Lecture Notes on Gaussian Processes with Examples

József Lőrinczi

Alfréd Rényi Institute of Mathematics

Budapest, Hungary

lorinczi@renyi.hu

Contents

1 Background material

1.1 Random variables

Probability space Let $\Omega \neq \emptyset$ be a given set in our context called *sample space*, and let $\mathcal{F}$ be a σ-field, i.e., a collection of subsets of Ω such that

(1) $\emptyset \in \mathcal{F}$,

(2) for every $A \in \mathcal{F}$ also $A^c := \Omega \setminus A \in \mathcal{F}$,

(3) for every $A_1, A_2, ... \in \mathcal{F}$ also $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$.

Properties (1) and (2) above immediately imply that $\Omega \in \mathcal{F}$, while a combination of (2) and DeMorgan's formulae imply $\cap_{i=1}^{\infty} A_i \in \mathcal{F}$. The elements of $\mathcal{F}$ are called *$\mathcal{F}$-measurable sets*, and the pair $(\Omega, \mathcal{F})$ is a *measurable space*. When Ω is a topological space with topology $\tau(\Omega)$, the Borel field $\mathcal{B}(\Omega)$ is the σ-field generated by $\tau(\Omega)$, and $(\Omega, \mathcal{B}(\Omega))$ is called a *Borel measurable space*.

Given a measurable space $(\Omega, \mathcal{F})$, a *probability measure* $P : \mathcal{F} \to [0, 1]$ is a set-function with the properties

(i) $P(\Omega) = 1$,

(ii) for any $(A_n)_{n \in \mathbb{N}} \subset \mathcal{F}$ such that $A_j \cap A_k = \emptyset$ whenever $j \neq k$, the property $P(\cup_{j=1}^{\infty} A_j) = \sum_{j=1}^{\infty} P(A_j)$ holds.

The triple $(\Omega, \mathcal{F}, P)$ is a *probability space*. The integral with respect to P of a Borel measurable function h is called *expectation* of h and we write

$$\int_{\Omega} h(\omega) dP(\omega) = \mathbb{E}_P[h].$$

Independence Two events $A, B \in \mathcal{F}$ are independent if and only if $P(A \cap B) = P(A)P(B)$. Two sub-σ-fields $\mathcal{F}_1, \mathcal{F}_2$ of $\mathcal{F}$ are independent if every $A \in \mathcal{F}_1$ and $B \in \mathcal{F}_2$ are pairwise independent. The following result is for an infinite number of events $A_1, A_2, ... \in \mathcal{F}$ and is fundamental in probability theory. Recall that the event $\limsup_{n \to \infty} A_n = \cap_{m=1}^{\infty} \cup_{n \geq m} A_n$ corresponds to the set of outcomes $\omega \in \Omega$ which occur infinitely many times (infinitely often) in the given infinite sequence of events.

Theorem 1.1 (Borel-Cantelli Lemma) *Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $(A_n)_{n \in \mathbb{N}} \subset \mathcal{F}$ be given.*

(1) If $\sum_{n=1}^{\infty} P(A_n) < \infty$, then $P(\limsup_{n \to \infty} A_n) = 0$.

(2) If $\sum_{n=1}^{\infty} P(A_n) = \infty$ and $(A_n)_{n \in \mathbb{N}}$ are independent, then $P(\limsup_{n \to \infty} A_n) = 1$.

Continuous random variables Let $(\Omega_1, \mathcal{F}_1), (\Omega_2, \mathcal{F}_2)$ be given measure spaces. A function $f : \Omega_1 \to \Omega_2$ is called *measurable* with respect to $\mathcal{F}_1$ and $\mathcal{F}_2$, denoted $f \in \mathcal{F}_1/\mathcal{F}_2$, if for every $A \in \mathcal{F}_2$ it follows that $f^{-1}(A) \in \mathcal{F}_1$. When $\mathcal{F}_2 = \mathcal{B}(\mathbb{R})$, then f is a real-valued *Borel measurable* function. Let $(\Omega_1, \mathcal{F}_1, P_1), (\Omega_2, \mathcal{F}_2, P_2)$ be given probability spaces. A function $X : \Omega_1 \to \Omega_2$, $f \in \mathcal{F}_1/\mathcal{F}_2$ is called an Ω_2-valued *random variable*. We will most of the time consider the case of $\Omega_1 = \Omega$, $\mathcal{F}_1 = \mathcal{F}$, with given Ω, and $\Omega_2 = \mathbb{R}$ so that

$$X : \Omega \to \mathbb{R} \quad \text{such that } X^{-1}(E) = \{\omega \in \Omega : X(\omega) \in E\} \in \mathcal{F}.$$

X is then a *real-valued random variable*. Intuitively, the inverse map identifies the sample points in Ω on which the observation of event A depends. Let X be a real-valued random variable. The measure P_X on $(\Omega, \mathcal{F})$ defined by

$$P_X(A) = P(X^{-1}(A)), \quad A \in \mathcal{F},$$

is called the *image measure* or the *probability distribution* of P under X. In many cases of interest of real-valued random variables X the probability measure P_X is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}$, i.e., it has a density (or Radon-Nikodým derivative) with respect to Lebesgue measure. We say that the absolutely integrable function $f_X : \mathbb{R} \to [0, \infty)$ is the *probability density function (pdf)* of the real-valued random variable X if

$$P_X(A) = \int_A f_X(x)dx, \quad \forall A \in \mathcal{F},$$

and equivalently write $dP_X(x) = f_X(x)dx$. Whenever the function $F_X(x) = P(X \leq x)$, called *probability distribution function* of X, is differentiable, we have the expression

$$f_X(x) = \frac{dF_X(x)}{dx}.$$

For two real-valued random variables X, Y not necessarily defined on the same probability space, we use the notation $X \sim Y$ whenever they are *identically distributed*, i.e., $P_X = P_Y$.

Two real-valued random variables X, Y on the same probability space are independent whenever the events $\{X \leq a\}$ and $\{Y \leq b\}$ are pairwise independent for all $a, b \in \mathbb{R}$.

Moments By the definition of the distribution of r.v. $X : \Omega \to \mathbb{R}$ we have the expectation

$$\mathbb{E}_P[h(X)] = \int_\Omega h(\omega)dP(\omega) = \int_\mathbb{R} h(x)dP_X(x) = \int_\mathbb{R} h(x)d(P \circ X^{-1})(x)$$

for any bounded Borel measurable function $h : \mathbb{R} \to \mathbb{R}$. Let X be a continuous real-valued random variable and $k \in \mathbb{N}$. The number

$$\mathbb{E}_P[X^k] = \int_\mathbb{R} x^k dP_X(x), \quad k \in \mathbb{N},$$

is called the *kth moment* of X whenever the integral $\int_\mathbb{R} |x^k| dP_X(x)$ exists. The first moment

$$\mathbb{E}_P[X] = \int_\mathbb{R} x dP_X(x)$$

is called the *expected value* or *mean* of X under the probability measure P_X. The *kth central moment* is

$$\mathbb{E}_P[(X - \mathbb{E}_P[X])^k] = \int_{\mathbb{R}} (x - \mathbb{E}_P[X])^k dP_X(x).$$

The second central moment

$$\operatorname{var} X = \mathbb{E}_P[(X - \mathbb{E}_P[X])^2] = \mathbb{E}_P[X^2] - \mathbb{E}_P[X]^2.$$

is called *variance* and its positive square root is called *standard deviation*. The fact that the variance is non-negative implies the *Cauchy-Schwarz inequality*

$$\mathbb{E}_P[X]^2 \leq \mathbb{E}_P[X^2]$$

for any random variable X. Note that $\operatorname{var} X = 0$ if and only if X is P-a.s. constant. The *covariance* of two real-valued random variables X and Y on the same probability space is defined by

$$\operatorname{cov}(X, Y) = \mathbb{E}_P[(X - \mathbb{E}_P[X])(Y - \mathbb{E}_P[Y])] = \mathbb{E}_P[XY] - \mathbb{E}_P[X]\mathbb{E}_P[Y],$$

which reduces to the variance when $X = Y$, i.e., $\operatorname{cov}(X, X) = \operatorname{var} X$.

Characteristic functions The Fourier transform

$$\widehat{P}_X(u) := \phi_X(u) := \mathbb{E}_P[e^{iuX}] = \int_{\mathbb{R}} e^{iux} dP_X(x)$$

of P_X is the *characteristic function* of X. The characteristic function is well defined at every point since

$$|\phi_X(u)| = \left|\mathbb{E}_P[e^{iuX}]\right| \leq \mathbb{E}_P[|e^{iuX}|] = 1.$$

Since ϕ_X is infinitely many times differentiable, it follows that

$$\mathbb{E}_P[X^k] = (-i)^k \phi_X^{(k)}(0), \quad k = 1, 2, ...$$

holds. By inverse Fourier transform it is possible to compute the probability density in terms of the characteristic function. As a result it can be proven for random variables X and Y that the characteristic functions $\phi_X = \phi_Y$ if and only if $X \sim Y$. Furthermore, X and Y are independent if and only if $\phi_{(X,Y)}(u, v) = \phi_X(u)\phi_Y(v)$, for all u, v. If X, Y are independent, then $\phi_{X+Y}(u) = \phi_X(u)\phi_Y(u)$. Also, if X, Y are independent, then $\operatorname{cov}(f(X), g(Y)) = 0$ for all Borel measurable functions f, g.

Definition 1.1 (Convergence of random variables) *Let $(X_n)_{n\geq 1}$ be a sequence of real valued random variables and X another real-valued random variable, all on a given probability space $(\Omega, \mathcal{F}, P)$. We speak of convergence of the sequence in the following senses:*

(1) $X_n \overset{\text{a.s.}}{\to} X$, i.e., X_n is convergent almost surely *to X if*

$$P\left(\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}\right) = 1.$$

(2) $X_n \overset{L^p}{\to} X$*, i.e.,* X_n *is convergent* in L^p-sense *to* X *for* $1 \le p < \infty$ *if*

$$\lim_{n\to\infty} \mathbb{E}_P\left[|X_n(\omega) - X(\omega)|^p\right] = 0.$$

(3) $X_n \overset{P}{\to} X$*, i.e.,* X_n *is convergent* in probability *to* X *if for every* $\varepsilon > 0$*,*

$$\lim_{n\to\infty} P\left(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| \ge \varepsilon\}\right) = 0.$$

(4) $X_n \overset{\mathrm{d}}{\to} X$*, i.e.,* X_n *is convergent* in distribution (*or* weakly) *to* X *if*

$$\lim_{n\to\infty} P\left(\{\omega \in \Omega : X_n(\omega) \le x\}\right) = P\left(\{\omega \in \Omega : X(\omega) \le x\}\right)$$

for every $x \in \mathbb{R}$ *at which the probability distribution function* $P(X \le x)$ *is continuous.*

Theorem 1.2 *Let* $(X_n)_{n\ge 1}$ *be a sequence of real-valued random variables and* X *another real-valued random variable, all on* $(\Omega, \mathcal{F}, P)$*.*

(1) If $X_n \overset{a.s.}{\to} X$ *or* $X_n \overset{L^p}{\to} X$ *for some* $p \ge 1$*, then* $X_n \overset{P}{\to} X$ *as* $n \to \infty$*.*

(2) If $X_n \overset{P}{\to} X$*, then* $X_n \overset{\mathrm{d}}{\to} X$ *as* $n \to \infty$*.*

(3) If $X_n \overset{P}{\to} X$*, then there exists a subsequence* $(X_{n_k})_{k\ge 1}$ *such that* $X_{n_k} \overset{a.s.}{\to} X$ *as* $n \to \infty$*.*

(4) If $X_n \overset{P}{\to} X$ *and* $(X_n)_{n\ge 1}$ *is uniformly integrable, i.e.,*

$$\lim_{N\to\infty} \sup_{n\in\mathbb{N}} \mathbb{E}_P\left[|X_n| 1_{\{|X_n| \ge N\}}\right] = 0,$$

then $X_n \overset{L^1}{\to} X$ *as* $n \to \infty$*.*

Converse statements to those in Theorem 1.2 are provided by two important standard theorems.

Theorem 1.3 (Dominated convergence) *Let* $X_n \overset{P}{\to} X$ *as* $n \to \infty$*. Suppose there is a random variable* Y *such that* $|X_n| \le Y$*,* $n \in \mathbb{N}$*, and* $\mathbb{E}[Y] < \infty$*. Then* $X_n \overset{L^1}{\to} X$ *as* $n \to \infty$*, and* $X \in L^1(\Omega, dP)$*.*

Theorem 1.4 (Monotone convergence) *Let* $X_n \overset{\text{a.s.}}{\to} X$ *as* $n \to \infty$*. Suppose the sequence* $(X_n)_{n\ge 1}$ *is monotone, and there is* $M > 0$ *such that* $\mathbb{E}[X_n] < M$*, for all* $n \in \mathbb{N}$*. Then* $\mathbb{E}_P[X_n] \to \mathbb{E}_P[X]$ *as* $n \to \infty$*, and* $X \in L^1(\Omega, dP)$*.*

As applications of Theorem 1.2 it is possible to prove standard limit theorems of sequences of random variables.

Theorem 1.5 (Limit theorems) *Let* $(X_n)_{n\geq 1}$ *be a sequence of iid real-valued random variables on* $(\Omega, \mathcal{F}, P)$. *Suppose* $\mathbb{E}[X_n] = \mu < \infty$ *and* $\operatorname{var} X_n = \sigma^2 < \infty$, *and consider* $S_n = X_1 + ... + X_n$.

(1) Central Limit Theorem:

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{\mathrm{d}} N(0,1) \quad \text{as } n \to \infty.$$

(2) Weak Law of Large Numbers:

$$\frac{S_n}{n} \xrightarrow{P} \mu \quad \text{as } n \to \infty.$$

(3) Strong Law of Large Numbers: $\mathbb{E}[|X_n|] < \infty$ *and* $\mathbb{E}[X_n] = \mu$, $n \in \mathbb{N}$, *if and only if*

$$\frac{S_n}{n} \xrightarrow{a.s.} \mu \quad \text{as } n \to \infty.$$

1.2 Random processes

Definition 1.2 (Random process) *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space and* S, T *non-empty sets. A family of* S*-valued random variables* $(X_t)_{t\in T}$ *is called a* random process with index set T.

We often use $S = \mathbb{R}^n$ for the set of values, in which case we speak of a *real-valued random process.* If the index set T is countable, then we speak of a *discrete time random process*, if T is uncountable, then we speak of a *continuous time random process.* Usually, we think of the elements of T as time, so intuitively a random process describes the time evolution of a random variable. For any fixed $t \in T$, the map $\omega \mapsto X_t(\omega)$ is a random variable, while for any fixed $\omega \in \Omega$, the map $t \mapsto X_t(\omega)$ is a function called *path.* This distinction justifies for any random process to address *distributional properties* on the one hand, and *path properties* on the other.

Definition 1.3 (Filtration) *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space and* $(\mathcal{F}_s)_{s\geq 0}$ *a family of sub-*σ*-fields of* $\mathcal{F}$. *The collection* $(\mathcal{F}_s)_{s\geq 0}$ *is called a* filtration *whenever* $\mathcal{F}_s \subset \mathcal{F}_t$ *for* $s \leq t$. *The given measurable space endowed with a filtration is called a* filtered space.

Intuitively, $\mathcal{F}_t$ contains the information known to an observer at time t. A basic example is the *natural filtration* $(\mathcal{F}_t^X)_{t\geq 0}$ given by $\mathcal{F}_t^X = \sigma(X_s, 0 \leq s \leq t)$, where $\sigma(X)$ stands for the minimal σ-field such that X is measurable. Intuitively, $\mathcal{F}_t^X$ contains the information obtained by observing X up to time t.

Definition 1.4 (Adapted process) *Let* $(\Omega, \mathcal{F})$ *be a filtered space,* $(\mathcal{F}_s)_{s\geq 0}$ *a given filtration and* $X = (X_t)_{t\geq 0}$ *a random process. The process* X *is called* $(\mathcal{F}_s)_{s\geq 0}$-adapted *whenever* X_t *is* $\mathcal{F}_t$*-measurable for every* $t \geq 0$.

If a process $(X_t)_{t\geq 0}$ is $(\mathcal{F}_t)_{t\geq 0}$-adapted it means that the process does not carry more information at time t than $\mathcal{F}_t$. Obviously, $(X_t)_{t\geq 0}$ is adapted to its natural filtration $(\mathcal{F}_t^X)_{t\geq 0}$. (Equivalently, the natural filtration $(\mathcal{F}_t^X)_{t\geq 0}$ is the smallest filtration making $(X_t)_{t\geq 0}$ adapted.)

Definition 1.5 (Markov process) *Let $(X_t)_{t\geq 0}$ be an adapted process on a given filtered space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$ The process $(X_t)_{t\geq 0}$ is called a* Markov process *with respect to $(\mathcal{F}_t)_{t\geq 0}$ whenever*

$$\mathbb{E}_P\left[f(X_t)|\mathcal{F}_s\right] = \mathbb{E}_P\left[f(X_t)|\sigma(X_s)\right], \quad 0 \leq s \leq t,$$

for all bounded Borel measurable functions f.

Markov processes can be characterized by their probability transition kernels.

Definition 1.6 (Probability transition kernel) *Let $(X_t)_{t\geq 0}$ be a real valued random process. A map $\mathbb{R}^+ \times \mathbb{R}^+ \times \mathbb{R} \times \mathcal{B}(\mathbb{R}) \ni (s,t,x,A) \mapsto p(s,t,x,A) \in \mathbb{R}$ is called a* probability transition kernel *if*

1. *$\forall x \in \mathbb{R}$, $s,t \geq 0$ the map $A \mapsto p(s,t,x,A)$ is a probability measure*
2. *$\forall A \in \mathcal{B}(\mathbb{R})$, $s,t \geq 0$ the map $x \mapsto p(s,t,x,A)$ is a Borel measurable function*
3. *for all $0 \leq r \leq s \leq t$ the* Chapman-Kolmogorov equality *holds, i.e.,*

$$\int_{-\infty}^{\infty} p(s,t,y,A)p(r,s,x,dy) = p(r,t,x,A).$$

The probability transition kernels may have densities $p(s,t,x,y)$ in the form

$$p(s,t,x,A) = \int_A p(s,t,x,y)dy.$$

Also, if a process is stationary, then $p(s,t,x,y) = p(|t-s|,x,y)$, and we write in general $p_t(x,y)$. Then the Chapman-Kolmogorov equality reduces to

$$\int_{-\infty}^{\infty} p_s(x,y)p_t(y,z)dz = p_{s+t}(x,y), \quad x,y \in \mathbb{R},\ s,t \geq 0.$$

The interpretation of the Markov property is that the future values of the process only depend on its present and not on its past values.

Let S be a set equipped with a σ-field $\mathcal{S}$. The *canonical realization* of an S-valued Markov process $(X_t)_{t\geq 0}$ is a random process on the space $(S^{[0,\infty)}, \mathcal{S}^{[0,\infty)}, P_\nu)$, where

$$S^{[0,\infty)} := \{\omega : \mathbb{R}^+ \to S\}$$
$$\mathcal{S}^{[0,\infty)} := \sigma\left(\omega(t) : t \in \mathbb{R}^+\right)$$

and P_ν is the unique probability measure on $(S^{[0,\infty)}, \mathcal{S}^{[0,\infty)})$ whose finite dimensional distributions are given by

$$P_\nu(X_0 \in A_0, ..., X_n \in A_n) = \int_{\mathbb{R}^{n+1}} \left(\prod_{i=1}^n \mathbf{1}_{A_i}(x_i) \right) \left(\prod_{i=1}^n p(t_{i-1}, t_i, x_{i-1}, dx_i) \right) \mathbf{1}_{A_0}(x_0)\nu(dx_0) \tag{1.1}$$

for all $0 = t_0 < t_1 < \ldots < t_n$, $n \in \mathbb{N}$, and with the notation $X_k := X_{t_k}$. The measure $\nu(A) = P(X_0 \in A)$ is the *initial distribution* describing the random variable at $t = 0$. The values x_k can be thought of the values recorded at time t_k, A_k are the "windows" through which the process is sampled. $X_t(\omega) = \omega(t)$ is called *coordinate process.*

Definition 1.7 (Conditional expectation) *Let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-field and X a random variable on $(\Omega, \mathcal{F}, P)$. The* conditional expectation $\mathbb{E}_P[X|\mathcal{G}]$ *with respect to $\mathcal{G}$ is defined by the unique $\mathcal{G}$-measurable random variable such that*

$$\mathbb{E}_P[\mathbf{1}_A X] = \mathbb{E}_P[\mathbf{1}_A \mathbb{E}_P[X|\mathcal{G}]], \quad A \in \mathcal{G}.$$

The left hand side of the equality above defines a probability measure $\tilde{P}(A) = \mathbb{E}_P[\mathbf{1}_A X]$ on $\mathcal{G}$, thus $\mathbb{E}_P[X|\mathcal{G}]$ is in fact the Radon-Nikodym derivative $d\tilde{P}/dP$.

Theorem 1.6 *Let X be a random variable on $(\Omega, \mathcal{F}, P)$, and $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-field. The following properties of conditional expectation hold:*

1. $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]] = \mathbb{E}[X]$.
2. *If X is $\mathcal{G}$-measurable, then $\mathbb{E}[X|\mathcal{G}] = X$, a.s.*
3. *If Y is $\mathcal{G}$-measurable and bounded, then $\mathbb{E}[XY|\mathcal{G}] = Y\mathbb{E}[X|\mathcal{G}]$, a.s.*
4. *If X is independent of $\mathcal{G}$ (i.e., X is an independent random variable of $\mathbf{1}_A$, $\forall A \in \mathcal{G}$), then $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}[X]$ a.s.*
5. Tower property: *If $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$, then $\mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbb{E}[X|\mathcal{H}]$, a.s.*
6. Linearity: $\mathbb{E}[\alpha X + \beta Y] = \alpha\mathbb{E}[X] + \beta\mathbb{E}[Y]$, *for all random variables X, Y, $\forall \alpha, \beta \in \mathbb{R}$.*
7. Monotonicity: *If $X \leq Y$ a.s., then $\mathbb{E}[X|\mathcal{G}] \leq \mathbb{E}[Y|\mathcal{G}]$, a.s.*
8. Conditional Jensen inequality: *If φ is a convex function and $\mathbb{E}[|X|], \mathbb{E}[|\varphi(X)|] < \infty$, then $\varphi(\mathbb{E}[X|\mathcal{G}]) \leq \mathbb{E}[\varphi(X)|\mathcal{G}]$ a.s.*
9. Fatou's Lemma: *Let $(X_n)_{n\geq 1}$ be a sequence of non-negative random variables. Then $\mathbb{E}[\liminf_{n\to\infty} X_n|\mathcal{G}] \leq \liminf_{n\to\infty} \mathbb{E}[X_n|\mathcal{G}]$ a.s.*

Definition 1.8 (Martingale) *Let $(\Omega, \mathcal{F}, P)$ be a filtered space with given filtration $(\mathcal{F}_t)_{t\geq 0}$. The random process $(X_t)_{t\geq 0}$ is an $(\mathcal{F}_t)_{t\geq 0}$-*martingale *whenever*

(1) $(X_t)_{t\geq 0}$ is $(\mathcal{F}_t)_{t\geq 0}$-adapted

(2) $\mathbb{E}_P[|X_t|] < \infty$ *for all* $t \geq 0$

(3) $\mathbb{E}_P[X_t|\mathcal{F}_s] = X_s$ *for each* $s \leq t$.

A martingale describes the model of a fair game, implying that the best prediction of the future net winnings/losses per unit stake over the interval $[s, t]$ is zero. Indeed,

$$\mathbb{E}[X_t - X_s|\mathcal{F}_s] = \mathbb{E}[X_t|\mathcal{F}_s] - \mathbb{E}[X_s|\mathcal{F}_s] = \mathbb{E}[X_t|\mathcal{F}_s] - X_s = 0.$$

The following result offers a simple method to disprove that a random process is a martingale.

Proposition 1.1 *The expectation of a martingale is constant (independent of time).*

PROOF. For every $0 \leq s \leq t$,

$$\mathbb{E}[X_s] = \mathbb{E}[\mathbb{E}[X_t|\mathcal{F}_s]] = \mathbb{E}[X_t].$$

□

2 Gaussian processes

2.1 Normal random variables

2.2 Gaussian vectors

2.3 Definitions and basic properties of Gaussian processes

2.4 Stationary Gaussian processes

2.5 Sample path properties

2.6 Markov property

2.7 Reproducing kernel Hilbert space

3 Brownian motion and related random processes

3.1 Definition and existence

3.2 Distributional properties

3.2.1 Gaussian properties

3.2.2 Invariance properties

3.3 Martingale and Markov properties

3.3.1 Martingale properties

3.3.2 Markov properties

3.4 Random processes related to Brownian motion

4 Solutions to the exercises