Introduction to stochastic analysis

Dario Gasbarra

November 29, 2011

Contents

Chapter 1

Why stochastic integration is needed ?

1.1 Introduction

Let x_t and y_t measurable functions $\mathbb{R}^+ \mapsto \mathbb{R}$, where x_t has finite variation and y_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

$$v_t(x) := x_t^{\oplus} + x_t^{\ominus} = \sup_{\Pi} \sum_{t_i \in \Pi} |x_{t_{i+1}} - x_{t_i}|$$

where in the left side the supremum is taken over all finite partitions of $[0,t]$ $\Pi = (0 = t_0 < t_1 < \cdots < t_n = t)$ with $n \in \mathbb{N}$. 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

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

is well defined as a Lebesgue Stieltjes integral. When the integrand y_s 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 $\Pi = \{0 = t_0 \le s_0 \le t_1 \le s_1 \le t_2 \le \dots \le t_{n-1} \le s_n \le t_n = t\}$ 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, Louis 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. $B_0 = 0$, and 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 Albert 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 A.N. Kolmogorov (1933) showed that the paths $B_t(\omega)$ are almosty surely Hölder continuous that is the random quantity

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

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.

1.1.1 Quadratic variation and Ito-Föllmer calculus

In 1979 Hans Föllmer published a short paper with title "Ito calculus without probabilities", where he showed how the stochastic calculus invented by Ito, using convergence in of Riemann sums in $L^2(\Omega, P)$ sense, applies surprisingly also pathwise for some non-random functions, using some special sequences of finite partitions.

We choose to start our journey into stochastic analysis from the modern pathwise result of Föllmer, which is rather minimalist.

Later in the following chapters we develop the classical Ito calculus based on martingales.

Note that in the real world is often the case that a random process say $(B_t(\omega) : t \in [0,1])$ is realized only once, and convergence in mean square sense or in probability remain rather abstract and unsatisfactory concepts, while almost sure convergence results are the most meaningful, since we are mainly interested in that single realized path.

This approach is also discussed by Dieter Sondermann in his book *Introduction to stochastic calculus for finance* .

Let (x_t) be the integrator and (y_t) integrand funktions

When (x_t) has *finite variation* , that is $x_t = (x_t^+ - x_t^-)$, jossa $x^\pm$ are Borel-measurable and non-decreasing, and (y_t) is Borel measurable and bounded, the Lebesgue-Stieltjes integral is well defined

$$\int_0^t y_s dx_s = \int_0^t y_s dx_s^+ - \int_0^t y_s dx_s^-$$

When y_s is also piecewise continuous, the Lebesgue-Stieltjes and Riemann-Stieltjes integrals coincide. The differential calculus is first order: for $F(\cdot) \in C^1(\mathbb{R})$,

$$F(x_t) = F(x_0) + \int_0^t F_x(x_s) dx_s + \sum_{s \le t} \big\{ F(x_s) - F(x_{s-}) - F_x(x_{s-})(x_s - x_{s-}) \big\}$$

with correction terms appear at the discontinuities of x_t.

What happens when the integrator is x_t has infinite total variation ? Can we make sense of the limit of Riemann sums for some class of integrands ?

For a path x_t of infinite total variation we can do the following:

by summing p-powers of small increments for some $p > 1$ and taking supremum we define the *p-power variation* of a continuous path x_t as

$$v_t^{(p)}(x) = \sup_{\Pi} \sum_{t_i \in \Pi} |x_{t_{i+1}} - x_{t_i}|^p$$

Since the increments are small, there is a chance that $v_t^{(p)}(x) < \infty$ even in the case were the total variation $v_t(x) = v_t^{(1)}(x) = \infty$.

In Ito calculus we consider $p = 2$ but we use a weaker notion of p-variation, where instead of taking a supremum over all finite partitions Π, we take the limit under a given sequence of partitions.

Consider a sequence of partitions $\{\Pi_n\}$ where

$$\Pi_n = \{t_0^n, \dots, t_{k^n}^n\}, \quad k_n < \infty, \quad 0 \le t_0^n < \dots t_{k^n}^n < \infty,$$
$$\Delta(\Pi_n, t) = \max_{t_i^n \in \Pi_n} (t_{i+1}^n \wedge t - t_i^n \wedge t) \to 0 \quad \text{for } n \to \infty \ .$$

$t \wedge s := \min\{t, s\}$.

Definition 2. *A continuous paths has $x : [0,\infty) \to \mathbb{R}$ pathwise quadratic variation among the sequence $\{\Pi_n\}$, is the sequence of discrete measures*

$$\xi_n(dt) = \sum_{t_i \in \pi_n} (x_{t_{i+1}} - x_{t_i})^2 \delta_{t_i}(dt)$$

converges weakly on compact intervals to a Radon measure $\xi(dt)$. The function $t \mapsto [x,x]_t := \xi([0,t])$ is continuous and defines the quadratic variation of x_t.

Here weak convergence on compacts (also called vague convergence) of $\xi_n \to \xi$ means that for all continuous functions y_s with compact support

$$\int y_s \xi_n(ds) \to \int y_s \xi(ds)$$

Lemma 1. *(Characterization): A continuous path x_t has quadratic variation among the sequence $\{\Pi_n\}$ if and only if*

$$\lim_{n\to\infty} \sum_{t_i \in \Pi_n} (x_{t_{i+1}\wedge t} - x_{t_i \wedge t})^2 = [x,x]_t \quad \forall t < \infty$$

pointwise.

Proof Consider a continuous integrand y_s. Since y is uniformly continuous on the compact $[0,1]$, $\forall \varepsilon > 0$, there are $k, m, \tau_1, \dots, \tau_m$ such that the piecewise constant function

$$y^\varepsilon(s) = \sum_{j=1}^m y_{\tau_j} \mathbf{1}_{(\tau_j, \tau_{j+1}]}(s) \quad \text{satisfies} \quad \sup_{s \le t} |y^\varepsilon(s) - y(s)| < \varepsilon$$

It follows

$$\left| \sum_{t_i \in \pi_n : t_i \le t} y_{t_i} (x_{t_{i+1}^n} - x_{t_i^n})^2 - \int_0^t y_s d[x,x]_s \right| \le$$
$$\left| \sum_{t_i \in \pi_n : t_i \le t} y_{t_i}^\varepsilon (x_{t_{i+1}^n} - x_{t_i^n})^2 - \int_0^t y_s d[x,x]_s \right| + \varepsilon \sum_{t_i \in \pi_n} (x_{t_{i+1}^n} - x_{t_i^n})^2$$
$$= \left| \sum_{j=1}^m y_{\tau_j} \sum_{t_i^n \in \pi_n : \tau_j < t_i^n \le \tau_{j+1} \wedge t} (x_{t_{i+1}^n} - x_{t_i^n})^2 - \int_0^t y_s d[x,x]_s \right| + \varepsilon \sum_{t_i \in \pi_n} (x_{t_{i+1}^n} - x_{t_i^n})^2$$
$$\longrightarrow \left| \sum_{j=1}^m y_{\tau_j} ([x,x]_{\tau_{j+1}\wedge t} - [x,x]_{\tau_j \wedge t}) - \int_0^t y_s d[x,x]_s \right| + \varepsilon [x,x]_t \quad \text{as } n \to \infty \ .$$

Since ε was arbitrary, as $\varepsilon \to 0$ from the definition of Riemann-Stieltjes integral it follows

$$\lim_{n\to\infty} \sum_{t_i \in \pi_n} y_{t_i} (x_{t_{i+1}^n} - x_{t_i})^2 = \int_0^t y_s d[x,x]_s \qquad \square.$$

Lemma 2. *When x_t is continuous and has quadratic variation among $\{\pi_n\}$, then $t \mapsto [x,x]_t$ is continuous.*

Proof: By definition $t \mapsto [x,x]_t$ is right-continuous, since it is the finite limit of right-continuous functions.

Let r be a point of discontinuity: there is $\Delta[x,x]_r := [x,x]_{r+} - [x,x]_{r-} > 0$. For each partition, let $(t^n_, t^n_+]$ be the interval containing r.

Then $(x_{t^n_+} - x_{t^n_-})^2 \to \Delta[x,x]_r$ by definition. On the other hand $(x_{t^n_+} - x_{t^n_-})^2 \to 0$ since x_t is uniformly continuous on compacts.

Remark 1. *Note that for $s < t < u$,*

$$|x_u - x_s| \le |x_u - x_t| + |x_t - x_s|$$

but

$$(x_u - x_s)^2 = (x_u - x_t)^2 + (x_t - x_s)^2 + 2(x_u - x_t)(x_t - x_s)$$

which is not necessarily smaller than $(x_u - x_t)^2 + (x_t - x_s)^2$.

The quadratic variation behaves differently than the first variation, by refining the partition the approximating sum is not necessarily non-increasing.

That's the reason while in the definition of first variation we can take the supremum over all partitions, while with this definition of quadratic variation we follow a given sequence of partitions.

Remark 2. *When x_t is continuous with finite total variation in $[0,t]$, it follows that $[x,x]_t = 0$:*

$$\sum_{t_i \in \pi_n : t_i \le t} (x_{t_{i+1}} - x_{t_i})^2 \le \sup_{t_i \in \pi_n : t_i \le t} |x_{t_{i+1}} - x_{t_i}| \sum_{t_i \in \pi_n : t_i \le t} |x_{t_{i+1}} - x_{t_i}|$$
$$\le \sup_{t_i \in \pi_n : t_i \le t} |x_{t_{i+1}} - x_{t_i}| \, Var_t(x) \to 0 \qquad kun\ n \to \infty,$$

since $Var_t(x) < \infty$. If for some sequence of partitions $\{\Pi_n\}$ exists strictly positive quadratic variation $[x,x]_t > 0$, necessarily $Var_t(x) = \infty$.

We show that for continuous paths with quadratic variation a second order differential calculus holds.

Proposition 1. *(Föllmer 1979): Let x_t a continuous path with pathwise quadratic variation among $\{\Pi_n\}$, and let $F(x) \in C^2(\mathbb{R})$. Then Ito formula holds:*

$$F(x_t) = F(x_0) + \int_0^t F_x(x_s)dx_s + \frac{1}{2}\int_0^t F_{xx}(x_s)d[x,x]_s\ , \quad t > 0$$

where the pathwise Ito-Föllmer integral with respect to x exists as the limit of Riemann sums among the sequence $\{\Pi_n\}$.

$$\int_0^t F_x(x_s)dx_s := \lim_n \sum_{t \ge t_i \in \pi_n} F_x(x_{t_i})(x_{t_{i+1}} - x_{t_i})$$

This is also called forward integral and denoted as

$$\int_0^t F_x(x_s)d\overrightarrow{x}_s$$

Proof: take telescopic sums

$$F(x_t) - F(x_0) = \lim_n \sum_{t \geq t_i \in \pi_n} \big(F(x_{t_{i+1}}) - F(x_{t_i}) \big)$$

and use Taylor expansion

$$\sum_{t \geq t_i \in \pi_n} \big(F(x_{t_{i+1}}) - F(x_{t_i}) \big) =$$

$$\sum F_x(x_{t_i})(x_{t_{i+1}} - x_{t_i}) + \frac{1}{2} \sum F_{xx}(x_{t_i})(x_{t_{i+1}} - x_{t_i})^2 + \sum r(x_{t_i}, x_{t_{i+1}})(x_{t_{i+1}} - x_{t_i})^2$$

where by the middle-point theorem

$$r(x_{t_i}, x_{t_{i+1}}) = \big(F_{xx}(x_i^*) - F_{xx}(x_{t_i}) \big)$$

for some $x_i^* \in (x_{t_i}, x_{t_{i+1}}]$. Note that $r(x_{t_i}, x_{t_{i+1}}) \to 0$ uniformly as $\Delta(\Pi_n) \to 0$ since the map $t \mapsto F_{xx}(x_t)$ is uniformly continuous on compacts.

As $n \uparrow \infty$, by definition of quadratic variation the second Riemann sums converges towards

$$\frac{1}{2} \int_0^t F_{xx}(x_s) d[x, x]_s$$

and the remainder term is dominated by

$$\max_{t_i \in \pi_n} \varphi\big(|x_{t_{i+1}} - x_{t_i}|\big) \sum_{t_i \in \pi_n, t_t \leq t} (x_{t_{i+1}} - x_{t_i})^2 \to 0 \cdot [x, x]_t \quad \text{kun } n \to \infty \ .$$

where $\varphi(\eta) \to 0$ as $\eta \to 0$,

Therefore the limit of Riemann sums among $\{\Pi_n\}$ exists, and it is given by

$$\int_0^t F_x(x_s) dx_s := \lim_n \sum_{t \geq t_i \in \pi_n} F_x(x_{t_i})(x_{t_{i+1}} - x_{t_i})$$
$$= F(x_t) - F(x_0) - \frac{1}{2} \int_0^t F_{xx}(x_s) d[x, x]_s \qquad \square$$

Remark 3.

1. *In general the existence and the value of such forward integral may depend on the particular sequence of partitions. However when $[x, x]$ exists for all $\{\pi_n\}$-sequences and its value does not depend on the sequence then also the forward integral $\int F_x(x_s) d\overrightarrow{x}_s$ is well defined independently of the sequence.*

2. *The existence of quadratic variation in the sense of weak convergence on compacts was the minimal assumption which we used to derive Ito formula.*

   oletus jolla johdetaan poluttainen Iton

3. *We have the following extension of Ito formula: if* $F(x,y) \in C^{2,1}$ *and* y_t *is continuous with finite variation, then*

$$\int_0^t F_x(x_s,y_s)dx_s := \lim_n \sum_{t\geq t_i \in \pi_n} F_x(x_{t_i},y_{t_i})(x_{t_{i+1}} - x_{t_i})$$
$$= F(x_t,y_t) - F(x_0,y_0) - \int_0^t F_y(x_s,y_s)dy_s - \frac{1}{2}\int_0^t F_{xx}(x_s)d[x,x]_s \qquad \square$$

4. *When* $F \in C^1(\mathbb{R})$ *and* x *is continuous with pathwise quadratic variation among* $\{\pi_n\}$*, then the function* $w_t := F(x_t)$ *has also quadratic variation among* $\{\pi_n\}$ *given by*

$$[w,w]_t = \int_0^t F_x(x_s)^2 d[x,x]_s$$

Proof: by Taylor expansion

$$\sum_{t_i\in\pi_n:t_i\leq t} \big\{F(x_{t_{i+1}}) - F(x_{t_i})\big\}^2 = \sum F_x(x_{t_i})^2(x_{t_{i+1}} - x_{t_i})^2 + \sum r(x_{t_i},x_{t_{i+1}})(x_{t_{i+1}} - x_{t_i})^2$$
$$\to \int_0^t F_x(x_s)^2 d[x,x]_s \qquad \text{as } n\to\infty$$

5. *We have defined the forward integral*

$$\int_0^t y_s d\overrightarrow{x}_s$$

for integrands $y_t = F(x_t,y_t)$ *with* $F \in C^{1,1}$ *and* y_t *of finite variation. What about more general integrands ?*

Let π_n a partition and $y \in C([0,t],\mathbb{R})$. Note that

$$I^n_t(y) := \sum_{t\geq t_i\in\pi_n} y_{t_i}(x_{t_{i+1}} - x_{t_i})$$

is a linear operator. We show that when x_t has infinite total variation the integral operator $I_t(y) = \int_0^t y_s dx_s$

$(C([0,t],\mathbb{R}), |\cdot|_\infty)$.

THIS CAN BE SKIPPED

Proposition 2. *(From Protter book) If for all* $y \in C(\mathbb{R})$ *exists*

$$I_t(y) := \lim_n I^n_t(y),$$

it follows that x_t *has finite first variation and therefore* $[x,x]_t = 0$.

Proof: $\forall n$ there is a continuos function $y_n(t)$ such that

$$y_n(t_i) = \mathrm{sign}(x_{t_{i+1}} - x_{t_i}) \quad \forall t_i \in \pi_n,$$

and $|y_n|_\infty = 1$.

For the operator norm

$$\| I_n \| \geq |I_n(y_n)| = \sum_{t \geq t_i \in \pi_n} \mathrm{sign}(x_{t_{i+1}} - x_{t_i})(x_{t_{i+1}} - x_{t_i}) = \sum_{t \geq t_i \in \pi_n} |x_{t_{i+1}} - x_{t_i}|,$$

and

$$\sup_n \| I_n \| \geq \mathrm{Var}(x)_t$$

If $\forall y \in C(\mathbb{R})$ there exists $I(y) = \lim_n I_n(y)$ among $\{\Pi_n\}$, necessarily $\sup_n |I_n(y)| < \infty$, and by the Banach Steinhaus theorem of functional analyis it follows that $\sup_n \| I_n \| < \infty$, which means $\mathrm{Var}(x)_t < \infty$.

We recall Banach-Steinhaus theorem: Let $(I_\nu : \nu \in J)$ a family of linear continuous operators, where $(X_i, |\cdot|_{X_i})$, $i = 1, 2$ are normed-spaces.

If

$$\sup_{\nu \in J} |I_\nu(y)|_{X_2} \leq C(y)$$

when we take supremum on both sides

$$\sup_{\nu \in J} \| I_\nu \| < \infty \quad , \quad \text{where } \| I_\nu \| = \sup_y \frac{|I_\nu(y)|_{X_2}}{|y|_{X_1}}$$

is the strong operator-norm.

1.1.2 Ito-Föllmer calculus for random paths

Definition 3. *Let $(X_t(\omega) : t \geq 0)$ a stochastic process with continuous paths defined on the probability space $(\Omega, \mathcal{F}, P)$. We say that X has stochastic quadratic variation process $([X, X]_t(\omega) : t \geq 0)$ when for all sequence of finite partitions $\{\Pi_n\}$ with $\Delta(\Pi_n, t) \to 0$*

$$\sum_{t \geq t_i \in \Pi_n} (X_{t_{i+1}} - X_{t_i})^2 \stackrel{P}{\to} [X, X]_t$$

with convergence in probability

It follows that for any sequence of finite partitions $\{\Pi_n\}$ with $\Delta(\Pi_n) \to 0$ there is a deterministic subsequence $\{\Pi_{n(m)}\}$ such that (first for all $t \in \mathbb{Q} \cap [0, \infty)$ and then by continuity of $[X, X]$ for all $t \geq 0$)

$$\sum_{t \geq t_i \in \Pi_{n(m)}} (X_{t_{i+1}}(\omega) - X_{t_i}(\omega))^2 \to [X, X]_t(\omega) \quad P\text{-almost surely } \omega \qquad (1.1)$$

In other words when we start with a deterministic sequence of finite partitions $\{\Pi_n\}$ for P-almost all paths $X.(\omega)$ the pathwise quadratic variation $[X(\omega), X(\omega)].$ among that subsquence $\{\Pi_{n(m)}\}$, which coincides with the stochastic quadratic variation $[X, X].(\omega)$.

This implies that Ito formula applies when we define the Ito integral as limit in probability of Riemann sums, which exists also P-almost surely when we take limit among the subsequence $\{\Pi_{n(m)}\}$.

Consider dyadic partitions

$$\Pi_n D_n = \{t_k^n = k 2^{-n} : k = 0, \dots, n 2^n\}$$

Proposition 3. *(by Paul Lévy) Brownian motion has P-almost surely quadratic variation $[B,B]_t = t$ among the dyadic sequence $\{D_n\}$.*

Proof: the variance of the approximating sums is

$$E\biggl(\Bigl\{\sum_{t_k^n \le t} (B_{t_{k+1}^n} - B_{t_k}^n)^2 - (t_{k+1}^n - t_k^n)\Bigr\}^2\biggr) = \sum_{t_k^n \le t} E\bigl(\{(B_{t_{k+1}^n} - B_{t_k}^n)^2 - (t_{k+1}^n - t_k^n)\}^2\bigr)$$

(since increments are independent the cross-product terms have zero expectation).

$$= \sum_{t_k^n \le t} \bigl\{E(\{\Delta B_{t_k^n}\}^4) + (\Delta t_k^n)^2 - 2(\Delta t_k^n) E(\{\Delta B_{t_k^n}\}^2)\bigr\} =$$

$$2 \sum_{t_k^n \le t} (t_{k+1}^n - t_k^n)^2 = 2\lfloor t 2^n \rfloor 2^{-2n} \le 2t2^{-n}$$

Let $\varepsilon > 0$ and

$$A_n^\varepsilon = \bigl\{\omega : |t - \sum_{t_k^n \le t} (B_{t_{k+1}}^n(\omega) - B_{t_k}^n(\omega))^2| > \varepsilon\bigr\}$$

by Chebychev inequality

$$P(A_n^\varepsilon) \le 2t2^{-n}\varepsilon^{-2}$$

Therefore

$$\sum_n P(A_n^\varepsilon) \le \varepsilon^{-2} 4t < \infty$$

Applying Borel Cantelli lemma, $\forall \varepsilon > 0$

$$P\bigl(\limsup_n A_n^\varepsilon\bigr) = 0$$

Taking $\varepsilon = 1/m,\ m \in \mathbb{N}$ and countable intersection of the complements

$$P\biggl(\bigcap_{m\ge 0} \bigcup_{k \ge 0} \bigcap_{n \ge k} A_n^{1/m}\biggr) = 1$$

which is the probability that exists $[B,B]_t = t$ by taking limits among the dyadic sequence.

Remark 4. *1. Essentially we used*

$$\sum_n \biggl(\sum_{t_k^n \le t} (t_{k+1}^n - t_k^n)^2\biggr) < \infty$$

which gives the rate of convergence of $\Delta(\Pi_n)$ to zero in order to obtain almost sure convergence from convergence in probability,

2. *The set of measure zero where convergence fails may well depend on the sequence of partitions. We cannot take supremum over partitions.*

3. *By a backward martingale argument his theorem extends to refining sequences of partitions with $\Pi_n \subseteq \Pi_{n+1}$, $\Delta(\Pi_n, t) \to 0$ kun $n \to \infty$ (you find in the book by Revuz and Yor, Continuous martingales and Brownian motion,Proposition 2.12) .*

1.1.3 Pathwise Stratonovich calculus

If in the approximating Riemann sums we evaluate the integrand at the midpoint rather than in theleft point we obtain

$$\begin{aligned}
&\sum_{t_i \in D_n : t_i \le t} F_x(B_{(t_{i+1}+t_i)/2})(B_{t_{i+1}} - B_{t_i}) = \\
&= \sum F_x(B_{t_i})(B_{t_{i+1}} - B_{t_i}) + \sum (F_x(B_{(t_{i+1}+t_i)/2}) - F_x(B_{t_i}))(B_{t_{i+1}} - B_{t_i}) \\
&= \sum F_x(B_{t_i})(B_{t_{i+1}} - B_{t_i}) + \sum F_{xx}(B_{t_i})(B_{(t_{i+1}+t_i)/2} - B_{t_i})(B_{t_{i+1}} - B_{t_i}) + \\
&+ \sum r(B_{(t_{i+1}+B_i)/2}, B_{t_i})(B_{(t_{i+1}+t_i)/2} - B_{t_i})(B_{t_{i+1}} - B_{t_i}) \\
&= \sum F_x(B_{t_i})(B_{t_{i+1}} - B_{t_i}) + \sum F_{xx}(B_{t_i})(B_{(t_{i+1}+B_i)/2} - x_{t_i})^2 + \\
&+ \sum F_{xx}(B_{t_i})(B_{(t_{i+1}+B_i)/2} - x_{t_i})(B_{t_{i+1}} - B_{(t_{i+1}+t_i)/2}) + \\
&+ \sum r(B_{(t_{i+1}+t_i)/2}, B_{t_i})(B_{(t_{i+1}+t_i)/2} - B_{t_i})(B_{t_{i+1}} - B_{t_i})
\end{aligned}$$

Lemma 3. *For the Brownian path*

$$\sum_{t_i \in D_n : t_i \le t} (B_{(t_{i+1}+t_i)/2} - B_{t_i})^2 \to \frac{1}{2}[B,B]_t = \frac{1}{2} t \ , \tag{1.2}$$

$$\sum_{t_i \in D_n : t_i \le t} (B_{(t_{i+1}+t_i)/2} - B_{t_i})(B_{t_{i+1}} - B_{(t_{i+1}+t_i)/2}) \to 0 \ , \tag{1.3}$$

Proof: exercise.

It follows that the Riemannin sums among the dyadics converge P-a.s. to the pathwise Stratonovich integral

$$\begin{aligned}
&\int_0^t F_x(x_s) \circ dx_s := \int_0^t F_x(x_s) dx_s + \frac{1}{2} \int_0^t F_{xx}(x_s) d[x,x]_s \\
&= F(x_t) - F(x_0) - \frac{1}{2} \int_0^t F_{xx}(x_s) d[x,x]_s + \frac{1}{2} \int_0^t F_{xx}(x_s) d[x,x]_s = F(x_t) - F(x_0) \ .
\end{aligned}$$

We see that

The Stratonovich integral follows the ordinary first order calculus:

$$\int_0^t F_x(B_s) \circ dB_s = \int_0^t F_x(B_s) dB_s + \frac{1}{2} \int_0^t F_{xx}(B_s) ds = F(B_t) - F(B_0)$$

By evaluating in the Riemann sums the integrand at the right point we obtain the pathwise *backward integral*

$$\begin{aligned}
&\int_0^t F_x(B_s) d\overleftarrow{B}_s = \lim_{n\to\infty} \sum_{t_i^n \in D_n} F_x(B_{t_{i+1}^n})(B_{t_{i+1}^n \wedge t} - B_{t_i^n}) \\
&= F(B_t) - F(B_0) + \frac{1}{2} \int_0^t F_{xx}(B_s) ds = \int_0^t F_x(B_s) d\overrightarrow{B}_s + \int_0^t F_{xx}(B_s) ds
\end{aligned}$$

Proof: exercise.

References H. Föllmer, "Calcul d Ito sans probabilites" (1980). Séminaire de Probabilités XV, pp 143-149 Springer

D. Sondermann, " Intoduction to stochastic calculus for finance " Springer.

1.1.4 Preliminaries on Gaussian random variables

Definition 4. *A random vector $X = (X_1, \dots, X_n)$ with values in $\mathbb{R}^n$ is jointly Gaussian iff there is a $\mu \in \mathbb{R}^n$ and a non-negative definite matrix K such that the joint characteristic function is given by*

$$\phi_X(\theta) := E(\exp(i\theta \cdot X)) = \exp(i\theta\mu - \frac{1}{2}\theta K \theta^T)$$

where $y \cdot x$ is the usual scalar product.

When the limit of a Gaussian random variable exists, it is necessarly Gaussian:

Lemma 4. *Let $\{\xi_n\}$ be a sequence of Gaussian r.v. with respective distributions $\mathcal{N}(\mu_n, \sigma_n^2)$, defined on the same probability space $(\Omega, \mathcal{F}, P)$, together with a r.v. ξ. If $\xi_n \overset{d}{\to} \xi$ (convergence in distribution) then ξ is Gaussian $\mathcal{N}(\mu, \sigma^2)$ where the limits $\mu = \lim_n \mu_n$ and $\sigma^2 = \lim_n \sigma_n^2$ exist.*

When $\sigma^2 = 0$, we agree that the constant random variable μ is Gaussian with zero variance.

Proof Since convergence in distribution is equivalent to the convergence of characteristic functions, it follows that

$$\phi_{\xi_n}(\theta) = \exp\left(i\mu_n\theta - \frac{1}{2}\theta^2\sigma_n^2\right) \to \phi_\xi(\theta) \quad \forall\theta$$

where $\forall\theta$

$$|\phi_{\xi_n}(\theta)| = \exp\left(-\frac{1}{2}\theta^2\sigma_n^2\right) \to |\phi_\xi(\theta)| = \exp\left(-\frac{1}{2}\theta^2\sigma^2\right)$$
$$\mathrm{Arg}(\phi_{\xi_n}(\theta)) = \mu_n\theta \to \mathrm{Arg}(\phi_\xi(\theta)) = \mu\theta$$

therefore

$$\phi_\xi(\theta) = \exp\left(i\mu\theta - \frac{1}{2}\theta^2\sigma^2\right) \quad \square$$

In particular if $\{\xi_n\}$ are Gaussian random variables with $\xi_n \overset{P}{\to} \xi$ in probability, then ξ is Gaussian and $\xi_n \to \xi$ in $L^p(\Omega)$ $\forall p < \infty$.

Remark We can replace convergence in distribution the lemma 4 with stronger convergence in probability or in L^p convergence,

Corollary 1. *If $X_n \to 0$ in probability and $X_n \sim \mathcal{N}(\mu_n, \sigma_n^2)$, then $\mu_n, \sigma_n^2 \to 0$ and $X_n \to 0$ in $L^p(\Omega)$ for all $p < \infty$.*

Definition 5. *A family of real valued random variables $\{\xi_t : t \in T\}$ is a Gaussian process if $\forall$ n, $t_1, \dots, t_n \in T$ the law of $(\xi_{t_1}, \dots, \xi_{t_n})$ is jointly Gaussian.*

Lemma 5. *(Gaussian integration by parts and tail probabilities)*

- *The standard Gaussian density*

$$\phi(x) := \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

satisfies

$$\frac{d\phi}{dx}(x) = -x\phi(x)$$

- *For a standard Gaussian random variable $G(\omega)$ with $E(G) = 0,\ E(G) = 1$ we have the Gaussian integration by parts formula:*

$$E_P\Big(f'(G)h(G) \Big) = E_P\Big(f(G)(Gh(G) - h'(G)) \Big)$$

In particular for $h(x) \equiv 1$

$$E_P\Big(f'(G) \Big) = E_P\Big(f(G)G \Big)$$

- *For $x > 0$ we have the upper bound*

$$P(G > x) = \int_x^\infty \phi(y)dy \le \int_x^\infty \frac{y}{x}\phi(y)dy = -\frac{1}{x}\int_x^\infty \phi'(y)dy =$$
$$\frac{1}{x}\{\phi(x) - \phi(\infty)\} = \frac{1}{x}\phi(x)$$

1.2 Paul Lévy's construction of Brownian motion

We have defined Brownian motion but we haven't yet shown that such stochastic process exists.

We construct recursively the Brownian motion on the dyadics $D_n \subseteq [0,1]$.

Given the values $(B_t : t \in D_n)$, we obtain by linear interpolation a continuous path $(B_t^{(n)}(\omega) : t \in [0,1])$.

Then we show that $B_t^{(n)}(\omega)$ converges uniformly for $t \in [0,1]$.

More precisely, let $(G_d(\omega) : d \in D)$ i.i.d. standard Gaussian random variables, where the dyadics $D = \bigcup_{n\in\mathbb{N}} D_n$ are countable.

At level $n = 0$, for $D_0 = \{0,1\}$ set

$$B_0(\omega) = 0,\ B_1(\omega) = G_0(\omega),$$
$$\text{and by linear interpolation } B_t^{(0)}(\omega) := tB_1(\omega),\ t \in [0,1]$$

Define the increasing sequence of σ-algebrae $\mathcal{G}_n = \sigma(B_d : d \in D_n)$.

Let $d \in D_n \setminus D_{n-1}$ and $d^-, d^+ \in D_{n-1}$ with $d^- < d < d^+$ and $d^+ - d^- = 2^{n-1}$. $d^\pm$ are the nearest neighbours of d at the previous level $(n-1)$.

Since the increments of (B_t) are independent,

$$P(B_d \in dx|\mathcal{G}_{n-1}) = P(B_d \in dx|B_d, B_{d^+})$$

which is a Gaussian law with mean $(B_{d^-} + B_{d^+})/2$ and variance

$$((d - d^-)^{-1} + (d^+ - d^-)^{-1})^{-1} = 2^{-(n+1)})$$

(Exercise: check this by using Bayes' formula).

We define inductively for $d \in D_n \setminus D_n$ and corresponding $d^\pm \in D_{n-1}$

$$B_d(\omega) := \frac{B_{d^-}(\omega) + B_{d^+}(\omega)}{2} + G_d(\omega)2^{-(n+1)/2} \tag{1.4}$$

Note that, for $t \in D$

$$B_t(\omega) := \sum_{d \in D} G_d(\omega)\eta_d(t) = \sum_{d \in D} G_d(\omega) \int_0^t \dot{\eta}_d(s)ds = \tag{1.5}$$

$$= \sum_{d \in D_n} G_d(\omega)\eta_d(t) = \sum_{d \in D_n} G_d(\omega) \int_0^t \dot{\eta}_d(s)ds, \quad \text{when } t \in D_n \tag{1.6}$$

where $\dot{\eta}_0(s) \equiv 0$, $\eta_1(s) := \mathbf{1}_{[0,1]}(s)$ and for $d \in D_n \setminus D_{n-1}$, $n > 0$,

$$\eta_d(s) = \left\{ \mathbf{1}_{[d^-,d)}(s) - \mathbf{1}_{[d,d^+)}(s) \right\} 2^{(n-1)/2}$$

and $d^{\pm}$ are the nearest neighbours of $d \in D_n \setminus D_{n-1}$ at level $(n-1)$.

To visualize the function $t \mapsto B_t(\omega)$, is the infinite sums of sawtooth function each with support on some dyadic interval $[k2^{-n}, (k+1)2^{-n})$ with independent Gaussian weights.

Note that for $d \in D_N \setminus D_{N-1}$ with neighbours $d_-, d_+ \in D_{N-1}$

$$0 = \int_0^1 \dot{\eta}_d(s)ds = \int_{d-}^{d+} \dot{\eta}_d(s)ds$$

so that

$$\int_0^t \dot{\eta}_d(s)ds = 0$$

for all $t \notin (d_-, d_+)$. Since $D_{N-1} \cap (d_-, d_+) = \emptyset$ necessarily

$$\int_0^t \dot{\eta}_d(s)ds = 0$$

for $d \in D_N \setminus D_{N-1}$ and $t \in D_{N-1}$. This shows that every $t \in D$ has a finite series expansion.

Let's show that for each $t \in D$ the series expansion (1.5) satisfies the recursion step (1.4).

Note first that for $t \in [0,1], \forall n \in N$, there is one and only one $d \in D_n \setminus D_{n-1}$ such that $t \in \text{support}(\eta_d)$.

Assume that $t \in D_N \setminus D_{N-1}$ with neighbours $t_-, t_+ \in D_{N-1}$.

Then

$$\frac{B_{t^-}(\omega) + B_{t^+}(\omega)}{2} + G_t(\omega)2^{-(N+1)/2} =$$
$$\sum_{n=0}^{N-1} \sum_{d \in D_n} G_d(\omega)\frac{1}{2}\left(\int_0^{t-} \eta_d(s)ds + \int_0^{t+} \dot{\eta}_d(s)ds \right) + G_t(\omega) \int_0^t \dot{\eta}_t(s)ds$$

where for $t \in D_N \setminus D_{N-1}$

$$\int_0^t \eta_t(s)ds = \int_{t_-}^t \dot{\eta}_t(s)ds = 2^{-N}2^{(N-1)/2} = 2^{-(N+1)/2}$$

and $\forall d \in D_{N-1}$, $t \in D_N \setminus D_{N-1}$

$$\frac{1}{2}\Big(\int_0^{t-} \dot\eta_d(s)ds + \int_0^{t+} \dot\eta_d(s)ds\Big) = \int_0^t \eta_d(s)ds$$

since when $d \in D_{N-1}$, $\dot\eta_d(s)$ is constant in the interval $(t-, t+)$, and we have obtained the series expansion (1.5) of $B_t(\omega)$.

We show that for P-almost surely the infinite series representation of $B_t(\omega)$ is converging uniformly on $t \in [0,1]$,

We use the Gaussian tail estimates: given $c > 0$ for n large enough, $G_d \sim \mathcal{N}(0,1)$

$$P(|G_d| > c\sqrt{n}) \le \exp\big(-\frac{c^2 n}{2}\big)$$

$$P\big(\omega : \exists d \in D_n \setminus D_{n-1} \text{ with } |G_d(\omega)| > c\sqrt{n}\big) \le \sum_{n \in D_n \setminus D_{n-1}} P(|G_d| > c\sqrt{n})$$

$$\le 2^{n-1}\exp\big(-\frac{c^2 n}{2}\big) \le \exp(-\alpha n)$$

when $c > \sqrt{\alpha + 2\log 2} > \sqrt{2\log 2}$, for some $\alpha > 0$.

For such c, since

$$\sum_{n\ge 0} \exp(-\alpha n) = (1 - \exp(-\alpha))^{-1} < \infty$$

by Borel Cantelli lemma

$$P\Big(\omega : \exists N(\omega) \text{ with } |G_d(\omega)| \le c\sqrt{n} \forall n \ge N(\omega), d \in D_n \setminus D_{n-1}\Big) = 1$$

Therefore for P-almost all ω and $n \ge N(\omega)$

$$\Big|\sum_{d \in D_n \setminus D_{n-1}} G_d(\omega) \int_0^t \dot\eta_d(s)ds\Big| \le c\sqrt{n} 2^{-(n+1)/2}$$

since for $d \in D_n \setminus D_{n-1}$, with neighbours $d^-, d^+ \in D_{n-1}$

$$\int_0^t \dot\eta_d(s)ds = 0$$

when $t \notin (d^-, d^+)$, and for $t \in (d^-, d^+)$

$$0 \le \int_0^t \dot\eta_d(s)ds \le \int_0^d \dot\eta_d(s)ds = 2^{-(n+1)/2}$$

so that P-almost surely the series

$$\sum_{n\ge 0} \sum_{d \in D_n \setminus D_{n-1}} G_d(\omega) \int_0^t \dot\eta_d(s)ds = \lim_{n\to\infty} B_t^{(n)}(\omega)$$

is absolutely convergent uniformly in $[0,1]$. This means that P-almost surely $\{t \mapsto B_t^{(n)}(\omega) : n \in \mathbb{N}\}$ is a Cauchy sequence on the space of continuous functions $C([0,1],\mathbb{R})$ equipped with the uniform norm. By completeness, for P-almost all ω a continuous limiting function $t \mapsto B_t(\omega)$ exists.

The set $(B_d(\omega) : d \in D)$ is a Brownian motion on the dyadics, since by construction at every dyadic level D_n the distribution of $(B_d : d \in D_n)$ coincides with the finite dimensional distribution of the Brownian motion.

Let's fix $k \geq 0$ and $0 = t_0 < t_1 < \cdots < t_k \leq 1$.

We find a sequence $(t_1^{(n)}, \ldots, t_k^{(n)}) \subseteq D_n$ such that $\max_{0 \leq i \leq k} |t_i^{(n)} - t_i| \leq 2^n$.

For P-almost all ω the path $t \mapsto B_t(\omega)$ is continuous, and

$$\left(B_{t_1^{(n)}}(\omega), \ldots, B_{t_k^{(n)}}(\omega)\right) \to \left(B_{t_1}(\omega), \ldots, B_{t_k}(\omega)\right)$$

Since $\left(B_{t_1^{(n)}}(\omega), \ldots, B_{t_k^{(n)}}(\omega)\right)$ is a jointly Gaussian vector and almost sure convergence implies convergence in distribution, by the multivariate version of lemma 4 it follows that the limit is a Gaussian random vector.

Morever since the increments are bounded in $L^2(\Omega)$

$$\delta_{ij}(t_i - t_{i-1}) = \lim_{n\to\infty} \delta_{ij}(t_i^{(n)} - t_{i-1}^{(n)}) =$$
$$\lim_{n\to\infty} E\left((B_{t_i^{(n)}} - B_{t_{i-1}^{(n)}})(B_{t_j^{(n)}} - B_{t_{j-1}^{(n)}}) \right) = E\left((B_{t_i} - B_{t_{i-1}})(B_{t_j} - B_{t_{j-1}}) \right)$$

where since Gaussian variables have moments of all order, in the last equality we can pass the limit inside the expectation by uniform integrability.

Therefore the increments of $B_t(\omega)$ over disjoint intervals are jointly Gaussian and uncorrelated, with $E((B_t - B_s)^2) = (t-s)$. We conclude that $(B_t(\omega) : t \in [0,1])$ is a Brownian motion.

1.3 Wiener integral, isonormal Gaussian processes, and white noise

Definition 6. *Define the Cameron-Martin space of absolutely continuous functions with square integrable derivative*

$$H = \left\{ t \mapsto h(t) = \int_0^t \dot{h}(s)ds : \dot{h} \in L^2([0,1], dt) \right\}$$

For $h, f \in H$ with $h(t) = \int_0^t \dot{h}(s)ds$, $f(t) = \int_0^t \dot{f}(s)ds$ we define the scalar product

$$(h,f)_H := (\dot{h}, \dot{f})_{L^2([0,1])} = \int_0^1 \dot{h}(s)\dot{f}(s)ds$$

H equipped with the scalar product is an Hilbert space. $\parallel h \parallel_H := \sqrt{(h,h)_H}$ is a norm.

The functions $\{\dot\eta_d(s) : d \in D\}$ used in Lévy construction form the Haar system, which is a complete orthonormal basis of the Hilbert space $L^2([0,1], dt)$, meaning that

$$(\eta_{d'}, \eta_{d``})_H = (\dot\eta_{d'}, \dot\eta_{d''})_{L^2([0,1])} = \int_0^1 \dot\eta_{d'}(s)\dot\eta_{d''}(s)ds = \delta_{d',d''}$$

and every $\dot h \in L^2([0,1], dt)$ has expansion

$$\dot h(t) = \sum_{n\geq 0}\sum_{d\in D_n} \dot\eta_d(t)(\dot\eta_d, \dot h)_{L^2([0,1])}$$

where the series converges in $L^2([0,1], dt)$-sense.

Equivalently the primitives

$$t \mapsto \eta_d(t) = \int_0^t \dot\eta_d(s)ds$$

form a complete orthonormal basis in H, so that every $h \in H$ has the expansion

$$h(t) = \sum_{n\geq 0}\sum_{d\in D_n} \eta_d(t)(\eta_d, h)_H$$

converging in $\| \cdot \|_H$ norm.

Definition 7. *An isonormal Gaussian space $\{B(h) : h \in H\}$ is a collection of zero mean jointly Gaussian random variables such that the covariance structure matches the scalar product in H*

$$E\big(B(h)B(f)\big) = (h, f)_H = \int_0^1 \dot h(s)\dot f(s)ds$$

for $h, f \in H$.

In particular we have the isometry between the subspace $\{B(h) : h \in H\}$ of $L^2(\Omega, \mathcal{F}, P)$ and H

$$\| B(h) \|^2_{L^2(\Omega,P)}\ E\big(B(h)^2\big) = \int_0^1 \dot h(s)^2 ds = \| h \|^2_H$$

Note that if $(h_n : n \in \mathbb{N}) \subseteq H$ is a Cauchy sequence in H-norm, then by the isometry the Gaussian variables $(B(h_n) : n \in \mathbb{N}) \subseteq L^2(\Omega, P)$ form a Cauchy sequence, and since L^2 is complete necessarily it has a limit in L^2 sense. Moreover the limit must be Gaussian, since limits in distribution of Gaussian variables are Gaussian, and L^2-convergence is stronger than convergence in probability which implies convergence in distribution.

In this way we define stochastic integrals of functions $\dot h(s) \in L^2([0,1], dt)$:

We approximate $\dot h(s)$ by piecewise constant functions

$$\dot h_n(s) = \sum_{t_i^n\in\Pi_n} y_i^n \mathbf{1}_{(t_{i-1}^n, t_i^n]}(s)$$

in $L^2([0,1], dt)$, for some $(y_1 \dots y_n)$ and Π_n finite partition of $[0,1]$ Equivalently

$$h_n(t) = \int_0^t \dot h_n(s)ds \text{ approximates } h(t) = \int_0^t \dot h(s)ds$$

in the Cameron Martin space H.

For such piecewise constant function we define the stochastic integral as the Riemann sum

$$B(h_n) := \int_0^t \dot h_n(s) dB_s = \sum_{t_i^n \in \Pi_n} y_i^n (W_{t_i^n} - W_{t_{i-1}^n})$$

we check that this satisfies the isometry, which then is used to define the stochastic integral

$$B(h) = \int_0^1 \dot h(s) dB_s$$

as the limit in $L^2(\Omega, P)$ of the Cauchy sequence $(B(h_n))$.

This was historically the first construction of a stochastic integral with deterministic integrands and it is due to Norbert Wiener. Using martingales, Kiyoshi Ito extended the construction to a much wider class of random integrand processes.

These Gaussian variables are identified with the *Wiener integrals*

$$B(h) = \int_0^1 \dot h(s) dB_s, \quad h \in H$$

Let $\{G_d(\omega) : d \in D\}$ i.i.d. standard Gaussian variables on the probability space $(\Omega, \mathcal{F}, P)$. We construct the isonormal Gaussian space indexed by $h \in H$ as follows:

For the elements of the Haar basis, define

$$\int_0^1 \eta_d(s) dB_s := G_d, \quad d \in D$$

For $h \in H$ By using the Haar expansion,

$$B(h) = \int_0^1 \dot h(s) dB_s := \sum_{n \geq 0} \sum_{d \in D_n \setminus D_{n-1}} G_d(\omega) (\dot h, \dot\eta_d)_{L^2([0,1])}$$

where the infinite sum converges in $L^2(\Omega, \mathcal{F}, P)$.

In particular for $t \in [0,1]$ and $\dot h(s) = \mathbf{1}_{[0,t]}(s)$

$$\begin{aligned} B(h) = \int_0^1 \mathbf{1}_{[0,t]}(s) dB_s = \int_0^t dB_s = B_t = \\ \sum_{n \geq 0} \sum_{d \in D_n \setminus D_{n-1}} G_d(\omega) \int_0^1 \dot\eta_d(s) \mathbf{1}_{[0,t]}(s) ds \\ = \sum_{n \geq 0} \sum_{d \in D_n \setminus D_{n-1}} G_d(\omega) \int_0^t \dot\eta_d(s) ds \end{aligned}$$

where the convergence is in $L^2(\Omega, \mathcal{F}, P)$.

Note this is exactly the series expansion used in Paul Lévy construction of Brownian motion, and it was shown that it converges P-almost surely in the Banach space of continuous functions equipped with uniform norm, which implied that P-almost surely $t \mapsto B_t(\omega)$ is continuous.

This construction works also by replacing the Haar system with any another complete orthonormal system in $L^2([0,1], dt)$.

Another insight is given by using white noise. Let $\{\dot{B}_t(\omega) : t \in [0,1]\}$ a zero-mean Gaussian *generalized process* with the covariance defined formally as the generalized function

$$E(\dot{B}_t \dot{B}_s) = \delta_0(t-s)$$

where $\delta_0(t-s)$ is the Dirac delta function of distribution theory, meaning that for $t \neq s$ $\dot{B}_t$ and $\dot{B}_s$ are uncorrelated while $\dot{B}_t$ has infinite variance. Such object does not exists pointwise since there are not Gaussian variables with infinite variance.

Formally $\dot{B}_t = \frac{dB_t}{dt}$ is the derivative of Brownian motion (whose paths are almost surely is nowhere differentiable as we will see).

Define for $h \in H$

$$B(h) = \int_0^1 \dot{h}(s) dB_s = \int_0^1 \dot{h}(s) \frac{dB_s}{ds} ds = \int_0^1 \dot{h}(s) \dot{B}(s) ds$$
$$= (\dot{h}, \dot{B})_{L^1([0,1])} = (h, B)_H$$

Note that $(h, B)_H$ is not defined ω-wise but it will be well define in $L^2(\Omega, P)$ sense as the limit of the smooth truncated series

We see using Fubini that

$$E(B(h)B(f)) = E\Big(\int_0^1 \dot{h}(s) dB_s \int_0^1 \dot{f}(t) dB_t \Big) = E\Big(\int_0^1 \dot{h}(s) \dot{B}(s) ds \int_0^1 \dot{f}(t) \dot{B}_t dt \Big)$$
$$= \int_0^1 \int_0^1 \dot{h}(s) \dot{f}(t) E\big(\dot{B}(s)\dot{B}(t)\big) dt \, ds = \int_0^1 \int_0^1 \dot{h}(s) \dot{f}(t) \delta_0(t-s) dt ds =$$
$$= \int_0^1 \dot{h}(s) \Big(\int_0^1 \dot{f}(t) \delta_0(t-s) dt \Big) ds =$$
$$\int_0^1 \dot{h}(s) \dot{f}(s) ds = (\dot{h}, \dot{f})_{L^2([0,1],dt)} = (h, f)_H$$

Note that for the Haar system $\{\eta_d : d \in D\}$

$$\dot{B}(s) := \sum_{n \geq 0} \sum_{d \in D_n} G_d(\omega) \eta_d(s)$$

satisfies formally the definition of white noise, since

$$E\Big(\sum_{d \in D} G_d \dot{\eta}_d(s) \sum_{d' \in D} G_{d'} \dot{\eta}_{d'}(t) \Big) = \sum_{d \in D} \sum_{d' \in D} \dot{\eta}_d(s) \dot{\eta}_{d'}(t) E\big(G_d G_{d'}\big)$$
$$= \sum_{d \in D} \dot{\eta}_d(s) \dot{\eta}_d(t) E(G_d^2) = \sum_{d \in D} \dot{\eta}_d(s) \dot{\eta}_d(t)$$

and by the Plancharel identity

$$\int_0^1 \int_0^1 \Big\{ \sum_{d\in D} \dot\eta_d(s)\dot\eta_d(t) \Big\} f(t)h(s)ds = \sum_{d\in D} \Big(\int_0^1 f(t)\eta_d(t)ds \Big) \Big(\int_0^1 h(s)\eta_d(s)ds \Big)$$

$$= \sum_{d\in D} (\dot f, \dot\eta_d)_{L^2([0,1])} (\dot h, \dot\eta_d)_{L^2([0,1])} = (\dot f, \dot h)_{L^2([0,1])}$$

$$= \int_0^1 \dot f(t)\dot h(t)dt = \int_0^1 \int_0^1 \dot f(t)\dot h(s)\delta_0(t-s)dtds$$

which shows that formally the covariance is the Dirac delta function

$$E(\dot B_t \dot B_s) = \sum_{d\in D} \dot\eta_d(s)\dot\eta_d(t) = \delta_0(t-s)$$

Conclusion the white noise $\dot B_t$ introduced formally as the derivative of Brownian motion is a generalized random process which does not exist pointwise but it makes sense to to integrate test function against it.

Chapter 2

Stochastic process: Kolmogorov's construction

2.1 Kolmogorov's extension

We skipped this section during the lectures since we have used Lévy's construction

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 8. *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 9. *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: $\forall k \in \mathbb{N}, t_1 \ldots t_k \in T$ $B_1, \ldots B_k \in \mathcal{B}(\mathbb{R}^d)$*

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

Definition 10. *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 11. *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 12. *We say that the family of finite dimensional distributions*

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

is consistent , when

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

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

Theorem 1. *(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 $\forall n \in \mathbb{N}, t_1, \dots, t_n \in \mathbb{N}, B_n \in \mathcal{B}(\mathbb{R}^n)$,

$$\mathbf{P}\left(\omega \in \mathbb{R}^T :\ (\omega_{t_1}, \dots, \omega_{t_n}) \in B_n\right) = P_{t_1,\dots,t_n}(B_n) \tag{2.1}$$

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 (2.1) 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 $\mathbf{P}(C_n) \ge \varepsilon > 0 \ \forall n$ 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 $\mathbf{P}(D_n) \ge \varepsilon > 0 \ \forall n$ 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

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

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, and $D^* \subseteq \bigcap_n F_n$ contradicting the hypothesis □

Definition 13. *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 6. *In a Borel space, the σ-algebra $\mathcal{S}$ is countably generated.*

Corollary 2. *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.

2.2 Continuity

We skipped also this section during the lectures since we have used Lévy's construction

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 2. *(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* $p, r > 0$

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

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

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

$$E\biggl(\sup_{(s,t)\in N_m} \| X_t - X_s \|^p\biggr) \le \sum_{(s,t)\in N_m} E\bigl(\| X_t - X_s \|^p\bigr) \le (2^{d(m+1)}d)(c2^{-m(d+r)}) = 2^d dc 2^{-mr} \tag{2.2}$$

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 (2.2)

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

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

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

which implies

$$K_\alpha(\omega) := \sup_{(s,t)\in D: s\neq t} \frac{\| X_t(\omega) - X_s(\omega) \|}{|t-s|^\alpha} < \infty \quad P\text{-almost surely} \tag{2.3}$$

Note that $\omega \mapsto K_\alpha(\omega)$ is measurable and $K_\alpha \in L^p(\Omega)$. By taking countable intersections of these events with $\alpha_n = \frac{r}{p}\big(\frac{n}{n+1}\big)$, almost surely (2.3) holds simultaneously for all $\alpha < r/p$ $\square$

Corollary 3. *Under the assumptions of Theorem 2, 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 $t \in [0,1]$

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

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

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

Corollary 4. *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 $T = \mathbb{R}$.

Chapter 3

Probability theory, complements

3.1 Change of measure

For a random variable $X(\omega)$ we say $X \in \mathcal{F}$, or $X \in L^0(\Omega, \mathcal{F})$, when X is $\mathcal{F}$-measurable.

For $X \in \mathcal{F}$ and $X(\omega) \geq 0 \forall \omega$ denote $X \in \mathcal{F}^+$.

If $X \in \mathcal{F}$ and $X(\omega) \geq 0$ P-a.s. denote $X \in L^0_+(\Omega, \mathcal{F})$.

Let

$$X(\omega) = \sum_{i=1}^{n} x_i \mathbf{1}_{A_i}(\omega)$$

for $x_i \in \mathbb{R}$ and $A_i \in \mathcal{F}$, $n \in \mathbb{N}$. We say that X is a **simple** r.v. and denote $X \in \mathcal{YF}$. Denote also $\mathcal{YF}^+ = \mathcal{YF} \cap \mathcal{F}^+$.

On the probability space $(\Omega, \mathcal{F}, P)$, let $Z(\omega) \geq 0$ P-a.s. with $0 < E_P(Z) < \infty$, which implies $P(\{\omega : Z(\omega) > 0\}) > 0$.

We introduce a new probability measure $Q : \mathcal{F} \to [0, 1]$

$$Q(A) := \frac{E_P(Z\mathbf{1}_A)}{E_P(Z)} \quad \forall A \in \mathcal{F}$$

Q is a probability: clearly it is additive and $Q(\Omega) = 1$. It is also σ-additive: $A_n \uparrow Omega$, (which means $A_n \subseteq A_{n+1}$ ja $\bigcup_n A_n = \Omega$), also $Z(\omega)\mathbf{1}_{A_n}(\omega) \uparrow Z(\omega)$ P-a.s. Using the monotone convergence theorem, it follows

$$Q(A_n)E_P(Z) = E_P\big(Z\mathbf{1}_{A_n}\big) \uparrow E_P(Z) = Q(\Omega)E_P(Z) \quad \Longrightarrow \quad Q(A_n) \uparrow 1$$

We can also use the normalized r.v.

$$\widetilde{Z}(\omega) := \frac{Z(\omega)}{E_P(Z)}$$

with $E_P(\tilde{Z}) = 1$, and write $Q(A) = E_P\big(\widetilde{Z}\mathbf{1}_A\big)$.

Theorem 3. *$\forall A \in \mathcal{F}$ $P(A) = 0 \implies Q(A) = 0$. We say that Q is absolutely continuous with respect to P, and denote $Q \ll P$.*

Proof $P(A) = 0 \Longrightarrow Z(\omega)\mathbf{1}_A(\omega) = 0$ P a.s.

Theorem 4. *When $X \in \mathcal{F}^+$, (which means $X(\omega) \geq 0$ P-a.s. and $\mathcal{F}$-measurable) ,*

$$E_Q(X) = \frac{E_P(XZ)}{E_P(Z)},$$

and $X \in L^1(\Omega, \mathcal{F}, Q)$ if and only if $(XZ) \in L^1(\Omega, \mathcal{F}, P)$.

Proof: when $X(\omega)$ is a simple random variable taking finitely many non-negative values (denote $X \in \mathcal{YF}^+$), it follows straight from the definition and linearity of the expectation. When $X \in \mathcal{F}^+$ there is monotone sequence of simple random variables such that $0 \leq X_n(\omega) \uparrow X(\omega)$ $\forall\omega$. By applying twice the monotonisen convergence theorem under Q and under P, we see that $E_Q(X_n) \uparrow E_Q(X)$ and

$$E_Q(X_n) = \frac{E_P(X_n Z)}{E_P(Z)} \uparrow \frac{E_P(XZ)}{E_P(Z)} \quad \square$$

Exercise 3. *Elementary conditional probability*

For $B \in \mathcal{F}$ with $P(B) > 0$, we change the probabity measure using the r.v. $Z(\omega) = P(B)^{-1}\mathbf{1}_B(\omega)$, obtaining

$$P(A|B) := E_P(Z\mathbf{1}_A) = \frac{E_P(\mathbf{1}_A\mathbf{1}_B)}{P(B)} = \frac{P(A \cap B)}{P(B)}, \qquad A \in \mathcal{F}$$

The map $P(\,\cdot\,|B) : A \in \mathcal{F} \mapsto P(A|B) \in [0,1]$ is a probability measure on $(\Omega, \mathcal{F})$, which is called the conditional probability given the event B.

The chain rule

$$P(A \cap B) = P(B)P(A|B) = P(A)P(B|A)$$

is very useful to evaluate the probabilities of complicated events.

The conditional expectatio of $X \in L^1(P)$ conditionally on B with $P(B) > 0$

$$E_P(X|B) := \frac{E_P(X\mathbf{1}_B)}{P(B)} = \int_\Omega X(\omega)P(d\omega|B)$$

Note that these elementary conditional probabilities are defined only when $P(B) > 0$ for the conditioning event. What about conditioning on P-null events ?

From an initial probability P on $(\Omega, \mathcal{F})$ We have built a probability measure $Q \ll P$ by using a random variable $0 \geq Z(\omega) \in L^1(P)$.

This works also in the opposite direction: when when $Q \ll P$ are probability measures on $(\Omega, \mathcal{F})$ there is a random variable $0 \leq Z(\omega) \in L^1(P)$ such that the change of measure formula $Q(A) = E_P(Z\mathbf{1}_A)$ holds.

Theorem 5. *(Radon-Nikodym) On a probability space $(\Omega, \mathcal{F})$ let P, Q probability measures (more in general P could be a σ-finite measure), such that $A \in \mathcal{F}$*

and $P(A) = 0$ imply $Q(A) = 0$. (notation: $Q \overset{\mathcal{F}}{\ll} P$). Then $\exists\ 0 \le Z(\omega) \in L^1(\Omega, \mathcal{F}, P)$ with $E_P(Z) = 1$ such that

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

$Z(\omega)$ is uniquely determined up to P-null sets. We denote

$$Z(\omega) = \frac{dQ}{dP}(\omega)$$

which is called likelihood ratio (finnish: uskottavuus-osamäärä) or Radon-Nikodym derivative

The proof will be given later by using martingales.

We write the change of measure formula as

$$E_Q(X) = \int_\Omega X(\omega)Q(d\omega) = \int_\Omega X(\omega)\,\frac{dQ}{dP}(\omega)\,P(d\omega)$$

Definition 14. *On a probability space $(\Omega, \mathcal{F})$ the probabilities P and P' are singular (notation: $P \perp P'$), when there is $A \in \mathcal{F}$ such that $P(A) = 0$ ja $P'(A) = P'(\Omega) = 1$.*

Exercise 4. *On a probability space $(\Omega, \mathcal{F}, P)$, let $\mathcal{F} = \sigma(X)$ where $X(\omega)$ is a standard Gaussian r.v. with $E(X) = 0, E(X^2) = 1$, and*

$$P(X \in dx) = \frac{1}{\sqrt{2\pi}} \exp\bigl(-\frac{x^2}{2}\bigr)dx$$

Let P' another probability such that

$$P'(X_i \in dx) = \frac{1}{\sqrt{2\pi}} \exp\bigl(-\frac{(x-\mu)^2}{2}\bigr)dx$$

We compute the likelihood ratio

$$Z'(\omega) = \frac{dP'}{dP}(\omega) \quad \text{and} \quad Z(\omega) = \frac{dP}{dP'}(\omega) = \frac{1}{Z'(\omega)}$$

From the R-N theorem it follows that $Z'(\omega)$ is $\sigma(X)$-measurable. There is a Borel-measurable function $z : \mathbb{R} \to \mathbb{R}^+$ such that $Z'(\omega) = z'(X(\omega))$.

For all Borel measurable $f(x) \ge 0$

$$\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) \exp\bigl(-\frac{(x-\mu)^2}{2}\bigr)dx = E_{P'}(f(X)) = E_P(f(X)Z')$$
$$= E_P(f(X)z'(X)) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x)z'(x) \exp\bigl(-\frac{x^2}{2}\bigr)dx$$

which implies

$$z'(x) = \exp\bigl(\mu x - \frac{1}{2}\mu^2\bigr),$$
$$Z'(\omega) = \exp\bigl(\mu X(\omega) - \frac{1}{2}\mu^2\bigr)$$

Since $E_P(Z') = 1$, it follows

$$E_P\bigl(\exp(\mu X)\bigr) = \exp\bigl(\frac{1}{2}\mu^2\bigr)$$

3.1.1 Lebesgue decomposition

Let P, P' probabilities on $(\Omega, \mathcal{F})$.

Then $Q := \frac{1}{2}(P + P')$ is a probabilty measure which satisfies $P \ll Q$ and $P' \ll Q$ on $\mathcal{F}$.

By the R-N theorem (5) the likelihood-ratio processes

$$\zeta(\omega) := \frac{dP}{dQ}(\omega) \text{ ja } \zeta'(\omega) := \frac{dP'}{dQ}(\omega) \ ,$$

do exist, non-negative and $\mathcal{F}$-measurable.

Note that $\forall \omega$

$$\zeta(\omega) + \zeta'(\omega) = \frac{2dP}{d(P+P')}(\omega) + \frac{2dP'}{d(P+P')}(\omega) = 2\,\frac{d(P+P')}{d(P+P')}(\omega) = 2 \ .$$

Since $\zeta(\omega) \geq 0, \zeta'(\omega) \geq 0$ it follows

$$\zeta(\omega) \leq 2, \zeta'(\omega) \leq 2 \quad Q \text{ a.s. }, \quad \text{and} \quad Q\big(\{\omega : \zeta(\omega) = 0\} \cap \{\omega : \zeta'(\omega) = 0\}\big) = 0.$$

We define $\forall \omega \in \Omega$

$$Z(\omega) = \frac{dP}{dP'}(\omega) := \frac{\zeta(\omega)}{\zeta'(\omega)} \quad \text{and } Z'(\omega) = \frac{dP'}{dP}(\omega) := \frac{\zeta'(\omega)}{\zeta(\omega)} = \frac{1}{Z(\omega)}$$

where by convention $0/0$ takes an arbitrary value, for example 0.

For $X \in \mathcal{F}^+$ we have the generalized change of measure formula

$$E_{P'}(X) = E_P(XZ') + E_{P'}(X\mathbf{1}(\zeta = 0))$$

Proof

$$\begin{aligned} E_{P'}(X) &= E_{P'}\big(X\{\mathbf{1}(\zeta > 0) + \mathbf{1}(\zeta = 0)\}\big) = E_Q\big(X\zeta'\mathbf{1}(\zeta > 0)\big) + E_{P'}\big(X\mathbf{1}(\zeta = 0)\big) \\ &= E_Q\left(X\frac{\zeta'}{\zeta}\zeta\mathbf{1}(\zeta > 0)\right) + E_{P'}\big(X\mathbf{1}(\zeta = 0)\big) = E_Q(XZ'\zeta) + E_{P'}\big(X\mathbf{1}(\zeta = 0)\big) \\ &= E_P(XZ') + E_{P'}\big(X\mathbf{1}(\zeta = 0)\big) = E_P(XZ') + E_{P^\perp}(X) \end{aligned}$$

where

$$P^\perp(d\omega) := \mathbf{1}(\zeta(\omega) = 0)P'(d\omega) \ ,$$

Therefore

$$P'(d\omega) = Z'(\omega)P(d\omega) + \mathbf{1}(\zeta(\omega) = 0)P'(d\omega) = Z'(\omega)P(d\omega) + P^\perp(d\omega)$$

P ja $P^\perp$ are singular, since for $A := \{\omega : \zeta(\omega) = 0\}$

$$P(A) = 0 \text{ and } P^\perp(A) = P^\perp(\Omega)$$

Since $P^\perp(\Omega) + E_P(Z') = P'(\zeta = 0) + E_P(Z') = 1$, $P^\perp$ is a probability measure if and only if $P \perp P'$, (equivalently $P^\perp = P'$). Also $E_P(Z') \leq 1$ and $E_P(Z') = 1$ if and only if $P' \ll P$, in such case $P^\perp = 0$.

3.2 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 7. *$X(\omega) \geq 0$ 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 4. *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 $\mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{F}$ 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 6. *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 $t \in \mathbb{R}$,

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

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 5. *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 decomposÃŋng $X(\omega) = (X^+(\omega) - X^-(\omega))$ with $X^\pm = (\pm X, 0)$ the result follows from linearity.

3.3 Conditional expectation as Radon-Nykodim derivative

Let $X \in L^1(\Omega, \mathcal{F}, P)$. We decompose $X(\omega) = X^+(\omega) - X^-(\omega)$ where $x^\pm = (\pm x) \vee 0 \geq 0$, and consider $X^\pm(\omega)$ separately. Without loss of generality, let $X(\omega) = X^+(\omega) \geq 0$.

We define a finite positive measure on $(\Omega, \mathcal{F})$:

$$\mu_X(A) = E_P(X\mathbf{1}_A) \quad \forall A \in \mathcal{F}$$

Note that $\mu_X(A) = 0$ when $P(A) = 0$, so that $\mu_X \overset{\ll}{\mathcal{F}} P$ (μ_X is dominated by P σ-algebra $\mathcal{F}$), and $X(\omega) = \frac{d\mu_X}{dP}(\omega)$ is the corresponding Radon-Nikodymin derivative.

Let $\mathcal{G} \subseteq \mathcal{F}$ a sub-σ-algebra. Obviously $\mu_X \overset{\ll}{\mathcal{G}} P$, μ_X is dominated by P on the σ-algebra $\mathcal{G}$. $\mu \ll P$ σ-algebrassa $\mathcal{G}$.

By the Radon-Nikodymin theorem a R-N derivative

$$Y(\omega) := \frac{d\mu_X|_{\mathcal{G}}}{dP|_{\mathcal{G}}}(\omega)$$

exists and it is an element of $L^1(\Omega, \mathcal{G}, P)$ which satisfies the change of measure formula

$$E_P(X\mathbf{1}_A) = \mu_X(A) = E_P(Y\mathbf{1}_A) \quad \forall A \in \mathcal{G}$$

by Kolmogorov's definition of conditional expectation $Y(\omega) = E_P(X|\mathcal{G})(\omega)$ P a.s.

Remark 5. *The existence of the conditional expectation of $X \in L^1(P)$ follows by RN-theorem. We have not proved yet RN-theorem but we will, using a martingale argument where we need conditional expectations. In order to avoid a circular proof, we showed that the conditional expectations by using approximating $L^2(P)$-projections.*

3.4 What can we say when $E_P(|X|) = \infty$?

Let $0 \le X(\omega) \in L^0(\Omega, \mathcal{F}, P)$ with $E_P(X) = \infty$. Also in this case we can truncate, take approximations in $L^2(P)$ and apply the monotone convergence theorem (which does not require integrability), to show that the conditional expectation

$$Y(\omega) = E_P(X|\mathcal{G})(\omega) \in [0, +\infty]$$

which is $\mathcal{G}$-measurable and satisfies $\forall A \in \mathcal{G}$.

$$E_P(X\mathbf{1}_A) = E_P(Y\mathbf{1}_A) \in [0, +\infty]$$

Note that $Y(\omega)$ could also take value $+\infty$, and in any case $E_P(Y) = E_P(X) = \infty$.

Consider the case $X(\omega) = (X(\omega)^+ - X(\omega)^-)$ with $E_P(|X|) = \infty$. Then the conditional expectation

$$E_P(X|\mathcal{G})(\omega) := E_P(X^+|\mathcal{G})(\omega) - E_P(X^+|\mathcal{G})(\omega) \in [-\infty, +\infty]$$

is well defined on the complement of

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

When $P(U) = 0$ the conditional expectation is well defined almost everywhere.

3.5 Regular conditional probability and kernels

The conditional probability of the event $A \in \mathcal{F}$ conditionally on the sub-σ-algebra $\mathcal{G}$ is defined P-almost surely as

$$P\big(A\big|\mathcal{G}\big)(\omega) = E_P\big(\mathbf{1}_A\big|\mathcal{G}\big)(\omega)$$

Since the conditional expectation is a non-negative operator, it follows that $P\big(A\big|\mathcal{G}\big)(\omega) \in [0,1]$ P-a.s.

Can we say that for P-almost all ω, the map $A \mapsto P\big(A\big|\mathcal{G}\big)(\omega) \in [0,1]$ is a probability measure on $(\Omega, \mathcal{F})$?

Let $\{A_n\} \subseteq \mathcal{F}$ with $A_n \downarrow \emptyset$. By the monotone convergence theorem conditional expectation that there is a set N with $P(N) = 0$ such that

$$P(A_n|\mathcal{G})(\omega) \downarrow 0 \quad \forall \omega \in N^c \tag{3.1}$$

The event N may depend on the sequence $\{A_n\}$, the set of such sequences ios not countable, it is not guaranteed that outside a P-null set (3.1) holds simultaneously for all sequences of events with $A_n \downarrow \emptyset$.

The conditional probabilities defined above are not always σ-additive.

Definition 15. *Let $(\Omega, \mathcal{F})$ and $(\widetilde{\Omega}, \widetilde{\mathcal{F}})$ probability spaces.*

A map $(A, \widetilde{\omega}) \mapsto K(A, \widetilde{\omega}) \in [0,1]$ is a probability kernel when

- *For every fixed $\widetilde{\omega} \in \widetilde{\Omega}$ the map $A \mapsto K(A, \widetilde{\omega})$ is a probability measure on $(\Omega, \mathcal{F})$*
- *For fixed $A \in \mathcal{F}$, the map $\widetilde{\omega} \mapsto K(A, \widetilde{\omega})$ is $\widetilde{\mathcal{F}}$-measurable.*

For the regular conditional probability consider $\widetilde{\Omega} = \Omega$ and $\widetilde{\mathcal{F}} = \mathcal{G} \subseteq \mathcal{F}$.

Definition 16. *The conditional probability has regular version when there is a $(\Omega, \mathcal{G})$ measurable kernel $K(A, \omega)$ on $(\Omega, \mathcal{F})$ such that for all $A \in \mathcal{F}$*

$$P(A|\mathcal{G})(\omega) = K(A, \omega) \qquad P \ a.s$$

Remark 6. *When the conditional probability $P(A|\mathcal{G})(\omega)$ has a regular version $K(A, \omega)$ we have*

$$E(X|\mathcal{G})(\omega) = \int_\Omega X(\omega') K(d\omega'|\omega)$$

Definition 17. *A probability space $(\Omega, \mathcal{F})$ is Borel if there is an 1-1 (injective) function $f : (\Omega, \mathcal{F}) \to [0,1], \mathcal{B}([0,1])$ such that on the image, the inverse f^{-1} is also measurable.*

Here $\mathcal{B}([0,1])$ is the Borel σ-algebra generated by the open sets.

Theorem 7. *Let $(\Omega, \mathcal{F}, P)$ a probability space, $\mathcal{G} \subseteq \mathcal{F}$ a sub-σ-algebra And $X(\omega)$ a random variable taking values in a Borel space $(\Omega', \mathcal{F}')$. Then the conditional probabilities*

$$P(X \in A'|\mathcal{G})(\omega), \qquad A' \in \mathcal{F}'$$

have regular version.

For a proof, see Kallenberg 'Foundations of Modern Probability', Thm 6.3, 6.4.

Remark 7. *A separable topological space (which contains a dense countable set) equipped with its Borel σ-algebra is a Borel space. In particular the euclidean space $\mathbb{R}^d$ is separable, and also the space $C([0,1],\mathbb{R}^d)$ of continous functions where the Brownian motion lives, and we can always work with the regular version of the conditional probability.*

3.6 Computation of conditional expectation under P-independence

Proposition 5. *On a probability space $(\Omega,\mathcal{F})$, let $\mathcal{G}\subseteq\mathcal{F}$ a sub-σ-algebra, $Y(\omega)$ $\mathcal{G}$-measurable r.v. with values in the measurable space $(S,\mathcal{S})$. Let also $X(\omega)\in(\widetilde{S},\widetilde{\mathcal{S}})$ P-independent from $\mathcal{G}$.*

Let $f:(\widetilde{S}\times S)\to\mathbb{R}^+$ a non-negative Borel-measurable function.

The conditional expectation has integral-representation

$$E_P\big(f(X,Y)\big|\mathcal{G}\big)(\omega)=E_P\big(f(X,y)\big)\Big|_{y=Y(\omega)}=\int_{\widetilde{S}}f(x,Y(\omega))P_X(dx) \tag{3.2}$$

with $P_X(B)=P(\{\omega:X(\omega)\in B\})$.

Proof: When $f(x,y)=f_1(x)f_2(y)$, $\forall G\in\mathcal{G}$ from P-independence follows

$$\begin{aligned}
&E_P\big(f_1(X)f_2(Y)\mathbf{1}_G\big)=E_P\big(f_1(X)\big)E_P(f_2(Y)\mathbf{1}_G)\\
&=E_P\Big(f_2(Y)E_P(f_1(X))\mathbf{1}_G\Big)=\int_\Omega\Big(\int_\Omega f_1(X(\omega'))f_2(\omega)P(d\omega')\Big)P(d\omega)\\
&=\int_\Omega E_P(f(X,y))\Big|_{y=Y(\omega)}\mathbf{1}_G(\omega)P(d\omega)
\end{aligned}$$

More in general by definition of jointly measurable functions we find a sequence

$$0\le f^{(n)}(x,y)=\sum_{k=1}^n f_1^{(n,k)}(x)f_2^{(n,k)}(y)\uparrow f(x,y),\qquad\text{as } n\to\infty$$

and the results follows by the monotone convergence theorem.

3.7 Computing conditional expectations by changing the measure: abstract Bayes' formula

Lemma 8. *The conditional expectation is a self-adjoint operator, meaning that for $X\in L^1(\Omega,\mathcal{F},P)$, and $\mathcal{G}\subseteq\mathcal{F}$ is a sub-σ-algebra, $\forall A\in\mathcal{F}$*

$$E_P\big(X\,E_P(\mathbf{1}_A|\mathcal{G})\big)=E_P\big(E_P(X|\mathcal{G})\,E_P(\mathbf{1}_A|\mathcal{G})\big)=E_P\big(E_P(X|\mathcal{G})\,\mathbf{1}_A\big)$$

Proof: straight from the definitions.

We have shown two cases where we are able to compute conditional expectations: when the σ-algebra $\mathcal{G}$ is generated by a countable set of atoms, or under independence using proposition (5).

When independence does not hold under the original measure P, is often possible to work with another simpler measure under which independence holds.

The next formula is a change of measure inside the conditional expectation.

Theorem 8. *(Abstract Bayes' formula). On the probability space $(\Omega,\mathcal{F})$, let $\mathcal{G}\subseteq\mathcal{F}$ and $P\overset{\mathcal{F}}{\ll}Q$ probability measures $Q(A)=0\Longrightarrow P(A)=0$ kun $A\in\mathcal{F}$.*

Radon-Nikodym it follwos that there is a R-N-derivative, which means a random variable

$$0\le Z(\omega):=\frac{dP}{dQ}(\omega)\ \in L^1(\Omega,\mathcal{F},Q)$$

for which the change of measure formula for the expectation holds:

$$E_P(X)=E_Q(XZ)\qquad\forall X\in L^1(\Omega,\mathcal{F},P)$$

Then the conditional expectation satisfies Bayes formula:

$$E_P\big(X\big|\mathcal{G}\big)(\omega)=\frac{E_Q\big(XZ\big|\mathcal{G}\big)(\omega)}{E_Q\big(Z\big|\mathcal{G}\big)(\omega)}\qquad\in L^1(\Omega,\mathcal{G},P)$$

Proof. Let $G\in\mathcal{G}$. From the change of measure formula and the deifinition of conditional expctation it follows

$$E_P\big(X\mathbf{1}_G\big)=E_Q\big(ZX\mathbf{1}_G\big)=E_Q\big(E_Q(ZX\mathbf{1}_G|\mathcal{G})\big)=E_Q\big(E_Q(ZX|\mathcal{G})\mathbf{1_G}\big)$$
$$=E_Q\left(\frac{E_Q(Z|\mathcal{G})}{E_Q(Z|\mathcal{G})}E_Q(ZX|\mathcal{G})\mathbf{1}_G\right)=E_Q\left(Z\frac{E_Q(ZX|\mathcal{G})}{E_Q(Z|\mathcal{G})}\mathbf{1}_G\right)=E_P\left(\frac{E_Q(ZX|\mathcal{G})}{E_Q(Z|\mathcal{G})}\mathbf{1}_G\right)\quad\square$$

Exercise 5. *(Bayes formula for densities) On a probability space $(\Omega,\mathcal{F})$, let and $X(\omega)\in\mathbb{R}^d, Y(\omega)\in\mathbb{R}^m$ random variables, let $\mathcal{F}=\sigma(X,Y)$ and $\mathcal{G}=\sigma(Y)$.*

Let $P\overset{\mathcal{F}}{\ll}Q$ probability measures such that $X\overset{Q}{\perp\!\!\!\perp}Y$ with RN-derivative

$$0\le Z(\omega):=z(X(\omega),Y(\omega))=\frac{dP}{dQ}(\omega)\ \in L^1(\Omega,\mathcal{F},Q)$$

where $z(x,y)\ge 0$ is Borel measurable.

Let $f(x,y)\ge 0$ Borel-measurable. From the abstract Bayes formula

$$\begin{aligned}
E_P\big(f(X,Y)\big|\mathcal{G}\big)(\omega)&=\frac{E_Q\big(f(X,Y)Z\big|\mathcal{G}\big)(\omega)}{E_Q\big(Z\big|\mathcal{G}\big)(\omega)}\\
&=\frac{\int_\Omega f(X(\widetilde{\omega}),Y(\omega))\ z(X(\widetilde{\omega}),Y(\omega))P(d\widetilde{\omega})}{\int_\Omega z(X(\widetilde{\omega}),Y(\omega))P(d\widetilde{\omega})}\\
&=\int_\Omega f(X(\widetilde{\omega}),Y(\omega))K(\omega,d\widetilde{\omega})\qquad\text{where}\\
K(\omega,d\widetilde{\omega})&=\frac{z(X(\widetilde{\omega}),Y(\omega))}{\int_\Omega\ z(X(\omega'),Y(\omega))P(d\omega')}P(d\widetilde{\omega})
\end{aligned}$$

is the regular version of the conditional probability. We can also integrate directly on the space $\mathbb{R}^d$ *where* $X(\omega)$ *takes values:*

$$E_P\big(f(X,Y)\big|\mathcal{G}\big)(\omega) = \frac{\int_{\mathbb{R}^d} f(x,Y(\omega))z(x,Y(\omega))P_X(dx)}{\int_{\mathbb{R}^d} z(x,Y(\omega))P_X(dx)} = \int_{\mathbb{R}^d} f(x,Y(\omega))k(Y(\omega),dx)$$

where

$$k(y,dx) = \frac{z(x,y)}{\int_{\mathbb{R}^d} z(x',y)P_X(dx')} P_X(dx)$$

When the distribution of the vector (X,Y) *has density with repect to the* $(d+m)$*-dimensional Lebesgue measure,*

$$P(X \in dx, Y \in dy) = p_{X,Y}(x,y)dxdy,$$

from Fubini's theorem it follows that also the marginal distributions P_X *and* P_Y *have densities*

$$P(X \in dx) = p_X(x)dx = \int_{\mathbb{R}^m} p_{X,Y}(x,y)dy$$
$$P(Y \in dy) = p_Y(y)dy = \int_{\mathbb{R}^d} p_{X,Y}(x,y)dx$$

Taking as probability space $\Omega = \mathbb{R}^d \times \mathbb{R}^m$ *and consider the probability measures*

$$Q_{X,Y}(dx,dy) := (P_X \otimes P_Y)(dx,dy) = p_X(x)p_Y(y)dxdy,\ P_{X,Y}(dx,dy) = p_{X,Y}(y)dxdy$$

From the assumption $P_{X,Y} \ll (P_X \otimes P_Y)$*, it follows that the Radon-Nykodim derivative is given by*

$$\frac{dP_{X,Y}}{dQ_{X,Y}}(x,y) = \frac{dP_{X,Y}}{d(P_X \otimes P_Y)}(x,y) = z(x,y) = \frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}$$

We write the regular transition probability in terms of the densities

$$k(y,dx) = \frac{z(x,y)}{\int_{\mathbb{R}^d} z(x',y)P_X(dx')} P_X(dx) = \frac{p_{X,Y}(x,y)}{p_Y(y)} dx = p_{X|Y}(x|y)dx$$

We have obtained the 'classical' Bayes' formula

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)} = \frac{p_X(x)p_{Y|X}(y|x)}{p_Y(y)} \quad .$$

3.8 Conditioning on P-null events : a warning

Let $X(\omega), Y(\omega)$ independent standard Gaussian, with $E_P(X) = E_P(Y) = 0$, $E_P(X^2) = E_P(Y^2) = 1$.

Let

$$W(\omega) = (X(\omega) - Y(\omega)), \quad Z(\omega) = \mathbf{1}(Y(\omega) \neq 0)\, \frac{X(\omega)}{Y(\omega)}$$

and $N := \{\omega :\ Y(\omega) = 0\}$.

Clearly $P(N) = 0$ and

$$N^c \cap \{\omega : X(\omega) = Y(\omega)\} = N^c \cap \{\omega : W(\omega) = 0\} = N^c \cap \{\omega : Z(\omega) = 1\}$$

Let $f : \mathbb{R} \to \mathbb{R}^+$ a non-negative Borel-measurable function.

$$i) \qquad E_P(f(X)|\{X = Y\}) = \frac{\iint\limits_{\mathbb{R}\times\mathbb{R}} f(x)\delta_0(x-y)p_X(x)p_Y(y)dxdy}{\iint\limits_{\mathbb{R}\times\mathbb{R}} \delta_0(x-y)p_X(x)p_Y(y)dxdy}$$

$$ii) \qquad E_P(f(X)|W = 0) = \int_{\mathbb{R}} f(x)p_{X|W}(x|0)dx$$

$$iii) \qquad E_P(f(X)|Z = 1) = \int_{\mathbb{R}} f(x)p_{X|Z}(x|1)dx$$

are not all equal !

Exercise 6. *Show that* $i) = ii) \neq iii)$.

A set of measure zero can be represented by using different random variables. The corresponding pointwise values of the conditional expectation may differ. This is not in contradiction with the theory, since we can always change the value of the conditional expectation on a set of probability zero.

Chapter 4

Martingale theory

4.1 Martingales

Definition 18. *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 19. *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 20. *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 21. *Let* $\tau(\omega)$ *an* $(\mathcal{F}_t)$*-stopping time, the stopped* σ*-algebra is defined as*

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

Exercise 7.

- *Check that* $\mathcal{F}_\tau$ *is a* σ*-algebra.*
- *If* $0 \leq \sigma(\omega) \leq \tau(\omega)$ $\forall \omega$ *where* σ, τ *are* $(\mathcal{F}_t)$*-stopping times then* $\mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$

  Proof of $\mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$:
  $A \in \mathcal{F}_\sigma \iff A \cap \{\sigma \leq t\} \in \mathcal{F}_\tau$, $\forall t \geq 0$,
  Also $\{\tau \leq t\} \in \mathcal{F}_t$, which implies

  $$A \cap \{\tau \leq t\} A \cap \{\sigma \leq t\} \cap \{\tau \leq t\} \in \mathcal{F}_t$$

Definition 22. *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* $s \leq t$

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

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

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

Exercise 8. *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 9. *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 10. *Let* $(B_t(\omega) : t \geq 0)$ *a Brownian motion. Consider the filtration* $\mathbb{F} = \{\mathcal{F}_t^B : t \geq 0\}$ *generated by* B *with* $\mathcal{F}_t^B = \sigma(B_s : 0 \leq s \leq t)$

Then $(B_t : t \geq 0)$ *and* $(B_t^2 - t : t \geq 0)$ *are* $\mathbb{F}$*-martingales.*

Exercise 11. *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 23. *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 6. *Let* (X_t) *be a martingale and* (Y_t) *a predictable process in the discrete-time filtration* $\mathbb{F} = (\mathcal{F}_t : t \in \mathbb{N})$*. 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 $p, q \in [1, +\infty]$, $p^{-1} + q^{-1} = 1$.

Corollary 6. *Let* $(M_t : t \in \mathbb{N})$ *an* $\mathbb{F}$*-martingale, and* $\tau(\omega) \in \mathbb{N}$ *a* $\mathbb{F}$*-stopping time. Then the stopped process*

$$M_t^\tau(\omega) = M_{t\wedge\tau}(\omega) = M_0 + \sum_{s=1} \mathbf{1}(\tau(\omega) \geq s)\big(M_s(\omega) - M_{s-1}(\omega)\big)$$

is a $\mathbb{F}$*-martingale.*

Proof: since $\mathbf{1}(\tau(\omega) \geq s) = \mathbf{1}(\tau(\omega) > s-1) \in \mathcal{F}_{s-1}$, we see that $M_{t\wedge\tau}$ is the martingale transform of a bounded $\mathbb{F}$-predictable integrand.

4.1.1 Martingale convergence

Theorem 9. *(Doob's forward convergence) Let* $(X_t : t \in \mathbb{N})$ *a supermartingale with*

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

Notation: $x^\pm = \max(\pm x, 0)$.

Then

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

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

Notes : although $X_\infty(\omega) \in L^1(\Omega)$ we don't have necessarily convergence in $L^1(P)$ sense. Joseph Leo Doob(1910-2004) American probabilist, is the father of martingale theory.

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

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

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

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

are $\mathcal{F}_t$-measurable 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 7. *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 8. *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 $(-X_t)$

4.2 Uniform integrability

Definition 24. *A collection of random variables $\mathcal{C} \subseteq L^1(\Omega, \mathcal{F}, P)$. is* **uniformly integrable (UI)** *with respect to P when*

$$\lim_{K\to\infty} \sup_{X\in\mathcal{C}} E_P\big(|X|\mathbf{1}(|X| > K)\big) = \int_{\{\omega:|X(\omega)|>K\}} |X(\omega)|P(d\omega) \longrightarrow 0 \textit{ kun } K \to \infty$$

Lemma 9. *A finite collection $\mathcal{C} = \{X_1, X_2, \dots, X_M\} \subset L^1(\Omega, \mathcal{F}, P), M \in \mathbb{N}$ is uniformly integrable. Proof: From the monotone convergence theorem it follows that a single random variable $X \in L^1(P)$ is uniformly integrable. A finite set $\{X_1, \dots, X_M\} \subset L^1(P)$ is uniformly integrable since*

$$\max_{k=1,\dots M} |X_k(\omega)| \le \sum_{k=1}^{N} |X_k(\omega)| \in L^1(P)$$

Remark 8. *To show that a sequence $\{X_n\}_{n\in\mathbb{N}}$ is uniformly integrable it is enough to find $Y \in L^1(P)$ such that*

$$\sup_{n\in\mathbb{N}} |X_n(\omega)| \le Y(\omega)$$

Lemma 10. *$X \in L^1(\Omega, \mathcal{F}, P)$, if and only if $\forall\varepsilon > 0\ \exists\delta$, such that $\forall A \in \mathcal{F}$,*

$$P(A) < \delta \Longrightarrow E_P\big(|X|\mathbf{1}_A\big) < \varepsilon$$

Proof, sufficiency: $\forall\omega$,

$$Y^{(K)}(\omega) := |X(\omega)|\mathbf{1}(|X(\omega)| \le K) \uparrow |X(\omega)|$$

and by (9)

$$E_P(|X|) - E_P(Y^{(K)}) = \int_{\{\omega:|X(\omega)|>K\}} |X(\omega)|P(d\omega) < \varepsilon$$

for K large enough so that $P(\{\omega : |X(\omega)| > K\}) < \delta$. It follows that

$$E_P(|X|) \le E_P(Y^{(K)}) + \varepsilon \le K + \varepsilon < \infty$$

Proof of necessity, by contradiction: otherwise there would be $\varepsilon > 0$ and a sequence of events $\{A_n : n \in \mathbb{N}\} \subseteq \mathcal{F}$ such that

$$P(A_n) < 2^{-n} \Longrightarrow E_P\big(|X|\mathbf{1}_{A_n}\big) \geq \varepsilon > 0$$

Denote $A = \limsup_n A_n$. Since

$$\sum_n P(A_n) \leq \sum_n 2^{-n} = 1 < \infty$$

$P(A) = 0$ by the Borel Cantelli lemma.

Let $B_n = \bigcup_{k \geq n} A_k$. By definition $A_n \subseteq B_n \downarrow A$, which means

$$|X(\omega)|\mathbf{1}_{A_n}(\omega) \leq |X(\omega)|\mathbf{1}_{B_n}(\omega) \downarrow |X(\omega)|\mathbf{1}_A(\omega) \quad \forall \omega$$

where the random variables above are integrable since $X \in L^1(P)$. It follows from the sufficiency part of the proof that

$$0 < \varepsilon \leq E_P(|X|\mathbf{1}_{A_n}) \leq E_P(|X|\mathbf{1}_{B_n}) \downarrow E_P(|X|\mathbf{1}_A) = 0$$

since $P(A) = 0$ □

Theorem 10. *(Characterization of convergence in $L^1(P)$) Consider $\{X_n : n \in \mathbb{N}\} \subseteq L^1(\Omega, \mathcal{F}, P)$, $n \in \mathbb{N}$ ja $X \in L^0(\Omega, \mathcal{F})$.*

$X_n \stackrel{P}{\to} X$ and $\{X_n : n \in \mathbb{N}\}$ is uniformly integrable,

if and only if $X_n \stackrel{L^1}{\to} X \in L^1(P)$,

Proof: When $X_n \stackrel{P}{\to} X$ there is a subsequence $n(k)$ such that $X_{n(k)}(\omega) \to X(\omega)$ P-a.s.

By Fatou's lemma

$$E_P(|X|) = E_P(\liminf_k |X_{n(k)}|) \leq \liminf_k E_P(|X_{n(k)}|) < \infty$$

since the random variables $\{X_n : n \in \mathbb{N}\}$ are uniformly integrable, which implies $X \in L^1(P)$.

Let $K \in \mathbb{N}$ and define the truncation

$$g^{(K)}(x) = \begin{cases} K & \text{kun } x > K \\ x & \text{kun } |x| \leq K \\ -K & \text{kun } x < -K \end{cases}$$

and the truncated random varianbles $X_n^{(K)}(\omega) = g^{(K)}(X_n(\omega))$, $X^{(K)}(\omega) = g^{(K)}(X(\omega))$. By lemma (9) and the uniform integrability assumptions it follows that $\forall \varepsilon > 0$ there is K such that

$$E_P(|X - X^{(K)}|) < \varepsilon \quad \text{and} \quad E_P(|X_n - X_n^{(K)}|) < \varepsilon \quad \forall n,$$

since

$$\sup_n E_P(|X_n - X_n^{(K)}|) = \sup_n \left\{ \int_{\{\omega:|X_n(\omega)|>K\}} |X(\omega)|P(d\omega) - KP\big(|X_n| > K\big) \right\}$$

$$\leq \sup_n \int_{\{\omega:|X_n(\omega)|>K\}} |X(\omega)|P(d\omega) \longrightarrow 0 \text{ for } K \to \infty .$$

We show first that

$$E_P(|X^{(K)} - X_n^{(K)}|) \to 0 \text{ for } K \to \infty.$$

Since $|g^{(K)}(x) - g^{(K)}(y)| < |x-y|$, it follows $X_n^{(K)} \stackrel{P}{\to} X^{(K)}$. There is $\bar{n}$ such that

$$P\big(|X_n^{(K)} - X^{(K)}| > \frac{\varepsilon}{3}\big) < \frac{\varepsilon}{3K} \text{ kun } n \geq \bar{n},$$

which implies

$$E_P(|X_n^{(K)} - X^{(K)}|) = E_P\left(|X_n^{(K)} - X^{(K)}|\, \mathbf{1}\left(|X_n^{(K)} - X^{(K)}| > \frac{\varepsilon}{3}\right)\right)$$
$$+E_P\left(|X_n^{(K)} - X^{(K)}|\, \mathbf{1}\left(|X_n^{(K)} - X^{(K)}| \leq \frac{\varepsilon}{3}\right)\right)$$
$$\leq 2K\; P\left(|X_n^{(K)} - X^{(K)}| > \frac{\varepsilon}{3}\right) + \frac{\varepsilon}{3} \leq 2K\frac{\varepsilon}{3K} + \frac{\varepsilon}{3} = \varepsilon \qquad \text{when } n \geq \bar{n}$$

By the triangle inequality, when $n \geq \bar{n}$

$$E_P(|X_n - X|) \leq E_P(|X_n - X_n^{(K)}|) + E_P(|X_n^{(K)} - X^{(K)}|) + E_P(|X^{(K)} - X|) \leq 3\varepsilon$$

In the other direction, when $E_P(|X_n - X|) \to 0$, convergence in probability $X_n \stackrel{P}{\to} X$ follows by Chebychev inequality.

Let $\varepsilon > 0$, and $N \in \mathbb{N}$ such that

$$E_P(|X - X_n|) < \frac{\varepsilon}{2} \qquad \text{when } n \geq N.$$

From lemma 10 $\exists\, \delta > 0$ jolla $\forall A \in \mathcal{F}$ jolla $P(A) < \delta$ it follows

$$\max_{n \leq N} E_P(|X_n|\mathbf{1_A}) < \varepsilon \qquad \text{ja} \qquad E_P(|X|\mathbf{1_A}) < \frac{\varepsilon}{2} \quad .$$

Since $E_P(|X_n|) \leq E_P(|X|) + E_P(|X_n - X|)$ where by assumption $E_P(|X_n - X|) \to 0$, there is $K > 0$ such that

$$\sup_n E_P(|X_n|) < K\delta < \infty\ .$$

By Chebychev inequality

$$P(|X_n| > K) \leq K^{-1} E_P(|X_n|) < \delta \quad \forall n \in \mathbb{N}\ .$$

When $n \geq N$,

$$E_P\big(|X_n|\mathbf{1}(|X_n| > K)\big) \leq E_P\big(|X|\mathbf{1}(|X_n| > K)\big) + E(|X - X_n|) < \varepsilon$$

When $n \leq N$ also $P(|X_n| > K) < \delta$ and

$$E_P\big(|X_n|\mathbf{1}(|X_n| > K)\big) < \varepsilon$$

which shows that $\{X_n : n \in \mathbb{N}\}$ is uniformly integrable $\square$

Uniform integrability is a compactness condition in $L^1(P)$ when we replace the norm topology by the so called weak-star topology:

Proposition 7. *(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 $\forall A \in \mathcal{F}$*

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

We skip this proof We proof $\Longrightarrow$, for the other implication see Kallenberg Foundations of Modern Probability Lemma 4.13.

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 4.2) 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.

Remark 9. *The stronger convergence of the subsequence in $L^1(P)$ does not follow.*

It is good to know the following characterization of uniform integrability:

Proposition 8. *$\mathcal{C} \subseteq L^1(P)$ is uniformly integrable if and only if*

$$\sup_{X\in\mathcal{C}} E_P(|X|) < \infty \quad \text{and} \quad \forall \varepsilon > 0 \quad \exists\, \delta : P(A) < \delta \implies \sup_{X\in\mathcal{C}} E_P\big(|X|\mathbf{1}_A\big) < \varepsilon$$

Proof. exercise

Remark 10. *When $\mathcal{C} \subseteq L^1(P)$ is uniformly integrable, for K large enough*

$$\sup_{X\in\mathcal{C}} E_P(|X|) < K + \sup_{X\in\mathcal{C}} E(|X|\mathbf{1}(|X| > K)) < K + \varepsilon < \infty$$

Nevertheless the unit ball $B_1 = \{X \in L^1(P) : E_P(|X|) \le 1\}$ is not uniformly integrable: let $\{A_n : n \in \mathbb{N}\} \subseteq \mathcal{F}$ such that $P(A_n) = n^{-1}$, and $X_n(\omega) = n\, \mathbf{1}_{A_n}(\omega)$. Clearlt $X_n \in B_1$ $\forall n$, and for all $K > 0$

$$\sup_n E_P\big(|X_n|\mathbf{1}(|X_n| > K)\big) = \sup_{n>K} E_P(|X_n|) = 1$$

However we have the following criteria:

Lemma 11. *Let $\mathcal{C} \subset L^p(\Omega)$ for some $p > 1$, with*

$$\sup_{X\in\mathcal{C}} E(|X|^p) < \infty$$

Then $\mathcal{C}$ is uniformly integrable.

Proof. Recall that $L^p(\Omega, \mathcal{F}, P) \subset L^1(\Omega, \mathcal{F}, P)$ for $p > 1$

$$E(|X|^p) \ge K^{p-1} E\big(|X|\mathbf{1}(|X| > K)\big) \implies$$
$$\sup_{X\in\mathcal{C}} E\big(|X|\mathbf{1}(|X| > K)\big) \le K^{1-p} \sup_{X\in\mathcal{C}} E(X^p) \quad \longrightarrow 0, \quad \text{as } K \longrightarrow \infty$$

Application: taking a derivative inside the expectation

Proposition 9. *On a probability space $(\Omega, \mathcal{F}, P)$ consider an uniformly integrable family of random variable $\{Y(t,\omega) : t \in [a,b]\} \subseteq L^1(\Omega, \mathcal{F}, P)$, with $a < b \in \mathbb{R}$. We also assume that*

- *For all $\omega \in \Omega$, the map $t \mapsto Y(t,\omega)$ is continuous*

It follow that:

1. *the map $t \mapsto E_P(Y(t))$ is continuous.*
2. *Let*

$$X(t,\omega) := \int_a^t Y(s,\omega)ds, \ t \in [a,b].$$

Then at all $t \in (a,b)$ the derivative exists

$$\frac{d}{dt} E_P\big(X(t)\big) = E_P\big(Y(t)\big) = E_P\left(\frac{d}{dt} X(t)\right)$$

and it is continous.

Proof. From the continuity assumption $\lim_{s\to t} Y_s(\omega) = Y_t(\omega)$ and by uniform integrability it follows

$$|E_P(Y_t) - E_P(Y_s)| \le E_P|Y_t - Y_s| \to 0 \quad \text{kun} \;\; s \to t.$$

Moreover

$$\sup_{t\in[a,b]} E_P\big(|Y_t|\big) < +\infty$$

and $|Y(t,\omega)| \in L^1\big([a,b]\times\Omega, \mathcal{B}([a,b])\otimes\mathcal{F}, dt\otimes P(d\omega)\big)$. By Fubini's theorem

$$E_P(X_t) = E_P\left(\int_a^t Y(s)ds\right) = \int_{[a,b]\times\Omega} Y(s,\omega)\, ds\otimes P(d\omega) = \int_a^t E_P(Y(s))ds$$

and since $t \mapsto E_P(Y(t))$ is continuous, by the mid-value theorem of analysis

$$\lim_{\Delta\to 0} \Delta^{-1}\big\{E_P(X_{t+\Delta}) - E_P(X_t)\big\}$$
$$= \lim_{\Delta\to 0} \Delta^{-1}\int_t^{t+\Delta} E_P(Y(s))ds = E_P(Y(t)) \quad \square$$

4.3 UI martingales

Lemma 12. *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

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

Proposition 10.

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

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

- *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 $K \geq 0$

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

  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 $t \geq 0$ 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 $T \geq t$,

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

- 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 (12)□.

4.3.1 Backward convergence of martingales

Definition 25. *A backward filtration is an increasing family of σ-algebrae* $(\mathcal{F}_t : t \in T)$ *where* $T = -\mathbb{N}, -\mathbb{R}, -\mathbb{N}\cup\{-\infty\} - \mathbb{R}\cup\{-\infty\}$. *For* $0 \geq t \geq u$

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

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

We consider a (sub,super)-martingale with respect to the backward filtration $(\mathcal{F}_t)_{t\leq 0}$ is an adapted and integrable process $(X_t : t \leq 0) \subseteq L^1(P)$ which satisfies the martingale property: for $0 \geq t \geq u$

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

(respectively $\geq, \leq$)

Theorem 11. *(Doob's martingale backward convergence) Let* $(X_t : -t \in \mathbb{N})$ *a be supermartingale in the backward filtration* $\mathbb{F} = (\mathcal{F}_t : t \in -\mathbb{N})$.

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 martingale in the backward filtration the assumption (3) holds automatically, $(X_t = E(X_0|\mathcal{F}_t), t \in -\mathbb{N})$ is uniformly integrable and*

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

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

Proof We copy the proof of the forward convergence theorem, where we play the same supermartingale game in the shifted time interval $\{t, t+1, \dots, -2, -1, 0\}$, with $t \in (-\mathbb{N})$. The profit given by the martingale transform

$$Y_s = (C \cdot X)_s = \begin{cases} 0 & \text{for } s \leq t \\ \sum_{r=t+1}^{s} C_r(X_r - X_{r-}) & \text{for } t < s \leq 0 \end{cases}$$

where $C_r(\omega) \in \{0,1\}$ is $\mathbb{F}$-predictable. It follows that $(Y_s : s \in -\mathbb{N})$ is a supermartingale as well, and

$$0 = E(Y_t) \geq E(Y_0) \geq E(U_{a,b}([t,0])(b-a) - (X_0 - a)^-$$

where $U_{a,b}([t,0])$ the number of upcrossing of $(X_s(\omega))$ in the interval $[t,0]$.

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

Since $U_{[a,b]}([t,0]) \uparrow U_{a,b}((-\infty,0])$ as $t \downarrow (-\infty)$, by monotone convergence theorem $E_P(U_{[a,b]}([t,0])) < \infty$, which implies $U_{[a,b]}([t,0]) < \infty$ P a.s. Since this holds for all $a < b \in \mathbb{Q}$, it follows as in the forward theorem that

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

When X_t is martingale by Fatou lemma

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

In the supermartingale case, we have only

$$E(|X_\infty|) = E(\liminf_t |X_t|) \leq \liminf_t E(|X_t|) = \liminf_t \{E(X_t^+) + E(X_t^-)\}$$

From the supermartingale property

$$X_t \geq E(X_0|\mathcal{F}_t) \quad t \leq 0$$

it follows

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

which implies $X_{-\infty}(\omega) > -\infty$ P-a.s. Since we dont' get for free an upper bound for $E(X_t^+)$, we need to assume (3).

Finally let $A \in \mathcal{F}_{-\infty} \subseteq \mathcal{F}_{-t}$ $\forall t \leq 0$. 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 expectation 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 $X_{-\infty} = E_P(X_t|\mathcal{F}_{-\infty})$.

Strong law of large numbers by martingale backward convergence

Lemma 13. *(Kolmogorov $0-1$ law) On a probability space $(\Omega, \mathcal{F}, P)$ consider a sequence of P-independent σ-algebrae $(\mathcal{G}_n : n \in \mathbb{N})$, $\mathcal{G}_n \subseteq \mathcal{F}$.*

This means that $\forall d \in \mathbb{N}$, $A_1 \in \mathcal{G}_1, \dots A_d \in \mathcal{G}_d$

$$P(A_1 \cap A_2 \cap \cdots \cap A_d) = P(A_1)P(A_2)\dots P(A_d)$$

We introduce the σ algebrae

$$\mathcal{F}_n = \bigvee_{k=0}^{n} \mathcal{G}_k, \quad \mathcal{F}_\infty = \bigvee_{k=0}^{\infty} \mathcal{G}_k, \quad \mathcal{T}_{-n} = \bigvee_{k=n}^{\infty} \mathcal{G}_k, \quad \mathcal{T}_{-\infty} = \bigcap_{n \in \mathbb{N}} \mathcal{T}_{-n}$$

Then the σ-algebra $\mathcal{T}_{-\infty}$ is P-trivial, i.e. $A \in \mathcal{T}_{-\infty} \Longrightarrow P(A) \in \{0, 1\}$

Proof *By assumption the σ-algebrae $\mathcal{F}_{n-1}$ and $\mathcal{T}_n$ are P-independent.*

Let $A \in \mathcal{T}_{-\infty} \subseteq \mathcal{F}_\infty$, then for all $n \in \mathbb{N}$ A is P-independent from $\mathcal{F}_n$.

It is easy to see that A is also P-independent from $\mathcal{F}_\infty$: for $B \in \mathcal{F}_\infty$, consider

$$E(\mathbf{1}_B|\mathcal{F}_n)(\omega) = P(B|\mathcal{F}_n)(\omega) \to \mathbf{1}_B(\omega) \ P \ a.s. \text{ and in } L^1(P)$$

Then

$$P(A \cap B) = E(\mathbf{1}_A \mathbf{1}_B) = E\big(\mathbf{1}_A \lim_{n\to\infty} E(\mathbf{1}_B|\mathcal{F}_n)\big) = \lim_{n\to\infty} E\big(\mathbf{1}_A E(\mathbf{1}_B|\mathcal{F}_n)\big)$$
$$= \lim_{n\to\infty} E(\mathbf{1}_A) E\big(E(\mathbf{1}_B|\mathcal{F}_n)\big) = \lim_{n\to\infty} P(A)P(B) = P(A)P(B)$$

Since $A \in \mathcal{F}_\infty$, A is P-independent from itself and

$$P(A) = P(A \cap A) = P(A)P(A) = P(A)^2 \Longrightarrow P(A) \in \{0, 1\}$$

Theorem 12. *(Kolmogorov's strong law of large numbers)*

Let $(X_t(\omega) : t \in \mathbb{N})$ i.i.d. with $X_1 \in L^1(P)$, and

$$S_t(\omega) = X_1(\omega) + \cdots + X_t(\omega)$$

Then

$$\lim_{t\to\infty} t^{-1} S_t(\omega) = E_P(X_1) \qquad P\text{-a.s. and in } L^1(P).$$

Proof Consider the backward filtration $\mathbb{F} = (\mathcal{F}_{-t} : t \in \mathbb{N})$ where for $t \le 0$

$$\mathcal{F}_{-t} = \sigma(S_t, S_{t+1}, \dots),$$

the $\mathbb{F}$-martingale

$$M_{-t} = E_P(X_1|\mathcal{F}_{-t}) \qquad t \in \mathbb{N}$$

The σ-algebra $\mathcal{F}_t$ is decreasing with respect to $t \in (-\mathbb{N})$.

By symmetry, the random pairs (S_t, X_r) ja (S_t, X_1) are identically distributed for $1 \le r \le t$, and by P-independence for $t \ge 0$

$$M_{-t} := E_P(X_1|\mathcal{F}_{-t}) = E_P(X_1|S_t, S_{t+1}, S_{t+2}, \dots)$$
$$= E_P(X_1|S_t, X_{t+1}, X_{t+2} \dots) = E_P(X_1|\sigma(S_t)) = E_P(X_r|\sigma(S_t)) \quad \forall 1 \le r \le t$$

which means

$$S_t = E_P(X_1 + \cdots + X_t|\sigma(S_t)) = \sum_{r=1}^{t} E_P(X_r|\sigma(S_t)) = tE_P(X_1|\sigma(S_-))$$

and $M_{-t}(\omega) = S_t(\omega)/t$ for $t \geq 0$.

By Doob's martingale backward convergence theorem

$$M_{-\infty}(\omega) = \lim_{t\to\infty} t^{-1}S_t(\omega) = M_{-\infty}(\omega) \quad P \text{ a.s. and in } L^1(P)$$

where we define for all $\omega \in \Omega$

$$M_{-\infty}(\omega) := \liminf_{t\to\infty} t^{-1}S_t(\omega)$$

Note also that $\forall\omega \in \Omega$, $\forall n \in \mathbb{N}$

$$\liminf_{t\to\infty} \frac{1}{t}S_t(\omega) = \liminf_{t\to\infty} \frac{1}{t}\sum_{i=1}^{n} X_i(\omega) + \liminf_{t\to\infty} \frac{1}{t}\sum_{i=(n+1)}^{t} X_i(\omega)$$

$$= 0 + \liminf_{t\to\infty} \frac{1}{t}\sum_{i=(n+1)}^{t} X_i(\omega)$$

is $\mathcal{T}_{-n} = \sigma(X_n, X_{n+1}, \dots)$-measurable $\forall n$, therefore it is measurable with respect to the tail σ-algebra $\mathcal{T}_{-\infty}$. Since the random variables $(X_t)_{t\in\mathbb{N}}$ are P-independent, by Kolmogorov's $0-1$ law it follows that $M_{-\infty}(\omega)$ is P-trivial: $P(t \leq M_{-\infty}) \in \{0,1\}$ $\forall t$ and $P(M_{-\infty} < \infty) = 1$, there is $c \in \mathbb{R}$ such that $P(M_{-\infty} = c) = 1$.

P almost surely and in $L^1(P)$

$$\frac{1}{t}S_t(\omega) \to c = E_P(X_1|\mathcal{F}_{-\infty})(\omega)$$

By taking expectation

$$c = E_P(M_{-\infty}) = E_P\big(E_P(X_1|\mathcal{F}_{-\infty})\big) = E_P(X_1).$$

Note $t^{-1}S_t(\omega) = E_P(X_1|\sigma(S_t))(\omega)$ follows from symmetry, and then we applied martingale backward convergence P-a.s. and in $L^1(P)$. Independence was needed to show that the limit

$$E_P(X_1|\sigma(S_t))(\omega) = E_P(X_1|\sigma(S_t, S_{t+1}, S_{t+2} \dots))(\omega)$$

is P-trivial. Without the independence assumption, we obtain the limit is a random variable. This extension is De Finetti's theorem. Bruno De Finetti (1906-1985) was an italian probabilist, economist and philosepher.

4.4 Exchangeability and De Finetti's theorem

Definition 26. *The sequence of random variables $(X_t)_{t\in\mathbb{N}}$ where $X_t(\omega)$ takes values in the measurable space $(S, \mathcal{S})$ is infinitely exchangeable (suomeksi äärettömästi vaihdettavissa) when $\forall n$, $t_1, \dots, t_n \in \mathbb{N}$ and any π permutation of $\{1, \dots, n\}$, the random vectors $(X_{t_1}, \dots, X_{t_n})$ and $(X_{t_{\pi(1)}}, \dots, X_{t_{\pi(n)}})$ have the same distribution under P.*

Note that that when $X_t(\omega)$ takes values in $\mathbb{R}$,

$$M_{-t}(\omega) = t^{-1} S_t(\omega) := E(X_1 | \mathcal{T}_{-t}), \quad t \in \mathbb{N}$$

is an uniformly integrable martingale in the backward filtration $\mathbb{F}$ which has a limit P-a.s. and in $L^1(P)$ as $t \to \infty$

$$M_{-\infty}(\omega) = E(X_1 | \mathcal{T}_{-\infty})(\omega) .$$

The tail σ-algebra $\mathcal{T}_{-\infty}$ is not necessarly trivial and $M_{-\infty}(\omega)$ is a random variable.

Definition 27. *The random variables $(X_t(\omega) : t \in \mathbb{N})$ taking values in $(S, \mathcal{S})$ are* **conditionally** *indendent and identically distributed given the σ-algebra $\mathcal{G}$ when, $\forall n$, $t_1, \dots, t_n$, $A_1 \dots A_n \in \mathcal{S}$,*

$$P(X_{t_1} \in A_1, \dots, X_{t_n} \in A_n | \mathcal{G})(\omega) = \prod_{i=1}^{n} P(X_1 \in A_i | \mathcal{G})(\omega) \quad P \text{ a.s.}$$

By taking expectation of the conditional expectation it follows that conditionally i.i.d. random variables are infinitely exchangeable. The reverse implication holds.

Theorem 13. *(De Finetti) Assume that $(S, \mathcal{S})$ is a Borel space, and the random sequence $(X_t(\omega) : t \in \mathbb{N}) \subseteq S$ is infinitely exchangeable w.r.t. P.*

Then $(X_t(\omega) : t \in \mathbb{N})$ are conditionally independent and identically distributed with respect to a tail σ-algebra $\mathcal{T}_{-\infty}$ to be defined below.

Proof Let consider the *empirical measure* of the first t- variables

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

which generated the σ-algebra

$$\sigma(\mu_t) = \sigma\{\mu_t(A) : A \in \mathcal{S}\} \subseteq \mathcal{F}.$$

Note that $\sigma(\mu_t) \subseteq \sigma(X_1, \dots, X_t)$, and for $t > 1$ it is strictly smaller because it contains the information about the realized values of the random variables but it forgets their time order.

Define the decreasing sequence of σ-algebrae

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

Let $1 \leq k \leq t \in \mathbb{N}$ and $f(x_1, \dots, x_k) : S^k \to \mathbb{R}$ a bounded measurable function. By symmetry we compute $E_P(f(X_1, \dots, X_k) | \mathcal{T}_{-t})(\omega)$:

Define the random probability measure

$$\mu_t^{\circ k} : \mathcal{S}^{\otimes k} \to [0, 1]$$

which is a regular version of the conditional distribution of the random vector $(X_1, \dots, X_k)$ conditionally on the σ-algebra $\sigma(\mu_t)$ (the regular version exists since $(S, \mathcal{S})$ is a Borel space).

By symmetry

$$
\begin{aligned}
&E_P(f(X_1,\dots,X_k)|\sigma(\mu_t))(\omega) \\
&= \mu_t^{\circ k}(f;\omega) := \int_{S^k} f(x)\mu_t^{\circ k}(dx;\omega) = \frac{1}{t!}\sum_{\pi} f(X_{\pi(1)}(\omega),\dots,X_{\pi(k)}(\omega)) \\
&= \frac{(t-k)!}{t!} \sum_{1\le i_1,\dots,i_k\le t \text{ distinct}} f(X_{i_1},X_{i_2},\dots,X_{i_k})
\end{aligned}
$$

where we sum over the permutation π of the set $\{1,\dots,t\}$.

Note that $\mu^{\circ k}(dx;\omega)$ is $\sigma(\mu_t)$-measurable, since it depends only on the values $\{X_1(\omega),\dots,X_t(\omega)\}$ and not by their ordering. Note also that $\mu_t^{\circ k}(dx)$ is not a product measure, in the sum there are not terms with repeated indexes.

For $k=1$

$$
\mu_t^{\circ 1}(A) = \mu_t(A) = \frac{1}{t}\sum_{k=1}^{t} \mathbf{1}(X_k \in A)
$$

is the empirical measure of $(X_1(\omega),\dots,X_t(\omega))$.

For $k\le t$ and any permutation π of $\{1,\dots,t\}$, by exchangeability $(X_1,\dots,X_k,\mu_t)$ and $(X_{\pi(1)},\dots,X_{\pi(k)},\mu_t)$ have the same distribution, which implies

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

By taking the normalized sum over the permutations,

$$
\mu_t^{\circ k}(f;\omega) = E_P(f(X_1,\dots,X_k)|\sigma(\mu_t))(\omega)
$$

Next we show that

$$
E_P(f(X_1,\dots,X_k)|\sigma(\mathcal{T}_{-t}))(\omega) = E_P(f(X_1,\dots,X_k)|\sigma(\mu_t))(\omega)
$$

Note also that

$$
\mathcal{T}_{-t} = \sigma(\mu_t,\mu_{t+1},\mu_{t+2},\dots) = \sigma(\mu_t,X_{t+1},X_{t+2},\dots)
$$

since the empirical measures $\mu_t(dx;\omega)$ and $\mu_{t+1}(dx;\omega)$ determine $X_{t+1}(\omega)$ by the identity

$$
(\mu_{t+1}-\mu_t)(dx) = \frac{1}{t+1}\Big(\mathbf{1}(X_{t+1}\in dx) - \mu_t(dx)\Big)
$$

By using infinite exchangeability follows the identity in law it follows that in law

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

for $m\in\mathbb{N}$ and any permutation π of $\{1,\dots,n\}$.

Exercise 12. *$(X_1,\dots,X_n)$ and $(X_{n+1},X_{n+2},\dots)$ are conditionally independent given $\sigma(\mu_n)$,*

Solution *Note that a random variable $W(\omega)$ is $\sigma(\mu_n)$-measurable if and only if $W(\omega)=g(X_1,\dots,X_n)$ where g is measurable and symmetric, i.e.*

$$g(x_1,\dots,x_m)=g(x_{\pi(1)},\dots,x_{\pi(m)})\quad \forall\pi \text{ permutations }.$$

Assume also that $g(x_1,\dots,x_n)$ is bounded, $Y(\omega)$ is a bounded and $\sigma(X_{n+1},X_{n+2},\dots)$-measurable random variable and $f(x_1,\dots,x_n)$ is bounded and $\mathcal{S}^{\otimes n}$-measurable, (not necessarly symmetric).

By infinite exchangeability it follows that $\forall n\in\mathbb{N}$ and for all permutations π of the indexes $\{1,\dots,n\}$, the sequences

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

have the same distribution,

$$\begin{aligned}
&E_P(Y\ W\ f(X_1,\dots,X_n))E_P(Y\ g(X_1,\dots,X_n)\ f(X_1,\dots,X_n))\\
&=E_P(Y\ g(X_{\pi(1)},\dots,X_{\pi(n)})\ f(X_{\pi(1)},\dots,X_{\pi(n)}))\\
&\quad(\text{ since the sequence is exchangeable })\\
&=E_P(Y\ g(X_1,\dots,X_n)\ f(X_{\pi(1)},\dots,X_{\pi(n)}))=E_P(Y\ W\ f(X_{\pi(1)},\dots,X_{\pi(n)}))\\
&\quad(\text{ since } g \text{ is symmetric })\\
&=\frac{1}{n!}\sum_{\pi}E_P\Big(YWf(X_{\pi(1)},\dots,X_{\pi(n)})\Big)=E_P\Big(YW\frac{1}{n!}\sum_{\pi}f(X_{\pi(1)},\dots,X_{\pi(n)})\Big)\\
&=E_P\big(Y\ W\ \mu_n^{\circ n}(f)\big)
\end{aligned}$$

By definition of conditional expectation

$$E_P\big(f(X_1,\dots,X_n)\big|\sigma(\mu_n,X_{n+1},X_{n+2},\dots)\big)(\omega)=\mu_n^{\circ n}(f;\omega)=E_P\big(f(X_1,\dots,X_n)\big|\sigma(\mu_n)\big)(\omega)$$

which means that under P, $(X_1,\dots,X_n)$ ja $(X_{n+1},X_{n+2},\dots)$ conditionally independent conditionally on $\sigma(\mu_n)$.

In other words, $\mathcal{T}_{-n}$ does not contain information about the time-order of the first n values of the sequence.

Since $M_{-t}^{(k)}(f):=\mu_t^{\circ k}(f)$ is a martingale in the filtration $(\mathcal{T}_{-t}:t\in\mathbb{N})$, by Doob's martingale backward convergence theorem as $t\to\infty$, the limit $M_{-\infty}^{(k)}(f)$ exists P-a.s. and in $L^1(P)$ sense.

Since $(X_1,\dots,X_k)$ takes values in a Borel space, the conditional probability

$$P((X_1,\dots,X_k)\in A|\mathcal{T}_{-\infty})(\omega),\quad A\in\mathcal{S}^{\otimes k}$$

has a regular version, which is a $\mathcal{T}_{-\infty}$-measurable probability kernel $\mu_\infty^{\circ k}(dx;\omega)$ on $(S_1\times\dots\times S_k)$ such that P-a.s., for all bounded measurable functions

$$\begin{aligned}
&M_{-\infty}^{(k)}(f;\omega)=E_P(f(X_1,\dots,X_k)|\sigma(\mathcal{T}_{-\infty}))(\omega)\\
&=\int_{S_1,\dots,S_k}f(x_1,\dots,x_k)\mu_\infty^{\circ k}(dx_1,\dots dx_k;\omega)
\end{aligned}$$

For $k = 1$ denote $\mu_\infty = \mu_\infty^{\circ 1}$, where

$$\lim_{t\to\infty} \frac{1}{t}\sum_{i=1}^{t} f(X_i(\omega)) = \int_S f(x)\mu_\infty(dx,\omega) \quad P\text{-a.s.}$$

Exercise 13. *Since $(S,\mathcal{S})$ is a Borel space there is a measurable injection $f : (S,\mathcal{S}) \to ([0,1],\mathcal{B}([0,1]))$ with measurable inverse f^{-1}. It follows that $A \subseteq \mathcal{S}$, $A \in \mathcal{S}$ if and only if $f(A)$ is Borel set. Since*

$$\sigma\big\{(a,b] : 0 \le a < b \le 1,\ a,b \in \mathbb{Q}\big\} = \mathcal{B}([0,1])$$

it follows that also $\mathcal{S}$ is countably generated, since

$$\mathcal{S} = \sigma\big\{f^{-1}((a,b]\cap f(S)) : 0 \le a < b \le 1, a,b \in \mathbb{Q}\big\} = \sigma\big\{A(\ell) : \ell \in \mathbb{N}\big\}$$

We know a priori that $\forall A \in \mathcal{S}$, $\exists \mathcal{N}_A \subseteq \Omega$ with $P(\mathcal{N}_A) = 0$ such that

$$\mu_t(A;\omega) \to \mu_\infty(A;\omega) \quad \forall \omega \notin \mathcal{N}_A$$

Since $P(\mathcal{N}) = 0$ where $\mathcal{N} = \bigcup_{\ell\in\mathbb{N}} \mathcal{N}_{A(\ell)}$, it follows that

$$\mu_t(A_\ell;\omega) \to \mu_\infty(A_\ell;\omega) \quad \forall \ell \in \mathbb{N} \quad \forall \omega \notin \mathcal{N}$$

and since $\sigma\{A_\ell : \ell \in \mathbb{N}\} = \mathcal{S}$ it follows that $\forall A \in \mathcal{S}$

$$\mu_t(A;\omega) \to \mu_\infty(A;\omega) \quad \forall A \in \mathcal{S} \quad \forall \omega \notin \mathcal{N} \tag{4.1}$$

Similarly we find a P-null set $\widetilde{\mathcal{N}} \subseteq \Omega$ such that $\forall k \in \mathbb{N}$, $\forall \{A_i\} \subseteq \mathcal{S}$

$$\mu_t^{\circ k}(A_1 \times \cdots \times A_k;\omega) \to \mu_\infty^{\circ k}(A_1 \times \cdots \times A_k;\omega) \quad \forall \omega \notin \widetilde{\mathcal{N}} \tag{4.2}$$

P-a.s. the collection of finite dimensional distributions

$$\Big\{\mu_\infty^{\circ k}(dx_1,\dots dx_k;\omega) : k \in \mathbb{N}\Big\}$$

is consistent (check !), and by Kolmogorov's extension theorem **??** is follows that for each ω outside a P-null set there is a random probability measure $\boldsymbol{\nu}_\infty(\,\cdot\,;\omega)$ on the space of sequences $(x_k : k \in \mathbb{N}) \subseteq S$ such that $\forall k$, $A_1,\dots,A_k \in \mathcal{S}$

$$P\big(X_1 \in A_1,\dots X_k \in A_k|\mathcal{T}_{-\infty})(\omega) =$$
$$\mu_\infty^{\circ k}(A_1 \times \cdots \times A_k;\omega) = \boldsymbol{\nu}_\infty\big(\big\{(x_l : l \in \mathbb{N}) : x_1 \in A_1,\dots,x_k \in A_k\big\};\omega\big)$$

We show that P-a.s. $\boldsymbol{\nu}_\infty(\cdot;\omega)$ is an product measure of infinite copies, that is $\forall k$

$$P\big(X_1 \in A_1,\dots X_k \in A_k|\mathcal{T}_{-\infty})(\omega) = \prod_{i=1}^{k} P(X_1 \in A_i|\mathcal{T}_{-\infty})(\omega)$$

Let $\mu_t^{\otimes k}$ be the k-fold product measure of the empirical measure μ_t. For every bounded and Borel measurable $f(x_1,\dots,x_k)$,

$$\mu_t^{\otimes k}(f) = t^{-k} \sum_{1\le i_1,\dots,i_k \le t} f(X_{i_1},\dots,X_{i_k})$$

where the sum contains also terms with repeated indexes. Then

$$(\mu_t^{\circ k} - \mu_t^{\otimes k})(f) = \mu_t^{\circ k}(f) - \mu_t^{\otimes k}(f) =$$
$$\mu_t^{\circ k}(f)\left(1 - \frac{t!}{t^k(t-k)!}\right) + t^{-k} \sum_{1 \le i_1,\dots i_k \le t:\ \exists l \ne m\ i_l = i_m} f(X_{i_1}, \dots, X_{i_k})$$

where in the first part we have terms without repeated indexes and in the second part all terms have at least on index repeated. Then $\forall k \in \mathbb{N}, \omega \in \Omega$,

$$|\mu_t^{\circ k}(f;\omega) - \mu_t^{\otimes k}(f;\omega)|$$
$$\le \| f \|_\infty \left(1 - \prod_{l=0}^{k-1} \frac{(t-l)}{t} + t^{-k}\binom{k}{2} t^{k-1}\right) \to 0 \text{ kun } t \to \infty$$

where $\| f \|_\infty = \sup_{x \in S} |f(x)|$ and the upper bound does not depend on ω.

For all $A_1, A_2 \dots \in \mathcal{S}$, $\forall k$ P-a.s. as $t \to \infty$

$$\mu_t^{\circ k}(A_1 \times A_2 \times \dots \times A_k) \to \mu_\infty^{\circ k}(A_1 \times A_2 \times \dots \times A_k)\ .$$

For $k = 1$

$$\mu_t^{\circ 1}(A_i) \to \mu_\infty(A_i),$$

and convergence follows also for the product measures

$$\mu_t^{\otimes k}(A_1 \times A_2 \times \dots \times A_k) = \prod_{i=1}^k \mu_t^{\circ 1}(A_i) \to \prod_{i=1}^k \mu_\infty(A_i) = \mu_\infty^{\otimes k}(A_1 \times A_2 \times \dots \times A_k).$$

By triangle inequality

$$|\mu_{-\infty}^{\circ k}(f) - \mu_{-\infty}^{\otimes k}(f)|$$
$$\le |\mu_{-\infty}^{\circ k}(f) - \mu_t^{\circ k}(f)| + |\mu_t^{\circ k}(f) - \mu_t^{\otimes k}(f)| + |\mu_t^{\otimes k}(f) - \mu_\infty^{\otimes k}(f)| \to 0$$

P-a.s. as $t \to \infty$, and

$$\mu_\infty^{\circ k}(f;\omega) = \mu_\infty^{\otimes k}(f;\omega) \qquad P\text{-a.s}$$

for all bounded measurable $f(x_1, \dots, x_k)$. It means that Kolmogorov's extension $\boldsymbol{\nu}_\infty$ is a product measure on the space of infinite sequences $S^{\mathbb{N}}$. For exery bounded measurable functions $g_1, \dots, g_k : S \to \mathbb{R}$

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

By taking expectations,

$$E_P(g_1(X_1)\dots g_k(X_k))$$
$$= E_P\left(\int_S g_i(x)\mu_\infty(dx) \right) = \int_{\mathcal{M}(S)} \left\{ \prod_{i=1}^k \int_S g_i(x)\mu(dx) \right\} Q(d\mu)$$

where Q is the distribution of the random measure $\mu_\infty(dx;\omega)$ in the space

$$\mathcal{M}(S) = \left\{ \text{ probability measures } \nu : \mathcal{S} \to [0,1] \right\}$$

In other words, a permutation symmetric (i.e. infinitely exchangeable) random sequence with values in a Borel space is the mixture of i.i.d. sequences
□

Exercise 14. *De Finetti original proof was for the simplest case of random binary sequences, where $S = \{0,1\}$ and the space of probability measures on S is $\mathcal{M}(S) = [0,1]$.*

Let $S_t(\omega) = (X_1(\omega) + \dots + X_t(\omega))$.

In coin-toss experiment, if the sequence of coin tosses is infinitely exchangeable under P, it has a limit $\vartheta(\omega) := \lim_{t\to\infty} t^{-1} S_t(\omega) \in [0,1]$ P-a.s. and in $L^1(P)$ sense.

Let $Q(d\theta) = P(\{\omega : \vartheta(\omega) \in d\theta\})$. By conditioning on the σ-algebra $\sigma(\vartheta)$, the coin-tosees are conditionally independent and Bernoulli distributed, with the same random probability-parameter $\vartheta(\omega) \in [0,1]$. The probability distribution of the limit $Q(d\theta)$ is interpreted as a priori probability on the parameter ϑ. It follows $\forall k$, $(x_i)_{i\in\mathbb{N}} \subseteq \{0,1\}$,

$$P\big(X_1 = x_1, \dots, X_k = x_k\big) = \int_0^1 \left\{ \prod_{i=1}^k P(X_1 = x_i | \vartheta = \theta) \right\} Q(d\theta)$$

$$= \int_0^1 \theta^{S_k} (1-\theta)^{(k-S_k)} Q(d\theta)$$

$$Q(B) = P\big(\big\{\omega : \lim_{t\to\infty} t^{-1} S_t(\omega) \in B\big\}\big), \quad B \in \mathcal{B}([0,1])$$

De Finetti's theorem is at the mathematical foundation of Bayesian statistical inference.

4.5 Doob optional sampling and optional stopping theorems

Lemma 14. *Let $(X_t : t \in \mathbb{N})$ a supermartingale and $0 \le \tau(\omega) \le k$ a bounded stopping time.*

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

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 14. *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 14 for the bounded stopping time $(\tau \wedge k) \leq k$ together with the fact that $A \cap \{\tau \leq k\}$ is also $\mathcal{F}_{(\tau \wedge k)}$-measurable. To check this, for all $t \in \mathbb{N}$ we have

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

Since $\mathbf{1}(\tau(\omega) \leq k) \uparrow \mathbf{1}(\tau(\omega) < \infty)$ 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 9. *Let $\tau(\omega) \geq \sigma(\omega)$ stopping times. and $(M_t : t \in \mathbb{N})$ an UI martingale.*

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

$$E_P(M_\tau | \mathcal{F}_\sigma) = M_\sigma \tag{4.3}$$

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

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

$$M_\sigma = E_P(M_\infty | \mathcal{F}_\sigma) = E_P(E_P(M_\infty | \mathcal{F}_\tau) | \mathcal{F}_\sigma) = E_P(M_\tau | \mathcal{F}_\sigma) \tag{4.4}$$

Corollary 10. *If $(M_t, t \in \mathbb{N})$ is a martingale and*

$$0 \leq \sigma(\omega) \leq \tau(\omega) \leq K \in N \tag{4.5}$$

are bounded stopping times, then

$$E_P(M_\tau | \mathcal{F}_\sigma) = M_\sigma \tag{4.6}$$

Proof apply corollary (9) to $(M_t : t = 1 \dots, K)$ which is uniformly integrable since it is a finite subset of $L^1(P)$.

Corollary 11. *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

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

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

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

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 15. *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*

We construct a martingale $(X_n(\omega) : n \in \mathbb{N})$ as follows: the filtration is the one generated by the sequence. $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$.

Let

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

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

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

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 $m > n$.

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

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 $m > n$,

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

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,

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

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

Corollary 12. *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*

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

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. and by the monotone convergence theorem also in $L^1(P)$.

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 15. *Let $(X_t(\omega) : t \in \mathbb{N})$ be a non-negative supermartingale. Since it is non-negative is automatically bounded in L^1, 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 $A \in \mathcal{F}_t$

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

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

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

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 □

4.6 Change of measure and Radon-Nikodym theorem

Definition 28. *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 $\nu \overset{\mathcal{G}}{\ll} \mu$.

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 16. *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 $P(A_n) \leq 2^{-n}$ and $Q(A_n) \geq \varepsilon > 0$ By Borel Cantelli lemma $P(\limsup A_n) = 0$, while by reverse Fatou lemma

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

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

Theorem 15. *(Radon-Nikodym) Let μ 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*) $\mathcal{F} = \sigma(F_n : n \in \mathbb{N})$ 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 $\limsup$.

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

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

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

Since $Q \ll P$, by lemma 16 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 29. *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 $\alpha \vee \beta$*

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

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 $\mathcal{G}' \supseteq \overline{\mathcal{G}}$, $\mathcal{G}'' \supseteq \overline{\mathcal{G}}$, $\mathcal{G}', \mathcal{G}'' \in \mathbb{G}$, then

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

By the triangle inequality this it is equivalent to

$$E_P\big(|Z_{\bar{\mathcal{G}}} - Z_{\mathcal{G}'}|\big) < \varepsilon/2$$

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 $(Z_{\mathcal{G}_n} : n \in \mathbb{N} \cup \{+\infty\})$ would be an uniformly integrable martingale in the filtration $\{\mathcal{G}_n\}$, which necessarily 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 $d(x^*, x_\alpha) \leq \varepsilon$ $\forall \alpha \geq \overline{\alpha}$.

Proof: for every n let $\overline{\alpha}_n$ such that $d(x_{\overline{\alpha}_n}, x_\alpha) \leq n^{-1}$ $\forall \alpha \geq \overline{\alpha}_n$, and we can choose $\overline{\alpha}_n \geq \overline{\alpha}_{n-1}$.

Therefore $(x_{\overline{\alpha}_n})$ 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).

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

Since

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

we have

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

where $\varepsilon > 0$ is arbitrarily small □

4.7 The Likelihood ratio process

Consider a probability space $(\Omega, \mathcal{F})$ equipped with a filtration $\mathbb{F} = (\mathcal{F}_t : t \in T)$, $(T = \mathbb{N}, \mathbb{R})$ and two probability measures P, Q. such that Q dominates P *locally*

$$P \overset{loc}{\ll} Q, \text{ which means } P_t \ll Q_t \quad \forall t \in T, t < \infty$$

where P_t, Q_t are the restriction of P, Q on the σ-algebra $\mathcal{F}_t$. In other words, if $A \in \mathcal{F}_t$ for some $t < \infty$ and $Q(A) = 0$, then $P(A) = 0$.

By the Radon-Nikodym theorem,there is a likelihood-ratio process

$$0 \leq Z_t(\omega) = \frac{dP_t}{dQ_t}(\omega) \in L^1(\Omega, \mathcal{F}_t, Q), \quad 0 \leq t < \infty,$$

such that $\forall A \in \mathcal{F}_t$, the change of measure formula holds

$$P(A) = E_Q(Z_t \mathbf{1}_A)$$

Proposition 11. *The process $(Z_t(\omega), 0 \leq t < \infty)$ is a $(Q, \mathbb{F})$-martingale.*

Proof. For $s \leq t$, $\forall A \in \mathcal{F}_s \subseteq \mathcal{F}_t$ the martingale propery follows:

$$P(A) = E_Q(Z_s \mathbf{1}_A) = E_P(Z_t \mathbf{1}_A)$$

Uniformly integrable likelihood-process Consider the discrete time case with $T = \mathbb{N}$. When $P \overset{loc}{\ll} Q$, $(Z_t : t \in \mathbb{N})$ is a non-negative $(Q, \mathbb{F})$-martingale and by Doob's convergence theorem there is $Z_\infty(\omega) \in L^1(Q)$ such that

$$Z_t(\omega) \to Z_\infty(\omega) \quad Q \text{ and } P \text{ almost surely .}$$

with $E(Z_\infty) \leq 1 = E(Z_0)$.

Moreover by lemma (15) $(Z_t : t \in \mathbb{N})$ is uniformly integrable with

$$Z_t(\omega) = E_Q(Z_\infty | \mathcal{F}_t)(\omega),$$
$$Z_t \overset{L^1(Q)}{\longrightarrow} Z_\infty,$$

if and only if $E(Z_\infty) = 1$. In this case $P \ll Q$ not just locally but also on the σ-algebra

$$\mathcal{F}_\infty = \bigvee_{t \in \mathbb{N}} \mathcal{F}_t$$

Martingales in mathematical statistics We continue with a probability space $(\Omega, \mathcal{F})$ equipped with the filtration $\mathbb{F} = (\mathcal{F}_t : t \geq 0)$, and consider a family of probability measures $(P_\theta(d\omega) : \theta \in \Theta)$, with parameter space $(\Theta \subseteq \mathbb{R}^d)$, such that $P_\theta \overset{loc}{\ll} Q \; \forall \theta \in \Theta$.

Denote

$$Z_t^\theta(\omega) = \frac{dP_t^\theta}{dQ_t}(\omega), \quad t \geq 0.$$

Assume

1. i) For $t > 0$ and $\forall \omega$, $Z_t^\theta(\omega)$ is continuously differentiable w.r.t. θ, with random gradient vector

$$V_t^\theta(\omega) = \nabla_\theta \log Z_t^\theta(\omega) = \left(\frac{\partial \log Z_t^\theta(\omega)}{\partial \theta_i} : i = 1, \ldots, d \right) = \left(\frac{1}{Z_t^\theta(\omega)} \frac{\partial Z_t^\theta(\omega)}{\partial \theta_i} : i = 1, \ldots, d \right)$$

   such that

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \{ Z_t^{\theta + \varepsilon h} - Z_t^\theta \} = (h, V_t(\theta)) Z_t^\theta \qquad \forall h \in \mathbb{R}^d, \omega \in \Omega$$

   $V_t^\theta(\omega)$ is called *score* .

   In order to interchange the order of differentiation and integration we also assume

2. $\nabla_\theta Z_t^\theta$ is locally uniformly dominated at θ, i.e. there is an U neighbourhood of θ and a random variable $0 \leq D_t(\theta, \omega) \in L^1(\Omega, \mathcal{F}_t, Q)$ and

$$|\nabla_\theta Z_t^\theta(\eta)| < D_t(\theta), \qquad \forall \eta \in U.$$

For $B \in \mathcal{F}_t$, by Fubini and dominated convergence

$$\frac{\partial}{\partial \theta_i} \int_\Omega \mathbf{1}_B Z_t^\theta dQ = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_\Omega \mathbf{1}_B (Z_t^{\theta + \varepsilon e_i} - Z_t^\theta) dQ =$$

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_\Omega \mathbf{1}_B \Big(\int_0^\varepsilon \frac{\partial}{\partial \theta_i} Z_t^{\theta + \epsilon e_i} d\epsilon \Big) dQ = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_0^\varepsilon \Big(\int_\Omega \mathbf{1}_B \frac{\partial}{\partial \theta_i} Z_t^{\theta + \epsilon e_i} dQ \Big) d\epsilon = \int_\Omega \mathbf{1}_B \frac{\partial}{\partial \theta_i} Z_t^\theta dQ.$$

Moreover $B \in \mathcal{F}_s$,

$$\int_\Omega \mathbf{1}_B E_Q\big(\frac{\partial}{\partial \theta_i} Z_t^\theta | \mathcal{F}_s \big) dQ = \int_\Omega \mathbf{1}_B \frac{\partial}{\partial \theta_i} Z_t^\theta dQ =$$

$$\frac{\partial}{\partial \theta_i} \int_\Omega \mathbf{1}_B Z_t^\theta dQ =$$

$$\frac{\partial}{\partial \theta_i} \int_\Omega \mathbf{1}_B Z_s^\theta dQ = \int_\Omega \mathbf{1}_B \frac{\partial}{\partial \theta_i} Z_s^\theta dQ = \int_\Omega \mathbf{1}_B \frac{\partial}{\partial \theta_i} E_Q(Z_t^\theta | \mathcal{F}_s) dQ$$

and we can change the order of derivation and integration

$$\frac{\partial}{\partial \theta_i} E_Q(Z_t^\theta | \mathcal{F}_s) = E_Q\big(\frac{\partial}{\partial \theta_i} Z_t^\theta | \mathcal{F}_s \big)$$

Proposition 12. *Under the previous assumption on the statistical model in a neighbourhood of θ, $\{V_t(\theta)\}_{t\geq 0}$ is a $(P^\theta,\mathbb{F})$-martingale: For $0\leq s\leq t$,*

$$E_{P^\theta}(V_t(\theta)|\mathcal{F}_s)=\frac{E_Q(Z_t^\theta V_t(\theta)|\mathcal{F}_s)}{E_Q(Z_t^\theta|\mathcal{F}_s)}=\frac{1}{Z_s^\theta}E_Q(\frac{\partial Z_t^\theta}{\partial\theta}|\mathcal{F}_s)=$$
$$\frac{1}{Z_s^\theta}\frac{\partial}{\partial\theta}E_Q(Z_t^\theta|\mathcal{F}_s)=\frac{1}{Z_s^\theta}\frac{\partial Z_s^\theta}{\partial\theta}=\frac{\partial\log Z_s^\theta}{\partial\theta}=V_s(\theta)$$

Essentially we had to assume that the limit $\nabla_\theta Z_t^\theta\in L^1(Q)$.

Since $\varepsilon^{-1}(Z_t^{\theta+\varepsilon h}-Z_t^\theta)\in L^1(Q)\ \forall\varepsilon>0$, is natural to use a weaker definition based on L^1-convergence instead of pointwise convergence.

Definition 30. *: A statistical experiment*
$(\Omega,\mathcal{F}_t,Q_t,(P_t^\theta)_{\theta\in\Theta})$ is L^1-differentiable at θ, if there is a random score-vector $V_t(\theta)\in L^1(P^\theta)$ such that $\forall h\in\mathbb{R}^d$

$$\lim_{\varepsilon\to 0}E_Q\left(\left|\frac{1}{\varepsilon}\{Z_t^{\theta+\varepsilon h}-Z_t^\theta\}-(h,V_t(\theta))Z_t^\theta\right|\right)=0$$

We show that under this generalized definition $V_t(\theta)$ as a random process is a $(P^\theta,\mathbb{F})$-martingale.

Proposition 13. *: If a time $t\geq 0$ the statistical experiment $(\Omega,\mathcal{F}_t,Q_t,(P_t^\theta)_{\theta\in\Theta})$ is L^1-differentiable at θ , then $\forall 0\leq s\leq t$ the statistical experiment $(\Omega,\mathcal{F}_s,Q_s,(P_s^\theta)_{\theta\in\Theta})$ is L^1-differentiable at θ , with random score-vector*

$$V_s(\theta)=E_{P_\theta}(V_t(\theta)|\mathcal{F}_s)$$

Proof: let $B\in\mathcal{F}_s$,

$$E_Q\left(\left\{\frac{1}{\varepsilon}\{Z_t^{\theta+\varepsilon h}-Z_t^\theta\}-(h,V_t(\theta))Z_t^\theta\right\}\mathbf{1}_B\right)$$
$$=E_Q\left(\left\{\frac{1}{\varepsilon}\{Z_s^{\theta+\varepsilon h}-Z_s^\theta\}-(h,E_Q(Z_t^\theta V_t(\theta)|\mathcal{F}_s))\right\}\mathbf{1}_B\right)$$
$$=E_Q\left(\left\{\frac{1}{\varepsilon}\{Z_s^{\theta+\varepsilon h}-Z_s^\theta\}-\big(h,\frac{E_Q(Z_t^\theta V_t(\theta)|\mathcal{F}_s)}{E_Q(Z_t^\theta|\mathcal{F}_s)}\big)Z_s^\theta\right\}\mathbf{1}_B\right)=$$
$$E_Q\left(\left\{\frac{1}{\varepsilon}\{Z_s^{\theta+\varepsilon h}-Z_s^\theta\}-\big(h,\ E_{P^\theta}(V_t(\theta)|\mathcal{F}_s)\big)Z_s^\theta\right\}\mathbf{1}_B\right)\to 0\text{ kun }\varepsilon\to 0$$

and since this holds $\forall B\in\mathcal{F}_s$,

$$E_Q\left(\left|\frac{1}{\varepsilon}\{Z_s^{\theta+\varepsilon h}-Z_s^\theta\}-\big(h,\ E_{P^\theta}(V_t(\theta)|\mathcal{F}_s)\big)Z_s^\theta\right|\right)\to 0\text{ kun }\varepsilon\to 0$$

Exercise 16. *(Laplace's two sided exponential distribution):*

For $P^\theta(dx)=\frac{1}{2}\exp(-|x-\theta|)dx$, the density $f^\theta(x)$ is not differentiable with respect to θ at the point $\theta_0=x$.

Nevertheless it is L^1-differentiable with score

$$V(\theta,x)=-sign(\theta-x)$$

Notes The story continues: since $Z_t^\theta \in L^1(Q)$, it follows that $\sqrt{Z_t^\theta} \in L^2(Q)$. When $\sqrt{Z_t^\theta}$ is L^2-differentiable, $V_t(\theta)$ is a square integrable $(P^\theta, \mathbb{F})$-martingale, we define *Fisher's information* as

$$I_t(\theta) = E_{P^\theta}(V_t(\theta)^\top V_t(\theta))$$

which is studied by using martingale theory.

4.8 Doob decomposition

Proposition 14. *Assume that $(X_t : t \in \mathbb{N})$ is an $\mathbb{F}$-adapted process. We always have the Doob decomposition*

$$\begin{aligned} X_t &= X_0 + M_t + A_t \textit{ where } A_0 = 0 \\ A_t &= \sum_{s=1}^t \Delta A_s = \sum_{s=1}^t \big(E(X_s|\mathcal{F}_{s-1}) - X_{s-1} \big) \textit{ is } \mathbb{F}\textit{-predictable,} \\ M_t &= \sum_{s=1}^t \Delta M_s = \big(X_s - E(X_s|\mathcal{F}_{s-1}) \big) \textit{ is a } \mathbb{F}\textit{-martingale} \end{aligned}$$

Proof write the telescopic sums with $\Delta X_t = \Delta M_t + \Delta A_t$.

When X_t is an $(\mathbb{F})$-submartingale (respectively supermartingale) A_t is non-decreasing (respectively non-increasing).

Consider the case where $(M_t : t \in \mathbb{N})$ and $(N_t : t \in \mathbb{N})$ are $\mathbb{F}$-martingales with $M_t, N_t \in L^2(\Omega)$ $\forall t \in \mathbb{N}$. For the product $N_t M_t$ we have

$$\begin{aligned} & M_t N_t - M_{t-1} N_{t-1} = N_{t-1}\Delta M_t + M_{t-1}\Delta N_t + \Delta M_t \Delta N_t \\ &= N_{t-1}\Delta M_t + M_{t-1}\Delta N_t + \Bigg(\Delta M_t \Delta N_t - E(\Delta M_t \Delta N_t|\mathcal{F}_{t-1}) \Bigg) + E(\Delta M_t \Delta N_t|\mathcal{F}_{t-1}) \end{aligned}$$

Denote

$$[N, M]_t = \sum_{s=1}^t \Delta N_s \Delta M_s, \quad \langle N, M\rangle_t = \sum_{s=1}^t E\big(\Delta N_s \Delta M_s \big| \mathcal{F}_{s-1}\big)$$

which are resepctively the (discrete) *quadratic covariation* and*predictable covariation* of the pair (N_t, M_t).

By writing the telescopic sum,

$$\begin{aligned} & N_t M_t - N_0 M_0 = \big(N_- \cdot M\big)_t + \big(M_- \cdot N\big)_t + [N, M]_t = \\ & \big(N_- \cdot M\big)_t + \big(M_- \cdot N\big)_t + \big([N, M]_t - \langle N, M\rangle_t\big) + \langle N, M\rangle_t = X_t + \langle N, M\rangle_t \end{aligned}$$

where the martingale transforms

$$\big(N_- \cdot M\big)_t = \sum_{s=1}^t N_{s-1}\Delta M_s, \quad \big(M_- \cdot N\big)_t = \sum_{s=1}^t M_{s-1}\Delta N_s$$

are $\mathbb{F}$-martingales (integrability follows by Cauchy-Schwartz inequality since $M_t, N_t \in L^(P))$. Also $\big([N, M]_t - \langle N, M\rangle_t\big)$ is an $\mathbb{F}$-martingale, and $\langle N, M\rangle_t$

is $\mathbb{F}$, predictable. Therefore the Doob decomposition is

$$N_t M_t = N_0 M_0 + X_t + \langle N, M\rangle_t, \quad \text{with martingale part}$$
$$X_t = \big(N_- \cdot M\big)_t + \big(M_- \cdot N\big)_t + \big([N,M]_t - \langle N, M\rangle_t\big)$$

Note that by taking expectation,

$$E(M_t N_t) - E(M_0 M_0) = E\big((M_t - M_0)(N_t - N_0)\big) = E\big(\langle M, N\rangle_t\big)$$

When $N_t = M_t$, by Jensen's inequality (M_t^2) is a $\mathbb{F}$-submartingale and the predictable variation

$$\langle M\rangle_t = \langle M, M\rangle_t = \sum_{s=1}^{t} E\big((\Delta M_s)^2 \big| \mathcal{F}_{s-1}\big)$$

is non-decreasing.

4.9 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 16. *Let $0 \le X_s(\omega), s \in \mathbb{N}$ a $(\mathcal{F}_t)$-submartingale.*
Then for $c > 0, T \in \mathbb{N}$,

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

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 $s \ne t$.
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 17. *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

$$\| X \|_p \le q \| Y \|_p \qquad \text{with } \left(\frac{1}{p} + \frac{1}{q}\right) = 1$$

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

$$E_P(X^p) = \int_\Omega \biggl(\int_0^{X(\omega)} p t^{p-1} dt \biggr) P(d\omega) = \int_0^\infty P(X \ge t) p t^{p-1} dt \le$$
$$\frac{p}{p-1} \int_0^\infty t P(X \ge t)(p-1) t^{p-2} dt \le q \int_0^\infty E_P\bigl(Y \mathbf{1}(X \ge t)\bigr)(p-1) t^{p-2} dt \le$$
$$q E_P\biggl(Y \int_0^{X(\omega)} (p-1) t^{p-2} dt \biggr) = q E_P(Y X^{p-1})$$
$$(\text{ Hölder }) \ \le q E_P(Y^p)^{1/p} E_P(X^{q(p-1)})^{1/q} = q \parallel Y \parallel_p \parallel X \parallel_p^{p-1} \ .$$

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 lemma holds for $X^{(n)}(\omega)$. The result follows by the monotone convergence theorem □

Theorem 17. *(Doob's L^p maximal inequality) Let $(M_t : t \in \mathbb{N})$ a martingale with $M_t \in L^p$ $\forall t \in \mathbb{N}$. Then for $1 < p < \infty, T \in \mathbb{N}$,*

$$\parallel M_T^* \parallel_p \le q \parallel M_T \parallel_p$$

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

$$c P\bigl(|M_T^*| > c\bigr) \le E_P\bigl(|M_T| \mathbf{1}(|M_T^*| > c)\bigr)$$

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

Corollary 13. *When $(M_t : t \in \mathbb{N})$ is a martingale in $L^2(P)$, we obtain*

$$E_P\bigl((M_T^*)^2\bigr) \le 4 E_P\bigl(M_T^2\bigr) = 4 \Bigl\{ E_P(M_0^2) + E_P\bigl(\langle M, M \rangle\bigr) \Bigr\}$$

Theorem 18. *(Kakutani) On a probability space $(\Omega, \mathcal{F}, P)$ let $(X_t : t \in \mathbb{N})$ P-independent random variables with $X_t(\omega) \ge 0$ and $E_P(X_t) = 1$.*
Let $\mathcal{F}_t = \sigma(X_1, \dots, X_t)$ and

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

M_t is a non-negative $(\mathcal{F}_t)$-martingale with $E(M_t) = 1$ and by Doob forward convergence theorem it has P-a.s. limit $M_\infty(\omega)$ as $t \to \infty$, with $M_\infty \in L^1(P)$, $E(M_\infty) \in [0, 1]$. 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: $\forall A \in \mathcal{F}_s$

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

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

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

which implies that (M_s) is an UI martingale:

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

3) $\Longrightarrow$ 2): Define

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

(N_t) is a non-negative martingale in $L^2(P)$.
By Doob L^p martingale inequality with $p = 2$,

$$E_P\Big(\sup_{s\leq t} M_s\Big) \leq (\text{ by Jensen's inequality}) \quad E_P\Big(\sup_{s\leq t} N_s^2\Big) \leq 4E(N_t^2) = \frac{4}{a_1^2 \dots a_t^2}$$

and by the monotone convergence theorem

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

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 $\big(\sup_{s\in\mathbb{N}} M_s\big) \in L^1(P)$.
In case $\prod_{t\in\mathbb{N}} a_t = 0$, by Fatou lemma

$$E_P\big(\sqrt{M_\infty}\big) = E_P\big(\liminf_t \sqrt{M_t}\big) \leq \liminf_t E_P\big(\sqrt{M_t}\big) = \lim_t a_1 a_2 \dots a_t = 0$$

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

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

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

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

Kakutani's theorem and 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

$$\prod_{n=1}^{\infty} E_P\big(\sqrt{z_n(X_n)}\big) > 0$$

$$\iff \sum_{n=1}^{\infty} \Big(1 - E_P\big(\sqrt{z_n(X_n)}\big)\Big) < \infty$$

Exercise 17. *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

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

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 18. *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

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

Chapter 5

Continuous martingales

5.1 Continuous time

Moving from discrete to continuous time, 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 18. *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 19. *We show a filtration which is not right-continuous, generated by a continuous process. Consider the probability space of continuous functions started at zero*

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

equipped with the Borel σ-algebra, where the canonical process is $X_t(\omega) = \omega_t$, Let $(\mathcal{F}_t^0)$ be the "raw" filtraton generated by X, with $\mathcal{F}_t^0 = \sigma(\omega_s : s \leq t)$.

Note that $A \in \mathcal{F}_t^0$ 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

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

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

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

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

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 $(\mathcal{F}_{t+}^0)$ stopping time.

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

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

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 $\{\widetilde{\tau} \leq t\}$ is $\mathcal{F}_{t+}^0$ measurable but not $\mathcal{F}_t^0$ 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 $\mathcal{F}_{t+}^0$

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

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

$\widetilde{\tau}$ 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 15. *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,

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

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 □

We need to extend the results for discrete time martingales to continuous time.

Lemma 19. *Let $\tau(\omega) \in \mathbb{R}^+ \cup \{+\infty\}$ a stopping time with respect to the filtration $\mathbb{F} = (\mathcal{F}_t : t \in \mathbb{R}^+)$.*

There is a sequence of stopping times $(\tau_n(\omega) : n \in \mathbb{N})$ where each τ_n takes finitely many values and $\tau_n(\omega) \geq \tau(\omega)$, approximating τ from above:

$$\tau_n(\omega) \downarrow \tau(\omega) \qquad \forall \omega \text{ as } n \uparrow \infty .$$

Proof: Define

$$\tau_n(\omega) = \begin{cases} +\infty & \text{if } \tau(\omega) \geq n \\ (k+1)/n & \text{otherwise, for } \tau(\omega) \in \big[k/n, (k+1)/n\big), \quad k \in \mathbb{N} \end{cases}$$

You see that τ_n is a $\mathbb{F}$-stopping time:

$$\{\omega : \tau_n(\omega) \leq t\} = \{\omega : \tau(\omega) \leq \lfloor tn \rfloor / n\} \in \mathcal{F}_{\lfloor tn \rfloor / n} \subseteq \mathcal{F}_t \quad \forall t \geq 0$$

where $\lfloor x \rfloor$ is the largest integer smaller than x.

Remark 11. *Note that corresponding random time approximating the $\mathbb{F}$-stopping time τ from below*

$$\widehat{\tau}_n(\omega) = \begin{cases} n & \text{if } \tau(\omega) \geq n \\ k/n & \text{otherwise, for } \tau(\omega) \in \big[k/n, (k+1)/n\big), \quad k \in \mathbb{N} \end{cases}$$

is not always a stopping time.

Definition 31. *A random time $\sigma(\omega) \in (\mathbb{R}^+ \cup \{+\infty\})$ is $\mathbb{F}$-predictable is there is an announcing sequence of $\mathbb{F}$-stopping times (τ_n) approximating σ from below*

$$\tau_n(\omega) \uparrow \sigma(\omega), \quad \forall \omega$$

and

$$\tau_n(\omega) < \tau(\omega) \quad \text{ on the set } \{\omega : \tau(\omega) > 0\}$$

Lemma 20. *A $\mathbb{F}$-predictable time is a $\mathbb{F}$-stopping time.*

Proof: $\forall t, \quad \big\{\omega : \sigma(\omega) \leq t\big\} = \bigcap_{n \in \mathbb{N}} \big\{\omega : \tau_n(\omega) \leq t\big\} \in \mathcal{F}_t.$

Lemma 21. *(Regularization) Let $(X_t : t \in \mathbb{Q}+)$ is a $\mathbb{F}$-submartingale, with $\mathbb{F} = (\mathcal{F}_t : t \in \mathbb{Q}^+)$. We can replace $\mathbb{Q}^+$ by any countable set dense in $\mathbb{R}^+$.*

Then P almost surely the left and right limits

$$X_{t-}(\omega) := \lim_{q \uparrow t, q \in \mathbb{Q}^+} X_q(\omega), \quad X_{t+}(\omega) := \lim_{q \downarrow t, q \in \mathbb{Q}^+} X_q(\omega)$$

exist simultaneously for all $t \in \mathbb{R}^+$.

Proof: It is enough to prove the lemma in a finite interval $[0,T]\cap\mathbb{Q}^+$, with $T\in Q$.

Let F_n a non-decreasing sequence of finite sets with $F_n\subseteq F_{n+1}$ and

$$\bigcup_{n\in\mathbb{N}} F_n = ([0,T]\cap\mathbb{Q}^+)$$

For each finite set F_n, $(X_q : q\in F_n)$ is a submartingale in the filtration $(\mathcal{F}_q : q\in F_n)$.

Define for $a<b\in\mathbb{R}$ the number of downcrossings of $[a,b]$ by $X(\omega)$

$$D_{[a,b]}\big(X_q(\omega) : q\in\mathbb{Q}\cap[0,T]\big) := \sup_F D_{[a,b]}\big(X_q(\omega) : q\in F\big)$$

where the supremum is over finite subsets $F\subseteq[0,T]\cap\mathbb{Q}^+$.

Note that for each finite F, $F\subseteq F_n$ for n large enough, therefore

$$D_{[a,b]}\big(X_q(\omega) : q\in F_n\big)\uparrow D_{[a,b]}\big(X_q(\omega) : q\in\mathbb{Q}\cap[0,T]\big)\quad\text{as } n\uparrow\infty,\ \forall\omega$$

By Doob submartingale inequality in discrete time , $\forall n$

$$E(D_{[a,b]}\big(X_q(\omega) : q\in F_n\big)\le\frac{E(X_T^+)+b^-}{b-a}\le\frac{E(|X_T|)+b^-}{b-a}<\infty$$

Therefore by monotone convergence, P almost surely

$$E\bigg(D_{[a,b]}\big(X_q(\omega) : q\in\mathbb{Q}\cap[0,T]\big)\bigg)<\infty \text{ and } D_{[a,b]}\big(X_q(\omega) : q\in\mathbb{Q}\cap[0,T]\big)<\infty\quad\forall a<b\in\mathbb{Q},$$

which means that P a.s. left and right limits exist simultaneously for all $t\in[0,T]$, and since $\mathbb{R}^+$ is covered by countably many finite intervals it holds also P a.s. simultaneously for all $t\in\mathbb{R}^+$ $\square$.

Remark 12. *Although the submartingale (X_q) was defined only on $\mathbb{Q}^+$, we can use the existence of the limit to redefine outside a P-null a version of the process which is right continuous at all $t\in\mathbb{R}^+$.*

In order to have adaptedness for the redefined process we need to work with the right continuous filtration completed by the P-null sets.

Lemma 22. *Let $D^+=\{k2^{-n} : k,n\in\mathbb{N}\}$ be the dyadic set (or another countable set dense in $\mathbb{R}^+$).*

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

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

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

Proof By definition, $(\mathcal{F}_t)_{t\in\mathbb{R}^+}$ is right continuous.

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 $(\widehat{\mathcal{F}}_{-n})$-martingale, and by Doob's backward convergence theorem (11) and since (M_{u_n}) is continuous on the dyadics, define

$$\begin{aligned} M_t(\omega) &:= \limsup_{n\to\infty} M_{u_n}(\omega) \quad \forall\omega, \\ &= \lim_{n\to\infty} M_{u_n}(\omega) \end{aligned}$$

where by definition M_t is $\mathcal{F}_t$-measurable and in the second equality the limit is P-almost surely and in $L^1(P)$, which implies $M_t \in L^1(P)$.

Let's check the martingale property: for $s, t \in \mathbb{R}$ with $s \le t$, 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 $r_n \le u_n$. Let $A \in \mathcal{F}_s \subseteq \mathcal{F}_{r_n}$, $\forall n$.

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 $(M_u)_{u\in D^+}$ $\square$

Proposition 16. *Doob' optional stopping theorem in continuous time.*

Let $(M_t : t \in [0, +\infty])$ *a right-continuous uniformly integrable* $\mathbb{F}$*-martingale where* $\mathbb{F}$ *is right continuous, and* $0 \le \sigma(\omega) \le \tau(\omega)$ $\mathbb{F}$*-stopping times.*

Then

$$E(M_\tau|\mathcal{F}_\sigma) = M_\sigma(\omega)$$

Proof: There are two non-increasing sequences of stopping times σ_n, τ_n with

$$\sigma(\omega) \le \sigma_n(\omega) \le \tau_n(\omega), \quad \tau(\omega) \le tau_n(\omega)$$

which for each fixed n take values in the dyadics $D_n = (k2^{-n} : k \in \mathbb{N})$ and

$$\sigma_n(\omega) \downarrow \sigma(\omega), \quad \sigma_n(\omega) \downarrow \tau_n(\omega) \quad \text{ as } n \uparrow \infty$$

To do this simply take

$$\begin{aligned} \tau_n(\omega) &:= (k+1)2^{-n} \text{ otherwise, for } \tau(\omega) \in \big[k2^{-n}, (k+1)2^{-n}\big), \quad k \in \mathbb{N} \\ \sigma_n(\omega) &= (k+1)2^{-n} \text{ otherwise, for } \sigma(\omega) \in \big[k2^{-n}, (k+1)2^{-n}\big), \quad k \in \mathbb{N} \end{aligned}$$

and $\tau_n(\omega) = +\infty$ and $\sigma_n(\omega) = +\infty$ when $\tau(\omega) = +\infty$ and $\sigma(\omega) = +\infty$ respectively.

and check that they are stopping times.

The backward fitrations $(\mathcal{F}_{\tau_n} : n \in N)$, $(\mathcal{F}_{\tau_n} : n \in N)$, are non-increasing as $n \to \infty$.

Therefore we apply Doob's backward convergence theorem,

$$M_{\tau_n}(\omega) \to M_\tau(\omega) \quad \text{and} \quad M_{\sigma_n}(\omega) \to M_\sigma(\omega)$$

not just P-almost surely (which is implied by the right continuity) but also in $L^1(P)$

For every fixed n, by the discrete time version of the optional sampling theorem with the filtration $(\mathcal{F}_d : d \in D_n)$ under the uniform integrability assumption

$$E_P(M_{\tau_n}|\mathcal{F}_{\sigma_n})(\omega) = M_{\sigma_n}(\omega)$$

Let $A \in \mathcal{F}_\sigma \subseteq \mathcal{F}_{\sigma_n} \subseteq \mathcal{F}_\tau \subseteq \mathcal{F}_{\tau_n}$.

$$E\big(M_\tau \mathbf{1}_A\big) = \lim_{n\to\infty} E\big(M_{\tau_n}\mathbf{1}_A\big) = \lim_{n\to\infty} E\big(M_{\sigma_n}\mathbf{1}_A\big) = E\big(M_\sigma \mathbf{1}_A\big)$$

where we used the convergence in $L^1(P)$ to take the limit in and out of the expectation.

Proposition 17. *Let (M_t) a right continuous martingale in the right continuous filtration $\mathbb{F}$, and $\tau(\omega)$ a $\mathbb{F}$-stopping time. Then the stopped process*

$$M_t^\tau(\omega) = M_{t\wedge\tau}(\omega) := M_t(\omega)\mathbf{1}(\tau(\omega) > t) + M_\tau(\omega)\mathbf{1}(\tau(\omega) \le t)$$

is a $\mathbb{F}$-martingale.

Proof Since τ is a stopping time it follows that $(M_{t\wedge\tau})$ is $\mathbb{F}$-adapted. Let's fix $0 \le s \le t < \infty$. Now in a finite interval $(M_s : s \le t)$ is uniformly integrable, and by Doob's optional stopping theorem applied to the bounded stopping times $(s\wedge\tau) \le (t\wedge\tau) \le t$,

$$E(M_{t\wedge\tau}|\mathcal{F}_{s\wedge\tau})(\omega) = M_{s\wedge\tau}$$

Next we show that

$$E(M_{\tau\wedge t}|\mathcal{F}_s) = M_\tau \mathbf{1}(\tau \le s) + E(M_{\tau\wedge t}|\mathcal{F}_{\tau\wedge s})\mathbf{1}(\tau > s)$$

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

$$E(M_{\tau\wedge t}\mathbf{1}_A) = E(M_\tau \mathbf{1}_A \mathbf{1}(\tau \le s)) + E(M_{\tau\wedge t}\mathbf{1}_A\mathbf{1}(\tau > s))$$

Note that $A \cap \{\tau > s\}$ is not only $\mathcal{F}_s$ measurable but also $\mathcal{F}_{\tau\wedge s}$ measurable since by definition for all $r \ge 0$

$$A \cap \{\tau > s\} \cap \{\tau\wedge s \le r\} = \begin{cases} \emptyset \in \mathcal{F}_s & \text{if } s > r \\ A \cap \{\tau > s\} \in \mathcal{F}_s & \text{if } s \le r \end{cases}$$

Therefore by taking conditional expectation w.r.t. $\mathcal{F}_{\tau\wedge s}$ inside the expectation we get

$$E(M_{\tau\wedge t}\mathbf{1}_A) = E\Big(\big(M_\tau \mathbf{1}(\tau \le s) + E(M_{\tau\wedge t}|\mathcal{F}_{t\wedge s})\mathbf{1}(\tau > s)\big)\mathbf{1}_A\Big)$$
$$= E\Big(\big(M_\tau \mathbf{1}(\tau \le s) + M_{\tau\wedge s}\mathbf{1}(\tau > s)\big)\mathbf{1}_A\Big) = E\big(M_{\tau\wedge s}\mathbf{1}_A\big)$$

which means

$$E(M_{t\wedge\tau}|\mathcal{F}_s)(\omega) = M_{s\wedge\tau}(\omega)$$

5.2 Localization

Definition 32. *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 and $X_0(\omega) = 0$, is locally bounded, with localizing sequence

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

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

5.3 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 $\Pi = (0 = t_0 \le t_1 \le \ldots, \le t_n = t)$ 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 23. *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 $(0 = t_0 \le t_1 \le t_2 \le \ldots)$, with $t_n \to \infty$. We have

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

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,

$$E_P(M_t^2) = 2E_P\left(\lim_{\Delta(\Pi)\to 0} \sum_{i=1}^{\infty} M_{t_{i-1}\wedge t}(M_{t_i\wedge t} - M_{t_{i-1}\wedge t}) \right)$$

$$= 2 \lim_{\Delta(\Pi)\to 0} 2E_P\left(\sum_{i=1}^{\infty} M_{t_{i-1}\wedge t}(M_{t_i\wedge t} - M_{t_{i-1}\wedge t}) \right) =$$

$$\lim_{\Delta(\Pi)\to 0} 2\sum_{i=1}^{\infty} E_P\Big(M_{t_{i-1}\wedge t} E_P(M_{t_i\wedge t} - M_{t_{i-1}\wedge t} | \mathcal{F}_{t_{i-1}\wedge t}) \Big) = 0$$

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, $E\big(M^2_{t\wedge\sigma_n}\big) = 0$ implies $M_{t\wedge\sigma_n} = 0$ P a.s,

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

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

Lemma 24. *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

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

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

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

Lemma 25. *Suppose $A_n^{(1)}$ and $A_n^{(2)}$ are two predictable processes satisfying the hypothesis of lemma 24 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 a constant $c > 0$ such that

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

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

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

where we used Cauchy-Schwartz inequality together with lemma 24.

Let $M_n := E_P(B_\infty|\mathcal{F}_n)$, $X_n := (M_n - B_n)$, satisfying

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

By Doob's L^p martingale maximal inequality

$$E\bigg(\sup_{n\in\mathbb{N}} X_n^2\bigg) \le E_P\bigg(\sup_{n\in\mathbb{N}} N_n^2\bigg) \le 4E_P(N_\infty^2) \le 4E_P(Y^2)$$

and

$$E\bigg(\sup_{n\in\mathbb{N}} M_n^2\bigg) \le 4E\bigg(M_\infty^2\bigg) = 4E(B_\infty^2)$$

Since $\sup_n |B_n| \le \sup_n |X_n| + \sup_n |M_n|$, by the inequality $(a+b)^2 \le 2(a^2+b^2)$

$$E(\sup_n B_n^2) \le 2\Big\{ E(\sup_n X_n^2) + E(\sup_n M_n^2)\Big\} \le 8\Big(E(Y^2) + E(B_\infty^2)\Big)$$
$$\le 8\Big(E(Y^2) + 2^{5/2} C E_P(Y^2)^{1/2}\Big) \quad \square$$

Theorem 19. *Suppose $(X_t : t \in \mathbb{R}^+)$ is a $(\mathcal{F}_t)$-submartingale 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 : The result holds also for continuous local submartingales (the localizing sequence is obtained by taking minimum of localizing sequences). It is also extended to processes with jumps.

Proof, Uniqueness: From the Bass, *Probabilistic techniques in analysis* .

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

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

is a continuous local martingale starting from 0 with paths of finite variation, and by lemma 23 it is constant P-almost surely.

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

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

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's submartingale decomposition

$$X_{k2^{-m}}(\omega) = M_k^{(m)} + A_k^{(m)}$$

In continuous time we define for each m piecewise constant filtrations

$$\overline{\mathcal{F}}_t^{(m)}(\omega) = \mathcal{F}_{k2^{-m}}(\omega) \quad \text{when } (k-1)2^{-m} < t \le k2^{-m}$$

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 $\overline{A}^{(m)}_t$ converges in $L^2(P)$ uniformly in t as $m\to\infty$.

For $m>n$, $\overline{A}^{(m)}_t$ and $\overline{A}^{(n)}_t$ are constant on the intervals $\big((k-1)2^{-m}, k2^{-m}\big]$, we have

$$\sup_t |\overline{A}^{(m)}_t - \overline{A}^{(n)}_t| = \sup_{k\in\mathbb{N}} |\overline{A}^{(m)}_{k2^{-m}} - \overline{A}^{(n)}_{k2^{-m}}|$$

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

$$E_P(\overline{A}^{(m)}_\infty - \overline{A}^{(m)}_t|\overline{\mathcal{F}}^{(m)}_t)(\omega) = E_P(A^{(m)}_\infty - A^{(m)}_k|\mathcal{F}_{k2^{-m}})(\omega) = E_P(X_\infty - X_t|\mathcal{F}_{k2^{-m}})(\omega) = E_P(X_\infty - X_t|\mathcal{F}_t)(\omega)$$

On the other hand

$$E_P(\overline{A}^{(n)}_\infty - \overline{A}^{(n)}_t|\overline{\mathcal{F}}^{(m)}_t)(\omega) = E_P(A^{(n)}_\infty - A^{(n)}_l|\mathcal{F}_t)(\omega) = E_P\left(E_P(A^{(n)}_\infty - A^{(n)}_l|\mathcal{F}_u)\Big|\mathcal{F}_t\right)(\omega) = E_P\left(E_P(X_\infty - X_u|\mathcal{F}_u)\Big|\mathcal{F}_t\right)(\omega) = E_P\big(X_\infty - X_u\big|\mathcal{F}_t\big)(\omega)$$

Then the difference of conditional expectations is bounded:

$$\begin{aligned}&\left|E_P(\overline{A}^{(m)}_\infty - \overline{A}^{(m)}_t|\mathcal{F}_t) - E_P(\overline{A}^{(n)}_\infty - \overline{A}^{(n)}_t|\mathcal{F}_t)\right|\\ &\le E_P\Big(|X_t - X_u|\big|\mathcal{F}_t\big) \le E_P\big(W(2^{-n})\big|\mathcal{F}_t\big)\end{aligned}$$

The assumptions of lemma 25 are satisfied, giving

$$E_P\left(\sup_t\big(\overline{A}^{(m)}_t - \overline{A}^{(n)}_t\big)^2\right) \le c\left\{E_P\big(W(2^{-n})^2\big) + 2CE_P\big(W(2^{-n})^2\big)^{1/2}\right\} \to 0 \quad \text{as } n\to\infty,\ m>n$$

We show the space of processes

$$\mathcal{S}_2 := \left\{Z(t,\omega)\ (\mathcal{F}_t)\text{-adapted with } \| Z \|^2_{\mathcal{S}_2} := E_P\left(\sup_t Z_t^2\right) < \infty\right\}$$

is complete under the $\|\cdot\|_{\mathcal{S}_2}$ norm.

Suppose $(Z^{(n)}_t : t\ge 0, n\in\mathbb{N})$ is a Cauchy sequence in $\mathcal{S}_2$. In particular there exists a sequence (N_k) with

$$E\left(\sup_t\big(Z^{(n)}_t - Z^{(m)}_t\big)^2\right) \le 2^{-k}, \quad \forall n,m\ge N_k$$

For each t define

$$Z_t^{(\infty)} = Z_t^{(N_0)} + \sum_{k=0}^{\infty} \big(Z_t^{(N_{k+1})}(\omega) - Z_t^{(N_k)}(\omega) \big)$$

where $\forall t$ the series converges in $L^2(\Omega, \mathcal{F}_t, P)$. Then by triangle inequality

$$\| Z^{(\infty)} - Z^{(m)} \|_{\mathcal{S}_2} = E\bigg(\sup_t \big(Z_t^{(\infty)} - Z_t^{(m)} \big)^2 \bigg)^{1/2}$$

$$\le E\bigg(\sup_t \big(Z_t^{(m)} - Z_t^{(N_k)} \big)^2 \bigg)^{1/2} + E\bigg(\sup_t \big(Z_t^{(\infty)} - Z_t^{(N_k)} \big)^2 \bigg)^{1/2} \le 2^{-k/2} + \sqrt{\sum_{h=k}^{\infty} 2^{-h}}$$

which is arbitrarily small for $m \ge N_k$ and k large enough.

By completeness, there is a $(\mathcal{F}_t)$-adapted process $A_t(\omega) \in S_2$ with

$$E_P\bigg(\sup_t \big\{ \overline{A}_t^{(n)} - A_t \big\}^2 \bigg) \to 0$$

From convergence in quadratic mean it follows that there is a subsequence (n_i) such that

$$\sup_t |\overline{A}_t^{(n_i)}(\omega) - A_t(\omega)| \to 0 \quad P\text{-almost surely .}$$

Next we show that $A_t(\omega)$ is continuous. For $t = k2^{-n}$,

$$\Delta \overline{A}_t^n = E_P\bigg(X_{(k)2^n} - X_{(k-1)2^n} \bigg| \mathcal{F}_{(k-1)2^{-n}} \bigg) \le E_P\big(W(2^{-n}) \big| \mathcal{F}_{(k-1)2^{-n}} \big)$$

where on the right hand side we have an uniformly integrable martingale. We have

$$E_P\bigg(\sup_t (\Delta \overline{A}_t^n)^2 \bigg) \le E_P\bigg(\sup_k E_P\big(W(2^{-n}) \big| \mathcal{F}_{(k-1)2^{-n}} \big)^2 \bigg) \le 4 E_P\bigg(W(2^{-n})^2 \bigg) \to 0 \quad \text{as } n \to \infty$$

by Doob L^p-martingale inequality. In particular there is a further subsequence (n_j) such that

$$\sup_t \Delta \overline{A}_t^{n_j}(\omega) \to 0 \qquad P\text{- almost surely as } j \to \infty$$

Almost sure continuity follows:

$$\sup_t |\Delta A_t(\omega)| \le \sup_t |\Delta A_t(\omega) - \Delta A_t^{(n_j)}(\omega)| + \sup_t |\Delta A_t^{(n_j)}(\omega)|$$

$$\le 2 \sup_t |A_t(\omega) - A_t^{(n_j)}(\omega)| + \sup_t |\Delta A_t^{(n_j)}(\omega)|$$

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 22 it is enough to show the martingale property for $s < t$ with $s, t \in D_N = \{k2^{-N} : k \in \mathbb{Z}\}$, and $B \in \mathcal{F}_s$:

$$E_P\big((M_t - M_s)\mathbf{1}_B\big) = E\big((X_t - X_s)\mathbf{1}_B\big) - E\big((A_t - A_s)\mathbf{1}_B\big)$$
$$= E\big((X_t - X_s)\mathbf{1}_B\big) - E\big((A_t^{(n)} - A_s^{(n)})\mathbf{1}_B\big) + E((A_t - A_t^{(n)})\mathbf{1}_B) - E\big((A_s - A_s^{(n)})\mathbf{1}_B\big)$$
$$= 0 + E((A_t - A_t^{(n)})\mathbf{1}_B) - E\big((A_s - A_s^{(n)})\mathbf{1}_B\big) \to 0 \text{ as } n \to \infty$$

where the last identity holds $\forall n \geq N$ by the discrete time martingale property, and by the Cauchy-Schwartz inequality,

$$\left| E_P\big((\overline{A}_t^{(n)} - A_t)\mathbf{1}_B\big) \right| \leq E_P\left(\sup_t (\overline{A}_t^{(n)} - A_t)^2\right)^{1/2} \sqrt{P(B)} \longrightarrow 0.$$

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 $(\tau_C : C \in \mathbb{N})$ □

Remark 13. *Note that without additional assumptions, it is not possible to show that M_t is a true martingale: for $t > s$ and $B \in \mathcal{F}_s$*

$$E_P\big((M_t - M_s)\mathbf{1}_B\big) = E_P\big(\lim_{C\to\infty}(M_{t\wedge\tau_C} - M_{s\wedge\tau_C})\mathbf{1}_B\big) \tag{5.1}$$
$$\stackrel{?}{=} \lim_{C\to\infty} E_P\big((M_{t\wedge\tau_C} - M_{s\wedge\tau_C})\mathbf{1}_B\big) = 0 \tag{5.2}$$

the interchange of limit and expectation is not always justified.

Definition 33. *1. the right continuous adapted process $(X_t(\omega))$ is in the class D (D is for Doob) is the family of random variables*

$$\Big\{ X_\tau(\omega) : \tau \textit{ is a stopping time} \Big\}$$

is uniformly integrable.

2. We say that a right continuous $(\mathcal{F}_t)$-adapted process $(X_t(\omega))$ is in the class DL (local Doob) if for each $t > 0$ the family of random variables

$$\Big\{ X_\tau(\omega) : \tau \textit{ is a stopping time with } \tau(\omega) \leq t \textit{ a.s.} \Big\}$$

is uniformly integrable,

Exercise 20. *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 14. *A continuous $(\mathcal{F}_t)$-submartingale 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 5.1.

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. $\square$.

5.4 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 34. *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 18. *$\langle M, \widetilde{M}\rangle_t$ is the unique continuous process of finite (total) variation such that $\langle M, \widetilde{M}\rangle_0 = 0$ and*

$$M_t\widetilde{M}_t = M_0\widetilde{M}_0 + \widehat{N}_t + \langle M, \widetilde{M}\rangle_t \tag{5.3}$$

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 (5.3) with $\widehat{N}_t = \big(N_t^{(+)} - N_t^{(-)}\big)/4$ □

Exercise 21. *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 $\widetilde{B}_t$ are square integrable martingales.

$$
\begin{aligned}
&E_P\big(B_t\widetilde{B}_t - B_s\widetilde{B}_s\big|\mathcal{F}_s\big) \\
&= B_s E_P\big(\widetilde{B}_t - \widetilde{B}_s\big|\big|\mathcal{F}_s\big) + \widetilde{B}_s E_P\big(B_t - B_s\big|\big|\mathcal{F}_s\big) + E_P\big((B_t - B_s)(\widetilde{B}_t - \widetilde{B}_s)\big|\mathcal{F}_s\big) = \\
&B_s E_P\big(\widetilde{B}_t - \widetilde{B}_s\big) + \widetilde{B}_s E_P\big(B_t - B_s\big) + E_P\big((B_t - B_s\big) E_P\big(\widetilde{B}_t - \widetilde{B}_s\big) = 0
\end{aligned}
$$

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

$$
\begin{aligned}
&E_P\big(B_t W_t - B_s W_s\big|\mathcal{F}_s\big) \\
&= B_s E_P\big(W_t - W_s\big|\big|\mathcal{F}_s\big) + \widetilde{W}_s E_P\big(W_t - W_s\big|\big|\mathcal{F}_s\big) + E_P\big((B_t - B_s)(W_t - W_s)\big|\mathcal{F}_s\big) \\
&= 0 + \sqrt{\alpha} E_P\big((B_t - B_s)^2\big|\mathcal{F}_s\big) + \sqrt{(1-\alpha)} E_P\big((B_t - B_s)(\widetilde{B}_t - \widetilde{B}_s)\big|\mathcal{F}_s\big) \\
&= \sqrt{\alpha}(\langle B\rangle_t - \langle B\rangle_s) = \sqrt{\alpha}(t-s)
\end{aligned}
$$

It follows that $\langle B, W\rangle_t = \sqrt{\alpha}\langle B\rangle_t = \sqrt{\alpha} t$

Theorem 20. *Let M be a continuous martingale with $|M_t(\omega)| \leq C < \infty \; \forall t > 0$. 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

$$\Delta = (0 \leq t_0 < t_1 < t_n \dots), \quad |\Delta| := \sup_i (t_i - t_{i-1})$$

$[M]_t$ 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 From Revuz-Yor *Continuous martingales and Brownian motion.*

Without loss of generality we assume $M_0 = 0$, otherwise consider $M_t = (M_t - M_0)$. Lets denote

$$T_t^{\Delta}(M) := \sum_{k=1}^{\infty} (M_{t\wedge t_k} - M_{t\wedge t_{k-1}})^2 \tag{5.4}$$

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

$$E\big((M_t - M_s)^2 | \mathcal{F}_s\big) = E\big(M_t^2 - M_s^2 | \mathcal{F}_s\big) \tag{5.5}$$

and for $s = s_0 < s_1 < \cdots < s_n = t$, it equals

$$= \sum_{k=1}^{n} E\big(M_{s_k}^2 - M_{s_{k-1}}^2 | \mathcal{F}_s\big) = \sum_{k=1}^{n} E\big(\{M_{s_k} - M_{s_{k-1}}\}^2 | \mathcal{F}_s\big) = E(T_t^{\Delta}(M) - T_s^{\Delta}(M) | \mathcal{F}_s)$$

In particular for fixed partitions Δ, Δ'

$$X_t^{\Delta,\Delta'} := T_t^{\Delta}(M) - T_t^{\Delta'}(M)$$

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 5.4 replacing the martingale M_t with the martingale X_t. (We don't need to write the explicit expression).

From 5.5 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

$$E(X_t^2) = E(T_t^{\Delta\Delta'}(X)) \le 2E_P\bigg(T_t^{\Delta\Delta'}(T^{\Delta}(M)) + T_t^{\Delta\Delta'}(T^{\Delta'}(M))\bigg)$$

We show that $E_P\bigg(T_t^{\Delta\Delta'}(T^{\Delta}(M))\bigg) \to 0$.

Let $s_k \in \Delta\Delta'$, $t_l \in \Delta$ such that $t_l \le s_k < s_{k+1} \le t_{l+1}$.

$$T_{s_{k+1}}^{\Delta}(M) - T_{s_k}^{\Delta}(M) = (M_{s_{k+1}} - M_{t_l})^2 - (M_{s_k} - M_{t_l})^2$$
$$= (M_{s_{k+1}} - M_{s_k})^2 + 2(M_{s_{k+1}} - M_{s_k})(M_{s_k} - M_{t_l}) = (M_{s_{k+1}} + M_{s_k} - 2M_{t_k})(M_{s_{k+1}} - M_{s_k})$$

and assuming that $t = s_n \in \Delta\Delta'$

$$T_t^{\Delta\Delta'}(T^{\Delta}(M)) = \sum_{k=0}^{n-1} \big(T_{s_{k+1}}^{\Delta}(M) - T_{s_k}^{\Delta}(M)\big)^2$$
$$\le \sup_{k \le n} (M_{s_{k+1}} + M_{s_k} - 2M_{t_k})^2 \sum_{k=0}^{n-1} (M_{s_{k+1}} - M_{s_k})^2$$
$$= \sup_{k \le n} (M_{s_{k+1}} + M_{s_k} - 2M_{t_k})^2 T_t^{\Delta\Delta'}(M)$$

By taking expectation and using Cauchy-Schwartz inequality

$$E_P\bigg(T_t^{\Delta\Delta'}(T^{\Delta}(M))\bigg) \le E_P\bigg(\sup_{k \le n}(M_{s_{k+1}} + M_{s_k} - 2M_{t_k})^4\bigg)^{1/2} E_P\big(\{T_t^{\Delta\Delta'}(M)\}^2\big)^{1/2}$$

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 $|M_t(\omega)| \le C$, convergence in $L^p(\Omega)$ follows as well.

In order to complete the proof we show that

$$E_P\big(\{T_t^{\Delta}(M)\}^2\big)$$

remains bounded as $|\Delta| \to 0$.

Assuming that $t = t_n \in \Delta$, denoting $\Delta M_k = (M_{t_k} - M_{t_{k-1}})$

$$\{T_t^{\Delta}(M)\}^2 = \sum_{k=1}^{n}(\Delta M_k)^4 + 2\sum_{k=1}^{n}\Big(\sum_{j>k}^{n}(\Delta M_j)^2\Big)(\Delta M_k)^2,$$

$$E_P\Big(\{T_t^{\Delta}(M)\}^2\Big) \le E_P\Big(T_t^{\Delta}(M)\sup_{k\le n}(\Delta M_k)^2\Big) + 2\sum_{k=1}^{n} E_P\Big((M_t - M_{t_k})^2(\Delta M_k)^2\Big)$$

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

$$E_P\Big(\{T_t^{\Delta}(M)\}^2\Big) \le E_P\Big(T_t^{\Delta}(M)\sup_{k\le n}\Big\{(\Delta M_k)^2 + 2(M_t - M_{t_k})^2\Big\}\Big)$$
$$\le E_P(T_t^{\Delta}(M))12C^2 = E_P(M_t^2)12C^2 \le 12C^4$$

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

$$E_P\Big(\sup_{s\le t}\big(T_s^{\Delta_n}(M) - T_s^{\Delta_k}(M)\big)^2\Big) \le 4E_P\Big(\big(T_t^{\Delta_n}(M) - T_t^{\Delta_k}(M)\big)^2\Big)$$

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

$$E_P\Big(\sup_{s\le t}\big([M]_s - T_s^{\Delta_n}(M)\big)^2\Big) \to 0$$

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

$$\sup_{s\le t}\big|[M]_s - T_s^{\Delta_{n(j)}}(M)\big| \to 0 \quad P\text{-almost surely .}$$

It follows that $[M]_s$ is non-decreasing since $T_s^\Delta(M)$ with $\Delta = \Delta_{n(j)}$ is non-decreasing.

Since the approximating processes $T_s^\Delta(M)$ with $\Delta = \Delta_{n(j)}$Âăare 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 $s \le t, \; A \in \mathcal{F}_s$

$$E_P\Big((M_t^2 - M_s^2)\mathbf{1}_A \Big) = E_P\Big((T_t^\Delta(M) - T_s^\Delta(M))\mathbf{1}_A \Big) \to E_P\Big(([M]_t - [M]_s))\mathbf{1}_A \Big)$$

as $\Delta \to 0$, 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!) $\Box$.

Remark 14.

$$[M]_t = \lim_{|\Delta| \to 0} \sum_{t_i \in \Delta} \big(M_{t_i \wedge t} - M_{t_{i-1} \wedge t} \big)^2$$
$$\langle M \rangle_t = \lim_{|\Delta| \to 0} \sum_{t_i \in \Delta} E\Big(\big(M_{t_i \wedge t} - M_{t_{i-1} \wedge t} \big)^2 \Big| \mathcal{F}_{t_{i-1}} \Big)$$

where the limit are taken in probability. These coincide when M is a continuous square integrable martingale but are different when M_t has jumps.

Corollary 15. *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 $\widetilde{M}_t$,*

$$[M, \widetilde{M}]_t = \lim_{|\Delta| \to 0} \sum_{k=1}^{\infty} \big(M_{t \wedge t_k} - M_{t \wedge t_{k-1}} \big) \big(\widetilde{M}_{t \wedge t_k} - \widetilde{M}_{t \wedge t_{k-1}} \big)$$

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 $|M_t^{\tau_n}| \le n$.

$N_t^{(n)} = \big(M_{t \wedge \tau_n}^2 - [M^{\tau_n}]_t \big)$ 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:

$$P\Big(\big|[M]_t - T_t^{\Delta}(M)\big| > \varepsilon \Big) =$$
$$P\Big(\{\tau_n \le t\} \bigcap \Big\{ \big|[M]_t - T_t^{\Delta}(M)\big| > \varepsilon \Big\}\Big) + P\Big(\{\tau_n > t\} \bigcap \Big\{ \big|[M]_{t\wedge\tau_n} - T_{t\wedge\tau_n}^{\Delta}(M)\big| > \varepsilon \Big\}\Big)$$
$$\le P\big(\tau_n \le t\big) + P\Big(\big|[M^{\tau_n}]_t - T_t^{\Delta}(M^{\tau_n})\big| > \varepsilon \Big)$$

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 $\square$.

Lemma 26. *Let $(M_t(\omega) : t \in \mathbb{N}) \subseteq L^2(P)$ a square integrable $\mathbb{F}$-martingale. The following conditions are equivalent:*

1. *$(M_t : t \in \mathbb{N})$ is bounded in $L^2(P)$, that is*

$$\sup_{t\in\mathbb{N}} E_P(M_t^2) < \infty$$

2.

$$\sum_{t=1}^{\infty} E\big((M_t - M_{t-1})^2\big) < \infty$$

3. *there is a r.v. $M_\infty \in L^2(P)$ such that $M_t = E(M_\infty|\mathcal{F}_t)$ and $M_t \to M_\infty$ in $L^2(P)$.*

Proof. Note that for $s \le t \in \mathbb{N}$, using telescoping sums, by the martingale property

$$E\big((M_t - M_s)^2\big) = E\Bigg(\Big\{ \sum_{n=s+1}^{t} \Delta M_n \Big\}^2\Bigg) = \sum_{n=s+1}^{t} E\big((\Delta M_n)^2\big)$$

For $s = 0$, we see that (1) $\Longleftrightarrow$ (2).

When (1) holds, $(M_t : t \in \mathbb{N})$ is an uniformly integrable martingale and $\exists M_\infty(\omega)$ such that $M_t = E(M_\infty|\mathcal{F}_t)$ and $M_t \to M_\infty$ P-almost surely and in $L^1(P)$. We show that $M_t \to M_\infty$ also in $L^2(P)$.

For $t, N \in \mathbb{N}$,

$$E\big((M_{t+N} - M_t)^2\big) = E\Bigg(\Bigg\{\sum_{s=t}^{t+N} \Delta M_s\Bigg\}^2\Bigg) = \sum_{s=t}^{t+N} E\big((\Delta M_s)^2\big)$$

where when we develop the square by the martingale property the cross terms have zero expectation. For fixed t as $N \to \infty$ by Fatou lemma

$$E\big((M_\infty - M_t)^2\big) \le \sum_{s=t}^{\infty} E\big((\Delta M_s)^2\big) \to 0$$

as $t \to \infty$ by the hypothesis (2).

We see also that

$$0 \le E\big((M_\infty - M_t)^2\big) = E\big((M_{t+N} - M_t)^2\big) + E\big((M_{t+N} - M_\infty)^2\big)$$
$$= \sum_{s=t+1}^{t+N} E\big((\Delta M_s)^2\big) + E\big((M_{t+N} - M_\infty)^2\big) \longrightarrow \sum_{s=t+1}^{\infty} E\big((\Delta M_s)^2\big) + 0$$

The inequality (5.6) is an equality, therefore (3) $\implies$ (1) □

Remark 15. *For continuous-time martingales $(M_t : t \in \mathbb{R}^+)$, we apply the result to the imbedded discrete time martingale $(M_t : t \in \mathbb{N})$.*

Proposition 19. *Let $(M_t : t \in \mathbb{R}^+)$ a continuous martingale with $E(M_t^2) < \infty$ $\forall t \ge 0$.*

Then $(M_t^2 - \langle M \rangle_t : t \in \mathbb{R}^+)$ is a true $\mathbb{F}$-martingale, in particular

$$E(M_t^2) = E(M_0^2) + E(\langle M \rangle_t)$$

By polarization, if $(\widetilde{M}_t : t \in \mathbb{R}^+) \subseteq L^2(P)$ is another continuous martingale, $(M_t\widetilde{M}_t - \langle M, \widetilde{M} \rangle_t : t \in \mathbb{R}^+)$ is a true $\mathbb{F}$-martingale, in particular

$$E(M_t\widetilde{M}_t) = E(M_0\widetilde{M}_0) + E(\langle M, \widetilde{M} \rangle_t)$$

Proof Let $\tau_0 = 0$ and $\tau_n(\omega) = \inf\{t : |M_t(\omega)| > n\}$, with $\tau_n(\omega) \uparrow \infty$ as $n \uparrow \infty$.

For fixed n, $(M_{t\wedge\tau_n} : t \ge 0)$ is a bounded martingale, and $(M^2_{t\wedge\tau_n} - \langle M \rangle_{t\wedge\tau_n} : t \in \mathbb{N})$ is a true martingale by theorem (20).

For fixed t consider the telescopic series

$$M_t(\omega) = M_0 + \sum_{n=1}^{\infty} (M_{t\wedge\tau_n} - M_{t\wedge\tau_{n-1}})$$

By Doob's optional stopping theorem $M_{t\wedge\tau_n} = E(M_t|\mathcal{F}_{t\wedge\tau_n}) \in L^2(P)$ and by lemma (26) applied with respect to the discrete time filtration $(\mathcal{F}_{t\wedge\tau_n} : n \in \mathbb{N})$

$$M_{t\wedge\tau_n} \to M_t \quad \text{in } L^2(P)$$

which implies

$$E(M_t^2) = \lim_{n\to\infty} E(M^2_{t\wedge\tau_n}) = \lim_{n\to\infty} E\big(\langle M \rangle_{t\wedge\tau_n}\big) = E(\langle M \rangle_t)$$

where the last equality follows by monotone convergence. This gives integrability we show the martingale property: for $s \leq t$, $A \in \mathcal{F}_s$,

Since $M^2_{t\wedge\tau_n} \to M^2_t$ in $L^1(P)$,

$$E((M^2_t - M^2_s)\mathbf{1}_A) = \lim_{n\to\infty} E\big((M^2_{t\wedge\tau_n} - M^2_{s\wedge\tau_n})\mathbf{1}_A\big)$$
$$= E\big((\langle M\rangle_{t\wedge\tau_n} - \langle M\rangle_{s\wedge\tau_n})\mathbf{1}_A\big) \to E\big((\langle M\rangle_t - \langle M\rangle_s)\mathbf{1}_A\big)$$

where we use monotone convergence again □

Remark The $L^2(P)$-isometry $E((M_t - M_0)^2) = E(\langle M\rangle_t)$ is the key step in the construction of the Ito integral.

Chapter 6

Ito calculus

6.1 Ito-isometry and stochastic integral

Proposition 20. *Let $\mathcal{M}^2$ be the space of continuous martingales $(M_t(\omega) : t \in \mathbb{R}^+)$ which are 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

$$(M, N)_{\mathcal{M}^2} := E_P\big(M_\infty N_\infty\big) = E_P\big(\langle M, N \rangle_\infty\big)$$

$$E_P\left(\sup_{t\geq 0} M_t^2\right)^{1/2} \leq 2 \| M \|_{\mathcal{M}^2}$$

by Doob's L^p martingale inequality

Proof When

$$\sup_{t\geq 0} E_P(M_t^2) < \infty$$

by lemma (26) $M_t \rightarrow M_\infty$ P-almost surely and in $L^2(P)$.

We show that $\mathcal{M}^2$ is complete.

If $(M^{(n)})_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathcal{M}^2$, then $(M^{(n)}_\infty)_{n\in\mathbb{N}}$ is a Cauchy sequence in the complete space $L^2(\Omega)$, and there is $M_\infty \in L^2(\Omega)$ such that $E_P\big((M^{(n)}_\infty - M_\infty)^2\big) \rightarrow 0$.

Define $M_t(\omega) := E_P(M_\infty|\mathcal{F}_t)(\omega)$, it follows that $M^{(n)} \rightarrow M \in \mathcal{M}^2$, equivalently

$$E_P\left(\sup_{t\geq 0}(M_t - M_t^{(n)})^2\right) \rightarrow 0$$

In particular there is a subsequence (n_j) such that for P-almost all ω

$$\sup_{t\geq 0}\big|M_t^{(n_j)}(\omega) - M_t(\omega)\big| \rightarrow 0$$

which implies that P-almost surely the path $t \mapsto M_t(\omega)$ is continuous. $\Box$.

Definition 35. *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 random 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 $0 \le a_1 < b_1 \le a_2 < b_2 \le \cdots < b_{n-1} \le a_n < b_n < \infty$ and $\eta_i(\omega)$ is $\mathcal{F}_{a_i}$-measurable.

Definition 36. *Given the filtration $\mathbb{F} = (\mathcal{F}_t)_{t\in\mathbb{R}}$, consider the measurable space $\Omega\times\mathbb{R}^+$ equipped with the predictable σ-algebra $\mathcal{P}$ generated by the left continuous $\mathbb{F}$-adapted processes.*

Exercise: the simple left-continuous $\mathbb{F}$-adapted processes generate also $\mathcal{P}$.

When $(\omega,t) \mapsto Y_t(\omega)$ is $\mathcal{P}$-measurable, we say that the process Y is $(\mathcal{F}_t)$-predictable.

Lemma 27. *Let $(M_t) \in \mathcal{M}^2$ a continuous martingale, and $Y_t \in \mathcal{S}$ a bounded simple predictable process with representation 6.1. 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 $L^2_a\big(\Omega\times\mathbb{R}^+, P(d\omega)\otimes \langle M\rangle(\omega,dt)\big)$ 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 $(N_t) \in \mathcal{M}^2$,

$$\langle (Y\cdot M), N\rangle_t := \int_0^t Y_s d\langle M,N\rangle_s := \sum_{i=1}^{n} \eta_i\big(\langle M,N\rangle_{b_i\wedge t} - \langle M,N\rangle_{a_i\wedge t}\big)$$

Proof By taking conditional expectation and using the martingale property

$$E_P\left(\left\{\int_0^\infty Y_s dM_s\right\}^2\right) =$$
$$\sum_{i=1}^{n} E_P\Big((\eta_i^2(M_{b_i} - M_{a_i})^2\Big) + 2\sum_{i=1}^{n}\sum_{1\le j<n} E_P\Big(\eta_i\eta_j(M_{b_i} - M_{a_i})(M_{b_j} - M_{a_j})\Big) =$$
$$\sum_{i=1}^{n} E_P\Big((\eta_i^2 E_P\big((M_{b_i} - M_{a_i})^2|\mathcal{F}_{a_i}\Big) + 2\sum_{i=1}^{n}\sum_{1\le j<n} E_P\Big(\eta_i\eta_j(M_{b_j} - M_{a_j})E_P(M_{b_i} - M_{a_i}|\mathcal{F}_{a_i}\Big) =$$
$$\sum_{i=1}^{n} E_P\Big((\eta_i^2(\langle M\rangle_{b_i} - \langle M\rangle_{a_i})\Big) = E_P\left(\int_0^\infty Y_s^2 d\langle M\rangle_s\right)$$

where the cross terms have zero expectation. Similarly,

$$E_P\biggl(\Bigl\{\int_0^t Y_s dM_s\Bigr\}(N_t-N_0)\biggr) =$$
$$\sum_{i=1}^n E_P\biggl(\eta_i(M_{b_i\wedge t}-M_{a_i\wedge t})(N_{b_i\wedge t}-N_{a_i\wedge t})\biggr)+2\sum_{i=1}^n\sum_{1\le j<n} E_P\biggl(\eta_i(M_{b_i\wedge t}-M_{a_i\wedge t})(N_{b_j\wedge t}-M_{a_j\wedge t})\biggr)=$$
$$\sum_{i=1}^n E_P\biggl((\eta_i E_P\bigl((M_{b_i\wedge t}-M_{a_i\wedge t})(N_{b_i\wedge t}-N_{a_i\wedge t})|\mathcal{F}_{a_i}\biggr)+$$
$$2\sum_{i=1}^n\sum_{1\le j<n} E_P\biggl(\eta_i(M_{b_j\wedge t}-M_{a_j}\wedge t)E_P(N_{b_i\wedge t}-N_{a_i\wedge t}|\mathcal{F}_{a_i}\biggr)=$$
$$\sum_{i=1}^n E_P\biggl((\eta_i(\langle M,N\rangle_{b_i\wedge t}-\langle M,N\rangle_{a_i\wedge t})\biggr)=E_P\Bigl(\int_0^t Y_s d\langle M,N\rangle_s\Bigr)$$

Theorem 21. *(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 $t\in[0,+\infty]$*

$$\int_0^t |Y_sU_s|\, d|\langle M,N\rangle|_s \le \Bigl(\int_0^t Y_s^2 d\langle M\rangle_s\Bigr)^{1/2}\Bigl(\int_0^t U_s^2 d\langle N\rangle_s\Bigr)^{1/2}$$

By Hölder inequality, we have also for $p,q>1,\ p^{-1}+q^{-1}=1$

$$E_P\Bigl(\int_0^t |Y_sU_s|\, d|\langle M,N\rangle|_s\Bigr)\le E_P\biggl(\Bigl\{\int_0^t Y_s^2 d\langle M\rangle_s\Bigr\}^{p/2}\biggr)^{1/p} E_P\biggl(\Bigl\{\int_0^t U_s^2 d\langle N\rangle_s\Bigr\}^{q/2}\biggr)^{1/q}$$

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

$$0\le\langle M+rN\rangle_t=\langle M\rangle_t+r^2\langle N\rangle_t+2r\langle N,M\rangle_t$$

The corresponding quadratic equation in the unknown r has at most one real solution, and the inequality for the discriminant follows:

$$\langle N,M\rangle_t^2-\langle M\rangle_t\langle N\rangle_t\le 0\iff \bigl|\langle N,M\rangle_t\bigr|\le\sqrt{\langle M\rangle_t}\sqrt{\langle N\rangle_t}$$

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

$$\Bigl|\int_0^t Y_sU_s d\langle M,N\rangle_s\Bigr|\le\Bigl(\int_0^t Y_s^2 d\langle M\rangle_s\Bigr)^{1/2}\Bigl(\int_0^t U_s^2 d\langle N\rangle_s\Bigr)^{1/2}$$

Assume that there is a finite Borel-measurable partition of $[0,t] = \bigcup_{i=1}^{n} B_i$ 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

$$\Delta V_i = \int_{B_i} dV_s$$

where $V_s = \langle M, N\rangle_s, \langle M\rangle_s, \langle N\rangle_s$, has paths of finite total 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 jointly measurable integrands by the monotone convergence theorem for the Lebesgue-Stieltjes integrals splitting first the integrands into positive and negative parts, and approximating from below by simple $\mathcal{F} \otimes \mathcal{B}(\mathbb{R}^+)$-measurable processes $\Box$

Remark 16. *The integrands $Y_s(\omega), U_s(\omega)$ were not assumed to be $\mathbb{F}$-adapted, just jointly measurable.*

Lemma 28. *(martingale characterization) 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.

$$E_P(M_0) = E_P\big(M_\tau\big) = E_P(\mathbf{1}_A M_s + \mathbf{1}_{A^c} M_t) =$$
$$E_P\big(M_t\big) + E_P\big(\mathbf{1}_A(M_s - M_t)\big) = E_P(\big(M_0\big) - E_P\big(\mathbf{1}_A(M_t - M_s)\big)$$
$$\Longrightarrow E_P\big(\mathbf{1}_A(M_t - M_s)\big) = 0$$

which gives the martingale property.

Definition 37. *On a probability space $(\Omega, \mathcal{F})$, a a stochastic process $(Y(s,\omega) : s \in \mathbb{R}^+)$ is jointly measurable when*

- *$\forall s$ the map $\omega \mapsto Y(s,\omega)$ is $\mathcal{F}$-measurable*
- *$\forall \omega$ the map $s \mapsto Y(s,\omega)$ is Borel measurable*

We say that $Y(s,\omega)$ is progressively measurable w.r.t. the filtration $\mathbb{F} = (\mathcal{F}_s)$, when $\forall t \geq 0$ the restricion

$$Y : [0,t] \times \Omega \mapsto \mathbb{R}^d$$

is $\mathcal{B}([0,t]) \otimes \mathcal{F}_t$-jointly measurable.

Theorem 22. *(Ito integral, from the Revuz and Yor's book) Let $(M_t) \in \mathcal{M}^2$ and $Y(s,\omega)$ a progressively measurable process with*

$$E_P\bigg(\int_0^\infty Y_s^2 d\langle M\rangle_s\bigg) < \infty$$

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 $\forall\ (N_t) \in \mathcal{M}^2$,

$$E_P\bigg((Y \cdot M)_\infty N_\infty\bigg) = E_P\bigg(\int_0^\infty Y_s d\langle M, N\rangle_s\bigg) = E_P\bigg(\langle Y \cdot M, N\rangle_\infty\bigg) \quad (6.1)$$

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, in particular

$$\bigg\langle (Y \cdot M), N \bigg\rangle_t = \int_0^t Y_s d\langle M, N\rangle_s$$

and for $N = (Y \cdot M)$

$$\langle Y \cdot M\rangle_t = \int_0^t Y_s^2 d\langle M, M\rangle_s \quad \forall t \in [0, +\infty]. \quad (6.2)$$

3. *By uniqueness it follows that for simple predictable integrands this definition of Ito integral coincides with the Riemann sums definition given in (27).*

Proof: The map

$$N_\infty \mapsto \varphi(N) := E_P\biggl(\int_0^\infty Y_s d\langle M, N\rangle_s\biggr)$$

is linear since the predictable covaration is bilinear. It is also continuous in $\mathcal{M}^2$ norm: by Kunita-Watanabe and Cauchy-Schwartz inequalities

$$|\varphi(N)| = \biggl|E_P\biggl(\int_0^\infty Y_s d\langle M, N\rangle_s\biggr)\biggr| \le E_P\biggl(\int_0^\infty Y_s^2 d\langle M\rangle_s\biggr)^{1/2} E_P\Bigl(\langle N\rangle_\infty\Bigr)^{1/2} =$$
$$E_P\biggl(\int_0^\infty Y_s^2 d\langle M\rangle_s\biggr)^{1/2} \parallel N \parallel_{\mathcal{M}^2}$$

When

$$E_P\biggl(\int_0^\infty Y_s^2 d\langle M\rangle_s\biggr) < \infty$$

by the Riesz representation theorem in the Hilbert space $\mathcal{M}^2$ there exists an unique continuous martingale $\{(Y\cdot M)_t\} \in \mathcal{M}^2$ such that

$$E_P\biggl(\int_0^\infty Y_s d\langle M, N\rangle_s\biggr) = \varphi(N) = \bigl((Y\cdot M), N\bigr)_{\mathcal{M}^2} =$$
$$E_P\Bigl((Y\cdot M)_\infty N_\infty\Bigr) = E_P\Bigl(\langle Y\cdot M, N\rangle_\infty\Bigr)$$

Note: up to now we did not need predictability or progressive measurability of (Y_s), in Kunita Watanabe inequality just joint measurability was required.

The progressive measurability of Y_s will be needed in to 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 $N \in \mathcal{M}^2$ which means, by definition of predictable covariation,

$$\bigl\langle (Y\cdot M), N\bigr\rangle_t = \int_0^t Y_s d\langle M, N\rangle_s.$$

By taking $N_t = (Y\cdot M)_t$ we obtain also (6.2)

$$\langle Y\cdot M\rangle_t = \int_0^t Y_s d\langle M, (Y\cdot M)\rangle_s = \int_0^t Y_s d(Y\cdot\langle M\rangle)_s = \int_0^t Y_s^2 d\langle M, M\rangle_s \; .$$

Let τ be a $(\mathcal{F}_t)$-stopping time. By Cauchy Schwartz and Kunita Watanabe inequalities $X_\tau \in L^1(P)$.

The martingales $(Y \cdot M)_t$ and (N_t) are uniformly integrable martingales (since they are bounded in $L^2(\Omega, \mathcal{F}, P)$), we write

$$E_P\big((Y \cdot M)_\tau N_\tau\big) = E_P\bigg(E_P\big((Y \cdot M)_\infty | \mathcal{F}_\tau\big) N_\tau\bigg) = E_P\bigg((Y \cdot M)_\infty N_\tau\bigg) =$$

$$E_P\bigg((Y \cdot M)_\infty N^\tau_\infty\bigg) = E_P\bigg(\langle (Y \cdot M), N^\tau \rangle_\infty\bigg) = \text{ by the defining property (6.3)}$$

$$= E_P\bigg(\int_0^\infty Y_s d\langle M, N^\tau \rangle_s\bigg) = E_P\bigg(\int_0^\tau Y_s d\langle M, N^\tau \rangle_s\bigg)$$

and by the martingale characterization lemma 28

$$X_t = (Y \cdot M)_t N_t - \int_0^t Y_s d\langle M, N^\tau \rangle_s$$

is a true martingale if it is $\mathbb{F}$-adapted.

Note that when $Y_s(\omega)$ is progressively measurable, $\forall t$

$$\int_0^t Y_s d\langle M, N \rangle_s$$

and X_t are $\mathcal{F}_t$-measurable.

To show that $(Y \cdot M)_0 = 0$, take a constant martingale $N_t \equiv N_0 \in L^2(\Omega, \mathcal{F}_0, P)$.

By Kunita-Watanabe inequality $|\langle M, N \rangle_t| \leq \sqrt{\langle M \rangle_t}\sqrt{\langle N \rangle_t} = 0$ since $[N, N]_t = \langle N, N \rangle_t = 0$.

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 $N_t = (Y \cdot M)_t$,

$$\langle (Y \cdot M), (Y \cdot M) \rangle_t = \int_0^t Y_s d\langle M, (Y \cdot M) \rangle_s = \int_0^t Y_s^2 d\langle M, M \rangle_s$$

Remark 17. *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 $L^2\big(\Omega \times \mathbb{R}^+, \mathcal{P}, P(d\omega)\langle M \rangle(dt, \omega)\big)$ 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).

Remark 18. *The Ito map $(Y, M) \mapsto (Y \cdot M) \in \mathcal{M}_2$ is bilinear.*

Lemma 29. *Under the assumption of Theorem (22), If τ is a stopping time, the stochastic integral with respect to the stopped martingale $M_t^\tau = M_{t\wedge\tau}$ satisfies*

$$(Y \cdot M^\tau)_t = \int_0^t Y_s dM_s^\tau = \int_0^t Y_s \mathbf{1}(\tau > s) dM_s =$$
$$(Y \cdot M)_t^\tau = (Y \cdot M)_{t\wedge\tau} = \int_0^{t\wedge\tau} Y_s dM_s$$

Proof. For $N \in \mathcal{M}_2$, since $\langle M, N^\tau \rangle_t = \langle M, N \rangle_{t\wedge\tau}$

$$E\biggl(\int_0^\infty Y_s d\langle M, N^\tau \rangle_s \biggr) = E(\biggl(\int_0^\infty Y_s \mathbf{1}(\tau > s) d\langle M, N \rangle_s \biggr)$$

implies by the uniqueness of the Riesz representation that

$$\int_0^\infty Y_s dM_s^\tau = \int_0^\infty Y_s \mathbf{1}(\tau > s) dM_s = \int_0^\tau Y_s dM_s$$

Proposition 21. *(Extension by localization)*

Let (M_t) a continuous local martingale and $(Y_t(\omega))$ a progressively measurable process such that $\forall t \geq 0$

$$P\biggl(\int_0^t Y_s^2 d\langle M \rangle_s < \infty \biggr) = 1$$

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

$$\langle (Y \cdot M), N \rangle_t = \int_0^\infty Y_s d\langle M, N \rangle_s \tag{6.3}$$

for every continuous local martingale N.

Proof Let (τ_n') a localizing sequence for M_t. Define the sequence of stopping times

$$\tau_n'' := \inf\biggl\{ t \geq 0 : \int_0^t Y_s^2 \langle M \rangle_s \geq n \biggr\}, \quad n \in \mathbb{N}$$

and $\tau_n = (\tau_n' \wedge \tau_n'')$. We see that $\tau_n(\omega) \uparrow \infty$ P a.s.

With this localization, for each n Y_t and the stopped process $M_t^{\tau_N}$ satisfy the assumptions of Theorem (22) and the Ito integral $(Y \cdot M^{\tau_n}) \in \mathcal{M}_2$ exists.

Note that $\forall 0 \leq k \leq n$ by lemma (29)

$$\int_0^t Y_s \mathbf{1}(\tau_k > s) dM_s^{\tau_k} = \int_0^t Y_s \mathbf{1}(\tau_k > s) dM_s^{\tau_n}$$

as elements of $\mathcal{M}_2$.

The sets $\Omega_k = \{\omega : \tau_{k-1}(\omega) \leq t < \tau_k(\omega)\}$ form a measurable partition of Ω.

Define

$$\int_0^t Y_s dM_s = \sum_{n=0}^{\infty}\Big(\int_0^t Y_s dM_t^{\tau_n} - \int_0^t Y_s dM_t^{\tau(n-1)}\Big) = \lim_{n\to\infty}\int_0^t Y_s dM_s^{\tau_n}$$

where for fixed t, P almost surely $\tau_n(\omega) \uparrow \infty$, and the telescopic sum contains only finitely many non-zero terms,

We see that P a.s. the trajectory $t \mapsto \int_0^t Y_s dM_s$ is continuous, and $\int_0^t Y_s dM_s$ is a local martingale with localizing sequence (τ_n) $\square$

Lemma 30. *(Dominated stochastic convergence) 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,*

$$|Y_s^{(n)}(\omega)| \to 0 \ P\text{-almost surely}$$

and there is a locally bounded process $X_s(\omega)$ such that P-almost surely, $|Y_s^{(n)}(\omega)| \leq X_s(\omega), \quad \forall s, n$. Then for all $t \geq 0$

$$\sup_{s\leq t}\Big|\int_0^t Y_s^{(n)} dM_s\Big| \to 0 \quad \text{in probability as } n \to \infty$$

Let $\tau(\omega)$ be a stopping time such that both stopped processes M_s^{τ} and X_s^{τ} are bounded. Then by the bounded convergence theorem

$$E_P\Big(\int_0^{\tau} (Y_s^{(n)})^2 d\langle M_s\rangle\Big) \to 0 \text{ as } n \to \infty$$

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 complete the argument we for any fixed t choose the localizing stopping time such that $P(\tau \leq t) < \varepsilon$ and conclude as in corollary (15).

Definition 38. *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 $\forall 0 \leq t < \infty$

$$\int_0^t Y_s^2 d\langle M\rangle_s < \infty \quad \text{and} \quad \int_0^t |Y_s||dA|_s < \infty \quad P\text{-almost surely}$$

where the integral on the right side is with respect to the total variation of A, we define

$$\int_0^t Y_s dX_s = \int_0^t Y_s dM_s + \int_0^t Y_s dA_s$$

We also have $[X, X] = [M, M] = \langle M\rangle = \langle X\rangle$

6.2 Ito formula for semimartingales

Proposition 22. *Let X_t, Y_t continuous semimartingales. Then we have the integration by parts formula*

$$X_tY_t = X_0Y_0 + \int_0^t X_s dY_s + \int_0^t Y_s dX_s + [X,Y]_t$$

Proof: By polarization it is enough to show

$$X_t^2 - X_0^2 - [X,X]_t = 2\int_0^t X_s dX_s$$

Since the formula is true when X has finite variation, it is enough to show

$$M_t^2 - M_0^2 - [M,M]_t = 2\int_0^t M_s dM_s$$

when M is a local martingale.

By taking telescopic sum for a grid $0 = t_0 < t_1 < \dots <$, by the discrete integration by parts formula

$$\sum_i \big(M_{t_i\wedge t} - M_{t_{i-1}\wedge t}\big)^2 = M_t^2 - M_0^2 - 2\sum_i M_{t_i}\big(M_{t_i\wedge t} - M_{t_{i-1}\wedge t}\big)$$

As $\Delta = \sup(t_i - t_{i-1}) \to 0$ the left side and right hand sides converges in probability uniformly on finite intervals respectively to $[M,M]_t$ and

$$M_t^2 - M_0^2 - 2\int_0^t M_s dM_s \quad \square$$

Theorem 23. *(Ito formula) When $X_t(\omega) \in \mathbb{R}^d$ is a continuous semimartingale and $f \in C^2(\mathbb{R}^d, \mathbb{R})$*

$$f(X_t) = f(X_0) + \sum_{i=1}^d \int_0^t \frac{\partial f}{\partial x_i}(X_s) dX_s^{(i)} + \frac{1}{2}\sum_{i,j}\int_0^t \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s) d\langle X^{(i)}, X^{(j)}\rangle_s$$

Proof When the result holds for the function $f(x_1,\dots,x_d)$, by the integration by parts formula is holds also for the function $g(x_1,\dots,x_d) = x_i f(x_1,\dots,x_d)$. It follows that Ito formula holds when $f(x)$ is a polynomial. By stopping it is enough to consider the case when $|X_t(\omega)| \le C < \infty$ P a.s. Since continuous functions are approximated uniformly on compacts by polynomials, we find a polynomial $f_n(x)$ such that

$$\sup_{|x|\le C} \big|(f_n - f)(x)\big| \le \frac{1}{n}, \ \sup_{|x|\le C}\left|\frac{\partial(f_n - f)}{\partial x_i}(x)\right| \le \frac{1}{n}, \ \sup_{|x|\le C}\left|\frac{\partial^2(f_n - f)}{\partial x_i \partial x_j}(x)\right| \le \frac{1}{n}$$

This implies P-almost sure convergence

$$f_n(X_t) \longrightarrow f(X_t), \quad \int_0^t \frac{\partial^2 f_n}{\partial x_i \partial x_j}(X_s) d\langle X^{(i)}, X^{(j)}\rangle_s \longrightarrow \int_0^t \frac{\partial^2 f}{\partial x_i \partial x_j}(X_s) d\langle X^{(i)}, X^{(j)}\rangle_s$$

uniformly on finite intervals, and by the dominated stochastic convergence lemma 30

$$\int_0^t \frac{\partial f}{\partial x_i}(X_s)dX_s^{(i)} \xrightarrow{P} \int_0^t \frac{\partial f}{\partial x_i}(X_s)dX_s^{(i)}$$

in probability, uniformly on finite intervals.

Theorem 24. *(Lévy characterization of Brownian motion) Let $M_t(\omega) \in \mathbb{R}^d$ a continuous $\mathbb{F}$-adapted process, with $M_0 = 0$. The following conditions are equivalent*

1. *M_t is a d-dimensional $\mathbb{F}$-Brownian motion: it has P a.s. continuous paths, $\forall s \le t$ the increment $(M_t - M_s)$ is P-independent from $\mathcal{F}_s$, and Gaussian with $E(M_t^{(k)} - M_s^{(k)}) = 0, E((M_t^{(k)} - M_s^{(k)})(M_t^{(h)} - M_s^{(h)})) = (t-s)\delta_{kh}$.*
2. *$M_t^{(k)}$ and $(M_t^{(k)} M_t^{(h)} - t\delta_{hk})$ are continuous $\mathbb{F}$-local martingales, $h, k = 1, \dots, d$.*

Proof we know already that 1) $\implies$ 2), and these local martingales are square integrable martingales (all moments of the Gaussian distribution are finite).

Assuming (2), we show that the increments are Gaussian independent from the past. The idea is to study the conditional distribution by usign the characteristic function.

Apply Ito formula to

$$f(M_t(\omega), t) = \exp\biggl(i\theta \cdot M_t(\omega) + \frac{1}{2}|\theta|^2 t\biggr) \in \mathbb{C}$$

(which means to apply separately Ito formula to real and imaginary parts), obtaining

$$f(M_t, t) - f(M_s, s) =$$
$$i\sum_{k=1}^d \theta_k \int_s^t f(M_r, r)dM_r^{(k)} + \frac{i^2}{2}\sum_{k,h} \theta_k\theta_h f(M_r, r)d\langle M^{(k)}, M^{(h)}\rangle_r + \frac{|\theta|^2}{2}\int_s^t f(M_r, r)dr =$$
$$= i\sum_{k=1}^d \theta_i \int_s^t f(M_r, r)dM_r^{(k)}$$

where the finite variation parts cancels since $\langle M^{(k)}, M^{(h)}\rangle_r = r\delta_{kh}$.

Therefore $f(M_t, t)$ is a local martingale. It is a true square integrable martingale since for all t

$$|f(M_t, t)| \le \exp\bigl(\frac{1}{2}|\theta|^2 t\bigr)$$

Let $s \le t$ and $A \in \mathcal{F}_s$. By the martingale property $\forall \theta \in \mathbb{R}^d$,

$$E\biggl(\bigl(f(M_t, t) - f(M_s, s)\bigr)\mathbf{1}_A\biggr) = 0$$
$$\implies E\biggl(\exp\bigl(i\theta \cdot (M_t - M_s)\bigr)\mathbf{1}_A\biggr) = E\biggl(E\biggl(\exp\bigl(i\theta \cdot (M_t - M_s)\bigr)\bigg|\mathcal{F}_s\biggr)\mathbf{1}_A\biggr) = \exp\biggl(-\frac{1}{2}|\theta|^2(t-s)\biggr)P(A)$$

which implies

$$E\Big(\exp\big(i\theta\cdot(M_t-M_s)\big)\Big|\mathcal{F}_s\Big)=\exp\Big(-\frac{1}{2}|\theta|^2 s\Big)\quad(\text{ deterministic })$$

Since the characteristic function characterizes the distribution, (M_t-M_s) is independent from $\mathcal{F}_s$ and Gaussian, with zero mean and covariance $(t-s)\mathrm{Id}$ □

6.3 Ito representation theorem

Let $B_t=(B_t^{(1)},\dots,B_t^{(d)})$ a d-dimensional Brownian motion.

Theorem 25. *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*

$$E_P\Big(\int_0^T H_s^2 ds\Big)<\infty$$

$$Y(\omega)=E_P(Y)+\int_0^T H_s dB_s=E_P(Y)+\sum_{i=1}^d\int_0^T H_s^{(i)}dB_s^{(i)}$$

$H_s(\omega)$ is unique $P(d\omega)\times ds$ almost surely.

Proof Uniqueness: if $\widetilde{H}_s$ has the same property, then by Ito isometry

$$\int_\Omega\Big(\int_0^T\big(H_s(\omega)-\widetilde{H}_s(\omega)\big)^2 ds\Big)P(d\omega)=0$$

Existence:

$$\mathcal{H}=\Big\{\int_0^T H_s dB_s:\ H\text{ is progressive and in }L^2(\Omega\times[0,T],dP\times dt)\Big\}$$

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 $E_P\Big(Y\int_0^T H_s dB_s\Big)=0$ for all progressive $H\in L^2(\Omega\times[0,T],dP\times dt)$, then $Y(\omega)=E_P(Y)$.

The random variable $\big(Y(\omega)-E_P(Y)\big)$ coincides with its orthogonal projection on the closed subspace $\mathcal{H}$, and the results follows.

Without loss of generality assume that $E_P(Y)=0$, otherwise take $\widetilde{Y}(\omega)=(Y(\omega)-E_P(Y))$. For $f(x)\in L^2([0,T],dt)$ with values in $\mathbb{R}^d$, consider the complex valued square integrable martingale

$$M_t^{(f)}=\exp\Big(i\int_0^t f(s)dB_s+\frac{1}{2}\int_0^t|f(s)|^2 ds\Big),\quad i=\sqrt{-1}$$

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 integrals in $\mathcal{H}$,

$$0 = E_P\Big(Y(M_T^{(f)} - 1)\Big) = E_P\Big(YM_T^{(f)}\Big) - E_P(Y) = E_P\Big(YM_T^{(f)}\Big)$$

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

$$\begin{aligned}
0 &= E_P\bigg(Y \exp\Big(i \sum_{k=1}^n \theta_k \cdot B_{t_k} + \frac{1}{2} \sum_{h,k=1}^n \theta_h \theta_k\ (t_h \wedge t_k)\Big)\bigg) \\
&= E_P\bigg(Y \exp\Big(i \sum_{k=1}^n \theta_k \cdot B_{t_k}\Big)\bigg) \exp\Big(\frac{1}{2} \sum_{h,k=1}^n \theta_h \theta_k\ (t_h \wedge t_k)\Big) \\
&\Longrightarrow E_P\bigg(Y \exp\Big(i \sum_{k=1}^n \theta_k \cdot B_{t_k}\Big)\bigg) = 0
\end{aligned}$$

By the Lévy inversion theorem, which holds not on only for probability measures but also for finite signed measures, the characteristic function characterizes the measure.

Since the characteristic function is identically zero, $\forall A_k \in \mathcal{B}(\mathbb{R}^d), k = 1, \dots, n$,

$$\mu_{t_1,\dots t_n}(A_1 \times \dots \times A_n) := E_P\bigg(Y\mathbf{1}\big(B_{t_1} \in A_1, \dots, B_{t_n} \in A_n\big)\bigg) = 0\ .$$

Since the cylinders generate the σ-algebra $\mathcal{F}_T^B$, by Dynkin extension theorem

$$E_P(Y\mathbf{1}_F) = 0 \qquad \forall F \in \mathcal{F}_T^B$$

By assumption $Y \in \mathcal{F}_T^B$ measurable, by taking $F^\pm = \{\omega : \pm Y(\omega) > 0\}$, we see that $Y(\omega) = 0$ P-a.s. □

Corollary 16. *Let (M_t) a martingale in the Brownian filtration bounded in L^2, i.e. $E_P(M_\infty^2) < \infty$. 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, $M_0 = E_P(M_0) = E_P(M_t) = E_P(M_\infty)$.

6.3.1 Computation of martingale representation

Let $F(\omega) = f(B_T(\omega))$ for some $f(x) \in L^2(\mathbb{R}, \gamma(x)dx)$.

$$E\big(f(B_T)\big|\mathcal{F}_t\big) = E\big(f(B_t+(B_T-B_t))\big|\mathcal{F}_t\big)$$
$$= E\bigg(f(x+G\sqrt{T-t})\bigg)\bigg|_{x=B_t(\omega)}$$
$$= \int_{\mathbb{R}} f(B_t(\omega)+y\sqrt{T-t})\gamma(y)dy =$$
$$\int_{\mathbb{R}} f(u)\frac{1}{\sqrt{T-t}}\gamma\bigg(\frac{B_t-u}{\sqrt{T-t}}\bigg)dy =$$

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

$$d\gamma(Y_t) = \gamma(Y_t)\bigg(-Y_t dY_t + \frac{1}{2}(Y_t^2-1)d\langle Y\rangle_t\bigg)$$

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

$$d\gamma(Y_t) = \gamma(Y_t)\bigg(-\frac{(B_t-u)}{T-t}dB_t - \frac{1}{2}\frac{(B_t-u)^2}{(T-t)^2}dt + \frac{1}{2}\bigg(\frac{(B_t-u)^2}{T-t}-1\bigg)\frac{1}{T-t}dt\bigg) =$$
$$-\gamma(Y_t)\bigg(\frac{B_t-u}{T-t}dB_t + \frac{1}{2(T-t)}dt\bigg)$$

Integrating by parts:

$$d\bigg(\frac{1}{\sqrt{T-t}}\gamma(Y_t)\bigg) = \frac{1}{\sqrt{T-t}}\gamma(Y_t)\bigg(-\frac{B_t-u}{T-t}dB_t - \frac{1}{2(T-t)}dt + \frac{1}{2(T-t)}dt\bigg)$$
$$= \frac{1}{\sqrt{T-t}}\gamma\bigg(\frac{B_t-u}{\sqrt{T-t}}\bigg)\bigg(\frac{u-B_t}{T-t}\bigg)dB_t$$

Therefore we have simply

$$g(B_t,u,t,T) = g(0,u,0,T) + \int_0^t g(B_s,u,s,T)\bigg(\frac{u-B_s}{T-s}\bigg)dB_s$$

and the stochastic exponential representation

$$g(B_t,u,t,T) = g(0,u,0,T)\mathcal{E}\left(\int_0^{\cdot}\left(\frac{u-B_s}{T-s}\right)dB_s\right)_t$$
$$= g(0,u,0,T)\exp\left(\int_0^t\left(\frac{u-B_s}{T-s}\right)dB_s - \frac{1}{2}\int_0^t\left(\frac{u-B_s}{T-s}\right)^2 ds\right)$$

Integrating with respect to du we get

$$E_P(f(B_T)|\mathcal{F}_t) =$$
$$\int_{\mathbf{R}} f(u)g(B_t,u,t,T)du = \int_{\mathbf{R}} f(u)g(0,u,0,T)du + \int_{\mathbf{R}}\left(\int_0^t f(u)\left(\frac{u-B_s}{T-s}\right)g(B_s,u,s,T)dB_s\right)du$$
$$= E_P(f(B_T)) + \int_0^t\left(\int_{\mathbf{R}} f(u)\left(\frac{u-B_s}{T-s}\right)g(B_s,u,s,T)du\right)dB_s$$
$$= E_P(f(B_T)) + \int_0^t \frac{E_P\big(f(B_T)(B_T-B_s)\big|\mathcal{F}_s\big)}{(T-s)}dB_s$$

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

$$F(\omega) = E_P\big(f(\| h \|_2 G)\big) + \int_0^T \frac{E_P\Big(f\Big(\int_0^T h(s)dB_s\Big)\int_t^T h(s)dB_s\Big|\mathcal{F}_s\Big)}{\int_t^T h(s)^2 ds} h(t)dB_t$$

Hint: define the deterministic time change

$$\tau(u) = \inf\Big\{t : \int_0^t h(s)^2 ds \ge u\Big\}$$

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 26. *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

$$\int_0^T \biggl(\int_\Theta \bigl(H^n(s,\omega,\theta) - H(s,\omega,\theta)\bigr)\alpha(d\theta)\biggr)^2 d\langle M\rangle_s$$
$$\le \int_{[0^T]\times\Theta} \bigl(H^n(s,\omega,\theta) - H(s,\omega,\theta)\bigr)^2 \alpha(d\theta)\otimes d\langle M\rangle_s \stackrel{P}{\to} 0$$

This implies

$$\int_\Theta \biggl(\int_0^T H^{(n)}(s,\theta)dB_s\biggr)\alpha(d\theta) =$$
$$\int_0^T \biggl(\int_\Theta H^{(n)}(s,\theta)\alpha(d\theta)\biggr)dB_s \stackrel{P}{\to} \int_0^T \biggl(\int_\Theta H^{(n)}(s,\theta)\alpha(d\theta)\biggr)dB_s$$

and since

$$\int_\Theta \biggl(\int_0^T \bigl(H^{(n)}(\omega,s,\theta) - H(\omega,s,\theta)\bigr)^2 d\langle M\rangle_s\biggr)\alpha(d\theta) \stackrel{P}{\to} 0$$

It follows that

$$\int_\Theta \biggl(\int_0^T (H^{(n)}(s,\theta)dB_s\biggr)\alpha(d\theta) \stackrel{P}{\to} \int_\Theta \biggl(\int_0^T (H(s,\theta)dB_s\biggr)\alpha(d\theta)$$

Proposition 23. *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*

$$E_P\bigl(f(G)G\bigr) = E_P\bigl(f'(G)\bigr)$$

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 $E_P\bigl(f'(G) - f(G)G\bigr)$.

More in general we approximate $f(x)$ with a sequence of compactly supported functions. Let $k_n(x) = (1-|x|/n)^+$. We have $0\le k_n(x)\le 1$, $\frac{d}{dx}k_n(x) = -n^{-1}\mathrm{sign}(x)bf1(|x|\le n)$, 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 24. *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

$$\frac{P(H_a \in d\ell)}{d\ell} = p_{H_a}(\ell) =$$
$$(2\pi)^{-1/2} \exp\left(-\frac{a^2}{2\ell}\right) a\, \ell^{-3/2}\, \mathbf{1}(\ell > 0), \qquad a > 0$$

Moreover

$$P(B_\ell^{\ge} a, B_\ell \in dx) = \frac{1}{\sqrt{\ell}} \gamma\left(\frac{a + |x - a|}{\sqrt{\ell}}\right) dx \tag{6.4}$$

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

$$\{B_\ell^* \ge a\} = \{B_\ell \ge a\} \cup \{\widetilde{B}_\ell \ge a\}$$

where $\{B_\ell \ge a\} \cap \{\widetilde{B}_\ell \ge a\} = \emptyset$

$$P\big(B_\ell^* \ge a\big) = P\big(\{B_\ell \ge a\} \cup \{\widetilde{B}_\ell \ge a\}\big)$$
$$= P\big(B_\ell \ge a\big) + P\big(\widetilde{B}_\ell \ge a\big) =$$
$$2P(B_\ell \ge a) = 2\big(1 - \Phi(a/\sqrt{\ell})\big) = 2\Phi(-a/\sqrt{\ell})$$

where $\Phi(x)$ is the cumulative distribution function of a standard Gaussian r.v.

By the same argument

$$P(B_\ell^* \geq a, B_\ell \in dx) = P(B_\ell^* \geq a, \widetilde{B}_\ell \in dx) = P(B_\ell^* \geq a, 2a - B_\ell \in dx)$$

now there are two case either $x \geq a$ or $x < a$. When $x \geq a$

$$\frac{P(B_\ell^* \geq a, B_\ell \in dx)}{dx}(x) = \frac{P(B_\ell \in dx)}{dx}(x)$$

otherwise $2a - x > a$. and

$$\frac{P(B_\ell^* \geq a, B_\ell \in dx)}{dx}(x) = \frac{P(B_\ell \in dx)}{dx}(2a - x)$$

In both cases this gives formula (6.4).

6.4 Barrier option in Black and Scholes model

Consider the Black and Scholes model for a risky asset and a riskless bond.

$$\begin{aligned}
S_t &= S_0 \exp\Big(\int_0^t \sigma_s dB_s + \int_0^t (\mu_t - \frac{\sigma_t^2}{2}) dt\Big), \\
U_t &= U_0 \exp\Big(\int_0^t \rho_s ds\Big) \\
& S_0 > 0,\ U_0 > 0 \\
& dS_t = S_t(\mu_t dt + \sigma_t dB_t), \quad dU_t = U_t \rho_t dt
\end{aligned}$$

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.

$$E_P(f(B_T)\mathbf{1}(B_T^* > a)) = \int_{\mathbb{R}} f(x)\frac{1}{\sqrt{T}}\gamma\biggl(\frac{a+|x-a|}{\sqrt{T}}\biggr)dx$$
$$E_P(f(B_T)\mathbf{1}(B_T^* > a)|\mathcal{F}_t) = E_P(f(B_T)\mathbf{1}(B_T^* > a)|B_t, B_t^*)$$
$$= \mathbf{1}(B_t^* > a)E_P(f(x+\sqrt{T-t}G))\Big|_{x=B_t} + \mathbf{1}(B_t^* \le a)E_P\bigl(f(x+W_{T-t})\mathbf{1}(W_{T-t}^* > (a-x))\bigr)\Big|_{x=B_t}$$
$$\mathbf{1}(B_t^* > a)\int_{\mathbb{R}} f(x)\frac{1}{\sqrt{T-t}}\gamma\biggl(\frac{x-B_t}{\sqrt{T-t}}\biggr)dx + \mathbf{1}(B_t^* \le a)\int_{\mathbb{R}} f(x)\frac{1}{\sqrt{T-t}}\gamma\biggl(\frac{a-B_t+|x-a|}{\sqrt{T-t}}\biggr)dx$$

By using Ito formula and stochastic Fubini theorem

$$E_P(f(B_T)\mathbf{1}(B_T^* > a)|\mathcal{F}_t) =$$
$$E_P(f(B_T)\mathbf{1}(B_T^* > a))$$
$$+\int_0^t \mathbf{1}(B_s^* > a)\biggl(\int_{\mathbb{R}} f(x)\frac{1}{\sqrt{T-s}}\gamma\biggl(\frac{x-B_s}{\sqrt{T-s}})\frac{x-B_s}{T-s}dx\biggr)dB_s$$
$$+\int_0^t \mathbf{1}(B_s^* \le a)\biggl(\int_{\mathbb{R}} f(x)\frac{1}{\sqrt{T-s}}\gamma\biggl(\frac{a-B_s+|x-a|}{\sqrt{T-s}})\frac{a-B_s+|x-a|}{T-s}dx\biggr)dB_s$$
$$= E_P(f(B_T)\mathbf{1}(B_T^* > a)) + \int_0^t \mathbf{1}(B_s^* > a)\frac{E_P\bigl(f(B_T)(B_T-B_s)\big|\mathcal{F}_s\bigr)}{(T-s)}dB_s$$
$$+\int_0^t \mathbf{1}(B_s^* \le a)\frac{E_P\bigl(f(B_T)(a-B_s+|B_T-a|)|\mathcal{F}_s\bigr)}{T-s}dB_s$$

We also write the joint law of B_t^*, B_t:

$$P\biggl(B_t^* > y, B_t \le x\biggr) = P\biggl(H_y \le t, (B_t - B_{H_y}) \le (x-y)\biggr)$$
$$= \int_0^t \Phi\biggl(\frac{x-y}{\sqrt{t-\ell}}\biggr)P(H_y \in d\ell)$$
$$= (2\pi)^{-1/2}\int_0^t \Phi\biggl(\frac{x-y}{\sqrt{t-\ell}}\biggr)\exp\biggl(-\frac{y^2}{2\ell}\biggr)y\,\ell^{-3/2}\,d\ell =$$
$$\int_0^t \Phi\biggl(\frac{x-y}{\sqrt{t-\ell}}\biggr)\frac{1}{\sqrt{\ell}}\gamma\biggl(\frac{y}{\sqrt{\ell}}\biggr)\frac{y}{\ell}d\ell$$

and the joint density is given by

$$\frac{P(B_t^* \in dy, B_t \in dx)}{dxdy} = -\frac{\partial^2}{\partial x\partial y}P\biggl(B_t^* > y, B_t \le x\biggr)$$
$$= \int_0^t \frac{1}{\sqrt{t-\ell}}\gamma\biggl(\frac{x-y}{\sqrt{t-\ell}}\biggr)\frac{1}{\sqrt{\ell}}\gamma\biggl(\frac{y}{\sqrt{\ell}}\biggr)\frac{1}{\ell}\biggl(\frac{y^2}{\ell}-1-\frac{y(x-y)}{(t-\ell)}\biggr)d\ell$$

By differentiating w.r.t. a we obtain the density of B_ℓ^*:

$$\frac{P(B_\ell^* \in da)}{da} = p_{B_\ell^*}(a) =$$
$$\frac{2}{\sqrt{2\pi\ell}}\exp\biggl(-\frac{a^2}{2\ell}\biggr)\mathbf{1}(a \ge 0) = \frac{2}{\sqrt{\ell}}\gamma\biggl(\frac{a}{\sqrt{\ell}}\biggr)\mathbf{1}(a \ge 0)$$

We now compute the regular conditional density given the σ-algebra $\mathcal{F}_t^B$, $t \geq 0$.

For any bounded measurable function g

$$E_P(g(H_a)|\mathcal{F}_t^B) = g(H_a)\mathbf{1}(H_a \leq t) + E_P(g(H_a)|B_t, H_a > t)\mathbf{1}(H_a > t) =$$
$$g(H_a)\mathbf{1}(H_a \leq t) + E_P(g(t + H_{a-x}))\Big|_{x=B_t} \mathbf{1}(H_a > t)$$

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

$$M(\ell, t) := \frac{P(H_a \in d\ell|B_t, H_a > t)}{d\ell} = (2\pi)^{-1/2} \exp\Big(-\frac{(B_t - a)^2}{2(\ell - t)}\Big)\frac{(a - B_t)}{(\ell - t)^{3/2}}\,\mathbf{1}(\ell > t)$$

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

$$M(\ell, t \wedge H_a) = M(\ell, 0)\mathcal{E}\Big(\int_0^{\cdot}\Big\{\frac{1}{(B_s - a)} + \frac{(a - B_s)}{\ell - s}\Big\}dB_s\Big)_{t\wedge H_a} =$$
$$M(\ell, 0)\exp\Big(\int_0^{t\wedge H_a}\Big\{\frac{1}{(B_s - a)} + \frac{(a - B_s)}{\ell - s}dB_s - \frac{1}{2}\int_0^{t\wedge H_a}\Big\{\frac{1}{(B_s - a)} + \frac{(a - B_s)}{\ell - s}\Big\}^2\Big)$$

Note that the process (B_t^*, B_t) is Markovian:

$$E_P(f(B_\ell^*)|\mathcal{F}_s) = \mathbf{1}(\ell \leq s)f(B_\ell^*) + \mathbf{1}(\ell > s)E_P\big(f(\max\{x, y + W_{\ell-s}^*\sqrt{\ell - s}\})\big)\Big|_{x=B_s^*, y=B_s}$$
$$= \mathbf{1}(\ell \leq s)f(B_\ell^*) + \mathbf{1}(\ell > s)\int_0^\infty f(\max\{B_s^*(\omega), B_s(\omega) + v)\frac{2}{\sqrt{\ell - s}}\gamma\Big(\frac{v}{\sqrt{\ell - s}}\Big)dv$$
$$= \mathbf{1}(\ell \leq s)f(B_\ell^*) + \mathbf{1}(\ell > s)\Big\{f(B_s^*)(2\Phi\Big(\frac{B_s^* - B_s}{\sqrt{\ell - s}}\Big) - 1) + \int_{B_s^*}^\infty f(v)\frac{2}{\sqrt{\ell - s}}\gamma\Big(\frac{v - B_s}{\sqrt{\ell - s}}\Big)dv\Big\}$$

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 $P(B_T \in dx, B_T^* \in dy|\mathcal{F}_t) = P(B_T \in dx, B_T^* \in dy|B_t, B_t^*)$.

6.5 Stochastic differential equation

Given a Brownian motion (B_t) we look for a stochatic process $(X_t : t \in [s, T])$ such that

$$X_t = \eta + \int_s^t b(u, X_u)du + \int_s^t \sigma(u, X_u)dB_u \quad 0 \le s \le t \tag{6.5}$$

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 (6.6) In differential notation we write

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t, \quad t \ge s \tag{6.6}$$

with initial condition $X_s(\omega) = \eta(\omega)$.

6.5.1 Generator of a diffusion

Lemma 31. *Assume that the SDE 6.6 has a strong solution and that $\varphi(t, x) \in C^{1,2}(\mathbb{R}^+ \times R^m; \mathbb{R})$. Then*

$$d\varphi(t, X_t) = \frac{\partial\varphi(t, X_t)}{\partial x}dX_t + \frac{1}{2}\frac{\partial^2\varphi(t, X_t)}{\partial x^2}d\langle X\rangle_t + \frac{\partial\varphi(t, X_t)}{\partial t}dt =$$

$$\frac{\partial\varphi(t, X_t)}{\partial x}\sigma(t, X_t)dB_t + \Bigg\{\frac{\partial\varphi(t, X_t)}{\partial x}b(t, X_t) + \frac{1}{2}\frac{\partial^2\varphi(t, X_t)}{\partial x^2}\sigma(t, X_t)^2 + \frac{\partial\varphi(t, X_t)}{\partial t}\Bigg\}dt$$

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 22. *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.

6.5.2 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

$$\begin{aligned}
&\sum_{t_i\in\Pi} X_{(t_i+t_{i+1})/2}(M_{t_{i+1}\wedge t} - M_{t_i\wedge t}) = \\
&\sum_{t_i\in\Pi} X_{t_i}(M_{t_{i+1}\wedge t} - M_{t_i\wedge t}) + \sum_{t_i\in\Pi}(X_{(t_i+t_{i+1})/2} - X_{t_i})(M_{(t_i+t_{i+1})/2\wedge t} - M_{t_i\wedge t}) \\
&+ \sum_{t_i\in\Pi}(X_{(t_i+t_{i+1})/2} - X_{t_i})(M_{t_{i+1}\wedge t} - M_{(t_i+t_{i+1})/2\wedge t}) \\
&\stackrel{P}{\to} \int_0^t X_s d^- M_s + \frac{1}{2}[M,X]_t + 0
\end{aligned}$$

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

$$\langle f'(M), M\rangle_t = \left\langle \int_0^{\cdot} f''(M_s)dM_s, M \right\rangle_t = \int_0^t f''(M_s)d\langle M, M\rangle_s$$

6.5.3 Doss-Sussman explicit solution of a SDE

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

$$\begin{aligned} &\frac{\partial}{\partial x} u(x, y) = \sigma(u(x, y)) \\ &dY_t = \left(\frac{\partial}{\partial y} u(W_t, Y_t)\right)^{-1} b(u(W_t, Y_t))dt \end{aligned}$$

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

$$\begin{aligned} &\frac{\partial}{\partial y} u(0, y) = 1, \\ &\frac{\partial}{\partial y} u(x, y) = 1 + \int_0^x \frac{\partial^2}{\partial x \partial y} u(\xi, y)d\xi = 1 + \int_0^x \frac{\partial}{\partial y} u(\xi, y)\sigma'(u(\xi, y))d\xi = \\ &= \exp\left(\int_0^x \sigma'(u(\xi, y))d\xi\right) \end{aligned}$$

Substituting

$$Y_t = Y_0 + \int_0^t \exp\Big(-\int_0^{W_s} \sigma'(u(\xi, Y_s))d\xi\Big)b(u(W_s, Y_s))ds$$

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,

$$\circ dX_t = \exp(W_t)dY_t + Y_t \circ d\exp(W_t)$$
$$\exp(W_t)\exp(-W_t)\cos(Y_t \exp(W_t))dt + Y_t \exp(W_t) \circ dW_t = \cos(X_t)dt + X_t \circ dW_t$$

6.6 Cameron-Martin-Girsanov theorem

6.6.1 Discrete time heuristics

Let $(\Delta B_1, \dots, \Delta B_n)$ i.i.d. Gaussian random variable with $E_P(\Delta B_1) = 0$, $E_P(\Delta B_1^2) = \Delta t$, let $\mathcal{F}_n = \sigma(\Delta B_i : i = 1 \dots, n)$.

Consider another measure Q on $(\Omega, \mathcal{F}_n)$ such that under Q the ΔB_i are i.i.d. with mean $E_P(\Delta B_i) = H_i \Delta t$ and variance $E_P(\Delta B_1^2) = \Delta t$.

On $(\Omega, \mathcal{F}_n)$ the likelihood ratio factorizes as

$$\frac{dQ|\mathcal{F}_n}{dP|\mathcal{F}_n} = \prod_{k=1}^n \exp\Big(-\frac{(\Delta B_k - A_k \Delta t)^2}{2\Delta t} + \frac{(\Delta B_k)^2}{2\Delta t}\Big) =$$
$$\exp\Big(\sum_{i=1}^n A_k \Delta B_i - \frac{1}{2}\sum_{i=1}^n A_k^2 \Delta t\Big)$$

This extends to the case when under Q the random variables ΔB_k are not i.i.d. but ΔB_k conditionally on $\mathcal{F}_{k-1}$ is conditionally gaussian, with conditional mean

$$E_Q(\Delta B_k|\mathcal{F}_{k-1}) = A_k \Delta t,$$

where A_k is predictable and conditional variance

$$E_Q((\Delta B_k)^2|\mathcal{F}_{k-1}) - A_k^2 \Delta t^2 = \Delta t$$

If $A_k \in L^1(P)$ $\forall k$ then under Q

$$M_k = \sum_{i=1}^k \Delta B_i - \sum_{i=1}^k A_i \Delta t$$

is a Q-martingale with predictable variation $\langle M \rangle_k = \sum_{i=1}^k \Delta t$.

6.6.2 Change of drift in continuous time

We denote by P_t the restriction of P on the σ-algebra $\mathcal{F}_t$.

Let (M_t) a continuous $\{\mathcal{F}_t\}$-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

$$\widetilde{M_t} = M_t + \int_0^t H_s d\langle M \rangle_s, \tag{6.7}$$

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 32. *Assume that $Q \overset{loc}{\ll} P$. The likelihood ratio process*

$$Z_t(\omega) := \frac{dQ_t}{dP_t}(\omega) \tag{6.8}$$

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 33. *Let* $Q \overset{loc}{\ll} P$ *probability measures on* $(\Omega, \mathcal{F})$ *equipped with the filtration* $\mathbb{F} = \{\mathcal{F}_t\}$ *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,

$$E_Q(X_{t\wedge\tau_n} \mathbf{1}_A) = E_P(Z_t X_{t\wedge\tau_n} \mathbf{1}_A) = E_P(Z_{t\wedge\tau_n} X_{t\wedge\tau_n} \mathbf{1}_A)$$
$$E_Q(X_{s\wedge\tau_n} \mathbf{1}_A) = E_P(Z_s X_{s\wedge\tau_n} \mathbf{1}_A) = E_P(Z_{s\wedge\tau_n} X_{s\wedge\tau_n} \mathbf{1}_A)$$

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 27. *(Cameron-Martin-Girsanov) Let* $Q \overset{loc}{\ll} P$ *probability measure on* $(\Omega, \mathcal{F})$ *equipped with the filtration* $\mathbb{F} = (\mathcal{F}_t : t \ge 0)$*, and* M_t *a continuous* $\mathbb{F}$*-local martingale such that change of drift formula (6.7) holds.*

Necessarily

$$Z_t = \frac{dQ_t}{dP_t} = Y_t \exp\biggl(\int_0^t H_s dM_s - \frac{1}{2} \int_0^t H_s^2 d\langle M \rangle_s \biggr)$$

where $X_t \ge 0$ *is a continuous* P*-martingale with* $E_P(X_0) = 1$ *and* $[M, X]_t = 0$ $\forall t$*.*

We rewrite the the change of drift formula (6.7) 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 33, 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

$$Z_t = Z_0 \mathcal{E}(H \cdot M)_t = Z_0 \mathcal{E}\Big(- \int_0^{\cdot} H_s dM_s \Big)_t :=$$
$$Z_0 \exp\Big(- \int_0^t H_s dM_s - \int_0^t H_s^2 d\langle M\rangle_s \Big)$$

Here $Z_0(\omega) = \frac{dQ_0}{dP_0}(\omega)$ is $\mathcal{F}_0$-measurable.

More in general

$$Z_t = Z_0 \mathcal{E}(H \cdot M + X)_t = Z_0 \mathcal{E}(H \cdot M)_t \mathcal{E}(X)_t \quad \square$$

Notes Igor Vladimirovich Girsanov (1934-1965) was a Russian mathematician.

6.7 Stochastic filtering

Lemma 34. *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 35. *Let (B_t) be a Brownian 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

$$E_P\Big(\int_0^t H_s dB_s \Big| \mathcal{F}_t^B \Big) = \int_0^T E_P(H_s | \mathcal{F}_s^B) dB_s$$

Moreover if M_t is a $(\mathcal{G}_t)$-martingale with $\langle M, B\rangle_s = 0$, $\forall 0 \le s \le t$ then

$$E_P\big(M_t - M_0|\mathcal{F}_t^B\big) = 0$$

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 $(\mathcal{F}_t^B)$.

$$E_P\bigg(\mathbf{1}_A\int_0^t H_s dB_s\bigg) = P(A)E_P\bigg(\int_0^t H_s dB_s\bigg) + E_P\bigg(\int_0^t K_s dB_s \int_0^t H_s dB_s\bigg)$$
$$= 0 + E_P\bigg(\langle K\cdot B, H\cdot B\rangle_t\bigg) = E_P\bigg(\int_0^t K_s H_s ds\bigg) =$$
$$\int_0^t E_P(K_s H_s)ds = \int_0^t E_P\big(K_s E_P(H_s|\mathcal{F}_s)\big)ds$$
$$= E_P\bigg(\bigg\langle\int_0^\cdot K_s dB_s, \int_0^\cdot E_P(H_s|\mathcal{F}_s)dB_s\bigg\rangle_t\bigg)$$
$$= 0 + E_P\bigg(\int_0^t K_s dB_s\int_0^t E_P(H_s|\mathcal{F}_s)dB_s\bigg) = E_P\bigg(\mathbf{1}_A\int_0^t E_P(H_s|\mathcal{F}_s)dB_s\bigg) =$$

where we used the Ito isometry and the definition of conditional expectation □

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,

$$E_P((M_t - M_0)\mathbf{1}_A) = P(A)E_P(M_t - M_0) + E_P((M_t - M_0)\int_0^t K_s dB_s) =$$
$$0 + E_P\bigg(\int_0^t K_s d\langle M, B\rangle_s\bigg) = 0$$

which means $E_P(M_t - M_0|\mathcal{F}_t^B) = 0$ □

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

$$dX_s = \big\{b(s,Y,X_s) - h(s,Y,X_s)g(s,Y,X_s)\big\}ds + f(s,Y,X_s)dV_s + g(s,Y,X_s)dY_s$$

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:

$$\sigma_t(\varphi) = \sigma_0(\varphi) + \int_0^t \sigma_s(L_{s,Y}\varphi)ds + \int_0^t \sigma_s(L^1_{s,Y}\varphi)dY_s \tag{6.9}$$

where $L_{s,Y}$ and $L^1_{s,Y}$ are differential operators on C^2 depending on time and on the past observations of Y:

$$\begin{aligned}
L_{s,Y}\ \varphi &= \frac{1}{2}(f^2(s,Y,\cdot) + g^2(s,Y,\cdot))\frac{\partial^2}{\partial^2 x}\varphi + b(s,Y,\cdot)\frac{\partial}{\partial x}\varphi \\
L^1_{s,Y}\ \varphi &= h(s,Y,\cdot)\varphi + g(s,Y,\cdot)\frac{\partial}{\partial x}\varphi
\end{aligned}$$

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 $\mathcal{Y}_t$.

$$\sigma_t(\varphi) := E_Q(\varphi(X_t)Z_t|\mathcal{Y}_t) =$$
$$E_Q(\varphi(X_0)|\mathcal{Y}_t)$$
$$+E_Q\bigg(\int_0^t Z_s\{\varphi'(X_s)g(s,Y,X_s)+\varphi(X_s)h(s,Y,X_s)\}dY_s\bigg|\mathcal{Y}_t\bigg)$$
$$+E_Q\bigg(\int_0^t Z_s\varphi'(X_s)f(s,Y,X_s)dV_s\bigg|\mathcal{Y}_t\bigg)$$
$$+E_Q\bigg(\int_0^t Z_s\{\varphi'(X_s)b(s,Y,X_s)+\frac{1}{2}\varphi''(X_s)\big(f(s,Y,X_t)^2+g(s,Y,X_t)^2\big)\}ds\bigg|\mathcal{Y}_t\bigg)$$

and 6.9 follows by lemma 35.

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

$$\sigma_t(1) = \exp\bigg(\int_0^t E_P(h(s,Y,X_s)|\mathcal{Y}_s)dY_s - \frac{1}{2}\int_0^t E_P(h(s,Y,X_s)|\mathcal{Y}_s)^2 ds\bigg)$$

Consequently by the Cameron Martin Girsanov theorem (27)

$$Y_t - \int_0^t E_P(h(s,Y,X_s)|\mathcal{Y}_s)ds$$

is a P Brownian motion in the $\{\mathcal{Y}_t\}$ filtration.

6.8 Final exam

: 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 characteristic 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.