approximate $f(x)$ with a sequence of compactly supported functions. Let $k_n(x) = (1 - |x|/n)^+$. We have $0 \le k_n(x) \le 1$, $\frac{d}{dx}k_n(x) = -n^{-1}\text{sign}(x)bf1(|x| \le n)$, and $\lim_{n\to\infty} k_n(x) = x$, $\forall x \in \mathbb{R}$.

Let $f_n(x) = f(x)k_n(x)$.

$$0 = E(f_n'(G) - f_n(G)G) = E((f'(G) - F(G)G)k_n(G)) + E(f(G)k_n'(G))$$

where we used the chain rule of differentiation. Since $|(f'(G) - F(G)G)k_n(G)| \le (f'(G) - F(G)G) \in L^1(P)$, by Lebesgue' dominated convergence theorem

$$E((f'(G) - F(G)G)k_n(G)) \to E(f'(G) - F(G)G)$$

and $E(|f(G)k_n'(G)|) \le n^{-1}E(|f(G)|) \to 0$

Example the maximum process

Let B_t be a standard Brownian motion starting from zero, $\mathcal{F}_t^B = \sigma(B_s : 0 \le s \le t)$. Define

$$B_t^* = \sup_{0\le s\le t}\{B_s\},$$
$$H_a = \inf\big\{t > 0 : B_t \ge a\big\}$$

respectively the running maximum and the first hitting time of level $a > 0$

Proposition 13. *For $a > 0$, by the reflection principle*

$$P(H_a \le \ell) = P(B_\ell^* \ge a) = 2P(B_\ell > a) = 2\big(1 - \Phi(a/\sqrt{\ell})\big)$$

where $\Phi(x) = P(B_1 \le x)$.

By differentiating with respect to ℓ we obtain the probability density of the hitting time H_a

$$\frac{P(H_a \in d\ell)}{d\ell} = p_{H_a}(\ell) =$$
$$(2\pi)^{-1/2}\exp\bigg(-\frac{a^2}{2\ell}\bigg)a\,\ell^{-3/2}\,\mathbf{1}(\ell > 0), \qquad a > 0$$

Moreover

$$P(B_\ell^{\ge} a, B_\ell \in dx) = \frac{1}{\sqrt{\ell}}\gamma\bigg(\frac{a + |x - a|}{\sqrt{\ell}}\bigg)dx \tag{21}$$

Proof We define a Brownian motion reflected after H_a

$$\widetilde{B}_t = \begin{cases} B_t & , t \le H_a \\ 2a - B_t & t > H_a \end{cases}$$

with representation

$$\widetilde{B}_t = \int_0^t \Big(\mathbf{1}(s \le H_a) - \mathbf{1}(s > H_a) \Big) dB_s$$

where the integrand is bounded anb adapted since H_a is a $(\mathcal{F}_t^B)$-stopping time Since

$$\langle \widetilde{B} \rangle_t = \int_0^t \Big(\mathbf{1}(s \le H_a) - \mathbf{1}(s > H_a) \Big)^2 ds = t$$

by Lévy characterization it follows that $\widetilde{B}_t$ is a Brownian motion.

By drawing a figure we see that

$$\{B_\ell^* \ge a\} = \{B_\ell \ge a\} \cup \{B_\ell \ge a\}$$

where $\{B_\ell \ge a\} \cap \{B_\ell \ge a\} = \emptyset$

$$P\big(B_\ell^* \ge a\big) = P\big(\{B_\ell \ge a\} \cup \{B_\ell \ge a\}\big)$$
$$= P\big(B_\ell \ge a\big) + P\big(B_\ell \ge a\big) =$$
$$2P(B_\ell \ge a\big) = 2\big(1 - \Phi(a/\sqrt{\ell})\big) = 2\Phi(-a/\sqrt{\ell})$$

where $\Phi(x)$ is the cumulative distribution function of a standard gaussian r.v.

By the same argument

$$P(B_\ell^* \ge a, B_\ell \in dx) = P(B_\ell^* \ge a, \widetilde{B}_\ell \in dx) = P(B_\ell^* \ge a, 2a - B_\ell \in dx)$$

now there are two case either $x \ge a$ or $x < a$. When $x \ge a$

$$\frac{P(B_\ell^* \ge a, B_\ell \in dx)}{dx}(x) = \frac{P(B_\ell \in dx)}{dx}(x)$$

otherwise $2a - x > a$. and

$$\frac{P(B_\ell^* \ge a, B_\ell \in dx)}{dx}(x) = \frac{P(B_\ell \in dx)}{dx}(2a - x)$$

In both cases this gives formula (21).

17 Barrier option in Black and Scholes model

Consider the Black and Scholes model for a risky asset and a riskless bond.

$$S_t = S_0 \exp\Big(\int_0^t \sigma_s dB_s + \int_0^t (\mu_t - \frac{\sigma_t^2}{2}) dt \Big),$$
$$U_t = U_0 \exp\Big(\int_0^t \rho_s ds \Big)$$
$$S_0 > 0, \; U_0 > 0$$

$dS_t = S_t(\mu_t dt + \sigma_t dB_t), \quad dU_t = U_t \rho_t dt$

here μ_t, σ_t, U_t are adapted to the Brownian filtration $\mathcal{F}_t^B$.

Denote the discounted process

$$\widetilde{S}_t = \frac{S_t}{U_t} = \widetilde{S}_0 \exp\Big(\int_0^t \sigma_s dB_s + \int_0^t (\mu_t - \rho_t - \frac{\sigma_t^2}{2}) dt \Big)$$

satisfying

$$d\widetilde{S}_t = \widetilde{S}_t \big(\sigma_t dB_t + (\mu_t - \rho_t) dt \big)$$

Denote

$$\widetilde{B}_t := B_t + \int_0^t \frac{(\mu_s - \rho_s)}{\sigma_s} ds = \int_0^t (\widetilde{S}_s \sigma_s)^{-1} d\widetilde{S}_u$$

We want to represent the discounted value of the option $\widetilde{F}(\omega) := F(\omega)(S_T(\omega))^{-1}$ as a stochastic integral with respect to the discounted stock $\widetilde{S}_t$, which is also a stochastic integral with respect $\widetilde{B}_t$. However $\widetilde{B}_t$ is not Brownian motion under the measure P since it has a drift.

In order to use the Ito representation theorem we must first change the measure in order to kill the drift of $\widetilde{B}_t$, which becomes a Brownian motion under the new measure Q.

$$E_P(f(B_T)\mathbf{1}(B_T^* > a)) = \int_{\mathbb{R}} f(x) \frac{1}{\sqrt{T}} \gamma\Big(\frac{a + |x-a|}{\sqrt{T}} \Big) dx$$
$$E_P(f(B_T)\mathbf{1}(B_T^* > a)|\mathcal{F}_t) = E_P(f(B_T)\mathbf{1}(B_T^* > a)|B_t, B_t^*)$$
$$= \mathbf{1}(B_t^* > a) E_P(f(x + \sqrt{T-t}G))\Big|_{x=B_t} + \mathbf{1}(B_t^* \le a) E_P\big(f(x + W_{T-t})\mathbf{1}(W_{T-t}^* \ge > (a-x))\big)\Big|_{x=B_t}$$
$$\mathbf{1}(B_t^* > a) \int_{\mathbb{R}} f(x) \frac{1}{\sqrt{T-t}} \gamma\Big(\frac{x - B_t}{\sqrt{T-t}} \Big) dx + \mathbf{1}(B_t^* \le a) \int_{\mathbb{R}} f(x) \frac{1}{\sqrt{T-t}} \gamma\Big(\frac{a - B_t + |x-a|}{\sqrt{T-t}} \Big) dx$$

By using Ito formula and stochastic Fubini theorem

$$E_P(f(B_T)\mathbf{1}(B_T^* > a)|\mathcal{F}_t) = E_P(f(B_T)\mathbf{1}(B_T^* > a))$$
$$+ \int_0^t \mathbf{1}(B_s^* > a) \Big(\int_{\mathbb{R}} f(x) \frac{1}{\sqrt{T-s}} \gamma\Big(\frac{x - B_s}{\sqrt{T-s}} \Big) \frac{x - B_s}{T-s} dx \Big) dB_s + \int_0^t \mathbf{1}(B_s^* \le a) \Big(\int_{\mathbb{R}} f(x) \frac{1}{\sqrt{T-s}} \gamma\Big(\frac{a - B_s}{\sqrt{}}$$
$$= E_P(f(B_T)\mathbf{1}(B_T^* > a)) + \int_0^t \mathbf{1}(B_s^* > a) \frac{E_P\big(f(B_T)(B_T - B_s)\big|\mathcal{F}_s\big)}{(T-s)} dB_s$$
$$+ \int_0^t \mathbf{1}(B_s^* \le a) \frac{E_P\big(f(B_T)(a - B_s + |B_T - a|)|\mathcal{F}_s\big)}{T-s} dB_s$$

We can also write the joint law of B_t^*, B_t.

$$
\begin{aligned}
P\Big(B_t^* > y, B_t \le x\Big) &= P\Big(H_y \le t, (B_t - B_{H_y}) \le (x-y)\Big) \\
&= \int_0^t \Phi\bigg(\frac{x-y}{\sqrt{t-\ell}}\bigg) P(H_y \in d\ell) \\
&= (2\pi)^{-1/2} \int_0^t \Phi\bigg(\frac{x-y}{\sqrt{t-\ell}}\bigg) \exp\bigg(-\frac{y^2}{2\ell}\bigg) y\, \ell^{-3/2}\, d\ell = \\
&\int_0^t \Phi\bigg(\frac{x-y}{\sqrt{t-\ell}}\bigg) \frac{1}{\sqrt{\ell}} \gamma\bigg(\frac{y}{\sqrt{\ell}}\bigg) \frac{y}{\ell} d\ell
\end{aligned}
$$

and the joint density is given by

$$
\begin{aligned}
\frac{P(B_t^* \in dy, B_t \in dx)}{dxdy} &= -\frac{\partial^2}{\partial x \partial y} P\Big(B_t^* > y, B_t \le x\Big) \\
= \int_0^t & \frac{1}{\sqrt{t-\ell}} \gamma\bigg(\frac{x-y}{\sqrt{t-\ell}}\bigg) \frac{1}{\sqrt{\ell}} \gamma\bigg(\frac{y}{\sqrt{\ell}}\bigg) \frac{1}{\ell} \bigg(\frac{y^2}{\ell} - 1 - \frac{y(x-y)}{(t-\ell}\bigg) d\ell
\end{aligned}
\tag{23}
$$

By differentiating w.r.t. a we obtain the density of B_ℓ^*:

$$
\begin{aligned}
&\frac{P(B_\ell^* \in da)}{da} = p_{B_\ell^*}(a) = \\
&\frac{2}{\sqrt{2\pi\ell}} \exp\bigg(-\frac{a^2}{2\ell}\bigg) \mathbf{1}(a \ge 0) = \frac{2}{\sqrt{\ell}} \gamma\bigg(\frac{a}{\sqrt{\ell}}\bigg) \mathbf{1}(a \ge 0)
\end{aligned}
$$

We now compute the regular conditional density given the σ-algebra $\mathcal{F}_t^B$, $t \ge 0$.

For any bounded measurable function g

$$
\begin{aligned}
&E_P(g(H_a)|\mathcal{F}_t^B) = g(H_a)\mathbf{1}(H_a \le t) + E_P(g(H_a)|B_t, H_a > t)\mathbf{1}(H_a > t) = \\
&g(H_a)\mathbf{1}(H_a \le t) + E_P(g(t + H_{a-x}))\bigg|_{x=B_t} \mathbf{1}(H_a > t)
\end{aligned}
$$

where have derived the Markov property of Brownian motion, and there is a regular version of the conditional probability which up to the stopping time H_a has density

$$
M(\ell, t) := \frac{P(H_a \in d\ell | B_t, H_a > t)}{d\ell} = (2\pi)^{-1/2} \exp\bigg(-\frac{(B_t - a)^2}{2(\ell - t)}\bigg) \frac{(a - B_t)}{(\ell - t)^{3/2}}\, \mathbf{1}(\ell > t)
$$

Note that since the process

$$
E_P(g(H_a)|\mathcal{F}_{t \wedge H_a}) = \int_0^\infty M(\ell, t \wedge H_a) g(\ell) d\ell
$$

is a martingale for every bounded measurable g, $M(\ell, t \wedge H_a)$ is a martingale for all values $\ell > 0$. We use Ito formula to find the martingale representation with respect to the Brownian motion:

$$dM(\ell,t)=(2\pi)^{-1/2}M(\ell,t)\Big\{(B_t-a)^{-1}dB_t+\frac{3}{2}(\ell-t)^{-1}dt-\frac{(B_t-a)}{(\ell-t)}dB_t-\frac{1}{2(\ell-t)}dt$$
$$-\frac{(B_t-a)^2}{2(\ell-t)^2}dt+\frac{1}{2}\frac{(B_t-a)^2}{(\ell-t)^2}dt-\frac{(B_t-a)}{(\ell-t)(B_t-a)}dt\Big\}=$$
$$M(\ell,t)\Big\{\frac{1}{(B_t-a)}+\frac{(a-B_t)}{\ell-t}\Big\}dB_t=M(\ell,t)F(\ell-t,a-B_t)dB_t$$

We have the stochastic exponential representation

$$M(\ell,t\wedge H_a)=M(\ell,0)\mathcal{E}\bigg(\int_0^{\cdot}\Big\{\frac{1}{(B_s-a)}+\frac{(a-B_s)}{\ell-s}\Big\}dB_s\bigg)_{t\wedge H_a}=$$
$$M(\ell,0)\exp\bigg(\int_0^{t\wedge H_a}\Big\{\frac{1}{(B_s-a)}+\frac{(a-B_s)}{\ell-s}dB_s-\frac{1}{2}\int_0^{t\wedge H_a}\Big\{\frac{1}{(B_s-a)}+\frac{(a-B_s)}{\ell-s}\Big\}^2\bigg)$$

Note that the process (B_t^*,B_t) is Markovian:

$$E_P(f(B_\ell^*)|\mathcal{F}_s)=\mathbf{1}(\ell\le s)f(B_\ell^*)+\mathbf{1}(\ell>s)E_P\big(f(\max\{x,y+W_{\ell-s}^*\sqrt{\ell-s}\})\big)\Big|_{x=B_s^*,y=B_s}$$
$$=\mathbf{1}(\ell\le s)f(B_\ell^*)+\mathbf{1}(\ell>s)\int_0^\infty f(\max\{B_s^*(\omega),B_s(\omega)+v)\frac{2}{\sqrt{\ell-s}}\gamma\Big(\frac{v}{\sqrt{\ell-s}}\Big)dv$$
$$=\mathbf{1}(\ell\le s)f(B_\ell^*)+\mathbf{1}(\ell>s)\Big\{f(B_s^*)\big(2\Phi\Big(\frac{B_s^*-B_s}{\sqrt{\ell-s}}\Big)-1\big)+\int_{B_s^*}^\infty f(v)\frac{2}{\sqrt{\ell-s}}\gamma\Big(\frac{v-B_s}{\sqrt{\ell-s}}\Big)dv\Big\}$$

Assume absolute continuity $f(x)=f(0)+\int_0^x f'(y)dy$.
For $s<\ell$ we use integration by parts obtaining

$$E_P\big(f'(B_T^*)\mathbf{1}(B_T^*>B_s^*)\big|\mathcal{F}_s\big)=\int_{B_s^*}f'(v)\frac{2}{\sqrt{\ell-s}}\gamma\Big(\frac{v-B_s}{\sqrt{\ell-s}}\Big)dv=$$
$$-f(B_s^*)\frac{2}{\sqrt{\ell-s}}\gamma\Big(\frac{B_s^*-B_s}{\sqrt{\ell-s}}\Big)+\int_{B_s^*}^\infty f(x)\frac{2}{\sqrt{\ell-s}}\gamma\Big(\frac{v-B_s}{\sqrt{\ell-s}}\Big)\Big(\frac{v-B_s}{\ell-s}\Big)dv=$$
$$-f(B_s^*)\frac{2}{\sqrt{\ell-s}}\gamma\Big(\frac{B_s^*-B_s}{\sqrt{\ell-s}}\Big)+E_P\bigg(f(B_T^*)\frac{(B_T^*-B_s)}{\ell-s}\mathbf{1}(B_T^*>B_s^*)\bigg|\mathcal{F}_s\bigg)$$

Therefore Ito representation gives

$$E_P\big(f(B_\ell^*)\big|\mathcal{F}_s\big)=$$
$$E_P(f(B_T^*))+\int_0^\ell\bigg\{E_P\bigg(f(B_\ell^*)\frac{(B_\ell^*-B_s)}{\ell-s}\mathbf{1}(B_\ell^*>B_s^*)\bigg|\mathcal{F}_s\bigg)$$
$$-f(B_s^*)\frac{P(W_{\ell-s}^*\in dv|W_0=B_s)}{dv}(B_s^*-B_s)\bigg\}dB_s$$
$$=E_P(f(B_\ell^*))+\int_0^T E_P\big(f'(B_\ell^*)\mathbf{1}(B_\ell^*>B_s^*)\big|\mathcal{F}_s\big)dB_s$$

where (W_t) is an independent Brownian motion. The last expression holds only when $f(x)$ is absolutely continuous.

Suppose now we want to compute the representation of $f(B_T(\omega),B_T^*(\omega))\in L^2(P)$ We need to compute the joint conditional laws $P(B_T\in dx,B_T^*\in dy|\mathcal{F}_t)=P(B_T\in dx,B_T^*\in dy|B_t,B_t^*)$.

18 Stochastic differential equation

Given a Brownian motion (B_t) we look for a stochatic process $(X_t : t \in [s,T])$ such that

$$X_t = \eta + \int_s^t b(u, X_u)du + \int_s^t \sigma(u, X_u)dB_u \quad 0 \le s \le t \tag{24}$$

with $\eta(\omega)$ $\mathcal{F}_s^B$-measurable. Of such process exists and it is adapted to the $(\mathcal{F}_t^B)$ we say that it is a *strong solution* of the stochastic differential equation (25) In differential notation we write

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t, \quad t \ge s \tag{25}$$

with initial condition $X_s(\omega) = \eta(\omega)$.

18.1 Generator of a diffusion

Lemma 18. *Assume that the SDE 25 has a strong solution and that* $\varphi(t,x) \in C^{1,2}(\mathbb{R}^+ \times R^m; \mathbb{R})$*. Then*

$$d\varphi(t, X_t) = \frac{\partial \varphi(t, X_t)}{\partial x} dX_t + \frac{1}{2}\frac{\partial^2 \varphi(t, X_t)}{\partial x^2} d\langle X \rangle_t + \frac{\partial \varphi(t, X_t)}{\partial t} dt =$$
$$\frac{\partial \varphi(t, X_t)}{\partial x}\sigma(t, X_t)dB_t + \left\{ \frac{\partial \varphi(t, X_t)}{\partial x} b(t, X_t) + \frac{1}{2}\frac{\partial^2 \varphi(t, X_t)}{\partial x^2}\sigma(t, X_t)^2 + \frac{\partial \varphi(t, X_t)}{\partial t} \right\} dt$$

Define the space-time generator operator

$$(L_t\phi)(t,x) = b(t,x)\frac{\partial \varphi(t,x)}{\partial x}\frac{1}{2}\sigma(t,x)^2\frac{\partial^2 \varphi(t,x)}{\partial x^2} + \frac{\partial \varphi(t,x)}{\partial t}$$

It follows that

$$M_t(\varphi) := \varphi(t, X_t) - \varphi(0, X_0) - \int_0^t (L_s\varphi)(s, X_s)ds = \int_0^t \frac{\partial \varphi(s, X_s)}{\partial x}\sigma(s, X_s)dB_s$$

is a continuous local martingale with $M_0(\varphi) = 0$*, such that for any local martingale* (N_t)

$$\langle M(\varphi), N \rangle_t = \int_0^t \frac{\partial \varphi(s, X_s)}{\partial x}\sigma(s, X_s)d\langle B, N \rangle_s$$

In particular for another $\psi(t,x) \in C^{2,1}$

$$\langle M(\varphi), M(\psi) \rangle_t = \int_0^t \frac{\partial \varphi(s, X_s)}{\partial x}\frac{\partial \psi(s, X_s)}{\partial x}\sigma(s, X_s)^2 ds$$

Exercise 16. *Using the definition show that*

$$\langle M(\varphi), M(\psi) \rangle_t = \int_0^t (L_s(\varphi\psi) - \varphi L_s\psi - \psi L_s\varphi)(s, X_s)ds$$

Hint: By polarization it is enough to consider the case $\psi(t,x) = \varphi(t,x)$ *For simplicity you can consider the time-homogeneous case with* $\sigma(t,x) = \sigma(x)$ $b(t,x) = b(x)$ *and* $\varphi(t,x) = \varphi(x)$*.*

Note that by construction for $H(s,\omega)$ progressively measurable the Ito integral $X_t = (H \cdot B)_t = \int_0^t H_s dB_s$ is the continuous local martingale (unique up to indistinguishability) such that

$$\langle (H \cdot B), M \rangle_t = \int_0^t H_s d\langle B, M \rangle_s$$

for any local martingale (M_t). This implies that for another progressively measurable $K(s,\omega)$

$$Y_t := (K \cdot X)_t = \int_0^t K_s dX_s = \int_0^t K_s H_s dB_s = ((KH) \cdot B)_t$$

since for any local martingale (M_t)

$$\langle Y, M \rangle_t = \int_0^t K_s d\langle X, M \rangle_s =$$
$$\int_0^t K_s H_s d\langle B, M \rangle = \big\langle ((KH) \cdot B), M \big\rangle_t$$

since this associative property holds for Lebesgue Stieltjes integrals.

18.2 Stratonovich integral

Let M_t be a continuous local martingale and X_t a semimartingale. We define the *Stratonovich integral* as

$$\int_0^t X_s \circ dM_s = \int_0^t X_s dM_s + \frac{1}{2}[X, M]_t$$

The idea is that the Ito integral corresponds with the forward integral which is the limit in probability of the approximating Riemann sums

$$\int_0^t X_s d^- M_s = (P) \lim_{\Delta(\Pi) \to 0} \sum_{t_i \in \Pi} X_{t_i} (M_{t_{i+1} \wedge t} - M_{t_i \wedge t})$$

This corresponds adapted piecewise constant approximating integrands

$$X_s^- = X_{t_i} \quad \text{when } s \in (t_i, t_{i+1}]$$

The choice

$$X_s^+ = X_{t_{i+1}} \quad \text{when } s \in (t_i, t_{i+1}]$$

does not give necessarily an adapted integrand. Nevertheless it is clear that since

$$X_{t_{i+1}} (M_{t_{i+1} \wedge t} - M_{t_i \wedge t}) = X_{t_i} (M_{t_{i+1} \wedge t} - M_{t_i \wedge t}) + (X_{t_{i+1}} - X_{t_i})(M_{t_{i+1} \wedge t} - M_{t_i \wedge t}) =$$

necessarily the backward integral

$$\int_0^t X_s d^+ M_s = (P) \lim_{\Delta(\Pi) \to 0} \sum_{t_i \in \Pi} X_{t_{i+1}} (M_{t_{i+1} \wedge t} - M_{t_i \wedge t}) = \int_0^t X_s d^- M_s + [X, M]_t$$

is also well defined.

The Stratonovich integral is approximated by picking the middle point

$$X_s^\circ = X_{(t_i+t_{i+1})/2} \quad \text{when } s \in (t_i, t_{i+1}]$$

We have

$$\begin{aligned}
&\sum_{t_i \in \Pi} X_{(t_i+t_{i+1})/2}(M_{t_{i+1}\wedge t} - M_{t_i \wedge t}) = \\
&\sum_{t_i \in \Pi} X_{t_i}(M_{t_{i+1}\wedge t} - M_{t_i\wedge t}) + \sum_{t_i\in\Pi}(X_{(t_i+t_{i+1})/2} - X_{t_i})(M_{(t_i+t_{i+1})/2\wedge t} - M_{t_i\wedge t}) \\
&+ \sum_{t_i\in\Pi}(X_{(t_i+t_{i+1})/2} - X_{t_i})(M_{t_{i+1}\wedge t} - M_{(t_i+t_{i+1})/2\wedge t}) \\
&\xrightarrow{P} \int_0^t X_s d^- M_s + \frac{1}{2}[M, X]_t + 0
\end{aligned}$$

as $\Delta(\Pi) \to 0$

Therefore

$$\int_0^t X_s \circ dM_s = \frac{1}{2}\Big(\int_0^t X_s d^- M_s + \int_0^t X_s d^+ M_s\Big)$$

the Stratonovich integral is the average of forward integral and a backward integral.

Note the Stratonovich integral obeys the law of standard calculus. Assuming for simplicity that $f \in C^3$, By Ito formula,

$$f(M_t) = f(M_0) + \int_0^t f'(M_s)d^- M_s + \frac{1}{2}f''(M_s)d\langle M\rangle_s = f(M_0) + \int_0^t f'(M_s)\circ dM_s$$

since

$$\langle f'(M), M\rangle_t = \Big\langle \int_0^{\cdot} f''(M_s)dM_s, M\Big\rangle_t = \int_0^t f''(M_s)d\langle M, M\rangle_s$$

18.3 Doss-Sussman explicit solutions

In the one-dimenstional case, sometimes we are able to proceed as follows:

Consider the SDE in Stratonovich sense

$$\begin{aligned}
dX_t &= b(X_t)dt + \sigma(X_t)\circ dW_t \\
&= b(X_t)dt + \sigma(X_t)dW_t + \frac{1}{2}d\langle\sigma(X), B\rangle_t = \big(b(X_t) + \frac{1}{2}\sigma'(X_t)\sigma(X_t)\big) + \sigma(X_t)dW_t
\end{aligned}$$

where in the first line the stochastic integral is in Stratonovich sense and on the second line in Ito sense. Here $\sigma'(x) = \frac{d}{dx}\sigma(x)$

We look for a solution of the form $X_t = u(W_t, Y_t)$ for some smooth function $u(x, y)$ and a continous process of finite variation Y_t.

Taking Stratonovich differential we get

$$dX_t = \frac{\partial}{\partial x}u(W_t, Y_t)\circ dW_t + \frac{\partial}{\partial y}u(W_t, Y_t)dY_t$$

which means that

$$\frac{\partial}{\partial x}u(x,y) = \sigma(u(x,y))$$
$$dY_t = \left(\frac{\partial}{\partial y}u(W_t,Y_t)\right)^{-1} b(u(W_t,Y_t))dt$$

We get also

$$\frac{\partial^2}{\partial x^2}u(x,y) = \sigma'(u(x,y))\sigma(u(x,y)), \quad \frac{\partial^2}{\partial x\partial y}u(x,y) = \sigma'(u(x,y))\frac{\partial}{\partial y}u(x,y),$$

We impose the additional condition $u(0,y) = y$, from which follows

$$\frac{\partial}{\partial y}u(0,y) = 1,$$
$$\frac{\partial}{\partial y}u(x,y) = 1 + \int_0^x \frac{\partial^2}{\partial x\partial y}u(\xi,y)d\xi = 1 + \int_0^x \frac{\partial}{\partial y}u(\xi,y)\sigma'(u(\xi,y))d\xi =$$
$$= \exp\left(\int_0^x \sigma'(u(\xi,y))d\xi\right)$$

Substituting

$$Y_t = Y_0 + \int_0^t \exp\left(-\int_0^{W_s} \sigma'(u(\xi,Y_s))d\xi\right) b(u(W_s,Y_s))ds$$

By solving these ODE we obtain the solution $X_t = u(W_t,Y_t)$.

Example Consider the SDE

$$dX_t = \cos(X_t)dt + X_t \circ dW_t = (\cos(X_t) + \frac{1}{2}X_t)dt + X_t dW_t$$

written respectively with Stratonovich and Ito differentials
the ODE

$$\frac{\partial}{\partial x}u(x,y) = u(x,y), \quad u(0,y) = y$$

has solution

$$u(x,y) = y\exp(x)$$

and

$$Y_t = Y_0 + \int_0^t \exp(-W_s)\cos(Y_s\exp(W_s))ds$$

The solution is $X_t = Y_t\exp(W_t)$. In fact by using integration by parts,

$$\circ dX_t = \exp(W_t)dY_t + Y_t \circ d\exp(W_t)$$
$$\exp(W_t)\exp(-W_t)\cos(Y_t\exp(W_t))dt + Y_t\exp(W_t) \circ dW_t = \cos(X_t)dt + X_t \circ dW_t$$

19 Cameron-Martin-Girsanov theorem

We denote by P_t the restriction of P on the σ-algebra $\mathcal{F}_t$.

Let (M_t) a $\{\mathcal{F}_t\}$-local martingale under the measure P and(H_t) an $\{\mathcal{F}_t\}$-adapted process such that for all $0 \le t < +\infty$

$$\int_0^t H_s^2 d\langle M\rangle_s < \infty \quad P \text{ almost surely}$$

We want to find a probability measure Q such that

$$\widetilde{M}_t = M_t + \int_0^t H_s d\langle M\rangle_s, \tag{26}$$

is a local martingale with respect to the measure Q and $Q_t \ll P_t \quad \forall t < \infty$. (notation $Q \overset{loc}{\ll} P$)

Lemma 19. *Assume that $Q \overset{loc}{\ll} P$. The likelihood ratio process*

$$Z_t(\omega) := \frac{dQ_t}{dP_t}(\omega) \tag{27}$$

is a true martingale with respect to the reference measure P.

Proof For $s < t$, if $A \in \mathcal{F}_s \subseteq \mathcal{F}_t$,

$$Q(A) = E_P(Z_t \mathbf{1}_A) = E_P(Z_s \mathbf{1}_A)$$

which gives the martingale property under P.

Note We recall also that a non-negative local martingale Z_t is a supermartingale, since if $\tau_n \uparrow \infty$ is a localizing sequence, for $s \le t$ by the Fatou lemma for conditional expectation

$$E_P(Z_t|\mathcal{F}_s) = E_P\big(\liminf_{n\uparrow\infty} Z_{t\wedge\tau_n}\big|\mathcal{F}_s\big) \le \liminf_{n\uparrow\infty} E_P(Z_{t\wedge\tau_n}|\mathcal{F}_s)$$
$$\le \liminf_{n\uparrow\infty} Z_{s\wedge\tau_n} = Z_s$$

Moreover Z_t is a true martingale if and only if $E_P(Z_t) = 1$, since in such case

$$Z_s - E_P(Z_t|\mathcal{F}_s) \ge 0 \quad \text{ and } E_P(Z_s) = E_P(Z_t) = 1$$

implies $Z_s = E_P(Z_t|\mathcal{F}_s)$ P-almost surely.

Lemma 20. *Let P probability measures on $(\Omega, \mathcal{F})$ equipped with the filtration $\{\mathcal{F}_t\}$ Then X_t is a Q (local)-martingale if and only if the product process $(X_t Z_t)$ is a P (local)-martingale.*

Proof for $s \le t$ $A \in \mathcal{F}_s$ we have

$$E_Q(X_t \mathbf{1}_A) = E_P(Z_t X_t \mathbf{1}_A)$$
$$E_Q(X_s \mathbf{1}_A) = E_P(Z_s X_s \mathbf{1}_A)$$

therefore the right hand sides coincide if and only if the left hand sides do.

Moreover if $\tau_n \uparrow \infty$ is a localizing sequence,

$$E_Q(X_{t\wedge\tau_n}\mathbf{1}_A) = E_P(Z_t X_{t\wedge\tau_n}\mathbf{1}_A) = E_P(Z_{t\wedge\tau_n} X_{t\wedge\tau_n}\mathbf{1}_A)$$
$$E_Q(X_{t\wedge\tau_n}\mathbf{1}_A) = E_P(Z_t X_{t\wedge\tau_n}\mathbf{1}_A) = E_P(Z_{t\wedge\tau_n} X_{t\wedge\tau_n}\mathbf{1}_A)$$

since the stopping time $(t\wedge\tau_n) \le t$ is bounded and by Doob optional sampling theorem

$$E_P(Z_t|\mathcal{F}_{t\wedge\tau_n}) = Z_{t\wedge\tau_n} \quad \square$$

Theorem 19. *(Cameron-Martin-Girsanov) Let Q a probability meausure such that $Q \overset{loc}{\ll} P$ and the change of drift formula (26) holds.*

Necessarily

$$Z_t = Y_t \exp\biggl(\int_0^t H_s dM_s - \frac{1}{2}\int_0^t H_s^2 d\langle M\rangle_s\biggr)$$

where $Y_t \ge 0$ is a continuous P-martingale with $E_P(Y_0) = 1$ and $[M,Y]_t = 0$ $\forall t$, and the change of drift formula (26) reads as

$$\widetilde{M}_t = M_t - \int_0^t \frac{1}{Z_s} d\langle M, Z\rangle_s$$

In particular when $Y_t \equiv 1$ $\forall t$, the change of measure is minimal, in the sense that every P-(local) martingale X_t such that $[X,M]_t \equiv 0$ is also a Q-(local) martingale.

Proof By the assumption and lemma 20, the product $(Z_t\widetilde{M}_t)$ is a local martingale under P. Using integration by parts, we obtain the martingale decomposition under Q

$$d(Z_t\widetilde{M}_t) = Z_t dM_t + Z_t H_t d\langle M\rangle_t + M_t dZ_t + d\langle \widetilde{M}, Z\rangle_t =$$
$$\bigl(Z_t dM_t + M_t dZ_t\bigr) + \bigl(Z_t H_t d\langle M\rangle_t + d\langle M, Z\rangle_t\bigr)$$

which implies

$$\langle M, Z\rangle_t = -\int_0^t Z_s H_s d\langle M\rangle_s$$

This is satisfied if and only if

$$\frac{1}{Z_t} dZ_t = -H_t dM_t + dX_t$$

where X_t is a P-martingale with $\langle M, X\rangle = 0$.

Let's assume first that $X_t = 0$. Then by Ito formula the solution of the linear stochatic differential equation $dZ_t = -Z_t H_t dM_t$ is the exponential martingale

$$Z_t = Z_0\mathcal{E}(H\cdot M)_t = Z_0\mathcal{E}\biggl(-\int_0^{\cdot} H_s dM_s\biggr)_t :=$$
$$Z_0 \exp\biggl(-\int_0^t H_s dM_s - \int_0^t H_s^2 d\langle M\rangle_s\biggr)$$

Here $Z_0(\omega) = \frac{dQ_0}{dP_0}(\omega)$ is $\mathcal{F}_0$-measurable.

More in general

$$Z_t = Z_0\mathcal{E}(H\cdot M + X)_t = Z_0\mathcal{E}(H\cdot M)_t\mathcal{E}(X)_t \quad \square$$

20 Stochastic filtering

Lemma 21. *Let M_t be a continuous local martingale under P with respect to a filtration $(\mathcal{G}_t)_{t\geq 0}$, and assume that (M_t) is adapted to a smaller filtration $(\mathcal{F}_t)_{t\geq 0}$, with $\mathcal{F}_t \subseteq \mathcal{G}$.*

Then M_t is also a $(\mathcal{F}_t)$-local martingale.

Proof

Let $\tau_n = \inf\{t : |M_t| \geq n\}$. Since M_t is $(\mathcal{F}_t)$-adapted, τ_n are stopping times in the $(\mathcal{F}_t)$- filtration, with $\tau_n \uparrow \infty$, and we know that for each n, the stopped process $M_t^{\tau_n} = M_{t\wedge\tau_n}$ is a true $(\mathcal{G}_t)$-martingale since it is bounded, which means that in particular for $0 \leq s \leq t$ $\forall A \in \mathcal{G}_s$

$$E_P\big((M_{t\wedge\tau_n} - M_{s\wedge\tau_n})\mathbf{1}_A\big) = 0$$

But this holds in particular $\forall A \in \mathcal{F}_s$, which means that $(M_t^{\tau_n})_{t\geq 0}$ is a true $(\mathcal{F}_t)$-martingale.

Note Without the continuity assumption we are not able to to produce a localizing sequence of $(\mathcal{F}_t)$-stopping times, just knowing that there is a localizing sequence of $(\mathcal{G}_t)$-stopping times.

Lemma 22. *Let (B_t) be aBrownian motion with the martingale property in the filtration $(\mathcal{G}_t)$ and obviously also with respect to the smaller filtration $(\mathcal{F}_t^B) \subseteq (\mathcal{G}_t)$ generated by itself.*

Let $H(s,\omega)$ a $(\mathcal{G}_t)$-adapted process which is not necessarily $(\mathcal{F}_t^B)$-adapted, such that

$$\int_0^t E_P(H_s^2)ds < \infty$$

Then

$$E_P\left(\int_0^t H_s dB_s \middle| \mathcal{F}_t^B\right) = \int_0^T E_P(H_s|\mathcal{F}_s^B)dB_s$$

Moreover if M_t is a $(\mathcal{G}_t)$-martingale with $\langle M, B\rangle_s = 0$, $\forall 0 \leq s \leq t$ then

$$E_P\big(M_t - M_0 \big| \mathcal{F}_t^B\big) = 0$$

Proof Let $A \in \mathcal{F}_t^B$. By the Ito-Clarck representation theorem

$$\mathbf{1}_A = P(A) + \int_0^t K_s dB_s$$

for some $K \in L^2([0,t] \times \Omega)$ adapted to $(\mathcal{F}_t^B)$.

$$E_P\Big(\mathbf{1}_A \int_0^t H_s dB_s\Big) = P(A) E_P\Big(\int_0^t H_s dB_s\Big) + E_P\Big(\int_0^t K_s dB_s \int_0^t H_s dB_s\Big)$$
$$= 0 + E_P\Big(\langle K \cdot B, H \cdot B\rangle_t\Big) = E_P\Big(\int_0^t K_s H_s ds\Big) =$$
$$\int_0^t E_P(K_s H_s) ds = \int_0^t E_P\big(K_s E_P(H_s|\mathcal{F}_s)\big) ds$$
$$= E_P\Big(\Big\langle \int_0^\cdot K_s dB_s, \int_0^\cdot E_P(H_s|\mathcal{F}_s) dB_s \Big\rangle_t\Big)$$
$$= 0 + E_P\Big(\int_0^t K_s dB_s \int_0^t E_P(H_s|\mathcal{F}_s) dB_s\Big) = E_P\Big(\mathbf{1}_A \int_0^t E_P(H_s|\mathcal{F}_s) dB_s\Big) =$$

where we used the Ito isometry and the definition of conditional expectation $\Box$

For the second part of the lemma, if $M_0 = 0,\ \langle M, B\rangle_s = 0,\ s \le t,\ A \in \mathcal{F}_t^B$ as before,

$$E_P((M_t - M_0)\mathbf{1}_A) = P(A) E_P(M_t - M_0) + E_P((M_t - M_0) \int_0^t K_s dB_s) =$$
$$0 + E_P\Big(\int_0^t K_s d\langle M, B\rangle_s\Big) = 0$$

which means $E_P(M_t - M_0|\mathcal{F}_t^B) = 0$ $\Box$

Consider the stochastic filtering settings in the St Flour lecture notes by E Pardoux :

$$dX_s = b(s, Y, X_s) ds + f(s, Y, X_s) dV_s + g(s, Y, X_s) dW_s$$
$$dY_s = h(s, Y, X_s) ds + dW_s$$

with (V, W) are independent P-Brownian motions and consider the filtration $\{\mathcal{F}_t\}$ with $\mathcal{F}_t = \mathcal{F}_t^{V,W,}$ and $\{\mathcal{Y}_t\}$ with $\mathcal{Y}_t = \mathcal{F}_t^Y$.

Here X_t is the state process, and the problem is to estimate "on-line" X_t using the information from the observation filtration $\{\mathcal{Y}_t\}$ which gives in noisy observations of the signals $h(s, Y, X_s)$.

For simplicity, it is assumed all all coefficient processes are bounded and Lipshitz.

We introduce a reference measure Q under which

$$dX_s = \{b(s, Y, X_s) - h(s, Y, X_s) g(s, Y, X_s)\} ds + f(s, Y, X_s) dV_s + g(s, Y, X_s) dY_s$$

and Y is a Brownian motion w.r.t Q in the $\{\mathcal{F}_t\}$ filtration. It follows that $P_t \ll Q_t$ with

$$Z_t := \frac{dP_t}{dQ_t} = \exp\Big(\int_0^t h(s, Y, X_s) dY_s - \frac{1}{2}\int_0^t h(s, Y, X_s)^2 ds\Big)$$

satisfying the linear SDE $dZ_t = Z_t h(t, Y, X_t) dY_t$.

For a function $\varphi \in C_B^2$, bounded and with bounded derivatives, by abstract Bayes formula

$$\pi_t(\varphi) := E_P(\varphi(X_t)|\mathcal{Y}_t) = \frac{E_Q(\varphi(X_t)Z_t|\mathcal{Y}_t)}{E_Q(Z_t|\mathcal{Y}_t)} = \frac{\sigma_t(\varphi)}{\sigma_t(1)}$$

Here π_t is the posterior probability measure process, and σ_t is the unnormalized posterior measure.

$\sigma_t(\varphi) = E_Q(\varphi(X_t)Z_t|\mathcal{Y}_t)$ satisfies the following SDE driven by the Q Brownian motion (Y_t) in the $(\mathcal{Y}_t)$ filtration:

$$\sigma_t(\varphi) = \sigma_0(\varphi) + \int_0^t \sigma_s(L_{s,Y}\varphi)ds + \int_0^t \sigma_s(L_{s,Y}^1\varphi)dY_s \qquad (28)$$

where $L_{s,Y}$ and $L_{s,Y}^1$ are differential operators on C^2 depending on time and on the past observations of Y:

$$L_{s,Y}\ \varphi = \frac{1}{2}(f^2(s,Y,\cdot) + g^2(s,Y,\cdot))\frac{\partial^2}{\partial^2 x}\varphi + b(s,Y,\cdot)\frac{\partial}{\partial x}\varphi$$
$$L_{s,Y}^1\ \varphi = h(s,Y,\cdot)\varphi + g(s,Y,\cdot)\frac{\partial}{\partial x}\varphi$$

To check this step, note that by the integration by parts formula

$$\begin{aligned}
&d(\varphi(X_t)Z_t) = Z_t d\varphi(X_t) + \varphi(X_t)dZ_t + d\langle\varphi(X_t), Z\rangle_t \\
&= Z_t\varphi'(X_t)dX_t + \frac{1}{2}Z_t\varphi''(X_t)d\langle X\rangle_t + Z_t\varphi(X_t)h(t,Y,X_t)dY_t + Z_t\varphi'(X_t)g(t,Y,X_t)h(t,Y,X_t)dt \\
&= Z_t\big\{\varphi'(X_t)g(t,Y,X_t) + \varphi(X_t)h(t,Y,X_t)\big\}dY_t + Z_t\varphi'(X_t)f(t,Y,X_t)dV_t + \\
&+Z_t\varphi'(X_t)\big\{b(t,Y,X_t) - h(t,Y,X_t)g(t,Y,X_t) + g(t,Y,X_t)h(t,Y,X_t)\big\}dt \\
&+\frac{1}{2}Z_t\varphi''(X_t)\big\{f(t,Y,X_t)^2 + g(t,Y,X_t)^2\big\}dt \\
&= Z_t\big\{\varphi'(X_t)g(t,Y,X_t) + \varphi(X_t)h(t,Y,X_t)\big\}dY_t + Z_t\varphi'(X_t)f(t,Y,X_t)dV_t \\
&+Z_t\big\{\varphi'(X_t)b(t,Y,X_t) + \frac{1}{2}Z_t\varphi''(X_t)\big(f(t,Y,X_t)^2 + g(t,Y,X_t)^2\big)\big\}dt
\end{aligned}$$

In integral form this means

$$\begin{aligned}
&\varphi(X_t)Z_t = \varphi(X_0) + \int_0^t Z_s\big\{\varphi'(X_s)g(s,Y,X_s) + \varphi(X_s)h(s,Y,X_s)\big\}dY_s + \\
&\int_0^t Z_s\varphi'(X_s)f(s,Y,X_s)dV_s \\
&+\int_0^t Z_s\big\{\varphi'(X_s)b(s,Y,X_s) + \frac{1}{2}\varphi''(X_s)\big(f(s,Y,X_t)^2 + g(s,Y,X_t)^2\big)\big\}ds
\end{aligned}$$

We take now conditional expectation under Q with respect to the σ-algebra $\mathcal{Y}_t$.

$$\sigma_t(\varphi) := E_Q(\varphi(X_t)Z_t|\mathcal{Y}_t) =$$
$$E_Q(\varphi(X_0)|\mathcal{Y}_t)$$
$$+E_Q\left(\int_0^t Z_s\{\varphi'(X_s)g(s,Y,X_s)+\varphi(X_s)h(s,Y,X_s)\}dY_s\middle|\mathcal{Y}_t\right)$$
$$+E_Q\left(\int_0^t Z_s\varphi'(X_s)f(s,Y,X_s)dV_s\middle|\mathcal{Y}_t\right)$$
$$+E_Q\left(\int_0^t Z_s\{\varphi'(X_s)b(s,Y,X_s)+\frac{1}{2}\varphi''(X_s)\big(f(s,Y,X_t)^2+g(s,Y,X_t)^2\big)\}ds\middle|\mathcal{Y}_t\right)$$

and 28 follows by lemma 22.

When $\varphi(x)\equiv 1$ we get a linear SDE for the random normalizing constant in Bayes formula:

$$\sigma_t(1) = 1+\int_0^t \sigma_s(1)E_P(h(s,Y,X_s)|\mathcal{Y}_s)dY_s$$

with solution

$$\sigma_t(1) = \exp\left(\int_0^t E_P(h(s,Y,X_s)|\mathcal{Y}_s)dY_s+\frac{1}{2}\int_0^t E_P(h(s,Y,X_s)|\mathcal{Y}_s)^2ds\right)$$

Consequently by the Cameron Martin Girsanov theorem (19)

$$Y_t-\int_0^t E_P(h(s,Y,X_s)|\mathcal{Y}_s)ds$$

is a P Brownian motion in the $\{\mathcal{Y}_t\}$ filtration.

21 Final exam :

These questions form the final exam, hopefully to be returned by the end of the summer. It is allowed to consult the literature and to collaborate with fellow students.

Question 1): Use the change of measure formula to show that

$$E_Q(Z_t|\mathcal{Y}_t) = \sigma_t(1) = \frac{dP|\mathcal{Y}_t}{dQ|\mathcal{Y}_t}$$

Question 2): Use integration by parts formula for the ratio $\pi_t(\varphi) = \sigma_t(\varphi)/\sigma_t(1)$ to prove the Zakai filter equation

$$\pi_t(\varphi) = \pi_0(\varphi)+\int_0^t \pi_s(L_{s,Y}\varphi)ds+\int_0^t\big\{\pi_s(L^1_{s,Y}\varphi)-\pi_s(h(s,Y,\cdot))\pi_s(\varphi)\big\}\big(dY_s-\pi_s(h(s,Y,\cdot))ds\big)$$

Question 3) Show that

$$Y_t-\int_0^t \pi_s(h(s,Y,\cdot))ds$$

is a Brownian motion with respect to the measure P and the filtration $(\mathcal{Y}_t)$.

Consider the linear gaussian case with

$$\begin{aligned} dX_s &= X_s b(s)ds + f(s)dV_s + g(s)dW_s \\ dY_s &= X_s h(s)ds + dW_s \end{aligned}$$

with $b(s), h(s), f(s), g(s)$ deterministic functions.

Question 4):Write down the Zakai filter equation for the prediction process

$$\hat{X}_t := E(X_t|\mathcal{Y}_t)$$

Question 5): Write down the equation for the prediction error variance

$$\hat{\sigma}_t^2 := E((X_t - \hat{X}_t)^2|\mathcal{Y}_t)$$

Since the process (X_t, Y_t) is jointly gaussian (why ? for example one can study the characterstic function) you should get a deterministic equation, called Riccati equation.

Since (X_t, Y_t) is jointly gaussian, it follows that conditionally on the σ-algebra $\mathcal{Y}_t$, X_t is conditionally gaussian with (random) conditional mean $\hat{X}_t$ and (deterministic) conditional variance $\hat{\sigma}_t^2$. You must use gaussianity in order to compute the conditional moments $\pi_t(x^k)$ for $k = 1, 2, 3$ which will appear in the Zakai equation.

For simplicity you can assume that the functions $b(s), h(s), f(s), g(s)$ are constant. If you want to simplify further, assume that $g(s) = 0$.

A standard reference on stochastic filtering theory is in Liptser and Shiryaev statistics of random processes.