Browninan Motion. Lecture Notes. October 2006

V. Kolokoltsov

These notes, revised and extended, form the basis for Chapter 2 and 3 of my book "Markov Processes, Semigroups and Generators", de Gryuter 2011.

Aims and Objectives. Brownian motion (BM) is an acknowledged champion in stochastic modeling for a wide variety of processes in physics (statistical mechanics, quantum fields, etc), biology (e.g. population dynamics, migration, disease spreading), finances (e.g. common stock prices). BM enjoys beautiful nontrivial properties deeply linked with various areas of mathematics. The general theory of modern stochastic process is strongly rooted in BM and was largely developed by extensions of its remarkable features. The aim of the course is to learn the basic properties of BM, its potential and limitation as a modeling tool, to understand its place among the main general classes of random processes such as martingales, Markov and Lévy processes, to learn to apply the general tools of stochastic analysis (e.g. stopping times) to BM and related diffusions, and to appreciate basic notions and methods of modern stochastic analysis itself through its application to BM.

General Remarks: 1) $\star$ denotes an additional material, 2) Sections 1 and 5 contain basic probability prerequisites for the course. 3) Recommended supplementary reading: *main*: I. Karatzas, S. Shreve. Browninan Motion and Stochastic Calculus. Springer 1998; D. Revuz, M. Yor. Continuous Martingales and Brownian Motion. Springer 1999; *for references in probability*: J. Jacod, Ph. Protter. Probability Essentials. Springer 2004; A.N. Shiryayev. Probability. Springer 1984.

Content.

CHAPTER 1. DEFINITION AND CONSTRUCTION of BM (p.1).

0. Overview, historical sketch, perspectives.

1. Review of measure and probability. 2. Brownian motion: construction via Hilbert space methods. 3. The construction of BM via Kolmogorov's continuity theorem.

CHAPTER 2. The LÉVY, MARKOV AND FELLER PROCESSES (p.12).

4. Processes with s.i. increments. 5. Conditioning. 6. Markov processes. 7. Feller processes and semigroups.

CHAPTER 3. MARTINGALE METHODS (p.31).

8. Martingales. 9. Stopping times and optional sampling theorem. 10. Strong Markov property. Diffusions as Feller processes with continuous paths. 11. Reflection principle and passage times for BM.

CHAPTER 4. HEAT CONDUCTION (OR DIFFUSION) EQUATION (p.45).

12. The Dirichlet problem for diffusion operators. 13. The stationary Feynman-Kac formula. 14. Diffusions with variable drift, Ornstein-Uhlenbeck processes.

CHAPTER 5. FINE PROPERTIES of BM (p.52).

Section 15. Zeros, excursions and local times. 16. Skorohod imbedding and invariance principle. 17. Sample path properties.

CHAPTER 1. DEFINITION AND CONSTRUCTION of BM.

Section 0. Overview, historical sketch, perspectives. Brown, Einstein, Smoluchovski, Langevin, Ornstein-Uhlenbeck, Chandrasekhar, Wiener, Feynman-Kac, Nelson, McKean, Dyson.

Section 1. Review of measure and probability

Def. A collection $\mathcal{F}$ of subsets of a given set S is called a *σ-algebra* if (i) $S \in \mathcal{F}$; (ii) $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$; (iii) ($\sigma$-additivity) $\cup_{n=1}^\infty A_n \in \mathcal{F}$ whenever $A_n \in \mathcal{F}$ for any $n \in \mathbf{N}$. The pair $(S, \mathcal{F})$ is called a *measurable space*. A *measure* on $(S, \mathcal{F})$ is a mapping $\mu : \mathcal{F} \mapsto [0, \infty]$ such that $\mu(\emptyset) = 0$ and *σ-additivity* holds:

$$\mu\left(\cup_{n=1}^\infty A_n\right) = \sum_{n=1}^\infty \mu(A_n)$$

for any sequence A_n of mutually disjoint sets in $\mathcal{F}$. The triple $(S, \mathcal{F}, \mu)$ is called a *measure space*. A measure μ is called *finite* if its *total mass* $\mu(S)$ is finite and respectively *σ-finite* if there exists a sequence A_n, $n \in \mathbf{N}$, of subsets of $\mathcal{F}$ such that $S = \cup_{n=1}^\infty A_n$ and $\mu(A_n) < \infty$ for all n. For a collection of subsets Γ of a set Ω σ-algebra $\sigma(\Gamma)$ *generated by* Γ is the minimal σ-algebra containing all sets from Γ.

Examples. 1. A measure space $(\Omega, \mathcal{F}, \mu)$ is called a *probability space* whenever $\mu(\Omega) = 1$. In this case μ is called a *probability measure* and the subsets from $\mathcal{F}$ are called *events*. 2. For a topological space, e.g. a subset of $\mathbf{R}^d$, the smallest σ-algebra $\mathcal{B}(S)$ containing all its open subsets is called the *Borel σ-algebra* of S. Its elements are called *Borel sets* and any measure on $(S, \mathcal{B}(S))$ is called a *Borel measure*. The simplest example of a Borel measure is given by Lebesgue measure on $\mathbf{R}^d$. 3. For a finite or a countable family of measure spaces $(S_i, \mathcal{F}_i, \mu_i)$, $i = 1, 2, ...,$, the *product measure space* $(S, \mathcal{F}, \mu)$ is defined, where $S = S_1 \times S_2 \times ...$, $\mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2 \otimes ...$ -the σ-algebra generated by the sets $A_1 \times ... \times A_n, A_i \in \mathcal{F}_i$, $n \in \mathbf{N}$, and $\mu = \mu_1 \times \mu_2 \times ...$ is the *product measure* uniquely specified by the prescription $\mu(A_1 \times ... \times A_n) = \mu_1(A_1)...\mu_n(A_n)$.

Borel-Cantelli lemma. *If a sequence of events A_n, $n \in \mathbf{N}$, on a probability space $(\Omega, \mathcal{F}, P)$ is such that $\sum_n P(A_n) < \infty$, then a.s. only a finite number of A_n can occur.*

Proof. Let $B = \{\omega \in \Omega : \text{infinite number } A_n \text{ occur}\}$. Then

$$B = \cap_n \left(\cup_{k \geq n} A_k\right)$$

and

$$P(B) \leq P\left(\cup_{k \geq n} A_k\right) \leq \sum_{k \geq n} P(A_k) \to 0,$$

as $n \to \infty$. Hence $P(B) = 0$.

Def. Completion. For a measure space $(S, \mathcal{F}, \mu)$ a subset of S is called *negligible* if it is a subset of a $N \in \mathcal{F}$ with $\mu(N) = 0$. The σ-algebra of subsets $\bar{\mathcal{F}}$ of S of the form $A \cup B$ with $A \subset \mathcal{F}$ and B being negligible and the measure $\bar{\mu}$ on it defined on these sets as

$\bar{\mu}(A \cup B) = \mu(A)$ are called respectively the *completion* of $\mathcal{F}$ and μ (with respect to μ). In particular, for $S \subset \mathbf{R}^d$ the completion of $\mathcal{B}(S)$ with respect to Lebesgue measure is called the σ-algebra of *Lebesgue measurable sets* in S.

Def. For a probability space $(\Omega, \mathcal{F}, \mu)$ one says that some property depending on $\omega \in \Omega$ holds *almost surely* or *with probability 1* if there exists a negligible set $N \in \mathcal{F}$ such that this property holds for all $\omega \in \Omega \setminus N$.

Def. If $(S_i, \mathcal{F}_i)$, $i = 1, 2$, are measurable spaces a mapping $f : S_1 \mapsto S_2$ is called $(\mathcal{F}_1, \mathcal{F}_2)$-*measurable* if $f^{-1}(A) \in \mathcal{F}_1$ whenever $A \in \mathcal{F}_2$. If S_1, S_2 are subsets of $\mathbf{R}^d$ equipped with their Borel σ-algebra such a mapping is said to be *Borel measurable.* Speaking about measurable mapping with values in $\mathbf{R}^d$ one usually means that $\mathbf{R}^d$ is equipped with its Borel σ-algebra.

$\star$ **Exercise and Def.** For $S \subset \mathbf{R}^d$ the *universal* σ-field $\mathcal{U}(S)$ is defined as the intersection of the completions of $\mathcal{B}(S)$ with respect to all probability measures on S. The $(\mathcal{U}(S), \mathcal{B}(S))$- measurable functions are called *universally measurable.* Show that a real valued function f is universally measurable if and only if for every probability measure μ on S there exists a Borel measurable function g_μ such that $\mu\{x : f(x) \neq g_\mu(x)\} = 0$. Hint for "only if" part: show that

$$f(x) = \inf\{r \in \mathbf{Q} : x \in U(r)\}, \quad \text{where}\, U(r) = \{x \in S : f(x) \leq r\}.$$

Since $U(r)$ belong to the completion of the Borel σ-algebra with respect to μ there exist $B(r)$, $r \in \mathbf{Q}$, such that

$$\mu\left(\cup_{r\in\mathbf{Q}}(B(r)\Delta U(r))\right) = 0.$$

Define

$$g_\mu(x) = \inf\{r \in \mathbf{Q} : x \in B(r)\}.$$

Def. For a probability space $(\Omega, \mathcal{F}, P)$ the measurable mappings $X : \Omega \mapsto \mathbf{R}^d$ are called *random variables* (shortly r.v.). The *law* (or *distribution*) of such a mapping is the Borel probability measure p_X on $\mathbf{R}^d$ defined as $p_X = P \circ X^{-1}$. In other words

$$p_X(A) = P(X^{-1}(A)) = P(\omega \in \Omega : X(\omega) \in A) = P(X \in A).$$

Two r.v. X and Y are called *identically distributed* if they have the same probability law. For a real (i.e. one-dimensional) r.v. X its *distribution function* is defined by $F_X(x) = p_X((-\infty, x])$. A real r.v. X has a *continuous distribution* with a *probability density function* f if $p_X(A) = \int_A f(x)dx$ for all Borel sets A. The σ-algebra $\sigma(X)$ *generated by a r.v.* X is the smallest σ-algebra containing the sets $\{X \subset B\}$ for all Borel sets B.

Exercise. Show that if X takes only finite number of values, then the law p_X is a sum of Dirac's δ- measures.

Def. Expectation and covariance. For a $\mathbf{R}^d$-valued r.v. X on a probability space $(\Omega, \mathcal{F}, \mu)$ and a Borel measurable function $f : \mathbf{R}^d \mapsto \mathbf{R}^m$ the *expectation* $\mathbf{E}$ of $f(X)$ is defined as

$$\mathbf{E}(f(X)) = \int_\Omega f(X(\omega))P(d\omega) = \int_{\mathbf{R}^d} f(x)p_X(dx). \tag{1}$$

X is called *integrable* if $\mathbf{E}(|X|) < \infty$. For two $\mathbf{R}^d$-valued r.v. $X = (X_1, ..., X_d)$ and $Y = (Y_1, ..., Y_d)$ the $d \times d$ matrix with the entries $\mathbf{E}[(X_i - \mathbf{E}(X_i))(Y_j - \mathbf{E}(Y_j)]$ is called the *covariance* of X and Y and is denoted $Cov(X, Y)$. In case $d = 1$ and $X = Y$ the number $Cov(X, Y)$ is called the *variance* of X and is denoted by $Var(X)$ and sometimes also by σ_X^2. The r.v. X and Y are called *uncorrelated* whenever $Cov(X, Y) = 0$.

Exercise. Show that the two expressions in the definition (1.1) really coincide. Hint: first choose f to be an indicator.

Def. Four basic notions of convergence of r.v. Let X and X_n, $n \in \mathbf{N}$ be $\mathbf{R}^d$-valued r.v. (defined in a probability space). One says that X_n converges to X (i) *in* L^p $(1 \leq p < \infty)$ if $\lim_{n\to\infty} \mathbf{E}(|X_n - X|^p) = 0$; (ii) *almost surely* if $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ almost surely; (iii) *in probability* if for any $\epsilon > 0$ $\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0$; (iv) *in distribution* if p_{X_n} *weakly converges* to p_X, i.e. if

$$\lim_{n\to\infty} \int_{\mathbf{R}^d} f(x) p_{X_n}(dx) = \int_{\mathbf{R}d} f(x) p_X(dx)$$

for all bounded continuous functions f.

Exer. Show that L^p-convergence $\Rightarrow$ convergence in probability $\Rightarrow$ weak convergence. Hint: for the first $\Rightarrow$ use Chebyshev inequality; for the second one decompose the integral $\int |f(X_n(\omega)) - f(X(\omega))| P(d\omega)$ into the three terms over the sets $\{|X_n - X| > \delta\}$, $\{|X_n - X| \leq \delta, |X| > 1/\delta\}$ and $\{|X_n - X| \leq \delta, |X| \leq 1/\delta\}$.

Exer. (i) Show that $X_n \to X$ in probability $\Leftrightarrow$

$$\lim_{n\to\infty} \mathbf{E}\left(\frac{|X_n - X|}{1 + |X_n - X|}\right) = 0.$$

(ii) Deduce from (i) that almost sure convergence implies convergence in probability. Hint for (i): it is enough to prove it for $X = 0$; for "if part" then use the inequality

$$\mathbf{E}\left(\frac{|X_n|}{1 + |X_n|}\right) \geq \frac{\epsilon}{1+\epsilon} P\left(\frac{|X_n|}{1 + |X_n|} > \epsilon\right);$$

for "only if part" decompose the integral $\mathbf{E}(|X_n|/(1 + |X_n|)$ into the two terms over the sets $|X_n| > \epsilon$ and $|X_n| < \epsilon$.

Example. Consider the following sequence of indicator functions $\{X_n\}$ on $[0, 1]$: $\mathbf{1}_{[0,1]}$, $\mathbf{1}_{[0,1/2]}$, $\mathbf{1}_{[1/2,1]}$, $\mathbf{1}_{[0,1/3]}$, $\mathbf{1}_{[1/3,2/3]}$, $\mathbf{1}_{[2/3,1]}$, $\mathbf{1}_{[0,1/4]}$, $\mathbf{1}_{[1/4,2/4]}$, etc. Then $X_n \to 0$ as $n \to \infty$ in probability and in all L^p, $p \geq 1$, but not a.s.; in fact $\limsup X_n(x) = 1$ and $\liminf X_n(x) = 0$ for each x so that $X_n(x) \to X(x)$ nowhere.

Exer. (i) Convince yourself that $X_n \to X$ in distribution does not imply $X_n - X \to 0$ in distribution.(ii) Show that $X_n \to 0$ in distribution $\Rightarrow$ $X_n \to 0$ in probability.

$\star$ **Exer.** Show that $X_n \to X$ a.s. $\Leftrightarrow$

$$\lim_{m\to\infty} P\{\sup_{n\geq m} |X_n - X| > \epsilon\} = 0$$

for all $\epsilon > 0$. Use this to give another proof of the fact that convergence a.s. implies convergence in probability. Hint: observe that the event $X_n \to X$ is complement to the event

$$B = \cup_{r\in \mathbf{Q}} B_r, \qquad B_r = \cap_{m\in\mathbf{Q}}\{\sup_{n\geq m} |X_n - X| > 1/r\},$$

i.e., a.s. convergence is equivalent to $P(B) = 0$ and hence to $P(B_r) = 0$ for all r.

Def. A family H of $L^1(\Omega, \mathcal{F}, \mu)$ is *uniformly integrable* if

$$\lim_{c\to\infty} \sup_{X\in H} \mathbf{E}(|X|\mathbf{1}_{|X|>c}) = 0.$$

Exer. If either (i) $\sup_{X\in H} \mathbf{E}(|X|^p) < \infty$ for a $p > 1$, or (ii) $\exists$ an integrable r.v. Y s.t. $|X| \leq Y$ for all $X \in H$, then H is uniformly integrable. Hint:

$$(i) \qquad \mathbf{E}(|X|\mathbf{1}_{|X|>c}) < \frac{1}{c^{p-1}}\mathbf{E}(|X|^p\mathbf{1}_{|X|>c}) < \frac{1}{c^{p-1}}\mathbf{E}(|X|^p),$$

$$(ii) \qquad \mathbf{E}(|X|\mathbf{1}_{|X|>c}) < \mathbf{E}(Y\mathbf{1}_{Y>c}).$$

Exer. If $X_n \to X$ a.s. and $\{X_n\}$ is uniformly integrable, then $X_n \to X$ in L^1. Hint: decompose the integral $\int |X_n - X|p(d\omega)$ into the sum of three over the domains $\{|X_n - X| > \epsilon\}$, $\{|X_n - X| \leq \epsilon, |X| \leq c\}$ and $\{|X_n - X| \leq \epsilon, |X| > c\}$, and they can be made small respectively because $X_n \to X$ in probability (as it holds a.s.), by dominated convergence an by uniform integrability.

Def. Characteristic functions. If p is a probability measure on $\mathbf{R}^d$ its *characteristic function* is the funtion $\phi_p(y) = \int e^{i(y,x)}p(dx)$. For a $\mathbf{R}^d$-valued r.v. X its *characteristic function* is defined as the characteristic function $\phi_X = \phi_{p_X}$ of its law p_X, i.e.

$$\phi_X(y) = \mathbf{E}(e^{i(y,x)}) = \int_{\mathbf{R}^d} e^{i(y,x)}p_X(dx).$$

Exer. Show that any ch.f. is a continuous function. Hint: use

$$|\phi_X(y+h) - \phi_X(y)| \leq \mathbf{E}|e^{ihX} - 1| \leq \max_{|x|\leq a} |e^{ihx} - 1| + 2P(|X| > a). \tag{2}$$

$\star$ **Riemann-Lebesgue Lemma.** *If a probability measure p has a density, then ϕ_p belongs to $C_\infty(\mathbf{R}^d)$ (continuous functions tending to 0 as its argument tends to ∞). In other words, the inverse Fourier transform*

$$f \to F^{-1}f(y) = (2\pi)^{-d/2} \int e^{i(y,x)} f(x)dx$$

is a bounded linear operator $L^1(\mathbf{R}^d) \mapsto C_\infty(\mathbf{R}^d)$.

Sketch of the proof. Reduce to the case, when f is a continuously differentiable function with a compact support, then use integration by part.

Exercise and Def. For a vector $m \in \mathbf{R}^d$ and a positive definite $d \times d$-matrix A a r.v. X is called *Gaussian* (or has *Gaussian distribution*) *with mean m and covariance A* and is denoted by $N(m, A)$ whenever its characteristic function is

$$\phi_{N(m,A)}(y) = \exp\{i(m, y) - \frac{1}{2}(y, Ay)\}.$$

(i) Show that if A is non-degenerate, a $N(m, A)$ r.v. have a distribution with the pdf

$$f(x) = \frac{1}{(2\pi)^{d/2}\sqrt{\det(A)}} \exp\{-\frac{1}{2}(x - m, A^{-1}(x - m))\}.$$

(ii) Show that $m = \mathbf{E}(X)$ and $A_{ij} = \mathbf{E}((X_i - m_i)(X_j - m_j))$.

Exercise and Def. suppose X_1 and X_2 are independent $\mathbf{R}^d$-valued r.v. with laws μ_1, μ_2 and characteristic functions ϕ_1 and ϕ_2. (i) Show that the r.v. $X_1 + X_2$ has the characteristic function $\phi_1\phi_2$ and the law given by the *convolution* $\mu_1 \star \mu_2$ defined by

$$(\mu_1 \star \mu_2)(A) = \int_{\mathbf{R}^d} \mu_1(A - x)\mu_2(dx) = \int_{\mathbf{R}^{2d}} \chi_{A-x}(y)\mu_1(dy)\mu_2(dx).$$

(ii) Extend this result to the case of n independent r.v. X_1,..., X_n.

$\star$ **Exer. and Def.** Show that if probability distributions p_n, $n \in \mathbf{N}$, converge weakly to a probability distribution p, then (i) the family p_n is *tight*, i.e.

$$\forall \epsilon > 0\, \exists K > 0 : \forall n,\, p_n(|x| > K) < \epsilon;$$

(ii) their characteristic functions ϕ_n converge uniformly on compact sets. Hint: for (ii) use tightness and representation (2) to show that the family of ch.f is equicontinuous, i.e.

$$\forall \epsilon\, \exists \delta : |\phi_n(y + h) - \phi(y)| < \epsilon \quad \forall h < \delta, n \in \mathbf{N}$$

which implies the uniform convergence.

Glivenko's Theorem. *If ϕ_n, $n \in \mathbf{N}$, and ϕ are the characteristic functions of probability distributions p_n and p on $\mathbf{R}^d$, then $\lim_{n\to\infty} \phi_n(y) = \phi(y)$ for each $y \in \mathbf{R}^d$ if and only if p_n converge to p weakly.*

Lévy's Theorem. *If ϕ_n, $n \in \mathbf{N}$, is a sequence of characteristic functions of probability distributions on $\mathbf{R}^d$ and $\lim_{n\to\infty} \phi_n(y) = \phi(y)$ for each $y \in \mathbf{R}^d$ and a function ϕ, which is continuous at the origin, then ϕ is itself a characteristic function.*

$\star$ **Exer.** Show that if a family of probability measures p_α is tight, then it is *relatively weakly compact*, i.e. any sequence of this family has a weakly convergent subsequence. Hint: tight $\Rightarrow$ family of characteristic functions is equicontinuous (by (2)), and hence is relatively compact in the topology of uniform convergence on compact sets. Finally use Levy's theorem.

Exercise. (i) Show that a finite linear combination of $\mathbf{R}^d$-valued Gaussian r.v. is again a Gaussian r.v. (ii) Show that if a sequence of $\mathbf{R}^d$-valued Gaussian r.v. converges in distribution to a r.v., then the limiting r.v. is again Gaussian. (iii) Show that if (X, Y)

is a $\mathbf{R}^2$-valued Gaussian r.v., then X and Y are uncorrelated if and only if they are independent.

$\star$ **Bochner's Theorem**. *A function* $\phi : \mathbf{R}^d \mapsto \mathbf{C}$ *is a characteristic function of a probability distribution if and only if it satisfies the following three properties: (i)* $\phi(0) = 1$*; (ii)* ϕ *is continuous at the origin; (iii)* ϕ *is positive definite, which means that*

$$\sum_{j,k=1}^{d} c_j \bar{c}_k \phi(y_j - y_k) \geq 0$$

for all real $y_1, ..., y_d$ *and all complex* $c_1, ..., c_d$.

Exercise. Prove the "only if" part of Bochner's theorem. Hint: for (iii) observe that

$$\sum_{j,k=1}^{d} c_j \bar{c}_k \phi_X(y_j - y_k) = \int_{\mathbf{R}^d} \sum_{j,k=1}^{d} c_j \bar{c}_k e^{i(y_j - y_k, x)} p_X(dx)$$

$$= \int_{\mathbf{R}^d} \left(\sum_{j=1}^{d} c_j e^{i(y_j, x)} \right) p_X(dx).$$

Def. Stochastic processes. A *stochastic process* is a collection $X = (X_t),\ t \geq 0)$ (or $t \in [0, T]$ for some $T > 0$) of $\mathbf{R}^d$- valued random variables defined on the same probability space. The *finite-dimensional distributions* of such a process are the collection of probability measures $p_{t_1,...,t_n}$ on $\mathbf{R}^{dn}$ (parametrized by finite collections of pairwise different non-negative numbers $t_1, ..., t_n$) defined as

$$p_{t_1,...,t_n}(H) = P((X_{t_1}, ..., X_{t_n}) \in H)$$

for each Borel subset H of $\mathbf{R}^{dn}$. These finite-dimensional distributions are (obviously) *consistent* (or satisfy *Kolmogorov's consistency criteria*): for any n, any permutation π of $\{1, ..., n\}$, any sequence $0 \leq t_1 < ... < t_{n+1}$, and any collection of Borel subsets $H_1, ..., H_n$ of $\mathbf{R}^d$ one has

$$p_{t_1,...,t_n}(H_1 \times ... \times H_n) = p_{t_{\pi(1)},...,t_{\pi(n)}}(H_{\pi(1)} \times ... \times H_{\pi(n)}),$$

$$p_{t_1,...,t_n,t_{n+1}}(H_1 \times ... \times H_n \times \mathbf{R}^d) = p_{t_1,...,t_n}(H_1 \times ... \times H_n).$$

Def. A stochastic process is called *Gaussian* if all its finite-dimensional distributions are Gaussian.

Kolmogorov's existence theorem. Given a family of probability measures $p_{t_1,...,t_n}$ (on $\mathbf{R}^{dn}$) satisfying the Kolmogorov consistency criteria, there exists a probability space $(\Omega, \mathcal{F}, P)$ and a stochastic process X on it having $p_{t_1,...,t_n}$ as its finite-dimensional distribution. In particular, one can choose Ω to be the set $(\mathbf{R}^d)^{\mathbf{R}^+}$ of all mappings from $\mathbf{R}^+$ to $\mathbf{R}^d$ and $\mathcal{F}$ to be the smallest σ-algebra containing all *cylinder sets*

$$I^H_{t_1,...,t_n} = \{\omega \in \Omega : (\omega(t_1), ..., \omega(t_n)) \in H\}, \quad H \in \mathcal{B}(\mathbf{R}^{dn}),$$

and X to be the *co-ordinate process* $X_t(\omega) = \omega(t)$.

Def. "Sameness" between processes. Suppose two processes X and Y are defined on the same probability space $(\Omega, \mathcal{F}, P)$. Then (i) X and Y are called *indistinguishable* if $P(\forall t X_t = Y_t) = 1$; (ii) X is a modification of Y if for each t $P(X_t = Y_t) = 1$.

Example. Consider a positive r.v. ξ with a continuous distribution (i.e. such that $P(\xi = x) = 0$ for any x). Put $X_t = 0$ for all t and let Y_t be 1 for $t = \xi$ and 0 otherwise. Then Y is a modification of X, but $P(\forall t X_t = Y_t) = 0$.

Exercise. Suppose Y is a modification of X and both processes have right-continuous sample paths. Then X and Y are indistinguishable. Hint: show that if X is a modification of Y, then $P(\forall t \in \mathbf{Q}\ X_t = Y_t) = 1$.

Monotone class theorem. *Let $\mathcal{S}$ be a collection of subsets of a set Ω s.t. (i) $\Omega \in \mathcal{S}$, (ii) $A, B \in \mathcal{S} \Rightarrow A \setminus B \in \mathcal{S}$, (iii) $A_1 \subset A_2 \subset ..., A_n \in \mathcal{S} \Rightarrow \cup_n A_n \in \mathcal{S}$. If a collection of subsets Γ belongs to S and is closed under pairwise intersection, then $\sigma(\Gamma) \in \mathcal{S}$.*

This result is routinely used in stochastic analysis to check a validity of a certain property for elements of $\sigma(\Gamma)$, where Γ is a collection of subsets closed under intersection. According to the theorem it is sufficient to check that the validity of this property is preserved under the set substraction and countable unions.

Theorem (strong law of large numbers). *If $\xi_1, \xi_2, ...$ is a collection of iid r.v. with $\mathbf{E}\xi_j = m$, then the means $(\xi_1 + ... + \xi_n)/n$ converge a.s. (and in L^1) to m.*

$\star$ **Riesz-Markov Theorem**. *Any positive bounded linear functional on the space $C_\infty(\mathbf{R}^d)$ has form $f \mapsto \int f(x)\mu(dx)$ for some finite Borel measure μ.*

Section 2. Brownian motion: construction via Hilbert space methods.

Main Def. A *Brownian motion (or a Wiener process) with variance σ^2* is a Gaussian process B_t (defined on a probability space $(\Omega, \mathcal{F}, P)$) satisfying the following conditions: (i) $B_0 = 0$ a.s.; (ii) the increments $B_t - B_s$ have distribution $N(0, \sigma^2(t-s))$ for all $0 \le s < t$; (iii) the r.v. $B_{t_2} - B_{t_1}$ and $B_{t_4} - B_{t_3}$ are independent whenever $t_1 \le t_2 \le t_3 \le t_4$; (iv) the trajectories $t \mapsto B_t$ are continuous a.s. Brownian motion with $\sigma = 1$ is called the *standard Wiener process or Brownian motion.*

Exer. 1. A Gaussian process B_t satisfying conditions (i) and (iv) of the above definition is a Brownian motion if and only if $\mathbf{E}B_t = 0$ and $\mathbf{E}(B_t B_s) = \sigma^2 \min(s,t)$ for any t, s. Hint: $\mathbf{E}(B_t B_s) = \sigma^2 \min(s,t)$ implies $\mathbf{E}((B_t - B_s)B_s) = 0$ for $t > s$. Hence $B_t - B_s, B_s$ are uncorrelated and consequently independent (being Gaussian).

Exer. 2 (elementary transformations of BM). Let B_t be a BM. Then so are the processes (i) $B_t^c = \frac{1}{\sqrt{c}} B_{ct}$ for any $c > 0$ (scaling), (ii) $-B_t$ (symmetry), (iii) $B_T - B_{T-t}, t \in [0, T]$ for any $T > 0$ (time reversal), (iv) $tB_{1/t}$ (time inversion). Hint: for (iv) in order to get continuity at the origin deduce from the law of large numbers that $B_t/t \to 0$ as $t \to \infty$ a.s.

Recall: Hilbert spaces, basis, Parceval.

Def. The *Haar functions* H_k^n, $n = 1, 2, ..., k = 0, 1, ..., 2^{n-1} - 1$, on $[0, 1]$ are defined as

$$H_k^n(t) = \begin{cases} 2^{(n-1)/2}, & k/2^{n-1} \le t < (k+1/2)/2^{n-1}, \\ -2^{(n-1)/2}, & (k+1/2)/2^{n-1} \le t < (k+1)/2^{n-1}, \\ 0, & \text{otherwise} \end{cases}$$

and the *Schauder functions* as $S_k^n(t) = \int_0^t H_k^n(u)\,du$. The system of Haar functions is known to be an orthonormal basis in $L^2[0,1]$.

Exer. 3. Check the orthogonality condition: $(H_k^n, H_l^m) = \int_0^1 H_k^n(x)H_l^m(x)\,dx = \delta_m^n \delta_l^k$. Hint: supports of H_k^n, H_l^n do not intersect for $k \neq l$.

Let ξ_k^n, $n = 1, 2, ...,$ $k = 0, 1, ..., 2^{n-1}$, be mutually independent $N(0,1)$ r.v. on a probability space $(\Omega, \mathcal{F}, P)$.

Exer. 4. Point out a probability space $(\Omega, \mathcal{F}, P)$, on which such a family can be defined.

Consider the partial sums

$$B_t^m = \sum_{n=1}^{m} f_n(t,\omega), \qquad f_n(t,\omega) = \sum_{k=0}^{2^{n-1}-1} \xi_k^n(\omega) S_k^n(t). \tag{1}$$

The main technical ingredient of the construction is the following

Lemma. There exists a subset $\Omega_0 \subset \Omega$ such that B_t^m converges as $m \to \infty$ uniformly on $[0,1]$ for all $\omega \in \Omega_0$ and $P(\Omega_0) = 1$.

Proof. Let

$$M_n(\omega) = \max\{|\xi_j^n| : 0 \leq j \leq 2^{n-1} - 1\}$$

Since

$$P(M_n > a) \leq \sum_{j=0}^{2^{n-1}-1} P(|\xi_j^n| > a)$$

$$= 2^n \frac{1}{\sqrt{2\pi}} \int_a^\infty e^{-x^2/2} dx \leq 2^n \frac{1}{\sqrt{2\pi}} \int_a^\infty \frac{x}{a} e^{-x^2/2} dx = 2^n \frac{1}{\sqrt{2\pi}} a^{-1} e^{-a^2/2},$$

one sees that

$$\sum_{n=1}^{\infty} P(M_n > n) \leq \frac{1}{\sqrt{2\pi}} \sum_{n=1}^{\infty} 2^n \frac{1}{n} e^{-n^2/2} < \infty.$$

Hence by Borel-Cantelli $P(\Omega_0) = 1$, where

$$\Omega_0 = \{\omega : M_n(\omega) \leq n \,\text{for large enough}\, n\}.$$

Consequently for $\omega \in \Omega_0$

$$|f_n(t,\omega)| \leq n \sum_{k=0}^{2^{n-1}-1} S_k^n(t) \leq n 2^{-(n+1)/2}$$

for all large enough n, because $\max_t S_k^n(t) = 2^{-(n+1)/2}$ and the functions S_k^n have non-intersecting supports for different k. This implies that

$$\sum_{n=0}^{\infty} \max_{0 \leq t \leq 1} |f_n(t,\omega)| < \infty$$

on Ω_0, which clearly implies the claim of the Lemma.

Main Theorem. Let B_t denote the limit of (1) for $\omega \in \Omega_0$ and let us put $B_t = 0$ for ω outside Ω_0. Then B_t is a standard Brownian motion on $[0, 1]$.

Proof. Since B_t is continuous in t as a uniform limit of continuous functions, the conditions (i) and (iv) of the definition hold. Moreover, the finite-dimensional distributions are clearly Gaussian and $\mathbf{E}B_t = 0$. Next, since

$$\sum (S_k^n)^2(t) = \sum (\mathbf{1}_{[0,t]}, H_k^n)^2 = (\mathbf{1}_{[0,t]}, \mathbf{1}_{[0,t]}) = t < \infty$$

(by Parceval) it follows that

$$\mathbf{E}[B_t - B_t^m]^2 = \sum_{n>m} \sum_{k=0}^{2^{n-1}-1} (S_k^n(t))^2 \to 0$$

as $m \to \infty$, and consequently B_t^m converge to B_t also in L_2. Hence one deduces that

$$\mathbf{E}(B_t B_s) = \lim_{m\to\infty} \mathbf{E}(B_t^m B_s^m) = \sum_{n=0}^{\infty} \sum_{j=0}^{2^{n-1}-1} (\mathbf{1}_{[0,t]}, H_j^n)(\mathbf{1}_{[0,s]}, H_j^n) = (\mathbf{1}_{[0,t]}, \mathbf{1}_{[0,s]}) = \min(t, s),$$

which completes the proof.

Corollary. *A standard Brownian motion exists on* $\{t \geq 0\}$.

Proof. By the main theorem there exists a sequence $(\Omega_n, \mathcal{F}_n, P_n)$, $n = 1, 2, ...$, of probability spaces with Brownian motions W_n on each of them. Take the product probability space Ω and define B on it recursively

$$B_t = B_n + W_{t-n}^{n+1}, \quad n \leq t \leq n+1.$$

Section 3. The construction of BM via Kolmogorov's continuity theorem.

The Kolmogorov-Chentsov Continuity Theorem. Suppose a process X_t, $t \in [0, T]$ on a probability space $(\Omega, \mathcal{F}, P)$ satisfies the condition

$$\mathbf{E}|X_t - X_s|^\alpha \leq C|t-s|^{1+\beta}, \quad 0 \leq s, t \leq T,$$

for some positive constants α, β, C. Then there exists a continuous modification $\tilde{X}_t$ of X_t, which is a.s. locally Hölder continuous with exponent γ for every $\gamma \in (0, \beta/\alpha)$, i.e.

$$P\left[\omega : \sup_{s,t\in[0,T]:|t-s|<h(\omega)} \frac{|\tilde{X}_t(\omega) - \tilde{X}_s(\omega)|}{|t-s|^\gamma} \leq \delta\right] = 1, \tag{1}$$

where $h(\omega)$ is an a.s. positive r.v. and $\delta > 0$ is a constant.

Proof. **Step 1.** By Chebyshev

$$P(|X_t - X_s| \geq \epsilon) \leq \epsilon^{-\alpha} E|X_t - X_s|^\alpha \leq C\epsilon^{-\alpha}|t-s|^{1+\beta}$$

and hence $X_s \to X_t$ in probability as $s \to t$.

Step 2. Setting $t = k/2^n$, $s = (k-1)/2^n$, $\epsilon = 2^{-\gamma n}$ in the above inequality yields

$$P(|X_{k/2^n} - X_{(k-1)/2^n}| \geq 2^{-\gamma n}) \leq C2^{-n(1+\beta-\alpha\gamma)}.$$

Hence

$$P\left(\max_{1\leq k\leq 2^n} |X_{k/2^n} - X_{(k-1)/2^n}| \geq 2^{-\gamma n}\right) \leq \sum_{k=1}^{2^n} P(|X_{k/2^n} - X_{(k-1)/2^n}| \geq 2^{-\gamma n}) \leq C2^{-n(\beta-\alpha\gamma)}.$$

By Borel-Cantelli (by the assumption $\beta - \alpha\gamma > 0$) there exists Ω_0 of measure 1 such that

$$\max_{1\leq k\leq 2^n} |X_{k/2^n} - X_{(k-1)/2^n}| < 2^{-\gamma n}, \qquad \forall n \geq n^\star(\omega), \tag{2}$$

where $n^\star(\omega)$ is a positive, integer-valued r.v.

Step 3. For each $n \geq 1$ define $D_n = \{k/2^n : k = 0, 1, ..., 2^n\}$ and $D = \cup_{n=1}^\infty D_n$. For a given $\omega \in \Omega_0$ and $n \geq n^\star(\omega)$ we shall show that $\forall m > n$

$$|X_t(\omega) - X_s(\omega)| \leq 2 \sum_{j=n+1}^{m} 2^{-\gamma j}, \quad \forall t, s \in D_m : 0 < t - s < 2^{-n}. \tag{3}$$

For $m = n+1$ necessarily $t - s = 2^{-(n+1)}$ and (3) follows from (2) with n replaced by $n+1$. Suppose (3) is valid for $m = n+1, ..., M-1$. Take $s < t : s, t \in D_M$ and define the numbers

$$\tau_{max} = \max\{u \in D_{M-1} : u \leq t\}, \quad \tau_{min} = \min\{u \in D_{M-1} : u \geq s\}$$

so that

$$s \leq \tau_{\min} \leq \tau_{\max} \leq t; \qquad \max(\tau_{\min} - s, t - \tau_{\max}) \leq 2^{-M}.$$

Hence from (2)

$$|X_{t_{\min}}(\omega) - X_s(\omega)| \leq 2^{-\gamma M}, \quad |X_{t_{\max}}(\omega) - X_t(\omega)| \leq 2^{-\gamma M},$$

and from (3) with $m = M-1$

$$|X_{\tau_{\max}}(\omega) - X_{\tau_{\min}}(\omega)| \leq 2 \sum_{j=n+1}^{M-1} 2^{-\gamma j},$$

which implies (3) with $m = M$.

Step 4. For $s, t \in D$ with

$$0 < t - s < h(\omega) = 2^{-n^\star(\omega)}$$

choose $n > n^\star(\omega)$ s.t.

$$2^{-(n+1)} \leq t - s < 2^{-n}.$$

By (3)

$$|X_t(\omega) - X_s(\omega)| \leq 2 \sum_{j=n+1}^{\infty} 2^{-\gamma j} \leq 2(1 - 2^{-\gamma})^{-1} 2^{-(n+1)\gamma} \leq 2(1 - 2^{-\gamma})^{-1} |t - s|^\gamma,$$

which implies the uniform continuity of X_t with respect to $t \in D$ for $\omega \in \Omega_0$.

Step 5. Define $\tilde{X}_t = \lim_{s \to t, s \in \mathbf{Q}} X_s$ for $\omega \in \Omega_0$ and zero otherwise. Then $\tilde{X}_t$ is continuous and satisfies (1) with $\delta = 2(1 - 2^{-\gamma})^{-1}$.

Step 6. $\tilde{X}_s = X_s$ for $s \in \mathbf{Q}$. Then $\tilde{X}_t = X_t$ a.s. for all t, because $X_s \to X_t$ in probability and $X_s \to \tilde{X}_t$ a.s.

Exer. Show that for any $n \in \mathbf{N}$ there exists a constant C_n s.t. $\mathbf{E}|X|^{2n} = C_n \sigma^{2n}$ for a r.v. X with the normal distribution $N(0, \sigma^2)$.

Corollary 1. *$\exists$ a probability measure P on $(\mathbf{R}^{[0,\infty)}, \mathcal{B}(\mathbf{R}^{[0,\infty)}))$ and a stochastic process W_t on it which is a BM under P.*

Proof. By Kolmogorov's existence theorem $\exists$ P s.t. co-ordinate process X_t satisfies all properties, but for continuity (if needed, details are given for general Markov processes in Section 6). By Kolmogorov's continuity theorem and the Exercise above, for each T there exists a continuous modification W^T on $[0, T]$. Set

$$\Omega_T = \{\omega : W_t^T(\omega) = X_t(\omega) \, \forall t \in [0, T] \cap \mathbf{Q}\}, \qquad \Omega_0 = \cap_{T=1}^\infty \Omega_t.$$

As $W_t^T = W_t^S$ for $t \in [0, \min(T, S)]$ (continuous modifications of each other), their common values define a required process on $t \geq 0$.

Corollary 2. *BM is a.s. Hölder continuous with any exponent $\gamma \in (0, 1/2)$.*

Proof. From Kolmogorov's theorem and above exercise it follows that BM is a.s. Hölder continuous with exponent γ whenever $\gamma < (n-1)/2n$ for some positive n.

CHAPTER 2. The LÉVY, MARKOV AND FELLER PROCESSES.

Section 4. Processes with s.i. increments.

Def. A probability measure μ on $\mathbf{R}^d$ with a ch.f. ϕ_μ is called *infinitely divisible* if, for all $n \in \mathbf{N}$, there exists a probability measure ν such that $\mu = \nu \star ... \star \nu$ (n times) $\Leftrightarrow \phi_\mu(y) = f^n(y)$ with f being a ch.f. of a probability measure.

Exer. 1. Convince yourself that two definitions above are actually equivalent.

Def. and Exer. A r.v. X is called *infinitely divisible* whenever its law p_X is infinitely divisible. Show that this is equivalent to the existence, for any n, of iid r.v. Y_j, $j = 1, ..., n$, s.t. $Y_1 + ... + Y_n$ has the law p_X.

Exer. 2. Convince yourself that any Gaussian distribution is infinitely divisible.

Examples. (i) A r.v. N with non-negative integers as a range is called *Poisson with the mean (or parameter)* $c > 0$ if

$$P(N = n) = \frac{c^n}{n!} e^{-c}.$$

Check (**Exer.!**) that $\mathbf{E}(N) = Var(N) = c$ and that the ch.f. of N is $\phi_N(y) = \exp\{c(e^{iy} - 1)\}$. This implies that N is infinitely divisible. (ii) Let now $Z(n)$, $n \in \mathbf{N}$, be a sequence of $\mathbf{R}^d$-valued iid r.v. with law μ_Z. The *compound Poisson r.v.* is $X = Z(1) + ... + Z(N)$ (random walk with a random number of steps).

Let us check that

$$\phi_X(y) = \exp\{\int_{\mathbf{R}^d} (e^{i(y,x)} - 1) c\mu_Z(dx)\}.$$

In fact,

$$\phi_X(y) = \sum_{n=0}^{\infty} \mathbf{E}(\exp\{i(y, Z(1) + ... + Z(N))\} | N = n) P(N = n)$$

$$= \sum_{n=0}^{\infty} \mathbf{E}(\exp\{i(y, Z(1) + ... + Z(n))\}) \frac{c^n}{n!} e^{-c} = \sum_{n=0}^{\infty} \phi_Z^n(y) \frac{c^n}{n!} e^{-c} = \exp\{c(\phi_Z(y) - 1)\}.$$

Def. A Borel measure ν on $\mathbf{R}^d \setminus \{0\}$ is called a *Lévy measure* if

$$\int_{\mathbf{R}^d \setminus \{0\}} \min(1, x^2) \nu(dx) < \infty.$$

Theorem 1 (the Lévy-Khintchine formula). For any $b \in \mathbf{R}^d$, a positive definite $d \times d$ matrix A and a Lévy measure ν the function

$$\phi(u) = \exp\{i(b, u) - \frac{1}{2}(u, Au) + \int_{\mathbf{R}^d \setminus \{0\}} [e^{i(u,y)} - 1 - i(u, y)\mathbf{1}_{B_1}(y)] \nu(dy)\} \tag{1}$$

is a characteristic function of an infinitely divisible measure, where B_a denotes a ball of radius a in $\mathbf{R}^d$. Conversely, any infinite divisible distribution has a characteristic function of this form.

Proof (in one direction only). If any function of form (1) is a ch.f., then it is infinitely divisible (as its roots have the same form). To show the latter we introduce the approximations

$$\phi_n(u) = \exp\{i \left(b - \int_{B_1 \setminus B_{1/n}} y\nu(dy), u \right) - \frac{1}{2}(u, Au) + \int_{\mathbf{R}^d \setminus B_{1/n}} (e^{i(u,y)} - 1) \nu(dy)\}. \tag{2}$$

Each ϕ_n is a ch.f. (of the convolution of a normal distribution and an independent compound Poisson) and $\phi_n(u) \to \phi(u)$ for any u. By Lévy theorem one needs to show only that ϕ is continuous at zero. This is easy (check it!).

Def. Writing $\phi(u) = e^{\eta(u)}$ in (4.1) the mapping η is called the *characteristic exponent* or *Lévy exponent* or *Lévy symbol* of ϕ (or of its distribution).

$\star$ **Theorem 2.** *Any infinitely divisible probability measure μ is a weak limit of a sequence of compound Poisson distributions.*

Proof. Let ϕ be a ch.f. of μ so that $\phi^{1/n}$ is the ch.f. of its "convolution root" μ_n. Define

$$\phi_n(u) = \exp\{n[\phi^{1/n}(u) - 1]\} = \exp\{\int_{\mathbf{R}^d} (e^{i(u,y)} - 1) n\mu_n(dy)\}.$$

Each ϕ_n is a ch.f.of a compound Poisson process and

$$\phi_n = \exp\{n(e^{(1/n)\ln\phi(u)} - 1)\} \to \phi(u), \quad n \to \infty.$$

Completes by Glivenko's theorem.

Def. Processes with s.i. increments. A process $X = X_t$, $t \geq 0$, has *independent increments* if for any collection of times $0 \leq t_1 < ... < t_{n+1}$ the r.v. $X_{t_{j+1}} - X_{t_j}$, $j = 1, ..., n$ are independent and it has *stationary increments* if $X_t - X_s$ is distributed like $X_{t-s} - X_0$ for any $t > s$. X is a *Lévy process* if (i) $X_0 = 0$ a.s., (ii) X has s.i. increments; (iii) X is *stochastically continuous*, i.e. $\forall\, a > 0$, $s \geq 0$

$$\lim_{t \to s} P(|X(t) - X(s)| > a) = 0.$$

Under (i), (ii), the latter is equivalent to $\lim_{t\to 0} P(|X(t)| > a) = 0$ for all $a > 0$.

Alternative version of the definition of the Lévy processes requires the right continuity of paths instead of stochastic continuity. At the end of the day this leads to the same class of processes, because, on the one hand, conclusions of Theorems 3 and 4 below are easily seen to remain valid under this assumption (which leads to stochastic continuity), and on the other hand, any Lévy process as defined above has a right continuous modification, as we shall see later. So we shall usually consider the right continuous modifications of the Lévy processes.

Theorem 3. *If X_t is stochastically continuous, then the map $t \mapsto \phi_{X_t}(u)$ is continuous for each u.*

Proof. Follows from

$$|\phi_{X_t}(u) - \phi_{X_s}(u)| = \int e^{i(u,X_s)}[e^{i(u,X_t - X_s)} - 1](\omega) P(d\omega)$$

$$\leq \int |e^{i(u,y)} - 1| P_{X_t - X_s}(dy) \leq \sup_{|y|<\delta} |e^{i(u,y)} - 1| + 2P(|X_t - X_s| > \delta).$$

Exer. 3. Let a right continuous function $f : \mathbf{R}_+ \mapsto \mathbf{C}$ satisfy $f(t+s) = f(t)f(s)$ and $f(0) = 1$. Show that $f(t) = e^{t\alpha}$ with some α. Hint: consider first $t \in \mathbf{N}$, then $t \in \mathbf{Q}$, then use continuity.

Theorem 4. *If X is a Lévy process, then X_t is infinitely divisible for all t and $\phi_{X_t}(u) = e^{t\eta(u)}$, where $\eta(u)$ is the Lévy symbol of X_1.*

Proof. $\phi_{X_{t+s}}(u) = \phi_{X_t}(u)\phi_{X_s}(u)$ and $\phi_{X_0}(u) = 1$. Hence by Exercise $\phi_{X_t} = \exp\{t\alpha(u)\}$. But $\phi_{X_1} = \exp\{\eta(u)\}$.

Example. Convince yourself that the Brownian motion is a Lévy process.

Def. *The Poisson process of intensity* $c > 0$ is a right continuous Lévy process s.t. each r.v. N_t is Poisson with the parameter tc.

Construction of Poisson processes. The existence of Poisson processes can be obtained by the following explicit construction. Let $\tau_1, \tau_2, ...$ be a sequence of iid exponential r.v. with parameter $c > 0$, i.e. $P(\tau_i > s) = e^{-cs}$, $s > 0$. Introduce the partial sums $S_n = \tau_1 + ... + \tau_n$. These sums have the *Gamma* (c, n) distributions

$$P(S_n \in ds) = \frac{c^n}{(n-1)!} s^{n-1} e^{-cs} ds$$

(**Exer.**: check it by induction taking into account that the distribution of S_n is the convolution of the distributions S_{n-1} and τ_n). Define N_t as the right continuous inverse to S_n, e.g.

$$N_t = \sup\{n \in \mathbf{N} : S_n \le t\},$$

so that $P(S_k \le t) = P(N_t \ge k)$ and

$$P(N_t = n) = P(S_n \le t, S_{n+1} > t) = \int_0^t \frac{c^n}{(n-1)!} s^{n-1} e^{-cs} e^{-c(t-s)}\, ds = e^{-ct} \frac{(ct)^n}{n!}.$$

Exer. 4. Prove that the process N_t constructed above is in fact a Lévy process by showing that

$$P(N_{t+r} - N_t \ge n, N_t = k) = P(N_r \ge n)P(N_t = k) = P(S_n \le r)P(N_t = k). \qquad (3)$$

Hint: take, say, $n > 1$ (the cases with $n = 0$ or 1 are even simpler) and observe that the l.h.s. of (3) is the probability of the event

$$(S_k \le t, S_{k+1} > t, S_{n+k} \le t + r)$$

$$= (S_k = s \le t, \tau_{k+1} = \tau > t - s, S_{n+k} - S_{k+1} = v \le (t + r) - (s + \tau)),$$

so that by independence the l.h.s. of (3) equals

$$\int_0^t \frac{c^k}{(k-1)!} s^{k-1} e^{-cs}\, ds \int_{t-s}^{\infty} ce^{-c\tau}\, d\tau \int_0^{(t+r)-(\tau+s)} \frac{c^{n-1}}{(n-2)!} v^{n-2} e^{-cv}\, dv,$$

which changing τ to $\tau + s$ and denoting it again by τ rewrites as

$$\int_0^t \frac{c^k}{(k-1)!} s^{k-1}\, ds \int_t^{\infty} ce^{-c\tau}\, d\tau \int_0^{t+r-\tau} \frac{c^{n-1}}{(n-2)!} v^{n-2} e^{-cv}\, dv.$$

By calculating the integral over ds and changing the order of v and τ this in turn rewrites as

$$\frac{(ct)^k}{k!} \int_0^r \frac{c^{n-1}}{(n-2)!} v^{n-2} e^{-cv}\, dv \int_t^{t+r-v} ce^{-c\tau}\, d\tau$$

$$= e^{-ct}\frac{(ct)^k}{k!}\int_0^r \frac{c^{n-1}}{(n-2)!}v^{n-2}(e^{-cv}-e^{-cr})\,dv.$$

It remains to see that by integration by parts the integral in this expression equals

$$\int_0^r \frac{c^n}{(n-1)!}s^{n-1}e^{-cs}\,ds,$$

and (3) follows.

Exer. 5 and Def. Let $Z(n)$, $n \in \mathbf{N}$, be a sequence of $\mathbf{R}^d$-valued iid r.v. with law μ_Z. The *compound Poisson process (with the distribution of jumps μ_Z and intensity λ)* is defined as

$$Y(t) = Z(1) + ... + Z(N_t),$$

where N_t is a Poisson process of intensity λ. The corresponding *compensated compound Poisson process* is defined as

$$\tilde{Y}_t = Y(t) - t\lambda \mathbf{E}Z(1).$$

From the above calculations of the ch.f. of a compound Poisson r.v. it follows that $Y(t)$ is a Lévy process with the Lévy exponent

$$\eta(u) = \int (e^{i(u,y)} - 1)\lambda \mu_Z(dy). \tag{4}$$

Check (i) that Y_t is a Lévy process and (ii) that $\mathbf{E}\tilde{Y}_t = 0$. Hint: to check condition (iii) in the definition of Lévy processes write

$$P(|Y_t| > a) = \sum_{n=0}^{\infty} P(|Z(1) + ... + Z(n)| > a)P(N_t = n)$$

and use dominated convergence (alternatively follows from obvious right continuity).

Remark. The existence of a Levy process with a given characteristic exponent can be proved by various constructions. The fastest way is based on carrying out on the level of processes the limiting procedure outlined for r.v. in our proof of Theorem 1 (in other words via Lévy-Ito decomposition, described below). But we shall obtain the existence (of a right continuous modification) later by a more general procedure (applied to all Feller processes) in three steps: (i) building finite-dimensional distributions via Markov property, (ii) using Kolmogorov' existence of a canonical process, (iii) defining right continuous modification via martingale methods.

Def. A Lévy process X_t with a characteristic exponent

$$\eta(u) = i(b,u) - \frac{1}{2}(u, Au) \tag{5}$$

(where A is a positive definite $d \times d$-matrix, $b \in \mathbf{R}^d$) and with a.s. continuous paths is called the d-dimensional *Brownian motion with covariance A and drift b*. It is called *standard* if $A = I$, $b = 0$.

Exer. 6. Show that this is equivalent to say that X_t is a Gaussian process s.t. (i) $B_0 = 0$ a.s.; (ii) the increments $B_t - B_s$ have normal distribution $N((t-s)b, (t-s)A)$ for all $0 \leq s < t$; (iii) the r.v. $B_{t_2} - B_{t_1}$ and $B_{t_4} - B_{t_3}$ are independent whenever $t_1 \leq t_2 \leq t_3 \leq t_4$; (iv) the trajectories $t \mapsto B_t$ are continuous a.s.

Exer. 7. Prove the existence of BM B_t with a given drift and covariance. Hint: first construct a standard d-dimensional BM W_t using product measure spaces, then define $B_t = bt + \sqrt{A}W_t$.

By $\Delta X_t = X_t - X_{t-}$ we shall denote the jumps of X_t.

Theorem 5 (Lévy-Ito decomposition). *Let X_t be a right continuous Lévy process with a characteristic exponent*

$$\eta(u) = i(b,u) - \frac{1}{2}(u, Au) + \int_{\mathbf{R}^d \setminus \{0\}} [e^{i(u,y)} - 1 - i(u,y)\mathbf{1}_{B_1}(y)]\nu(dy). \tag{6}$$

Then X_t can be represented as the sum of three independent Lévy processes $X_t = X_t^1 + X_t^2 + X_t^3$, where X_t^1 is the BM with a drift specified by Lévy exponent (5),

$$X_t^2 = \sum_{s \leq t} \Delta X_t \mathbf{1}_{|\Delta X_t| > 1}$$

is a compound Poisson process with the exponent

$$\eta_2(u) = \int_{\mathbf{R}^d \setminus B_1} [e^{i(u,y)} - 1]\nu(dy) \tag{7}$$

obtained by summing the jumps of X_t of size exceeding 1 and X_t^3 is the limit of the compensated compound Poisson processes $X_t^3(n)$ with the exponents

$$\eta(u) = \int_{B_1 \setminus B_{1/n}} [e^{i(u,y)} - 1]\nu(dy) - i(u, \int_{B_1 \setminus B_{1/n}} y\nu(dy)).$$

The process X_t^3 has jumps only of the size not exceeding 1 and has all finite moments $\mathbf{E}|X_t^3|^m$, $m > 0$.

Proof. Straightforward from (1) and (2). In particular, the product form of ch.f. (1) ensures the independence of X^i, $i = 1, 2, 3$, formula (7) comes by comparison with (4), the moments of X_t^3 are given by

$$\mathbf{E}|X_t^3|^{2k} = \int_{B_1} |y|^{2k}\nu(dy), \quad k = 1, 2, ...$$

Corollary 1. *The only continuous Lévy processes are BM with drifts or deterministic processes (pure drifts).*

Corollary 2. *For any collection of disjoint Borel sets A_i, $i = 1, ..., n$ not containing zero in their closures the processes*

$$X_t^{A_i} = \sum_{s \leq t} \Delta X_t \mathbf{1}_{\Delta X_t \in A_i}$$

are independent compound Poisson process with characteristic exponent

$$\eta_{A_i}(u) = \int_{A_i} (e^{i(u,y)} - 1)\nu_(dy). \tag{8}$$

and $X_t - \sum_{j=1}^n X_t^{A_j}$ *is a Lévy process independent of all* X^{A_j} *with the jumps only outside* $\cup_j A_j$*. Moreover, the processes* $N(t, A_i)$ *that count the number of jumps of* X_t *or* $X_t^{A_i}$ *in* A_i *up to time* t *are independent Poisson processes of intensity* $\nu(A_i)$*.*

Def. Let μ be a σ-finite measure on a metrical space E (we need only the case with E being a Borel subset of $\mathbf{R}^d$). The collection of r.v. $\phi(B)$ parametrized by Borel subsets of E is a *Poisson random measure with intensity* μ if each $\phi(B)$ is a Poisson r.v. with parameter $\mu(B)$ and if $\phi(B_1), ..., \phi(B_n)$ are independent whenever $B_1, ..., B_n$ are disjoint.

Corollary 3. *The collection of r.v.* $N((s,t], A) = N(t, A) - N(s, A)$ *(notations from Corollary 2) counting the number of jumps of* X_t *of size* A *that occur in the time interval* $(s, t]$ *specifies a Poisson random measure on* $(0, \infty) \times (\mathbf{R}^d \setminus \{0\})$ *with intensity* $dt \otimes \nu$*.*

Remark. To prove the existence of a Lévy process in the spirit of Theorem 5 one needs two additional ingredients: existence of a Poisson random measure with an arbitrary intensity (which is rather easy) to construct the processes $N(t, A)$ of jumps of a Lévy process and the proof of convergence of the approximation $X_t^3(n)$ (see Theorem 5), for which one needs Doob's maximal inequality for martingales (which can be derived as a consequence of Doob's optional sampling given Section 8).

Def. The non-decreasing Lévy processes with values in $\mathbf{R}_+$ are called *subordinators.*

Theorem 6. *A real valued Lévy process* X_t *is a subordinator iff its characteristic exponent has the form*

$$\eta(u) = ibu + \int_0^\infty (e^{iuy} - 1)\nu(dy), \tag{9}$$

where $b \geq 0$ *and the Lévy measure* ν *has support in* $\mathbf{R}_+$ *and satisfies the additional condition*

$$\int_0^1 x\nu(dx) < \infty. \tag{10}$$

Moreover

$$X_t = tb + \sum_{s \leq t} (\Delta X_s).$$

Proof. First if X is positive, then it can only increase from $X_0 = 0$. Hence by iid property it is a non-decreasing process and consequently the Lévy measure has support in $\mathbf{R}_+$ and X contains no Brownian part, e.g. $A = 0$ in (6). Next,

$$\sum_{s \leq t} (\Delta X_s) \mathbf{1}_{|X_s| \leq 1} = \sum_{s \leq t} |\Delta X_s| \mathbf{1}_{|X_s| \leq 1} \leq X_t^2,$$

implying that

$$\mathbf{E} \sum_{s \leq t} (\Delta X_s) \mathbf{1}_{|X_s| \leq 1} \leq \mathbf{E} X_t^2 < \infty.$$

But

$$\mathbf{E}\sum_{s\le t}(\Delta X_s)\mathbf{1}_{|X_s|\le 1}=\lim_{\epsilon\to 0}\mathbf{E}\sum_{s\le t}(\Delta X_s)\mathbf{1}_{\epsilon\le|X_s|\le 1}=\int_0^1 x\nu(dx),$$

implying (10).

Def. Clearly for a subordinator X_t the Laplace transform is well defined and

$$\mathbf{E}e^{-\lambda X_t}=\exp\{-t\Phi(\lambda)\},\tag{11}$$

where

$$\Phi(\lambda)=-\eta(i\lambda)=b\lambda+\int_0^\infty(1-e^{-\lambda y})\nu(dy)\tag{12}$$

is called the *Laplace exponent* or *cumulant.*

Def. A subordinator X_t is a *one-sided stable process*, if to each $a\ge 0$ there corresponds a constant $b(a)\ge 0$ s.t. aX_t and $X_{tb(a)}$ have the same law.

Exer. 8 (exponents of stable subordinators). (i) Show that $b(a)$ in this definition is continuous and satisfies the equation $b(ac)=b(a)b(c)$, hence deduce that $b(a)=a^\alpha$ with some $\alpha>0$ called the *index of stability* or *stability exponent.* (ii) Deduce further that $\Phi(a)=b(a)\Phi(1)$ and hence

$$\mathbf{E}e^{-uX_t}=\exp\{-tru^\alpha\}\tag{13}$$

with a constant $r>0$, called the *rate.* Taking into account that Φ from (12) is increasing and concave, deduce that necessarily $\alpha\in(0,1)$. (iii) Show that for $\alpha\in(0,1)$

$$\int_0^\infty(1-e^{-uy})\frac{dy}{y^{1+\alpha}}=\frac{\Gamma(1-\alpha)}{\alpha}u^\alpha\tag{14}$$

by using the integration by parts in order to rewrite the l.h.s. of this equation as

$$\frac{u}{\alpha}\int_0^\infty e^{-uy}y^{-\alpha}\,dy.$$

Deduce that stable subordinators with index α and rate r described by (13) has the Laplace exponent (12) with the Lévy measure

$$\nu(dy)=r\frac{\alpha}{\Gamma(1-\alpha)}y^{-(1+\alpha)}.\tag{15}$$

Exer. 9 Prove the law of large number for a Poisson process N_t of intensity c: $N_t/t\to c$ a.s. as $t\to\infty$. Hint: use the construction of N_t given above and the fact that $S_n/n\to\frac{1}{c}$ as $n\to\infty$ according to the usual law of large numbers.

Section 5. Conditioning.

Def. For a given measure space $(S,\mathcal{F},\mu)$, a measure ν on $(S,\mathcal{F})$ is called *absolutely continuous* with respect to μ if $\nu(A)=0$ whenever $A\in\mathcal{F}$ and $\mu(A)=0$. Two measures are called *equivalent* if they are mutually absolutely continuous.

The Radon-Nikodym Theorem. If μ is σ-finite and ν is finite and absolutely continuous with respect to μ, then there exists a unique (up to almost sure equality) non-negative measurable function g on S such that for all $A \in \mathcal{F}$

$$\nu(A) = \int_A g(x)\mu(dx).$$

This g is called the *Radon-Nikodym derivative* of ν with respect to μ and is often denoted $d\nu/d\mu$.

Def. Conditional expectation. Let X be a integrable r.v. on a probability space $(\Omega, \mathcal{F}, P)$ and let $\mathcal{G}$ be a sub- σ-algebra of $\mathcal{F}$. If $X \geq 0$ everywhere, the formula $Q_X(A) = \mathbf{E}(X\mathbf{1}_A)$ for $A \in \mathcal{G}$ defines a measure Q_X on $(\Omega, \mathcal{G})$ that is obviously absolutely continuous with respect to P. The r.v. $\mathbf{E}(X|\mathcal{G}) = dQ_X/dP$ on $(\Omega, \mathcal{G}, P)$ is called the *conditional expectation of X with respect to $\mathcal{G}$*. If X is not supposed to be positive one defines the *conditional expectation* as $\mathbf{E}(X|\mathcal{G}) = \mathbf{E}(X^+|\mathcal{G}) - \mathbf{E}(X^-|\mathcal{G})$. In other words $Y = \mathbf{E}(X|\mathcal{G})$ is a r.v. on $(\Omega, \mathcal{G}, P)$ such that

$$\int_A Y(\omega)P(d\omega) = \int_A X(\omega)P(d\omega) \tag{1}$$

for all $A \in \mathcal{G}$. If $X = (X_1, ..., X_d) \in \mathbf{R}^d$, then

$$\mathbf{E}(X|\mathcal{G}) = (\mathbf{E}(X_1|\mathcal{G}), ..., \mathbf{E}(X_n|\mathcal{G}).$$

Exer. 1. Let a σ-algebra $\mathcal{G}$ be finite and defined as the set of unions of a finite collection of disjoint sets $G_i \in \mathcal{F}$, $i = 1, ..., n$ such that $\Omega = \cup_{i=1}^n G_i$. Show that for any i the function $\mathbf{E}(X|\mathcal{G})$ is a constant on G_i that equals $\int_{G_i} X(\omega)P(d\omega)$.

Theorem 1 (key properties of the conditional expectation).

(i) $\mathbf{E}(\mathbf{E}(X|\mathcal{G})) = \mathbf{E}(X)$;

(ii) *if Y is $\mathcal{G}$-measurable, then* $\mathbf{E}(XY|\mathcal{G}) = Y\mathbf{E}(X|\mathcal{G})$ *a.s.;*

(iii) if Y is $\mathcal{G}$-measurable and X is independent of $\mathcal{G}$, then $\mathbf{E}(XY|\mathcal{G}) = Y\mathbf{E}(X)$ *a.s. and*

$$\mathbf{E}(f(X,Y)|\mathcal{G}) = G_f(Y) \tag{2}$$

a.s. for a Borel function f, where $G_f(y) = \mathbf{E}(f(X,y))$ *a.s.;*

(iv) if $\mathcal{H}$ is a sub- σ-algebra of $\mathcal{G}$ then $\mathbf{E}(\mathbf{E}(X|\mathcal{G})|\mathcal{H}) = \mathbf{E}(X|\mathcal{H})$ *a.s.;*

(v) the mapping $X \mapsto \mathbf{E}(X|\mathcal{G})$ is on orthogonal projection $L^2(\Omega, \mathcal{F}, P) \mapsto L^2(\Omega, \mathcal{G}, P)$.

(vi) $X_1 \leq X_2 \to \mathbf{E}(X_1|\mathcal{G}) \leq \mathbf{E}(X_2|\mathcal{G})$ *a.s.*

(vii) the mapping $X \mapsto \mathbf{E}(X|\mathcal{G})$ is a linear contraction $L^1(\Omega, \mathcal{F}, P) \mapsto L^1(\Omega, \mathcal{G}, P)$.

Exer. 2. Prove the above theorem. Hint: (ii) consider first the case with Y being an indicator function of a $\mathcal{G}$-measurable set; (v) assume $X = Y + Z$ with Y from $L^2(\Omega, \mathcal{G}, P)$ and Z from its orthogonal complement and show that $Y = \mathbf{E}(X|\mathcal{G})$. (vi) Follows from an obvious remark that $X \geq 0 \Rightarrow \mathbf{E}(X|\mathcal{G}) \geq 0$.

Exer. 3. Give an alternative construction of conditional expectation (proving all its properties) by passing Radon-Nikodym: define it by the property (v) from the above theorem.

Def. If Z is a r.v. on $(\Omega, \mathcal{F}, P)$ one calls $\mathbf{E}(X|\sigma(Z))$ the *conditional expectation of X with respect to Z* and denotes it shortly by $\mathbf{E}(X|Z)$.

Exercise 4 and Def. Show that the r.v. $\mathbf{E}(X|Z)$ is a constant on any Z-level set $\{\omega : Z(\omega) = z\}$. One denotes this constant by $\mathbf{E}(X|Z = z)$ and calls it the *conditional expectation of X given $Z = z$*. Show that

$$\mathbf{E}(X) = \int \mathbf{E}(X|Z)(\omega)P(d\omega) = \int \mathbf{E}(X|Z = z)p_Z(dz). \tag{3}$$

Hint: use (1.1) with the function $f(Z(\omega)) = \mathbf{E}(X|Z)(\omega) = \mathbf{E}(X|Z = z(\omega))$.

Def. Let X and Z be $\mathbf{R}^d$ and respectively $\mathbf{R}^m$-valued r.v. on $(\Omega, \mathcal{F}, P)$, and let $\mathcal{G}$ be a sub-*sigma*-algebra of $\mathcal{F}$. *Conditional probability of X given $\mathcal{G}$* and *X given $Z = z$* respectively are defined as

$$P_{X|\mathcal{G}}(B;\omega) \equiv P(X \subset B|\mathcal{G})(\omega) = \mathbf{E}(\mathbf{1}_B(X)|\mathcal{G})(\omega), \quad \omega \in \Omega;$$

$$P_{X|Z=z}(B) \equiv P(X \subset B|Z = z) = \mathbf{E}(\mathbf{1}_B(X)|Z = z),$$

for Borel sets B, or equivalently through the equations

$$\mathbf{E}(f(X)|\mathcal{G})(\omega) = \int_{\mathbf{R}^d} f(x)P_{X|\mathcal{G}}(dx;\omega) \tag{4}$$

$$\mathbf{E}(f(X)|Z = z) = \int_{\mathbf{R}^d} f(x)P_{X|Z=z}(dx)$$

for bounded Borel functions f. Of course $P_{X|Z=z}(B)$ is just the common value of $P_{X|Z}(B;\omega)$ on the set $\{\omega : Z(\omega) = z\}$.

It is possible to show (though this is not obvious) that *regular conditional probability of X given $\mathcal{G}$* exists, i.e. such a version of conditional probability that $P_{X|\mathcal{G}}(B,\omega)$ is a probability measure on $\mathbf{R}^d$ as a function of B for each ω (notice that from the above discussion the required additivity of conditional expectations hold a.s. only so that they may fail to define a probability even a.s.) and is $\mathcal{G}$-measurable as a function of ω. Hence one can define *conditional r.v.* $X_\mathcal{G}(\omega)$, $X_Z(\omega)$ and $X_{Z=z}$ as r.v. with the corresponding conditional distributions.

Exer. 5. For a Borel function h

$$\mathbf{E}h(X,Z) = \int h(x,z)P_{X|Z=z}(dx)p_Z(dz) \tag{5}$$

(if the l.h.s. is well defined). Hint: From the above definition

$$\int_{A\in\mathcal{G}} \int_{\mathbf{R}^d} f(x)P_{X|\mathcal{G}}(dx;\omega)P(d\omega) = \int_A f(X(\omega))P(d\omega)$$

and in particular

$$\int_{C\in\mathbf{R}^m} \int_{\mathbf{R}^d} f(x)P_{X|Z=z}(dx)P_Z(dz) = \int \mathbf{1}_{Z\subset C}(\omega)f(X(\omega))P(d\omega).$$

Hence

$$\int_{\mathbf{R}^m} \int_{\mathbf{R}^d} g(z) f(x) P_{X|Z=z}(dx) P_Z(dz) = \mathbf{E}(f(X)g(Z))$$

for Borel f, g, which implies (5).

Exer. 6. Deduce from (5) that (i) if X, Z are r.v. with a joint probability density function $f_{X,Z}(x,z)$, then the conditional r.v. $X_{Z=z}$ has a probability density function

$$f_{X_{Z=z}}(x) = f_{X,Z}(x,z)/f_Z(z)$$

whenever f_Z does not vanish, (ii) if X, Z are discrete r.v. with joint probability $P(X = i, Z = j) = p_{ij}$, then the conditional probabilities $p(X = i|Z = j)$ are given by the usual formula $p_{ij}/P(Z = j)$.

Theorem 2. *Let X be a integrable variable on $(\Omega, \mathcal{F}, P)$ and let $\mathcal{G}_n$ be (i) an increasing sequence of sub-σ-algebras of $\mathcal{F}$ with $\mathcal{G}$ being the minimal σ -algebra containing all $\mathcal{G}_n$, or (ii) decreasing sequence of sub-σ-algebras of $\mathcal{F}$ with $\mathcal{G} = \cap_{n=1}^\infty \mathcal{G}_n$. Then a.s. and in L^1*

$$\mathbf{E}(X|\mathcal{G}) = \lim_{n\to\infty} \mathbf{E}(X|\mathcal{G}_n). \tag{6}$$

Furthermore, if $X_n \to X$ a.s. and $|X_n| < Y$ for all n, where Y is an integrable r.v., then a.s. and in L^1

$$\mathbf{E}(X|\mathcal{G}) = \lim_{n\to\infty} \mathbf{E}(X_n|\mathcal{G}_n). \tag{7}$$

Sketch of the proof of the convergence in L^1 (a.s. convergence is a bit more involved, and we shall neither prove, nor use it). Any r.v. of the form χ_B with $B \in \mathcal{G}$ can be approximated in L^2 by a $\mathcal{G}_n$-measurable r.v. ξ_n. Hence the same holds for any r.v. from $L^2(\Omega, \mathcal{F}, P)$. As $E(X|\mathcal{G}_n)$ is the best approximation (L^2-projection) one obtains (6) for $X \in L^2(\Omega, \mathcal{F}, P)$, and hence for $X \in L^1(\Omega, \mathcal{F}, P)$ by density arguments. Next,

$$\mathbf{E}(X_n|\mathcal{G}_n) - \mathbf{E}(X|\mathcal{G}) = \mathbf{E}(X_n - X|\mathcal{G}_n) + (\mathbf{E}(X|\mathcal{G}_n) - \mathbf{E}(X|\mathcal{G}).$$

Since $|X_n| < Y$ and $X_n \to X$ a.s. one concludes that $X_n \to X$ in L^1 by dominated convergence. Hence

$$\mathbf{E}(\mathbf{E}|X_n - X||\mathcal{G}_n) = \mathbf{E}|X_n - X| \to 0.$$

Theorem 3. *If $X \in L^1(\Omega, \mathcal{F}, P)$, the family of r.v. $\mathbf{E}(X|\mathcal{G})$, $\mathcal{G}$ runs through all sub-σ-algebra of $\mathcal{F}$, is uniformly integrable.*

Proof.

$$\mathbf{1}_{|\mathbf{E}(X|\mathcal{G})|>c}\mathbf{E}(X|\mathcal{G}) = \mathbf{E}(X\mathbf{1}_{|\mathbf{E}(X|\mathcal{G})|>c}|\mathcal{G}),$$

because $\{|\mathbf{E}(X|\mathcal{G})| > c\} \in \mathcal{G}$. Hence

$$\mathbf{E}\left(\mathbf{1}_{|\mathbf{E}(X|\mathcal{G})|>c}\mathbf{E}(X|\mathcal{G})\right) \le \mathbf{E}\left(\mathbf{1}_{|\mathbf{E}(X|\mathcal{G})|>c}|X|\right)$$

$$\le \mathbf{E}(|X|\mathbf{1}_{|X|>d}) + dP(|\mathbf{E}(X|\mathcal{G})| > c) \le \mathbf{E}(|X|\mathbf{1}_{|X|>d}) + \frac{d}{c}\mathbf{E}(|X|).$$

First choose d to make the first term small, then c to make the second one small.

Theorem 4 (locality of conditional expectation). *Let the σ-algebras $\mathcal{G}_1, \mathcal{G}_2 \in \mathcal{F}$ and r.v. $X_1, X_2 \in L^1(\Omega, \mathcal{F}, P)$ be such that $\mathcal{G}_1 = \mathcal{G}_2$ and $X_1 = X_2$ on a set $A \in \mathcal{G}_1 \cap \mathcal{G}_1$. Then $\mathbf{E}(X_1|\mathcal{G}_1) = \mathbf{E}(X_2|\mathcal{G}_2)$ a.s. on A.*

Proof. Note that $\mathbf{1}_A \mathbf{E}(X_1|\mathcal{G}_1)$ and $\mathbf{1}_A \mathbf{E}(X_2|\mathcal{G}_2)$ are both $\mathcal{G}_1 \cap \mathcal{G}_2$-measurable, and for any $B \subset A$ s.t. $B \in \mathcal{G}_1$ (and hence $B \in \mathcal{G}_2$)

$$\int_B \mathbf{E}(X_1|\mathcal{G}_1) P(d\omega) = \int_B X_1 P(d\omega) = \int_B X_2 P(d\omega) = \int_B \mathbf{E}(X_2|\mathcal{G}_2) P(d\omega).$$

Section 6. Markov processes.

Def. Let $(\Omega, \mathcal{F})$ be a measurable space. A family $\mathcal{F}_t$, $t \geq 0$, of sub-σ-algebras of $\mathcal{F}$ is called a *filtration* if $\mathcal{F}_s \subset \mathcal{F}_t$ whenever $s \leq t$. By $\mathcal{F}_\infty$ one denotes the minimal σ-algebra containing all $\mathcal{F}_t$. A probability space $(\Omega, \mathcal{F}, P)$ with a filtration is said to be *filtered*. A process $X = X_t$ defined on a filtered probability space $(\Omega, \mathcal{F}, P)$ is *adapted* (or $\mathcal{F}_t$-adapted) if X_t is $\mathcal{F}_t$-measurable for each t. Any process X defines its own *natural filtration* $\mathcal{F}_t^X = \sigma\{X_s : 0 \leq s \leq t\}$ and X is clearly adapted to it.

Main Def. An adapted process $X = X_t$ on a filtered probability space $(\Omega, \mathcal{F}, P)$ is called *Markov process* if for all $f \in B_b(\mathbf{R}^d)$, $0 \leq s \leq t$ it satisfies the following *Markov property*:

$$\mathbf{E}(f(X_t)|\mathcal{F}_s) = \mathbf{E}(f(X_t)|X_s) \quad \text{a.s.} \tag{1}$$

and moreover the function

$$\Phi^{s,t} f(x) = \mathbf{E}(f(X_t)|X_s = x). \tag{2}$$

belongs to $B_b(\mathbf{R}^d)$ whenever f so does for any $0 \leq s \leq t$.

Theorem 1. *Any Lévy process X (e.g. Browninan motion) is Markov with respect to its natural filtration. Moreover*

$$\mathbf{E}(f(X_t)|\mathcal{F}_s^X) = \int_{\mathbf{R}^d} f(X_s + z) p_{t-s}(dz) \tag{3}$$

for $f \in B_b(\mathbf{R}^d)$, $0 \leq s < t$, where p_t is the law of X_t.

Proof. By (5.2)

$$\mathbf{E}(f(X_t)|\mathcal{F}_s^X) = \mathbf{E}(f(X_t - X_s + X_s)|\mathcal{F}_s^X) = G_f(X_s),$$

where

$$G_f(y) = \mathbf{E}(f(X_t - X_s + y)) = \int f(z + y) p_{t-s}(dz),$$

and (3) follows. Similarly the r.h.s. of (1) equals the r.h.s. of (3) implying (1) with the filtration $\mathcal{F}_t^X$.

Def. A Lévy process X_t on a probability space $(\Omega, \mathcal{F}, P)$ equipped with a filtration $\mathcal{F}_t$ is called *$\mathcal{F}_t$-Lévy process* if it is $\mathcal{F}_t$-adapted and the increments $X_t - X_s$ are independent of $\mathcal{F}_s$ for all $0 \leq s < t$.

Theorem 2 (properties of transition).

(i) $\Phi^{s,s} = I$ (identity operator); (ii) (*positivity*) $f \geq 0 \Rightarrow \Phi^{s,t} f \geq 0$; (iii) (*conservativity*) $\Phi^{s,t}(1) = 1$; (iv) (*propagator property*) $\Phi^{r,s}\Phi^{s,t} = \Phi^{r,t}$ for $r \leq s \leq t$.

Proof. (i)-(iii) are obvious and do not depend on the Markov property. (iv) By (6.1)

$$\Phi^{r,t} = \mathbf{E}(f(X_t)|X_r = x) = \mathbf{E}(\mathbf{E}(f(X_t)|\mathcal{F}_s)|X_r = x)$$

$$= \mathbf{E}(\mathbf{E}(f(X_t)|X_s)|X_r = x) = \mathbf{E}(\Phi^{s,t} f(X_s)|X_r = x) = (\Phi^{r,s}(\Phi^{s,t} f))(x).$$

Def. For a Markov process X *transition probabilities* are defined by

$$p_{s,t}(x, A) = (\Phi^{s,t}\mathbf{1}_A)(x) = P(X_t \in A|X_s = x),$$

so that

$$(\Phi^{s,t} f)(x) = \int_{\mathbf{R}^d} f(y) p_{s,t}(x, dy), \quad f \in B_b(\mathbf{R}^d).$$

A Markov process has *transition densities* whenever the measures $p_{s,t}(x, .)$ have densities, say $\rho_{s,t}(x, y)$ so that $p_{s,t}(x, A) = \int_A \rho_{s,t}(x, y)\, dy$.

Exer. 1. (i) Show that for a Lévy process $p_{s,t}(x, A) = q_{t-s}(A - x)$, where q_t is the law of X_t. (ii) Write down the probability density of the Brownian motion.

Theorem 3 (the Chapman-Kolmogorov equations). *If X is a Markov process, then for any Borel A*

$$p_{r,t}(x, A) = \int_{\mathbf{R}^d} p_{s,t}(y, A) p_{r,s}(x, dy).$$

Proof. Apply the operator equation $\Phi^{r,s}\Phi^{s,t} = \Phi^{r,t}$ to the indicator function $\mathbf{1}_A$.

Exer. 2. (Chapman- Kolmogorov for processes with transition densities.) If a Markov process has transition densities, then Chapman-Kolmogorov rewrites as

$$\rho_{r,t}(x, z) = \int_{\mathbf{R}^d} \rho_{r,s}(x, y) \rho_{s,t}(y, z)\, dy.$$

Def. A family of mappings $\{p_{s,t} : 0 \leq s \leq t < \infty\}$ from $\mathbf{R}^d \times \mathcal{B}(\mathbf{R}^d)$ to $[0, 1]$ is said to be a *transition family* (shortly t.f.) if (i) $p_{s,t}(x, A)$ is measurable as a function of x and is a probability measure as a function of A, (ii) the Chapman-Kolmogorov equations hold.

Theorem 4. *A process X is Markov with respect to its natural filtration $\mathcal{F}_t^X$ with t.f. $p_{s,t}$ and initial measure ν $\Leftrightarrow$ for any $0 = t_0 < t_1 < .... < t_k$ and positive Borel f_i*

$$\mathbf{E}\prod_{i=0}^{k} f_i(X_{t_i}) = \int \nu(dx_0) f_0(x_0) \int p_{0,t_1}(x_0, dx_1) f_1(x_1) ... \int p_{t_{k-1},t_k}(x_{k-1}, dx_k) f_k(x_k). \quad (4)$$

Proof. Let X be Markov with t.f. $p_{s,t}$. Then

$$\mathbf{E}\prod_{i=0}^{k} f_i(X_{t_i}) = \mathbf{E}\left(\prod_{i=0}^{k-1} f_i(X_{t_i})\mathbf{E}(f_k(X_{t_k}|\mathcal{F}_{t_{k-1}})\right)$$

$$= \mathbf{E}\left(\prod_{i=0}^{k-1} f_i(X_{t_i})\Phi^{t_{k-1},t_k} f_k(X_{t_{k-1}})\right) = \mathbf{E}\left(\prod_{i=0}^{k-1} f_i(X_{t_i})\int p_{t_{k-1},t_k}(X_{t_{k-1}},dx_k)f_k(x_k)\right)$$

and repeating this inductively one arrives to the r.h.s. of (4). Conversely, as (1), (2) is equivalent to

$$\int_{A\in\mathcal{F}_s} f(X_t(\omega))P(d\omega) = \int_{A\in\mathcal{F}_s} (\Phi^{s,t}f)(X_s(\omega))P(d\omega)$$

(here one uses that $\mathbf{E}(f(X_t)|X_s)(\omega)$ is constant on a level set of X_s and hence it can be written as $(\Phi^{s,t}f)(X_s(\omega)))$, and because $\mathcal{F}_s$ is generated by the sets $\prod_{i=1}^k \mathbf{1}_{X_{t_i}\in A_i}$, $t_1 < ... < t_k \le s$, to prove that X is Markov one has to show that for any $t_1 < .... < t_k \le s < t$ and Borel functions $f_1,...,f_k$, g

$$\mathbf{E}\left(\prod_{i=0}^{k} f_i(X_{t_i})g(X_t)\right) = \mathbf{E}\left(\prod_{i=0}^{k} f_i(X_{t_i})\Phi^{s,t}g(X_t)\right),$$

and this follows by applying (4) to both sides of this equation.

Theorem 5. *Let $\{p_{s,t} : 0 \le s \le t < \infty\}$ be a transition family and μ a probability measure on $\mathbf{R}^d$. Then there exists a probability measure P on the measure space $(\mathbf{R}^d)^{\mathbf{R}^+}$ equipped with its natural filtration $\mathcal{F}_t^0 = \sigma(X_u : u \le t)$ generated by the co-ordinate process X s.t. the co-ordinate process X_t is Markov with initial distribution μ and with t.f. $p_{s,t}$.*

Proof. On cylinder sets define

$$p_{t_0,t_1,...,t_n}(A_0 \times A_1 \times ... \times A_n)$$

$$= \int_{A_0} \mu(dx_0)\int_{A_1} p_{0,t_1}(x_0,dx_1)\int_{A_2} p_{t_1,t_2}(x_1,dx_2)...\int_{A_n} p_{t_{n-1},t_n}(x_{n-1},dx_n).$$

Chapman-Kolmogorov $\Rightarrow$ consistency, which implies (Kolmogorov's theorem) the existence of a process X_t with such finite dimensional distributions. Clearly X_0 has law μ and X_t is adapted to its natural filtration. Theorem 4 ensures that this process is Markov.

Def. A Markov process constructed in the above theorem is called *canonical process* corresponding to t.f. $p_{s,t}$.

Def. A Markov process is called (time) *homogeneous* if $p_{s,t}$ depend on the difference $t - s$ only. One then writes p_{t-s} for $p_{s,t}$ and Φ_{t-s} for $\Phi^{s,t}$.

We shall deal only with homogeneous Markov processes.

Exer. 3. If X is a canonical Markov process and Z is a $\mathcal{F}_\infty^0$-measurable bounded (or positive) function on $(\mathbf{R}^d)^{\mathbf{R}^+}$. Then the map $x \mapsto E_x(Z)$ is (Borel) measurable and

$$\mathbf{E}_\nu(Z) = \int \nu(dx)\mathbf{E}_x(Z)$$

for any probability measure ν (initial distribution of X). Hint: extend by the monotone class theorem from the mappings Z being indicators of cylinders, for which this is equivalent to (3).

⋆ **Theorem 6 (a more powerful formulation of Markov property).** *Coordinate process on* $((\mathbf{R}^d)^{\mathbf{R}^+}, \mathcal{F}^0_\infty, P)$ *is Markov* $\Leftrightarrow$ *for any bounded (or positive) r.v.* Z *on* $(\mathbf{R}^d)^{\mathbf{R}^+}$, *every* $t > 0$ *and starting measure* ν

$$\mathbf{E}_\nu(Z \circ \theta_t | \mathcal{F}^0_t) = \mathbf{E}_{X_t}(Z) \qquad P_\nu - \text{a.s.},$$

where θ *is the canonical shift operator* $X_s(\theta_t(\omega)) = X_{t+s}(\omega)$.

Proof. One needs to show that

$$\mathbf{E}_\nu((Z \circ \theta_t)Y) = \mathbf{E}_\nu(\mathbf{E}_{X_t}(Z)Y)$$

for $\mathcal{F}^0_t$-measurable r.v. Y. By usual extension arguments it is enough to do it for $Y = \prod_{i=1}^k f_i(X_{t_i})$ and $Z = \prod_{j=1}^n g_j(X_{s_j})$, where $t_i \leq t$ and f_i, g_j are positive Borel. Thus one has to show that

$$\mathbf{E}_\nu\left(\prod_{j=1}^n g_j(X_{s_j+t}) \prod_{i=1}^k f_i(X_{t_i})\right) = \mathbf{E}_\nu\left(\mathbf{E}_{X_t}\left(\prod_{j=1}^n g_j(X_{s_j})\right) \prod_{i=1}^k f_i(X_{t_i})\right).$$

But the l.h.s. equals

$$\mathbf{E}_\nu\left(\mathbf{E}\left(\prod_{j=1}^n g_j(X_{s_j+t}) | \mathcal{F}^0_t\right) \prod_{i=1}^k f_i(X_{t_i})\right),$$

which coincides with the r.h.s. by the homogeneous Markov property.

Section 7. Feller processes and semigroups.

Recall: Banach spaces: $L^p(\Omega, \mathcal{F}, P), p \geq 1$, $L^\infty(\Omega, \mathcal{F}, P)$, $B_b(X), C_b(X), C_\infty(X)$, convergence, linear operators and their norms, dense subspaces.

Def. A *semigroup of linear contractions* on a Banach space B is a family Φ_t, $t \geq 0$, of bounded linear operators on B with norm not exceeding one s.t. Φ_0 is the identity operator and $\Phi_t \Phi_s = \Phi_{t+s}$ for all $t, s \geq 0$. Such semigroup on the Banach space $B_b(X)$ (X-subset of $\mathbf{R}^d$) is called a *sub-Markov semigroup*, if it *preserves positivity* (or is *positive*), i.e. if $f \geq 0$ always implies $\Phi_t f \geq 0$, and a *Markov semigroup*, if additionally it preserves constants, i.e. $\Phi_t 1 = 1$.

Theorem 1. *For a Markov process with homogeneous t.f.* p_t *the operators*

$$\Phi_t f(x) = \int p_t(x, dy) f(y) = E_x f(X_t)$$

form a Markov semigroup in $B_b(X)$.

Proof. A direct consequence of definitions and Chapman-Kolmogorov equations.

Def. (i) A semigroup Φ_t of linear contractions on a Banach space B is called strongly continuous, if $\|\Phi_t f - f\| \to 0$ as $t \to 0$ for any $f \in B$. (ii) A strongly continuous semigroup of positive linear contractions on $C_\infty(\mathbf{R}^d)$ is called a *Feller semigroup*. It is called *conservative* if it extends to a semigroup of contractions on $B_b(\mathbf{R}^d)$ preserving constants.

We shall discuss only conservative Feller semigroups.

Def. A (homogeneous) Markov process is called a *Feller process*, if its Markov semigroup reduced to $C_\infty(\mathbf{R}^d)$ is a (conservative) Feller semigroup.

$\star$ **Proposition.** *Any Feller semigroup arises in this way, i.e. it is given by*

$$\Phi_t f(x) = \int p_t(x, dy) f(y)$$

with a certain t.f. p_t.

Sketch of the Proof. Follows more or less directly from the Riesz-Markov theorem.

Exer. 1. (i) Show that if A is a bounded linear operator in a Banach space, then

$$T_t = e^{tA} = \sum_{n=0}^{\infty} \frac{t^n}{n!} A^n$$

defines a strongly continuous semigroup. (ii) Show that the process of BM is Feller. (iii) Show that the semigroup of shifts $T_t f(x) = f(x+t)$ is strongly continuous in $C_\infty(\mathbf{R})$ (and hence is Feller there), as well as in $L^1(\mathbf{R})$ or $L^2(\mathbf{R})$, but is not strongly continuous in $C_b(\mathbf{R})$. Observe also that for analytic functions

$$f(x+t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} (D^n f)(x),$$

which can be formally written as $e^{tD} f(x)$. (iv) Let $\eta(y)$ be a complex-valued continuous function on $\mathbf{R}^d$ s.t. $Re\, \eta \le 0$. Convince yourself that

$$T_t f(y) = e^{t\eta(y)} f(y) \tag{1}$$

is a semigroups of contraction in all our Banach spaces $L^p(\mathbf{R}^d)$, $L^\infty(\mathbf{R}^d)$, $B_b(\mathbf{R}^d), C_b(\mathbf{R}^d)$, $C_\infty(\mathbf{R}^d)$. Show that it is strongly continuous in $L^p(\mathbf{R}^d)$ and $C_\infty(\mathbf{R}^d)$, but not necessarily in other three spaces.

Theorem 2. *Let X_t be a Lévy process with Levy symbol η. Then X_t is a Feller process with semigroup Φ_t s.t.*

$$\Phi_t f(x) = \int f(x+y) p_t(dy), \quad f \in C_b(\mathbf{R}^d), \tag{2}$$

where p_t is the law of X_t.

Sketch of the proof. Formula (2) was established earlier. Notice that any $f \in C_\infty$ is uniformly continuous (check it!). For any such f

$$\Phi_t f(x) - f(x) = \int (f(x+y) - f(x)) p_t(dy)$$

$$= \int_{|y|>K} (f(x+y) - f(x)) p_t(dy) + \int_{|y|\le K} (f(x+y) - f(x)) p_t(dy),$$

and the first (resp. the second) term is small for small t and any K by stochastic continuity of X (resp. for small K and arbitrary t by uniform continuity of f). Hence $\|\Phi_t f - f\| \to 0$ as $t \to 0$. Check that $\Phi_t f \in C_\infty$ for any t (**Exer.**)

$\star$ **Exer. 2.** Recall the inversion formula for the Fourier transform on $S(\mathbf{R}^d)$ and check that the Fourier transform takes the semigroup Φ_t to a multiplication semigroup, i.e.

$$\Phi_t f(x) = F^{-1}(e^{t\eta} F f), \quad f \in S(\mathbf{R}^d).$$

Use this representation in conjunction with Exer. 1 (iv) to give another proof of the Feller property of the semigroup Φ_t. Hint: by inversion

$$\Phi_t f(x) = \mathbf{E}(f(X_t + x) = (2\pi)^{d/2} \mathbf{E} \left(\int_{\mathbf{R}^d} e^{i(u, x+X_t)} F f(u)\, du \right)$$

which yields (justify by Fubini's)

$$\Phi_t f(x) = (2\pi)^{d/2} \int_{\mathbf{R}^d} e^{i(u,x)} \mathbf{E} e^{i(u,X_t)} F f(u)\, du = (2\pi)^{d/2} \int_{\mathbf{R}^d} e^{i(u,x)} e^{t\eta(u)} F f(u)\, du.$$

Feller property follows then from the exercise (iv) above and density arguments.

Def. Let T_t be a strongly continuous semigroup of linear contractions on a Banach space B. The *generator of* T_t is defined as the operator

$$Af = \lim_{t\to 0} \frac{T_t f - f}{t}$$

on the linear subspace $D_A \subset B$ (the *domain* of A), where this limit exists (in the topology of B). The *resolvent* of T_t (or of A) is defined for any $\lambda > 0$ as the operator

$$R_\lambda f = \int_0^\infty e^{-\lambda t} T_t f\, dt.$$

Theorem 3 (basic properties of the generator and the resolvent).
(i) $T_t D_A \subset D_A$ *for each* $t \ge 0$.
(ii) $T_t A f = A T_t f$ *for each* $t \ge 0, f \in D_A$.
(iii) R_λ *is a bounded operator in* B *with* $\|R_\lambda\| \le \lambda^{-1}$ *(for any* $\lambda > 0$*).*
(iv) $\lambda R_\lambda f \to f$ *as* $\lambda \to \infty$.
(v) $R_\lambda f \in D_A$ *for any* f *and* $\lambda > 0$ *and* $(\lambda - A) R_\lambda f = f$, *i.e.* $R_\lambda = (\lambda - A)^{-1}$.

(vi) If $f \in D_A$, then $R_\lambda f A f = A R_\lambda f$.
(vii) D_A is dense in B.
Proof. (i) and (ii) Observe that for $\psi \in D_A$

$$AT_t\psi = \left[\lim_{h\to 0}\frac{1}{h}(T_h - I)\right] T_t\psi = T_t\left[\lim_{h\to 0}\frac{1}{h}(T_h - I)\right]\psi = T_t A\psi.$$

(iii) $\|R_\lambda f\| \le \int_0^\infty e^{-\lambda t}\|f\|\,dt = \lambda^{-1}\|f\|$.
(iv) Follows from the equation

$$\lambda\int_0^\infty e^{-\lambda t}T_t f\,dt = \lambda\int_0^\infty e^{-\lambda t}f\,dt + \lambda\int_0^\epsilon e^{-\lambda t}(T_t f - f)\,dt + \lambda\int_\epsilon^\infty e^{-\lambda t}(T_t f - f)\,dt$$

observing that the first term on the r.h.s. is f, the second (resp. the third) term is small for small ϵ (resp. for any ϵ and large λ).
(v) From definitions

$$AR_\lambda f = \lim_{h\to 0}\frac{1}{h}(T_h - \mathbf{1})R_\lambda f = \frac{1}{h}\int_0^\infty e^{-\lambda t}(T_{t+h}f - T_t f)\,dt$$

$$= \lim_{h\to 0}\left[\frac{e^{\lambda h}-1}{h}\int_0^\infty e^{-\lambda t}T_t f\,dt - \frac{e^{\lambda h}}{h}\int_0^h e^{-\lambda t}T_t f\,dt\right] = \lambda R_\lambda f - f.$$

(vi) Follows from definitions and (ii).
(vii) Follows from (iv) and (v).

Exer. 3. Give another proof of (vii) above (by-passing the resolvent) by showing that $\forall\psi \in B$ the vector $\psi_t = \int_0^t T_u\psi du$ belongs to D_A and $A\psi_t = T_t\psi - \psi$.

Exer. 4. The generator A of the semigroup $T_t f = e^{t\eta}f$ from Exer. 1 (iv) above is given by the multiplication operator $Af = \eta f$ on functions f s.t. $\eta^2 f \in C_\infty(\mathbf{R}^d)$ (or respectively $\eta^2 f \in L^p(\mathbf{R}^d)$) .

Theorem 4. *If X_t is a Lévy process with a characteristic exponent*

$$\eta(u) = i(b,u) - \frac{1}{2}(u,Au) + \int_{\mathbf{R}^d\setminus\{0\}}[e^{i(u,y)} - 1 - i(u,y)\chi_{B_1}(y)]\nu(dy), \tag{3}$$

its generator is given by

$$Lf(x) = \sum_{j=1}^d b_j\frac{\partial f}{\partial x_j} + \frac{1}{2}\sum_{j,k=1}^d A_{jk}\frac{\partial^2 f}{\partial x_j\partial x_k} + \int_{\mathbf{R}^d\setminus\{0\}}[f(x+y) - f(x) - \sum_{j=1}^d y_j\frac{\partial f}{\partial x_j}\chi_{B_1}(y)]\nu(dy). \tag{4}$$

For instance for a Brownian motion with a drift the generator is given by the differential part (first two terms) of (4).

Sketch of the Proof. Let us check (4) on the exponential functions. General case follows then by approximation arguments. For $f(x) = e^{i(u,x)}$

$$\Phi_t f(x) = \int f(x+y)p_t(dy) = e^{i(u,x)}\int e^{i(u,y)}p_t(dy) = e^{i(u,x)}e^{t\eta(u)}.$$

Hence

$$Lf(x) = \frac{d}{dt}|_{t=0}\Phi_t f(x) = \eta(u)e^{i(u,x)},$$

which is given by (4) due to the elementary properties of the exponent.

Remark. Of course $e^{i(u,x)}$ does not belong to $C_\infty(\mathbf{R}^d)$ and some attention should be payed to an appropriate choice of the domain of the generator.

$\star$ **Exer. 5.** Give an alternative proof of Theorem 4 using the representation of Φ_t by the Fourier transform given in a Exer. 2.

Def. An operator A on in $C_b(\mathbf{R}^d)$ defined on a domain D_A (i) is *conditionally positive*, if $Af(x) \geq 0$ for any $f \in D_A$ s.t. $f(x) = 0 = \min_y f(y)$, (ii) satisfies the *positive maximum principle (PMP)*, if $Af(x) \leq 0$ for any $f \in D_A$ s.t. $f(x) = \max_y f(y) \geq 0$, (iii) is *local* if $Af(x) = 0$ whenever $f \in D_A$ vanishes in a neighborhood of x, (iv) satisfies a *local PMP*, if $Af(x) \leq 0$ for any $f \in D_A$ having a local non-negative maximum at x.

Theorem 5. *Let A be a generator of a Feller semigroup Φ_t. Then (i) A is conditionally positive and (ii) satisfies the PMP on D_A. (iii) If moreover A is local and D_A contains C^∞_{comp}, then it satisfies the local PMP on C^∞_{comp}.*

Sketch of the proof. For (i)

$$Af(x) = \lim_{t\to 0}\frac{\Phi_t f(x) - f(x)}{t} = \lim_{t\to 0}\frac{\Phi_t f(x)}{t} \geq 0$$

by positivity preservation. For (ii) apply (i) to the function $f_x(y) = f(x) - f(y)$.

Theorem 6. *If the generator L of a (conservative) Feller semigroup Φ_t with t.f. $p_t(x, dy)$ is local and $C^\infty_{comp} \subset D_L$, then*

$$Lf(x) = \sum_{j=1}^d b_j(x)\frac{\partial f}{\partial x_j} + \frac{1}{2}\sum_{j,k=1}^d a_{jk}(x)\frac{\partial^2 f}{\partial x_j \partial x_k} \tag{5}$$

for certain $b_i, a_{ij} \in C(\mathbf{R}^d)$ s.t. $A = (a_{ij})$ is a positive definite matrix.

Proof. To shorten the formulas, assume $d = 1$. Let χ be a smooth function $\mathbf{R} \mapsto [0, 1]$ that equals 1 (resp. 0) for $|x| \leq 1$ (resp. $|x| > 2$). For an $f \in C^\infty_{comp}$ one can write

$$f(y) = f(x) + f'(x)(y-x)\chi(y-x) + \frac{1}{2}f''(x)(y-x)^2\chi(y-x) + g_x(y),$$

where $g_x(y) = o(1)(y-x)^2$ as $y \to x$. By conservativity $L1 = 0$. Hence

$$Lf(x) = b(x)f'(x) + \frac{1}{2}a(x)f''(x) + (Lg_x)(x)$$

with

$$b(x) = L[(.-x)\chi(.-x)](x) = \lim_{t\to 0}\frac{1}{t}\int (y-x)\chi(y-x)p_t(x, dy),$$

$$a(x) = L[(.-x)^2\chi(.-x)](x) = \lim_{t\to 0}\frac{1}{t}\int (y-x)^2\chi(y-x)p_t(x, dy).$$

But $\pm g_x(y)+\epsilon(y-x)^2$ vanishes at $y=x$ and has a local minimum there for any ϵ so that $Lg(x)=0$, which completes the proof.

Def. Feller process with a generator of type (5) is called a *(Feller) diffusion.*

Exer. 6. Show that the coefficients b_j and a_{ij} can be defined as

$$b_j(x)=\lim_{t\to 0}\frac{1}{t}\int (y-x)_j \mathbf{1}_{\{|y-x|\le\epsilon\}}(y)p_t(x,dy), \tag{6}$$

$$a_{ij}(x)=\lim_{t\to 0}\frac{1}{t}\int (y-x)_i(y-x)_j \mathbf{1}_{\{|y-x|\le\epsilon\}}(y)p_t(x,dy) \tag{7}$$

for any $\epsilon>0$. Conversely, if these limits exist and are independent of ϵ, then the generator is local so that the process is a diffusion.

Exer. 7 (a mathematical version of "Einstein's style" of the analysis of BM). If the generator L of a (conservative) Feller semigroup Φ_t with t.f. $p_t(x,dy)$ is such that $C^\infty_{comp}\subset D_L$ and

$$p_t(x;\{y:|y-x|\ge\epsilon\})=o(t), \quad t\to 0,$$

for any $\epsilon>0$, then L is local (and hence of diffusion type).

Conclusion about BM. A BM (possibly with a drift) can be characterized as (i) a diffusion with iid increments or as (ii) a Lévy process with a local generator.

Exer. 8. (i) Show that the resolvent of the standard BM is given by the formula

$$R_\lambda f(x)=\int_{-\infty}^\infty R^1_\lambda(|x-y|)f(y)\,dy=\frac{1}{\sqrt{2\lambda}}\int_{-\infty}^\infty e^{-\sqrt{2\lambda}|y-x|}f(y)\,dy. \tag{8}$$

Hint: Check this identity for the exponential functions $f(x)=e^{i\theta x}$ using the known ch.f. of the normal r.v. $N(0,t)$. (ii) Show that for the standard BM in $\mathbf{R}^3$

$$R_\lambda f(x)=\int_{\mathbf{R}^3} R^3_\lambda(|x-y|)f(y)\,dy=\int_{\mathbf{R}^3}\frac{1}{2\pi|x-y|}e^{-\sqrt{2\lambda}|y-x|}f(y)\,dy. \tag{9}$$

Hint: observe that

$$R^3_\lambda(|z|)=\int_0^\infty e^{-\lambda t}(2\pi t)^{-3/2}e^{-|z|^2/(2t)}\,dt=-\frac{1}{2\pi|z|}(R^1_\lambda)'(|z|). \tag{10}$$

CHAPTER 3. MARTINGALES METHODS.

Section 8. Martingales.

Def. An adapted integrable process on a filtered probability space is called *submartingale* if, for all $0\le s\le t<\infty$,

$$\mathbf{E}(X_t|\mathcal{F}_s)\ge X_s,$$

a *supermartingale*, if the reverse inequality holds, and a *martingale* if

$$\mathbf{E}(X_t|\mathcal{F}_s)=X_s.$$

Def. A filtration $\mathcal{F}_t$ is said to satisfy the *usual hypotheses*, if (i) (completeness) $\mathcal{F}_0$ contains all sets of P-measure zero (all P-negligible sets), (ii) (right continuity) $\mathcal{F}_t = \mathcal{F}_{t^+} = \cap_{\epsilon>0}\mathcal{F}_{t+\epsilon}$. Adding to all $\mathcal{F}_t$ (of an arbitrary filtration) all P-negligible sets leads to a new filtration called the *augmented filtration*.

Theorem (on regularity of submartingales) (without a proof). *Let M be a submartingale. (i) The following left and right limits exist and are a.s. finite for each $t > 0$:*

$$M_{t_-} = \lim_{s\in\mathbf{Q}, s\to t, s<t} M_s; \qquad M_{t_+} = \lim_{s\in\mathbf{Q}, s\to t, s>t} M_s.$$

$\star$ *(ii) If the filtration satisfies the usual hypotheses and if the map $t \mapsto \mathbf{E}M_t$ is right-continuous, then M has a cadlag (right continuous with finite left limits everywhere) modification.*

Theorem 1. *If X is a Levy process with Lévy symbol η, then $\forall u \in \mathbf{R}^d$, the process*

$$M_u(t) = \exp\{i(u, X_t) - t\eta(u)\}$$

is a complex $\mathcal{F}_t^X$-martingale.

Proof. $\mathbf{E}|M_u(t)| = \exp\{-t\eta(u)\} < \infty$ for each t. Next, for $s \le t$

$$M_u(t) = M_u(s)\exp\{i(u, X_t - X_s) - (t-s)\eta(u)\}.$$

Then

$$\mathbf{E}(M_u(t)|\mathcal{F}_s^X) = M_u(s)\mathbf{E}(\exp\{i(u, X(t-s))\})\exp\{-(t-s)\eta(u)\} = M_u(s).$$

Exer. 1. Show that the following processes are martingales:

(1) standard BM B_t, $B_t^2 - t$, $B_t^3 - 3tB_t$, $B_t^4 - 6tB_t^2 + 3t^2$;

(2) d-dimensional Brownian motion $B(t)$ with covariance A, $|B(t)|^2 - tr(A)t$, and $\exp\{(u, B(t)) - (u, Au)/2\}$ for any u;

(3) compensated Poisson process $\tilde{N}_t = N_t - \lambda t$ with an intensity λ and $\tilde{N}_t^2 - \lambda t$;

(4) *closed martingales*: $\mathbf{E}(Y|\mathcal{F}_t)$, where Y is an arbitrary integrable r.v. in a filtered probability space. Hint: for (4) use Theorem 5.1 (iv).

Theorem 2 (Dynkin's formula). *Let $f \in D$ -domain of a Feller process X_t. Then the process*

$$M_t^f = f(X_t) - f(X_0) - \int_0^t Af(X_s)\,ds, \quad t \ge 0,$$

is a martingale under any initial distribution ν, often called Dynkin's martingale.

Proof.

$$\mathbf{E}(M_{t+h}^f|\mathcal{F}_t) - M_t^f = \mathbf{E}(f(X_{t+h}) - \int_0^{t+h} Af(X_s)\,ds|\mathcal{F}_t) - (f(X_t) - \int_0^t Af(X_s)\,ds)$$

$$= \Phi_h f(X_t) - \mathbf{E}\left(\int_t^{t+h} Af(X_s)\,ds|\mathcal{F}_t\right) - f(X_t) = \Phi_h f(X_t) - f(X_t) - \int_0^h A\Phi_s f(X_t)\,ds = 0.$$

Theorem 3. *A Feller process X_t admits a cadlag modification.*

Remarks. (i) Unlike martingales, we do not need the right continuity of the filtration here. (ii) Proving the result for Lévy processes only one often utilizes the special martingales $M_u(t) = \exp\{i(u, X_t) - t\eta(u)\}$ instead of Dynkin's one used in the proof below.

Proof. Let f_n be a sequence in C_∞ that separates points. By Dynkin's formula and Theorem on martingale, there exists a set Ω of full measure s.t. $f_n(X_t)$ has right and left limits on it for all n along all rational numbers $\mathbf{Q}$. Hence X_t has right and left limits on Ω. Define

$$\tilde{X}_t = \lim_{s \to t, s > t, s \in \mathbf{Q}} X_s.$$

Then $X_s \to \tilde{X}_t$ a.s. and $X_s \to X_t$ weakly (by Feller property). Hence $\tilde{X}_t$ has the same distributions as X_t. Moreover, for any $h, g \in C_\infty$

$$\mathbf{E}_\nu(g(X_t)h(\tilde{X}_t)) = \lim_{s \to t, s > t} \mathbf{E}_\nu(g(X_t)\Phi_{s-t}h(X_t))$$

$$= \lim_{s \to t, s > t} \mathbf{E}_\nu(g(X_t)h(X_s)) = \mathbf{E}_\nu(g(X_t)h(X_t)),$$

where the uniform convergence was used. This implies

$$\mathbf{E}f(X_t, \tilde{X}_t) = \mathbf{E}f(X_t, X_t)$$

for all bounded positive Borel functions f. Choosing f to be the indicator of the set $\{(x, y) : x \neq y\}$ yields $\tilde{X}_t = X_t$ a.s.

Our final result here the following:

Theorem 4. *The augmented filtration $\mathcal{F}_t^\nu$ of the canonical filtration $\mathcal{F}_t^0$ is right continuous.*

Proof. Because $\mathcal{F}_t^\nu$ and $\mathcal{F}_{t+}^\nu$ are P_ν-complete, it is enough to show

$$\mathbf{E}_\nu(Z|\mathcal{F}_t^\nu) = \mathbf{E}_\nu(Z|\mathcal{F}_{t+}^\nu) \quad P_\nu - \text{a.s.}$$

for $\mathcal{F}_\infty^0$-measurable and positive Z. By the monotone class theorem, it is sufficient to show this for $Z = \prod_{i=1}^n f_i(X_{t_i})$ with $f \in C_\infty$ and $t_1 < .... < t_n$. We shall use the observation that

$$\mathbf{E}_\nu(Z|\mathcal{F}_t^\nu) = \mathbf{E}_\nu(Z|\mathcal{F}_t^0) \quad P_\nu - \text{a.s.}$$

For a $t > 0$ choose an integer k: $t_{k-1} \le t < t_k$ so that for $h < t_k - t$

$$\mathbf{E}_\nu(Z|\mathcal{F}_{t+h}^\nu) = \prod_{i=1}^{k-1} f_i(X_{t_i}) g_h(X_{t+h}) \qquad P_\nu - \text{a.s.}$$

where

$$g_h(x) = \int p_{t_k - t - h}(x, dx_k) f_k(x_k) \int p_{t_{k+1} - t_k}(x_k, dx_{k+1}) f_{k+1}(x_{k+1}) ... \int p_{t_n - t_{n-1}}(x_{n-1}, dx_n) f_n(x_n).$$

As $h \to 0$, g_h converges uniformly (Feller!) to

$$g(x) = \int p_{t_k - t}(x, dx_k) f_k(x_k) \int p_{t_{k+1}-t_k}(x_k, dx_{k+1}) f_{k+1}(x_{k+1}) ... \int p_{t_n - t_{n-1}}(x_{n-1}, dx_n) f_n(x_n).$$

Moreover, $X_{t+h} \to X_t$ a.s. (right continuity!) and by Theorem 5.2

$$\mathbf{E}_\nu(Z|\mathcal{F}^\nu_{t+}) = \lim_{h\to 0} \mathbf{E}_\nu(Z|\mathcal{F}^\nu_{t+h}) = \prod_{i=1}^{k-1} f_i(X_{t_i}) g(X_t) = \mathbf{E}_\nu(Z|\mathcal{F}^\nu_t).$$

Remark. The Markov property is preserved by augmentation (we shall not address this issue in detail).

Exer. 2. Show that if a random process X_t is left continuous (e.g. is a Brownian motion), then its natural filtration $\mathcal{F}^X_t$ is left continuous. Hint: $\mathcal{F}^X_t$ is generated by the sets $\Gamma = \{(X_{t_1}, ..., X_{t_n}) \in B\}$, $0 \le t_1 < ... < t_n = t$.

Exer.3. Let X_t be a Markov chain on $\{1, ..., n\}$ with transition probabilities $q_{ij} > 0$, $i \neq j$, which can be defined via the semigroup of stochastic matrices Φ_t with the generator

$$(Af)_i = \sum_{j\neq i} (f_j - f_i) q_{ij}.$$

Let $N_t = N_t(i)$ denote the number of transitions during time t of a process starting at some point i. Show that $N_t - \int_0^t q(X_s)\, ds$ is a martingale, where $q(l) = \sum_{j\neq l} q_{lj}$ denote the *intensity of the jumps*. Hint: to check that $\mathbf{E}N_t = \mathbf{E}\int_0^t q(X_s)\, ds$ show that the function $\mathbf{E}N_t$ is differentiable and

$$\frac{d}{dt}\mathbf{E}(N_t) = \sum_{j=1}^n P(X_t = j) q_j.$$

Exer. 4 (Poisson integrals). Recall first that right continuous functions of bounded variation on $\mathbf{R}_+$ (=differences of increasing functions) are i one-to-one correspondence with signed Radon measures on $\mathbf{R}_+$ according to the formulas $f_t = \mu([0, t])$, $\mu((s, t]) = f_t - f_s$, and the *Stieltjes integral* of a locally bounded Borel function g

$$\int_0^t g_s\, df_s = \int_{(0,t]} g_s\, df_s$$

is defined as the Lebesgue integral of g with respect to the corresponding measure μ. Let N_t be a Poisson process of intensity $c > 0$ with respect to a right continuous filtration $\mathcal{F}_t$. (i) Show that

$$\int_0^t N_s dN_s = \frac{1}{2} N_t(N_t + 1), \quad \int_0^t N_{s-} dN_s = \frac{1}{2} N_t(N_t - 1)$$

(integration in the sense of Stieltjes). (ii) Let H be a left continuous bounded adapted process. Show that the processes

$$M_t = \int_0^t H_s\, dN_s - c\int_0^t H_s\, ds, \tag{1}$$

$$M_t^2 - c\int_0^t H_s^2\, ds \tag{2}$$

are martingales. Hint: For (ii) check this first for simple left continuous processes $H_s = \xi(\omega)\mathbf{1}_{(a,b]}(s)$, where from adapted-ness ξ is $\mathcal{F}_t$-measurable for any $t \in (a, b]$ and hence $\mathcal{F}_a$-measurable by right continuity. Then, say

$$M_t = \xi[(N_{\min(t,b)} - N_a) - c(\min(t, b) - a)], \quad t \geq a,$$

and one conclude that $\mathbf{E}M_t = 0$ by the independence of ξ and $N_{a+u} - N_a$ and the properties of the latter.

Section 9. Stopping times and optional sampling theorem.

Def. Let $(\Omega, \mathcal{F}, P)$ be a probability space equipped with a filtration $\mathcal{F}_t$. A*stopping time* (respectively *optional time*) is a r.v. $T : \Omega \mapsto [0, \infty]$ s.t. $\forall t \geq 0$, $(T \leq t) \in \mathcal{F}_t$ (respectively $(T < t) \in \mathcal{F}_t$).

Prop. 1. (i) T is a stopping time $\Rightarrow$ T is an optional time. (ii) if $\mathcal{F}_t$ is right continuous, the two notions coincide.

Proof. (i) $\{T < t\} = \cup_{n=1}^{\infty}\{T \leq t - 1/n\} \in \mathcal{F}_{t-1/n} \subset \mathcal{F}_t$.

(ii) $\{T \leq t\} = \cap_{n=m}^{\infty}\{T < t + 1/n\} \in \mathcal{F}_{t+1/m}$. Hence $\{T \leq t\} \in \mathcal{F}_{t^+}$.

Def. *Hitting time*: $T_A = \inf\{t \geq 0 : X_t \in A\}$, where X_t is a process and A is a Borel set.

Exer. 1. Show that if X is a $\mathcal{F}_t$-adapted and right continuous and A is (i) open or (ii) closed, then T_A is (i) a optional or (ii) a stopping time respectively. Hint:

$$(i) \quad \{T < t\} = \cup_{s<t, s\in\mathbf{Q}}\{X_s \in A\} \subset \mathcal{F}_t,$$

$$(ii) \quad \{T > t\} = \cap_{s\leq t, s\in\mathbf{Q}}\{X_s \notin A\} \subset \mathcal{F}_t.$$

Prop. 2. *If* T, S *are stopping times, then so are* $\min(T, S)$, $\max(T, S)$ *and* $T + S$.

Proof. (i) $\{\min(T, S) \leq t\} = \{T \leq t\} \cup \{S \leq t\}$. (ii) $\{\max(T, S) \leq t\} = \{T \leq t\} \cap \{S \leq t\}$. At last for (iii)

$$\{T + S > t\} = \{T = 0, S > t\} \cup \{T > t, S = 0\} \cup \{T \geq t, S > 0\} \cup \{0 < T < t, T + S > t\}.$$

The first three events are in $\mathcal{F}_t$ trivially or by Prop. 1. To see that the same holds for the last one, it can be written as

$$\cup_{r\in(0,t)\cap\mathbf{Q}}\{t > T > r, S > t - r\}.$$

Def. If T is a stopping time and X is a adapted process, the *stopped σ-algebra* $\mathcal{F}_T$ (of events determined prior to T) is

$$\mathcal{F}_t = \{A \in \mathcal{F} : A \cap \{T \leq t\} \in \mathcal{F}_t, \forall t \geq 0\}$$

and the *stopped r.v.* X_T is $X_T(\omega) = X_{T(\omega)}$.

Exer. 2. (i) Convince yourself that $\mathcal{F}_T$ is a σ-algebra. (ii) Show that if S, T are stoping times s.t. $S \leq T$ a.s., then $\mathcal{F}_S \subset \mathcal{F}_T$.

Exer. 3. If X is an adapted process and T a stopping time taking finitely many values, then X_T is $\mathcal{F}_T$-measurable. Hint: Say, range of T is $t_1 < ... < t_n$. Then

$$\{X_T \in B\} \cap \{T \leq t_j\} = \cup_{k=1}^{j} \{X_T \in B\} \cap \{T = t_k\} = \cup_{k=1}^{j} \{X_{t_k} \in B\} \cap \{T = t_k\} \in \mathcal{F}_{t_j}.$$

Exer. 4. (i) If T_n is a sequence of $\mathcal{F}_t$ stopping times, then $\sup_n T_n$ is a stopping time. (ii) If $\mathcal{F}_t$ is right continuous, then $\inf_n T_n$ is a stopping time. (iii) If additionally T_n is decreasing and converging to T, then $\mathcal{F}_T = \cap_n \mathcal{F}_{T_n}$. Hint: (i) $\{\sup T_n \leq t\} = \cap\{T_n \leq t\}$. (ii) $\{\inf T_n < t\} = \cup\{T_n < t\} \in \mathcal{F}_t$ and use Prop. 1 (ii).

Def. A process X is *progressively measurable* (or *progressive*) if $\forall t$ the map $(s, \omega) \mapsto X_s(\omega)$ from $[0, t] \times \Omega$ into $\mathbf{R}^d$ is $\mathcal{B}([0, t]) \otimes \mathcal{F}_t$ -measurable.

Prop. 3. *An adapted process with right or left continuous paths is progressive.*

Proof. Say, X_t is right continuous. Define $X_0^{(n)}(\omega) = X_0(\omega)$ and

$$X_s^{(n)}(\omega) = X_{(k+1)t/2^n}(\omega) \quad \text{for} \quad \frac{kt}{2^n} < s \leq \frac{k+1}{2^n} t$$

where $t > 0$, $n > 0$, $k = 0, 1, ..., 2^n - 1$. The map $(s, \omega) \mapsto X_s^{(n)}(\omega)$ is $\mathcal{B}([0, t]) \otimes \mathcal{F}_t$-measurable. Hence the same holds for X_s, since $X_s^{(n)} \to X_s$ by right continuity.

Prop. 4. *If X is progressive and T is a stopping time, then the stopped r.v. X_T is $\mathcal{F}_T$-measurable on $\{T < \infty\}$.*

Proof. The r.v. $(s, \omega) \mapsto X_s(\omega)$ is $\mathcal{B}([0, t]) \otimes \mathcal{F}_t$ measurable, and the mapping $\omega \mapsto (T(\omega), \omega)$ is $\mathcal{F}_t, \mathcal{B}([0, t]) \otimes \mathcal{F}_t$ measurable, and so is its restriction on the set $\{T \leq t\}$. Hence the composition $X_T(\omega)$ of this maps is $\mathcal{F}_t$-measurable on the set $\{T \leq t\}$, which means $\{\omega : X_T(\omega) \in B, T \leq t\} \in \mathcal{F}_t$ for a Borel set B, as required.

Exer. 5. For an optional time T define the sequence (T_n), $n \in \mathbf{N}$ of decreasing random times converging to T as

$$T_n(\omega) = \begin{cases} T(\infty), & \text{if} \quad T(\omega) = \infty \\ k/2^n, & \text{if} \quad (k-1)/2^n \leq T(\omega) < k/2^n \end{cases}$$

Show that all T_n are stopping times converging monotonically to T.

Def. (predictability and martingale transform (discrete stochastic integral)). A process H_n, $n = 1, 2, ...$, is called *predictable* with respect to a discrete filtration $\mathcal{F}_n$, $n = 0, 1, ...$, if H_n is $\mathcal{F}_{n-1}$-measurable for all n. Let (X_n), $n = 0, 1, ...$ be a stochastic process adapted to $\mathcal{F}_n$, and H_n a positive bounded predictable process. The process $H \circ X$ defined inductively by

$$(H \circ X)_0 = X_0, \quad (H \circ X)_n = (H \circ X)_{n-1} + H_n(X_n - X_{n-1})$$

is called the *transform of X by H* and a *martingale transform* if X is a martingale.

Prop. 5. *$(H \circ X)$ is a (sub)martingale whenever X so is.*
Proof. Follows from

$$\mathbf{E}((H \circ X)_n | \mathcal{F}_{n-1}) = (H \circ X)_{n-1} + H_n \mathbf{E}(X_n - X_{n-1} | \mathcal{F}_{n-1}).$$

Exer. 6. Let T, S be bounded stopping times s.t. $S \leq T \leq M$. Let

$$H_n = \mathbf{1}_{n \leq T} - \mathbf{1}_{n \leq S} = \mathbf{1}_{S < n \leq T}.$$

Show that H is predictable and $(H \circ X)_n - X_0 = X_T - X_S$ for $n > M$. Hint:

$$(H \circ X)_n - X_0 = \mathbf{1}_{S<1 \leq T}(X_1 - X_0) + ... + \mathbf{1}_{S<n \leq T}(X_n - X_0).$$

Prop. 6 (discrete optional sampling and martingale characterization). *(i) Let (X_n), $n = 0, 1, ...$ be a $\mathcal{F}_n$-adapted integrable process. The following three statements are equivalent:*
(i) X_t is a submartingale (respectively a martingale),
(ii) for any bounded stopping times $S \leq T$

$$\mathbf{E}(X_S) \leq \mathbf{E}(X_T) \tag{1}$$

(respectively with the equality sign),
(iii) for any bounded stopping times $S \leq T$

$$X_S \leq \mathbf{E}(X_T | \mathcal{F}_s) \quad \text{a.s.} \tag{2}$$

(respectively with equality).
Proof. (iii) $\Rightarrow$ (i) is obvious. (i) $\Rightarrow$ (ii) follows from Exer. 6 and Prop. 5. Finally, to get (ii) $\Rightarrow$ (iii) one applies (1) to the stopping times

$$S^B = S\mathbf{1}_B + M(1 - \mathbf{1}_B), \quad T^B = T\mathbf{1}_B + M(1 - \mathbf{1}_B)$$

with $B \in \mathcal{F}_S$ (check they are stopping times!) yielding

$$\mathbf{E}(X_S \mathbf{1}_B + X_M(1 - \mathbf{1}_B)) \leq \mathbf{E}(X_T \mathbf{1}_B + X_M(1 - \mathbf{1}_B)),$$

which implies $\mathbf{E}(X_S \mathbf{1}_B) \leq \mathbf{E}(X_T \mathbf{1}_B)$ and hence (2).
As an easy application one gets the following fundamental estimate.
$\star$ **Prop. 7.** *(i) If X_n is a submartingale, $n = 1, ..., N$, then*

$$\lambda P(\sup |X_n| \geq \lambda) \leq \mathbf{E}(|X_N| \mathbf{1}_{\sup_n |X_n| \geq \lambda}) \leq \mathbf{E}(|X_N|).$$

(ii) If X_t is a right continuous submartingale on $t \in [0, T]$ or $t \geq 0$, then

$$\lambda P(\sup |X_t| \geq \lambda) \leq \sup_t \mathbf{E}(|X_t|).$$

Proof. As $|X_n|$ is again a submartingale, it is enough to consider the case of positive X. Define a stopping time S being equal to N if $\sup_n X_n < \lambda$ and $S = \inf\{n : X_n \geq \lambda\}$ otherwise. Then

$$\mathbf{E}(X_N) \geq \mathbf{E}(X_S) = \mathbf{E}(X_S \mathbf{1}_{\sup_n |X_n| \geq \lambda}) + \mathbf{E}(X_S \mathbf{1}_{\sup_n |X_n| < \lambda})$$

$$\geq \lambda P(\sup X_n \geq \lambda) + \mathbf{E}(X_N \mathbf{1}_{\sup_n |X_n| < \lambda}),$$

and the required estimate follows by subtraction.

(ii) From a finite index set one directly extends it to a countable index set, and then use right continuity to obtain the general estimate.

Doob's optional stopping (or sampling) theorem. *If X is a right continuous (sub)martingale, $S \leq T$ are two stopping times and either (i) T is bounded or (ii) the family X_τ with τ running through all stopping times is uniformly integrable (the latter occurs e.g. if $X_t = \mathbf{E}(X_\infty|\mathcal{F}_t)$ for some integrable X_∞), then X_S and X_T are integrable with*

$$X_S \leq \mathbf{E}(X_T|\mathcal{F}_S)$$

with equality in case X is a martingale.

Proof. Let $S_n \leq T_n$ be a sequences of decreasing stopping times with countably many values (see Exer. 5) converging to S and T. Then

$$\int_A X_{S_n} dP \leq \int_A X_{T_n} dp \tag{3}$$

for all $A \in \mathcal{F}_{S_n}$ and in particular for $A \in \mathcal{F}_S$. By right continuity X_{T_n} (respectively X_{S_n}) converge to X_T (respectively X_S) point-wise and by uniform integrability also in L^1 (use Theorem 5.3 in case (i))). Hence (3) implies

$$\int_A X_S dP \leq \int_A X_T dp$$

for $A \in \mathcal{F}_S$, i.e. the required result.

Example: "violation of optional sampling". (η_n), $n \in \mathbf{N}$, are iid Bernoulli r.v. s.t. η_n equals 1 (success) or -1 (loss) with probability p and $q = 1 - p$. The player's stake at nth turn is V_n. Naturally V_n is $\mathcal{F}_{n-1} = \sigma(\eta_1, ..., \eta_{n-1})$-measurable. Then the total gain is

$$X_n = \sum_{i=1}^n V_i \eta_i = \sum_{i=1}^n V_i \Delta Y_i = (V \circ Y)_n,$$

where $Y_n = \eta_1 + ... + \eta_n$. The game is *fair* (or *favorable*, or *unfavorable*) if $p = q$ (or $p > q$, or $p < q$) $\Leftrightarrow$ $(X_n, \mathcal{F}_n)$ is a martingale (or submartingale, or supermartingale).

Consider a strategy V (called the *martingale strategy*) s.t. $V_1 = 1$ and further

$$V_n = \begin{cases} 2^{n-1}, & \text{if} \quad \eta_1 = ... = \eta_{n-1} = -1, \\ 0 & \text{otherwise} \end{cases}.$$

Thus if $\eta_1 = ... = \eta_n = -1$, the total loss after n turns will be $\sum_{i=1}^n 2^{i-1} = 2^n - 1$ and if then $\eta_{n+1} = 1$, $X_{n+1} = 2^n - (2^n - 1) = 1$. Denoting $T = \inf\{n : X_n = 1\}$ and assuming $p = q = 1/2$ yields

$$P(T = n) = (1/2)^n, \quad P(T < \infty) = 1(\text{Borel} - \text{Cantelli!}),$$

and consequently

$$\mathbf{E} X_T = P(X_T = 1) = 1 > X_0 = 0,$$

though X_n is a martingale and $EX_n = 0$ for all n.

Section 10. Strong Markov property. Diffusions as Feller processes with continuous paths.

Main Def. A time homogeneous Markov process with t.f. p_t is called *strong Markov* if

$$\mathbf{E}_\nu(f(X_{S+t})|\mathcal{F}_S) = (\Phi_t f)(X_S) \quad P_\nu - \text{a.s.on} \quad \{S < \infty\} \tag{1}$$

for any $\{\mathcal{F}_t\}$-stopping time S, initial distribution ν and positive Borel f.

Exer. 1. If (1) holds for bounded stopping times, then it holds for all stopping times. Hint: For any n and a stopping time S

$$\mathbf{E}_\nu(f(X_{\min(S,n)+t})|\mathcal{F}_{\min(S,n)}) = (\Phi_t f)(X_{\min(S,n)}) \quad P_\nu - \text{a.s.}$$

Hence by locality (Theorem 5.4)

$$\mathbf{E}_\nu(f(X_{S+t})|\mathcal{F}_S) = (\Phi_t f)(X_S) \quad P_\nu - \text{a.s.} \quad \text{on}\,\{S \le n\}.$$

To complete the proof take $n \to \infty$ thus exhausting the set $\{S < \infty\}$.

Exer. 2. A Markov process X_t is strong Markov $\Leftrightarrow$ $\forall$ a.s. finite stopping time T the process $Y_t = X_{T+t}$ is a Markov process with respect to $\mathcal{F}_{T+t}$ with the same t.f. Hint: strong Markov $\Leftrightarrow$

$$\mathbf{E}_\nu(f(X_{T+t+s})|\mathcal{F}_{T+t}) = (\Phi_s f)(X_{T+t}) \quad P_\nu - \text{a.s.}$$

$\star$ **Exer. 3.** A canonical Markov process is strong Markov $\Leftrightarrow$

$$\mathbf{E}_\nu(Z \circ \theta_t|\mathcal{F}_S) = \mathbf{E}_{X_S}(Z) \quad P_\nu - \text{a.s.}$$

for any $\{\mathcal{F}_t\}$-optional time S, initial distribution ν and $\mathcal{F}_\infty$-measurable r.v. Z, where θ is the canonical shift.

Theorem 1. *Any Feller process X_t is strong Markov. Proof.* Let T take values on a countable set D. Then

$$\mathbf{E}_\nu(f(X_{T+t})|\mathcal{F}_T) = \sum_{d\in D} \mathbf{1}_{T=d}\mathbf{E}_\nu f(X_{d+t})|\mathcal{F}_d) = \sum_{d\in D} \mathbf{1}_{T=d}\Phi_t f(X_d) = \Phi_t f(X_T).$$

For a general T take a decreasing sequence of T_n with only finitely many values converging to T. Then

$$\mathbf{E}_\nu(f(X_{T_n+t})|\mathcal{F}_{T_n}) = \Phi_t f(X_{T_n}),$$

for all n, i.e.

$$\int_A f(X_{T_n+t})P(d\omega) = \int_A \Phi_t f(X_{T_n})P(d\omega)$$

for all $A \in \mathcal{F}_{T_n}$, in particular for $A \in \mathcal{F}_T$, as $\mathcal{F}_T \subset \mathcal{F}_{T_n}$. Hence by right continuity of X_t and dominated convergence passing to limit $n \to \infty$ yields

$$\int_A f(X_{T+t})P(d\omega) = \int_A \Phi_t f(X_T)P(d\omega)$$

for all $A \in \mathcal{F}_T$, as required.

Theorem 2. *If X is a Lévy process, then the process $X_T(t) = X_{T+t} - X_T$ is again a Lévy process, which is independent of $\mathcal{F}_t$, and its law under P_ν is the same as that of X under P_0.*

First proof (as a corollary of the strong Markov property of Feller processes). For a positive Borel functions f_i

$$\mathbf{E}_\nu\left(\prod_i f_i(X_{T+t_i} - X_T)|\mathcal{F}_T\right) = \mathbf{E}_{X_T}\left(\prod_i f_i(X_{t_i} - X_0)\right),$$

but this is a constant not depending on X_T.

2nd proof (direct using special martingales $M_u(t) = \exp\{i(u, X_t) - t\eta(u)\}$). Assume T is bounded. Let $A \in \mathcal{F}_T$, $u_i \in \mathbf{R}^d$, $0 = t_0 < t_1 < ... < t_n$. Then

$$\mathbf{E}\left(\mathbf{1}_A \exp\{i\sum_{j=1}^n (u_j, X_T(t_j) - X_T(t_{j-1}))\}\right) = \mathbf{E}\left(\mathbf{1}_A \prod_{j=1}^n \frac{M_{u_j}(T+t_j)}{M_{u_j}(T+t_{j-1})}\right)\prod_{j=1}^n \phi_{t_j - t_{j-1}}(u_j),$$

where $\phi_t(u) = \mathbf{E}e^{i(u,X_t)}$. By conditioning for $s < t$

$$\mathbf{E}\left(\mathbf{1}_A \frac{M_u(T+t)}{M_u(T+s)}\right) = \mathbf{E}\left(\frac{\mathbf{1}_A}{M_u(T+s)}\mathbf{E}(M_u(T+t)|\mathcal{F}_{T+s})\right) = P(A).$$

Repeating this argument yields

$$\mathbf{E}\left(\mathbf{1}_A \exp\{i\sum_{j=1}^n (u_j, X_T(t_j) - X_T(t_{j-1}))\}\right) = P(A)\prod_{j=1}^n \phi_{t_j - t_{j-1}}(u_j),$$

which implies the statement of the Theorem by means of the following fact.

Exer. 4. Suppose X is a r.v. on $(\Omega, \mathcal{F}, P)$, $\mathcal{G}$ is a sub-σ-algebra of $\mathcal{F}$ and

$$\mathbf{E}(e^{i(u,X)}\mathbf{1}_A) = \phi_p(u)P(A)$$

for any $A \in \mathcal{G}$, where ϕ_p is the ch.f. of a probability law p. Then X is independent of $\mathcal{G}$ and the distribution of X is p.

Def. Denote

$$\tau_h = \inf\{t \geq 0 : |X_t - X_0| > h\}, \quad h > 0.$$

A point x is called *absorbing*, if $\tau_h = \infty$ a.s. for every h.

Lemma (intuitively clear, omit a technical proof). *A point is absorbing iff $\Phi f(x) = f(x)$ for any $f \in D_A$. Otherwise $\mathbf{E}_x(\tau_h) < \infty$ for all sufficiently small h.*

Theorem 3 (Dynkin's formula for the generator). *Let X be a Feller process with continuous paths and generator A. For any $f \in D_A$ and a non absorbing x*

$$Af(x) = \lim_{h \to 0} \frac{\mathbf{E}_x f(X_{\tau_h}) - f(x)}{\mathbf{E}_x \tau_h}. \tag{2}$$

For absorbing points x and all f: $Af(x) = 0$.

Proof. By Dynkin's martingale and optional stopping

$$\mathbf{E}_x f(X_{\min(t,\tau_h)}) - f(x) = \mathbf{E}_x \int_0^{\min(t,\tau_h)} Af(X_s)\, ds, \quad t, h > 0.$$

As $\mathbf{E}_x \tau_h < \infty$ for small h, this extends to $t = \infty$ by dominated convergence. This implies (1) by the continuity of Af taking into account that $\mathbf{E}_x(\tau_h) > 0$ by continuity of paths.

The next beautiful result is a direct consequence of (2).

Theorem 4. *Let A be a generator of a Feller process X_t s. t. $C^\infty_{comp} \subset D_A$. If X_t is a.s. continuous P_ν for every ν, then A is local on C^∞_{comp} and hence X_t is a diffusion.*

Remark. The inverse statement holds as well, e.g. the Feller processes with local generators have a.s. continuous paths.

Conclusion about BM. We have got another proof that BM (possibly with a drift) is the only Lévy process with continuous paths.

Section 11. Reflection principle and passage times for BM.

Def. Let B be a Brownian motion on $(\Omega, \mathcal{F}, P)$. The *passage time* T_b to a level b is

$$T_b(\omega) = \inf\{t \geq 0 : B_t(\omega) = b\}.$$

The (intuitively clear) equation

$$P(T_b < t, B_t \geq b) = P(T_b < t, B_t < b)$$

for $b > 0$ is called the *reflection principle.*

Since

$$P(T_b < t) = P(T_b < t, B_t \geq b) + P(T_b < t, B_t < b),$$

and $P(T_b < t, B_t \geq b) = P(B_t \geq b)$ it implies

$$P(T_b < t) = 2P(B_t \geq b) = \sqrt{2/(t\pi)} \int_b^\infty e^{-x^2/2t}\, dx = 2/\pi \int_{bt^{-1/2}}^\infty e^{-x^2/2}\, dx.$$

Differentiating yields the density

$$P(T_b \in dt) = \frac{|b|}{\sqrt{2\pi t^3}} e^{-b^2/2t}\, dt. \tag{1}$$

The necessity to justify the reflection principle and hence these calculations was one of the reason to introduce the strong Markov property.

Theorem 1 (reflection principle). *For a BM B_t*

$$P(T_b \le t) = P(M_t \ge b) = 2P(B_t \ge b) = P(|B_t| \ge b), \tag{2}$$

where $M_t = \inf\{b : T_b \ge t\} = \sup\{B_s : s \le t\}$. In particular, distribution (1) holds.

Proof.

$$P(M_t \ge b, B_t < b) = P(T_b \le t, B_{T_b+(t-T_b)} - B_{T_b} < 0) = P(T_b \le t)P(B_s < 0)$$

$$= \frac{1}{2}P(T_b \le t) = \frac{1}{2}P(M_t \ge b),$$

and the result follows as

$$P(M_t \ge b) = P(B_t \ge b) + P(M_t \ge b, B_t < b).$$

Theorem 2. *Process T_a is a left continuous non-decreasing Levy process (i.e. it is a subordinator), and $T_{a^+} = \inf\{t : B_t > a\}$ is its right continuous modification.*

Proof. Since $T_b - T_a = \inf\{t \ge 0 : B_{T_a+t} - B_{T_a} \ge b - a\}$, this difference is independent of $\mathcal{F}_{T_a}$ by the strong Markov property of BM. Stochastic continuity follows from the density (1). Clearly the process T_a (respectively T_{a^+}) is non-decreasing and left continuous (respectively right continuous) and $T_{a^+} = \lim_{s\to 0, s>0} T_{a+s}$. At last it follows from the continuity of BM that $T_a = T_{a^+}$ a.s.

Theorem 3. *For the process T_a*

$$\mathbf{E}e^{-uT_a} = e^{-a\sqrt{2u}}, \tag{3}$$

which implies by (4.13), (4.15) that T_a is a stable subordinator with the index $\alpha = 1/2$ and Lévy measure $\nu(dx) = (2\pi x^3)^{-1/2}dx$.

First proof. Compute directly from density (1) using the integral calculated in (7.10).

Second proof. As $M_s(t) = \exp\{sB_t - s^2t/2\}$ is a martingale one concludes from optional sampling that

$$1 = \mathbf{E}\exp\{sB_{T_a} - s^2T_a/2\} = e^{sa}\mathbf{E}\exp\{-s^2T_a/2\},$$

and (3) follows by substituting $u = s^2/2$. (Remark. As Doob's theorem is stated for bounded stopping times, in order to be precise here one has to consider first the stopping times $\min(n, T_a)$ and then pass to the limit $n \to \infty$.)

Third proof. For any $a > 0$ the process $\frac{1}{b}T_{a\sqrt{b}}$ is the first hitting time of the level a for the process $b^{-1/2}B_{bt}$. As by the scaling property of BM the latter is again a BM, $\frac{1}{b}T_{a\sqrt{b}}$ and T_a are identically distributed, and thus the subordinator T_a is stable. Comparing expectations one identifies the rate leading again to (3).

Theorem 4. *The joint distribution of M_t and B_t is given by the density*

$$\phi(t,a,b) = P(B_t \in da, M_t \in db) = \frac{2(2b-a)}{\sqrt{2\pi t^3}} \exp\left\{-\frac{(2b-a)^2}{2t}\right\} dadb. \tag{4}$$

Proof. Let $a \leq b$. Then

$$P(B_t < a, M_t \geq b) = P(M_t \geq b, B_{T_b+(t-T_b)} - B_{T_b} < -(b-a))$$

$$= P(M_t \geq b, B_{T_b+(t-T_b)} - B_{T_b} \geq b-a) = P(M_t \geq b, B_t \geq 2b-a) = P(B_t \geq 2b-a),$$

and (4) follows by differentiation.

Theorem 5. *The reflected Brownian motion $|B_t|$ and the process $Y_t = M_t - B_t$ are both Markov with the same probability density*

$$p_t^+(x,y) = p_t(x-y) + p_t(x+y), \tag{5}$$

where $p_t(x-y)$ is the transition density of the standard BM.

Proof. To prove the statement for $|B_t|$ one has to show that

$$P(|B_t + x| \in [a,b]) = P(|B_t - x| \in [a,b]) = \int_a^b p_t^+(x,y)\, dy \tag{6}$$

for all $b > a \geq 0$. This holds, because

$$P(|B_t + x| \in [a,b]) = P(B_t \in [a-x, b-x]) + P(-B_t \in [a+x, b+x])$$

$$= P(B_t \in [a-x, b-x]) + P(B_t \in [a+x, b+x]) = \int_a^b (p_t(y+x) + p_t(y-x))\, dy.$$

Turning to Y_t let $m = M_t > 0$, $b = B_t < m$ and $r = m - b$. Then by the strong Markov:

$$P(M_{t+h} - B_{t+h} < \xi | \mathcal{F}_t) = P(M_{t+h} - B_{t+h} < \xi | B_t = b, M_t = m)$$

$$= \mathbf{E}(\mathbf{1}_{M_{t+h}-B_{t+h}<\xi} \mathbf{1}_{M_{t+h}=m} | B_t = b, M_t = m) + \mathbf{E}(\mathbf{1}_{M_{t+h}-B_{t+h}<\xi} \mathbf{1}_{M_{t+h}>m} | B_t = b, M_t = m)$$

$$= \mathbf{E}(\mathbf{1}_{r-B_h<\xi} \mathbf{1}_{M_h<r}) + \mathbf{E}(\mathbf{1}_{M_h-B_h<\xi} \mathbf{1}_{M_h \geq r}),$$

and this is seen (by inspection) to be the integral of $\phi(t,x,y)$ from (4) over the domain $r - \xi < x < y < x + \xi$, i.e. it equals

$$\int_{r-\xi}^{\infty} dx \int_x^{x+\xi} dy \phi(t,x,y) = -\int_{r-\xi}^{\infty} p_t(2y-x)|_{y=x}^{y=x+\xi},$$

because $\phi(t,x,y) = -\frac{\partial}{\partial y} p_t(2y-x)$. Hence

$$P(M_{t+h} - B_{t+h} < \xi | \mathcal{F}_t) = \int_{r-\xi}^{\infty} p_t(x)\,dx - \int_{r-\xi}^{\infty} p_t(2\xi + x)\,dx$$

$$= \int_{r-\xi}^{\infty} p_t(x)\,dx - \int_{r+\xi}^{\infty} p_t(y)\,dy.$$

Differentiating with respect to ξ yields (5).

Def. The *arcsin law* is the distribution of $\xi = \sin^2 X$ when X is $U(0,2\pi)$ (uniformly distributed on $[0,2\pi]$). Clearly

$$P(\xi \le t\} = P\{|\sin X| \le \sqrt{t}\} = \frac{2}{\pi} \arcsin \sqrt{t}, \quad t \in [0,1]. \tag{7}$$

Theorem 6. *Let B_t be a Brownian motion on $[0,1]$ with the maximum process M_t. Then the random times $\tau = \inf\{t : B_t = M_1\}$ (when B_t first attains its maximum), $\tilde{\tau} = \sup\{t : B_t = M_1\}$ (when B_t for the last time attains its maximum) and the time $\theta = \sup\{t : B_t = 0\}$ of the last exit from the origin obey all the arcsin law. In particular, as $\tau \le \tilde{\tau}$ it implies that $\tau = \tilde{\tau}$ a.s.*

Proof.

$$P(\tilde{\tau} \le t) = P(\tau \le t) = P\left(\sup_{s \le t}(B_s - B_t) \ge \sup_{s \ge t}(B_s - B_t)\right) = P(|B_t| \ge |B_1 - B_t|)$$

$$= P(t\xi^2 \ge (1-t)\eta^2) = P\left(\frac{\eta^2}{\xi^2 + \eta^2} \le t\right) = P(\sin^2 X \le t),$$

where ξ, η are independent $N(0,1)$ r.v. and X is uniformly distributed on $[0,2\pi]$. (**Exer.**: use Theorems 1,4 to explain the reasoning behind all this equivalences!). And

$$P(\theta < t) = P(\sup_{s \ge t} B_s < 0) + P(\inf_{s \ge t} B_s > 0) = 2P(\sup_{s \ge t}(B_s - B_t) < -B_t)$$

$$= 2(|B_1 - B_t| < B_t) = P(|B_1 - B_t| < |B_t|) = P(\tau \le t).$$

Exer. Show (either directly or applying the scaling transformation to (7)) that for $\tau_t = \inf\{s \in (0,t) : B_s = M_t\}$

$$P(\tau_t \le r) = \int_0^r \frac{dy}{\pi\sqrt{y(t-y)}} = \frac{2}{\pi} \arcsin \sqrt{\frac{r}{t}}.$$

CHAPTER 4. HEAT CONDUCTION (OR DIFFUSION) EQUATION.

Section 12. The Dirichlet problem for diffusion operators.

Assume a_{ij}, b_j are continuous bounded function s.t. the matrix (a_{ij}) is positive definite and the operator

$$Lf(x) = \sum_{j=1}^d b_j(x)\frac{\partial f}{\partial x_j} + \frac{1}{2}\sum_{j,k=1}^d a_{jk}(x)\frac{\partial^2 f}{\partial x_j \partial x_k}$$

generates a Feller diffusion X_t. Assume Ω is an open subset of $\mathbf{R}^d$ with the boundary $\partial\Omega$ and closure $\bar{\Omega}$. The *Dirichlet problem for L in Ω* consists in finding an $u \in C_b(\bar{\Omega}) \cap C_b^2(\Omega)$ s.t.

$$Lu(x) = f(x), x \in \Omega, \quad u|_{\partial\Omega} = \psi \tag{1}$$

for given $f \in C_b(\Omega)$, $\psi \in C_b(\partial\Omega)$. A fundamental link between probability and PDE is given by the following

Theorem 1. *Let Ω be bounded and $\mathbf{E}_x\tau_\Omega < \infty$ for all $x \in \Omega$, where $\tau_\Omega = \inf\{t \geq 0 : X_t \in \partial\Omega\}$ (e.g. if X is a BM), and let $u \in C_b(\bar{\Omega}) \cap C^2(\Omega)$ be a solution to (1). Then*

$$u(x) = \mathbf{E}_x\left[\psi(X_{\tau_\Omega}) - \int_0^{\tau_\Omega} f(X_t)\,dt\right]. \tag{2}$$

In particular, such a solution u is unique.

Proof. (i) Assume first that u can be extended to the whole $\mathbf{R}^d$ as a function $u \in C_\infty(\mathbf{R}^d) \cap C_b^2(\mathbf{R}^d)$. Then $u \in D_L$ and applying the stopping time τ_Ω to Dynkin's martingale yields

$$\mathbf{E}\left[u(X_{\tau_\Omega}) - u(x) - \int_0^{\tau_\Omega} Lu(X_t)\,dt\right] = 0, \tag{2a}$$

implying (2). (ii) In general case choose an expanding sequence of domains $\Omega_n \subset \Omega$ with smooth boundaries tending to Ω as $n \to \infty$. The solution u to the problem

$$Lu_n(x) = f(x), x \in \Omega_n, \quad u_n|_{\partial\Omega_n} = u$$

can be extended to $\mathbf{R}^d$ as in (i) and hence is unique and has the representation

$$u(x) = u_n(x) = \mathbf{E}_x\left[u(X_{\tau_{\Omega_n}}) - \int_0^{\tau_{\Omega_n}} f(X_t)\,dt\right], \quad x \in \Omega_n.$$

Taking the limit as $\to \infty$ yields (2), because $\tau_{\Omega_n} \to \tau_\Omega$, as $n \to \infty$.

Example. Take $\Omega = (\alpha, \beta) \subset \mathbf{R}$ and

$$L = \frac{1}{2}a(x)\frac{d^2}{dx^2} + b(x)\frac{d}{dx}$$

with $a, b \in C(\bar{\Omega})$, $a > 0$. Then $u(x) = P_x(X_{\tau_\Omega} = \beta)$ is the probability that X_t starting at a point $x \in (\alpha, \beta)$ reaches β before α and represents a solution to the problem

$$\frac{1}{2}a(x)\frac{d^2u(x)}{dx^2} + b(x)\frac{du(x)}{dx} = 0, x \in (\alpha, \beta), \quad u(\alpha) = 0, u(\beta) = 1. \tag{3}$$

On the other hand, $u(x) = \mathbf{E}_x \tau_\Omega$ is the mean exit time from Ω that solves the problem

$$\frac{1}{2} a(x) \frac{d^2 u(x)}{dx^2} + b(x) \frac{du(x)}{dx} = -1, x \in (\alpha, \beta), \quad u(\alpha) = u(\beta) = 0. \tag{4}$$

Exer.1. (i) Solve problem (3) analytically showing that

$$P_x(X_{\tau_\Omega} = \beta) = \int_\alpha^x \exp\{g(y)\}\, dy \left(\int_\alpha^\beta \exp\{g(y)\}\, dy \right)^{-1}, \tag{5}$$

where $g(x) = -\int_\alpha^x (2b/a)(y)\, dy$. In particular, for a standard BM B_t starting at x this gives $P_x(B_{\tau_\Omega} = \beta) = (x - \alpha)/(\beta - \alpha)$. (ii) Solve (4) with $b = 0$ showing that in this case

$$\mathbf{E}_x \tau_\Omega = 2 \frac{x - \alpha}{\beta - \alpha} \int_\alpha^\beta \frac{\beta - y}{a(y)}\, dy - 2 \int_\alpha^x \frac{x - y}{a(y)}\, dy. \tag{6}$$

In particular, for a BM this turns to $(x - \alpha)(\beta - x)$. Hint for (ii): show first that the solution to the Cauchy problem

$$\frac{1}{2} a(x) u''(x) = -1, \quad u(\alpha) = 0$$

is given by formula

$$u(x) = \omega(x - \alpha) - 2 \int_\alpha^x (x - y) a^{-1}(y)\, dy$$

with a constant ω.

Exer. 2. Check that $\Delta \phi = h''(|x|) + \frac{d-1}{|x|} h'(|x|)$ for $\phi(x) = h(|x|)$. Deduce that if such ϕ is harmonic (i.e. satisfies the Laplace equation $\Delta \phi = 0$) in $\mathbf{R}^d$, then

$$h(r) = \begin{cases} A + Br^{-(d-2)}, & d > 2 \\ A + B \ln r, & d = 2 \end{cases} \tag{7}$$

with some constants A, B.

Exer. 3. Solve the equation $\Delta \phi = 0$ in the shell $S_{r,R} = \{x \in \mathbf{R}^d : r < |x| < R\}$ with boundary conditions $\phi(x)$ being 1 (respectively zero) on $|x| = R$ (resp. $|x| = r$). Hence compute the probability that the standard Browninan motion started from a point $x \in S_{r,R}$ leaves the shell via the outer part of the boundary. Hint: choosing appropriate A, B from (7) one finds

$$\phi(x) = \begin{cases} \frac{|x|^{2-d} - r^{2-d}}{R^{2-d} - r^{2-d}}, & d > 2 \\ \frac{\ln |x| - \ln r}{\ln R - \ln r}, & d = 2 \end{cases}. \tag{8}$$

This describes the required probability due to Theorem 1.

Exer. 4. Calculate the probability of the Brownian motion W_t ever hitting the ball B_r if started at a distance $a > r$ from the origin. Hint: Let T_R (resp. T_r) be the first time $\|W_t\| = R$ (resp. r). By letting $R \to \infty$ in (8)

$$P_x(T_r < \infty) = \lim_{R \to \infty} P_x(T_r < T_R) = \begin{cases} (r/a)^{d-2}, & d > 2 \\ 1, & d = 2 \end{cases} \tag{9}$$

Exer. 5. Use Borel-Cantelli and Exer. 4 to deduce that for $d > 2$ and any starting point $x \neq 0$ there exists a.s. a positive $r > 0$ s.t. W_t^x starting at x never hits the ball B_r. Hint: For any $r < a$ let A_n be the event that W_t^x ever hits the ball $B_{r/2^n}$. Then $\sum P(A_n) < \infty$.

Exer. 6. Show that BM in dimension $d > 2$ is *transient*, i.e. that a.s. $\lim_{t\to\infty} \|W_t\| = \infty$. Hint: As W_t is a.s. unbounded (why?), the event that W_t does not tend to infinity means that there exists a ball B_r s.t. infinitely many events A_n occur, where A_n means that the trajectory returns to B_r after being outside $B_{2^n r}$. This leads to a contradiction by Borel-Cantelli and (9).

$\star$ **Theorem 2.** *Let L be a generator of a Feller diffusion X_t. Given a domain $\Omega \subset \mathbf{R}^d$ assume that there exists a two times continuously differentiable function $f \geq 0$ in $\mathbf{R}^d \setminus \Omega$ s.t. $Lf(x) \leq 0$ and for some $a > 0$ and a point $x_0 \in \mathbf{R}^d \setminus \Omega$ one has*

$$f(x_0) < a < \inf\{f(x) : x \in \partial\Omega\}.$$

Then X_t started at x_0 will never hit Ω with a positive probability (this actually means that the diffusion X_t is transient).

Proof. Let $N > \|x_0\|$, and let τ_Ω and τ_N denote the hitting times of Ω and the sphere $\|y\| = N$ respectively. Put $T_N = \min(\tau_N, \tau_\Omega)$. From (2a) it follows that

$$\mathbf{E}_{x_0} f(X_{T_n}) \leq f(x_0) < a.$$

Hence

$$a > \inf\{f(x) : x \in \partial\Omega\} P_{x_0}(\tau_\Omega < \tau_N) > a P_{x_0}(\tau_\Omega < \tau_N).$$

passing to the limit as $n \to \infty$ yields

$$a > a P_{x_0}(\tau_\Omega < \infty)$$

implying $P_{x_0}(\tau_\Omega < \infty) < 1$.

Section 13. The stationary Feynman-Kac formula.

Recall that the equation

$$\lambda g = Ag + f, \tag{1}$$

where A is the generator of a Feller semigroup Φ_t, $f \in C_\infty(\mathbf{R}^d)$, $\lambda > 0$, is solved uniquely by the formula

$$g(x) = R_\lambda f(x) = \mathbf{E}_x \int_0^\infty e^{-\lambda t} f(X_t)\, dt.$$

This suggests a guess that a solution to the more general equation

$$(\lambda + k)g = Ag + f, \tag{2}$$

where the additional letter k denotes a bounded continuous function could look like

$$g(x) = \mathbf{E}_x \int_0^\infty \exp\{-\lambda t - \int_0^t k(X_s)\, ds\} f(X_t)\, dt. \tag{3}$$

This is the *stationary Feynman-Kac formula* that we are going to discuss now. The fastest way of proving it (at least for diffusions) is by means of Ito's stochastic calculus. Not having this tool at our disposal, we shall use a different method by first rewriting it in terms of the resolvents (thus rewriting the differential equation (2) in an integral form).

Theorem 1. *Let X_t be a Feller process with the semigroup Φ_t and the generator A. Suppose $f \in C_\infty(\mathbf{R}^d)$, k is a continuous bounded non-negative function and $\lambda > 0$. Then $g \in D_A$ and satisfies (2) iff $g \in C_\infty(\mathbf{R}^d)$ and*

$$R_\lambda(kg) = R_\lambda f - g. \tag{4}$$

Proof. Applying R_λ to both sides of (2) and using $R_\lambda(\lambda - A)g = g$ yields (4). Conversely, subtracting the resolvent equations for f and kg

$$AR_\lambda f = \lambda R_\lambda f - f, \quad AR_\lambda(kg) = \lambda R_\lambda(kg) - kg, \tag{5}$$

and using (4) yields (2).

Theorem 2. *Under the assumptions of Theorem 1 the function (3) yields a solution to (4) and hence to (2).*

Proof. Using the Markov property one writes

$$R_\lambda(kg) = \mathbf{E}_x \int_0^\infty e^{-\lambda s} k(X_s) g(X_s)\, ds$$

$$= \mathbf{E}_x \int_0^\infty e^{-\lambda s} k(X_s) \int_0^\infty \exp\{-\lambda t - \int_0^t k(X_{u+s})\, du\} f(X_{t+s})\, dt\, ds.$$

Changing the variables of integration t, u to $\tilde{t} = s + t$ and $\tilde{u} = s + u$ and denoting them again by t and u respectively leads to

$$R_\lambda(kg) = \mathbf{E}_x \int_0^\infty e^{-\lambda t} f(X_t) \int_0^t k(X_s) \exp\{-\int_s^t k(X_u)\, du\}\, ds\, dt,$$

which by the integration by parts rewrites as

$$\mathbf{E}_x \int_0^\infty e^{-\lambda t} f(X_t) \left[1 - \exp\{-\int_0^t k(X_s)\, ds\}\right] dt = (R_\lambda f - g)(x),$$

as required.

In many interesting particular situations the validity of formula (3) can be extended beyond the general conditions of Theorem 2. Let us consider one of this extensions for a one-dimensional BM.

Theorem 3. *Assume $k \geq 0$ and f are piecewise-continuous bounded functions on $\mathbf{R}$ with the finite sets of discontinuity being $Disc_k$ and $Disc_f$. Then the (clearly bounded) function g given by (3) with X_t being a BM B_t is continuously differentiable, has a piecewise continuous second derivative and satisfies*

$$(\lambda + k)g = \frac{1}{2}g'' + f \quad \text{outside} \quad Disc_k \cup Disc_f. \tag{6}$$

Proof. The calculations in the proof of Theorem 2 remains valid for all bounded measurable f and k showing that g satisfies (4). Moreover, for piecewise continuous f and k one sees from dominated convergence that this g is continuous. Next, from Exercise 7.7 one finds that

$$R_\lambda f(x) = \frac{1}{\sqrt{2\lambda}} \left[\int_{-\infty}^{x} e^{\sqrt{2\lambda}(y-x)} f(y)\, dy + \int_{x}^{\infty} e^{\sqrt{2\lambda}(x-y)} f(y)\, dy \right].$$

Hence $R_\lambda f$ is continuously differentiable for any bounded measurable f with

$$(R_\lambda f)'(x) = \int_{x}^{\infty} e^{\sqrt{2\lambda}(x-y)} f(y)\, dy - \int_{-\infty}^{x} e^{\sqrt{2\lambda}(y-x)} f(y)\, dy.$$

This implies in turn that $(R_\lambda f)''$ is piece-wise continuous for a piecewise continuous f and the resolvent equations (5) hold outside $Disc_f \cup Disc_k$. Hence one shows as in Theorem 2 that g satisfies (6), which by integration implies the continuity of g'.

Exer. 2. Show that for $\alpha, \beta > 0$ and a BM B_t

$$\mathbf{E}_x \int_0^\infty \exp\left\{ -\alpha t - \beta \int_0^t \mathbf{1}_{(0,\infty)}(B_s)\, ds \right\} dt = \frac{1}{\alpha+\beta} \left[1 + \frac{\sqrt{\alpha+\beta} - \sqrt{\alpha}}{\sqrt{\alpha}} e^{-\sqrt{2(\alpha+\beta)}x} \right] \tag{7}$$

for $x \geq 0$. Hint: by Theorem 3 the function $z(x)$ on the l.h.s. of (7) is a bounded solution to the equation

$$\begin{cases} \alpha z(x) = \frac{1}{2} z''(x) - \beta z(x) + 1, & x > 0 \\ \alpha z(x) = \frac{1}{2} z''(x) + 1, & x < 0 \end{cases} \tag{8}$$

with the boundary conditions

$$z(0_+) = z(0_-), \quad z'(0_+) = z'(0_-).$$

The bounded solution to (8) have the form

$$z(x) = \begin{cases} A \exp\{-\sqrt{2(\alpha+\beta)}x\} + \frac{1}{\alpha+\beta}, & x > 0 \\ B \exp\{\sqrt{2\alpha}x\} + \frac{1}{\alpha}, & x < 0 \end{cases}.$$

Theorem 4 (arcsin law for the occupation time). *The law for the occupation time* $O_t = \int_0^t \mathbf{1}_{(0,\infty)}(B_s)\, ds$ *of* $(0,\infty)$ *by a standard BM* B_t *has the density*

$$P(O_t \in dy) = \frac{dy}{\pi\sqrt{y(t-y)}}. \tag{9}$$

Proof. By the uniqueness of the Laplace transform it is enough to show that

$$\mathbf{E} e^{-\beta O_t} = \int_0^t e^{-\beta y} \frac{dy}{\pi\sqrt{y(t-y)}}. \tag{10}$$

But from (7)

$$\int_0^\infty e^{-\alpha t}\mathbf{E}e^{-\beta O_t}\,dt = z(0) = \frac{1}{\sqrt{\alpha(\alpha+\beta)}},$$

and on the other hand

$$\int_0^\infty e^{-\alpha t}\int_0^t e^{-\beta y}\frac{dy}{\pi\sqrt{y(t-y)}}dydt = \frac{1}{\pi}\int_0^\infty \frac{e^{-(\alpha+\beta)y}}{\sqrt{y}}dy\int_0^\infty \frac{e^{-\alpha s}}{\sqrt{s}}ds = \frac{1}{\sqrt{\alpha(\alpha+\beta)}},$$

which implies (10) again by the uniqueness of the Laplace transform.

Exer. 3. From formula (9) yielding the solution to eq. $(\lambda-\Delta)g = f$, $\lambda > 0$, in $\mathbf{R}^3$ deduce that the solution to the Poisson equation $\Delta g = -f$ in $\mathbf{R}^3$ is given by formula

$$g = \frac{1}{2\pi}\int \frac{f(y)}{|x-y|}dy$$

whenever f decreases quickly enough at infinity.

Section 14. Diffusions with variable drift, Ornstein-Uhlenbeck processes.

In order to be able to solve probabilistically equations involving second order differential operators, one has to know that these operators generate Markov (Feller) semigroups. Here we show how BM can be used to construct processes with the generators of the form

$$Lf(x) = \frac{1}{2}\Delta f(x) + (b(x), \frac{\partial f}{\partial x}), \quad x \in \mathbf{R}^d. \tag{1}$$

Let b be a bounded Lipshitz continuous function, i.e. $|b(x)-b(y)| \le C|x-y|$ with a constant C. Let B_t be a $\mathcal{F}_t$ BM on a filtered probability space. Then the equation

$$X_t = x + \int_0^t b(X_s)\,ds + B_t$$

has a unique global continuous solution $X_t(x)$ for any x depending continuously on x (proof by fixed point arguments literally the same as for usual ODE). Clearly $X_t(x)$ is a $\mathcal{F}_t$-Markov process starting at x.

Theorem 1. *X_t is a Feller process with the generator (1).*

Proof. Clearly $\Phi_t f(x) = \mathbf{E}f(X_t(x))$ is a semigroup of positive contractions on $C_b(\mathbf{R}^d)$. Let $f \in C^\infty_{comp}(\mathbf{R}^d)$. Then

$$\Phi_t f(x) - f(x) = \mathbf{E}\frac{\partial f}{\partial x}(x)(B_t + \int_0^t b(X_s)\,ds)$$

$$+\frac{1}{2}\mathbf{E}\left(\frac{\partial^2 f}{\partial x^2}(x)(B_t + \int_0^t b(X_s)\,ds), B_t + \int_0^t b(X_s)\,ds\right) + ...,$$

where dots denote the correcting term of McLaurent (or Taylor) series. Taking into account that $\mathbf{E}|B_t^k| = O(t^{k/2})$ it follows that the r.h.s. of this expression is

$$(\frac{\partial f}{\partial x}(x), \mathbf{E}(B_t + tb(x))) + \frac{1}{2}\mathbf{E}\left(\frac{\partial^2 f}{\partial x^2}(x)B_t, B_t\right) + o(t), \quad t \to 0,$$

so that

$$\frac{1}{t}(\Phi_t f(x) - f(x)) \to Lf(x), \quad t \to 0.$$

Hence any $f \in C_{comp}^\infty(\mathbf{R}^d)$ belongs to the domain of the generator L and Lf is given by formula (1). As clearly $\Phi_t f \to f$ for any such f, $t \to 0$, it follows that the same holds for all $f \in C_\infty$ by density arguments.

Exer 1. Convince yourself that the assumption that b is bounded can be dispensed with (only Lipshitz continuity is essential).

Example 1. Solution to the *Langevin* equation

$$v_t = v - b\int_0^t v_s\, ds + B_t$$

with a given constant $b > 0$ defines a Feller process called *Ornstein-Uhlenbeck (velocity) process* with the generator

$$LF(v) = \frac{1}{2}\Delta f(v) - b(v, \frac{\partial f}{\partial v}), \quad v \in \mathbf{R}^d. \tag{2}$$

The pair $(v_t, x_t = x_0 + \int v_s\, ds)$ describes the evolution of a (Newton) particle subject to white noise driving force and friction and is also called sometimes the *Ornstein-Uhlenbeck process*.

Example 2. Solution to the system

$$\begin{cases} \dot{x}_t = y_t \\ y_t = -\int_0^t \frac{\partial V}{\partial x}(x_s)\, ds - b\int_0^t y_s\, ds + B_t \end{cases} \tag{3}$$

describes the evolution of a Newton particle in the potential field V subject to friction and white noise driving force.

Exer. 2. Assume $b = 0$ and that the potential V is bounded below, say $V \geq 1$ everywhere, and is increasing to ∞ as $|x| \to \infty$.

(i) Write down the generator L of the pair process (x_t, y_t). Answer:

$$Lf(x, y) = (y, \frac{\partial f}{\partial x}) - (\frac{\partial V}{\partial x}, \frac{\partial f}{\partial y}) + \frac{1}{2}\Delta f.$$

(ii) Check that $L(H^{-\alpha}) \leq 0$ for $0 < \alpha < (d/2) - 1$, where $H(x, y) = V(x) + y^2/2$ is the energy function (Hamiltonian). (iii) Applying Dynkin's formula with $f = H^{-\alpha}$ for the process starting at (x, y) with the stopping time

$$\tau_h = \inf\{t \geq 0 : H(x_t, y_t) = h\}$$

with $h < H(x,y)$, show that $\mathbf{E}_{x,y} f((x,y)(\tau_h)) < f(x,y)$ and consequently

$$P_{x,y}(\tau_h < \infty) \leq (h/H(x,y))^\alpha.$$

(iv) Follow the same reasoning as in Exer. 12.6 to establish that the process (x_t, y_t) is transient in dimension $d \geq 3$ (this result is remarkable, as it holds for all (smooth) V.

Open problem. Under which condition on V the process specified by (3) with $b = 0$ is transient in dimension $d = 1, 2$ (for $d \geq 3$ the answer is fully settled by Exer. 2; in $d = 1$ only a necessary (but not a sufficient) condition for transience is known).

CHAPTER 5. FINE PROPERTIES of BM.

Section 15. Zeros, excursions and local times.

From Theorem 11.5 one has for a BM W_t that

$$|W_t| = M_t - B_t, \tag{1}$$

where B_t is another BM and M_t is its maximum. Hence the times where W_t is away from the origin coincide with times where $M_t \neq B_t$ and remains constant so that one can interpret M_t as a measure of time W_t spends at the origin. This motivates the Lévy definition of the process measuring the *local times $L_t(0)$ of a BM W_t spend at the origin* by the equation $2L_t(0) = M_t$ (some authors do not include the multiplier 2 in this equation).

As for each τ the time $s \leq \tau$ for which $B_s = M_\tau$ is a.s. unique (see Theorem 11.6), one can choose a set of full measure ω_0 s.t. for all $\omega \in \Omega_0$ this holds for all rational τ. For any such ω and a $t > 0$ define

$$\gamma_t(\omega) = \sup\{s \in [0,t] : W_s = 0\} = \sup\{s \in [0,t] : B_s = M_t\},$$

$$\beta_t(\omega) = \inf\{s \in [t,\infty) : W_s = 0\} = \inf\{s \in [t,\infty) : B_s = M_t\}$$

so that

$$\gamma_t(\omega) < t < \beta_t(\omega) \tag{2}$$

whenever $W_t \neq 0$. Assumption $\omega \in \Omega_0$ implies that the maximum of B_s on $[0,t]$ is attained uniquely at $s = \gamma_t(\omega)$ so that

$$T_{M_t(\omega)}(\omega) = \gamma_t(\omega), \quad T_{M_t(\omega)+}(\omega) = \beta_t(\omega)$$

and thus

$$T_{M_t(\omega)+}(\omega) - T_{M_t(\omega)}(\omega) = \beta_t(\omega) - \gamma_t(\omega)$$

– the size of the jump in $T_b(\omega)$ at $b = M_t(\omega)$ equals the length of the *excursion interval* $(\gamma_t(\omega), \beta_t(\omega))$ *straddling* t.

Let $N(b, [\delta, \epsilon))$ denote the number of jumps of size $l \in [\delta, \epsilon)$ of the Levy subordinator T_a (or its right continuous modification T_{a+}) that occur on the time interval $(0, b]$ (cf. notations in Corollary 3 to Theorem 4.5), and let $N^\delta(b) = N(b, [\delta, \infty))$. According to

Corollary 3 to Theorem 4.5 and Theorem 11.3 the process $b \to N(b, [\delta, \epsilon))$ is a Poisson process with the intensity

$$\nu([\delta, \epsilon)) = \int_\delta^\epsilon (2\pi x^3)^{-1/2} dx = \sqrt{\frac{2}{\pi}}(\delta^{-1/2} - \epsilon^{-1/2}). \tag{3}$$

Theorem 1. *A.s.*

$$\lim_{\delta\to 0} \sqrt{\frac{\pi\delta}{2}} N^\delta(b) = b \quad \forall b \in [0, \infty). \tag{4}$$

Proof. According to (3) the r.v. $Q_t = N^{1/t^2}(b)$ is Poisson with parameter $\sqrt{2/\pi}bt$. Hence, as the process Q_t has non-decreasing right continuous paths and independent increments, it is a Poisson process and by the law of large numbers (see Exer. 4.9) $Q_t/t \to b\sqrt{2/\pi}$ a.s. as $t \to \infty$ implying (4) for a given b. Then one deduce that it holds a.s. for all rational b, and then by continuity one extends it to all b.

Theorem 2 (Lévy, 1948).

$$L_t(0) = \lim_{\delta\to 0} \sqrt{\frac{\pi\delta}{8}} n_t(\delta), \tag{5}$$

where n_t is the number of excursion intervals away from the origin, of duration $\geq \delta$, completed by $W_s, s \leq t$.

Proof. It follows from (4) and the above definition of Lévy's local times that

$$L_t(0) = \lim_{\delta\to 0} \sqrt{\frac{\pi\delta}{8}} \tilde{n}_t(\delta),$$

where $\tilde{n}_t(\delta)$ denotes the number of excursion intervals away from the origin, of duration $\geq \delta$, completed by $W_s, s \leq T_{M_t+}$. But according to (2) $\beta_t = T_{M_t+}$ is the time of completion of the excursion straddling t. Hence $n_t(\delta)$ and $\tilde{n}_t(\delta)$ differs at most by one, and (5) follows.

Exer. Show that the *zero set of BM* $\{t : W_t = 0\}$ is a (i) closed set of Lebesgue measure zero, and (ii) is unbounded and has an accumulation point at the origin. Hint: (i) use Fubini's theorem and continuity of BM, (ii) maximum and minimum of BM are both a.s. not equal to zero on any finite interval, and are both a.s. unbounded on $t \geq 0$.

Section 16. Skorohod imbedding and invariance principle.

For $a \leq 0 \leq b$ let $\nu_{a,b}$ be the unique probability measure on the two point set $\{a, b\}$ with mean zero so that $\nu_{a,b} = \delta_0$ for $ab = 0$ and

$$\nu_{a,b} = \frac{b\delta_a - a\delta_b}{b - a} \tag{1}$$

otherwise.

Prop. 1 (Randomization Lemma). *For any distribution μ on $\mathbf{R}$ of zero mean denote $\mu_\pm$ its restriction on $\mathbf{R}_+$ and $\mathbf{R}_-$ respectively and put $c = \int x\mu_+(dx) = -\int x\mu_-(dx)$. Then*

$$\mu = \int \tilde{\mu}(dx\,dy)\nu_{x,y}, \tag{2}$$

where the distribution $\tilde{\mu}$ on $\mathbf{R}_- \times \mathbf{R}_+$ is given by

$$\tilde{\mu}(dx\,dy) = \mu(\{0\})\delta_{0,0}(dx\,dy) + c^{-1}(y-x)\mu_-(dx)\mu_+(dy).$$

Proof. Direct calculations applying both sides of (2) to a continuous function f.

Prop. 2. *Let τ be a stopping time for BM B_t such that $B_{\min(t,\tau)}$ is uniformly bounded. Then*

$$\mathbf{E}B_\tau = 0, \quad \mathbf{E}\tau = \mathbf{E}B_\tau^2.$$

Proof. By optional stopping (and basic martingales)

$$\mathbf{E}B_{\min(t,\tau)} = 0, \quad \mathbf{E}(\min(t,\tau)) = \mathbf{E}B^2_{\min(t,\tau)},$$

and desired result is obtained by dominated and monotone convergence as $t \to \infty$.

Prop. 3 (embedding of r.v.). *For a probability measure μ on $\mathbf{R}$ with mean zero choose a random pair (a,b) with distribution $\tilde{\mu}$ from Prop. 1 and an independent BM B_t. Then (i) the random time $T = \inf\{t : B_t \in \{a,b\}\}$ is optional for filtration $\sigma\{a,b; B_s, s \le t\}$, (ii) the law of B_τ is μ, (iii) the expectation of τ coincides with the second moment (variance) of μ.*

Proof. By Exer. 1 of Section 12 the r.v. B_τ for fixed a,b would have the distribution (1). Hence

$$\mathbf{E}f(B_\tau) = \mathbf{E}\mathbf{E}(f(B_\tau)|a,b) = \int\int f(z)\nu_{x,y}(dz)\tilde{\mu}(dx\,dy) = \int f(x)\mu(dx),$$

yielding (ii). Then (iii) follows from Prop. 2.

Theorem 1 (Skorohod embedding). *Let $\xi_1, \xi_2, \dots$ be iid r.v. with mean 0 and $S_n = \xi_1 + \dots + \xi_n$. Then there exist a filtered probability space with a BM B_t and stopping times $0 = T_0 \le T_1 \le \dots$ s.t. the differences $\Delta T_n = T_n - T_{n-1}$ are iid with $\mathbf{E}\Delta T_n = \mathbf{E}\xi_1^2$ and B_{T_n} are distributed like S_n for all n.*

Remark. $\tau_n = \inf\{t \ge \tau_{n-1} : B_t = S_n\}$ would give a trivial solution if the moment requirement would not be imposed.

Proof. Let μ denote the common law of ξ_j. Take iid pairs (a_n, b_n), n=1,2,..., with the distribution $\tilde{\mu}$ from Prop. 1 and an independent BM. Everything follows from the recursively definition of random times $0 = T_0 \le T_1 \le T_2 \le \dots$ by

$$T_n = \inf\{t \ge T_{n-1} : B_t - B_{T_{n-1}} \in \{a,b\}\}.$$

Theorem 2 (Approximation of random walks). *Let $\xi_1, \xi_2, \dots$ be iid r.v. with mean 0 and variance 1, and let $S_n = \xi_1 + \dots + \xi_n$. Then there exists a BM B_t s.t.*

$X_t = t^{-1/2} \sup_{s \le t} |S_{[s]} - B_s|$ *converges to zero in probability as* $t \to \infty$ *(*$[s]$ *denotes the integer part of* s*).*

Proof. Choose T_n and B as in Theorem 1. Then $T_n/n \to 1$ a.s. by LLN, hence $T_{[t]}/t \to 1$ a.s. and hence (check it!) $\delta_t/t \to 0$ a.s., where $\delta_t = \sup_{s \le t} |T_{[s]} - s|$. For any t, h, ϵ by scaling property of BM

$$P(X_t > \epsilon) \le P(\delta_t > th) + P(\sup_{u - v \le th, u, v \le t + th} |B_u - B_v| > \epsilon\sqrt{t})$$

$$= P(\delta_t/t > h) + P(\sup_{u - v \le h, u, v \le 1 + h} |B_u - B_v| > \epsilon),$$

which can be made arbitrary small by choosing small h and large t.

Corollary (Functional CLT, invariance principle, two formulations). *(i) For all* $C, \epsilon > 0$ *there exists* N *s.t. for all* $n > N$ *there exists a BM* B_t *(depending on* n*) s.t.*

$$P\left(\sup_{t \le 1} \left| \frac{S_{[tn]}}{\sqrt{n}} - B_t \right| > C \right) < \epsilon.$$

(ii) Let F *be a uniformly continuous function on the space* $D[0,1]$ *of cadlag functions on* $[0,1]$ *equipped with the sup-norm topology. Then* $F(\frac{S_{[tn]}}{\sqrt{n}})$ *converges in distribution to* $F(B)$ *with* $B = B_t$ *being a standard BM.*

Proof. (i) Applying Theorem 2 with $t = n$ yields

$$P\left(n^{-1/2} \sup_{s \le n} |S_{[s]} - B_s| > C\right) \to 0$$

as $n \to \infty$ for any C. With $s = tn$ this rewrites as

$$P\left(\sup_{t \le 1} \left| \frac{S_{[tn]}}{\sqrt{n}} - \frac{B_{tn}}{\sqrt{n}} \right| > C\right) \to 0.$$

But by scaling $B_{tn}/\sqrt{n}$ is again a BM and (i) follows.

(ii) One has to show that

$$\mathbf{E}\left(g(F(\frac{S_{[.n]}}{\sqrt{n}})) - g(F(B_.))\right) \to 0 \tag{3}$$

as $n \to 0$ for any bounded uniformly continuous g. Choosing for each n a version of B from (i) one decomposes (3) into the sum of two terms with the function under the expectation multiplied by the indicators $\mathbf{1}_{Y_n > C}$ and $\mathbf{1}_{Y_n \le C}$ respectively, where

$$Y_n = \sup_{t \le 1} \left| \frac{S_{[tn]}}{\sqrt{n}} - B_t \right|.$$

Then the first term is small by (i) for any C and n large enough, and the second term is small for small C by uniform continuity of F and g.

Examples. 1. Applying statement (i) with $t = 1$ yields the usual CLT for random walks. 2. Applying (ii) with $F(h(.)) = \sup_{t \in [0,1]} h(t)$ and taking into account the distribution of the maximum of BM (obtained by the reflection principle) yields

$$P\left(\frac{\max\{S_k : k \leq n\}}{\sqrt{n}} \leq x\right) \to 2P(N \leq x), \quad x \geq 0,$$

where N is a standard normal r.v. $N(0,1)$.

Section 17. Sample path properties. Non-differentiability, quadratic variation, module of continuity, iterated log, rate of escape.