Probability

Jan Dereziński
Department of Mathematical Methods in Physics,
Faculty of Physics, University of Warsaw,
Pasteura 5,
02-093 Warszawa, Poland,
email: jan.derezinski@fuw.edu.pl

June 16, 2022

1 Random processes

1.1 Random variables

A probability space $(\Omega, \mathcal{F}, \mathrm{P})$ is a set Ω equipped with a σ-algebra $\mathcal{F} \subset 2^\Omega$ and a probability measure $\mathrm{P} : \mathcal{F} \to [0, 1]$. If $(\mathcal{S}, \mathcal{B})$ is another set equipped with a σ-algebra, then X is a S-valued random variable if it is a measurable transformation $X : \Omega \to S$, modulo sets of measure zero. We usually take $S = \mathbb{R}$, so that when we do not specify the target set we mean a real valued random variable.

1.2 Finite distributions of random processes

Let T be a set and $\{X_t\}_{t\in T}$ a family of random variables with values in S on a probability space $(\Omega, \mathcal{F}, \mathrm{P})$. Such a family is called a *random process*. Often $T = [0, \infty[$ or $T = \{0, 1, 2, \dots\}$, and T is called the time.

If $t_1, \dots, t_k$ are distinct elements of T, then $(X_{t_1}, \dots X_{t_k}) : \Omega \to S^k$ defines a measurable transformation. This transformation defines a probability measure on S^k: if A is a measurable set in S^k, then

$$\mathrm{P}_{t_1,\dots,t_k}(A) := \mathrm{P}\Big(\big(X_{t_1}(\omega), \dots X_{t_k}(\omega)\big) \in A\Big). \tag{1.1}$$

The family of measures $\mathrm{P}_{t_1,\dots,t_k}$ is called *finite dimensional distributions of* P. It satisfies the *consistency conditions*:

$$\mathrm{P}_{t_1,\dots,t_k}(A) = \mathrm{P}_{t_{\sigma(1)},\dots,t_{\sigma(k)}}(\sigma(A)), \qquad \text{for every permutation} \quad \sigma; \tag{1.2}$$

$$\mathrm{P}_{t_1,\dots,t_k}(A) = \mathrm{P}_{t_1,\dots,t_k,t_{k+1}}(A \times S). \tag{1.3}$$

One can ask whether for every familly of measures $\mathrm{P}_{t_1,\dots,t_k}$ satisfying the consistency conditions is derived from a certain random process. The answer is (partly) positive due to the Kolmogorov Theorem, which we describe below.

Let $\tilde{\Omega}$ be the set of all functions $\tilde{\omega} : T \to S$. (Other possible notations for $\tilde{\Omega}$ are $\underset{t\in T}{\times} S$, the Cartesian product of many copies of S indexed by T). We say that $B \subset \tilde{\Omega}$ is a *cylindrical set* if there exist $\{t_1, \dots, t_k\} \subset T$ and a measurable set $A \subset S^k$ such that

$$\{\tilde{\omega} \mid \big(\tilde{\omega}(t_1), \dots, \tilde{\omega}(t_k)\big) \in A\} = B. \tag{1.4}$$

Let $\tilde{\mathcal{F}}$ be the σ-algebra generated by cylindrical sets in $\tilde{\Omega}$.

Theorem 1.1 (The Kolmogorov Consistency Theorem) *If S is a Polish space (e.g. a countable space or a closed subset of $\mathbb{R}^d$) and $\mathrm{P}_{t_1,\dots,t_k}$ satisfies the consistency conditions, then there exists a unique measure $\tilde{\mathrm{P}}$ on $(\tilde{\Omega}, \tilde{\mathcal{F}})$ such that $\mathrm{P}_{t_1,\dots,t_k}$ are its finite distributions.*

For instance, for $t \in T$, let P_t be a probability measure on S. Then

$$\mathrm{P}_{t_1,\dots,t_k} := \mathrm{P}_{t_1} \otimes \cdots \otimes \mathrm{P}_{t_k} \tag{1.5}$$

is a familty satisfying the consistency condition. The resulting measure on $\tilde{\Omega}$ can be called the product measure $\underset{t\in T}{\otimes} \mathrm{P}_t$.

1.3 Gaussian processes

Let $[\sigma_{ij}]$ be a positive definite matrix $n \times n$. Let $X_1, \dots, X_n$ be the (real) random variables on $\mathbb{R}^n$ with the density

$$\rho(x_1, \dots, x_n) = \frac{\sqrt{\det \sigma}}{(2\pi)^{\frac{n}{2}}} \exp\Big(- \frac{x_i \sigma_{ij} x_j}{2} \Big). \tag{1.6}$$

Then $\rho(x)\mathrm{d}x$ is a probability measure. Let $[\sigma^{ij}]$ be the inverse of $[\sigma_{ij}]$. Then

$$\mathrm{Cov}(X_i X_j) = \sigma^{ij}. \tag{1.7}$$

Clearly, the positive matrix $[\sigma^{ij}]$ determines uniquely a Gaussian measure on $\mathbb{R}^n$ with mean zero and satisfying (1.7).

Let $\{\sigma^{t,s}\}_{t,s\in T}$ be a family of numbers such that for any $t_1, \dots, t_k$ the matrix $[\sigma^{t_i,t_j}]_{i,j=1,\dots k}$ is positive definite. Let $\mathrm{P}_{t_1,\dots,t_k}$ be the Gaussian measure on $\mathbb{R}^k$ with the covariance matrix $[\sigma^{t_i,t_j}]_{i,j=1,\dots k}$. Then the consistency condition is satisfied and we can define the Gaussian measure on $\underset{t\in T}{\times} \mathbb{R}$.

1.4 Conditional expectation

Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space and $A, B \in \mathcal{F}$. Assume that $\mathrm{P}(B) \neq 0$. Recall that the conditional probability of A given B is

$$\mathrm{P}(A|B) := \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)} = \frac{\int_B \mathbb{1}_A(\omega)\mathrm{d}\mathrm{P}(\omega)}{\mathrm{P}(B)}. \tag{1.8}$$

Let $X \in L^1(\Omega, \mathcal{F}, \mathrm{P})$ (X is an integrable random variable). Then we can define the *conditional expectation of X given B* by

$$\mathrm{E}(X|B) := \frac{\int_B X(\omega)\mathrm{d}\mathrm{P}(\omega)}{\mathrm{P}(B)}. \tag{1.9}$$

Suppose now $\mathcal{G}$ is a σ-subalgebra of $\mathcal{F}$. The *conditional expectation of X given $\mathcal{G}$*, denoted $\mathrm{E}(X|\mathcal{G})$, is the random variable $Y \in L^1(\Omega, \mathcal{G}, \mathrm{P})$ such that for any $A \in \mathcal{G}$

$$\int_A X\mathrm{d}\mathrm{P} = \int_A Y\mathrm{d}\mathrm{P}. \tag{1.10}$$

Note that the lhs of (1.10) defines a measure on $(\Omega, \mathcal{G})$ continuous wrt $\mathrm{P}\Big|_{\mathcal{G}}$. Hence the existence of Y follows by the Radon-Nikodym Theorem. Y is unique up to zero measure sets wrt $\mathrm{P}\Big|_{\mathcal{G}}$.

1. If X is measurable wrt $\mathcal{G}$, then

$$\mathrm{E}(X|\mathcal{G}) = X. \tag{1.11}$$

2. If X is independent of $\mathcal{G}$, then
$$\mathrm{E}(X|\mathcal{G}) = \mathrm{E}(X). \tag{1.12}$$
In fact, to prove this it is enough to assume that $X = \mathbb{1}_B$, with B independent of $A \in \mathcal{G}$. Then
$$\int_A \mathbb{1}_B \mathrm{dP} = \int_{A\cap B} \mathrm{dP} = \mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B) = \mathrm{E}(\mathbb{1}_B) \int_A \mathrm{dP}. \tag{1.13}$$

3. $\mathrm{E}(\mathrm{E}(X|\mathcal{G})) = \mathrm{E}(X)$.

4. $\mathrm{E}\big(\mathrm{E}(X|\mathcal{G})|\mathcal{G}\big) = \mathrm{E}(X|\mathcal{G})$.

5. If $\mathcal{G} = \{B, \Omega\backslash B, \emptyset, \Omega\}$ and $0 < \mathrm{P}(B) < 1$, then
$$\mathrm{E}(X|\mathcal{G}) = \mathrm{E}(X|B)\mathbb{1}_B + E(X|\Omega\backslash B)\mathbb{1}_{\Omega\backslash B}. \tag{1.14}$$

6. If $X, XY \in L^1(\Omega, \mathcal{F}, \mathrm{P})$ and Y is measurable wrt $\mathcal{G}$, then
$$\mathrm{E}(XY|\mathcal{G}) = Y\mathrm{E}(X|\mathcal{G}). \tag{1.15}$$

(We say that Y is measurable wrt $\mathcal{G}$ if $Y^{-1}(B) \in \mathcal{G}$ for all Borel sets B.)

If $Y_1, \dots, Y_n$ are random variables on $(\Omega, \mathcal{F})$, then $\sigma(Y_1, \dots, Y_n)$ will denote the smallest σ-algebra wrt which $Y_1, \dots, Y_n$ are measurable. We will write
$$\mathrm{E}(X|Y_1, \dots, Y_n) := \mathrm{E}\big(X|\sigma(Y_1, \dots, Y_n)\big). \tag{1.16}$$

1.5 Markov chains

Let S be a discrete set, $T = \{0, 1, 2, \dots\}$ and $[P^n_{ij}]_{i,j\in S}$ be a family of stochastic matrices, that is
$$P^n_{ij} \geq 0, \quad \sum_j P^n_{ij} = 1. \tag{1.17}$$

Let ρ_j be a probability distribution on S, that is,
$$\rho_j \geq 0, \quad \sum_j \rho_j = 1. \tag{1.18}$$

We define the family of finite dimensional probability distributions $\mathrm{P}^{0,1,\dots,n}$:
$$\mathrm{P}^{0,1,\dots,n}(X_n = s_n, X_{n-1} = s_{n-1}, \dots, X_0 = s_0) := \rho_{s_0} P^1_{s_0 s_1} \cdots P^n_{s_{n-1}s_n}. \tag{1.19}$$

It is easy to see that this family is consistent. In fact, it is enough to check
$$\sum_{s_n \in S} \mathrm{P}^{0,\dots,n}(X_n = s_n, X_{n-1} = s_{n-1}, \dots, X_0 = s_0) = \mathrm{P}^{0,\dots,n-1}(X_{n-1} = s_{n-1}, \dots, X_0 = s_0). \tag{1.20}$$

By using Kolmogorov's Theorem there exists a probability measure P on the probability space $\times_{n=0}^{\infty} S$, such that X_n, $n = 0, 1, \dots$, becomes a random process.

A random process constructed this way has special properties and is called a *Markov chain* (with discrete time $\{0, 1, 2, \dots\}$ on a discrete state space S). It is determined by the *initial distribution* ρ and the *nth step transition matrix* P^n.

Suppose we adopt a converse point of view. We start from a probability measure P on the probability space $\times_{n=0}^{\infty} S$ and random S-valued variables X_n, $n = 0, 1, \ldots$. We say that it is a Markov chain if the following condition holds: If $\mathrm{P}(X_{n-1} = s_{n-1}, \ldots, X_0 = s_0) > 0$, then

$$\mathrm{P}(X_n = s_n | (X_{n-1} = s_{n-1}) = \mathrm{P}(X_n = s_n \mid X_{n-1} = s_{n-1}, \ldots, X_0 = s_0). \tag{1.21}$$

Setting

$$P_{ij}^n := \mathrm{P}(X_n = j | X_{n-1} = i), \tag{1.22}$$

$$\rho_i := \mathrm{P}(X_0 = i), \tag{1.23}$$

we retrieve the construction described in (1.20).

If we have a random process with values in, say, $\mathbb{R}$, then we can reformulate the above definition using the conditional expectation. We say that a random process $\{X_n\}_{n=0,1,\ldots}$ with values in $\mathbb{R}$ is a Markov chain if for any $n = 1, 2, \ldots$

$$\mathrm{E}(X_n | X_{n-1}) = \mathrm{E}(X_n | X_{n-1}, \ldots, X_0). \tag{1.24}$$

It is easy to generalize the above definitions and constructions to stochatic transformations on more general measure spaces.

1.6 Examples of Markov semigroups

Example 1.2 *2-state transition matrix.*

$$P = \begin{bmatrix} 1 - p_1 & p_1 \\ p_2 & 1 - p_2 \end{bmatrix}. \tag{1.25}$$

Stationary distribution: $\left[\frac{p_2}{p_1+p_2}, \frac{p_1}{p_1+p_2}\right]$.

Example 1.3 *Random walk absorbing on the left and reflecting on the right.*

The transition matrix:

$$\begin{bmatrix} 1 & & & & \\ q & 0 & p & & \\ & q & 0 & p & \\ & & q & 0 & p \\ & & & 1 & 0 \end{bmatrix} \tag{1.26}$$

The stationary state is $[1, 0, 0, 0, 0]$.

Example 1.4 *Permutation.*

The transition matrix:

$$\begin{bmatrix} & 1 & \\ & & 1 \\ 1 & & \end{bmatrix} \tag{1.27}$$

The stationary distribution: $[\frac{1}{3}, \frac{1}{3}, \frac{1}{3}]$.

Example 1.5 *The Bernoulli scheme.*

State space: $\{0, 1, 2, \dots\}$. The transition matrix:

$$\begin{bmatrix} q & p & & & \\ & q & p & & \\ & & q & p & \\ & & & q & p \\ & & & & \end{bmatrix} \tag{1.28}$$

$$p_{j,j+m}(n) = p^m q^{n-m} \binom{n}{m}. \tag{1.29}$$

Example 1.6 *Random walk in 1 dimension.*

State space: $\mathbb{Z}$. The transition matrix:

$$\begin{bmatrix} & & & & & & \\ q & 0 & p & & & & \\ & q & 0 & p & & & \\ & & q & 0 & p & & \\ & & & q & 0 & p & \\ & & & & & & \end{bmatrix} \tag{1.30}$$

$$p_{j,j+m}(n) = p^{\frac{n+m}{2}} q^{\frac{n-m}{2}} \binom{n}{\frac{(n+m)}{2}}, \tag{1.31}$$

for even $n - m$, otherwise it is 0.

Using the Stirling formula $n! \sim \sqrt{2\pi n} n^n \mathrm{e}^{-n}$ we obtain

$$P_{0,0}(2n) \sim \frac{1}{\sqrt{\pi n}} (4pq)^n. \tag{1.32}$$

Example 1.7 *The Ehrenfest Model.*

We have two vessels and n particles. At random we transfer a particle from one vessel to the other.

This can be described by the state space $\{0, 1, \dots, n\}$, where the number corresponds to the number of particles in the 1st vessel. The transition matrix is given by $p_{j,j+1} = \frac{n-j}{n}$, $p_{j,j-1} = \frac{j}{n}$:

$$\begin{bmatrix} 0 & \frac{n}{n} & 0 & \\ \frac{1}{n} & 0 & \frac{n-1}{n} & \\ 0 & \frac{2}{n} & 0 & \\ & & & \end{bmatrix} \tag{1.33}$$

Using the identity

$$\binom{n}{j} = \binom{n}{j-1}(n-j+1) + \binom{n}{j+1}(j+1), \tag{1.34}$$

we obtain the stationary distribution

$$2^{-n} \binom{n}{j}. \tag{1.35}$$

1.7 Detailed balance condition

We say that a stochastic matrix P satisfies the Detailed Balance Condition if there exist $\rho_i > 0$ such that

$$\rho_i P_{ij} = \rho_j P_{ji}. \tag{1.36}$$

Then $\rho = [\rho_i]$ satisfies $\rho P = \rho$. Besides, if we define $l^2(S, \rho)$,, then P is self-adjoint in the sense of $l^2(S, \rho)$.

Equivalent condition: for any i_1, i_2, i_3,

$$P_{i_1 i_2} P_{i_2 i_3} P_{i_3 i_1} = P_{i_1 i_3} P_{i_3 i_2} P_{i_2 i_1}. \tag{1.37}$$

1.8 Markov processes with continuous time

Let us now generalize the definition of a Markov chain to the continuous time, that is to $T = [0, \infty[$. We will then say a "Markov process", not a "Markov chain".

Let $\{P^{t,s}\}_{t,s\geq 0}$ be a family of stochastic matrices with the state space S satisfying

$$P^{t,t} = \mathbb{1}, \quad P^{t,s}P^{s,u} = P^{t,u}, \quad 0 \leq t \leq s \leq u. \tag{1.38}$$

Let ρ^0 be a probability distribution on S and

$$\rho^t = \rho^0 P^{0,t}. \tag{1.39}$$

For any $0 \leq t_1 < \cdots < t_k$ we define the probability distribution on S^k:

$$\begin{aligned} &\mathrm{P}^{t_0,t_1,\dots,t_n}(X_{t_n} = s_n, \dots, X_{t_1} = s_1, X_0 = s_0) \\ =&\rho^{t_0} P^{t_0,t_1}_{s_0 s_1} \cdots P^{t_{n-1},t_n}_{s_{n-1} s_n}. \end{aligned} \tag{1.40}$$

We easily check that the family is consistent. To see this we use

$$\begin{aligned} \sum_{s_i} P^{t_{i-1}t_i}_{s_{i-1}s_i} P^{t_i t_{i+1}}_{s_i s_{i+1}} &= P^{t_{i-1}t_{i+1}}_{s_{i-1}s_{i+1}}, \\ \sum_{s_0} \rho^{t_0}_{s_0} P^{t_0 t_1}_{s_0 s_1} &= \rho^{t_1}_{s_1}. \end{aligned} \tag{1.41}$$

Therefore, it defines a measure P on $\underset{t\in[0,\infty[}{\times} S$. The random process $\{X_t\}_{t\in[0,\infty[}$ is said to be Markov.

Suppose that the values of X_t are in $\mathbb{C}^d$. Then we have an equivalent definition: For any $0 \leq t_0 < t_1 < \cdots < t_k$,

$$\mathrm{E}(X_{t_k}|X_{t_{k-1}}) = \mathrm{E}(X_{t_k}|X_{t_{k-1}}, \dots, X_{t_0}). \tag{1.42}$$

1.9 Infinitesimally stochastic matrices

Suppose that $K = [K_{ij}]$ is a real matrix. We say that it is infinitesimally stochastic (or Markovian) if

$$K_{ij} \geq 0, \quad i \neq j; \quad \sum_j K_{ij} = 0. \tag{1.43}$$

Note that necessarily $K_{ii} \leq 0$. If K is infinitesimally stochastic, then e^{tK} is stochastic. In fact, set $P(t) = \mathrm{e}^{tK}$.

$$i \neq j, \quad \frac{\mathrm{d}}{\mathrm{d}t} P(t)_{ij}\Big|_{t=0} = K_{ij} \geq 0, \quad P(0)_{ij} = 0, \quad \text{hence } P_{ij}(t) > 0 \quad \text{for small } t; \tag{1.44}$$

$$P(0)_{ii} = 1, \quad \text{hence } P_{ij}(t) > 0 \quad \text{for small } t; \tag{1.45}$$

$$\frac{\mathrm{d}}{\mathrm{d}t} \sum_j P_{ij}(t) = \sum_{kj} P_{ik}(t) K_{kj} = 0, \qquad \text{hence } \sum_j P_{ij}(t) = 1. \tag{1.46}$$

Example:

$$Z = \begin{bmatrix} -\lambda & \lambda \\ 0 & 0 \end{bmatrix}. \tag{1.47}$$

Then

$$Z \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 0 \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 \end{bmatrix} Z = \begin{bmatrix} 0 & 1 \end{bmatrix} 0, \tag{1.48}$$

$$Z \begin{bmatrix} 1 \\ 0 \end{bmatrix} = -\lambda \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & -1 \end{bmatrix} Z = \begin{bmatrix} 0 & 1 \end{bmatrix} (-\lambda). \tag{1.49}$$

Hence

$$\mathrm{e}^{tZ} = \begin{bmatrix}1\\1\end{bmatrix}\begin{bmatrix}0 & 1\end{bmatrix} + \mathrm{e}^{-t\lambda}\begin{bmatrix}1\\0\end{bmatrix}\begin{bmatrix}1 & -1\end{bmatrix} = \begin{bmatrix}\mathrm{e}^{-t\lambda} & 1-\mathrm{e}^{-t\lambda}\\ 0 & 1\end{bmatrix}. \tag{1.50}$$

1.10 Poisson Process

Let $\lambda > 0$. Recall that

$$p_\lambda(n) := \mathrm{e}^{-\lambda}\frac{\lambda^n}{n!}, \quad n = 0, 1, 2, \dots \tag{1.51}$$

defines a probability distribution called the Poisson distribution.

Consider the state space $\{0, 1, 2, \dots\}$ and the infinitesimally stochastic matrix

$$Z := \begin{bmatrix} -\lambda & \lambda & & & & \\ & -\lambda & \lambda & & & \\ & & -\lambda & \lambda & & \\ & & & -\lambda & \lambda & \\ & & & & -\lambda & \\ & & & & & \end{bmatrix} = -\lambda 1\!\!1 + \lambda N, \tag{1.52}$$

where N is the right unilateral shift. Clearly,

$$\mathrm{e}^{tZ} = \sum_{n=0}^{\infty} \mathrm{e}^{-\lambda t}\frac{\lambda^n (t-s)^n \mathrm{e}^{-\lambda(t-s)}}{n!} N^n =: \sum_{n=0}^{\infty} p_{\lambda(t-s)}(n) N^n. \tag{1.53}$$

If we denote by $(m|$ the mth basis vector, then

$$(m|\mathrm{e}^{tZ} = \sum_{n=0}^{\infty}(m+n|p_{\lambda t}(n). \tag{1.54}$$

Using the family of stochastic matrices $\mathrm{e}^{(t-s)Z}$ and the initial distribution $[1, 0, \dots]$, we can construct the Markov process X_t with time $[0, \infty[$ and values in $\{0, 1, 2, \dots\}$. Clearly, for $0 \leq t_1 < \cdots < t_n$

$$\begin{aligned} &\mathrm{P}(X_{t_n} = s_n, \dots, X_{t_1} = s_1) \\ =&p_{\lambda(t_n - t_{n-1})}\big(s_n - s_{n-1}\big)\cdots p_{\lambda(t_2-t_1)}\big(s_2 - s_1\big)p_{\lambda t_1}\big(s_1\big). \end{aligned} \tag{1.55}$$

The Poisson process has the following properties, which can be used as its definition:

1. $X_0 = 0$ almost everywhere.
2. For $0 \leq t_1 < \cdots < t_n$, the random variables $X_{t_n} - X_{t_{n-1}}, \dots X_{t_2} - X_{t_1}, X_{t_1}$ are independent.
3. For $0 \leq s < t$ the random variable $X_t - X_s$ has the Poisson distribution with the parameter $\lambda(t-s)$, that is

$$\mathrm{P}(X_t - X_s = n) = \begin{cases} p_{\lambda(t-s)}(n), & n = 0, 1, 2, \dots; \\ 0, & n = -1, -2, \dots. \end{cases} \tag{1.56}$$

1.11 Brownian motion

There are several ways to define the Brownian motion. One possible definition is to define it as a Gaussian process with the covariances

$$\sigma^{t_1,t_2} := \mathrm{Cov}(W_{t_1}, W_{t_2}) = \min(t_1, t_2), \quad t_1, t_2 \in [0, \infty[. \tag{1.57}$$

To see that the matrix made of (1.57) is positive definite, consider $0 \le t_1 < \cdots < t_n$ and set $s_j = t_j - t_{j-1}$, $s_1 = t_1$. Then this matrix is

$$\begin{bmatrix} t_1 & t_1 & \dots & t_1 \\ t_1 & t_2 & \dots & t_2 \\ & & \dots & \\ t_1 & t_2 & \dots & t_k \end{bmatrix} = s_1 \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ & & \dots & \\ 1 & 1 & \dots & 1 \end{bmatrix} + s_2 \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 1 & \dots & 1 \\ & & \dots & \\ 0 & 1 & \dots & 1 \end{bmatrix} + \cdots + s_k \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & 1 \end{bmatrix}. \tag{1.58}$$

Clearly, each matrix in (1.58) is positive definite.

This process has the following properties, which can be used as its definition:

1. $W_0 = 0$.
2. For $0 \le s \le t$ the random variable $W_t - W_s$ is Gaussian with zero mean and variance $t - s$.
3. For $0 \le t_0 < t_1 < \cdots < t_n$ the random variables W_{t_0}, $W_{t_1} - W_{t_0}$,..., $W_{t_n} - W_{t_{n-1}}$ are independent.

It is easy to see 1-3. For instance, if $0 \le s \le t \le u$ then

$$\mathrm{E}(W_t - W_s)^2 = \mathrm{E}(W_t^2) - 2\mathrm{E}(W_t W_s) + \mathrm{E}(W_s^2) = t - 2s + s = t - s; \tag{1.59}$$

$$\mathrm{E}(W_s(W_t - W_u)) = \mathrm{E}(W_s W_t) - \mathrm{E}(W_s W_u) = s - s = 0. \tag{1.60}$$

Then we use the fact that for Gaussian variables vanishing of the correlation implies independence.

The Brownian motion is a Markov process with the state space $\mathbb{R}$ equipped with its Borel structure andwit time in $[0, \infty[$. In fact, consider the difffusion semigroup given by its integral kernel wrt the Lebesgue measure

$$p_t(x, y) := \mathrm{e}^{\frac{t\Delta}{2}}(x, y) = \frac{1}{\sqrt{2\pi t}} \exp\Big(- \frac{(x-y)^2}{2t}\Big). \tag{1.61}$$

Using the family of stochastic operators $\mathrm{e}^{\frac{t\Delta}{2}}$ and the initial distribution $\delta(x)$ we define the finite dimensional distributions

$$\begin{aligned} &\mathrm{P}(X_{t_n} = x_n, \dots, X_{t_1} = x_1)\mathrm{d}x_n \cdots \mathrm{d}x_1 \\ =& p_{t_1}(0, x_1) p_{t_2 - t_1}(x_1, x_2) \cdots p_{t_n - t_{n-1}}(x_{n-1}, x_n). \end{aligned} \tag{1.62}$$

From these finite dimensional distribution we can construct a Markov process with help of the Kolmogorov Theorem.

Acturally, one usually prefers a slightly different construction, which yield much smaller probability spaces. The typical textbook definition of the Brownian motion involves the space of continuous functions $[0, \infty[\ni t \mapsto W(t)$. Note that this requires a separate construction, since continuous functions do not form a measurable subset of $\tilde{\Omega}$.

2 Law of large numbers

2.1 The Jensen Inequality

We say that $g : \mathbb{R} \to \mathbb{R}$ is convex if

$$g(\tau x + (1-\tau)y) \le \tau g(x) + (1-\tau)g(y), \quad x, y \in \mathbb{R}, \quad 0 \le \tau \le 1. \tag{2.63}$$

As a consequence, if $p_1, \dots, p_n \ge 0$, $p_1 + \cdots + p_n = 1$, $x_1, \dots, x_n \in \mathbb{R}$, then

$$g(p_1 x_1 + \cdots + p_n x_n) \le p_1 g(x_1) + \cdots + p_n g(x_n). \tag{2.64}$$

Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space.

Proposition 2.1 *Let g be convex, X a real random variable such that $g(X) \in L^1$. Then*

$$g(\mathrm{E}(X)) \le \mathrm{E}(g(X)). \tag{2.65}$$

Proof. If $A_1, \dots A_n$ is a partition of Ω into measurable sets and $X = x_1 \mathbb{1}_{A_1} + \cdots x_n \mathbb{1}_{A_n}$ with $\mathrm{P}(A_n) = p_n$, then

$$\mathrm{E}(X) = \sum_{i=1}^{n} x_i p_i, \quad \mathrm{E}\big(g(X)\big) = \sum_{i=1}^{n} p_i g(x_i). \tag{2.66}$$

Therefore for elementary functions the Jensen inequality coincides with (2.64). □

Proposition 2.2 $[0,\infty[\ni p \mapsto \big(\mathrm{E}|X|^p\big)^{\frac{1}{p}}$ *is an increasing function.*

Proof. Let $g(t) = t^r$ and $r \ge 1$. Then $g''(t) = r(r-1)t^{r-2} \ge 0$. Hence g is convex.
Let $0 \le q \le p$. We use the Jensen inequality with $g = t^r$ and $r := \frac{p}{q} \ge 1$:

$$\mathrm{E}(|X|^q)^{\frac{p}{q}} \le \mathrm{E}(|X|^{q\frac{p}{q}}) = \mathrm{E}(|X|^p). \qquad \square \tag{2.67}$$

2.2 Law of large numbers by the Chebyshev Inequality

Proposition 2.3

$$\mathrm{P}(|X| \ge \epsilon) \le \frac{\mathrm{E}(|X|)}{\epsilon}. \tag{2.68}$$

Therefore,

$$\mathrm{P}\big(|X - \mathrm{E}(X)| \ge c\big) \le \frac{1}{c^2}\mathrm{Var}(X). \tag{2.69}$$

Proof. Clearly

$$\epsilon \mathbb{1}(|X| \ge \epsilon) \le |X|. \tag{2.70}$$

This proves (2.68). By applying it to $|X - \mathrm{E}(X)|^2$ and $\epsilon = c^2$ we obtain (2.69). □

Suppose that X_1, X_2,... are independent random variables with $\mathrm{E}(X_n) = m$. Set

$$S_n := X_1 + \cdots + X_n. \tag{2.71}$$

Clearly,

$$\mathrm{E}\Big(\frac{S_n}{n}\Big) = m. \tag{2.72}$$

Theorem 2.4 *Let*

$$\mathrm{Var}(X_n) \le v. \tag{2.73}$$

Then

$$\mathrm{P}\Big(\Big|\frac{S_n}{n} - m\Big|\Big) > \delta\Big) \le \frac{v}{\delta^2 n}. \tag{2.74}$$

Proof. In fact,

$$\mathrm{Var}\Big(\frac{S_n}{n}\Big) = \frac{1}{n^2}\sum_{j=1}^{n} \mathrm{Var}(X_j) \le \frac{v}{n}. \tag{2.75}$$

Applying the Chebyshev inequality, more precisely (2.69), we obtain (2.74). □

2.3 Improved Law of Large Numbers

Theorem 2.5 *Let X_n, S_n be as above and*

$$\mathrm{E}(X_n - m)^4 \le K. \tag{2.76}$$

Then

$$\mathrm{P}\Big(\frac{S_n}{n} \to m\Big) = 1. \tag{2.77}$$

Proof. We can replace X_n with $X_n - \mathrm{E}(X_n)$, so that $\mathrm{E}(X_n) = 0$. Then, using

$$\mathrm{E}(X_i X_j^3) = \mathrm{E}(X_i X_j^2 X_k) = \mathrm{E}(X_i X_j X_k X_l) = 0 \tag{2.78}$$

for distinct i, j, k, l, we obtain

$$\mathrm{E}(S_n^4) = \sum_{k=1}^{n} \mathrm{E}(X_k^4) + 6 \sum_{j \ne k} \mathrm{E}(X_j^2 X_k^2) \tag{2.79}$$

Now

$$\mathrm{E}(X_j^2) \le \mathrm{E}(X_j^4)^{\frac{1}{2}}. \tag{2.80}$$

Hence

$$\mathrm{E}(X_i^2 X_j^2) = \mathrm{E}(X_i^2)\mathrm{E}(X_k^2) \le \mathrm{E}(X_i^4)^{\frac{1}{2}} \mathrm{E}(X_k^4)^{\frac{1}{2}} \le K. \tag{2.81}$$

Therefore,

$$\mathrm{E}\Big(\frac{S_n}{n}\Big)^4 \le \frac{1}{n^4}\big(nK + 3n(n-1)K\big) \le \frac{3K}{n^2}, \tag{2.82}$$

$$\text{hence} \quad \mathrm{E}\left(\sum_{n=1}^{\infty} \left(\frac{S_n}{n}\right)^4\right) \le \sum_{n=1}^{\infty} \frac{3K}{n^2} < \infty. \tag{2.83}$$

Therefore, by Lemma 2.6, $\sum\limits_{n=1}^{\infty} \left(\frac{S_n}{n}\right)^4$ is convergent almost everywhere. Hence

$$\frac{S_n}{n} \to 0 \quad \text{almost everywhere.} \tag{2.84}$$

□

Lemma 2.6 *Let $X \ge 0$ and $\mathrm{E}(X) < \infty$. Then $X < \infty$ almost everywhere (there exists a measurable set N of measure zero such that $X < \infty$ outside N).*

Proof. Let $N := \{X(\omega) = \infty\}$. Then $X \ge \infty \mathbb{1}_N$. Now

$$\mathrm{E}(X) \ge \infty \mathrm{P}(N). \qquad \square \tag{2.85}$$

2.4 Convergence of random variables

We say that $X_n \to X$

$$\text{almost surely} \quad \text{iff } \mathrm{P}\big(\omega \mid \lim_{n\to\infty} X_n(\omega) = X(\omega)\big) = 1; \tag{2.86}$$

$$\text{in probability} \quad \text{iff for any } \epsilon > 0, \quad \lim_{n\to\infty} \mathrm{P}\big(|X_n - X| > \epsilon\big) = 0; \tag{2.87}$$

$$\text{in the } p\text{th moment (in } L^p) \quad \text{iff } \lim_{n\to\infty} \mathrm{E}|X_n - X|^p = 0. \tag{2.88}$$

Proposition 2.7 *$X_n \to X$ almost surely implies $X_n \to X$ in probability.*

Proof. Let $N := \{\omega \mid X_n(\omega) \not\to X(\omega)\}$. Then $\mathrm{P}(N) = 0$.

Let $\epsilon > 0$. For any $\omega \in \Omega \backslash N$ we have $|X_n(\omega) - X(\omega)| > \epsilon$ finitely many times. Hence

$$N \supset \{|X_n - X| > \epsilon \text{ infinitely often }\} = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} \{|X_m - X| > \epsilon\}, \tag{2.89}$$

$$\text{hence } 0 = \mathrm{P}(N) \geq \mathrm{P}\left(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} \{|X_m - X| > \epsilon\} \right) \tag{2.90}$$

$$= \lim_{n\to\infty} \mathrm{P}\Big(\bigcup_{m=n}^{\infty} \{|X_m - X| > \epsilon\} \Big) \geq \lim_{n\to\infty} \mathrm{P}\Big(\{|X_n - X| > \epsilon\} \Big), \tag{2.91}$$

$$\text{hence } 0 = \lim_{n\to\infty} \mathrm{P}(|X_n - X| > \epsilon). \qquad \square \tag{2.92}$$

Proposition 2.8 *$X_n \to X$ in L^p implies $X_n \to X$ in probability.*

Proof.

$$0 \leftarrow \mathrm{E}(|X - X_n|^p) \geq \epsilon^p \mathrm{P}(|X - X_n| > \epsilon). \qquad \square \tag{2.93}$$

3 Characteristic functions and Central Limit Theorem

3.1 Real random variables

Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space and X a real valued random variable on Ω. X defines a probability measure μ on Borel sets in $\mathbb{R}$ by

$$\mu(A) := \mathrm{P}(X(\omega) \in A). \tag{3.94}$$

μ is sometimes called the *law of X*.

The distribution function of X is defined as

$$F(t) := \mu(]-\infty, t]) = \mathrm{P}(X \leq t). \tag{3.95}$$

If μ is continuous wrt the Lebesgue measure, then there exists $f \in L^1(\mathbb{R})$ with $\int f = 1$, $f \geq 0$ such that

$$\mu(A) = \int_A f(x)\mathrm{d}x. \tag{3.96}$$

In general, we will write $\mathrm{d}\mu(x) = f(x)\mathrm{d}x$ even if such a density does not exist, e.g. δ_a will be written as $\delta(x-a)\mathrm{d}x$. We have $F'(x) = f(x)$ in the distributional sense. We say that μ possesses an atom at $a \in \mathbb{R}$ if $\mathrm{P}(X = a) > 0$. Note that the number of atoms is countable.

3.2 Characteristic functions

Let X be a random variable, μ its "law". Probabilists use the term "characteristic function of X" as the name of the Fourier transform of μ. Thus, using various notations, the characteristic function is defined as

$$\phi_X(\xi) = \hat{\mu}(-\xi) = \int \mathrm{e}^{\mathrm{i}\xi x}\mathrm{d}\mu(x) = \mathrm{E}(\mathrm{e}^{\mathrm{i}\xi X}) = \int \mathrm{e}^{\mathrm{i}\xi X(\omega)}\mathrm{dP}(\omega). \tag{3.97}$$

Here are some properties:

1. $\phi_X(0) = 1$.
2. $|\phi_X(\xi)| \leq 1$.
3. $\xi \mapsto \phi_X(\xi)$ is continuous.
4. $\phi_{-X}(\xi) = \overline{\phi_X(\xi)}$.
5. $\phi_{aX+b} = \mathrm{e}^{\mathrm{i}b\xi}\phi_X(a\xi)$.
6. If X, Y are independent, then $\phi_{X+Y}(\xi) = \phi_X(\xi)\phi_Y(\xi)$. (Because $\mathrm{E}(\mathrm{e}^{\mathrm{i}(X+Y)\xi}) = \mathrm{E}(\mathrm{e}^{\mathrm{i}X\xi})\mathrm{E}(\mathrm{e}^{\mathrm{i}Y\xi})$.)

3.3 Examples of characteristic functions

In the following table $x \in \mathbb{R}$ and $n \in \mathbb{N}$. On the left we give the density of a probability measure, on the right its characteristic function:

$$\begin{array}{rl}
\frac{1}{\sigma\sqrt{2\pi}}\exp\Big(-\frac{(x-\mu)}{2\sigma^2}\Big) & \exp\Big(\mathrm{i}\mu\xi - \frac{1}{2}\sigma^2\xi^2\Big), \\
\frac{1}{2}\mathbb{1}_{[-1,1]} & \frac{\sin\xi}{\xi}; \\
\frac{1}{2}\mathrm{e}^{-|x|} & \frac{1}{1+\xi^2}; \\
\frac{1}{\pi(1+x^2)} & \mathrm{e}^{-|\xi|}; \\
(1-|x|)\mathbb{1}_{[-1,1]} & 2\frac{(1-\cos\xi)}{\xi^2}; \\
\frac{(1-\cos x)}{\pi x^2} & (1-|\xi|)\mathbb{1}_{[-1,1]}(\xi); \\
\mathrm{e}^{-\lambda}\frac{\lambda^n}{n!} & \mathrm{e}^{\lambda(\mathrm{e}^{\mathrm{i}\langle x\rangle}-1)}; \\
p^n(1-p)^{N-n}\binom{N}{n} & (p\mathrm{e}^{\mathrm{i}\xi}+1-p)^N; \\
(1-p)^n p^\alpha\binom{\alpha+n-1}{n} & \big(1-(1-p)\mathrm{e}^{\mathrm{i}\xi}\big)^{-\alpha}p^\alpha.
\end{array}$$

3.4 Weak convergence of probability measure

Let $\mathrm{Prob}(\mathbb{R})$ denote the set of probability measures on $\mathbb{R}$. Let $\mu_n, \mu \in \mathrm{Prob}(\mathbb{R})$. We say that $\mu_n \to \mu$ weakly if for any $h \in C_{\mathrm{b}}(\mathbb{R})$

$$\int h\mathrm{d}\mu_n \to \int h\mathrm{d}\mu. \tag{3.98}$$

Example 3.1 *1. Let $f_n, f \in L^1_+(\mathbb{R})$ and $\|f_n - f\|_1 \to 0$. Then $f_n\mathrm{d}x \to f\mathrm{d}x$ weakly.*

2. Let $x_n, x \in \mathbb{R}$ and $x_n \to x$. Then $\delta_{x_n} \to \delta_{x_0}$ weakly.

3. Let $x_n \to \infty$. Then δ_{x_n} does not have a weak limit.

Proposition 3.2 *Let F_n, F be the distribution functions of μ_n, μ. Then $\mu_n \to \mu$ weakly iff $F_n(x) \to F(x)$ for every $x \in \mathbb{R}$ which is not an atom of F.*

Let X_n, X be random variables, possibly on different probability spaces $(\Omega_n, \mathcal{F}_n)$, $(\Omega, \mathcal{F})$. We say that $X_n \to X$ *in law* if for the corresponding measures on $\mathbb{R}$ we have the weak convergence $\mu_n \to \mu$.

3.5 Convergence of characteristic functions

Theorem 3.3 (Levy-Cramer) *Let $\mu_n \in \mathrm{Prob}(\mathbb{R})$, ϕ_n their characteristic functions. Suppose that for all $\xi \in \mathbb{R}$ there exists*

$$\phi(\xi) := \lim_{n\to\infty} \phi_n(\xi). \tag{3.99}$$

and ϕ is continuous in 0. Then there exists $\mu \in \mathrm{Prob}(\mathbb{R})$ such that ϕ is the characteristic function of μ and $\mu_n \to \mu$ weakly.

One can see that the condition of the continuity is necessary from the following example. Suppose that

$$\mathrm{d}\mu_n = \frac{1}{n\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2n^2}} \mathrm{d}x. \tag{3.100}$$

Then

$$\phi_n(\xi) = \mathrm{e}^{-\frac{\xi^2 n^2}{2}} \to \begin{cases} 1 & \xi = 0; \\ 0 & \xi \neq 0. \end{cases} \tag{3.101}$$

Clearly, μ_n does not converge to any measure.

3.6 Central Limit Theorem

Theorem 3.4 *Suppose that X_n are independent random variables with the same distribution as X. Let*

$$\mathrm{E}(X) = 0, \quad \sigma^2 := \mathrm{Var}(X) < \infty. \tag{3.102}$$

Set

$$G_n := \frac{X_1 + \cdots + X_n}{\sigma\sqrt{n}}. \tag{3.103}$$

Then G_n converges in law to the normal distribution. In other words, (noting that the normal distribution has no atoms),

$$\mathrm{P}(G < x) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \mathrm{e}^{-\frac{y^2}{2}} \mathrm{d}y =: \mathrm{Erf}(x). \tag{3.104}$$

Proof. (We follow Williams). Set

$$R_n(x) := \mathrm{e}^{\mathrm{i}x} - \sum_{k=0}^{n} \frac{(\mathrm{i}x)^k}{k!}. \tag{3.105}$$

We have $R_0(x) = \mathrm{e}^{\mathrm{i}x} - 1 = \int_0^x \mathrm{i}\mathrm{e}^{\mathrm{i}y}\mathrm{d}y$. Therefore,

$$|R_0(x)| \leq \min(2, |x|). \tag{3.106}$$

Next $R_n(x) = \int_0^x \mathrm{i}R_{n-1}(y)\mathrm{d}y$. Hence

$$|R_n(x) \leq \min\Big(\frac{2|x|^n}{n!}, \frac{|x|^{n+1}}{(n+1)!}\Big). \tag{3.107}$$

Therefore, if $\mathrm{E}(X) = 0$ and $\sigma^2 = \mathrm{Var}(X) < \infty$,

$$\begin{aligned} &\left|\mathrm{E}\Big(\mathrm{e}^{\mathrm{i}\xi X} - 1 + \frac{1}{2}\sigma^2\Big)\right| = |\mathrm{E}R_2(\xi X)| \\ &\leq \mathrm{E}|R_2(\xi X)| \leq \xi^2 \mathrm{E}\Big(\min\Big(|X|^2, \frac{|\xi||X|^3}{6}\Big)\Big). \end{aligned} \tag{3.108}$$

We have the pointwise convergence $\min\Big(|X|^2, \frac{|\xi||X|^3}{6}\Big) \to 0$ as $\xi \to 0$. Besides, $|X|^2$ is integrable. Therefore, by the Lebesque Dominated Convergence Theorem

$$\mathrm{E}\Big(\min\Big(|X|^2, \frac{|\xi||X|^3}{6}\Big) \to 0 \tag{3.109}$$

Hence

$$\phi_X(\xi) = 1 - \frac{1}{2}\sigma^2\xi^2 + o(\xi^2). \tag{3.110}$$

Now, as $n \to \infty$,

$$\phi_{G_n}(\xi) = \phi_X\Big(\frac{\xi}{\sigma\sqrt{n}}\Big)^n \tag{3.111}$$

$$= \Big(1 - \frac{\xi^2}{2n} + o\Big(\frac{\xi^2}{\sigma^2 n}\Big)\Big)^n \tag{3.112}$$

$$= \exp\Big(n\log\Big(1 - \frac{\xi^2}{2n} + o\Big(\frac{\xi^2}{\sigma^2 n}\Big)\Big) \tag{3.113}$$

$$= \exp\Big(n\Big(-\frac{\xi^2}{2n} + o\Big(\frac{\xi^2}{\sigma^2 n}\Big)\Big) = \exp\Big(-\frac{\xi^2}{2} + no\Big(\frac{1}{n}\Big)\Big) \to \mathrm{e}^{-\frac{1}{2}\xi^2}. \tag{3.114}$$

$\mathrm{e}^{-\frac{1}{2}\xi^2}$ is clearly continuous at 0. Hence we can invoke the Levy-Cramer Theorem to get the result. □

3.7 Stable distributions

We say that the distribution of a random variable X is *stable* if the following holds: If $X_1, \dots, X_n$ are independent random variables idedntically distributed as X, then $S_n := X_1 + \cdots + X_n$ has the same distribution as $c_n X$ for some $c_n \in \mathbb{R}$.

One can show that the only possible c_n are $c_n = n^{\frac{1}{\alpha}}$ for some $\alpha \in]0,2]$. We then say that X is *α-stable.*

If $\phi(\xi)$ is the characteristic function of X, then S_n has the characteristic fanction $\phi^n(\xi)$. Therefore, the α-stability s equivalent to

$$\phi^n(\xi) = \phi(n^{\frac{1}{\alpha}}\xi). \tag{3.115}$$

Here are examples of stable distributions:

The Gaussian distribution:.

$$p(x) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{x^2}{2}}, \quad \phi(\xi) = \mathrm{e}^{-\frac{\xi^2}{2}}. \tag{3.116}$$

Clearly, it is 2-stable:

$$\Big(\mathrm{e}^{-\frac{\xi^2}{2}}\Big)^n = \mathrm{e}^{-\frac{(\sqrt{n}\xi)^2}{2}}. \tag{3.117}$$

The Cauchy distribution.

$$p(x) = \frac{1}{\pi(1+x^2)}, \quad \phi(\xi) = \mathrm{e}^{-|\xi|}. \tag{3.118}$$

Clearly, it is 1-stable:

$$\Big(\mathrm{e}^{-|\xi|}\Big)^n = \mathrm{e}^{-|n\xi|}. \tag{3.119}$$

More generally, distributions with the characteristic functions $e^{-c|\xi|^\alpha}$, $0 < \alpha \leq 2$ are α-stable. Note that among them only the Gaussian distribution has a finite variance. To see this we use

$$\mathrm{E}(X^2) = -\frac{\mathrm{d}^2}{\mathrm{d}\xi^2}\phi(\xi)\Big|_{\xi=0} \tag{3.120}$$

Now $e^{-|\xi|^\alpha}$ for $0 < \alpha < 2$ is not twice differentiable at zero:

$$\frac{\mathrm{d}^2}{\mathrm{d}\xi^2}e^{-|\xi|^\alpha} = \alpha(\alpha-1)|\xi|^{\alpha-2}e^{-|\xi|^\alpha}, \tag{3.121}$$

which for $\alpha \in]0,2[\setminus\{1\}$ has no limit at $\xi = 0$, and for $\alpha = 1$, we get

$$\frac{\mathrm{d}^2}{\mathrm{d}\xi^2}e^{-|\xi|} = (2\delta(\xi)+1)e^{-|\xi|}. \tag{3.122}$$