1.1. Probability spaces and σ-fields.

Definition (σ-field)**.** *We say that $\mathcal{F} \subset 2^{\Omega}$ is a σ-field (or a σ-algebra), if*
(a) $\Omega \in \mathcal{F}$,
(b) *If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$ as well,*
(c) *If $A_i \in \mathcal{F}$ for $i = 1, 2, \dots$, then also $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$.*

Remark. *Using DeMorgan's law, you can easily check that if $A_i \in \mathcal{F}$ for $i = 1, 2, \dots$ and $\mathcal{F}$ is a σ-field, then also $\cap_i A_i \in \mathcal{F}$. Similarly, you can show that a σ-field is closed under countable many elementary operations.*

Definition. *A pair $(\Omega, \mathcal{F})$ with $\mathcal{F}$ a σ-field of subsets of Ω is called a* measurable space. *Given a measurable space, a* probability measure $\mathbf{P}$ *is a function $\mathbf{P} : \mathcal{F} \to [0, 1]$, having the following properties:*
(a) $0 \le \mathbf{P}(A) \le 1$ *for all $A \in \mathcal{F}$,*
(b) $\mathbf{P}(\Omega) = 1$,
(c) *(Countable additivity) $\mathbf{P}(A) = \sum_{n=1}^{\infty} \mathbf{P}(A_n)$ whenever $A = \cup_{n=1}^{\infty} A_n$ is a countable union of disjoint sets $A_n \in \mathcal{F}$.*
A probability space *is a triplet $(\Omega, \mathcal{F}, \mathbf{P})$ with $\mathbf{P}$ a probability measure on the measurable space $(\Omega, \mathcal{F})$.*

Definition. *Given a collection of subsets $A_\alpha \subseteq \Omega$, where $\alpha \in \Gamma$ is a not necessarily countable index set, we denote the smallest σ-field $\mathcal{F}$ such that $A_\alpha \in \mathcal{F}$ for all $\alpha \in \Gamma$ by $\sigma(\{A_\alpha\})$ (or sometimes by $\sigma(A_\alpha, \alpha \in \Gamma)$, and call $\sigma(\{A_\alpha\})$ the* σ-field generated by the collection *$\{A_\alpha\}$. That is $\sigma(\{A_\alpha\}) - \cap\{\mathcal{G} : \mathcal{G} \subset 2^{\Omega}$ is a σ-field, $A_\alpha \in \mathcal{G} \forall \alpha \in \Gamma\}$.*

Example. *An example of a generated σ-field is the* Borel σ-field *on $\mathbb{R}$. It may be defined as $\mathcal{B} = \sigma(\{(a, b) : a, b \in \mathbb{R}\})$. \$*

Lemma. *If two different collections of generators $\{A_\alpha\}$ and $\{B_\beta\}$ are such that $A_\alpha \in \sigma(\{B_\beta\})$ for each α and $B_\beta \in \sigma(\{A_\alpha\})$ for each β, then $\sigma(\{A_\alpha\}) = \sigma(\{B_\beta\})$.*

Proposition. *There exists a subset of $\mathbb{R}$ that is not in $\mathcal{B}$. That is, not all sets are Borel sets.*

1.2: Random variables and their expectation.

Definition. *A* Random Variable *(R.V.) is a function $X : \Omega \to \mathbb{R}$ such that $\forall \alpha \in \mathbb{R}$ the set $\{\omega : X(\omega) \le \alpha\}$ is in $\mathcal{F}$ (such a function is also called a* $\mathcal{F}$-measurable or, simply, measurable function*).*

Remark. *Note that $X(\omega) = \sum_{n=1}^{N} c_n I_{A_n}(\omega)$ is a R.V. for any finite N, non-random $c_n \in \mathbb{R}$ and sets $A_n \in \mathcal{F}$. We call any such X a simple function, denoted $X \in$ SF.*

Proposition. *For every R.V. $X(\omega)$ there exists a sequence of simple functions $X_n(\omega)$ such that $X_n(\omega) \to X(\omega)$ as $n \to \infty$, for each fixed $\omega \in \Omega$.*

Definition. *We say that a R.V. X and Y are defined on the same probability space $(\Omega, \mathcal{F}, \mathbf{P})$ are almost surely the same if $\mathbf{P}(\{\omega : X(\omega) \ne Y(\omega)\}) = 0$. This shall be denoted by $X \overset{a.s.}{=} Y$. We also use the terms* almost surely, almost everywhere, *and* with probability 1 *interchangeably.*

Definition. *Given a R.V. X we denote $\sigma(X)$ the smallest σ-field $\mathcal{G} \subseteq \mathcal{F}$ such that $X(\omega)$ is measure on $(\Omega, \mathcal{G})$. One can show that $\sigma(X) = \sigma(\{\omega : X(\omega) \le \alpha\})$. We call $\sigma(X)$ the σ-field generated by X and interchangeably use the notations $\sigma(X)$ and $\mathcal{F}_X$. Similarly, given R.V. $X_1, \dots, X_n$ on the same measurable space $(\Omega, \mathcal{F})$, denote $\sigma(X_k, k \le n)$ the smallest σ-field $\mathcal{F}$ such that all $X_k(\omega)$ are measurable on $(\Omega, \mathcal{F})$.*

Exercise. *Let $(\Omega, \mathcal{F})$ be a measurable space and let X_n be a sequence of random variables on it. Assume that for each $\omega \in \Omega$, the limit $X_\infty(\omega) = \lim_{n\to\infty} X_n(\omega)$ exists and is finite. Prove that X_∞ is a random variable on $(\Omega, \mathcal{F})$.*

Definition. *A function $g : \mathbb{R} \to \mathbb{R}$ is called* Borel (measurable) function *if g is a R.V. on $(\mathbb{R}, \mathcal{B})$. if g is a R.V. on $(\mathbb{R}, \mathcal{B})$. We shall extend the notion of Borel sets and functions to $\mathbb{R}^n$ by defining the Borel σ-field on $\mathbb{R}^n$ as $\mathcal{B}_n = \sigma(\{[a_1, b_1] \times \cdots \times [a_n, b_n] : a_i, b_i \in \mathbb{R}, i = 1, \dots, n\})$ and calling $g : \mathbb{R}^n \to \mathbb{R}$ a Borel function if g is a R.V. on $(\mathbb{R}^n, \mathcal{B}_n)$.*

Proposition. *If $g : \mathbb{R}^n \to \mathbb{R}$ is a Borel function and $X_1, \dots, X_n$ are R.V. on $(\Omega, \mathcal{F})$, then $g(X_1, \dots, X_n)$ is also a R.V. on $(\Omega, \mathcal{F})$.*

Theorem. *If Z is a R.V. on $(\Omega, \sigma(Y_1, \dots, Y_n))$, then $Z = g(Y_1, \dots, Y_n)$ for some Borel function $g : \mathbb{R}^n \to \mathbb{R}$.*

Proposition. *For any $n < \infty$, any Borel function $g : \mathbb{R}^n \to \mathbb{R}$ and R.V. $Y_1, \dots, Y_n$ on the same measurable space we have the inclusion $\sigma(g(Y_1, \dots, Y_n)) \subseteq \sigma(Y_1, \dots, Y_n)$.*

Corollary. *Suppose R.V. $Y_1, \dots, Y_n$ and $Z_1, \dots, Z_m$ defined on the same measurable space are such that $Z_k = g_k(Y_1, \dots, Y_n)$, $k = 1, \dots, m$ and $Y_i = h_i(Z_1, \dots, Z_m)$, $i = 1, \dots, n$ for some Borel functions $g_k : \mathbb{R}^n \to \mathbb{R}$ and $h_i : \mathbb{R}^m \to \mathbb{R}$. Then $\sigma(Y_1, \dots, Y_n) = \sigma(Z_1, \dots, Z_m)$.*

Definition. *The (mathematical)* expectation *of a R.V. $X(\omega)$ is denote $\mathbf{E}X$. With $x_{k,n} = k2^{-n}$ and the intervals $I_{k,n} = (x_{k,n}, x_{k+1,n}]$ for $k = 0, 1, \dots,$ the expectation of $X(\omega) \ge 0$ is defined as: $\mathbf{E}X = \lim_{n\to\infty} \left[\sum_{k=0}^{\infty} x_{k,n} \mathbf{P}(\{\omega : X(\omega) \in I_{k,n}\})\right]$.*

Example. *Though note detailed in these notes, it is possible to show that for Ω countable and $\mathcal{F} = 2^{\Omega}$ our definition coincides with the well know elementary definition $\mathbf{E}X = \sum_\omega X(\omega) p_\omega$ (where $X(\omega) \ge 0$). More generally, the formula $\mathbf{E}X = \sum_i x_i \mathbf{P}(\{\omega : X(\omega) = x_i\})$ applies whenever the range of X is a bounded below countable set $\{x_1, x_2, \dots\}$ of real numbers e.g. whenever $X \in$ SF.*

Remark. *Using the elementary formula $\mathbf{E}Y = \sum_{m=1}^{N} c_m \mathbf{P}(A_m)$ for the simple function $Y(\omega) = \sum_{m=1}^{N} c_m I_{A_m}(\omega)$, it can be shown that our definition of the expectation of $X \ge 0$ coincides with $\mathbf{E}X = \sup\{\mathbf{E}Y : Y \in \text{SF}, 0 \le Y \le X\}$.*

Definition. *We say that a R.V. $X(\omega)$ has a* probability density function *f_X if $\mathbf{P}(a \le X \le b) = \int_a^b f_X(x)\,dx$ for every $a < b \in \mathbb{R}$. Such f_X must be a non-negative function with $\int_{\mathbb{R}} f_X(x)\,dx = 1$.*

Proposition. *When a non-negative R.V. $X(\omega)$ has a probability density function f_X, our definition of the expectation coincides with the well known elementary formula $\mathbf{E}X = \int_0^\infty x f_X(x)\,dx$.*

Remark. *The first definition of expectation is also called the* Lebesgue integral *of X with respect to the probability measure $\mathbf{P}$ and consequently denoted $\mathbf{E}X = \int X(\omega) d\mathbf{P}(\omega)$ (or $\int X(\omega)\mathbf{P}(d\omega)$). It is based on splitting the range of $X(\omega)$ to finite many small intervals. This allows us to deal with rather general domain Ω, in contrast to Riemann's integral where the* domain *of integration is split into finitely many small intervals and is hence limited to $\mathbb{R}^d$. Even when $\Omega = [0, 1]$ it allows us to deal with measures $\mathbf{P}$ for which $\omega \to \mathbf{P}([0, \omega])$ is not smooth (and hence Riemann's integral fails to exist). If the corresponding Riemann integral exists, then it necessarily coincides with our Lebesgue integral.*

Definition. *For a general R.V. X consider the non-negative R.V.-s $X_+ = max(X, 0$ and $X_- = -\min(X, 0)$, so $X = X_+ - X_-$. Let $\mathbf{E}X = \mathbf{E}X_+ - \mathbf{E}X_-$, provided either $\mathbf{E}X_+ < \infty$ or $\mathbf{E}X_- < \infty$.*

Definition. *We say that a random variable X is* integrable *(or has finite expectation) if $\mathbf{E}|X| < \infty$, that is, both $\mathbf{E}X_+ < \infty$ and $\mathbf{E}X_- < \infty$.*

Remark. *Suppose $X = Y - Z$ for some non-negative R.V.-s Y and Z. Then, necessarily $Y = D + X_+$ and $Z = D + X_-$ for some $D \ge 0$ and all $\omega \in \Omega$. It can be shown that the expectation is linear with respect to the addition of non-negative R.V. In particular, $\mathbf{E}Y = \mathbf{E}D + \mathbf{E}X_+$ and $\mathbf{E}Z = \mathbf{E}D + \mathbf{E}X_-$. Therefore, $\mathbf{E}Y - \mathbf{E}Z = \mathbf{E}X_+ - \mathbf{E}X_- = \mathbf{E}X$, provided either $\mathbf{E}Y$ or $\mathbf{E}Z$ is finite. We conclude in this case that $\mathbf{E}X = \mathbf{E}Y - \mathbf{E}Z$. However, it is possible to have X integrable while $\mathbf{E}D = \infty$ resulting with $\mathbf{E}Y = \mathbf{E}Z = \infty$.*

Proposition. *If a R.V. X has the probability density function f_X and $h : \mathbb{R} \to \mathbb{R}$ is a Borel measurable function, then the R.V. $Y = h(X)$ is integrable if and only if $\int_{-\infty}^{\infty} |h(x)| f_X(x)\,dx < \infty$, in which case $\mathbf{E}Y = \int_{-\infty}^{\infty} h(x) f_X(x)\,dx$.*

Definition. *A R.V. X with probability density function $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ for $x \in \mathbb{R}$ where $\mu \in \mathbb{R}$ and $\sigma > 0$ is called a non-degenerate Gaussian (or Normal) R.V. with mean μ and variance σ^2, denoted by $X \sim N(\mu, \sigma^2)$.*

Proposition. *The expectation has the following properties.*
(1) $\mathbf{E}I_A = \mathbf{P}(A)$ *for any $A \in \mathcal{F}$.*
(2) *If $X(\omega) = \sum_{n=1}^{N} c_n I_{A_n}$ is a simple function, then $\mathbf{E}X = \sum_{n=1}^{N} c_n \mathbf{P}(A_n)$.*
(3) *If X and Y are integrable R.V. then for any constants α, β the R.V. $\alpha X + \beta Y$ is integrable and $\mathbf{E}(\alpha X + \beta Y) = \alpha\mathbf{E}X + \beta\mathbf{E}Y$.*
(4) $\mathbf{E}X = c$ *if $X(\omega) = c$ with probability 1.*
(5) *Monotonicity: If $X \ge Y$ a.s., then $\mathbf{E}X \ge \mathbf{E}Y$. Further, if $X \ge Y$ a.s. and $\mathbf{E}X = \mathbf{E}Y$, then $X = Y$ a.s.*

Proposition (Jensen's inequality)**.** *Suppose $g(\cdot)$ is a convex function. If X is an integrable R.V. and $g(X)$ is also integrable, then $\mathbf{E}(g(x)) \ge g(\mathbf{E}X)$.*

Theorem (Markov's inequality)**.** *Suppose f is a non-decreasing, Borel measurable function with $f(x) > 0$ for any $x > 0$. Then, for any random variable X and all $\epsilon > 0$, $\mathbf{P}(|X(\omega)| > \epsilon) \le \frac{1}{f(\epsilon)} \mathbf{E}(f(|X|))$.*

Proposition. *Suppose Y and Z are random variables on the same probability space with both $\mathbf{E}[Y^2]$ and $\mathbf{E}[Z^2]$ finite. Then, $\mathbf{E}|YZ| \le \sqrt{\mathbf{E}Y^2 \mathbf{E}Z^2}$.*

1.3: Convergence of random variables.

Definition. *We say that random variables X_n converge to X almost surely, denoted $X_n \xrightarrow{a.s.} X$, if there exists $A \in \mathcal{F}$ with $\mathbf{P}(A) = 1$ such that $X_n(\omega) \to X(\omega)$ as $n \to \infty$ for each fixed $\omega \in A$.*

Exercise. *Show that if $X_n \xrightarrow{a.s.} X$ and f is a continuous function then $f(X_n) \xrightarrow{a.s.} f(X)$ as well.*

Definition. *We say that $(\Omega, \mathcal{F}, \mathbf{P})$ is a* complete probability space *if any subset N of $B \in \mathcal{F}$ with $\mathbf{P}(B) = 0$ is also in $\mathcal{F}$.*

Definition. *We say that X_n converges to X in probability, denoted $X_n \to_p X$, if $\mathbf{P}(\{\omega : |X_n(\omega) - X(\omega)| > \epsilon\}) \to 0$ as $n \to \infty$, for any fixed $\epsilon > 0$.*

Theorem. *We have the following relations:*

- *If $X_n \xrightarrow{a.s} X$, then $X_n \to_p X$.*
- *If $X_n \to_p X$, then there exists a subsequence n_k such that $X_{n_k} \xrightarrow{a.s.} X$ for $k \to \infty$.*

Proposition. *In general, $X_n \to_p X$ does not imply $X_n \xrightarrow{a.s.} X$.*

Definition. *Let $\{A_n\}$ be a sequence of events, and $B_n = \cup_{k=n}^{\infty} A_k$. Define $A^\infty = \cap_{n=1}^{\infty} B_n$, so $\omega \in A^\infty$ if and only if $\omega \in A_k$ for infinitely many values of k.*

Lemma (Borel-Cantelli I). *Suppose $A_k \in \mathcal{F}$ and $\sum_{k=1}^{\infty} \mathbf{P}(A_k) < \infty$. Then, necessarily $\mathbf{P}(A^\infty) = 0$.*

Lemma (Borel-Cantelli II). *If $\{A_k\}$ are independent and $\sum_{k=1}^{\infty} \mathbf{P}(A_k) = \infty$, then $\mathbf{P}(A^\infty) = 1$.*

Proposition. *Suppose $\mathbf{E}[X_n^2] \leq 1$ for all n. Then $n^{-1}X_n(\omega) \to 0$ a.s. for $n \to \infty$.*

Definition. *Fixing $1 \leq q < \infty$, we denote $L^q(\Omega, \mathcal{F}, \mathbf{P})$ the collection of random variables X on $(\Omega, \mathcal{F})$ for which $\mathbf{E}(|X|^q) < \infty$. L^1 denotes the space of integrable functions, and L^2 denotes the space of square-integrable functions.*

Proposition. *The sequence $||X||_q = [\mathbf{E}(|X|^q)]^{1/q}$ is non-decreasing in q.*

Definition. *We say that X_n converges in q-mean, or in L^q to X, denoted $X_n \xrightarrow{q.m.} X$ if $X_n, X \in L^q$ and $||X_n - X||_q \to 0$ as $n \to \infty$ (i.e. $\mathbf{E}(|X_n - X|^q) \to 0$) as $n \to \infty$.*

Corollary. *If $X_n \xrightarrow{q.m.} X$ and $q \geq r$, then $X_n \xrightarrow{r.m.} X$.*

Proposition. *$L^q(\Omega, \mathcal{F}, \mathbf{P})$ is a complete, normed (topological) vector space with the norm $|| \cdot ||_q$. That is $\alpha X + \beta Y \in L^q$ whenever $X, Y \in L^q$, $\alpha, \beta \in \mathbb{R}$, with $X \mapsto ||X||_q$ a norm on L^q and if $X_n \in L^q$ and if $X_n \in L^q$ are such that $||X_n - X_m|| \to 0$ as $n, m \to \infty$ then $X_n \xrightarrow{q.m.} X$ for some $X \in L^q$.*

Proposition. *If $X_n \xrightarrow{q.m.} X$, then $X_n \to_p X$.*

Example. *The converse of the above does not hold in general.*

Exercise. *Give a **counterexample** to each of the following claims:*

- *If $X_n \to X$ a.s., then $X_n \to X$ in L^q, $q \geq 1$;*
- *If $X_n \to X$ in L^q then $X_n \to X$ a.s.;*
- *If $X_n \to X$ in probability then $X_n \to X$ a.s.*

Example. *For $n \in \mathbb{N}$ find the largest k s.t. $2^k \leq n$ and write $n = 2^k + m$. Then let $X_n = \mathbf{1}_{[m/2^k, (m+1)/2^k]} : [0, 1] \to \mathbb{R}$. $X_n \to 0$ in probability and L^q but not a.s.*

Example. *On $[0, 1]$ let $X_n = n\mathbf{1}_{[0,1/n]}$. Then $X_n \to 0$ a.s. but not in L^q.*

Proposition. *If $X_n \xrightarrow{q.m.} X$ and $X_n \xrightarrow{a.s.} Y$, then $X = Y$ a.s.*

1.4: Independence, weak convergence and uniform integrability.

Definition. *The law of a R.V. X, denoted $\mathcal{P}_X$, is the probability measure on $(\mathbb{R}, \mathcal{B})$ such that $\mathcal{P}_X(B) = \mathbf{P}(\{\omega : X(\omega) \in B\})$ for any Borel set B.*

Proposition. *Let X be a R.V. on $(\Omega, \mathcal{F}, \mathbf{P})$ and let g be a Borel function of $\mathbb{R}$. Suppose either g is non-negative or $\mathbf{E}|g(X)| < \infty$. Then, $\mathbf{E}[g(X)] = \int_{\mathbb{R}} g(x) d\mathcal{P}_X(x)$, where the integral on the rhs merely denotes the expectation of the random variables $g(x)$ on the (new) probability space $(\mathbb{R}, \mathcal{B}, \mathcal{P}_X)$.*

Definition. *The distribution function F_X of a real-valued R.V. X is $F_X(\alpha) = \mathbf{P}(\{\omega : X(\omega) < \alpha\}) = \mathcal{P}_X((-\infty, \alpha])$ for all $\alpha \in \mathbb{R}$.*

Proposition. *The distribution function F_X uniquely determines the law $\mathcal{P}_X$ of X.*

Proposition. *A R.V. X has a (probability) density (function) f_X if and only if its distribution function F_X can be expressed as $F_X(\alpha) = \int_{-\infty}^{\alpha} f_X(x)\, dx$, for all $\alpha \in \mathbb{R}$. Such F_X is continuous and almost everywhere differentiable with $\frac{dF_X}{dx}(x) = f_X(x)$ for almost every x.*

Definition. *We say that R.V.-s X_n converge in law (or weakly) to a R.V. X, denoted by $X_n \xrightarrow{\mathcal{L}} X$ if $F_{X_n}(\alpha) \to F_X(\alpha)$ as $n \to \infty$ for each fixed α which is a continuity point of F_X. In other books, this is called convergence in distribution, and denoted $X_n \xrightarrow{\mathcal{D}} X$.*

Proposition. *$X_n \xrightarrow{\mathcal{L}} X$ if and only if for each bounded h that is continuous on the range of X we have that $\mathbf{E}h(X_n) \to \mathbf{E}h(X)$ as $n \to \infty$.*

Remark. *Note that $F_X(\alpha) = \mathcal{P}_X((-\infty, \alpha]) = \mathbf{E}[I_{(-\infty,\alpha]}(X)]$ involves the function $h(x) = 1_{(-\infty,\alpha]}(x)$. Restricting the convergence in law to continuity points of F_X is what makes the above possible.*

Proposition. *If $X_n \to_p X$, then $X_n \xrightarrow{\mathcal{L}} X$.*

Proposition. *If $X_n \xrightarrow{\mathcal{L}} X$ and X is a non-random constant (almost surely), then $X_n \to_p X$ as well.*

Definition. *We say that a sequence of probability measure Q_n on a topological space $\mathbb{S}$ (i.e. a set with a notion of open sets, or topology) and its Borel σ-field $\mathcal{B}_{\mathbb{S}}$ (= the σ-field generated by the open subsets of $\mathbb{S}$), converges weakly to a probability measure Q if for each fixed g continuous and bounded on $\mathbb{S}$, $\int_{\mathbb{S}} h(\omega) dQ_n(\omega) \to \int_{\mathbb{S}} h(\omega) dQ(\omega)$ as $n \to \infty$ (the integrals denote the expectation of the R.V. $h(\omega)$ in their respective probability spaces). We shall use $Q_n \Rightarrow Q$ to denote weak convergence.*

Example. *$X_n \xrightarrow{\mathcal{L}} X$ if and only if $\mathcal{P}_{X_n} \Rightarrow \mathcal{P}_X$. That is, when $\int_{\mathbb{R}} h(\xi) d\mathcal{P}_{X_n}(\xi) \to \int_{\mathbb{R}} h(\xi) d\mathcal{P}_X(\xi)$, for each fixed $h : \mathbb{R} \to \mathbb{R}$ continuous and bounded.*

Definition (Uniform integrability). *A collection of R.V.-s $\{X_\alpha, \alpha \in \mathcal{I}\}$ is called U.I. if $\lim_{M\to\infty} \sup_\alpha \mathbf{E}[|X_\alpha| I_{|X_\alpha|>M}] = 0$.*

Theorem (Dominated Convergence). *If there exists a random variable Y such that $\mathbf{E}Y < \infty$, $|X_n| \leq Y$ for all n and $X_n \to_p X$, then $\mathbf{E}X_n \to \mathbf{E}X$.*

Corollary (Bounded Convergence). *Suppose $|X_n| \leq C$ for some finite constance C and all n. If $X_n \to X$, then also $\mathbf{E}X_n \to \mathbf{E}X$.*

Theorem (Monotone Convergence). *If $X_n \geq 0$ and $X_n(\omega) \uparrow X(\omega)$ for a.e. ω, then $\mathbf{E}X_n \uparrow \mathbf{E}X$. This applies even if $X(\omega) = \infty$ for some $\omega \in \Omega$.*

Definition. *Events $A_i 1\mathcal{F}$ are $\mathbf{P}$-mutually independent if for any $L < \infty$ and distinct indices $i_1, i_2, \ldots, i_L$, $\mathbf{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_L}) = \prod_{k=1}^{L} \mathbf{P}(A_{i_k})$.*

Definition. *Two σ-fields $\mathcal{H}, \mathcal{G} \subseteq \mathcal{F}$ are $\mathbf{P}$-independent if $\mathbf{P}(G \cap H) = \mathbf{P}(G)\mathbf{P}(H)$ for all $G \in \mathcal{G}, H \in \mathcal{H}$. The random vectors vectors $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_m)$ are independent if their corresponding generated σ-fields are independent.*

Proposition. *For any finite $n, m \geq 1$, two random vector $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_m)$ with values in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively, are independent if and only if $\mathbf{E}(h(X_1, \ldots, X_n)g(Y_1, \ldots, Y_m)) = \mathbf{E}(h(X_1, \ldots, X_n))\mathbf{E}(g(Y_1, \ldots, Y_m))$, for all bounded, Borel measurable function $g : \mathbb{R}^m \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}$.*

Definition. *Square-integrable random variables X and Y defined on the same probability space are called uncorrelated if $\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y)$.*

Remark. *Independent random variables are uncorrelated, but the converse is not neccessarily true.*

Proposition. *Any two square-integrable independent randome variables X and Y are also uncorrelated.*

Chapter 2. Conditional expectation (C.E.) and Hilbert spaces

2.1. Condition expectation: existence and uniqueness.

Proposition. *There exists a unique (a.s.) optimal $A \in \mathcal{H}_Y$ such that $\mathbf{d}^2 = \mathbf{E}[(X - Z)^2]$. Further, the optimality of Z is equivalent to the orthogonality property $\mathbf{E}[(X - Z)V] = 0$ for all $V \in \mathcal{H}_Y$.*

Definition. *For $X \in L^2(\Omega, \mathcal{F}, \mathbf{P})$, the* conditional expectation *$Z = \mathbf{E}(X|Y)$ is the unique R.V. in $\mathcal{H}_Y$ satisfying the orthogonality property above.*

Definition. *The* conditional expectation *of $X \in L^1(\Omega, \mathcal{F}, \mathbf{P})$ given a σ-field $\mathcal{G} \subseteq \mathcal{F}$, is the R.V. Z on $(\Omega, \mathcal{G})$ such that $\mathbf{E}[(X - Z)I_A] = 0$ for all $A \in \mathcal{G}$. $\mathbf{E}(X|Y)$ correspond to the special case of $\mathcal{G} = \mathcal{F}_Y$.*

Theorem. *The C.E. of integrable R.V. X given any σ-field $\mathcal{G}$ exists and is a.s. unique. That is, there exists Z measurable on $\mathcal{G}$ that satisfies the above definition, and if Z_1 and Z_2 are both measurable on $\mathcal{G}$ satisfying this then $Z_1 \overset{a.s.}{=} Z_2$. Further, if we also have $\mathbf{E}X^2 < \infty$, then Z also satisfies $\mathbf{E}[(X - Z)V] = 0$ for all $V \in L^2(\Omega, \mathcal{G}, \mathbf{P}$, hence for $\mathcal{G} = \mathcal{F}_Y$, Z coincides with the first definition.*

Proposition. *If X is a non-negative R.V. then a.s. $\mathbf{E}(X|\mathcal{G}) \geq 0$.*

2.2. Hilbert spaces.

2.3. Properties of the conditional expectation.

Remark. *Note that X may have the same law as Y while $\mathbf{E}(X|\mathcal{G})$ does not have the same law as $\mathbf{E}(Y|\mathcal{G})$. For example, take $\mathcal{G} = \sigma(X)$ with X and Y square integrable, independent, of the same distribution and positive variance. Then, $\mathbf{E}(X|\mathcal{G}) = X$ and $\mathbf{E}(Y|\mathcal{G}) = \mathbf{E}Y$ have different laws.*

Proposition. *Conditional expectation is linear, i.e. $X, Y \in L^1(\Omega, \mathcal{F}, \mathbf{P})$ implies $\mathbf{E}(\alpha X + \beta Y|\mathcal{G}) = \alpha \mathbf{E}(X|\mathcal{G}) + \beta \mathbf{E}(Y|\mathcal{G})$.*

Proposition (Tower property)**.** *Suppose $\mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{F}$ and $X \in L^1(\Omega, \mathcal{F}, \mathbf{P})$. Then, $\mathbf{E}(X|\mathcal{H}) = \mathbf{E}(\mathbf{E}(X|\mathcal{G})|\mathcal{H})$.*

Proposition (Jensen's inequality)**.** *Let $f : \mathbb{R} \to \mathbb{R}$ be a convex function (that is, $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$ for any x, y and $0 \leq \lambda \leq 1$). Suppose $X \in L^1(\Omega, \mathcal{F}, \mathbf{P})$ is such that $\mathbf{E}\,|f(X)| < \infty$. Then, $\mathbf{E}(f(X)|\mathcal{G}) \geq f(\mathbf{E}(X|\mathcal{G}))$.*

Corollary. *For each $q \geq 1$, the norm of the conditional expectation of $X \in L^q(\Omega, \mathcal{F}, \mathbf{P})$ given a σ-field $\mathcal{G}$ never exceeds the(L^q)-norm of X. ($f(x) = |x|^q$).*

Theorem (Monotone Convergence for C.E.)**.** *If $0 \leq X_m \nearrow X$ and $\mathbf{E}(X) < \infty$ then $\mathbf{E}(X_m|\mathcal{G}) \nearrow \mathbf{E}(X|\mathcal{G})$ a.s.*

Theorem (Dominated Convergence for C.E.)**.** *If $|X_m| \leq Y \in L^1(\Sigma, \mathcal{F}, \mathbf{P})$ and $X_m \overset{a.s.}{\to} X$, then $\mathbf{E}(X_m|\mathcal{G}) \overset{a.s.}{\to} \mathbf{E}(X|\mathcal{G})$.*

Remark. *In contrast to regular DCT, convergence in* probability *to X by X_m that are dominated does not imply a.s. convergence of $\mathbf{E}(X_m|\mathcal{G})$. Taking $\mathcal{G} = \mathcal{F}$ this contradicts "$X_n \to_p X$ does not imply $X_n \overset{a.s.}{\to} X$."*

Theorem. *Suppose $X_n \overset{q.m.}{\to} X$, that is $X_n, X \in L^q$ with $\mathbf{E}\,|X_n - X| \to 0$. Then, $\mathbf{E}(X_n|\mathcal{G}) \overset{q.m.}{\to} \mathbf{E}(X|\mathcal{G})$.*

Proposition. *Suppose Y is bounded and measurable on $\mathcal{G}$, and that $X \in L^1(\Omega, \mathcal{F}, \mathbf{P})$. Then, $\mathbf{E}(XY|\mathcal{G}) = Y\mathbf{E}(X|G)$.*

Proposition. *If X is integrable and σ-fields $\mathcal{G}$ and $\sigma(\sigma(X), \mathcal{H})$ are independent, then $\mathbf{E}[X|\sigma(\mathcal{G}, \mathcal{H})] = \mathbf{E}[X|\mathcal{H}]$.*

2.4. Regular conditional probability.

Definition. *The* conditional expectation *of an integrable random variable X given a σ-field $\mathcal{G}$ is $\mathbf{E}[X|\mathcal{G}] = \int_{\mathbb{R}} x \hat{\mathbf{P}}_{X|\mathcal{G}}(dx, \omega)$.*

Example. *The R.C.P.D. is explicit when $\mathcal{G} = \sigma(Y)$ and the random vector (X, Y) has a* probability density function *$f_{X,Y}$. For all $x, y \in \mathbb{R}$, $\mathbf{P}(X \leq x, Y \leq y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(u, v)\,du\,dv$. In this case the R.C.P.D. of X given $\mathcal{G} = \sigma(Y)$ has density function $f_{X|Y}(x|Y(\omega))$ where $f_{X|Y}(x, y) = f_{X,Y}(x, y)/f_Y(y)$ and $f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(v, y)\,dv$. Hence $\mathbf{E}(X|Y) = \int_{\mathbb{R}} x f_{X|Y}(x|Y)\,dx$.*

Chapter 3. Stochastic Processes: general theory

3.1: Definition, distribution, and versions.

Definition. *Given $(\Omega, \mathcal{F}, \mathbf{P})$, a* stochastic process *(S.P.) $\{X_t\}$ is a collection of R.V.-s indexed by $t \in I$. We call $t \mapsto X_t(\omega)$ the* sample path *of the S.P.*

Definition. *A* random walk *is the sequence $S_n = \sum_{i=1}^{n} \xi_i$, where ξ_i are i.i.d. real-valued R.V.-s defined on the same $(\Omega, \mathcal{F}, \mathbf{P})$. When $\xi_i \in \mathbb{Z}$ we say it's a random walk on the integers, and we call $\xi_i \in \{-1, 1\}$ a simple random walk.*

Theorem. *Consider the random walk S_n when $\mathbf{E}\xi_i = 0$ and $\mathbf{E}\xi_i^2 = 1$. Take the linear interpolation of S_n, scale space by $n^{-1/2}$ and time by n^{-1}. Taking $n \to \infty$ we get what we call* Brownian motion *on $0 \leq t \leq 1$.*

Definition. *Given $N < \infty$ and $t_1, \ldots, t_N \in \mathcal{I}$, we denote the (joint) distribution of $(X_{t_1}, \ldots, X_{t_N})$ by $F_{t_1,\ldots,t_N}(\cdot)$, i.e. $F_{t_1,\ldots,t_N}(\alpha_1, \ldots, \alpha_N) = \mathbf{P}(X_{t_1 \leq \alpha_1}, \ldots, X_{t_N \leq \alpha_N})$ for all $\alpha_i \in \mathbb{R}$. We call the collection of functions $F_{t_1,\ldots,t_N}(\cdot)$ the finite dimensional distributions (f.d.d.) of the S.P.*

Definition. *With $\mathcal{G}_t$ the smallest σ-field containing $\sigma(X_s)$ for any $0 \leq s \leq t$, we say that a S.P. $\{X_t\}$ has* indepedent increments *if $X_{t+h} - X_t$ is independent of $\mathcal{G}_t$ for any $h > 0$ and all $t \geq 0$. This property is determined by the f.d.d. That is, if $X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$ are mutually independent for all $n < \infty$ and $0 \leq t_1 < t_2 < \cdots < t_n < \infty$ then the S.P. $\{X_t\}$ has independent increments.*

Remark. *For example, both the random walk and the Brownian motion have indepedent increments.*

Example. *Consider $\Omega = [0, 1]$ with Borel σ-field and Uniform law. Define $Y_t(\omega) \equiv 0$ and $X_t(\omega) = \mathbf{1}_{\{\omega\}}(t)$. Note that $\mathbf{P}(X_t = Y_t) = 1$ for any fixed t but $\mathbf{P}(X_t = Y_t \forall t \in [0, 1]) = 0$ and no sample paths $t \mapsto X_t(\omega)$ are continuous.*

Definition. *Two S.P. $\{X_t\}$ and $\{Y_t\}$ are called* versions *of one another if they have the same f.d.d.s.*

Definition. *A S.P. $\{Y_t\}$ is called a* modification *of another S.P. $\{X_t\}$ if $\mathbf{P}(X_t = Y_t) = 1$ for all t.*

Exercise. *If $\{Y_t\}$ is a modification of $\{X_t\}$ then $\{Y_t\}$ is also a version of $\{X_t\}$.*

Example. *Consider $\Omega = \{HH, TT, HT, TH\}$ with uniform probability measure, corresponding to two indepdent fair coin tosses with outcome $\omega = (\omega_1, \omega_2)$. Let $X_t(\omega) = \mathbf{1}_{[0,1)}(t) I_H(\omega_1) + \mathbf{1}_{[1,2)}(t) I_H(\omega_2)$ for $0 \leq t < 2$, and let $Y_t(\omega) = 1 - X_t(\omega)$. These are version of each other but not modifications of each other.*

Definition. *We say that a collection of f.d.d.s is* consistent *if*

$$\lim_{\alpha_k \uparrow \infty} F_{t_1,\ldots,t_N}(\alpha_1, \ldots, \alpha_N) = F_{t_1,\ldots,t_{k-1},t_{k+1},\ldots,t_N}(\alpha_1, \ldots, \alpha_{k-1}, \alpha_{k+1}, \ldots, \alpha_N),$$

for any $1 \leq k \leq N$, $t_1 < t_2 < \cdots < t_N$ and $\alpha_i \in \mathbb{R}$.

Proposition. *The f.d.d.s of any S.P. must be consistent. Conversely, for any consistent collection f.d.d.s, there exists a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and a stochastic process $\{X_t(\omega)\}$ on it, whose f.d.d.s are in agreement with the given collection. Further, the restriction of $\mathbf{P}$ to the σ-field $\mathcal{F}_X$ is uniquely determined by the given collection of f.d.d.*

3.2: Characteristic functions, Gaussian variables and processes.

Definition. *A random vector $\underline{X} = (X_1, \ldots, X_n)$ with values in $\mathbb{R}^n$ has the* characteristic function *$\Phi_{\underline{X}}(\underline{\theta}) = \mathbf{E}[\exp(i \sum_{k=1}^{n} \theta_k X_k)]$, where $\underline{\theta} = (\theta_1, \ldots, \theta_n) \in \mathbb{R}^n$.*

Remark. *The characteristic function exists for any $\underline{X}$ because trig functions are bounded.*

Proposition. *The characteristic function determines the law of a random vector. That is, if $\Phi_{\underline{X}}(\underline{\theta}) = \Phi_{\underline{Y}}(\underline{\theta})$ for all $\underline{\theta}$ then $\underline{X}$ has the same law (= probability measure on $\mathbb{R}^n$) as $\underline{Y}$.*

Exercise. *If $X_n \overset{\mathcal{L}}{\to} X$ then $\Phi_{X_n}(\theta) \to \Phi_X(\theta)$ for any θ.*

Remark. *The converse of the previous exercise is also true.*

Definition. *We say that a random vector $\underline{X} = (X_1, \ldots, X_n)$ has a* probability density function *$f_{\underline{X}}$ if*

$$\mathbf{P}(a_i \leq X_i \leq b_i, i = 1, \ldots, n) = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f_{\underline{X}}(x_1, \ldots, x_n)\,dx_n \cdots dx_1$$

for every $a_i < b_i, i = 1, \ldots, n$. Such density $f_{\underline{X}}$ must be a non-negative Borel measurable function with $\int_{\mathbb{R}^n} f_{\underline{X}}(\underline{x})\,d\underline{x} = 1$. $f_{\underline{X}}$ can be called the joint density *of X_i.*

Proposition. *If $\underline{X} = (X_1, \ldots, X_n)$, with X_i R.V.s, then X_i are mutually independent if and only if $\Phi_{\underline{X}}(\underline{\theta}) = \mathbf{E}\left[\prod_{k=1}^{n} e^{i\theta_k X_k}\right] = \prod_{k=1}^{n} \Phi_{X_k}(\theta_k)$ for all $\underline{\theta} \in \mathbb{R}^n$.*

Definition. *We say that a random vector $\underline{X} = (X_1, \ldots, X_n)$ has a* Gaussian *(or multivariate Normal) distribution if $\Phi_{\underline{X}}(\underline{\theta}) = \exp\left(-\frac{1}{2}\langle\underline{\theta}, \Sigma\underline{\theta}\rangle + i\langle\underline{\theta}, \underline{\mu}\rangle\right)$ for some positive semidefinite $n \times n$ matrix Σ, some $\underline{\mu} \in \mathbb{R}^n$, and all $\underline{\theta} \in \mathbb{R}^n$.*

Remark. *When $n = 1$ we say that a R.V. X is Gaussian if $\mathbf{E}[e^{i\theta X}] = e^{-\frac{1}{2}\theta^2\sigma^2 + i\theta\mu}$.*

Definition. *We say that $\underline{X}$ has a* non-degenerate *Gaussian distribution if Σ is invertible.*

Proposition. *A random vector $\underline{X}$ with a non-degenerate Gaussian distribution has the density $f_{\underline{X}}(\underline{x}) = (2\pi)^{-n/2}(\det\Sigma)^{-1/2}\exp\left(-\frac{1}{2}\langle\underline{x} - \underline{\mu}, \Sigma^{-1}(\underline{x} - \underline{\mu})\rangle\right)$. In particular, if $\sigma^2 > 0$, then a Gaussian R.V. X has density $f_X(x) = (2\pi)^{-1/2}\sigma^{-1}\exp\left(-\frac{1}{2}(x - \mu)^2/\sigma^2\right)$.*

Proposition. *The* parameters of the Gaussian distribution *are $\mu_j = \mathbf{E}X_j$ and $\Sigma_{jk} = \mathbf{E}[(X_j - \mu_j)(X_k - \mu_k)]$.*

Proposition. *If a Gaussian random vector $\underline{X} = (X_1, \ldots, X_n)$ has uncorrelated coordinates, then its coordinates are also mutually independent.*

Proposition. *Suppose a sequence of n-dimensional Gaussian random vectors $\underline{X}^{(k)}$ converges in 2-mean to an n-dimensional random vector $\underline{X}$, that is, $\mathbf{E}[(X_i - X_i^{(k)})^2] \to 0$ as $k \to \infty$ for each i. Then, $\underline{X}$ is a Gaussian random vector, whose parameters are the limits of the corresponding parameters of $\underline{X}^{(k)}$.*

Proposition. *A random vector $\underline{X}$ has the Gaussian distribution if and only if $(\sum_{i=1}^{n} a_{ji} X_i, j = 1, \ldots, m)$ is a Gaussian random vector for any non-random coefficients $a_{ij} \in \mathbb{R}$.*

Definition. *A stochastic process (S.P.) $\{X_t\}$ is* Gaussian *if for all $n < \infty$ and all $t_1, \ldots, t_n \in \mathcal{I}$ the random vector $(X_{t_1}, \ldots, X_{t_n})$ has a Gaussian distribution, that is, all f.d.d.s of the process are Gaussian*

Corollary. *All distributional properties of Gaussian processes are determined by the mean $\mu(t) = \mathbf{E}X_t$ of the process and its autocovariance function $\rho(t, s) = \mathbf{E}[(X_t - \mu(t))(X_s - \mu(s))]$.*

Proposition. *If $Cov(Y_{t+h} - Y_t, Y_s) = 0$ for a Gaussian stochastic process $\{Y_t\}$, all $t \geq s$ and $h > 0$, then the S.P. $\{Y_t\}$ has uncorrelated hence independent increments (which is thus also equivalent to $\mathbf{E}(Y_{t+h} - Y_t|\sigma(Y_s, s \leq t)) = \mathbf{E}(Y_{t+h} - Y_t)$ for any $t \geq 0$ and $h > 0$).*

Proposition. *If the S.P. $\{X_t, t \in \mathcal{I}\}$ and the Gaussian S.P. $\{X_t^{(k)}, t \in \mathcal{I}\}$ are such that $\mathbf{E}[(X_t - X_t^{(k)})^2] \to 0$ as $k \to \infty$, for each fixed $t \in \mathcal{I}$, then X_t is a Gaussian S.P. with mean and auto-covariance functions that are the pointwise limits of those for $X_t^{(k)}$.*

Definition. *A stochastic process $\{X_t\}$ indexed by $t \in \mathbb{R}$ is called* (strong sense) stationary *if its f.d.d. satisfy $F_{t_1,\ldots,t_N}(\alpha_1, \ldots, \alpha_N) = F_{t_1+\tau,\ldots,t_N+\tau}(\alpha_1, \ldots, \alpha_N)$ for all $\tau \in \mathbb{R}$, $N < \infty$, $\alpha_i \in \mathbb{R}$, and $t_1 < \cdots < t_N$. A similar definition applies to discrete time S.P. indexed by t on the integers, just then t_i and τ take only integer values.*

Proposition. *A Gaussian S.P. is stationary if and only if $\mu(t) = \mu$ (a constant) and $\rho(t, s) = r(|t - s|)$, where $r : \mathbb{R} \to \mathbb{R}$ is a function of the time difference. (A stochastic process whose mean and auto-covariance function satisfy these two properties is called* weak sense (or covariance) stationary*. In general, a weak sense stationary process is not a strong sense stationary process. However, for Gaussians, they are the same.*

Definition. *A process $\{X_t, t \geq 0\}$ has stationary increments if $X_{t+h} - X_t$ and $X_{s+h} - X_s$ have the same law for all $s, t, h \geq 0$. The same definition applies to discrete time S.P. except for integer values.*

Example. *A sequence of i.i.d. r.v.s $\{X_n, n \in \mathbb{Z}\}$ is a discrete time stationary process. However, many processes are not stationary. For example, the random walk $S_n = \sum_{i=1}^n X_i$ is a non-stationary S.P. when $\mathbf{E}X_1 = 0$ and $\mathbf{E}X_1^2 = 1$. Otherwise the law of S_n and in particular its second moment would not depend on n, but clearly $\mathbf{E}S_n = n$. Also, every stationary process has stationary increments, but the random walk S_n has stationary increments but is not stationary.*

3.3: Sample path continuity.

Definition. *$\{X_t\}$ has* continuous sample path *with probability 1 if $\mathbf{P}(\{\omega : t \mapsto X_t(\omega)$ is continuous$\}) = 1$. Similarly, we use the term* continuous modification *to denote a modification $\{\tilde{X}_t\}$ of $\{X_t\}$ that has continous sample path with probability 1.*

Definition. *A S.P. Y_t is* locally Hölder continuous *with exponent γ if for some $c < \infty$ and a R.V. $h(\omega) > 0$,*

$$\mathbf{P}\left(\left\{\omega : \sup_{0 \leq s,t \leq T, |t-s| \leq h(\omega)} |Y_t(\omega) - Y_s(\omega)| \leq c\,|t - s|^\gamma\right\}\right) = 1.$$

Note the "locally" refers to $h(\omega)$; if it holds for unrestricted t, s we say Y_t is globally/uniformly Hölder continuous *with exponent γ.*

Theorem (Kolmogorov's continuity theorem). *Given a S.P. $\{X_t, t \in [0, T]\}$, suppose there exist $\alpha, \beta, c, h_0 > 0$ so that $\mathbf{E}(|X_{t+h} - X_t|^\alpha) \leq ch^{1+\beta}$ for all $0 \leq t, t + h \leq T$, $0 < h < h_0$. Then, there exists a continuous modification Y_t of X_t so that Y_t is also locally Hölder continuous with exponent γ for any $0 < \gamma < \beta/\alpha$.*

Example. *Consider the S.P. $X_t(\omega) = I_{\{U \leq t\}}(\omega)$ for $t \in [0, 1]$ where U is Uniform on $[0, 1]$ (jumps up at a random point). Note that $|X_{t+h} - X_t| = 1$ if $0 \leq t < U \leq t + h$ and equals zero otherwise. So, $\mathbf{E}(|X_{t+h} - X_t|^\alpha) = U((t, t + h]) \leq h$ for any $h > 0$ and $t \geq 0$. However, $X_t(\omega)$ is discontinuous as $t = U(\omega)$ when $\omega \neq 0$. So $\{X_t\}$ a.s. has discontinuous sample paths (and it can be shown it has no continuous modification).*

Exercise. *Suppose the S.P. X_t has zero mean and variance one always. Then $|\mathbf{E}(X_t X_{t+h})| \leq 1$ for any $h > 0$. Also, if we have for $p > 1$ $\mathbf{E}(X_t X_{t+h}) \geq 1 - \lambda h^p$ for all $0 < h \leq h_0$, then by Kolmogorov, X_t has a continuous modification. Now suppose X_t a Gaussian S.P. with $\mathbf{E}X_t = 0$ and $EX_t^2 = 0$. Then if X_t satisfies $\mathbf{E}(X_t X_{t+h}) \geq 1 - \lambda h^p$ for all $0 < h \leq h_0$, for some finite λ, $p > 0$ (not $p > 1$), then it has a modification which is locally Hölder continuous for all $\gamma < p/2$.*

Example (Random telegraph noise). *Let $\tau_i, i = \in \mathbb{N}$ be independent random times each having the Exponential(1) distribution, i.e. $\mathbf{P}(\tau_i \leq x) = 1 - e^{-x}$ for all i, $x \geq 0$. Starting at $R_0 \in \{-1, 1\}$ such that $\mathbf{P}(R_0 = 1) = 1/2$, the S.P. R_t alternately jumps between ± 1 at the random times $s_k = \sum_{i=1}^k \tau_i$, so R_t is constant on each (s_k, s_{k+1}). Since almost surely $s_1 < \infty$ this S.P. does not have a continuous modification. However $\mathbf{E}(R_t R_{t+\varepsilon}) = \mathbf{P}(R_t = R_{t+\varepsilon}) - \mathbf{P}(R_t \neq R_{t+\varepsilon}) = 1 - 2\mathbf{P}(R_t \neq R_{t+\varepsilon})$, and for any $t \geq 0$ and $\varepsilon > 0$, $\varepsilon^{-1}\mathbf{P}(R_t \neq R_{t+\varepsilon}) \leq \varepsilon^{-1}\mathbf{P}(\tau_i \leq \varepsilon) = \varepsilon^{-1}(1 - e^{-\varepsilon}) \leq 1$, so it satisfies $\mathbf{E}(R_t R_{t+\varepsilon}) \geq 1 - \lambda h^p$ for $p = 1$ and $\lambda = 2$ and therefore has a continuous modification.*

Definition. *An S.P. X_t has* right-continuous with left limits *(RCLL) sample paht if for a.e. ω, the path $X_t(\omega)$ is RCLL at any t, i.e. for $h \downarrow 0$ both $X_{t+h}(\omega) \to X_t(\omega)$ and the limit of $X_{t-h}(\omega)$ exists). Similarly, a modification having RCLL sample path with probability one is called an* RCLL modification *of the S.P.*

Definition. *Two S.P.s X_t and Y_t have* indistinguishable *sample path if $\mathbf{P}(X_t = Y_t$ for all $t) = 1$.*

Exercise. *If X_n, Y_n are discrete time S.P.s that are modifications of each other then they have indistinguishable sample path. If X_t, Y_t are continuous time S.P.s that are modifications of each other, and both have RCLL sample paths a.s., then they are indistinguishable. However, if you consider $Y_t(\omega) \equiv 0$ and $X_t(\omega) = I_{\{}\omega\}(t)$ (where $\omega \in [0, 1]$ uniformly) then X_t, Y_t are not indistinguishable.*

Proposition. *For sample paths of any S.P.: Hölder continuity $\Longrightarrow$ Continuous with probability one $\Longrightarrow$ RCLL.*

Theorem (Fubini's theorem). *If X_t has RCLL sample path and for some interval I and σ-field $\mathcal{H}$, almost surely $\int_I \mathbf{E}[|X_t|\,|\mathcal{H}]\,dt$ is finite then almost surely $\int_I X_t\,dt$ is finite and $\int_I \mathbf{E}[X_t|\mathcal{H}]\,dt = \mathbf{E}\left[\int_I X_t\,dt|\mathcal{H}\right]$.*

Chapter 4. Martingales and Stopping Times

4.1: Discrete time martingales and filtrations.

Definition. *A filtration is a non-decreasing family of sub-σ-fields $\{\mathcal{F}_n\}$ of our measurable space $(\Omega, \mathcal{F})$. That is, $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots \subseteq \mathcal{F}_n \cdots \subseteq \mathcal{F}$ and $\mathcal{F}_n$ is a σ-field for each n.*

Definition. *A S.P. $\{X_n, n = 0, 1, \ldots\}$ is adapted to a filtration $\{\mathcal{F}_n\}$ if $\omega \mapsto X_n)(\omega)$ is a R.V. on $(\Omega, \mathcal{F}_n)$ for each n, that is, if $\sigma(X_n) \subseteq \mathcal{F}_n$ for each n.*

Definition. *A filtration $\{\mathcal{G}_n\}$ with $\mathcal{G}_n = \sigma(X_0, X_1, \ldots, X_n)$ is the minimal filtration with respect to which $\{X_n\}$ is adapted. We therefore call it the canonical filtration for the S.P. $\{X_n\}$.*

Definition. *A martingale (denoted MG) is a pair $(X_n, \mathcal{F}_n)$ where $\{\mathcal{F}_n\}$ is a filtration and X_n is an integrable (i.e. $\mathbf{E}|X_n| < \infty$) S.P. adapted to this filtration such that $\mathbf{E}[X_{n+1}|\mathcal{F}_n] = X_n$ for rall n a.s.*

Proposition. *If $X_n = \sum_{i=1}^n D_i$ then the canonical filtration for $\{X_n\}$ is the same as the canonical filtration for $\{D_n\}$. Further, $(X_n, \mathcal{F}_n)$ is a martingale if and only if $\{D_n\}$ is an integrable S.P., adapted to $\{\mathcal{F}_n\}$, such that $\mathbf{E}(D_{n+1}|\mathcal{F}_n) = 0$ a.s. for all n.*

Definition. *We call a sequence $\{V_n\}$ previsible predictable for the filtration $\{\mathcal{F}_n\}$ if V_n is measureable on $\mathcal{F}_{n-1}$ for all $n \geq 1$.*

Theorem. *Let $(X_n, \mathcal{F}_n)$ be a MG and $\{V_n\}$ be a previsible sequence for the same filtration. The sequence of RV $Y_n = \sum_{k=1}^n V_k(X_k - X_{k-1})$, called the martingale transform of V with respect to X, is then a MG with respect to the filtration $\{\mathcal{F}_n\}$, provided $|V_n| \leq C_n$ for some non-random constants $C_n < \infty$, or more generally $\mathbf{E}|V_n|^q < \infty$ and $\mathbf{E}|X_n|^p < \infty$ for all n and some $1 \leq p, q < \infty$ such that $\frac{1}{q} + \frac{1}{p} = 1$.*

Remark. *The integrability conditions imposed in the above Theorem ensure that $\mathbf{E}|V_k||X_k| < \infty$, hence that the MG transform $\{Y_n\}$ is an integrable SP. Once this is established, $\{Y_n\}$ would be a MG, so we can state different versions of the theorem by further varying our integrability conditions.*

Definition. *A sub-martingale (denoted subMG) is an integrable SP $\{X_n\}$, adapted to the filtration $\{\mathcal{F}_n\}$, such that $\mathbf{E}[X_{n+1}|\mathcal{F}_n] \geq X_n$ for all n almost surely. A super-martingale (denoted supMG) is an integrable SP $\{X_n\}$, adapted to the filtration $\{\mathcal{F}_n\}$ such that $\mathbf{E}[X_{n+1}|\mathcal{F}_n] \leq X_n$ for all n almost surely.*

Remark. *Note that $\{X_n\}$ is a subMG iff $\{-X_n\}$ is a supMG, so they share many of the same properties.*

Remark. *If $\{X_n\}$ a subMG, then necessarily $n \mapsto \mathbf{E}X_n$ is non-decreasing, since by the tower property $\mathbf{E}[X_n] = \mathbf{E}[\mathbf{E}[X_n|\mathcal{F}_{n-1}]] \geq \mathbf{E}[X_{n-1}]$ for all $n \geq 1$.*

4.2: Continuous time martingales and right continuous filtrations.

Definition. *The pair $(X_t, \mathcal{F}_t)$, $t \geq 0$ real-valued, is called a continuous time martingale if*

- *The σ-fields $\mathcal{F}_t \subseteq \mathcal{F}$, $t \geq 0$, form a continuous time filtration, that is $\mathcal{F}_t \subseteq \mathcal{F}_{t+h}$ for all $t \geq 0$ and $h > 0$.*
- *The continuous time SP $\{X_t\}$ is integrable and adapted to this filtration. That is, $\mathbf{E}|X_t| < \infty$ and $\sigma(X_t) \subseteq \mathcal{F}_t$ for all $t \geq 0$.*
- *For any fixed $t \geq 0$ and $h > 0$, the identity $\mathbf{E}(X_{t+h}|\mathcal{F}_t) = X_t$ holds a.s.*

Replacing equality in c with $\geq$ and $\leq$ defines the continuous time subMG and supMG respectively. Similar to a previous remark, if $\{X_t\}$ is a subMG, then $\mathbf{E}[X_t] \geq \mathbf{E}[X_s]$ for all $t \geq s$ (and the reverse for supMG).

Remark. *Let $\sigma(X_s, 0 \le s \le t)$ denote the smallest σ-field containing $\sigma(X_s)$ for each $s \le t$. The canonical filtration for a continuous time SP $\{X_t\}$ is $\sigma(X_s, 0 \le s$ let). Further, if $(X_t, \mathcal{F}_t)$ is a MG, then $(X_t, \sigma(X_s, 0 \le s \le t))$ is also a MG.*

Proposition. *Any integrable SP $\{M_t\}$ of independent increments and constant mean (i.e. $\mathbf{E}[M_t] = \mathbf{E}[M_0]$), is a MG.*

Definition. *A filtration is called right-continuous if for any $t \ge 0$, $\cap_{h>0}\mathcal{F}_{t+h} = \mathcal{F}_t$. (We assume "usual conditions" with $N \in \mathcal{F}_)$ whenever $\mathbf{P}(N) = 0$.)*

Example. *Consider the uniform probability measure on $\Omega = \{-1, 1\}$ and $\mathcal{F} = 2^\Omega$. The process $X_t(\omega) = \omega t$ clearly has continuous sample path. It is easy to see that its canonical filtration has $\mathcal{G}_0 = \{\emptyset, \Omega\}$ while $\mathcal{G}_h = \mathcal{F}$ for all $h > 0$ and is evidently not right-continuous at $t = 0$.*

Theorem. *If $(X_t, \mathcal{F}_t)$ is a MG with a right-continuous filtration $\{\mathcal{F}_n\}$, then $\{X_t\}$ has a RCLL modification.*

4.3: Stopping times and the optional stopping theorem.

Definition. *A random variable τ taking values in $\{0, 1, \dots, n, \dots, \infty\}$ is a stopping time for the filtration $\{\mathcal{F}_n\}$ if the event $\{\tau \le n\}$ is in $\mathcal{F}_n$ for each finite $n \ge 0$.*

Definition. *Using the notation $n \wedge \tau = \min(n, \tau(\omega))$, the stopped at τ stochastic process $\{X_{n\wedge\tau}\}$ is given by $X_{n\wedge\tau} = X_n(\omega)$ if $n \le \tau(\omega)$ and $X_{n\wedge\tau} = X_{\tau(\omega)}(\omega)$ if $n > \tau(\omega)$.*

Theorem. *If $(X_n, \mathcal{F}_n)$ is a subMG (or supMG or a MG), and τ is a stopping time for $\{\mathcal{F}_n\}$, then $(X_{n\wedge\tau}, \mathcal{F}_n)$ is also a subMG, or subMG or MG, respectively.*

Corollary. *If $(X_n, \mathcal{F}_n)$ is a subMG and τ is a stopping time for $\{\mathcal{F}_n\}$, then $\mathbf{E}(X_{n\wedge\tau} \ge \mathbf{E}(X_0)$ for all n. If in addition $(X_n, \mathcal{F}_n)$ is a MG, then $\mathbf{E}[X_\tau] = \mathbf{E}[X_0]$.*

Theorem (Doob's optional stopping)**.** *If $(X_n, \mathcal{F}_n)$ is a subMG and $\tau < \infty$ a.s. is a stopping time for the filtration $\{\mathcal{F}_n\}$ such that the sequence $\{X_{n\wedge\tau}$ is uniformly integrable, then $\mathbf{E}(X_\tau) \ge \mathbf{E}(X_0)$. If in addition $(X_n, \mathcal{F}_n)$ is a MG, ten $\mathbf{E}[X_\tau] = \mathbf{E}[X_0]$.*

Definition. *A non-negative random variable $\tau(\omega)$ is called stopping time with respect to the continuous time filtration $\{\mathcal{F}_t\}$ if $\{\omega : \tau(\omega) \le t\} \in \mathcal{F}_t$ for all $t \ge 0$.*

Proposition. *If right-continuous SP $\{X_t\}$ is adapted to the filtration $\{\mathcal{F}_t\}$ then $\tau_B(\omega) = \inf\{t \ge 0 : X_t(\omega) \in B\}$ is a stopping time for $\mathcal{F}_t$ when either (a) B is an open set and $\mathcal{F}_t$ is a right continuous filtration, or (b) B is a closed set and the sample path $t \mapsto X_t(\omega)$ is continuous for all $\omega \in \Omega$.*

Example. *Consider the open set $B = (0, \infty)$ and the SP $X_t(\omega) = \omega t$ of continuous sample path and canonical filtration that is not right continuous. If is easy to check that $\tau_B(1) = 0$ and $\tau_B(-1) = \infty$. As the $\{\omega : \tau_B(\omega) \le 0\} = \{1\}$ is not in $\mathcal{G}_0 = \{\emptyset, \Omega\}$, we see that in this case, τ_B is not a stopping time for $\mathcal{G}_t$.*

Theorem. *If τ is a stopping time for the filtration $\{F_t\}$ and the SP $\{X_t\}$ of right-continuous sample path is a subMG (or supMG or MG) for $\{\mathcal{F}_t\}$, then $X_{t\wedge\tau} = X_{t\wedge\tau}(\omega)$ is also a subMG (or supMG or MG, respectively), for this filtration.*

Theorem (Doob's optional stopping)**.** *If $(X_t, \mathcal{F}_t)$ is a subMG with right-continuous sample path and $\tau < \infty$ a.s. is a stopping time for the filtration $\{\mathcal{F}_t\}$ such that $\{X_{t\wedge\tau}\}$ is uniformly integrable, then $\mathbf{E}(X_\tau) \ge \mathbf{E}(X_0)$. If in addition $(X_t, \mathcal{F}_t)$ is a MG, then $\mathbf{E}(X_\tau) = \mathbf{E}(X_0)$.*

Definition. *The stopped σ-field $\mathcal{F}_\tau$ associated with the stopping time τ for a filtration $\{\mathcal{F}_t\}$ is the collection of events $A \in \mathcal{F}$ such that $A \wedge \{\omega : \tau(\omega) \le t\} \in \mathcal{F}_t$ for each $t \ge 0$.*

4.4: Martingale representations and inequalities.

Theorem (Doob's decomposition)**.** *Given an integrable SP $\{X_n\}$, adapted to the discrete parameter filtration $\{\mathcal{F}_n\}$, $n \ge 0$, there exists a decomposition $X_n = Y_n + A_n$ such that $(Y_n, \mathcal{F}_n)$ is a MG and $\{A_n\}$ is a previsible SP. This decomposition is unique up to the value of Y_0, a RV measurable on $\mathcal{F}_0$.*

Theorem (Doob-Meyer decomposition)**.** *Suppose $\{\mathcal{F}_t\}$ is a right-continuous filtration and the martingale $(M_t, \mathcal{F}_t)$ of continuous sample path is such that $\mathbf{E}M_t^2 < \infty$ for each $t \ge 0$. Then, there exists a unique (integrable) SP such that*

- *$A_0 = 0$,*
- *$\{A_t\}$ has continuous sample path,*
- *$\{A_t\}$ is adapted to $\{\mathcal{F}_t\}$,*
- *$t \mapsto A_t$ is non-decreasing,*
- *$(M_t^2 - A_t, \mathcal{F}_t)$ is a MG.*

Definition. *A SP $\{A_t\}$ in the Doob-Meyer decomposition (of $\{M_t^2\}$) is called the increasing part or the increasing process associated with the MG $(M_t, \mathcal{F}_t)$.*

Remark. *If a square-integrable martingale of continuous sample path has a zero increasing path, then it is almost surely constant.*

Theorem (Doob's inequality)**.**

- *Suppose $\{X_n\}$ is a subMG. Then, for all $x > 0$ and $N < \infty$, $\mathbf{P}(\max_{0\le n\le N} X_n > x) \le x^{-1}\mathbf{E}|X_N|$.*
- *Suppose $\{X_n, n \le \infty\}$ is a subMG. Then, for all $x > 0$, $\mathbf{P}(\sup_{0\le n<\infty} X_n > x) \le x^{-1}\mathbf{E}|X_\infty|$.*
- *Suppose $\{X_t\}, t \in [0, T]$ is a continuous-parameter, right continuous subMG (that is, each sample path $t \mapsto X_t(\omega)$ is right continuous). Then, for all $x > 0$, $\mathbf{P}(\sup_{0\le t\le T} X_t > x) \le x^{-1}\mathbf{E}|X_T|$.*

Example. *Suppose $\{X_t\}$ is a right continuous subMG for $t \in [0, T]$ such that $\mathbf{E}[(X_t)_+^p] < \infty$ for some $p > 1$ and all $t \ge 0$. Then, for $q = p/(p-1)$, any $x > 0$ and $t \le T$, $\mathbf{P}(\sup_{0\le u\le t} X_u > x) \le x^{-p}\mathbf{E}[(X_t)_+^p]$ and $\mathbf{E}[(\sup_{0\le u\le t} X_u)_+^p] \le q^p\mathbf{E}[(X_t)_+^p]$, where $(y)_+^p$ denotes the function $(\max(y, 0))^p$.*

4.5: Martingale convergence theorems.

Theorem (Doob's convergence theorem)**.** *Suppose $(X_t, \mathcal{F}_t)$ is a right continuous subMG.*

- *If $\sup_{t\ge 0}\mathbf{E}[(X_t)_+] < \infty$, then $X_\infty = \lim_{t\to\infty} X_t$ exists w.p. 1. Further, in this case $\mathbf{E}|X_\infty| \le \lim_{t\to\infty}\mathbf{E}|X_t| < \infty$.*
- *If $\{X_t\}$ is uniformly integrable then $X_t \to X_\infty$ also in L^1. Further, the L^1 convergence $X_t \to X_\infty$ implies that $X_t \le \mathbf{E}(X_\infty|\mathcal{F}_t)$ for any fixed $t \ge 0$.*

Corollary. *If $(X_t, \mathcal{F}_t)$ is a right continuous MG and $\sup_t \mathbf{E}|X_t| < \infty$ then $X_\infty = \lim_{t\to\infty} X_t$ exists w.p.1 and is integrable. If $\{X_t\}$ is also U.I. then $X_t = \mathbf{E}(X_\infty|\mathcal{F}_t)$ for all t (such a martingale, namely $X_t = \mathbf{E}(X|\mathcal{F}_t)$ for an integrable RV X and a filtration $\{\mathcal{F}_t\}$, is called Doob's martingale of X with respect to $\{\mathcal{F}_t\}$).*

Proposition. *If the right continuous MG $\{Y_t\}$ is such that $\mathbf{E}Y_t^2 \le C$ for some $C < \infty$ and all $t \ge 0$, then there exists a RV Y_∞ such that $Y_t \to Y_\infty$ a.s. and in L^2. Moreover, $\mathbf{E}Y_\infty^2 \le C < \infty$ and the corresponding result holds in the context of discrete parameter MGs.*

Remark. *The above proposition does not have an L^1 analog. Namely, there exists a non-negative MG $\{Y_n\}$ such that $\mathbf{E}Y_n = 1$ for all n and $Y_n \to Y_\infty = 0$ a.s., so obviously Y_n does not converge to Y_∞ in L^1.*

4.6: Branching processes: extinction probabilities.

Definition. *The Branching process is a discrete time SP $\{Z_n\}$ taking non-negative integer values, such that $Z_0 = 1$ and for any $n \ge 1$, $Z_n = \sum_{j=1}^{Z_{n-1}} N_j^{(n)}$, where N and $N_j^{(n)}$ for $j = 1, 2, \dots$ are iid, non-negative integer valued RV with finite mean $m = \mathbf{E}(N) < \infty$, and where we use the convention that if $Z_{n-1} = 0$ then also $Z_n = 0$.*

Remark. *We use the filtration $\mathcal{F}_n = \sigma(\{N_j^{(k)}, k \le n, j = 1, 2, \dots\})$. (Quickly note that $\mathcal{G}_n = \sigma(\{Z_k, k \le n\})$ is a strict subset of $\mathcal{F}_n$.)*

Proposition. *The SP $X_n = m^{-n}Z_n$ is a martingale for the filtration $\mathcal{F}_n$.*

Definition. *$p_{ex} = \mathbf{P}(\{\omega : Z_n(\omega) = 0$ for all n large enough$\})$. p_{ex} is called the extinction probability.*

Proposition.

- *(Sub-critical) If $m < 1$, then $p_{ex} = 1$.*
- *(Critical) If $m = 1$ and $\mathbf{P}(N = 1) < 1$ then $p_{ex} = 1$.*

Proposition. *(Super-critical) If $m > 1$ and $\mathbf{P}(N = 0) = 0$, extinction is impossible. For other cases, to compute p_{ex}, consider the function $\varphi(p) = \mathbf{P}(N = 0) + \sum_{k=1}^\infty \mathbf{P}(N = k)p^k$. There exists a unique solution $\rho \in (0, 1)$ to the equation $p = \varphi(p)$. Upon verifying that ρ^{Z_n} is a martingale, we can show that $p_{ex} = \rho$. (Shown in exercise.)*

Chapter 5. The Brownian motion

5.1. Brownian motion: definition and construction.

Definition. *A stochastic process $(W_t, 0 \le t \le T)$ is called a* Brownian motion *(or Wiener Process if) (a) W_t is a Gaussian process, (b) $\mathbf{E}W_t = 0$, $\mathbf{E}(W_tW_s) = \min(t, s)$, and (c) For almost every ω, the sample path $t \mapsto W_t(\omega)$ is continuous on $[0, T]$.*

Remark. *Brownian motion is martingale and is not stationary but has stationary increments*

Proposition. *The Brownian motion has independent increments of zero mean. (proof: look at $\mathbf{E}[(W_{t+h} - W_t)W_s]$).*

Remark. *A Gaussian R.V. Y with $\mathbf{E}Y = 0$ and $\mathbf{E}Y^2 = \sigma^2$ has moments $\mathbf{E}(Y^{2n}) = \frac{(2n)!}{2^n n!}\sigma^{2n}$.*

Exercise. *Suppose W_t is a B.M. and $\alpha, s, T > 0$ are non-random constants. Then, the following are also Brownian motions: $\{-W_t, t \ge 0\}$, $\{W_{s+t} - W_s, t \ge 0\}$, $\{W_T - W_{T-t}, 0 \le t \le T\}$, $\{\sqrt{\alpha}W_{t/\alpha}, t \ge 0\}$, if $\tilde{W}_0 = 0$ and $\tilde{W}_t = tW_{1/t}$ then $\{\tilde{W}_t\}$ is B.M.*

Exercise. *Almost surely,* $t^{-1}W_t \to 0$ *for* $t \to \infty$.

Exercise. *Define the following: the* Brownian bridge $B_t = W_t - \min(t,1)W_{e^t}$, *the* Geometric Brownian motion $Y_t = e^{W_t}$, *the* Ornstein-Uhlenbeck process $U_t = e^{-t/2}W_{e^t}$, *and* $X_t = x + \mu t + \sigma W_t$ *a* Brownian motion with drift μ *and diffusion coefficient* $\sigma > 0$ *starting from* x. *Of these,*

(1) W_t, B_t, U_t, *and* X_t *are Gaussian, but* Y_t *is not* ($Y_1 > 0$).
(2) U_t *is Gaussian with constant mean and autocovariance* $\rho(t,s) = \rho(|t-s|)$ *so it is stationary. The others do not have constant mean or homogeneous-ish autocovariance so are not stationary.*
(3) All have continuous sample path.
(4) B_t *and* U_t *are not adapted to* $\sigma(W_s, s \le t)$. Y_t *and* X_t *are;* Y_t *is a submartingale for this filtration and* X_t *is if* $\mu \ge 0$ *(super if otherwise).*

Exercise. *For* $0 \le t \le 1$, *these S.P. have the same distribution as the Brownian bridge and have continuous modifications:* $\hat{B}_t = (1-t)W_{t/(1-t)}$ *for* $t < 1$ *with* $\hat{B}_1 = 0$, *and* $Z_t = tW_{1/t-1}$ *for* $t > 0$ *with* $Z_0 = 0$.

5.2 The reflection principle and Brownian hitting times.

Theorem (Levy's martingale characterization)**.** *Suppose a square-integrable MG* $(X_t, \mathcal{F}_t)$ *of right-continuous filtration and continuous sample path is such that* $(X_t^2 - t, \mathcal{F}_t)$ *is also a MG. Then,* X_t *is a Brownian motion. (note: continuity is essential, e.g.* $X_t = N_t - t$ *(Poisson-ish) satisfies the other properties)*

Proposition. *Suppose* $(X_t, \mathcal{F}_t)$ *is a square-integrable martingale with* $X_0 = 0$, *right-continuous filtration and continuous sample path. If the increasing part* A_t *in the corresponding Doob-Meyer decomposition is almost surely unbounded then* $W_s = X_{\tau_s}$ *is a Brownian motion, where* $\tau_s = \inf\{t \ge 0 : A_t > s\}$ *are* $\mathcal{F}_t$*-stopping times such that* $s \mapsto \tau_s$ *is a non-decreasing and right-continuous mapping of* $[0,\infty)$ *to* $[0,\infty)$ *with* $A_{\tau_s} = s$ *and* $X_t = W_{A_t}$.

Proposition. *If* τ *is a stopping time for the canonical filtration* $\mathcal{G}_t$ *of the Brownian motion* W_t *then the S.P.* $X_t = W_{t+\tau} - W_\tau$ *is also a Brownian motion, which is independent of the stopped* σ*-field* $\mathcal{G}_\tau$. *Note: this implies B.M. is a strong Markov process.*

Remark (Reflection principle)**.** *We use the last proposition to calculate the pdf of the* first hitting time $\tau_\alpha = \inf\{t > 0 : W_t = \alpha\}$. *Now,* $\{\omega : W_T(\omega) \ge \alpha\} \subseteq \{\omega : \max_{0\le s\le T} W_s(\omega)\} = \{\omega : \tau_\alpha(\omega) \le T\}$. *Recall that* $X_t = W_{t+\tau_\alpha} - W_{\tau_\alpha}$ *is a B.M. independent of* τ_α *(which is measurable on* $\mathcal{G}_{\tau_\alpha}$*). The law of* X_t *is invariant to a sign-chang so we have the* reflection principle *stating that* $\mathbf{P}(\max_{0\le s\le T} W_s \ge \alpha, W_T \ge \alpha) = \mathbf{P}(\tau_\alpha \le T, X_{T-\tau_\alpha} \ge 0) = \mathbf{P}(\tau_\alpha \le T, X_{T-\tau_\alpha} \le 0) = \mathbf{P}(\max_{0\le s\le T} W_s \ge \alpha, W_T \le \alpha)$. *Since* $\mathbf{P}(W_T = \alpha) = 0$ *we have* $\mathbf{P}(\max_{0\le s\le T} W_s \ge \alpha) = \mathbf{P}(\max_{0\le s\le T} W_s \ge \alpha, W_T \ge \alpha) + \mathbf{P}(\max_{0\le s\le T} W_s \ge \alpha, W_T \le \alpha) = 2\mathbf{P}(\max_{0\le s\le T} W_s \ge \alpha, W_T \ge \alpha) = 2\mathbf{P}(W_T \ge \alpha)$.

Remark. *Using a similar argument for the simple random walk* S_n *gets us* $\mathbf{P}(\max_{0\le k\le n} S_k \ge r) = 2\mathbf{P}(S_n > r) + \mathbf{P}(S_n = r)$.

Remark. *Let* $\tau_{\beta,\alpha} = \inf\{t : W_t \ge \alpha \text{ or } W_t \le \beta\}$. *Applying Doob's optional stopping theorem for the uniformly integrable stopped martingale* $W_{t\wedge\tau_{\beta,\alpha}}$ *of continuous sample path we get* $\mathbf{P}(W_{\tau_{\beta,\alpha}} = \alpha) = \beta/(\alpha+\beta)$.

Exercise. $\mathbf{E}(\tau_{\beta,\alpha}) = \alpha\beta$ *(apply Doob's optional stopping theorem on* $W^2_{t\wedge\tau_{\beta,\alpha}} - t\wedge\tau_{\beta,\alpha}$.

5.3. Smoothness and variation of the Brownian sample path.

Definition. *For any finite partition* π *of* $[a,b]$, *i.e.* $\pi = \{a = t_0^{(\pi)} < \cdots < t_k^{(\pi)} = b\}$. *Let* $\|\pi\| = \max_i\{t_{i+1}^{(\pi)} - t_i^{(\pi)}\}$ *denote the length of the longest interval in* π *and* $V_\pi^{(q)} = \sum_i \left|f(t_{i+1}^{(\pi)}) - f(t_i^{(\pi)})\right|^q$ *denote the q-th variation of* $f(\cdot)$ *on* π. *The q-th variation of* $f(\cdot)$ *on* $[a,b]$ *is then* $V^{(q)}(f) = \lim_{\|\pi\|\to 0} V_{(\pi)}^{(q)}(f)$ *provided it exists.*

Definition. *The q-th variation of a S.P.* X_t *on the interval* $[a,b]$ *is the random variable* $V^{(q)}(X)$ *obtained when replacing* $f(t)$ *by* $X_t(\omega)$ *in the above definition when the limit makes sense.*

Definition. *The* quadratic variation *of a stochastic process* X, *denoted* $V_t^{(2)}(X)$ *is the non-decreasing, non-negative S.P. corresponding to the quadratic variation of* X *on the intervals* $[0,t]$.

Proposition. *For a Brownian motion* $W(t)$, *as* $\|\pi\| \to 0$ *we have that* $V_{(\pi)}^{(2)}(W) \to (b-a)$ *in 2-mean.*

Corollary. *The quadratic variation of the Brownian motion is the S.P.* $V_t^{(2)}(W) = t$, *which is the same as the* increasing process *in the Doob-Meyer decomposition of* W_t^2. *More generally, the quadratic variation equals the increasing process for any square-integrable martingagle of continuous sample path and right-continuous filtration.*

Corollary. *With probability one, the sample path of the Brownian motion* $W(t)$ *is* not Lipschitz continuous $(|W(t) - W(s)| \le L|t-s|)$ *in any interval* $[a,b]$.

Exercise. *Fixing* $\gamma > 1/2$, *with probability one, the sample path of B.M. is not globally Holder continuous of exponent* γ *in* $[a,b]$.

Chapter 6. Markov, Poisson and Jump processes

6.1: Markov chains and processes.

Definition. *A discrete time stochastic process* $\{X_n, n = 0,1,\dots\}$ *with each RV* X_n *taking values in a measurable space* $(\mathbb{S},\mathcal{B})$ *is called a Markov chain if for every non-negative integer* n *and any set* $A \in \mathcal{B}$, *almost surely* $\mathbf{P}(X_{n+1} \in A|X_0,\dots,X_n) = \mathbf{P}(X_{n+1} \in A|X_n)$. *The set* $\mathbb{S}$ *is called the state space of the Markov chain.*

Remark. *This above definition is equivalent to the identity* $\mathbf{E}(f(X_{n+1})|X_1,\dots,X_n) = \mathbf{E}(f(X_{n+1})|X_n)$ *holding almost surely for each bounded measurable function* $f(\cdot)$.

Definition. *A homogeneous Markov chain is a Markov chain that has a modification for which* $\mathbf{P}(X_{n+1} \in A|X_n)$ *does not depend on* n *(except via the value of* X_n*).*

Definition. *To each homogeneous Markov chain* $\{X_n\}$ *with values in a closed subset* $\mathbb{S}$ *of* $\mathbb{R}$ *correspond its stationary transition probabilities* $p(A|x)$ *such that* $p(\cdot|x)$ *is a probability measure on* $(\mathbb{S},\mathcal{B})$ *for any* $x \in \mathbb{S}$; $p(A|\cdot)$ *is measurable on* $\mathcal{B}$ *for any* $A \in \mathcal{B}$, *and almost surely* $p(A|X_n) = \mathbf{P}(X_{n+1} \in A|X_n)$ *for all* $n \ge 0$.

Remark. *Many Markov chains are not martingales. For example, a sequence of independent variables* X_n *is a Markov chain, but unless* $\mathbf{P}(X_n = c) = 1$ *for some non-random* c *and all* n, *it is not a martingale. Similarly, many martingales do no have the Markov property. For example, the sequence* $X_n = X_0(1+S_n)$ *with* X_0 *uniformly chosen in* $\{1,3\}$ *independently of the simple random walk* S_n *of zero mean, is a martingale of zero-mean is a martingale, but not a Markov chain.*

Remark. *(Notation) Let* $\mathbf{P}_x$ *denote the law of the homogeneous Markov chain starting at* $X_0 = x$.

Definition. *The initial distribution of a Markov chain is the probability measure* $\pi(A) = P(X_0 \in A)$ *on* $(\mathbb{S},\mathcal{B})$.

Proposition. *(Strong Markov Property) Let* X_n *be a homogeneous Markov chain. Then,* $\mathbf{P}((X_\tau, X_{\tau+1},\dots,X_{\tau+k}) \in B|\mathcal{G}_\tau) = \mathbf{P}_{X_\tau}((X_0,\dots,X_k) \in B)$ *holds for any almost surely finite stopping time* τ *with respect to its canonical filtration* $\mathcal{G}_n$, *with* $\mathcal{G}_\tau$ *denoting the corresponding stopped* σ*-field.*

Definition. *A stochastic process* $X(t)$ *indexed by* $t \in [0,\infty)$ *and taking values in a measurable space* $(\mathbb{S},\mathcal{B})$ *is called a Markov process if for any* $t,u \ge 0$ *and* $A \in \mathcal{B}$ *we have that almost surely* $\mathbf{P}(X(t+u) \in A|\sigma(X(s), s \le t)) = \mathbf{P}(X(t+u) \in A|X(t))$. *Equivalently, we call* $X(t)$ *a Markov process if for any* $t,u \ge 0$ *and any bounded measurable function* $f(\cdot)$ *on* $(\mathbb{S},\mathcal{B})$, *almost surely,* $\mathbf{E}(f(X(t+u))|\sigma(X(s), s \le t)) = \mathbf{E}(f(X(t+u))|X(t))$. *The set* $\mathbb{S}$ *is called the state space of the Markov process.*

Definition. *For each* $t > s$ *and fixed* $s \in \mathbb{S}$ *there exists a probability measure* $p_{t,s}(\cdot|x)$ *on* $(\mathbb{S},\mathcal{B})$ *such that for each fixed* $A \in \mathcal{B}$, *the function* $p_{t,s}(A|\cdot)$ *is measurable and* $\mathbf{P}(X(t) \in A|X(s)) = \mathbf{E}(I_{X(t)\in A}|X(s)) = p_{t,s}(A|X(s))$ *almost surely. Such a collection* $p_{t,s}(A|x)$ *is called the transition probabilities for the Markov process* $\{X(t)\}$.

Definition. *We say that* $p_{t,s}(A|x)$ *for* $t > s \ge 0$, $x \in \mathbb{S}$ *and* $A \in \mathcal{B}$ *are regular transition probabilities if* $p_{t,s}(\cdot|x)$ *are probability measures on* $(\mathbb{S},\mathcal{B})$, *the functions* $p_{t,s}(A|\cdot)$ *are Borel measurable and Chapman-Kolmogorov equations* $p_{t,s}(A|x) = \int p_{t,u}(A|y)p_{u,s}(dy|x)$, *hold for every* $t > u > s \ge 0$, $x \in \mathbb{S}$ *and* $A \in \mathcal{B}$.

Theorem. *Given regular transition probabilities* $p_{t,s}(\cdot|\cdot)$ *and a probability measure* $\pi(\cdot)$ *on* $(\mathbb{S},\mathcal{B})$, *the identities* $\mathbf{P}(X(t_k) \in A_k,\dots,X(t_0) \in A_0) = \int_{A_0}\cdots\int_{A_k} p_{t_k,t_{k-1}}(dx_k|x_{k-1}) \cdot p_{t_1,t_0}(dx_1|x_0)\pi(dx_0)$ *define the finite dimensional distributions of a Markov process* $\{X(t)\}$ *having the specified transition probabilities and the initial distribution* π.

Definition. *A homogeneous Markov process is a Markov process with regular transition probabilities of the form* $p_{t,s}(\cdot|\cdot) = p_{t-s}(\cdot|\cdot)$, *which in turn are called the stationary regular transition probabilities.*

Proposition. *Suppose* $X(t)$ *is a homogeneous Markov process, with* $\mathcal{G}_t$ *denoting its canonical filtration* $\sigma(X(s), s \le t$ *and* $\mathbf{P}_x$ *the law of the process starting at* $X(0) = x$. *Then, any such process has the regular Markov property. That is,* $\mathbf{P}_x(X(\cdot+\tau) \in \Gamma|\mathcal{G}_\tau) = \mathbf{P}_{X(\tau)}(X(\cdot) \in \Gamma)$ *almost surely for any* $x \in \mathbb{R}$, *non-random* $\tau \ge 0$ *and* Γ *in the cylindrical* σ*-field* $\mathcal{B}^{[0,\infty)}$.

Exercise. *Let $\{X_t, t \geq 0\}$ be a Markov process of state state $\mathbb{S}$. Suppose $h_t : \mathbb{S} \to \mathbb{S}'$ are measurable and invertible for any fixed $t \geq 0$ and $g : [0,\infty) \to [0,\infty)$ is invertible and strictly increasing.*

- *Verify that $Y_t = h_t(X_{g(t)})$ is a Markov process.*
- *Show that if $\{X_t\}$ is a homogeneous Markov process then so if $\{h(X_t)\}$.*

Proposition. *Every continuous time stochastic process of independent increments is a Markov process. Further, every continuous time SP of stationary independent increments is a homogeneous Markov process.*

Remark. *To re-cap, we have seen three main ways of showing that a SP $\{X_t, t \geq 0\}$ is a Markov process:*

- *Computing $\mathbf{P}(X_{t+h} \in A|\mathcal{G}_t)$ directly and checking that is only depends on X_t (and not on X_s for $s < t$).*
- *Showing that the process has independent increments and applying the above proposition.*
- *Showing that it is an invertible function of another Markov process, and appealing to the above Exercise.*

Proposition. *If a Markov process or a Markov chain is also a stationary process, then it is a homogeneous Markov process, or Markov chain, respectively.*

Remark. *Note however that many homogeneous Markov processes and Markov chains are not stationary processes. Convince yourself that among such examples are the Brownian motion (in continuous time) and the random walk (in discrete time).*

Definition. *A homogeneous Markov process is called a strong Markov process if $\mathbf{P}_x(X(\cdot + \tau) \in \Gamma|\mathcal{G}_\tau) = \mathbf{P}_{X(\tau)}(X(\cdot) \in \Gamma)$ holds for any almost surely finite stopping time τ with respect to its canonical filtration $\mathcal{G}_\tau$.*

Corollary. *The Brownian motion is a strong Markov process.*

Example. *With X_0 independent of the Brownian motion W_t, consider the S.P. $X_t = X_0 + W_t I_{\{X_0 \neq 0\}}$ of continuous sample path. Noting that $I_{X_0} = 0 = I_{X_t=0}$ almost surely (as the difference occurs on the event $\{\omega : W_t(\omega) = -X_0(\omega) \neq 0\}$ which is of zero probability), by the independence of increments of W_t, hence of X_t in case $X_0 \not 0$, we have that almost surely $\mathbf{P}(X_{t+u} \in A|\sigma(X_s, s \leq t)) = I_{0\in A} I_{X_0=0} + \mathbf{P}(W_{t+u} - W_t + X_t \in A|X_t) I_{X_0\neq 0} = I_{0\in A} I_{X_t=0} + p_t(A|X_t) I_{X_t\neq 0}$ for $p_t(A|x)$ the stationary regular transition probabilities of B.M., i.e. $p_t(A|x) = \int_A \frac{e^{-(y-x)^2/2t}}{\sqrt{2\pi t}} dy$. X_t is a homogeneous Markov process and satisfies the regular Markov property, but it is not strong Markov; consider the a.s. finite s.t. $\tau = \inf\{t \geq 0 : X_t = 0\}$ and $\Gamma = \{x(\cdot) : x(1) > 0\}$ (left side is $\frac{1}{2}\mathbf{1}_{x\neq 0}$ while right side is zero).*

Proposition. *The Markov property (same equation as the definition above) holds for any stopping time τ (with respect to the canonical filtration of the homogeneous Markov process $\{X(t)\}$), provided τ assumes at most a countable number of non-random values.*

6.2: Poisson process, Exponential inter-arrivals and order statistics.

Condition. $\boxed{C_0}$. *Each sample path $N_t(\omega)$ is piecewise constant, nondecreasing, right continuous, with $N_0(\omega) = 0$, all jump discontinuities are of size one, and there are infinitely many of them.*

Associated with each sample path $N_t(\omega)$ satisfying $\boxed{C_0}$ are the jump times $0 = T_0 < T_1 < \ldots$ with $T_k = \inf\{t \geq 0 : N_t \geq k\}$ for each k, or equivalently $N_t = \sup\{k \geq 0 : T_k \leq t\}$.

Recall that N has the Poisson(μ) law if $P(N = k) = \frac{\mu^k}{k!} e^{-\mu}$.

Condition. $\boxed{C_1}$. *For any k and any $0 < t_1 < \cdots < t_k$, the increments $N_{t_1}, N_{t_2} - N_{t_1}, \ldots, N_{t_k} - N_{t_{k-1}}$ are independent random variables and fro some $\lambda > 0$ and all $t > s \geq 0$ the increment $N_t - N_s$ has the Poisson($\lambda(t-s)$) law.*

Definition. *Among the processes satisfying $\boxed{C_0}$ the Poisson process is the unique one satisfying $\boxed{C_1}$.*

Corollary. *$M_t = N_t - \lambda t$ is a MG. Also, it is square-integrable and $M_t^2 - \lambda t$ is a martingale for the (right-continuous) filtration $\sigma(N_s, s \leq t)$.*

Proposition. *The Poisson processes are the only stochastic processes with stationary independent increments. Also, the Poisson process is a homogeneous Markov process; it has stationary regular transition probabilities $p_t(x + k|x) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, k, x \geq 0$. Also the Poisson process is a strong Markov process.*

Proposition (Memoryless property of the Exponential law)**.** *We say that a r.v. T has Exponential(λ) law if $\mathbf{P}(T > t) = e^{-\lambda t}$. Except for $T = 0$ w.p.1 these are the only laws for which $\mathbf{P}(T > x + y|T > y) = \mathbf{P}(T > x)$ for all $x, y \geq 0$.*

Condition. $\boxed{C_2}$. *The gaps between jump times $T_k - T_{k-1}$ are i.i.d. r.v.s each of Exponential(λ) law.*

Proposition. *A stochastic process N_t that satisfies $\boxed{C_0}$ is a Poisson process of rate λ iff it satisfies $\boxed{C_2}$.*

Remark. *Obviously the sample path of the Poisson process are never continuous. However, $\mathbf{P}(N_{t+h} - N_t \geq 1) = 1 - e^{-\lambda h} \to 0$ as $h \downarrow 0$, so $\mathbf{P}(T_k = t) = 0$ so there are no fixed discontinuities (i.e. occuring at non-random times).*

Proposition. *Fixing positive t and a positive integer n let U_i be i.i.d. random variables, each uniform on $[0, t]$ and consider their order statisticsc U_i^*. That is, permute the order of U_i so that $U_1^* \leq \cdots \leq U_n^*$. The joint distribution of $(U_1^*, \ldots, U_n^*)$ is precisely that of the first n arrival times $(T_1, \ldots, T_n)$ of a Poisson process, conditional on the event $N_t = n$.*

Math 136 - Stochastic Processes

Homework Set 1, Autumn 2013. Due: October 2

1. Exercise 1.1.3. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and A, B, A_i events in $\mathcal{F}$. Prove the following properties of $\mathbb{P}$.

   (a) *Monotonicity.* If $A \subseteq B$ then $\mathbb{P}(A) \le \mathbb{P}(B)$.
   ANS: $A \subseteq B$ implies that $B = A \cup (B \setminus A)$. Hence, $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A)$. Thus since $\mathbb{P}(B \setminus A) \ge 0$, we get $\mathbb{P}(A) \le \mathbb{P}(B)$.

   (b) *Subadditivity.* If $A \subseteq \cup_i A_i$ then $\mathbb{P}(A) \le \sum_i \mathbb{P}(A_i)$.
   ANS: For each i set $B_i = A_i \setminus \bigcup_{j=1}^{i-1} A_j$. Then the B_i are disjoint and we let $C = \bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} B_i$. Since $A \subseteq C$, from part (a), $\mathbb{P}(A) \le \mathbb{P}(C)$. Also, $\mathbb{P}(C) = \sum_{i=1}^{\infty} \mathbb{P}(B_i)$ and $B_i \subseteq A_i$ therefore $\mathbb{P}(B_i) \le \mathbb{P}(A_i)$ so $\mathbb{P}(C) \le \sum_{i=1}^{\infty} \mathbb{P}(A_i)$ and hence $\mathbb{P}(A) \le \sum_{i=1}^{\infty} \mathbb{P}(A_i)$.

   (c) *Continuity from below:* If $A_i \uparrow A$, that is, $A_1 \subseteq A_2 \subseteq \ldots$ and $\cup_i A_i = A$, then $\mathbb{P}(A_i) \uparrow \mathbb{P}(A)$.
   ANS: Construct the disjoint sets $B_1 = A_1$ and $B_i = A_i \setminus A_{i-1}$ for $i \ge 2$, noting that $A_i = \cup_{j \le i} B_j$ and $A = \cup_j B_j$. Therefore, $\mathbb{P}(A_i) = \sum_{j=1}^{i} \mathbb{P}(B_j) \uparrow \sum_{j=1}^{\infty} \mathbb{P}(B_j) = \mathbb{P}(\cup_j B_j) = \mathbb{P}(A)$.

   (d) *Continuity from above:* If $A_i \downarrow A$, that is, $A_1 \supseteq A_2 \supseteq \ldots$ and $\cap_i A_i = A$, then $\mathbb{P}(A_i) \downarrow \mathbb{P}(A)$.
   ANS: Apply part (c) to the sets $A_i^c \uparrow A^c$ to have that $1 - \mathbb{P}(A_i) = \mathbb{P}(A_i^c) \uparrow \mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.

   (e) *Inclusion-exclusion rule:*
   $$\mathbb{P}(\bigcup_{i=1}^{n} A_i) = \sum_i \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) + \sum_{i<j<k} \mathbb{P}(A_i \cap A_j \cap A_k) - \cdots + (-1)^{n+1} \mathbb{P}(A_1 \cap \cdots \cap A_n).$$
   ANS: The proof is by induction on n. The case where $n = 1$ is immediate. For $n = 2$, we observe
   $$\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1 \cup [A_2 \setminus (A_1 \cap A_2)]) = \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2).$$
   Suppose the result holds for some $n \ge 2$. Applying the result to the two sets $\cup_{i=1}^{n} A_i$ and A_{n+1}, we see
   $$\mathbb{P}(A_1 \cup \cdots \cup A_{n+1}) = \mathbb{P}(A_1 \cup \cdots \cup A_n) + \mathbb{P}(A_{n+1}) - \mathbb{P}((A_1 \cap A_{n+1}) \cup \cdots \cup (A_n \cap A_{n+1})).$$
   Inclusion-exclusion for $n+1$ now follows by applying the case for n to the first and last probabilities on the right hand side and rearranging.

2. Exercise 1.1.9. Verify the alternative definitions of the Borel σ-field $\mathcal{B}$:
$$\begin{aligned}\sigma(\{(a,b) : a < b \in \mathbb{R}\}) &= \sigma(\{[a,b] : a < b \in \mathbb{R}\}) = \sigma(\{(-\infty, b] : b \in \mathbb{R}\}) \\ &= \sigma(\{(-\infty, b] : b \in \mathcal{Q}\}) = \sigma(\{O \subseteq \mathbb{R} \text{ open }\})\end{aligned}$$
Hint: Any $O \subseteq \mathbb{R}$ open is a countable union of sets (a, b) for $a, b \in \mathcal{Q}$ (rational).
ANS: Let $\sigma_1 = \sigma(\{(a,b) : a < b \in \mathbb{R}\})$, $\sigma_2 = \sigma(\{[a,b] : a < b \in \mathbb{R}\})$, $\sigma_3 = \sigma(\{(-\infty, b] : b \in \mathbb{R}\})$, $\sigma_4 = \sigma(\{(-\infty, b] : b \in \mathcal{Q}\})$ and $\sigma_5 = \sigma(\{O \subseteq \mathbb{R} \text{ open }\})$, be the five σ-fields appearing in the problem. Recall that if a collection of sets $\mathcal{A}$ is a subset of a σ-field $\mathcal{F}$, then also $\sigma(\mathcal{A}) \subseteq \mathcal{F}$. For this reason we have that $\sigma_1 \subseteq \sigma_5$ and defining $\sigma_0 = \sigma(\{(a,b) : a < b \in \mathcal{Q}\})$, we have for same reason that $\sigma_0 \subseteq \sigma_1$. By the hint provided we see that any open set O is a countable union of sets in σ_0, hence also in σ_0. Therefore, $\sigma_5 \subseteq \sigma_0$, forcing in view of the above $\sigma_0 = \sigma_1 = \sigma_5$. Since $(-\infty, a]$ is the countable union of $[b - i, b]$, $i = 1, 2, \ldots$, it follows that $(-\infty, b] \in \sigma_2$ for any $b \in \mathbb{R}$, hence $\sigma_4 \subseteq \sigma_3 \subseteq \sigma_2$. Since each set $[a, b]$ can be expressed as the countable intersection $\cap_{i=1}^{\infty}(a - 1/i, b + 1/i)$, we see that $\sigma_2 \subseteq \sigma_1$. Further, since $[b, \infty)$ is the countable intersection of the complements of $(-\infty, b - 1/i]$, $i = 1, 2, \ldots$, it follows that $[b, \infty) \in \sigma_4$ for $b \in \mathcal{Q}$, hence (a, b) which is the complement of the union of $(-\infty, a]$ and $[b, \infty)$ is in σ_4 when $a, b \in \mathcal{Q}$, resulting with $\sigma_0 \subseteq \sigma_4$. Recall we have shown that $\sigma_0 = \sigma_1 = \sigma_5$ and just now saw that $\sigma_0 \subseteq \sigma_4 \subseteq \sigma_3 \subseteq \sigma_2 \subseteq \sigma_1$, implying all six σ-fields considered are the same.

3. Exercise 1.1.12 Check that the following are Borel sets and find the probability assigned to each by the uniform measure from Example 1.1.11: $(0, 1/2) \cup (1/2, 3/2)$, $\{1/2\}$, a countable subset A of $\mathbb{R}$, the set of irrational numbers in $(0, 1)$, $[0, 1]$, and $\mathbb{R}$.
ANS: $(0, 1/2) \cup (1/2, 3/2)$ is open and hence Borel. By countable additivity,
$$U((0, 1/2) \cup (1/2, 3/2)) = U((0, 1/2)) + U((1/2, 3/2)) = 1/2 + 1/2 = 1.$$
The singleton $\{1/2\}$ is closed and hence Borel. There are two easy ways to see that $U(\{1/2\}) = 0$. First, fixing $\epsilon > 0$ arbitrary, we see that
$$U(\{1/2\}) \le U((1/2 - \epsilon/2, 1/2 + \epsilon/2)) = \epsilon.$$
Second,
$$1 = U((0,1)) = U((0, 1/2) \cup (1/2, 1) \cup \{1/2\}) = 1/2 + 1/2 + U(\{1/2\}).$$
If $A \subseteq \mathbb{R}$ is countable then we can write $A = \cup_{n=1}^{\infty}\{a_n\}$ for $a_n \in \mathbb{R}$. Since each $\{a_n\}$ is closed, A is a countable union of closed sets and hence Borel. Either $a_n \in (0, 1)$ or $a_n \notin (0, 1)$. In the former case we

can argue as before to get that $U(\{a_n\}) = 0$ and in the latter case that $U(\{a_n\}) = 0$ is trivial. Hence by countable subadditivity,
$$U(A) \le \sum_{n=1}^{\infty} U(\{a_n\}) = 0.$$
Let J denote the set of rationals in $(0, 1)$. Then J is countable and hence Borel with $U(J) = 0$. The set I of irrationals in $(0, 1)$ is thus Borel since we can write $I = (0, 1) \setminus J$. We have,
$$U(I) = U((0,1) \setminus J) = U((0,1)) - U(J) = 1.$$
The set $[0, 1]$ is Borel since it is closed. We have,
$$U([0,1]) = U((0,1)) + U(\{0\}) + U(\{1\}) = 1.$$
Finally, the set of reals $\mathbb{R}$ is Borel since it is open. We have,
$$U(\mathbb{R}) = U(\mathbb{R} \cap (0,1)) = U((0,1)) = 1.$$

4. Exercise 1.2.5. Let $\Omega = \{1, 2, 3\}$. Find a σ-field $\mathcal{F}$ such that $(\Omega, \mathcal{F})$ is a measurable space, and a mapping X from Ω to $\mathbb{R}$, such that X is not a random variable on $(\Omega, \mathcal{F})$.
ANS: Let $\mathcal{F} = \sigma(\{1,2,3\}) = \{\{1,2,3\}, \emptyset\}$ be the trivial σ-field. Together $(\Omega, \mathcal{F})$ form a measurable space. Let $X(\omega) = \omega$ where $\omega \in \Omega$. Then $\{\omega : X(\omega) \le 1\} = \{1\} \notin \mathcal{F}$, so X is not a random variable.

5. Exercise 1.2.18 Provide an example of a measurable space, a R.V. on it, and:

   (a) A function $g(x) \not\equiv x$ such that $\sigma(g(X)) = \sigma(X)$.
   ANS: Take $\Omega = \mathbb{R}$, $\mathcal{B}$ the Borel sets on $\mathbb{R}$, $X(x) = x$, and $g(x) = -x$. Then $\sigma(X) = \sigma(-X) = \mathcal{B}$.

   (b) A function f such that $\sigma(f(X))$ is strictly smaller than $\sigma(X)$ and is not the trivial σ-field $\{\emptyset, \Omega\}$.
   ANS: Take $\Omega, \mathcal{B}$, and X as before and set $f(x) = 1_{(0,1)}(x)$. Then $\sigma(X) = \mathcal{B}$ but $\sigma(f(X)) = \sigma((0,1)) = \{\emptyset, \mathbb{R}, (0,1), (0,1)^c\} \neq \mathcal{B}$.

6. Exercise 1.2.40. Show that if $\mathbf{E}[X^2] = 0$ then $X = 0$ almost surely.
ANS: For $n \in \mathbb{N}$, let $A_n = \{|X| > 1/n\}$. Note that $\{X \neq 0\} = \cup_n A_n$. Hence by countable subadditivity it suffices to show that $\mathbb{P}(A_n) = 0$ for all n. This follows immediately by applying Markov's inequality (Theorem 1.2.38) to the function $f(x) = x^2$:
$$\mathbb{P}(A_n) \le n^2 \mathbf{E}[X^2] = 0.$$

Math 136 - Stochastic Processes

Homework Set 2, Autumn 2013, Due: October 9

1. Exercise 1.3.14. Suppose that T_n are independent Exponential(1) random variables (that is, $\mathbf{P}(T_n > t) = e^{-t} 1_{\{t \ge 0\}}$).

   (a) Using both Borel-Cantelli lemmas, show that
   $$\mathbf{P}(T_k(\omega) > \alpha \log k \text{ for infinitely many values of } k) = 1_{\alpha \le 1}.$$
   ANS: Let
   $$A_k = A_k(\alpha) = \{T_k > \alpha \log k\}.$$
   Our aim is to show that $\mathbf{P}(A_k \text{ i.o}) = 1_{\alpha \le 1}$. We have,
   $$\sum_{k=1}^{\infty} \mathbf{P}(A_k) = \sum_{k=1}^{\infty} e^{-k \log \alpha} = \sum_{k=1}^{\infty} k^{-\alpha}.$$
   If $\alpha > 1$ this series is convergent, hence by the first Borel-Cantelli lemma (Lemma 1.3.10), $\mathbf{P}(A_k \text{ i.o}) = 0$. If $\alpha \le 1$ this series is divergent. Thus since the events $\{A_k\}$ are independent, the second Borel-Cantelli lemma (Lemma 1.3.11) implies $\mathbf{P}(A_k \text{ i.o.}) = 1$.

   (b) Deduce that $\limsup_{n \to \infty}(T_n / \log n) = 1$ almost surely.
   ANS: Note that,
   $$\begin{aligned} 1_{\alpha \le 1} &= \mathbf{P}(T_k > \alpha \log k \text{ i.o.}) \le \mathbf{P}(\limsup_{k \to \infty} T_k / \log k \ge \alpha) \le \mathbf{P}(\cap_{m=1}^{\infty}\{T_k > (\alpha - 1/m) \log k \text{ i.o}\}) \\ &= \lim_{m \to \infty} 1_{\alpha - 1/m \le 1} = 1_{\alpha \le 1}. \end{aligned}$$
   Therefore,
   $$\mathbf{P}(\limsup_{k \to \infty} T_k / \log k \ge \alpha) = 1_{\alpha \le 1},$$
   from which the desired conclusion is immediate.

2. Exercise 1.3.21. Fixing $q \ge 1$, use the triangle inequality for the norm $\|\cdot\|_q$ on L^q to show that if $X_n \stackrel{q.m.}{\to} X$, then $\mathbf{E}|X_n|^q \to \mathbf{E}|X|^q$. Using Jensen's inequality for $g(x) = |x|$, deduce that also $\mathbf{E}X_n \to \mathbf{E}X$. Finally, provide an example to show that $\mathbf{E}X_n \to \mathbf{E}X$ does not necessarily imply $X_n \to X$ in L^1.
ANS: By the triangle inequality $\|X_n - X + X\|_q \le \|X_n - X\|_q + \|X\|_q$ and rearranging terms we also have that $-\|X_n - X\|_q \le \|X_n\|_q - \|X\|_q \le \|X_n - X\|_q$. If $X_n \to X$ in L^q then $\|X_n - X\|_q = [\mathbf{E}(|X_n - X|^q)]^{1/q} \to 0$, so by the above, $\lim_{n\to\infty}(\|X_n\|_q - \|X\|_q) = 0$. Now, we just got that $\lim_{n\to\infty}[\mathbf{E}(|X_n|^q)]^{1/q} = [\mathbf{E}(|X|^q)]^{1/q}$, hence also $\lim_{n\to\infty} \mathbf{E}(|X_n|^q) = \mathbf{E}(|X|^q)$. By Corollary 1.3.19, we have $X_n \stackrel{L^1}{\to} X$, i.e.,

$\mathbf{E}|X_n - X| \to 0$. Using Jensen's inequality for $g(x) = |x|$, we have $|\mathbf{E}(X_n - X)| \le \mathbf{E}|X_n - X|$. Thus, $|\mathbf{E}(X_n - X)| \to 0$, or equivalently, $\mathbf{E}X_n \to \mathbf{E}X$. Let $X \equiv 0$ and
$$X_n = \begin{cases} 1 & \text{with probability } 1/2 \\ -1 & \text{with probability } 1/2 \end{cases}$$
Then $\mathbf{E}X_n = \mathbf{E}X = 0$ while $\mathbf{E}|X_n - X| = 1$ for all n.

3. Exercise 1.4.2. For a R.V. defined on $(\Omega, \mathcal{F}, \mathbf{P})$ verify that $\mathcal{P}_X$ is a probability measure on $(\mathbb{R}, \mathcal{B})$.
Hint: First show that for $B_i \in \mathcal{B}$, $\{\omega : X(\omega) \in \cup_i B_i\} = \cup_i \{\omega : X(\omega) \in B_i\}$ and that if the B_i are disjoint then so are the sets $\{\omega : X(\omega) \in B_i\}$.
ANS: We'll first justify the two statements to which the hint refers. Note that $\omega_0 \in \{\omega : X(\omega) \in \cup_i B_i\}$ iff $X(\omega_0) \in \cup_i B_i$ iff $X(\omega_0) \in B_i$ for some i iff $\omega_0 \in \cup_i\{\omega : X(\omega) \in B_i\}$. This proves the equality $\{\omega : X(\omega) \in \cup_i B_i\} = \cup_i\{\omega : X(\omega) \in B_i\}$. Suppose that the sets $B_i \in \mathcal{B}$ are disjoint. If $i \neq j$, $\omega_0 \in \{\omega : X(\omega) \in B_i\} \cap \{\omega : X(\omega) \in B_j\}$ iff $X(\omega_0) \in B_i \cap B_j = \emptyset$. Therefore $\{\omega : X(\omega) \in B_i\} \cap \{\omega : X(\omega) \in B_j\} = \emptyset$.
Using these two facts it is now easy to show that $\mathcal{P}_X$ is a probability measure on $(\mathbb{R}, \mathcal{B})$. Indeed, it is completely obvious that $0 \le \mathcal{P}_X(A) \le 1$ for all $A \in \mathcal{B}$ since $\mathbf{P}$ is itself a probability. Furthermore,
$$\mathcal{P}_X(\mathbb{R}) = \mathbf{P}(\omega : X(\omega) \in \mathbb{R}) = \mathbf{P}(\Omega) = 1.$$
Finally, suppose B_i is a countable collection of pairwise disjoint subsets of $\mathcal{B}$. Using the hint and the countable additivity of $\mathbf{P}$,
$$\mathcal{P}_X(\cup_i B_i) = \mathbf{P}(\omega : X_i(\omega) \in \cup_i B_i) = \mathbf{P}(\cup_i\{\omega : X_i(\omega) \in B_i\}) = \sum_i \mathbf{P}(\omega : X_i(\omega) \in B_i) = \sum_i \mathcal{P}_X(B_i).$$

4. Exercise 1.4.14. Let $M_n = \max_{1 \le i \le n}\{T_i\}$, where T_i, $i = 1, 2, \ldots$ are independent Exponential(λ) random variables (i.e. $F_{T_i}(t) = 1 - e^{-\lambda t}$ for some $\lambda > 0$, all $t \ge 0$ and any i). Find non-random numbers a_n and a non-zero random variable M_∞ such that $(M_n - a_n)$ converges in law to M_∞.
Hint: Explain why $F_{M_n - a_n}(t) = (1 - e^{-\lambda t} e^{-\lambda a_n})^n$ and find $a_n \to \infty$ for which $(1 - e^{-\lambda t} e^{-\lambda a_n})^n$ converges per fixed t and its limit is strictly between 0 and 1.
ANS: Let $a_n = \lambda^{-1} \log n$ and let the distribution function of M_∞ be $F_{M_\infty}(x) = \exp(-e^{-\lambda x})$ (this function is monotone increasing from 0 to 1 and differentiable everywhere, hence a distribution function of a R.V. with density). Indeed, since M_n is the maximum of n I.I.D. random variables T_i, each of which having the distribution function $F_{T_i}(t) = 1 - e^{-\lambda t}$ for $t \in [0, \infty)$, we have that
$$\mathbf{P}(M_n \le \lambda^{-1} \log n + x) = \prod_{i=1}^{n} \mathbf{P}(T_i \le \lambda^{-1} \log n + x) = (1 - n^{-1} e^{-\lambda x})^n,$$

for all $x \ge -\lambda^{-1} \log n$. Fixing any real-valued x, in the limit $n \to \infty$ we thus get that
$$\mathbf{P}(M_n \le \lambda^{-1} \log n + x) \to \exp(-e^{-\lambda x}).$$
This amounts to $(M_n - a_n)$ converging in law to M_∞.

5. Exercise 1.4.17.

   (a) Give an example of random variables X and Y on the same probability space, such that $\mathcal{P}_X = \mathcal{P}_Y$ while $\mathbf{P}(\{\omega : X(\omega) \neq Y(\omega)\}) = 1$.
   ANS: Take $(\Omega, \mathcal{F}) = (\{A, B\}, 2^{\Omega})$ and $\mathbf{P}$ defined by $\mathbf{P}(\{A\}) = \mathbf{P}(\{B\}) = 1/2$. Let $X, Y : \Omega \to \{0, 1\}$ be defined by $X(A) = Y(B) = 0$ and $X(B) = Y(A) = 1$. Then X and Y have the same distribution, hence the same law. But they are never equal, i.e. the set on which they are not equal is Ω itself, which has probability 1.

   (b) Give an example of random variables $X_n \stackrel{\mathcal{L}}{\to} X_\infty$ where each X_n has a probability density function, but X_∞ does not have such.
   ANS: Take the probability space of Example 1.1.11 and let $X_n(\omega) = \omega/n$ and $X_\infty \equiv 0$. Then $X_n \stackrel{a.s.}{\to} X_\infty$, which implies that $X_n \stackrel{\mathcal{L}}{\to} X_\infty$. Yet, each X_n has a p.d.f. (namely $f_{X_n}(x) = n\mathbf{1}_{[0,1/n]}$), while X_∞ does not, since F_{X_∞} is not continuous at 0 (criterion from Proposition 1.4.8 in the notes).

   (c) Suppose Z_p denotes a random variable with a Geometric distribution of parameter $p > 0$, that is $\mathbf{P}(Z_p = k) = p(1-p)^{k-1}$ for $k = 1, 2, \ldots$. Show that $\mathbf{P}(pZ_p > t) \to e^{-t}$ as $p \to 0$, for each $t \ge 0$ and deduce that pZ_p converge in law to the *Exponential* random variable T, whose density is $f_T(t) = e^{-t}\mathbf{1}_{t \ge 0}$.
   ANS: We calculate, evaluating a geometric series at the second equality
   $$\mathbf{P}(Z_p > t/p) = \sum_{k=[t/p]}^{\infty} p(1-p)^k = p(1-p)^{[t/p]} \frac{1}{1 - (1-p)} = (1-p)^{[t/p]}$$
   Now $(1-p)^{t/p-1} \le (1-p)^{[t/p]} \le (1-p)^{t/p}$ and $(1-p)^{1/p} \to e^{-1}$ as $p \to 0$ (a result from a course in analysis). Hence both the RHS and the LHS of the above inequality tends to e^{-t} and we get $\mathbf{P}(pZ_p > t) \to e^{-t}$ for all t as desired. Now this implies $F_{pZ_p}(t) \to F_T(t)$ for all t, which is our definition of convergence in law of pZ_p to T.

   (d) Suppose R.V.-s X_n and X_∞ have (Borel measurable) densities $f_n(s)$ and $f_\infty(s)$, respectively, such that $f_n(s) \to f_\infty(s)$ as $n \to \infty$, for each fixed $s \in \mathbb{R}$ and further that f_∞ is strictly positive on $\mathbb{R}$. Let $g_n(s) = 2\max(0, 1 - f_n(s)/f_\infty(s))$. Explain why (recall Definition 1.2.23)
   $$\int_{\mathbb{R}} |f_n(s) - f_\infty(s)| ds = \int_{\mathbb{R}} g_n(s) f_\infty(s) ds,$$

   why it follows from Corollary 1.4.29 that $\int_{\mathbb{R}} g_n(s) f_\infty(s) ds \to 0$ as $n \to \infty$ and how you deduce from this that $X_n \stackrel{\mathcal{L}}{\to} X_\infty$.
   ANS: First note that since the total integral of a p.d.f. is always 1, we have $\int_{\{s: f_n(s) < f_\infty(s)\}} (f_\infty(s) - f_n(s)) ds = \int_{\{s: f_n(s) \ge f_\infty(s)\}} (-f_\infty(s) + f_n(s)) ds$. This gives the second equality in the following computation while the first equality comes from g_n being zero on the set $\{s : f_n(s) \ge f_\infty(s)\}$ and $2(1 - f_n(s)/f_\infty(s))$ on its complement
   $$\int_{\mathbb{R}} g_n(s) f(s) ds = 2 \int_{\{s: f_n(s) < f_\infty(s)\}} (f_\infty(s) - f_n(s)) ds = \int_{\mathbb{R}} |f_\infty(s) - f_n(s)| ds$$
   Now define a random variable $Y_n(s) = g_n(s)$ on the probability space $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, Q)$ with Q defined by $Q(B) = \int_B f_\infty(s) ds$ for $B \in \mathcal{B}$. Note that $|Y_n| \le 2$ and that for all $s \in \mathbb{R}$ we have $Y_n(s) \to 0$ as $n \to \infty$ (since $f_n(s) \to f_\infty(s)$), hence the Bounded Convergence Theorem applies (Corollary 1.4.29) and we can conclude that $\mathbf{E}_Q Y_n = \int_{\mathbb{R}} g_n(s) f_\infty(s) ds \to 0$. Finally, let h be a continuous and bounded ($|h| < K$ for some $K < \infty$) function on $\mathbb{R}$. Then $\mathbf{E}h(X_n) - \mathbf{E}h(X_\infty) = \int_{\mathbb{R}} h(s)(f_n(s) - f_\infty(s)) ds$ and taking the absolute value
   $$|\mathbf{E}h(X_n) - \mathbf{E}h(X_\infty)| \le \int_{\mathbb{R}} |h(s)||f_n(s) - f_\infty(s)| ds \le K \int_{\mathbb{R}} |f_n(s) - f_\infty(s)| ds = K \int_{\mathbb{R}} g_n(s) f_\infty(s) ds \to 0$$
   as $n \to \infty$. Now Proposition 1.4.11 implies $X_n \stackrel{\mathcal{L}}{\to} X_\infty$.

6. Exercise 1.4.30. Use Monotone Convergence to show that
$$\mathbf{E}(\sum_{n=1}^{\infty} Y_n) = \sum_{n=1}^{\infty} \mathbf{E}Y_n,$$
for any sequence of non-negative R.V. Y_n. Deduce that if $X \ge 0$ and A_n are disjoint sets with $\mathbf{P}(\cup_n A_n) = 1$, then
$$\mathbf{E}(X) = \sum_{n=1}^{\infty} \mathbf{E}(X I_{A_n}).$$
Further, show that this applies also for any $X \in L^1$.
ANS: For each m let $X_m = \sum_{n=1}^{m} Y_n$. Since the Y_n are non-negative it follows that $\{X_m\}$ is a non-negative non-decreasing sequence with (possibly infinite) limit $\sum_{n=1}^{\infty} Y_n$. Hence by monotone convergence (Theorem 1.4.29) and the linearity of the expectation,
$$\mathbf{E}(\sum_{n=1}^{\infty} Y_n) = \mathbf{E}(\lim_{m\to\infty} X_m) = \lim_{m\to\infty} \mathbf{E}(X_m) = \lim_{m\to\infty} \left(\sum_{n=1}^{m} \mathbf{E}(Y_n)\right) = \sum_{n=1}^{\infty} \mathbf{E}(Y_n).$$
Suppose that $X \ge 0$ and A_n are disjoint with $\mathbf{P}(\cup_n A_n) = 1$. Then the random variables $Y_n = X I_{A_n} \ge 0$ satisfy the criterion of the first part of the problem. Using that $\mathbf{P}(\cup_n A_n) = 1$, we have
$$\mathbf{E}(X) = \mathbf{E}(X I_{\cup_n A_n}) = \mathbf{E}(X \sum_{n=1}^{\infty} I_{A_n}) = \mathbf{E}(\sum_{n=1}^{\infty} X I_{A_n}) = \sum_{n=1}^{\infty} \mathbf{E}(X I_{A_n}).$$

Finally, suppose $X \in L^1$. Let $X_+ = \max(X, 0)$ and $X_- = -\min(X, 0) = \max(-X, 0)$ denote the positive and negative parts of X, respectively. Applying the previous part to the non-negative random variables X_+ and X_-, we get
$$\mathbf{E}X = \mathbf{E}X_+ - \mathbf{E}X_- = \sum_{n=1}^{\infty} \mathbf{E}X_+ I_{A_n} - \sum_{n=1}^{\infty} \mathbf{E}X_- I_{A_n} = \sum_{n=1}^{\infty} \mathbf{E}(X_+ - X_-) I_{A_n} = \sum_{n=1}^{\infty} \mathbf{E}X I_{A_n}.$$
Note that this could also have been accomplished just as easily by applying dominated convergence to the sequence $X_n = \sum_{k=1}^{n} X I_{A_k}$ (with $|X_n| \le |X|$ for all n).

Math 136 - Stochastic Processes

Homework Set 3, Autumn 2013, Due: October 16

1. Exercise 1.4.31. Prove Proposition 1.4.3 using the following steps.

   (a) Verify that the identity (1.4.1) holds for indicator functions $g(x) = I_B(x)$ for $B \in \mathcal{B}$.
   ANS: Let $B \in \mathcal{B}$ be an arbitrary Borel set and let $g(x) = I_B(x)$. Note that $I_B(X) = I_{\{X \in B\}}$. Hence,
   $$\mathbf{E}(g(X)) = \mathbf{E}(I_B(X)) = \mathbf{E}(I_{\{X \in B\}}) = \mathcal{P}_X(B) = \int_{\mathbb{R}} I_B(x) d\mathcal{P}_X(x) = \int_{\mathbb{R}} g(x) d\mathcal{P}_X(x).$$
   Therefore the desired result holds for indicators.

   (b) Using the linearity of the expectation, check that this identity holds whenever $g(x)$ is a (non-negative) simple function on $(\mathbb{R}, \mathcal{B})$.
   ANS: Let $g(x)$ be a non-negative simple function. Then there exists constants $c_1, \ldots, c_n \ge 0$ and Borel sets $B_1, \ldots, B_n$ such that
   $$g(x) = \sum_{i=1}^{n} c_i I_{B_i}(x).$$
   Hence, by the linearity of the expectation and the integral (which denotes an expectation of $g(x)$ on $(\mathbb{R}, \mathcal{B}, \mathcal{P}_X)$), we have
   $$\begin{aligned}\mathbf{E}[g(X)] &= \mathbf{E}\left[\sum_{i=1}^{n} c_i I_{B_i}(X)\right] = \sum_{i=1}^{n} c_i \mathbf{E}[I_{B_i}(X)] = \sum_{i=1}^{n} c_i \int_{\mathbb{R}} I_{B_i}(x) d\mathcal{P}_X(x) \text{ (by part (a))} \\ &= \int_{\mathbb{R}} \sum_{i=1}^{n} c_i I_{B_i}(x) d\mathcal{P}_X(x) = \int_{\mathbb{R}} g(x) d\mathcal{P}_X(x).\end{aligned}$$

   (c) Combine the definition of the expectation via the identity (1.2.2) with Monotone Convergence to deduce that (1.4.1) is valid for any non-negative Borel function $g(x)$.
   ANS: Let $g(x)$ be a non-negative Borel function. Then there exists a sequence $\{g_n\}$ of simple functions such that $g_1 \ge 0$, $g_n \le g_{n+1}$, and $g_n(x) \uparrow g(x)$ as $n \to \infty$ (for example, take $g_n(x) = f_n(g(x))$ for $f_n(\cdot)$ of Proposition 1.2.6). Hence,
   $$\begin{aligned}\mathbf{E}[g(X)] &= \lim_n \mathbf{E}[g_n(X)] \text{ (Monotone Convergence for } g_n(X(\omega))) \\ &= \lim_n \int_{\mathbb{R}} g_n(x) d\mathcal{P}_X(x) \text{ (part (b))} \\ &= \int_{\mathbb{R}} g(x) d\mathcal{P}_X(x) \text{ (Monotone Convergence for } g_n(x))\end{aligned}$$

(d) Recall that $g(x)=g_+(x)-g_-(x)$ for $g_+(x)=\max(g(x),0)$ and $g_-(x)=-\min(g(x),0)$ non-negative Borel functions. Thus, using Definition 1.2.25 conclude that (1.4.1) holds whenever $\mathbf{E}|g(X)|<\infty$.
ANS: Let g be an arbitrary Borel function and let g_+ and g_- be the corresponding positive and negative parts of g. By part (c), we have both

$$\mathbf{E}[g_+(X)]=\int_{\mathbb{R}} g_+(x)d\mathcal{P}_X(x) \text{ and } \mathbf{E}[g_-(X)]=\int_{\mathbb{R}} g_-(x)d\mathcal{P}_X(x).$$

Hence if

$$\mathbf{E}[g_+(X)]+\mathbf{E}[g_-(X)]=\mathbf{E}|g(X)|<\infty,$$

then $\mathbf{E}[g_+(X)]<\infty$ and $\mathbf{E}[g_-(X)]<\infty$. In particular, their difference makes sense. So, by linearity of the expectation and part (c),

$$\begin{aligned}\mathbf{E}[g(X)]&=\mathbf{E}[g_+(X)-g_-(X)]=\mathbf{E}[g_+(X)]-\mathbf{E}[g_-(X)]\\&=\int_{\mathbb{R}} g_+(x)d\mathcal{P}_X(x)-\int_{\mathbb{R}} g_-(x)d\mathcal{P}_X(x) \qquad \text{(part (c))}\\&=\int_{\mathbb{R}}(g_+(x)-g_-(x))d\mathcal{P}_X(x)=\int_{\mathbb{R}} g(x)d\mathcal{P}_X(x).\end{aligned}$$

2. Exercise 1.4.33. Suppose a R.V. W on a probability space $(\Omega,\mathcal{F},\mathbf{P})$ has the $N(\mu,1)$ law of Definition 1.2.30.

(a) Check that $Z=\exp(-\mu W+\mu^2/2)$ is a positive random variable with $\mathbf{E}Z=1$.
ANS: Since $x\mapsto\exp(x)$ is positive whenever $x\in\mathbb{R}$ and normal random variables are finite a.s. it immediately follows that $\mathbf{P}(Z>0)=1$. Note that the random variable $-\mu W+\mu^2/2$ has distribution $N(-\mu^2/2,\mu^2)$. Hence by Exercise 1.2.31,

$$\mathbf{E}Z=\exp(-(\mu^2/2)+(\mu^2/2))=1.$$

As an aside, we comment that there are a number of different ways to justify Exercise 1.2.31, the most elementary of which is to compute the expectation directly using Proposition 1.2.29 and "completing the square" in the exponential term.

(b) Show that under the corresponding equivalent probability measure $\widetilde{\mathbf{P}}$ of Exercise 1.4.32 the R.V. W has the $N(0,1)$ law.
ANS: Fixing $t\in\mathbb{R}$, we compute

$$\begin{aligned}\widetilde{\mathbf{P}}(W\le t)&=\mathbf{E}ZI_{\{W\le t\}}=\mathbf{E}\exp(-\mu W+\mu^2/2)I_{\{W\le t\}}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^t\exp(-\mu x+\mu^2/2)\exp(-(x-\mu)^2/2)dx\\&=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^t\exp(-x^2/2)dx,\end{aligned}$$

which by Definition 1.2.30 and Proposition 1.4.8 is precisely the value at t of the distribution function of a $N(0,1)$ random variable.

3. Exercise 2.1.4. Let $\Omega=\{a,b,c,d\}$, with event space $\mathcal{F}=2^\Omega$ and let $\mathbf{P}$ be a probability measure on $\mathcal{F}$ such that $\mathbf{P}(\{a\})=1/2$, $\mathbf{P}(\{b\})=1/4$, $\mathbf{P}(\{c\})=1/6$ and $\mathbf{P}(\{d\})=1/12$.

(a) Find $\sigma(I_A)$, $\sigma(I_B)$ and $\sigma(I_A,I_B)$ for subsets $A=\{a,d\}$ and $B=\{b,c,d\}$ of Ω.
ANS: Visibly, $\sigma(I_A)=\{\emptyset,A,A^c,\Omega\}$. Likewise, $\sigma(I_B)=\{\emptyset,B,B^c,\Omega\}$. In our case, $A\cap B=\{d\}$ and $A\cap B^c=\{a\}$ are both in $\sigma(I_A,I_B)$. It is not hard to check that

$$\sigma(I_A,I_B)=\sigma(\{a\},\{d\})=\{\{a\},\{d\},\{a,d\},\{b,c\},\{a,b,c\},\{b,c,d\},\{a,b,c,d\},\emptyset\}.$$

(b) Let $\mathcal{H}=L^2(\Omega,\sigma(I_B),\mathbf{P})$. Find the conditional expectation $\mathbf{E}(I_A|I_B)$ and the value of $\mathbf{d}^2=\inf\{\mathbf{E}[(I_A-W)^2]:W\in\mathcal{H}\}$.
ANS: Any R.V. measurable on $\sigma(I_B)$ is of the form $Z=\alpha I_B+\beta I_{B^c}$ for some non-random α and β. Thus, the same applies for $Z=\mathbf{E}(I_A|I_B)$. Using the definition (and characterization) of conditional expectation we can directly "solve" for α and β to see that

$$\alpha=\mathbf{E}(I_AI_B)/\mathbf{E}(I_B) \text{ and } \beta=\mathbf{E}(I_AI_{B^c})/\mathbf{E}(I_{B^c}).$$

We can readily evaluate these expressions to arrive at

$$\alpha=\mathbf{P}(\{d\})/(\mathbf{P}(\{b\})+\mathbf{P}(\{c\})+\mathbf{P}(\{d\}))=(1/12)/(1/2)=1/6.$$

Similarly, $\beta=1$, so $\mathbf{E}(I_A|I_B)=(1/6)I_B+I_{B^c}$. By Definition 2.1.3, $\mathbf{d}^2=\mathbf{E}[V^2]$ for $V=I_A-(1/6)I_B-I_{B^c}$. It is not hard to check that $V(a)=0$, $V(b)=V(c)=-1/6$ and $V(d)=5/6$, leading to

$$\begin{aligned}\mathbf{d}^2&=V(a)^2\mathbf{P}(\{a\})+V(b)^2\mathbf{P}(\{b\})+V(c)^2\mathbf{P}(\{c\})+V(d)^2\mathbf{P}(\{d\})\\&=\tfrac{1}{36}\cdot\tfrac14+\tfrac{1}{36}\cdot\tfrac16+\tfrac{25}{36}\cdot\tfrac{1}{12}=\tfrac{5}{72}.\end{aligned}$$

4. Exercise 2.3.3. Let $\mathcal{F}_0=\{\Omega,\emptyset\}$. Show that if $Z\in L^1(\Omega,\mathcal{F}_0,\mathbf{P})$ then Z is necessarily a non-random constant and deduce that $\mathbf{E}(X|\mathcal{F}_0)=\mathbf{E}X$ for any $X\in L^1(\Omega,\mathcal{F},\mathbf{P})$.
ANS: It is immediate from the definition of measurability that any $\mathcal{F}_0$-measurable random variable is constant. Indeed, suppose that $X(\omega_0)=\alpha$ for some $\omega_0\in\Omega$ and $\alpha\in\mathbb{R}$. Then $\{\omega:X(\omega)=\alpha\}\neq\emptyset$ and therefore $\{\omega:X(\omega)=\alpha\}=\Omega$.

Obviously $\mathbf{E}X$ which is non-random is measurable on $\mathcal{F}_0$. By definition of C.E. suffices to show that $\mathbf{E}[XI_A]=\mathbf{E}[\mathbf{E}[X]I_A]=\mathbf{E}[X]\mathbf{P}(A)$ for any $A\in\mathcal{F}_0$, that is for $A=\Omega$ and for $A=\emptyset$. Both are trivial since $I_\Omega=1$ and $I_\emptyset=0$ for all ω and $\mathbf{P}(\Omega)=1$ while $\mathbf{P}(\emptyset)=0$.

5. Exercise 2.3.6. Give an example of a R.V. X and two σ-fields $\mathcal{F}_1$ and $\mathcal{F}_2$ on $\Omega=\{a,b,c\}$ in which

$$\mathbf{E}(\mathbf{E}(X|\mathcal{F}_1)|\mathcal{F}_2)\neq\mathbf{E}(\mathbf{E}(X|\mathcal{F}_2)|\mathcal{F}_1).$$

ANS: Take $\Omega=\{a,b,c\}$ and $\mathbf{P}(a)=\mathbf{P}(b)=\mathbf{P}(c)=1/3$. Let $X=I_{\{b,c\}}(\omega)$, which is measurable on $\mathcal{F}_1=\{\Omega,\emptyset,\{a\},\{b,c\}\}$, so $\mathbf{E}(X|\mathcal{F}_1)=X$. Let $\mathcal{F}_2=\{\Omega,\emptyset,\{a,b\},\{c\}\}$, and note that $\mathbf{E}(X|\mathcal{F}_2)=Y=I_{\{c\}}(\omega)+\frac12I_{\{a,b\}}(\omega)$. Since $Y=\mathbf{E}(\mathbf{E}(X|\mathcal{F}_1)|\mathcal{F}_2)$ is not measurable on $\mathcal{F}_1$, necessarily $Y\neq\mathbf{E}(Y|\mathcal{F}_1)$.

6. Exercise 2.3.16. Let $Z=(X,Y)$ be a uniformly chosen point on $(0,1)^2$. That is, X and Y are independent random variables, each having the $U(0,1)$ measure of Example 1.1.11. Set $T=I_A(Z)+5I_B(Z)$ where $A=\{0<x<1/4,3/4<y<1\}$ and $B=\{3/4<x<1,0<y<1/2\}$.

(a) Find an explicit formula for the conditional expectation $W=\mathbf{E}(T|X)$ and use it to determine the conditional expectation $U=\mathbf{E}(TX|X)$.
ANS: Note $A=A_1\times A_2$ for $A_1=\{x\in(0,1/4)\}$, $A_2=\{y\in(3/4,1)\}$ hence $I_A(x,y)=I_{A_1}(x)I_{A_2}(y)$. Similarly $I_B(x,y)=I_{B_1}(x)I_{B_2}(y)$ for $B_1=\{x\in(3/4,1)\}$, $B_2=\{y\in(0,1/2)\}$. Consequently, $T=I_{A_1}(X)I_{A_2}(Y)+5I_{B_1}(X)I_{B_2}(Y)$. Thus, by the linearity of the C.E. and "taking out what is known" (Proposition 2.3.15) we have that

$$W=\mathbf{E}(T|X)=I_{A_1}(X)\mathbf{E}(I_{A_2}(Y)|X)+5I_{B_1}(X)\mathbf{E}(I_{B_2}(Y)|X).$$

Further, since X and Y are independent, $I_{A_2}(Y)$ and $I_{B_2}(Y)$ are independent of X. Thus, we have that

$$\mathbf{E}(I_{A_2}(Y)|X)=\mathbf{E}I_{A_2}(Y)=\mathbf{P}(Y\in A_2)=\frac14,$$

with the right-most identity due to Y being uniformly chosen on $(0,1)$ with A_2 an interval of length 1/4. Similarly, $\mathbf{E}(I_{B_2}(Y)|X)=1/2$, so we have that

$$W=\frac14I_{A_1}(X)+\frac52I_{B_1}(X).$$

Since X is bounded, we know that $U=\mathbf{E}(TX|X)=XW$ by Proposition 2.3.15.

(b) Find the value of $\mathbf{E}((T-W)\sin(e^X))$.
ANS: Since $\sin(e^X)\in L^2(\Omega,\sigma(X),\mathbf{P})=\mathcal{H}_X$ and $W=\mathbf{E}(T|X)$ for T square-integrable, this is zero by Proposition 2.1.2.

(c) Without any computation decide whether $\mathbf{E}W^2-\mathbf{E}T^2$ is negative, zero, or positive. Explain your answer.
ANS: Recall Proposition 2.1.2 that $\mathbf{E}((T-W)W)=0$ (since $W\in\mathcal{H}_X$). Hence, with $T=W+(T-W)$ we have that

$$\mathbf{E}T^2=\mathbf{E}W^2+2\mathbf{E}(T-W)W+\mathbf{E}(T-W)^2=\mathbf{E}W^2+\mathbf{E}(T-W)^2.$$

By part (a) we already know that $\mathbf{P}(T\neq W)>0$, hence $\mathbf{E}(T-W)^2>0$, implying that $\mathbf{E}W^2-\mathbf{E}T^2$ is negative. We end by remarking that a perhaps shorter and more geometric argument can be made by recalling that the conditional expectation W of a square integrable T given $\mathcal{G}=\sigma(X)$ is just an orthogonal projection in a Hilbert space; T not being measurable with respect to $\mathcal{G}$ is equivalent this projection being strictly norm-reducing.

Math 136 - Stochastic Processes
Homework Set 4, Autumn 2013, Due: October 23

1. Exercise 2.3.19. Suppose that X and Y are square integrable random variables.

(a) Show that if $\mathbf{E}(X|Y)=\mathbf{E}(X)$ then X and Y are uncorrelated.
ANS: By the tower property and "taking out what is known,"

$$\mathbf{E}(XY)=\mathbf{E}(\mathbf{E}(XY|Y))=\mathbf{E}(\mathbf{E}(X|Y)Y)=\mathbf{E}(\mathbf{E}(X)Y)=\mathbf{E}(X)\mathbf{E}(Y).$$

(b) Provide and example of uncorrelated X and Y for which $\mathbf{E}(X|Y)\neq\mathbf{E}(X)$.
ANS: Suppose that Y is a standard normal random variable and $X=Y^2$. Then,

$$\mathbf{E}(XY)=\mathbf{E}Y^3=0=\mathbf{E}X\mathbf{E}Y,$$

but $\mathbf{E}(X|Y)=X$ and $\mathbf{P}(X\neq\mathbf{E}(X))>0$ (so $\mathbf{E}(X|Y)\neq\mathbf{E}(X)$).

(c) Provide and example where $\mathbf{E}(X|Y)=\mathbf{E}(X)$ but X and Y are not independent (this is also an example of uncorrelated but not independent R.V.).
ANS: Suppose that S takes the values 1 and -1 each with probability 1/2, Y a standard normal random variable independent of S, and $X=SY$. Then,

$$\mathbf{E}(X|Y)=\mathbf{E}(SY|Y)=\mathbf{E}(S|Y)Y=(\mathbf{E}S)Y=0=\mathbf{E}X.$$

Obviously, X and Y are not independent since $|X|=|Y|$.

2. Exercise 2.4.6.

(a) Suppose that the joint law of (X,Y,Z) has a density. Express the R.C.P.D. of Y given X,Z in terms of this density.
ANS: Let $f_{X,Y,Z}(x,y,z)$ denote the joint density of (X,Y,Z). Then the R.C.P.D. of Y given X,Z has the density $f_{Y|X,Z}(y|X(\omega),Z(\omega))$, where

$$f_{Y|X,Z}(y|x,z)=\frac{f_{X,Y,Z}(x,y,z)}{f_{X,Z}(x,z)}$$

and $f_{X,Z}(x,z)=\int_{\mathbb{R}}f_{X,Y,Z}(x,v,z)dv$.

(b) Using this expression, show that if X is independent of the pair (Y,Z) then

$$\mathbf{E}(Y|X,Z)=\mathbf{E}(Y|Z).$$

ANS: If X is independent of (Y,Z) then $f_{X,Y,Z}(x,y,z)=f_X(x)f_{Y,Z}(y,z)$ for all x,y,z. It follows that $f_{X,Z}(x,z)=f_X(x)f_Z(z)$ and so by Definition 2.4.4 we have similarly to Example 2.4.5 that

$$\begin{aligned}\mathbf{E}(Y|X,Z)&=\int_{\mathbb{R}}yf_{Y|X,Z}(y|X,Z)dy=\int_{\mathbb{R}}y\frac{f_{X,Y,Z}(X,y,Z)}{f_{X,Z}(X,Z)}dy\\&=\int_{\mathbb{R}}y\frac{f_X(X)f_{Y,Z}(y,Z)}{f_X(X)f_Z(Z)}dy=\int_{\mathbb{R}}y\frac{f_{Y,Z}(y,Z)}{f_Z(Z)}dy=\int_{\mathbb{R}}yf_{Y|Z}(y|Z)dy=\mathbf{E}(Y|Z).\end{aligned}$$

(c) Give an example of random variables X,Y,Z, such that X is independent of Y and

$$\mathbf{E}(Y|X,Z)\neq\mathbf{E}(Y|Z).$$

ANS: Let X and Y be independent $N(0,1)$ random variables and $Z=X+Y$. Note that $Y=Z-X$ is measurable on $\sigma(X,Z)$ hence $\mathbf{E}(Y|X,Z)=Y$ (see Example 2.3.2). In contrast, $\mathbf{E}(Y|Z)$ is by definition measurable on $\sigma(Z)$ whereas $Y=Z-X$ is not (can't be expressed as a non-random function of Z). Consequently, $Y\neq\mathbf{E}(Y|Z)$.
Alternatively, elementary computation with densities shows that $\mathbf{E}[Y|Z]=Z/2\neq Y$. Indeed, $X=Z-Y$ so

$$f_{Y,Z}(y,z)=f_Y(y)f_X(z-y)=\frac{1}{2\pi}\exp\left[-\frac{y^2+(y-z)^2}{2}\right]=\frac{1}{2\pi}\exp\left[-\frac{2y^2-2yz+z^2}{2}\right].$$

Further, $Z\sim N(0,2)$ with $f_Z(z)=1/\sqrt{4\pi}\exp(-z^2/4)$, resulting with

$$\begin{aligned}\mathbf{E}[Y|Z]&=\int_{\mathbb{R}}y\frac{f_{Y,Z}(y,Z)}{f_Z(Z)}dy=\frac{\sqrt2}{\sqrt{2\pi}}\int_{\mathbb{R}}y\exp\left[-\frac{2y^2-2yZ+Z^2/2}{2}\right]dy\\&=\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}}(\sqrt2y)\exp\left[-\frac{(\sqrt2y-Z/\sqrt2)^2}{2}\right]dy\\&=Z/2.\end{aligned}$$

3. Exercise 3.1.12. To practice your understanding you should at this point check that the processes X_t and Y_t of Example 3.1.11 are versions of each other but are not modifications of each other.
ANS: For any $t\ge0$,

$$\begin{aligned}Y_t(\omega)=1-X_t(\omega)&=\mathbf{1}_{[0,1)}(t)I_T(\omega_1)+\mathbf{1}_{[1,2)}(t)I_T(\omega_2)\\&\overset{\mathcal{L}}{=}\mathbf{1}_{[0,1)}(t)I_H(\omega_1)+\mathbf{1}_{[1,2)}(t)I_H(\omega_2)=X_t(\omega)\end{aligned}$$

Similarly, we have for any $n<\infty$ and $0\le t_1<t_2<\ldots<t_n<2$,

$$\mathbf{P}(\omega:X_{t_1}(\omega)\le\alpha_1,\ldots,X_{t_n}(\omega)\le\alpha_n)=\mathbf{P}(\omega:Y_{t_1}(\omega)\le\alpha_1,\ldots,Y_{t_n(\omega)}\le\alpha_n)$$

However, for any $t\ge0$ we have that $X_t+Y_t=1$ so $\mathbf{P}(X_t=Y_t)=0$.

4. Exercise 3.2.12. Let X be a Gaussian R.V. independent of S, with $\mathbf{E}(X)=0$ and $\mathbf{P}(S=1)=\mathbf{P}(S=-1)=1/2$.

(a) Check that SX is Gaussian.
ANS: Note that $\mathbf{E}[e^{i\theta SX}]=\frac12(\mathbf{E}[e^{i\theta X}]+\mathbf{E}[e^{-i\theta X}])=e^{-\theta^2\sigma^2/2}=\mathbf{E}[e^{i\theta X}]$ and recall the remark below Definition 3.2.8.

(b) Give an example of uncorrelated, zero-mean, Gaussian R.V. X_1 and X_2 such that the vector $\underline{X}=(X_1,X_2)$ is not Gaussian and where X_1 and X_2 are not independent.
ANS: Consider the Gaussian variables $X_1=X$ and $X_2=SX$. Then, $\mathbf{E}[X_2]=\mathbf{E}[S]\mathbf{E}[X_1]=0$ since $\mathbf{E}[X_1]=0$ and $\mathbf{E}[X_1X_2]=\mathbf{E}[SX^2]=\mathbf{E}[S]\mathbf{E}[X^2]=0$. That is, X_1 and X_2 are zero-mean and uncorrelated. Fixing $a>0$ note that $\mathbf{P}(X\ge a)>0$ and

$$\mathbf{P}(S=1)=\frac12=\mathbf{P}(X\ge0)>\mathbf{P}(X\ge a)=\mathbf{P}(SX\ge a)$$

(since SX has the same zero-mean Gaussian law as X). Therefore,

$$\mathbf{P}(SX\ge a,X\ge a)=\mathbf{P}(S=1,X\ge a)=\mathbf{P}(S=1)\mathbf{P}(X\ge a)>\mathbf{P}(SX\ge a)\mathbf{P}(X\ge a),$$

and in particular, SX and X are not independent. But, if (SX,X) is a Gaussian random vector then by Proposition 3.2.14 SX and X must also be independent. Thus, we deduce that (SX,X) is not a Gaussian random vector.

5. Exercise 3.2.13. Suppose (X,Y) has a bivariate Normal distribution (per Definition 3.2.8) with mean vector $\underline{\mu}=(\mu_X,\mu_Y)$ and the covariance matrix $\Sigma=\begin{pmatrix}\sigma_X^2&\rho\sigma_X\sigma_Y\\\rho\sigma_X\sigma_Y&\sigma_Y^2\end{pmatrix}$, with $\sigma_X,\sigma_Y>0$ and $|\rho|\le1$.

(a) Show that (X,Y) has the same law as $(\mu_X+\sigma_X\rho U+\sigma_X\sqrt{1-\rho^2}V,\mu_Y+\sigma_YU)$, where U and V are independent Normal R.V.-s of mean zero and variance one. Explain why this implies that $Z=X-(\rho\sigma_X/\sigma_Y)Y$ is independent of Y.
ANS: Since (U,V) has a bivariate Normal distribution, so does its linear transformation $(\widetilde{X},\widetilde{Y})$, where $\widetilde{X}=\mu_X+\sigma_X\rho U+\sigma_X\sqrt{1-\rho^2}V$ and $\widetilde{Y}=\mu_Y+\sigma_YU$ (see Proposition 3.2.16). To show that (X,Y) has the same law as $(\widetilde{X},\widetilde{Y})$, it then suffices to show that they have the same mean vector and covariance matrix. It is obvious that $(\widetilde{X},\widetilde{Y})$ has mean vector $\underline{\mu}=(\mu_X,\mu_Y)$ and its covariance matrix equals Σ because:

$$\mathbf{E}(\widetilde{X}-\mu_X)^2=\sigma_X^2\rho^2\mathbf{E}(U^2)+2\sigma_X^2\rho\sqrt{1-\rho^2}\mathbf{E}(UV)+\sigma_X^2(1-\rho^2)\mathbf{E}(V^2)=\sigma_X^2\rho^2+\sigma_X^2(1-\rho^2)=\sigma_X^2\,,$$

$$\mathbf{E}(\widetilde{Y}-\mu_Y)^2=\sigma_Y^2\mathbf{E}(U^2)=\sigma_Y^2\,,$$

$$\mathbf{E}[(\widetilde{X}-\mu_X)(\widetilde{Y}-\mu_Y)]=\rho\sigma_X\sigma_Y\mathbf{E}(U^2)+\sigma_X\sigma_Y\sqrt{1-\rho^2}\mathbf{E}(UV)=\rho\sigma_X\sigma_Y\,.$$

Next note that the independence of U and V implies that

$$\widetilde{Z}=\widetilde{X}-(\rho\sigma_X/\sigma_Y)\widetilde{Y}=\mu_X-\rho\sigma_X\mu_Y/\sigma_Y+\sigma_X\sqrt{1-\rho^2}V$$

is independent of $\widetilde{Y}=\mu_Y+\sigma_YU$. Since (Z,Y) has the same law as $(\widetilde{Z},\widetilde{Y})$ it follows (from the definition of independence of random variables) that Z is independent of Y.

(b) Explain why such X and Y are independent whenever they are uncorrelated (hence also whenever $\mathbf{E}(X|Y)=\mathbf{E}X$).
ANS: Clearly, $\rho=0$ when X and Y are uncorrelated and from part (a) we see that in this case $X=Z$ is independent of Y.

(c) Verify that $\mathbf{E}(X|Y)=\mu_X+\frac{\rho\sigma_X}{\sigma_Y}(Y-\mu_Y)$.
ANS: Using part (a), that is, $X=Z+\frac{\rho\sigma_X}{\sigma_Y}Y$ and Z independent of Y, we have by linearity of the C.E. that

$$\mathbf{E}(X|Y)=\mathbf{E}(Z+\frac{\rho\sigma_X}{\sigma_Y}Y|Y)=\mathbf{E}Z+\frac{\rho\sigma_X}{\sigma_Y}Y=\mu_X-\frac{\rho\sigma_X}{\sigma_Y}\mu_Y+\frac{\rho\sigma_X}{\sigma_Y}Y=\mu_X+\frac{\rho\sigma_X}{\sigma_Y}(Y-\mu_Y).$$

6. Exercise 3.2.26. Suppose $\{X_t\}$ is a zero-mean, (weak sense) stationary process with auto-covariance function $r(t)$.

(a) Show that $|r(h)|\le r(0)$ for all $h>0$.
ANS: By Proposition 1.2.41 and stationarity,

$$|r(h)|=|\mathbf{E}X_hX_0|\le(\mathbf{E}X_h^2)^{1/2}(\mathbf{E}X_0^2)^{1/2}=\mathbf{E}X_0^2=r(0).$$

(b) Show that if $r(h)=r(0)$ for some $h>0$ then $X_{t+h}\overset{a.s.}{=}X_t$ for each t.
ANS: If $r(h)=r(0)$, then by stationarity

$$\mathbf{E}(X_h-X_0)^2=\mathbf{E}X_h^2-2\mathbf{E}X_0X_h+\mathbf{E}X_0^2=2(r(0)-r(h))=0\,,$$

implying that $X_h\overset{a.s.}{=}X_0$.

(c) Explain why part (c) of Exercise 3.2.13 implies that if $\{X_t\}$ is a zero-mean, stationary, Gaussian process with auto-covariance function $r(t)$ such that $r(0)>0$, then $\mathbf{E}(X_{t+h}|X_t)=\frac{r(h)}{r(0)}X_t$ for any t and $h\ge0$.
ANS: If $\{X_t\}$ is such a Gaussian process, then the random vector (X_{t+h},X_t) has a Gaussian distribution with mean vector $(0,0)$ and covariance matrix $\begin{pmatrix}r(0)&r(h)\\r(h)&r(0)\end{pmatrix}$. Hence, $\mathbf{E}(X_{t+h}|X_t)=\frac{r(h)}{r(0)}X_t$ by part (c) of Exercise 3.2.13.

(d) Conclude that there is no zero-mean, stationary, Gaussian process of independent increments other than the trivial process $X_t\equiv X_0$.
ANS: Suppose $\{X_t\}$ is a zero-mean, stationary, Gaussian process of independent increments and auto-covariance function $r(t)$. If $r(0)=0$ then $X_t\equiv0=X_0$ for all t, as claimed. Next, if $r(0)>0$ then $X_t\neq0$ with positive probability and for any $h>0$ by the assumed independence of $X_{t+h}-X_t$ and X_t we have from part (c) that

$$0=\mathbf{E}(X_{t+h}-X_t)=\mathbf{E}(X_{t+h}-X_t|X_t)=\frac{r(h)}{r(0)}X_t-X_t\,,$$

implying that $r(h)=r(0)$. By part (b) we thus conclude that $X_h=X_0$ a.s. for any fixed $h>0$, as claimed.

Math 136 - Stochastic Processes

Homework Set 5, Autumn 2013, Due: October 30

1. Exercise 3.2.21. Consider the random variables $\widehat{S}_k$ of Example 1.4.13.

(a) Applying Proposition 3.2.6 verify that the corresponding characteristic functions are

$$\Phi_{\widehat{S}_k}(\theta) = [\cos(\theta/\sqrt{k})]^k\,.$$

ANS: Let X_i for $i = 1 \dots k$ be i.i.d. RVs with $\mathbf{P}(X_i = -1) = \mathbf{P}(X_i = 1) = 1/2$. Then using Proposition 3.2.6 for the first equality we have

$$\Phi_{\widehat{S}_k}(\theta) = \prod_{i=1}^{k} \Phi_{X_i/\sqrt{k}}(\theta) = \{\Phi_{X_1/\sqrt{k}}(\theta)\}^k = \{\mathbf{E}(e^{i\theta X_1/\sqrt{k}})\}^k = \{(e^{-\theta/\sqrt{k}} + e^{\theta/\sqrt{k}})/2\}^k = \{\cos(\theta/\sqrt{k})\}^k$$

(b) Recalling that $\delta^{-2}\log(\cos\delta) \to -0.5$ as $\delta \to 0$, find the limit of $\Phi_{\widehat{S}_k}(\theta)$ as $k \to \infty$ while $\theta \in \mathbb{R}$ is fixed.

ANS: Note that $\Phi_{\widehat{S}_k}(\theta) = \exp\{k \log[\cos(\theta/\sqrt{k})]\}$. Taking $\delta = \theta/\sqrt{k}$ and exploiting the continuity of the exponential function we get $\Phi_{\widehat{S}_k}(\theta) \to e^{-\theta^2/2}$.

(c) Suppose random vectors $\underline{X}^{(k)}$ and $\underline{X}$ in $\mathbb{R}^n$ are such that $\Phi_{\underline{X}^{(k)}}(\underline{\theta}) \to \Phi_{\underline{X}}(\underline{\theta})$ as $k \to \infty$, for any fixed $\underline{\theta}$. It can be shown that then the laws of $\underline{X}^{(k)}$, as probability measures on $\mathbb{R}^n$, must converge weakly in the sense of Definition 1.4.20 to the law of $\underline{X}$. Explain how this fact allows you to verify the C.L.T. statement $\widehat{S}_n \xrightarrow{\mathcal{L}} G$ of Example 1.4.13.

ANS: From the previous part we see that $\Phi_{\widehat{S}_k}(\theta) \to \Phi_G(\theta)$ for all θ, where G is a standard normal random variable. Then what has been stated above implies that $\widehat{S}_k \xrightarrow{\mathcal{L}} G$.

2. Exercise 3.2.22. Consider the random vectors $\underline{X}^{(k)} = (\frac{1}{\sqrt{k}}S_{k/2}, \frac{1}{\sqrt{k}}S_k)$ in $\mathbb{R}^2$, where $k = 2, 4, 6, \dots$ is even, and S_k is the *simple random walk* of Definition 3.1.2, with $\mathbf{P}(\xi_1 = -1) = \mathbf{P}(\xi_1 = 1) = 0.5$.

(a) Verify that $\Phi_{\underline{X}^{(k)}}(\underline{\theta}) = [\cos((\theta_1+\theta_2)/\sqrt{k})]^{k/2}[\cos(\theta_2/\sqrt{k})]^{k/2}$, where $\underline{\theta} = (\theta_1, \theta_2)$.

ANS: Here $\Phi_{\underline{X}^{(k)}}(\underline{\theta}) = \mathbf{E}\exp(\theta_1 S_{k/2}/\sqrt{k} + \theta_2 S_k/\sqrt{k})$ and since $S_k = S_{k/2} + \tilde{S}_{k/2}$ where $\tilde{S}_{k/2}$ is independent, identically distributed copy of $S_{k/2}$, we have

$$\mathbf{E}\exp(\theta_1 S_{k/2}/\sqrt{k} + \theta_2 S_k/\sqrt{k}) = \mathbf{E}\exp[(\theta_1+\theta_2)S_{k/2}/\sqrt{k}]\mathbf{E}\exp[\theta_2 S_{k/2}/\sqrt{k}]$$

The required result now follows by noting that $S_k/\sqrt{k}$ has the same distribution as $\widehat{S}_k$ from Exercise 3.2.21, so their characteristic functions are equal.

(b) Find the mean vector $\underline{\mu}$ and the covariance matrix Σ of a Gaussian random vector $\underline{X}$ for which $\Phi_{\underline{X}^{(k)}}(\underline{\theta})$ converges to $\Phi_{\underline{X}}(\underline{\theta})$ as $k \to \infty$.

ANS: Same approach as in part (b) of the Exercise 3.2.21 gives $\Phi_{\underline{X}^{(k)}}(\underline{\theta}) \to e^{-(\theta_1+\theta_2)^2/4}e^{-\theta_2^2/4}$. We now need $\underline{\mu}$ and Σ such that

$$\exp[-(\underline{\theta}, \Sigma\underline{\theta})/2 + i(\underline{\theta}, \underline{\mu})] = \exp[(-\theta_1^2/2 - \theta_1\theta_2 - \theta_2^2)/2]$$

which gives $\underline{\mu} = (0,0)$, $\Sigma_{11} = \Sigma_{12} = \Sigma_{21} = 1/2$ and $\Sigma_{22} = 1$.

(c) Upon appropriately generalizing what you did in part (b), I claim that the *Brownian motion* of Theorem 3.1.3 must be a Gaussian stochastic process. Explain why, and guess what is the mean $\mu(t)$ and auto-covariance function $\rho(t,s)$ of this process (if needed take a look at Chapter 5).

ANS: The Brownian motion of Theorem 3.1.3 arises as a weak limit as $k \to \infty$ of linear interpolations of a scaled random walk, which at $t = \frac{i}{k}$ with $i = 0 \dots k$ has values $X_t^{(k)} := \frac{1}{\sqrt{k}}S_{kt}$ (where S_n as in part (b) above). Now, for any $0 \le t_1 < t_2 < \cdots < t_n \le 1$ the random vector $(X_{t_1}^{(k)}, \dots, X_{t_n}^{(k)})$ will be a generalization of the random vector $\underline{X}^{(k)}$ from part (b) above (there $n = 2$, $t_1 = 1/2$ and $t_2 = 1$) and it will converge weakly to a Gaussian RV. But this weak limit is also a f.d.d. of the process obtained as a weak limit in Theorem 3.1.3, i.e. the Brownian motion. So all its f.d.d.'s must be Gaussian, which means that the Brownian motion is a Gaussian process. Guided by part (b) we guess $\mu(t) = 0$ and $\rho(t,s) = \min(t,s)$.

3. Exercise 3.3.5. Suppose that the stochastic process X_t is such that $\mathbf{E}[X_t] = 0$ and $\mathbf{E}[X_t^2] = 1$ for all $t \in [0,T]$.

(a) Show that $|\mathbf{E}[X_t X_{t+h}]| \le 1$ for any $h > 0$ and $t \in [0, T-h]$.

ANS: By Jensen's inequality for $g(x) = |x|$ and Proposition 1.2.41:

$$|\mathbf{E}[X_t X_{t+h}]| \le \mathbf{E}[|X_t X_{t+h}|] \le \sqrt{\mathbf{E}X_t^2}\sqrt{\mathbf{E}X_{t+h}^2} = 1.$$

(b) Suppose that for some $\lambda < \infty$, $p > 1$, and $h_0 > 0$,

$$\mathbf{E}[X_t X_{t+h}] \ge 1 - \lambda h^p$$

for all $0 < h \le h_0$. Using Kolmogorov's continuity theorem show that then X_t has a continuous modification.

ANS: $\mathbf{E}(|X_{t+h} - X_t|^2) = \mathbf{E}(X_{t+h}^2) + \mathbf{E}(X_t^2) - 2\mathbf{E}(X_{t+h}X_t) = 2(1 - \mathbf{E}[X_{t+h}X_t)] \le (2\lambda)h^p$. By Kolmogorov's theorem with $\alpha = 2$, $c = 2\lambda$ and $\beta = p - 1 > 0$ the process X_t has a continuous modification.

(c) Suppose that X_t is a Gaussian stochastic process such that $\mathbf{E}[X_t] = 0$ and $\mathbf{E}[X_t^2] = 1$ for all $t \in [0,T]$. Show that if X_t satisfies the inequality (3.3.2) for some $\lambda < \infty$, $p > 0$, and $h_0 > 0$, then for any $0 < \gamma < p/2$, the process X_t has a modification which is locally Hölder continuous with exponent γ. (Hint: see Section 5.1 for the moments of Gaussian R.V.).

ANS: Since $\{X_t\}$ is a zero-mean Gaussian stochastic process, $X_{t+h} - X_t$ is a zero mean Gaussian random variable, so by (3.3.2),

$$\mathbf{E}[|X_{t+h} - X_t|^{2n}] = \frac{(2n)!}{2^n n!}\left[\mathbf{E}(X_{t+h} - X_t)^2\right]^n \le [(2n)!/n!]\lambda^n h^{pn}$$

for any integer n, $0 < h \le h_0$ and $t \in [0, T-h]$. Fix an integer n large enough so $\gamma < \beta/\alpha$ when $\alpha = 2n$ and $\beta = pn - 1$ (i.e. $\gamma < p/2 - 1/(2n)$), and set $c = [(2n)!\lambda^n]/n!$ finite. By the preceding, Kolmogorov's continuity theorem applies for these parameters and yields the existence of a modification of X_t that is locally Hölder continuous with exponent γ.

4. Exercise 3.3.8

(a) Let $\{X_n\}, \{Y_n\}$ be discrete time S.P.s that are modifications of each other. Show that $\mathbf{P}(X_n = Y_n$ for all $n \ge 0) = 1$.

ANS: For each n let $A_n = \{\omega : X_n(\omega) = Y_n(\omega)\}$. Since $\{X_n\}, \{Y_n\}$ are modifications of each other we know that $\mathbf{P}(A_n) = 1$. Hence $\mathbf{P}(\cap_{n=1}^\infty A_n) = 1$ since a *countable* intersection of sets that occur with probability one also occurs with probability one. Noting that $\cap_{n=1}^\infty A_n = \{\omega : X_n(\omega) = Y_n(\omega)$ for all $n \ge 0\}$ gives the desired result.

(b) Let $\{X_t\}, \{Y_t\}$ be continuous time S.P.s that are modifications of each other. Suppose that both processes have right-continuous sample paths a.s. Show that $\mathbf{P}(X_t = Y_t$ for all $t \ge 0) = 1$.

ANS: Without loss of generality we assume that the sample paths of $\{X_t\}$ and $\{Y_t\}$ are right-continuous *for all* ω. For each $t \ge 0$, let $A_t = \{\omega : X_t(\omega) = Y_t(\omega)\}$. Since $\{X_t\}, \{Y_t\}$ are modifications of each other we know that $\mathbf{P}(A_t) = 1$. The set $\mathbb{Q}$ of rational numbers is countable, so $A = \cap_{r \in \mathbb{Q}, r \ge 0} A_r$ is a countable intersection of sets A_r such that $\mathbf{P}(A_r) = 1$ and consequently $\mathbf{P}(A) = 1$ as well. It thus suffices to show that $A_t \supseteq \cap_{r \in \mathbb{Q}, r \ge t} A_r$ for all $t \ge 0$ since then $B = \cap_{t \ge 0} A_t \supseteq A$ so $\mathbf{P}(B) \ge \mathbf{P}(A)$ yielding that $\mathbf{P}(B) = 1$ as is claimed. Thus, it suffices to show that if $\omega \in \cap_{r \in \mathbb{Q}, r \ge t} A_r$ for $t \ge 0$ irrational, then $\omega \in A_t$ as well. Indeed, by right continuity of the sample path of both processes,

$$X_t(\omega) = \lim_{r \in \mathbb{Q}, r \downarrow t} X_r(\omega) = \lim_{r \in \mathbb{Q}, r \downarrow t} Y_r(\omega) = Y_t(\omega),$$

which gives the desired result.

(c) Provide an example of two S.P.-s which are modifications of one another but which are not indistinguishable.

ANS: The underlying probability space is $(\mathbb{R}, \mathcal{B}, U)$ with U the uniform measure on $(0,1)$. Let $X_t = 0$ be a constant stochastic process and $Y_t(\omega) = 0$ if $t \neq \omega$ and $Y_t(\omega) = 1$ if $t = \omega$, for $t \in [0,1]$. Then,

$$\mathbf{P}(X_t = Y_t) = U(\{\omega \in (0,1) : \omega \neq t\}) = 1$$

but

$$\mathbf{P}(\{X_t = Y_t \text{ for all } t \in [0,1]\}) = 0$$

since for every $t \in [0,1]$, $X_t(t) \neq Y_t(t)$.

5. Exercise 5.1.4. Suppose W_t is a Brownian motion and $\alpha, s, T > 0$ are non-random constants. Show the following.

(a) (Symmetry) $\{-W_t, t \ge 0\}$ is a Brownian motion.

ANS: Obviously $-W_t$ remains Gaussian, continuous, and has the same mean function and auto-covariance functions as W_t. Indeed,

$$\mathbf{E}(-W_t) = -\mathbf{E}W_t = 0$$

and

$$\mathbf{E}(-W_t)(-W_s) = \mathbf{E}W_t W_s = \min(t,s).$$

(b) (Time homogeneity) $\{W_{s+t} - W_s, t \ge 0\}$ is a Brownian motion.

ANS: Again, it is clear that $W_{s+t} - W_s$ is a continuous Gaussian process. Its mean and auto-covariance functions are,

$$\mathbf{E}(W_{s+t} - W_s) = \mathbf{E}W_{s+t} - \mathbf{E}W_s = 0,$$

and

$$\begin{aligned}\mathbf{E}(W_{s+t} - W_s)(W_{s+\sigma} - W_s) &= \mathbf{E}(W_{s+t}W_{s+\sigma} - W_{s+t}W_s - W_{s+\sigma}W_s + W_s^2)\\ &= \min(s+\sigma, s+t) - 2s + s = \min(\sigma, t).\end{aligned}$$

These agree with that of Brownian motion which gives the desired conclusion.

(c) (Time reversal) $\{W_T - W_{T-t}, 0 \le t \le T\}$ is a Brownian motion.

ANS: Clearly, $W_T - W_{T-t}$ is continuous and Gaussian. We compute,

$$\mathbf{E}(W_T - W_{T-t}) = 0$$

and

$$\begin{aligned}\mathbf{E}(W_T - W_{T-t})(W_T - W_{T-s}) &= \mathbf{E}(W_T^2 - W_T W_{T-t} - W_T W_{T-s} + W_{T-t}W_{T-s})\\ &= T - (T-t) - (T-s) + \min(T-t, T-s)\\ &= t + s - T + \min(T-t, T-s)\\ &= \min(s,t),\end{aligned}$$

giving the desired result.

(d) (Scaling, or self-similarity) $\{\sqrt{\alpha}W_{t/\alpha}, t \ge 0\}$ is a Brownian motion.

ANS: Since both spatial and time scaling are continuous and spatial scaling preserves the Gaussian distribution, $\sqrt{\alpha}W_{t/\alpha}$ is a continuous Gaussian process. Its mean and auto-covariance functions are,

$$\mathbf{E}\sqrt{\alpha}W_{t/\alpha} = \sqrt{\alpha}\mathbf{E}W_{t/\alpha} = 0$$

and

$$\mathbf{E}(\sqrt{\alpha}W_{t/\alpha}\sqrt{\alpha}W_{s/\alpha}) = \alpha\mathbf{E}W_{t/\alpha}W_{s/\alpha} = \alpha\min(t/\alpha, s/\alpha) = \min(t,s).$$

Hence $\sqrt{\alpha}W_{t/\alpha}$ is a Brownian motion.

(e) (Time inversion) If $\widetilde{W}_0 = 0$ and $\widetilde{W}_t = tW_{1/t}$, then $\{\widetilde{W}_t, t \ge 0\}$ is a Brownian motion.

ANS: Again, it is clear that $\widetilde{W}_t$ is a Gaussian process with mean function $\mu(t) \equiv 0$. Fixing $t > 0$ we note that $\mathbf{E}\widetilde{W}_0\widetilde{W}_t = 0 = \min(0,t)$. Further, for any $s > 0$,

$$\mathbf{E}\widetilde{W}_s\widetilde{W}_t = (st)\min(1/s, 1/t) = \min(s,t).$$

Hence, $\widetilde{W}_t$ has the same auto-covariance function as Brownian motion. At this point we know that $\widetilde{W}_t$ has the same f.d.d. as Brownian motion and that there exists an event Γ with $\mathbf{P}(\Gamma) = 0$ such that $t \mapsto \widetilde{W}_t(\omega)$ is continuous at any $t > 0$ provided $\omega \notin \Gamma$. So, $\widetilde{W}_t$ is a Brownian motion if almost surely $\widetilde{W}_t \to 0$ when $t \downarrow 0$. The most direct way to show this is to recall that as $\widetilde{W}_t$ has the f.d.d. of a Brownian motion, by Kolmogorov's continuity theorem $\widetilde{W}_t$ has a continuous modification V_t. With both $t \mapsto \widetilde{W}_t(\omega)$ and $t \mapsto V_t(\omega)$ continuous at any $t > 0$ and all $\omega \notin \Gamma'$ such that $\mathbf{P}(\Gamma') = 0$, necessarily $\mathbf{P}(V_t = \widetilde{W}_t$ for all $t > 0) = 1$ (by the same argument you used in solving part (b) of Exercise 3.3.8). Since almost surely both $V_t \to V_0 = 0$ when $t \to 0$ and $\widetilde{W}_t = V_t$ for all $t > 0$, it follows that also $\widetilde{W}_t \to 0$ a.s. An alternative proof of the a.s. convergence to 0 of $\widetilde{W}_t$ is by invoking the strong law of large numbers to have that $\widetilde{W}_{1/n} = n^{-1}W_n \to 0$ as $n \to \infty$ (since W_n is a sum of n i.i.d. $N(0,1)$ random variables) then arguing that the Gaussian process $t^{-1}W_t$ does not fluctuate much on $t \in [n, n+1]$ (via standard bounds for the tail of $N(0,1)$ random variables).

(f) With W_t^i denoting independent Brownian motions find the constants c_n such that $c_n\sum_{i=1}^n W_t^i$ are also Brownian motions.

ANS: Let $B_t = c_n\sum_{i=1}^n W_t^i$ which is obviously a zero-mean, continuous, Gaussian process. The constants c_n are thus determined so the requirement that $\mathbf{E}B_tB_s = \min(s,t)$. Indeed, by the independence of the Brownian motions W_t^i,

$$\mathbf{E}B_tB_s = c_n^2\sum_{i,j=1}^n \mathbf{E}W_t^iW_s^j = c_n^2\sum_{i=1}^n \mathbf{E}W_t^iW_s^i = c_n^2 n\min(s,t)\,,$$

so we get the stated result for $c_n = 1/\sqrt{n}$.

6. Exercise 5.1.12 Fix $H \in (0,1)$. A Gaussian stochastic process $\{X_t,\ t \ge 0\}$ is called a fractional Brownian motion (or in short, fBM), of Hurst parameter H if $\mathbf{E}(X_t) = 0$ and

$$\mathbf{E}(X_tX_s) = \frac{1}{2}[|t|^{2H} + |s|^{2H} - |t-s|^{2H}],\ s,t \ge 0.$$

(a) Show that an fBM of Hurst parameter H has a continuous modification that is also locally Hölder continuous with exponent γ for any $0 < \gamma < H$.

ANS: Fix $0 < \gamma < H$. We have for all t, s and any positive integer n that

$$\mathbf{E}|X_t - X_s|^{2n} = C_n(\mathbf{E}|X_t - X_s|^2)^n = C_n(\mathbf{E}X_t^2 + \mathbf{E}X_s^2 - 2\mathbf{E}X_tX_s)^n = C_n|t-s|^{2Hn},$$

where C_n are some non-random finite constants (c.f. the explicit formula for moments of a normal random variable, immediately after the proof of Proposition 5.1.3). So from Kolmogorov's continuity theorem (with $\alpha = 2n$ and $\beta = 2Hn - 1$) we see that X_t possesses a continuous modification with any Hölder exponent in $(0, (2Hn-1)/2n)$. With $(2Hn-1)/2n = H - 1/(2n)$ we get the desired result by taking n large enough so that $H - \frac{1}{2n} > \gamma$.

(b) Verify that in case $H = 1/2$ such a modification yields the (standard) Brownian motion.

ANS: Since such a modification is a continuous Gaussian process, we just need to show that for $H = 1/2$ the process has the same mean and auto-covariance as the standard Brownian motion. The former is obvious and for the latter, we compute,

$$\mathbf{E}(X_tX_s) = \frac{1}{2}[t + s - |t-s|] = \min(t,s).$$

(c) Show the self-similarity property, whereby for any non-random $\alpha > 0$ the process $\{\alpha^H X_{t/\alpha}\}$ is an fBM of the same Hurst parameter H.

ANS: With $\{X_t\}$ a Gaussian S.P. visibly so is $\{\alpha^H X_{t/\alpha}\}$. It thus suffices to show that the rescaled

process has the same mean and auto-covariance functions as $\{X_t\}$. The former is obvious and for the latter, we compute,

$$\mathbf{E}\alpha^H X_{t/\alpha}\alpha^H X_{s/\alpha} = \alpha^{2H}(\frac{1}{2}[(t/\alpha)^{2H} + (s/\alpha)^{2H} - |t/\alpha - s/\alpha|^{2H}]) = \frac{1}{2}[t^{2H} + s^{2H} - |t-s|^{2H}].$$

(d) For which values of H is the fBM a process of stationary increments and for which values of H is it a process of independent increments?

ANS: Recall that we have already seen in part (a) that $\mathbf{E}(X_{t+h} - X_t)^2 = h^{2H}$. Since the distributional properties of Gaussian random variables are determined entirely by their mean and variance we thus conclude that the fBM process has stationary increments for all H. As $\{X_t\}$ is Gaussian, it has independent increments if and only if

$$\mathbf{E}(X_t - X_s)(X_{t'} - X_{s'}) = 0$$

for all $t > s \ge t' > s'$. We compute,

$$\begin{aligned}\mathbf{E}(X_t - X_s)(X_{t'} - X_{s'}) &= \mathbf{E}(X_tX_{t'} + X_sX_{s'} - X_tX_{s'} - X_sX_{t'})\\ &= \frac{1}{2}((t-s')^{2H} + (s-t')^{2H} - (t-t')^{2H} - (s-s')^{2H})\end{aligned}$$

If $H = 1/2$ it is easy to see that the above vanishes[1]. Suppose $H \neq 1/2$. Then setting $t = 2, s = t' = 1, s' = 0$, the above becomes

$$\frac{1}{2}(2^{2H} - 2) \neq 0.$$

Hence $\{X_t\}$ has independent increments if and only if $H = 1/2$.

[1] Note that we could have immediately conclude the independence of increments when $H = 1/2$ from part (b) since the standard Brownian motion has this property; however, we went through the above computation since it useful for the next part.

Math 136 - Stochastic Processes

Homework Set 6, Autumn 2013, Due: November 6

1. Exercise 4.1.6 Provide an example of a probability space $(\Omega, \mathcal{F}, \mathbf{P})$, a filtration $\{\mathcal{F}_n\}$ and a stochastic process $\{X_n\}$ adapted to $\{\mathcal{F}_n\}$ such that:

(a) $\{X_n\}$ is a martingale with respect to its canonical filtration but $(X_n, \mathcal{F}_n)$ is not a martingale.

ANS: Take $\Omega = \{a,b\}$, $\mathcal{F}_0 = \mathcal{F} = 2^\Omega$, $X_0 = 0$, $X_1 = \pm 1$ with probability 1/2 and $X_n = X_1$ for all $n \ge 2$. Then $\{X_n\}$ is a martingale with respect to its canonical filtration since:

$$X_0 = 0 = \mathbf{E}(X_1) = \mathbf{E}(X_1|X_0)$$

and

$$X_n = \mathbf{E}(X_n|X_n) = \mathbf{E}(X_{n+1}|X_n) = \mathbf{E}(X_{n+1}|\sigma(X_0, \dots, X_n))$$

for all $n \ge 1$. Now consider the filtration $\{\mathcal{F}_n\}$ where $\mathcal{F}_n = 2^\Omega$ for all n. Then,

$$X_0 = 0 \neq X_1 = \mathbf{E}(X_1|\mathcal{F}_0),$$

so that $(X_n, \mathcal{F}_n)$ is not a martingale.

(b) Provide a probability measure $\mathbf{Q}$ on $(\Omega, \mathcal{F})$ under which $\{X_n\}$ is not a martingale even with respect to its canonical filtration.

ANS: Let $\mathbf{Q}$ be a probability measure on $(\Omega, \mathcal{F})$ such that $X_1 = 1$ with probability $p > 1/2$ and $X_1 = -1$ with probability $1 - p < 1/2$. Then

$$\mathbf{E}X_1 = (2p-1) > 0 \neq 0 = X_0$$

so that $\{X_n\}$ is not a martingale with respect to its canonical filtration.

2. Exercise 4.1.23 Let $\xi_1, \xi_2, \dots$ be independent with $\mathbf{E}\xi_i = 0$ and $\mathbf{E}\xi_i^2 = \sigma_i^2$.

(a) Let $S_n = \sum_{i=1}^n \xi_i$ and $s_n^2 = \sum_{i=1}^n \sigma_i^2$. Show that $\{S_n^2\}$ is a sub-martingale and $\{S_n^2 - s_n^2\}$ is a martingale.

ANS: Using the same argument of Example 4.1.8 we know that $\{S_n\}$ is a martingale with respect to its canonical filtration. Moreover, from the fact that $S_n^2 = \sum_{i=1}^n \xi_i^2 + 2\sum_{1 \le i < j \le n}\xi_i\xi_j$ it is clear that $\mathbf{E}|S_n^2| < \infty$ for all n. Thus since $x \mapsto x^2$ is a convex function it follows from the conditional Jensen inequality that S_n^2 is a sub-martingale. Letting $\mathcal{F}_n = \sigma(S_1, \dots, S_n)$ and using that ξ_{n+1} is independent of $\mathcal{F}_n$, we have

$$\mathbf{E}[S_{n+1}^2|\mathcal{F}_n] = \mathbf{E}[(S_n + \xi_{n+1})^2|\mathcal{F}_n] = \mathbf{E}[S_n^2 + 2\xi_{n+1}S_n + \xi_{n+1}^2|\mathcal{F}_n] = S_n^2 + \sigma_{n+1}^2.$$

Hence,

$$\mathbf{E}[S_{n+1}^2 - s_{n+1}^2|\mathcal{F}_n] = S_n^2 - s_{n+1}^2 + \sigma_{n+1}^2 = S_n^2 - s_n^2.$$

Thus $\{S_n^2 - s_n^2\}$ is a martingale as desired.

(b) Suppose also that $m_n = \prod_{i=1}^n \mathbf{E}[e^{\xi_i}] < \infty$. Show that $\{e^{S_n}\}$ is a sub-martingale and $M_n = e^{S_n}/m_n$ is a martingale.

ANS: By assumption $m_n < \infty$ giving us that $\{e^{S_n}\}$ is an integrable SP. Since $\{S_n\}$ is a martingale and $x \mapsto e^x$ is convex it follows from the conditional Jensen inequality that $\{e^{S_n}\}$ is a sub-martingale. Letting $\mathcal{F}_n = \sigma(M_1, \dots, M_n)$, the independence of ξ_{n+1} and $\mathcal{F}_n$ gives us that

$$\mathbf{E}[M_{n+1}|\mathcal{F}_n] = \frac{1}{m_{n+1}}\mathbf{E}[e^{S_n}e^{\xi_{n+1}}|\mathcal{F}_n] = \frac{e^{S_n}}{m_{n+1}}\mathbf{E}[e^{\xi_{n+1}}] = \frac{e^{S_n}}{m_n} = M_n.$$

Therefore $\{M_n\}$ is a martingale.

3. Exercise 4.2.5. Let $\mathcal{G}_t$ denote the canonical filtration of a Brownian motion W_t.

(a) Show that for any $\lambda \in \mathbb{R}$, the S.P. $M_t(\lambda) = \exp(\lambda W_t - \lambda^2 t/2)$, is a continuous time martingale with respect to $\mathcal{G}_t$.

ANS: Note that $\mathbf{E}|M_t(\lambda)| = e^{-\lambda^2(t/2)}\mathbf{E}[e^{\lambda W_t}]$ which since W_t is a Gaussian random variable, we know to be finite. Further, $\mathbf{E}e^{\lambda(W_{t+h} - W_t)} = e^{\lambda^2 h/2}$ yielding the identity $\mathbf{E}[M_{t+h}(\lambda)|\mathcal{G}_t] = e^{-\lambda^2(t/2) + \lambda W_t} = M_t(\lambda)$, so $M_t(\lambda)$ is a martingale.

(b) Explain why $\frac{d^k}{d\lambda^k}M_t(\lambda)$ are also martingales with respect to $\mathcal{G}_t$.

ANS: Fixing $\lambda \in \mathbb{R}$, let $\lambda_m = \lambda + m^{-1}$ and recall that $M_t(\lambda_m)$ is a MG with respect to $\mathcal{G}_t$. The adapted process $X_t(m, \lambda) := m(M_t(\lambda_m) - M_t(\lambda))$ is then also a MG with respect to $\mathcal{G}_t$. That is, $\mathbf{E}[X_{t+h}(m,\lambda) - X_t(m,\lambda)|\mathcal{G}_t] = 0$ for any non-random $h, t \ge 0$. Further, considering $m \to \infty$ we get by definition of the derivative that $X_t(m, \lambda)$ converges a.s. to the adapted S.P. $Z_t(\lambda) := \frac{d}{d\lambda}M_t(\lambda)$. Now, by the mean value theorem

$$\sup_m\{|X_t(m,\lambda)|\} \le Y_t := \sup\{|Z_t(\lambda + u)| : 0 \le u \le 1\}\,.$$

Computing explicitly $Z_t(\lambda + u)$, it is not hard to check that $Y_t \le (|W_t| + (|\lambda| + 1)t)e^{(|\lambda|+1)|W_t|}$ is integrable (per fixed $t \ge 0$). From the preceding we thus deduce by dominated convergence for C.E. that a.s. $\mathbf{E}[Z_{t+h}(\lambda) - Z_t(\lambda)|\mathcal{G}_t] = 0$. Consequently, given $\lambda \in \mathbb{R}$ non-random, the process $Z_t(\lambda)$ is a MG with respect to $\mathcal{G}_t$. Applying the same reasoning with $M_t(\lambda)$ replaced by $Z_t(\lambda)$ extends our claim from $k = 1$ to $k = 2$, and arguing inductively in k, the same applies for $k = 3, 4, \dots$.

(c) Compute the first three derivatives in λ of $M_t(\lambda)$ at $\lambda = 0$ and deduce that the S.P. $W_t^2 - t$ and $W_t^3 - 3tW_t$ are also MGs.

ANS: Fixing $x, t \in \mathbb{R}$, the derivative of $M(\lambda) := e^{\lambda x - \lambda^2 t/2}$ is $M'(\lambda) = (x - \lambda t)M(\lambda)$, yielding that $M''(\lambda) = [(x-\lambda t)^2 - t]M(\lambda)$ and $M'''(\lambda) = (x-\lambda t)[(x-\lambda t)^2 - 3t]M(\lambda)$. In case $\lambda = 0$ we have $M(0) = 1$ resulting with $M'(0) = x$, $M''(0) = x^2 - t$ and $M'''(0) = x^3 - 3tx$. Setting $x = W_t$ we deduce by the preceding that $W_t^2 - t$ and $W_t^3 - 3tW_t$ are also MGs.

4. Exercise 4.2.10 Given a positive MG $(Z_t, \mathcal{F}_t)$ with $\mathbf{E}Z_0 = 1$ consider for each $t \geq 0$ the probability measure $\widetilde{\mathbf{P}}_t \colon \mathcal{F}_t \to \mathbb{R}$ given by $\widetilde{\mathbf{P}}_t(A) = \mathbf{E}[Z_t I_A]$.

(a) Show that $\widetilde{\mathbf{P}}_t(A) = \widetilde{\mathbf{P}}_s(A)$ for any $A \in \mathcal{F}_s$ and $0 \leq s \leq t$.
ANS: Since Z_t is a martingale and I_A is $\mathcal{F}_s$-measurable, we have by the tower property and taking out what is known that

$$\widetilde{\mathbf{P}}_t(A) = \mathbf{E}[Z_t I_A] = \mathbf{E}[\mathbf{E}[Z_t I_A \,|\, \mathcal{F}_s]] = \mathbf{E}[\mathbf{E}[Z_t \,|\, \mathcal{F}_s] I_A] = \mathbf{E}[Z_s I_A] = \widetilde{\mathbf{P}}_s(A).$$

(b) Fixing $0 \leq u \leq s \leq t$ and $Y \in L^1(\Omega, \mathcal{F}_s, \widetilde{\mathbf{P}}_t)$, set $X_{s,u} = \mathbf{E}(YZ_s|\mathcal{F}_u)/Z_u$. With $\widetilde{\mathbf{E}}_t$ denoting the expectation under $\widetilde{\mathbf{P}}_t$, deduce that $\widetilde{\mathbf{E}}_t(Y|\mathcal{F}_u) = X_{s,u}$ almost surely under $\widetilde{\mathbf{P}}_t$ (hence also under $\mathbf{P}$, by Exercise 1.4.32).
ANS: First note that $YZ_s \in L^1(\Omega, \mathcal{F}_s, \mathbf{P})$ since

$$\mathbf{E}(|Y|Z_s) = \mathbf{E}(|Y|\mathbf{E}(Z_t \,|\, \mathcal{F}_s)) = \mathbf{E}(\mathbf{E}(|Y|Z_t \,|\, \mathcal{F}_s)) = \mathbf{E}(|Y|Z_t) = \widetilde{\mathbf{E}}_t(|Y|) < \infty.$$

Consequently, the $\mathcal{F}_u$ measurable random variable $X_{s,u} = \mathbf{E}(YZ_s \,|\, \mathcal{F}_u)/Z_u$ is well defined. Further, fixing $A \in \mathcal{F}_u$, recall that YI_A is $\mathcal{F}_s$ measurable and $(Z_t, \mathcal{F}_t)$ a martingale. Hence, using the tower property, taking out what is known and applying part (a) for the $\mathcal{F}_u$ measurable $X_{s,u}I_A$ we get that

$$\begin{aligned}\widetilde{\mathbf{E}}_t[YI_A] &= \mathbf{E}[Z_t Y I_A] = \mathbf{E}[\mathbf{E}(Z_t Y I_A \,|\, \mathcal{F}_s)] = \mathbf{E}[YI_A \mathbf{E}(Z_t \,|\, \mathcal{F}_s)] = \mathbf{E}[YZ_s I_A] \\ &= \mathbf{E}[\mathbf{E}(YZ_s I_A \,|\, \mathcal{F}_u)] = \mathbf{E}[\mathbf{E}(YZ_s \,|\, \mathcal{F}_u) I_A] = \mathbf{E}[Z_u X_{s,u} I_A] = \widetilde{\mathbf{E}}_u[X_{s,u}I_A] = \widetilde{\mathbf{E}}_t[X_{s,u}I_A].\end{aligned}$$

Since this applies for any $A \in \mathcal{F}_u$, we have by definition of the conditional expectation in the probability space $(\Omega, \mathcal{F}, \widetilde{\mathbf{P}}_t)$ that $X_{s,u} = \widetilde{\mathbf{E}}_t(Y|\mathcal{F}_u)$ up to a set $N \in \mathcal{F}$ such that $\widetilde{\mathbf{P}}_t(N) = 0$. Recall Exercise 1.4.32 that $\widetilde{\mathbf{P}}_t(N) = 0$ if and only if $\mathbf{P}(N) = 0$, so the identity $X_{s,u} = \widetilde{\mathbf{E}}_t(Y|\mathcal{F}_u)$ holds for $\mathbf{P}$ almost every ω, as claimed.

5. Exercise 5.1.8. Compute the mean and the auto-covariance functions of the processes B_t, Y_t, U_t, and X_t.

ANS: We compute,

$$\begin{aligned}
&\mathbf{E}(B_t) = 0,\\
&\mathbf{E}(B_tB_s) = s(1-t) \text{ when } 0 \leq s \leq t \leq 1,\\
&\mathbf{E}(B_tB_s) = s-1 \text{ when } 0 \leq 1 \leq s \leq t \text{ and}\\
&\mathbf{E}(B_tB_s) = 0 \text{ when } 0 \leq s \leq 1 \leq t;\\
&\mathbf{E}(Y_t) = e^{t/2} \text{ and } \mathbf{E}[(Y_t - e^{t/2})(Y_s - e^{s/2})] = e^{(t+s)/2}(e^{\min(t,s)} - 1);\\
&\mathbf{E}(U_t) = 0 \text{ and } \mathbf{E}(U_tU_s) = e^{-|t-s|/2}.\\
&\mathbf{E}X_t = x + \mu t \text{ and,}\\
&\mathbf{E}[(X_t - x - \mu t)(X_s - x - \mu s)] = \sigma^2 \mathbf{E}(W_tW_s) = \sigma^2 \min(t,s).
\end{aligned}$$

Justify your answers to:

(a) Which of the processes W_t, B_t, Y_t, U_t, X_t is Gaussian?
ANS: We know that W_t is a Gaussian process. The f.d.d. of the S.P. B_t and U_t correspond to deterministic linear combinations of the joint Gaussian r.v. W_{t_i}, hence both B_t and U_t are Gaussian processes. Since $Y_1 = e^{W_1}$ is strictly positive and not almost surely a constant, it can not be a Gaussian r.v, hence Y_t is not a Gaussian process. Finally, X_t is just an affine (time-dependent) translate of a Gaussian process and hence Gaussian.

(b) Which of these processes is stationary?
ANS: Stationarity implies the process has constant mean and its auto-covariance $\rho(t,s)$ is a function only of $|t-s|$. The S.P. W_t, B_t, Y_t and X_t fail to have this property so are non-stationary. The S.P. U_t satisfies these conditions and being also Gaussian, this suffices for U_t being a stationary process.

(c) Which of these processes has continuous sample paths?
ANS: W_t has continuous sample paths by the definition of Brownian motion so B_t, Y_t, U_t, X_t are finite compositions of functions continuous in t. Therefore, all five processes have continuous sample paths.

(d) Which of these processes is adapted to the filtration $\sigma(W_s, s \leq t)$ and which is also a sub-martingale for this filtration?
ANS: Recall that W_t is adapted and is a martingale for its canonical filtration. The processes B_t and U_t depend on values of W_s for $s > t$ so they are not adapted to this filtration. The S.P. Y_t is the composition of the convex function e^x and a martingale and hence a submartingale. Finally, as X_t is an affine translate of W_t, it is visibly adapted to the filtration and is a submartingale provided

that $\mu \geq 0$:

$$\mathbf{E}[X_t|\sigma(W_s : s \leq t)] = x + \mu t + \sigma W_s \geq x + \mu s + \sigma W_s = X_s.$$

Note that if $\mu < 0$ we get the reverse inequality.

6. Exercise 5.1.11. Suppose W_t is a Brownian motion.

(a) Compute the probability density function of the random vector (W_s, W_t). Then compute $\mathbf{E}(W_s|W_t)$ and $\mathrm{Var}(W_s|W_t)$, first for $s > t$, then for $s < t$.
Hint: Consider Example 2.4.5.
ANS: Suppose first that $t < s$. Then, $W_s - W_t$ is independent of W_t, having a Gaussian distribution of zero mean and variance $s-t$ Therefore, $\mathbf{E}(W_s|W_t) = W_t$ and $\mathrm{Var}(W_s|W_t) = \mathbf{E}((W_s - W_t)^2|W_t) = s - t$. Moving to deal with $t > s$, note that (W_s, W_t) is a Gaussian random vector, of zero mean and covariance matrix Σ whose entries are $\Sigma_{11} = \Sigma_{12} = \Sigma_{21} = s$, $\Sigma_{22} = t$. Upon finding that Σ is invertible and computing its inverse, we get that (W_s, W_t) has the (joint) probability density function $f_{W_s,W_t}(x,y) = \exp(-x^2/(2s) - (y-x)^2/(2(t-s)))/(2\pi\sqrt{s(t-s)})$. With the density of W_t being $g_{W_t}(y) = \exp(-y^2/2t)/\sqrt{2\pi t}$, we have by Example 2.4.5 that the conditional density of W_s given W_t is $f_{W_s|W_t}(x|W_t)$ for

$$f_{W_s|W_t}(x|y) = f_{W_s,W_t}(x,y)/g_{W_t}(y) = \exp(-(x - sy/t)^2/(2\sigma^2))/(\sqrt{2\pi}\sigma)$$

where $\sigma^2 = s(t-s)/t$. The latter is the density of a Gaussian random variable of mean sy/t and variance σ^2, so as explained in Example 2.4.5 we have that $\mathbf{E}(W_s|W_t) = (s/t)W_t$ and $\mathrm{Var}(W_s|W_t) = s - s^2/t$.

(b) Explain why the Brownian Bridge B_t, $0 \leq t \leq 1$ has the same distribution as $\{W_t, 0 \leq t \leq 1$, conditioned upon $W_1 = 0\}$ (which is the reason for naming B_t a Brownian bridge).
Hint: Both Exercise 2.4.6 and parts of Exercise 5.1.8 may help here.
ANS: For $s \leq t \leq 1$ we know that $X = W_1 - W_t$ is independent of the random vector $(Y, Z) = (W_s, W_t)$. Consequently, combining part (a) with Exercise 2.4.6 we have that $\mathbf{E}(W_s|W_t, W_1 - W_t) = \mathbf{E}(W_s|W_t) = (s/t)W_t$. Further, $\sigma(W_t, W_1) = \sigma(W_t, W_1 - W_t)$, so also $\mathbf{E}(W_s|W_t, W_1) = (s/t)W_t$. Thus, applying the tower property for $\sigma(W_1) \subseteq \sigma(W_t, W_1)$ and taking out what is known, we see that

$$\mathbf{E}[W_sW_t|W_1] = \mathbf{E}[W_t\mathbf{E}(W_s|W_t, W_1)|W_1] = (s/t)\mathbf{E}(W_t^2|W_1).$$

Recall that by part (a), $\mathbf{E}(W_t|W_1) = tW_1$ and

$$\mathbf{E}(W_t^2|W_1) = \mathrm{Var}(W_t|W_1) + [\mathbf{E}(W_t|W_1)]^2 = t - t^2 + t^2W_1^2,$$

implying that

$$\mathbf{E}(W_sW_t|W_1) = s(1-t) + stW_1^2.$$

Though we shall not do so in detail, fixing $0 < s_1 < \ldots < s_n < 1$ one can compute the density of $(W_{s_1}, \ldots, W_{s_n})$ conditional on $\{W_1 = 0\}$, per Example 2.4.5, and verify that it is the density of a (zero-mean) non-degenerate Gaussian random vector. Consequently, $\{W_t, 0 \leq t \leq 1\}$ conditional on the event $\{W_1 = 0\}$ is a Gaussian S.P. Recall Exercise 5.1.8, that $\mathbf{E}(B_t) = 0$ and $\mathbf{E}(B_sB_t) = s(1-t)$ for all $0 \leq s \leq t \leq 1$. In conclusion, we have established that the Gaussian S.P. $\{W_t, 0 \leq t \leq 1\}$ conditional on the event $\{W_1 = 0\}$, has the same mean and auto-covariance functions as the Gaussian S.P. B_t. Therefore, these two S.P. have the same distribution (i.e. the same f.d.d.).

Math 136 - Stochastic Processes
Homework Set 7, Autumn 2013, Due: November 13

1. Exercise 4.3.4. Show that the *first hitting time* $\tau(\omega) = \min\{k \geq 0 : X_k(\omega) \in B\}$ of a Borel set $B \subseteq \mathbb{R}$ by a sequence $\{X_k\}$, is a stopping time for the canonical filtration $\mathcal{F}_n = \sigma(X_k, k \leq n)$. Provide an example where the *last hitting time* $\theta = \sup\{k \geq 0 : X_k \in B\}$ of a set B by the sequence, is not a stopping time (not surprising, since we need to know the whole sequence $\{X_k\}$ in order to verify that there are no visits to B after a given time n).
ANS: We have, $\{\tau \leq n\} = \bigcup_{k=0}^n \{X_k \in B\} \in \mathcal{F}_n$ since it is a finite union of elements in $\mathcal{F}_n$. This verifies that τ is a stopping time for the filtration $\mathcal{F}_n$. Consider the stochastic process corresponding to two coin flips: $\Omega = \{HH, HT, TH, TT\}$, $\mathcal{F}_1 = \{\phi, \{HH, HT\}, \{TH, TT\}, \Omega\}$, $\mathcal{F}_2 = 2^\Omega$ and $X_k(\omega) = I_{\{\omega_k = H\}}$, $k = 1, 2$ for any $\omega = \omega_1\omega_2 \in \Omega$. Let $B = \{1\}$. Then $\{\theta \leq 1\} = \{HT, TT\} \notin \mathcal{F}_1$. So θ is not a stopping time.

2. Exercise 4.3.15 Let $\mathcal{G}_t$ denote the canonical filtration of the S.P. $\{X_t\}$.

(a) Verify that $\mathcal{G}_{t^+} = \bigcap_{u>0} \mathcal{G}_{t+u}$ is a right-continuous filtration.
ANS: Simply note that

$$\bigcap_{h>0} \mathcal{G}_{(t+h)^+} = \bigcap_{h>0}\Big[\bigcap_{u>0} \mathcal{G}_{t+h+u}\Big] = \bigcap_{v>0} \mathcal{G}_{t+v} = \mathcal{G}_{t^+},$$

so by definition $\{\mathcal{G}_{t^+}\}$ is a right-continuous filtration.

(b) Considering part (a) of Proposition 4.3.13 for filtration $\mathcal{G}_{t^+}$, deduce that for any fixed $b > 0$ and $\delta > 0$ the random variable $\tau_b^{(\delta)} = \inf\{t \geq \delta : X_{t-\delta} > b\}$ is a stopping time for $\{\mathcal{G}_t\}$, provided $\{X_t\}$ has right-continuous sample path.
ANS: Note that $\mathcal{G}_t \subseteq \mathcal{G}_{t^+}$ implying that the S.P. $Z_t = X_{t-\delta}$, $t \geq \delta$, is adapted to $\{\mathcal{G}_{(t-\delta)^+}\}$ for any fixed $\delta > 0$. Hence, $\{\tau_b^{(\delta)} \leq t\} \in \mathcal{G}_{(t-\delta)^+}$ for any $t \geq \delta$, by part (a) of Proposition 4.3.13. Further, with $\delta > 0$, it follows that $\mathcal{G}_{(t-\delta)^+} \subseteq \mathcal{G}_t$ for any $t \geq \delta$, hence $\{\tau_b^{(\delta)} \leq t\} \in \mathcal{G}_t$ for all t (the case $t < \delta$ is trivial, for then the relevant event is empty). We conclude that $\tau_b^{(\delta)}$ is a stopping time for $\{\mathcal{G}_t\}$, as claimed.

(c) With $Y_t = \int_0^t X_s^2 ds$ use part (b) of Proposition 4.3.13 to show that $\theta_1 = \inf\{t \geq 0 : Y_t = b\}$ is another stopping time for $\{\mathcal{G}_t\}$. Then explain why $\theta_2 = \inf\{t \geq 0 : Y_{2t} = b\}$, is in general not a stopping time for this filtration.
ANS: That θ_1 is a stopping time for $\{\mathcal{G}_t\}$ is immediate from the continuity of the sample path $t \mapsto Y_t$ and the fact that the singleton $\{b\}$ is a closed set (where part (b) of Proposition 4.3.13

does not require right-continuity of the filtration). There are many counterexamples to θ_2 not being a stopping time with respect to $\{\mathcal{G}_t\}$. For example, consider the right continuous process $X_t = 2\xi\mathbf{1}_{\{t\geq 1\}}$ where $\mathbf{P}(\xi = 0) = \mathbf{P}(\xi = 1) = 1/2$. For this process and $b = 1$ it is easy to verify that the event $\{\theta_2 \leq 3/4\}$ is merely $\{\xi = 1\}$. Since in this case $\mathcal{G}_t = \{\emptyset, \Omega\}$ when $t < 1$, it follows that $\{\xi = 1\} \notin \mathcal{G}_{3/4}$.

3. Exercise 4.3.18. Let W_t be a Brownian motion. Fixing $a > 0$ and $b > 0$ let $\tau_{a,b} = \inf\{t \geq 0 : W_t \notin (-a, b)\}$. We will see in Section 5.2 that $\tau_{a,b}$ is finite with probability one.

(a) Check that $\tau_{a,b}$ is a stopping time and that $W_{t\wedge\tau_{a,b}}$ is uniformly integrable.
ANS: Since W_t has continuous sample paths $|W_{t\wedge\tau_{a,b}}| \leq \max(a,b)$ is uniformly (in t and ω) bounded, hence U.I. Further, $(-a,b)^c$ is a closed set so $\tau_{a,b}$ is a stopping time by part (b) of Proposition 4.3.13.

(b) Applying Doob's optional stopping theorem for this stopped martingale, compute the probability that W_t reaches level b before it reaches level $-a$.
ANS: Since $W_{\tau_{a,b}} \in \{-a, b\}$, applying the optional stopping theorem (we can do this because of part (a) and the assumption $\tau_{a,b} < \infty, a.s.$), we have that $0 = \mathbf{E}[W_0] = \mathbf{E}[W_{\tau_{a,b}}] = -a\mathbf{P}[W_{\tau_{a,b}} = -a] + b\mathbf{P}[W_{\tau_{a,b}} = b]$. Consequently, $\mathbf{P}[W_{\tau_{a,b}} = b] = a/(b+a)$.

(c) Justify using the optional stopping theorem for $\tau_{b,b}$ and the martingales $M_t(\lambda)$ of Exercise 4.2.5. Deduce from it the value of $\mathbf{E}(e^{-\theta\tau_{b,b}})$ for $\theta > 0$.
Hint: In part (c) you may use the fact that the S.P. $\{-W_t\}$ has the law as $\{W_t\}$.
ANS: Let $X = e^{-\lambda^2\tau_{b,b}/2}$ and $A = \{W_{\tau_{b,b}} = b\}$. Noting that the non-negative $M_{t\wedge\tau_{b,b}}(\lambda) \leq e^{|\lambda|b}$ is a U.I. process and $W_{\tau_{b,b}} \in \{-b, b\}$, it follows by Doob's optional stopping theorem that

$$1 = \mathbf{E}M_0(\lambda) = \mathbf{E}(M_{\tau_{b,b}}(\lambda)) = e^{\lambda b}\mathbf{E}[XI_A] + e^{-\lambda b}\mathbf{E}[XI_{A^c}]. \tag{1}$$

Suppose we change the sign of the Brownian sample path, from $W_t(\omega)$ to $-W_t(\omega)$. The value of $\tau_{b,b}$, and hence that of X, is invariant under such a change, while the events A and A^c are exchanged by it. With the S.P. $\{-W_t\}$ having the same distribution (i.e. f.d.d.) as $\{W_t\}$, we thus deduce that $\mathbf{E}[XI_A] = \mathbf{E}[XI_{A^c}]$, and hence both are equal to $\mathbf{E}[X]/2$. Plugging this into (1) we get that $1 = \cosh(\lambda b)\mathbf{E}[e^{-\lambda^2\tau_{b,b}/2}]$. Setting $\lambda = \sqrt{2\theta}$ we conclude that $\mathbf{E}[e^{-\theta\tau_{b,b}}] = 1/\cosh(\sqrt{2\theta}b)$ from which the law of $\tau_{b,b}$ can be computed.

4. Exercise 4.3.20. Consider $M_t = \exp(\lambda Z_t)$ for non-random constants λ and r, where $Z_t = W_t + rt, t \geq 0$, and W_t is a Brownian motion.

(a) Compute the conditional expectation $\mathbf{E}(M_{t+h}|\mathcal{G}_t)$ for $\mathcal{G}_t = \sigma(Z_u, u \leq t)$ and $t, h \geq 0$.
ANS: Noting that $\mathcal{G}_t = \sigma(W_u, u \leq t)$, we have that $W_{t+h} - W_t$ is independent of $\mathcal{G}_t$ and hence

$$\begin{aligned}
\mathbf{E}(M_{t+h}|\mathcal{G}_t) &= \mathbf{E}(\exp(\lambda Z_{t+h})|\mathcal{G}_t)\\
&= \exp(\lambda Z_t)\mathbf{E}(\exp[\lambda(Z_{t+h} - Z_t)]|\mathcal{G}_t)\\
&= M_t\mathbf{E}(\exp[\lambda(W_{t+h} - W_t + rh)]|\mathcal{G}_t)\\
&= e^{\lambda rh}M_t\mathbf{E}(\exp[\lambda(W_{t+h} - W_t)])\\
&= e^{\lambda rh + \lambda^2 h/2}M_t
\end{aligned}$$

(b) Find the value of $\lambda \neq 0$ for which $(M_t, \mathcal{G}_t)$ is a martingale.
ANS: By part (a), $(M_t, \mathcal{G}_t)$ is a martingale if and only if $e^{\lambda rh + \lambda^2 h/2} = 1$ for any $h \geq 0$, which gives $\lambda = -2r$ (when $r \neq 0$).

(c) Fixing $a, b > 0$, apply Doob's optional stopping theorem to find the law of $Z_{\tau_{a,b}}$ for $\tau_{a,b} = \inf\{t \geq 0 : Z_t \notin (-a, b)\}$.
ANS: As the case $r = 0$ has been discussed in Exercise 4.3.18, we assume hereafter that $r \neq 0$ and let $\tau_c = \inf\{t \geq 0 : W_t = c\}$ for any $c \in \mathbb{R}$. We show in Section 5.2 that almost surely $\tau_c < \infty$ for each fixed $c \in \mathbb{R}$. When $r > 0$, we have $Z_t \geq W_t$ resulting with $\tau_{a,b} \leq \tau_b$; when $r < 0$, we have $Z_t \leq W_t$ resulting with $\tau_{a,b} \leq \tau_{-a}$. Therefore, $\tau_{a,b} < \infty$ almost surely. By continuity of W and hence of Z, $Z_{\tau_{a,b}} \in \{-a, b\}$. Part (b) tells us that $(M_t, \mathcal{G}_t)$ is a martingale for $M_t = \exp(-2rZ_t)$. Since $M_{t\wedge\tau_{a,b}}$ is uniformly bounded (by $e^{2|r|\max(a,b)}$), hence U.I., we can apply Doob's optional stopping theorem and get

$$1 = \mathbf{E}(M_0) = \mathbf{E}(M_{\tau_{a,b}}) = e^{2ra}\mathbf{P}(Z_{\tau_{a,b}} = -a) + e^{-2rb}\mathbf{P}(Z_{\tau_{a,b}} = b).$$

Consequently, with $1 = \mathbf{P}(Z_{\tau_{a,b}} = -a) + \mathbf{P}(Z_{\tau_{a,b}} = b)$ we get that

$$\mathbf{P}(Z_{\tau_{a,b}} = b) = \frac{e^{2ra} - 1}{e^{2ra} - e^{-2rb}} \quad \text{and} \quad \mathbf{P}(Z_{\tau_{a,b}} = -a) = \frac{1 - e^{-2rb}}{e^{2ra} - e^{-2rb}}.$$

5. Exercise 5.2.4. Let W_t be a Brownian motion.

(a) Show that $-\min_{0\leq t\leq T} W_t$ and $\max_{0\leq t\leq T} W_t$ have the same distribution which is also the distribution of $|W_T|$.
ANS: Recall that W_t is a Gaussian process of zero mean. Since its f.d.d. have densities which are symmetric around the origin, it follows that the S.P. W_t and $-W_t$ have the same law. With $-\min_{t\in[0,T]} W_t = \max_{t\in[0,T]}(-W_t)$, we see that the latter two R.V. have the same distribution. We know that $\mathbf{P}(\max_{t\in[0,T]} W_t \geq \alpha) = 2\mathbf{P}(W_T \geq \alpha) = \mathbf{P}(|W_T| \geq \alpha)$ for all $\alpha \geq 0$. So, the three R.V. $|W_T|$, $\max_{0\leq t\leq T} W_t$ and $-\min_{0\leq t\leq T} W_t$ have the same distribution.

(b) Show that the probability α that the Brownian motion W_u attains the value zero at some $u \in (s, s+t)$ is given by $\alpha = \int_{-\infty}^{\infty} p_t(|x|)\phi_s(x)dx$, where $p_t(x) = \mathbf{P}(|W_t| \geq x)$ for $x, t > 0$ and $\phi_s(x)$ denotes the probability density of the R.V. W_s for $s > 0$.
Remark: The explicit formula $\alpha = (2/\pi)\arccos(\sqrt{s/(s+t)})$ is obtained in [**KT75**, page 348] by computing this integral.
ANS: Let $\mathcal{H}_s = \sigma(W_s)$ and A denote the event $\{\exists u \in (s, s+t) : W_u = 0\}$. Then, by the tower property $\alpha = \mathbf{P}(A) = \mathbf{E}(\mathbf{P}(A|\mathcal{H}_s))$. Since $\mathbf{P}(W_{s+t} = 0|\mathcal{H}_s) = 0$ and the Brownian path is continuous, we have that

$$I_{\{W_s<0\}}\mathbf{P}(A|\mathcal{H}_s) = I_{\{W_s<0\}}\mathbf{P}(\max_{u\in[0,t]}(W_{s+u} - W_s) \geq -W_s|\mathcal{H}_s).$$

We know that conditional on $\mathcal{H}_s$, the S.P. $\{W_{s+u} - W_s : u \geq 0\}$ has the original Brownian law (for example, see Proposition 5.2.3). Applying part (a), we deduce that $I_{\{W_s<0\}}\mathbf{P}(A|\mathcal{H}_s) = I_{\{W_s<0\}}p_t(-W_s)$. The same considerations yield in case $W_s > 0$ that

$$I_{\{W_s>0\}}\mathbf{P}(A|\mathcal{H}_s) = I_{\{W_s>0\}}\mathbf{P}(-\min_{u\in[0,t]}(W_{s+u} - W_s) \geq W_s|\mathcal{H}_s).$$

It follows by part (a) then that $I_{\{W_s>0\}}\mathbf{P}(A|\mathcal{H}_s) = I_{\{W_s>0\}}p_t(W_s)$. With $W_s \neq 0$ almost surely, combining these two formulas we have that

$$\alpha = \mathbf{E}[\mathbf{P}(A|\mathcal{H}_s)] = \mathbf{E}[p_t(|W_s|)] = \int_{-\infty}^{\infty} p_t(|x|)\phi_s(x)dx$$

as stated.

6. Exercise 5.2.5. Show that $\mathbf{E}(\tau_{\beta,\alpha}) = \alpha\beta$ by applying Doob's optional stopping theorem for the uniformly integrable stopped martingale $W^2_{t\wedge\tau_{\beta,\alpha}} - t \wedge \tau_{\beta,\alpha}$.
ANS: We have seen en-route to (5.2.2) that $\tau_{\beta,\alpha} \leq \tau_\alpha < \infty$ almost surely. Considering the martingale $X_t = W_t^2 - t$ of continuous sample path we have further assumed in the statement of the exercise that $X_{t\wedge\tau_{\beta,\alpha}}$ is U.I. Thus, Doob's optional stopping theorem (Theorem 4.3.16) applies here, leading to the identity $\mathbf{E}(W^2_{\tau_{\beta,\alpha}} - \tau_{\beta,\alpha}) = \mathbf{E}(W_0^2 - 0) = 0$. That is,

$$\mathbf{E}\tau_{\beta,\alpha} = \mathbf{E}W^2_{\tau_{\beta,\alpha}} = \alpha^2\mathbf{P}(W_{\tau_{\beta,\alpha}} = \alpha) + \beta^2\mathbf{P}(W_{\tau_{\beta,\alpha}} = -\beta) = \frac{\alpha^2\beta}{\alpha+\beta} + \frac{\beta^2\alpha}{\alpha+\beta} = \alpha\beta.$$

Math 136 - Stochastic Processes
Homework Set 8, Autumn 2013, Due: November 20

1. Exercise 4.4.10. Find a non-random $f(t)$ such that $X_t = e^{W_t - f(t)}$ is a martingale, and for this value of $f(t)$ find the increasing process associated with the martingale X_t via the Doob-Meyer decomposition. Hint: Try an increasing process $A_t = \int_0^t e^{2W_s - h(s)}ds$ and use Fubini's theorem to find the non-random $h(s)$ for which $M_t = X_t^2 - A_t$ is a martingale with respect to the filtration $\mathcal{G}_t = \sigma(W_s, s \leq t)$.
ANS: By Exercise 4.2.5 we know that $e^{W_t - t/2}$ is a martingale, hence we take $f(t) = t/2$. We assume that the increasing process in the Doob-Meyer decomposition has the form $A_t = \int_0^t e^{2W_u - h(u)}du$. Clearly, $A_0 = 0$. Also, A_t has continuous sample paths, since W_t does; A_t depends only on the values of W_u for $0 \leq u \leq t$ so it is $\{\mathcal{G}_t\}$-adapted; and A_t is nondecreasing since $e^x > 0$. Further, $\mathbf{E}[e^{2W_u}|\mathcal{G}_s] = e^{2W_s + 2(u-s)}$ for all $u \geq s$, hence $\mathbf{E}[X_t^2|\mathcal{G}_s] = e^{2W_s - 2s + t}$ and by Fubini's theorem also,

$$\mathbf{E}[A_t - A_s|\mathcal{G}_s] = \int_s^t \mathbf{E}[e^{2W_u - h(u)}|\mathcal{G}_s]du = e^{2W_s - 2s}\int_s^t e^{2u - h(u)}du,$$

when $t \geq s$. The remaining condition of $(X_t^2 - A_t, \mathcal{G}_t)$ a martingale thus amounts to

$$\mathbf{E}[X_t^2 - (A_t - A_s)|\mathcal{G}_s] - X_s^2 = e^{2W_s - 2s}(e^t - e^s - \int_s^t e^{2u - h(u)}du) = 0,$$

which evidently holds for $h(u) = u$. In conclusion, the increasing part associated with the MG $(X_t, \mathcal{G}_t)$ is $A_t = \int_0^t e^{2W_s - s}ds$.

2. Exercise 4.5.4. Consider an urn that at stage 0 contains one red ball and one blue ball. At each stage a ball is drawn at random from the urn, with all possible choices being equally likely, and it and one more ball of the same color are then returned to the urn. Let R_n denote the number of red balls at stage n and $M_n = R_n/(n+2)$ the corresponding fraction of red balls.

(a) Find the law of R_{n+1} conditioned on $R_n = k$ and use it to compute $\mathbf{E}(R_{n+1}|R_n)$.
ANS: At time n, there are k red balls and $(n+2-k)$ blue balls if $R_n = k$. So we have that R_{n+1} can only take the values k and $k+1$ with non-zero probabilities $(n+2-k)/(n+2)$ and $k/(n+2)$, respectively. Thus, $\mathbf{E}(R_{n+1}|R_n) = (n+2-R_n)R_n/(n+2) + R_n(R_n+1)/(n+2) = \frac{n+3}{n+2}R_n$.

(b) Check that M_n is a martingale with respect to its canonical filtration.
ANS: We have that M_n is bounded so it is integrable. Note that the canonical filtration $\mathcal{G}_n$ for $\{M_n\}$ is the same as that of $\{R_n\}$. Further, per fixed given value of R_n, the value of R_{n+1} is independent of $(R_0, R_1, \ldots, R_{n-1})$. Hence,

$$\mathbf{E}(M_{n+1}|\mathcal{G}_n) = \frac{1}{n+3}\mathbf{E}(R_{n+1}|\mathcal{G}_n) = \frac{1}{n+3}\mathbf{E}(R_{n+1}|R_n) = \frac{1}{n+2}R_n = M_n,$$

so M_n is indeed a martingale with respect to its canonical filtration.

(c) Applying Proposition 4.5.3 conclude that $M_n \to M_\infty$ in L^2 and that $\mathbf{E}(M_\infty) = \mathbf{E}(M_0) = 1/2$.

ANS: Since $0 \le M_n \le 1$, we have $\mathbf{E}M_n^2 \le 1, n = 1, 2, \ldots$. By Proposition 4.5.3 there exists a R.V. M_∞ such that $M_n \to M_\infty$ a.s. and in L^2. Consequently, as shown for example in Exercise 1.3.21 (or by the bounded convergence of Corollary 1.4.29), $\mathbf{E}(M_\infty) = \lim_{n\to\infty} \mathbf{E}(M_n) = \mathbf{E}(M_0) = 1/2$.

(d) Using Doob's (maximal) inequality show that $\mathbf{P}(\max_{k\ge1} M_k > 3/4) \le 2/3$.

ANS: By part (c) and Doob's inequality, we have that $\mathbf{P}(\max_{k\ge1} M_k > 3/4) \le (4/3)\mathbf{E}(M_\infty) = 2/3$.

3. Exercise 4.6.8. Suppose $\{Z_n\}$ is a branching process with $\mathbf{P}(N = 1) < 1$ and $Z_0 = 1$. Show that

$$\mathbf{P}(\lim_{n\to\infty} Z_n = \infty) = 1 - p_{ex},$$

first in case $m \le 1$, then in case $\mathbf{P}(N = 0) = 0$ and finally using the preceding exercise, for $m > 1$ and $\mathbf{P}(N = 0) > 0$.

ANS: Since $\mathbf{P}(N = 1) < 1$ we have by Propositions 4.6.3 and 4.6.5 that $p_{ex} = 1$ when $m \le 1$. That is, in this case w.p.1. $Z_n = 0$ for all n large enough, yielding the stated claim.

In contrast, if $\mathbf{P}(N = 0) = 0$ then Z_n is non-decreasing, so $p_{ex} = 0$. Further, in this case Z_n is bounded only if $N_1^{(k)} = 1$ for all k large enough, which with $\mathbf{P}(N = 1) < 1$ occurs with zero probability, again resulting with the stated claim.

Finally, for $m > 1$ and $\mathbf{P}(N = 0) > 0$ we have from Exercise 4.6.7 that $p_{ex} = \rho \in (0, 1)$ and further,

$$1 - p_{ex} = 1 - \mathbf{P}(M_\infty = 1) = \mathbf{P}(M_\infty = 0) = \mathbf{P}(\lim_{n\to\infty} \rho^{Z_n} = 0) = \mathbf{P}(\lim_{n\to\infty} Z_n = \infty)\,,$$

as claimed.

4. Exercise 4.6.9. Let $\{Z_n\}$ be a branching process with $Z_0 = 1$. Compute p_{ex} in each of the following situations and specify for which values of the various parameters the extinction is certain.

(a) The offspring distribution satisfies, for some $0 < p < 1$,

$$\mathbf{P}(N = 0) = p,\ \mathbf{P}(N = 2) = 1 - p.$$

ANS: We have that $m = \mathbf{E}N = 2(1 - p)$ with $m \le 1$ if and only if $p \ge 1/2$. Thus, $p_{ex} = 1$ when $p \ge 1/2$ by Proposition 4.6.3 (for $p > 1/2$) and Proposition 4.6.5 (for $p = 1/2$, applicable since here $\mathbf{P}(N = 1) = 0 < 1$). Finally, if $p < 1/2$ then $m > 1$ so $\{Z_n\}$ is super-critical with $\mathbf{P}(N = 0) = p > 0$. We have shown in Exercise 4.6.7 that in this case p_{ex} is the unique solution in $(0, 1)$ of

$$0 = x - \phi(x) = x - \mathbf{P}(N = 0) - \mathbf{P}(N = 2)x^2 = x - p - (1 - p)x^2$$

(taking the function $\phi(x)$ per equation (4.6.2) that corresponds to our law of N). As

$$x - p - (1 - p)x^2 = (1 - p)(1 - x)(x - p/(1 - p))\,,$$

we conclude that $p_{ex} = p/(1 - p) < 1$ when $p < 1/2$.

(b) The offspring distribution is (shifted) Geometric, i.e. for some $0 < p < 1$,

$$\mathbf{P}(N = k) = p(1 - p)^k,\ k = 0, 1, 2, \ldots$$

ANS: We have now that $m = \mathbf{E}N = \sum_{k=1}^\infty kp(1 - p)^k = (1 - p)/p$ (where to get the last identity differentiate in p the identity $\sum_{k=0}^\infty (1 - p)^k = 1/p$ and multiply both sides by $-p(1 - p)$). As in part (a), if $p \ge 1/2$ then $m \le 1$ and consequently $p_{ex} = 1$ (for here too $\mathbf{P}(N = 1) = p(1 - p) < 1$). In contrast, $p < 1/2$ yields a super-critical branching process with $\mathbf{P}(N = 0) = p > 0$, so again from Exercise 4.6.7 we have that p_{ex} is the unique solution in $(0, 1)$ of

$$0 = x - \phi(x) = x - \sum_{k=0}^\infty \mathbf{P}(N = k)x^k = x - p\sum_{k=0}^\infty (1 - p)^k x^k = x - \frac{p}{1 - (1 - p)x}\,.$$

Thus, p_{ex} is the unique root in $(0, 1)$ of the quadratic equation

$$0 = x(1 - (1 - p)x) - p = x - p - (1 - p)x^2\,,$$

and as you have seen in part (a), it follows that $p_{ex} = p/(1-p)$. Thus, though the law of N in part (b) is different from its law in part (a), both result with same values of p_{ex} (for all choices of p).

5. Exercise 5.3.10. Suppose $(W_t, \mathcal{F}_t)$ satisfies Lévy's characterization of the Brownian motion. Namely, it is a square-integrable martingale of right-continuous filtration and continuous sample path such that $(W_t^2 - t, \mathcal{F}_t)$ is also a martingale. Suppose X_t is a bounded $\mathcal{F}_t$-adapted simple process. That is,

$$X_t = \eta_0 \mathbf{1}_{\{0\}}(t) + \sum_{i=0}^\infty \eta_i \mathbf{1}_{(t_i, t_{i+1}]}(t),$$

where the non-random sequence $t_k > t_0 = 0$ is strictly increasing and unbounded (in k), while the (discrete time) S.P. $\{\eta_n\}$ is uniformly (in n and ω) bounded and adapted to $\mathcal{F}_{t_n}$. Provide an explicit formula for $A_t = \int_0^t X_u^2 du$, then show that both

$$I_t = \sum_{j=0}^{k-1} \eta_j (W_{t_{j+1}} - W_{t_j}) + \eta_k(W_t - W_{t_k}),\ \text{when } t \in [t_k, t_{k+1}),$$

and $I_t^2 - A_t$ are martingales with respect to $\mathcal{F}_t$ and explain why this implies that $\mathbf{E}I_t^2 = \mathbf{E}A_t$ and $V_t^{(2)}(I) = A_t$.

ANS: Since the intervals $(t_i, t_{i+1}]$ are pairwise disjoint,

$$X_t^2 = \eta_0^2 \mathbf{1}_{\{0\}}(t) + \sum_{i=0}^\infty \eta_i^2 \mathbf{1}_{(t_i, t_{i+1}]}(t).$$

Consequently,

$$A_t = \int_0^t X_u^2 du = \sum_{j=0}^{k-1} \eta_j^2 (t_{j+1} - t_j) + \eta_k^2(t - t_k),\ \text{when } t \in [t_k, t_{k+1}),$$

Next note that I_t is adapted to $\mathcal{F}_t$ (on account of the adaptedness of $\{\eta_n\}$ to $\mathcal{F}_{t_n}$ and that of $\{W_t\}$ to $\{\mathcal{F}_t\}$), and is integrable (for each summand is integrable due to boundedness of η_n and the integrability of W_t). Further, by the tower property, if $(I_t - I_{t_k}, \mathcal{F}_t)$ satisfies the martingale property for $t \in [t_k, t_{k+1}]$ and each fixed k then also $(I_t, \mathcal{F}_t)$ is a martingale (for all $t \ge 0$). Fixing k and $t_k \le s < t \le t_{k+1}$, note that taking out η_k which is measurable on $\mathcal{F}_{t_k} \subseteq \mathcal{F}_s$, we get by the martingale property of $(W_t, \mathcal{F}_t)$ that

$$\mathbf{E}[(I_t - I_{t_k}) - (I_s - I_{t_k})|\mathcal{F}_s] = \mathbf{E}[\eta_k(W_t - W_s)|\mathcal{F}_s] = \eta_k(\mathbf{E}[W_t|\mathcal{F}_s] - W_s) = 0,$$

as needed for proving that $(I_t, \mathcal{F}_t)$ is a martingale.

Similarly, note that $J_t = I_t^2 - A_t$ is $\mathcal{F}_t$-adapted and integrable (on account of square integrability of $\{W_t\}$ and boundedness of η_n). As before, to show that $(J_t, \mathcal{F}_t)$ is a martingale it suffices to verify the martingale property for $(J_t - J_{t_k}, \mathcal{F}_t)$ with $t \in [t_k, t_{k+1}]$ and k fixed. To this end, note that

$$J_t - J_{t_k} = 2I_{t_k}(I_t - I_{t_k}) + \eta_k^2[(W_t - W_{t_k})^2 - (t - t_k)]\,,$$

and recall that we have shown this property already for $(I_t - I_{t_k}, \mathcal{F}_t)$. Since I_{t_k} is measurable on $\mathcal{F}_{t_k} \subseteq \mathcal{F}_t$, the same applies for $(I_{t_k}(I_t - I_{t_k}), \mathcal{F}_t)$. Further, η_k^2 is also measurable on $\mathcal{F}_{t_k} \subseteq \mathcal{F}_t$ and by the preceding, linearity of the C.E. and taking out what is known, we only need to verify that $(\widehat{W}_u^2 - u, \mathcal{F}_{t_k+u})$ has the martingale property for $\widehat{W}_u = W_{t_k+u} - W_{t_k}$ and $0 \le u \le t_{k+1} - t_k$. This in turn follows from our assumption that W_t is Brownian motion (with respect to $\mathcal{F}_t$), hence by Proposition 5.2.3 so is $\widehat{W}_u$ (now with respect to $\mathcal{F}_{t_k+u}$).

Clearly, the martingale property of J_t implies that $\mathbf{E}J_t = \mathbf{E}J_0 = 0$, that is $\mathbf{E}I_t^2 = \mathbf{E}A_t$. We have proved that both $(I_t, \mathcal{F}_t)$ and $(I_t^2 - A_t, \mathcal{F}_t)$ are martingales of continuous sample path and right-continuous filtration, with $A_0 = 0$ and $t \mapsto A_t$ non-decreasing. Thus, A_t is the increasing process associated with $(I_t, \mathcal{F}_t)$ via the Doob-Meyer decomposition. As stated in Corollary 5.3.5, A_t must then be also the quadratic variation $V_t^{(2)}(I)$ of the "stochastic integral" I_t.

6. Exercise 5.3.14. Consider the stochastic process $Y(t) = W(t)^2$, for $0 \le t \le 1$, with $W(t)$ a Brownian motion.

(a) Show that for any $\gamma < 1/2$ the sample path of $Y(t)$ is locally Hölder continuous of exponent γ with probability one.

ANS: Suppose that f is a function on $[0, 1]$ that is locally Hölder continuous of exponent $\gamma > 0$. Then, the same holds for the function f^2. Indeed, here $M = \sup_{x\in[0,1]} f(x)$ is finite (since f is continuous on the bounded interval $[0, 1]$) and we have for all $s, t \in [0, 1]$ that

$$\frac{|f^2(t) - f^2(s)|}{|t - s|^\gamma} = |f(t) + f(s)|\frac{|f(t) - f(s)|}{|t - s|^\gamma} \le 2M\frac{|f(t) - f(s)|}{|t - s|^\gamma}.$$

Our claim follows from this fact in view of part (a) of Exercise 5.1.12 (in case $H = 1/2$ there; see also Exercise 5.3.7).

Alternatively, noting that $Y(s + h) - Y(s) = (2Z + X)X$ for the independent Gaussian $X = W(s + h) - W(s)$ and $Z = W(s)$ of zero mean and variances h and s, respectively, it is not hard to show that $\mathbf{E}[(Y(s + h) - Y(s))^{2p}] \le C(p)h^p$ for any positive integer p, some finite $C(p)$ and all $h, s \in [0, 1]$. The claim then follows by an application of Kolmogorov's continuity theorem (where $\gamma < \beta/\alpha = (p - 1)/2p$ once p is large enough).

(b) Compute $\mathbf{E}V_{(\pi)}^{(2)}(Y)$ for a finite partition π of $[0, t]$ to k intervals, and find its limit as $\|\pi\| \to 0$.

ANS: Using notations of part (a) it is not hard to check that for all $h > 0$, $s \ge 0$,

$$(Y(s + h) - Y(s))^2 = 4Z^2X^2 + 4ZX^3 + X^4\,,$$

where by independence of $X \sim N(0, h)$ and $Z \sim N(0, s)$ it follows that

$$\mathbf{E}[|Y(s + h) - Y(s)|^2] = 4\mathbf{E}Z^2\mathbf{E}X^2 + 4\mathbf{E}Z\mathbf{E}X^3 + \mathbf{E}X^4 = 4sh + 3h^2\,.$$

With $\Delta t_i = t_{i+1} - t_i$ for the partition $\pi = \{0 = t_0, t_1, \ldots, t_k = t\}$, using this identity and the linearity of the expectation we have

$$\mathbf{E}V_{(\pi)}^{(2)}(Y) = \mathbf{E}[\sum_{i=0}^{k-1}(Y(t_{i+1}) - Y(t_i))^2] = 4\sum_{i=0}^{k-1} t_i\Delta t_i + 3\sum_{i=0}^{k-1} \Delta t_i^2$$

Note that $\sum_{i=0}^{k-1} t_i\Delta t_i$ is a Riemann sum approximation for the integral $\int_0^t udu = t^2/2$ that thus converges to $t^2/2$ as $||\pi|| \to 0$. Further, with

$$\sum_{i=0}^{k-1} \Delta t_i^2 \le ||\pi|| \sum_{i=0}^{k-1} \Delta t_i = ||\pi||t \to 0$$

as $||\pi|| \to 0$, we conclude that $\mathbf{E}V_{(\pi)}^{(2)}(Y)$ converges to $2t^2$ in the limit $||\pi|| \to 0$.

Indeed, we note in passing that for the partition π of $[0, t]$ to k intervals of equal length $\Delta t_i = t/k$, i.e. taking $t_i = i(t/k)$, the expectation we consider simplifies to $2t^2 + t^2/k$ that evidently converges to $2t^2$ as $k \to \infty$.

(c) Show that the total variation of $Y(t)$ on the interval $[0, 1]$ is infinite.

ANS: In Proposition 5.3.12 we saw that the Brownian motion has infinite total variation in any fixed interval $[a, b]$, $b > a$. Inside any open interval (a, b) there is a sub-interval $[r, q]$ with $q > r$ rational numbers. As there are only countably many such sub-intervals, we deduce that there exists an event A of probability one such that if $\omega \in A$ then $t \mapsto W(t, \omega)$ is continuous and has an infinite total variation in every open interval in $(0, 1)$. In particular, fixing $\omega \in A$ this implies that there exists $t \in (0, 1)$ such that $W(t, \omega) \ne 0$ (since otherwise $W(t, \omega) = 0$ for all t and such a path would have finite total variation). Fixing such $t = t(\omega)$ we assume with out loss of generality that $W(t, \omega) = 2\delta > 0$ and note that by continuity of the sample path there exists $\epsilon > 0$ such that $|W(s, \omega) - W(t, \omega)| \le \delta$ for all $s \in (t - \epsilon, t + \epsilon)$. This implies that for any s_1 and s_2 in $(t - \epsilon, t + \epsilon)$ we have $W(s_1, \omega) + W(s_2, \omega) \ge 2\delta$ and hence

$$|Y(s_1, \omega) - Y(s_2, \omega)| = |W(s_1, \omega) + W(s_2, \omega)||W(s_1, \omega) - W(s_2, \omega)| \ge 2\delta|W(s_1) - W(s_2)|.$$

It follows that the total variation of $Y(s, \omega)$ on the interval $(t - \epsilon, t + \epsilon)$ is bounded below by 2δ times the total variation of $W(s, \omega)$ on the same interval. Our claim follows since we already know that the latter quantity is infinite for $\omega \in A$.