Chapter 2
Some Mathematical Preliminaries

2.1 Probability Spaces, Random Variables and Stochastic Processes

Having stated the problems we would like to solve, we now proceed to find reasonable mathematical notions corresponding to the quantities mentioned and mathematical models for the problems. In short, here is a first list of the notions that need a mathematical interpretation:

(1) A random quantity
(2) Independence
(3) Parametrized (discrete or continuous) families of random quantities
(4) What is meant by a "best" estimate in the filtering problem (Problem 3)
(5) What is meant by an estimate "based on" some observations (Problem 3)?
(6) What is the mathematical interpretation of the "noise" terms?
(7) What is the mathematical interpretation of the stochastic differential equations?

In this chapter we will discuss (1)–(3) briefly. In the next chapter we will consider (6), which leads to the notion of an Itô stochastic integral (7). In Chapter 6 we will consider (4)–(5).

The mathematical model for a random quantity is a *random variable*. Before we define this, we recall some concepts from general probability theory. The reader is referred to e.g. Williams (1991) for more information.

Definition 2.1.1 *If Ω is a given set, then a σ-*algebra *$\mathcal{F}$ on Ω is a family $\mathcal{F}$ of subsets of Ω with the following properties:*

(i) $\emptyset \in \mathcal{F}$
(ii) $F \in \mathcal{F} \Rightarrow F^C \in \mathcal{F}$, *where* $F^C = \Omega \setminus F$ *is the complement of* F *in* Ω
(iii) $A_1, A_2, \ldots \in \mathcal{F} \Rightarrow A := \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

The pair $(\Omega, \mathcal{F})$ *is called a* measurable space. *A* probability measure P *on a measurable space* $(\Omega, \mathcal{F})$ *is a function* $P: \mathcal{F} \longrightarrow [0,1]$ *such that*

(a) $P(\emptyset) = 0, \ P(\Omega) = 1$

(b) *if* $A_1, A_2, \ldots \in \mathcal{F}$ *and* $\{A_i\}_{i=1}^{\infty}$ *is* disjoint *(i.e.* $A_i \cap A_j = \emptyset$ *if* $i \neq j$*) then*

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \ .$$

The triple $(\Omega, \mathcal{F}, P)$ *is called a* probability space. *It is called a* complete *probability space if* $\mathcal{F}$ *contains all subsets* G *of* Ω *with* P*-outer measure zero, i.e. with*

$$P^*(G) := \inf\{P(F); F \in \mathcal{F}, G \subset F\} = 0 \ .$$

Any probability space can be made complete simply by adding to $\mathcal{F}$ all sets of outer measure 0 and by extending P accordingly. From now on we will assume that all our probability spaces are complete.

The subsets F of Ω which belong to $\mathcal{F}$ are called $\mathcal{F}$*-measurable* sets. In a probability context these sets are called *events* and we use the interpretation

$$P(F) = \text{“the probability that the event } F \text{ occurs”} \ .$$

In particular, if $P(F) = 1$ we say that “F occurs with probability 1”, or “almost surely (a.s.)”.

Given any family $\mathcal{U}$ of subsets of Ω there is a smallest σ-algebra $\mathcal{H}_{\mathcal{U}}$ containing $\mathcal{U}$, namely

$$\mathcal{H}_{\mathcal{U}} = \bigcap\{\mathcal{H}; \mathcal{H} \ \sigma\text{-algebra of } \Omega, \ \mathcal{U} \subset \mathcal{H}\} \ .$$

(See Exercise 2.3.)

We call $\mathcal{H}_{\mathcal{U}}$ *the* σ*-algebra generated by* $\mathcal{U}$.

For example, if $\mathcal{U}$ is the collection of all open subsets of a topological space Ω (e.g. $\Omega = \mathbf{R}^n$), then $\mathcal{B} = \mathcal{H}_{\mathcal{U}}$ is called the *Borel* σ*-algebra* on Ω and the elements $B \in \mathcal{B}$ are called *Borel sets*. $\mathcal{B}$ contains all open sets, all closed sets, all countable unions of closed sets, all countable intersections of such countable unions etc.

If $(\Omega, \mathcal{F}, P)$ is a given probability space, then a function $Y: \Omega \to \mathbf{R}^n$ is called $\mathcal{F}$*-measurable* if

$$Y^{-1}(U) := \{\omega \in \Omega; Y(\omega) \in U\} \in \mathcal{F}$$

for all open sets $U \in \mathbf{R}^n$ (or, equivalently, for all Borel sets $U \subset \mathbf{R}^n$).

If $X: \Omega \to \mathbf{R}^n$ is any function, then *the* σ*-algebra* $\mathcal{H}_X$ *generated by* X is the smallest σ-algebra on Ω containing all the sets

$$X^{-1}(U) \ ; \qquad U \subset \mathbf{R}^n \ \text{ open} \ .$$

It is not hard to show that

$$\mathcal{H}_X = \{X^{-1}(B);\ B \in \mathcal{B}\}\ ,$$

where $\mathcal{B}$ is the Borel σ-algebra on $\mathbf{R}^n$. Clearly, X will then be $\mathcal{H}_X$-measurable and $\mathcal{H}_X$ is the smallest σ-algebra with this property.

The following result is useful. It is a special case of a result sometimes called the *Doob-Dynkin lemma*. See e.g. M. M. Rao (1984), Prop. 3, p. 7.

Lemma 2.1.2 *If $X, Y\colon \Omega \to \mathbf{R}^n$ are two given functions,then Y is $\mathcal{H}_X$-measurable if and only if there exists a Borel measurable function $g\colon \mathbf{R}^n \to \mathbf{R}^n$ such that*

$$Y = g(X)\ .$$

In the following we let $(\Omega, \mathcal{F}, P)$ denote a given complete probability space. A *random variable* X is an $\mathcal{F}$-measurable function $X\colon \Omega \to \mathbf{R}^n$. Every random variable induces a probability measure μ_X on $\mathbf{R}^n$, defined by

$$\mu_X(B) = P(X^{-1}(B))\ .$$

μ_X is called the *distribution of* X.

If $\int_\Omega |X(\omega)| dP(\omega) < \infty$ then the number

$$E[X]\colon = \int_\Omega X(\omega) dP(\omega) = \int_{\mathbf{R}^n} x d\mu_X(x)$$

is called *the expectation* of X (w.r.t. P).

More generally, if $f\colon \mathbf{R}^n \to \mathbf{R}$ is Borel measurable and $\int_\Omega |f(X(\omega))| dP(\omega) < \infty$ then we have

$$E[f(X)]\colon = \int_\Omega f(X(\omega)) dP(\omega) = \int_{\mathbf{R}^n} f(x) d\mu_X(x)\,.$$

The L^p-spaces

If $X : \Omega \to \mathbf{R}^n$ is a random variable and $p \in [1, \infty)$ is a constant we define the L^p-norm of X, $||X||_p$, by

$$\|X\|_p = \|X\|_{L^p(P)} = \Big(\int_\Omega |X(\omega)|^p dP(\omega) \Big)^{\frac{1}{p}}.$$

If $p = \infty$ we set

$$\|X\|_\infty = \|X\|_{L^\infty(P)} = \inf\{N \in \mathbf{R}; |X(\omega)| \le N \text{ a. s.}\}.$$

The corresponding L^p-spaces are defined by

$$L^p(P) = L^p(\Omega) = \{X : \Omega \to \mathbf{R}^n ; \|X\|_p < \infty\}.$$

With this norm the L^p-spaces are *Banach spaces*, i.e. complete (see Exercise 2.19), normed linear spaces. If $p = 2$ the space $L^2(P)$ is even a *Hilbert space*, i.e. a complete inner product space, with inner product

$$(X, Y)_{L^2(P)} := E[X \cdot Y]; \qquad X, Y \in L^2(P).$$

The mathematical model for *independence* is the following:

Definition 2.1.3 *Two subsets $A, B \in \mathcal{F}$ are called* independent *if*

$$P(A \cap B) = P(A) \cdot P(B) .$$

A collection $\mathcal{A} = \{\mathcal{H}_i ; i \in I\}$ of families $\mathcal{H}_i$ of measurable sets is independent *if*

$$P(H_{i_1} \cap \cdots \cap H_{i_k}) = P(H_{i_1}) \cdots P(H_{i_k})$$

for all choices of $H_{i_1} \in \mathcal{H}_{i_1}, \cdots, H_{i_k} \in \mathcal{H}_{i_k}$ with different indices $i_1, \ldots, i_k$.

A collection of random variables $\{X_i ; i \in I\}$ is independent *if the collection of generated σ-algebras $\mathcal{H}_{X_i}$ is independent.*

If two random variables $X, Y : \Omega \to \mathbf{R}$ are independent then

$$E[XY] = E[X]E[Y] ,$$

provided that $E[|X|] < \infty$ and $E[|Y|] < \infty$. (See Exercise 2.5.)

Definition 2.1.4 *A* stochastic process *is a parametrized collection of random variables*

$$\{X_t\}_{t \in T}$$

defined on a probability space $(\Omega, \mathcal{F}, P)$ and assuming values in $\mathbf{R}^n$.

The parameter space T is usually (as in this book) the halfline $[0, \infty)$, but it may also be an interval $[a, b]$, the non-negative integers and even subsets of $\mathbf{R}^n$ for $n \geq 1$. Note that for each $t \in T$ fixed we have a random variable

$$\omega \to X_t(\omega) ; \qquad \omega \in \Omega .$$

On the other hand, fixing $\omega \in \Omega$ we can consider the function

$$t \to X_t(\omega) ; \qquad t \in T$$

which is called a *path* of X_t.

It may be useful for the intuition to think of t as "time" and each ω as an individual "particle" or "experiment". With this picture $X_t(\omega)$ would represent the position (or result) at time t of the particle (experiment) ω. Sometimes it is convenient to write $X(t, \omega)$ instead of $X_t(\omega)$. Thus we may also regard the process as a function of two variables

$$(t,\omega) \to X(t,\omega)$$

from $T \times \Omega$ into $\mathbf{R}^n$. This is often a natural point of view in stochastic analysis, because (as we shall see) there it is crucial to have $X(t,\omega)$ jointly measurable in (t,ω).

Finally we note that we may identify each ω with the function $t \to X_t(\omega)$ from T into $\mathbf{R}^n$. Thus we may regard Ω as a subset of the space $\widetilde{\Omega} = (\mathbf{R}^n)^T$ of all functions from T into $\mathbf{R}^n$. Then the σ-algebra $\mathcal{F}$ will contain the σ-algebra $\mathcal{B}$ generated by sets of the form

$$\{\omega; \omega(t_1) \in F_1, \cdots, \omega(t_k) \in F_k\}\,, \qquad F_i \subset \mathbf{R}^n \text{ Borel sets}$$

Therefore one may also adopt the point of view that a stochastic process is *a probability measure P on the measurable space* $((\mathbf{R}^n)^T, \mathcal{B})$.

The *(finite-dimensional) distributions* of the process $X = \{X_t\}_{t\in T}$ are the measures $\mu_{t_1,\ldots,t_k}$ defined on $\mathbf{R}^{nk}$, $k = 1, 2, \ldots$, by

$$\mu_{t_1,\ldots,t_k}(F_1 \times F_2 \times \cdots \times F_k) = P[X_{t_1} \in F_1, \cdots, X_{t_k} \in F_k]\,; \qquad t_i \in T\,.$$

Here $F_1, \ldots, F_k$ denote Borel sets in $\mathbf{R}^n$.

The family of all finite-dimensional distributions determines many (but not all) important properties of the process X.

Conversely, given a family $\{\nu_{t_1,\ldots,t_k}; k \in \mathbf{N}, t_i \in T\}$ of probability measures on $\mathbf{R}^{nk}$ it is important to be able to construct a stochastic process $Y = \{Y_t\}_{t\in T}$ having $\nu_{t_1,\ldots,t_k}$ as its finite-dimensional distributions. One of Kolmogorov's famous theorems states that this can be done provided $\{\nu_{t_1,\ldots,t_k}\}$ satisfies two natural consistency conditions: (See Lamperti (1977).)

Theorem 2.1.5 (Kolmogorov's extension theorem)
For all $t_1, \ldots, t_k \in T$, $k \in \mathbf{N}$ let $\nu_{t_1,\ldots,t_k}$ be probability measures on $\mathbf{R}^{nk}$ s.t.

$$\nu_{t_{\sigma(1)},\cdots,t_{\sigma(k)}}(F_1 \times \cdots \times F_k) = \nu_{t_1,\cdots,t_k}(F_{\sigma^{-1}(1)} \times \cdots \times F_{\sigma^{-1}(k)}) \tag{K1}$$

for all permutations σ on $\{1, 2, \ldots, k\}$ and

$$\nu_{t_1,\ldots,t_k}(F_1 \times \cdots \times F_k) = \nu_{t_1,\ldots,t_k,t_{k+1},\ldots,t_{k+m}}(F_1 \times \cdots \times F_k \times \mathbf{R}^n \times \cdots \times \mathbf{R}^n) \tag{K2}$$

for all $m \in \mathbf{N}$, where (of course) the set on the right hand side has a total of $k + m$ factors.

Then there exists a probability space $(\Omega, \mathcal{F}, P)$ and a stochastic process $\{X_t\}$ on $\Omega, X_t\colon \Omega \to \mathbf{R}^n$, s.t.

$$\nu_{t_1,\ldots,t_k}(F_1 \times \cdots \times F_k) = P[X_{t_1} \in F_1, \cdots, X_{t_k} \in F_k]\,,$$

for all $t_i \in T$, $k \in \mathbf{N}$ and all Borel sets F_i.

2.2 An Important Example: Brownian Motion

In 1828 the Scottish botanist Robert Brown observed that pollen grains suspended in liquid performed an irregular motion. The motion was later explained by the random collisions with the molecules of the liquid. To describe the motion mathematically it is natural to use the concept of a stochastic process $B_t(\omega)$, interpreted as the position at time t of the pollen grain ω. We will generalize slightly and consider an n-dimensional analog.

To construct $\{B_t\}_{t\geq 0}$ it suffices, by the Kolmogorov extension theorem, to specify a family $\{\nu_{t_1,\dots,t_k}\}$ of probability measures satisfying (K1) and (K2). These measures will be chosen so that they agree with our observations of the pollen grain behaviour:

Fix $x \in \mathbf{R}^n$ and define

$$p(t,x,y) = (2\pi t)^{-n/2} \cdot \exp\left(-\frac{|x-y|^2}{2t}\right) \qquad \text{for } y \in \mathbf{R}^n,\ t>0\ .$$

If $0 \leq t_1 \leq t_2 \leq \cdots \leq t_k$ define a measure $\nu_{t_1,\dots,t_k}$ on $\mathbf{R}^{nk}$ by

$$\nu_{t_1,\dots,t_k}(F_1 \times \cdots \times F_k) = \tag{2.2.1}$$

$$= \int\limits_{F_1\times\cdots\times F_k} p(t_1,x,x_1)p(t_2-t_1,x_1,x_2)\cdots p(t_k-t_{k-1},x_{k-1},x_k)dx_1\cdots dx_k$$

where we use the notation $dy = dy_1 \cdots dy_k$ for Lebesgue measure and the convention that $p(0,x,y)dy = \delta_x(y)$, the unit point mass at x.

Extend this definition to all finite sequences of t_i's by using (K1). Since $\int\limits_{\mathbf{R}^n} p(t,x,y)dy = 1$ for all $t\geq 0$, (K2) holds, so by Kolmogorov's theorem there exists a probability space $(\Omega, \mathcal{F}, P^x)$ and a stochastic process $\{B_t\}_{t\geq 0}$ on Ω such that the finite-dimensional distributions of B_t are given by (2.2.1), i.e.

$$P^x(B_{t_1} \in F_1, \cdots, B_{t_k} \in F_k) =$$

$$= \int\limits_{F_1\times\cdots\times F_k} p(t_1,x,x_1)\cdots p(t_k-t_{k-1},x_{k-1},x_k)dx_1\dots dx_k\ . \tag{2.2.2}$$

Definition 2.2.1 *Such a process is called (a version of)* Brownian motion starting at x *(observe that* $P^x(B_0 = x) = 1$*).*

The Brownian motion thus defined is not unique, i.e. there exist several quadruples $(B_t, \Omega, \mathcal{F}, P^x)$ such that (2.2.2) holds. However, for our purposes this is not important, we may simply choose any version to work with. As we shall soon see, the paths of a Brownian motion are (or, more correctly, can be chosen to be) continuous, a.s. Therefore we may identify (a.a.) $\omega \in \Omega$ with a continuous function $t \to B_t(\omega)$ from $[0,\infty)$ into $\mathbf{R}^n$. Thus we may adopt the point of view that Brownian motion is just the space $C([0,\infty), \mathbf{R}^n)$ equipped with certain probability measures P^x (given by (2.2.1) and (2.2.2) above).

This version is called the *canonical* Brownian motion. Besides having the advantage of being intuitive, this point of view is useful for the further analysis of measures on $C([0,\infty),\mathbf{R}^n)$, since this space is Polish (i.e. a complete separable metric space). See Stroock and Varadhan (1979).

We state some basic properties of Brownian motion:

(i) B_t is a *Gaussian process*, i.e. for all $0 \le t_1 \le \cdots \le t_k$ the random variable $Z = (B_{t_1},\dots,B_{t_k}) \in \mathbf{R}^{nk}$ has a *(multi)normal distribution.* This means that there exists a vector $M \in \mathbf{R}^{nk}$ and a non-negative definite matrix $C = [c_{jm}] \in \mathbf{R}^{nk\times nk}$ (the set of all $nk \times nk$-matrices with real entries) such that

$$E^x\Big[\exp\Big(i\sum_{j=1}^{nk} u_j Z_j\Big)\Big] = \exp\Big(-\tfrac{1}{2}\sum_{j,m} u_j c_{jm} u_m + i\sum_j u_j M_j\Big) \tag{2.2.3}$$

for all $u = (u_1,\dots,u_{nk}) \in \mathbf{R}^{nk}$, where $i = \sqrt{-1}$ is the imaginary unit and E^x denotes expectation with respect to P^x. Moreover, if (2.2.3) holds then

$$M = E^x[Z] \quad \text{is the mean value of } Z \tag{2.2.4}$$

and

$$c_{jm} = E^x[(Z_j - M_j)(Z_m - M_m)] \quad \text{is the covariance matrix of } Z\,. \tag{2.2.5}$$

(See Appendix A).

To see that (2.2.3) holds for $Z = (B_{t_1},\dots,B_{t_k})$ we calculate its left hand side explicitly by using (2.2.2) (see Appendix A) and obtain (2.2.3) with

$$M = E^x[Z] = (x, x, \cdots, x) \in \mathbf{R}^{nk} \tag{2.2.6}$$

and

$$C = \begin{pmatrix} t_1 I_n & t_1 I_n & \cdots & t_1 I_n \\ t_1 I_n & t_2 I_n & \cdots & t_2 I_n \\ \vdots & \vdots & & \vdots \\ t_1 I_n & t_2 I_n & \cdots & t_k I_n \end{pmatrix}\,. \tag{2.2.7}$$

Hence

$$E^x[B_t] = x \qquad \text{for all } t \ge 0 \tag{2.2.8}$$

and

$$E^x[(B_t - x)^2] = nt,\, E^x[(B_t - x)(B_s - x)] = n\,\min(s,t)\,. \tag{2.2.9}$$

Moreover,

$$E^x[(B_t - B_s)^2] = n(t-s) \text{ if } t \ge s\,, \tag{2.2.10}$$

since

$$\begin{aligned} E^x[(B_t - B_s)^2] &= E^x[(B_t - x)^2 - 2(B_t - x)(B_s - x) + (B_s - x)^2] \\ &= n(t - 2s + s) = n(t-s), \text{when } t \ge s\,. \end{aligned}$$

(ii) B_t has *independent increments*, i.e.

$$B_{t_1}, B_{t_2} - B_{t_1}, \cdots, B_{t_k} - B_{t_{k-1}} \text{ are independent} \\ \text{for all } 0 \le t_1 < t_2 \cdots < t_k \,. \qquad (2.2.11)$$

To prove this we use the fact that normal random variables are independent iff they are uncorrelated. (See Appendix A). So it is enough to prove that

$$E^x[(B_{t_i} - B_{t_{i-1}})(B_{t_j} - B_{t_{j-1}})] = 0 \qquad \text{when } t_i < t_j \,, \qquad (2.2.12)$$

which follows from the form of C:

$$E^x[B_{t_i}B_{t_j} - B_{t_{i-1}}B_{t_j} - B_{t_i}B_{t_{j-1}} + B_{t_{i-1}}B_{t_{j-1}}] \\ = n(t_i - t_{i-1} - t_i + t_{i-1}) = 0 \,.$$

From this we deduce that $B_s - B_t$ is independent of $\mathcal{F}_t$ if $s > t$.

(iii) Finally we ask: Is $t \to B_t(\omega)$ continuous for almost all ω? Stated like this the question does not make sense, because the set $H = \{\omega; t \to B_t(\omega) \text{ is continuous}\}$ is not measurable with respect to the Borel σ-algebra $\mathcal{B}$ on $(\mathbf{R}^n)^{[0,\infty)}$ mentioned above (H involves an uncountable number of t's). However, if modified slightly the question can be given a positive answer. To explain this we need the following important concept:

Definition 2.2.2 *Suppose that $\{X_t\}$ and $\{Y_t\}$ are stochastic processes on $(\Omega, \mathcal{F}, P)$. Then we say that $\{X_t\}$* is a version of *(or a* modification of*)* $\{Y_t\}$ *if*

$$P(\{\omega; X_t(\omega) = Y_t(\omega)\}) = 1 \quad \text{for all } t \,.$$

Note that if X_t is a version of Y_t, then X_t and Y_t have the same finite-dimensional distributions. Thus from the point of view that a stochastic process is a probability law on $(\mathbf{R}^n)^{[0,\infty)}$ two such processes are the same, but nevertheless their path properties may be different. (See Exercise 2.9.)

The continuity question of Brownian motion can be answered by using another famous theorem of Kolmogorov:

Theorem 2.2.3 (Kolmogorov's continuity theorem) *Suppose that the process $X = \{X_t\}_{t\ge 0}$ satisfies the following condition: For all $T > 0$ there exist* positive *constants α, β, D such that*

$$E[|X_t - X_s|^\alpha] \le D \cdot |t-s|^{1+\beta} \,; \qquad 0 \le s,\ t \le T \,. \qquad (2.2.13)$$

Then there exists a continuous version of X.

For a proof see for example Stroock and Varadhan (1979, p. 51).

For Brownian motion B_t it is not hard to prove that (See Exercise 2.8)

$$E^x[|B_t - B_s|^4] = n(n+2)|t-s|^2 \,. \qquad (2.2.14)$$

So Brownian motion satisfies Kolmogorov's condition (2.2.13) with $\alpha = 4$, $D = n(n+2)$ and $\beta = 1$, and therefore it has a continuous version. From now on we will assume that B_t is such a continuous version.

Finally we note that

If $B_t = (B_t^{(1)}, \cdots, B_t^{(n)})$ is n-dimensional Brownian motion, then the 1-dimensional processes $\{B_t^{(j)}\}_{t\geq 0}$, $1 \leq j \leq n$ are independent, 1-dimensional Brownian motions . (2.2.15)

Exercises

2.1. Suppose that $X\colon \Omega \to \mathbf{R}$ is a function which assumes only countably many values $a_1, a_2, \ldots \in \mathbf{R}$.

a) Show that X is a random variable if and only if

$$X^{-1}(a_k) \in \mathcal{F} \qquad \text{for all } k = 1, 2, \ldots \tag{2.2.16}$$

b) Suppose (2.2.16) holds. Show that

$$E[|X|] = \sum_{k=1}^{\infty} |a_k| P[X = a_k] \,. \tag{2.2.17}$$

c) If (2.2.16) holds and $E[|X|] < \infty$, show that

$$E[X] = \sum_{k=1}^{\infty} a_k P[X = a_k] \,.$$

d) If (2.2.16) holds and $f\colon \mathbf{R} \to \mathbf{R}$ is measurable and bounded, show that

$$E[f(X)] = \sum_{k=1}^{\infty} f(a_k) P[X = a_k] \,.$$

2.2. Let $X\colon \Omega \to \mathbf{R}$ be a random variable. The *distribution function* F of X is defined by

$$F(x) = P[X \leq x] \,.$$

a) Prove that F has the following properties:

(i) $0 \leq F \leq 1$, $\lim\limits_{x\to-\infty} F(x) = 0$, $\lim\limits_{x\to\infty} F(x) = 1$.

(ii) F is increasing (= non-decreasing).

(iii) F is right-continuous, i.e. $F(x) = \lim_{\substack{h\to 0\\ h>0}} F(x+h)$.

b) Let $g: \mathbf{R} \to \mathbf{R}$ be measurable such that $E[|g(X)|] < \infty$. Prove that

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) dF(x) ,$$

where the integral on the right is interpreted in the Lebesgue-Stieltjes sense.

c) Let $p(x) \geq 0$ be a measurable function on $\mathbf{R}$. We say that X *has the density* p if

$$F(x) = \int_{-\infty}^{x} p(y) dy \qquad \text{for all } x .$$

Thus from (2.2.1)–(2.2.2) we know that 1-dimensional Brownian motion B_t at time t with $B_0 = 0$ has the density

$$p(x) = \frac{1}{\sqrt{2\pi t}} \exp(-\frac{x^2}{2t}); \quad x \in \mathbf{R} .$$

Find the density of B_t^2.

2.3. Let $\{\mathcal{H}_i\}_{i \in I}$ be a family of σ-algebras on Ω. Prove that

$$\mathcal{H} = \bigcap \{\mathcal{H}_i; i \in I\}$$

is again a σ-algebra.

2.4.* a) Let $X: \Omega \to \mathbf{R}^n$ be a random variable such that

$$E[|X|^p] < \infty \qquad \text{for some } p,\ 0 < p < \infty .$$

Prove *Chebychev's inequality*:

$$P[|X| \geq \lambda] \leq \frac{1}{\lambda^p} E[|X|^p] \qquad \text{for all } \lambda \geq 0 .$$

Hint: $\int_\Omega |X|^p dP \geq \int_A |X|^p dP$, where $A = \{\omega: |X| \geq \lambda\}$.

b) Suppose there exists $k > 0$ such that

$$M = E[\exp(k|X|)] < \infty .$$

Prove that $P[|X| \geq \lambda] \leq M e^{-k\lambda}$ for all $\lambda \geq 0$.

2.5. Let $X, Y: \Omega \to \mathbf{R}$ be two independent random variables and assume for simplicity that X and Y are bounded. Prove that

$$E[XY] = E[X]E[Y] .$$

(Hint: Assume $|X| \le M$, $|Y| \le N$. Approximate X and Y by simple functions $\varphi(\omega) = \sum_{i=1}^{m-1} a_i \mathcal{X}_{F_i}(\omega)$, $\psi(\omega) = \sum_{j=1}^{n-1} b_j \mathcal{X}_{G_j}(\omega)$, respectively, where $F_i = X^{-1}([a_i, a_{i+1}))$, $G_j = Y^{-1}([b_j, b_{j+1}))$, $-M = a_0 < a_1 < \ldots < a_m = M$, $-N = b_0 < b_1 < \ldots < b_n = N$. Then

$$E[X] \approx E[\varphi] = \sum_i a_i P(F_i), \quad E[Y] \approx E[\psi] = \sum_j b_j P(G_j)$$

and
$$E[XY] \approx E[\varphi\psi] = \sum_{i,j} a_i b_j P(F_i \cap G_j) \ldots \Big) .$$

2.6.* Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $A_1, A_2, \ldots$ be sets in $\mathcal{F}$ such that

$$\sum_{k=1}^{\infty} P(A_k) < \infty .$$

Prove the *Borel-Cantelli* lemma:

$$P(\bigcap_{m=1}^{\infty} \bigcup_{k=m}^{\infty} A_k) = 0 ,$$

i.e. the probability that ω belongs to infinitely many $A_k' s$ is zero.

2.7.* a) Suppose $G_1, G_2, \ldots, G_n$ are disjoint subsets of Ω such that

$$\Omega = \bigcup_{i=1}^{n} G_i .$$

Prove that the family $\mathcal{G}$ consisting of $\emptyset$ and all unions of some (or all) of $G_1, \ldots, G_n$ constitutes a σ-algebra on Ω.

b) Prove that any *finite* σ-algebra $\mathcal{F}$ on Ω is of the type described in a).

c) Let $\mathcal{F}$ be a *finite* σ-algebra on Ω and let $X: \Omega \to \mathbf{R}$ be $\mathcal{F}$-measurable. Prove that X assumes only finitely many possible values. More precisely, there exists a disjoint family of subsets $F_1, \ldots, F_m \in \mathcal{F}$ and real numbers $c_1, \ldots, c_m$ such that

$$X(\omega) = \sum_{i=1}^{m} c_i \mathcal{X}_{F_i}(\omega) .$$

2.8. Let B_t be Brownian motion on $\mathbf{R}$, $B_0 = 0$. Put $E = E^0$.

a) Use (2.2.3) to prove that

$$E[e^{iuB_t}] = \exp(-\tfrac{1}{2}u^2 t) \quad \text{for all } u \in \mathbf{R} .$$

b) Use the power series expansion of the exponential function on both sides, compare the terms with the same power of u and deduce that

$$E[B_t^4] = 3t^2$$

and more generally that

$$E\left[B_t^{2k}\right] = \frac{(2k)!}{2^k \cdot k!} t^k \ ; \qquad k \in \mathbf{N} \ .$$

c) If you feel uneasy about the lack of rigour in the method in b), you can proceed as follows: Prove that (2.2.2) implies that

$$E[f(B_t)] = \frac{1}{\sqrt{2\pi t}} \int_{\mathbf{R}} f(x) e^{-\frac{x^2}{2t}} dx$$

for all functions f such that the integral on the right converges. Then apply this to $f(x) = x^{2k}$ and use integration by parts and induction on k.

d) Prove (2.2.14), for example by using b) and induction on n.

2.9.* To illustrate that the (finite-dimensional) distributions alone do not give all the information regarding the continuity properties of a process, consider the following example:
Let $(\Omega, \mathcal{F}, P) = ([0,\infty), \mathcal{B}, \mu)$ where $\mathcal{B}$ denotes the Borel σ-algebra on $[0,\infty)$ and μ is a probability measure on $[0,\infty)$ with no mass on single points. Define

$$X_t(\omega) = \begin{cases} 1 & \text{if } t = \omega \\ 0 & \text{otherwise} \end{cases}$$

and

$$Y_t(\omega) = 0 \quad \text{for all } (t,\omega) \in [0,\infty) \times [0,\infty) \ .$$

Prove that $\{X_t\}$ and $\{Y_t\}$ have the same distributions and that X_t is a version of Y_t. And yet we have that $t \to Y_t(\omega)$ is continuous for all ω, while $t \to X_t(\omega)$ is discontinuous for all ω.

2.10. A stochastic process X_t is called *stationary* if $\{X_t\}$ has the same distribution as $\{X_{t+h}\}$ for any $h > 0$. Prove that Brownian motion B_t has stationary increments, i.e. that the process $\{B_{t+h} - B_t\}_{h \ge 0}$ has the same distribution for all t.

2.11. Prove (2.2.15).

2.12. Let B_t be Brownian motion and fix $t_0 \ge 0$. Prove that

$$\widetilde{B}_t := B_{t_0+t} - B_{t_0} \ ; \qquad t \ge 0$$

is a Brownian motion.

2.13.* Let B_t be 2-dimensional Brownian motion and put

$$D_\rho = \{x \in \mathbf{R}^2; |x| < \rho\} \qquad \text{for } \rho > 0 .$$

Compute

$$P^0[B_t \in D_\rho] .$$

2.14.* Let B_t be n-dimensional Brownian motion and let $K \subset \mathbf{R}^n$ have zero n-dimensional Lebesgue measure. Prove that the expected total length of time that B_t spends in K is zero. (This implies that the *Green measure* associated with B_t is absolutely continuous with respect to Lebesgue measure. See Chapter 9).

2.15.* Let B_t be n-dimensional Brownian motion starting at 0 and let $U \in \mathbf{R}^{n\times n}$ be a (constant) orthogonal matrix, i.e. $UU^T = I$. Prove that

$$\widetilde{B}_t := UB_t$$

is also a Brownian motion.

2.16. **(Brownian scaling).** Let B_t be a 1-dimensional Brownian motion and let $c > 0$ be a constant. Prove that

$$\widehat{B}_t := \frac{1}{c} B_{c^2 t}$$

is also a Brownian motion.

2.17.* If $X_t(\cdot)\colon \Omega \to \mathbf{R}$ is a continuous stochastic process, then for $p > 0$ the *p'th variation process* of X_t, $\langle X, X\rangle_t^{(p)}$ is defined by

$$\langle X, X\rangle_t^{(p)}(\omega) = \lim_{\Delta t_k \to 0} \sum_{t_k \le t} \left|X_{t_{k+1}}(\omega) - X_{t_k}(\omega)\right|^p \quad \text{(limit in probability)}$$

where $0 = t_1 < t_2 < \ldots < t_n = t$ and $\Delta t_k = t_{k+1} - t_k$. In particular, if $p = 1$ this process is called the *total variation process* and if $p = 2$ this is called the *quadratic variation process*. (See Exercise 4.7.) For Brownian motion $B_t \in \mathbf{R}$ we now show that the quadratic variation process is simply

$$\langle B, B\rangle_t(\omega) = \langle B, B\rangle_t^{(2)}(\omega) = t \quad \text{a.s.}$$

Proceed as follows:

a) Define

$$\Delta B_k = B_{t_{k+1}} - B_{t_k}$$

and put

$$Y(t,\omega) = \sum_{t_k \le t} (\Delta B_k(\omega))^2 ,$$

Show that

$$E[(\sum_{t_k \le t}(\Delta B_k)^2 - t)^2] = 2\sum_{t_k \le t}(\Delta t_k)^2$$

and deduce that $Y(t,\cdot) \to t$ in $L^2(P)$ as $\Delta t_k \to 0$.

b) Use a) to prove that a.a. paths of Brownian motion do not have a bounded variation on $[0,t]$, i.e. the total variation of Brownian motion is infinite, a.s.

2.18. a) Let $\Omega = \{1,2,3,4,5\}$ and let $\mathcal{U}$ be the collection

$$\mathcal{U} = \{\{1,2,3\},\{3,4,5\}\}$$

of subsets of Ω. Find the smallest σ-algebra containing $\mathcal{U}$ (i.e. the σ-algebra $\mathcal{H}_{\mathcal{U}}$ generated by $\mathcal{U}$).

b) Define $X : \Omega \to \mathbf{R}$ by

$$X(1) = X(2) = 0, \quad X(3) = 10, \quad X(4) = X(5) = 1.$$

Is X measurable with respect to $\mathcal{H}_{\mathcal{U}}$?

c) Define $Y : \Omega \to \mathbf{R}$ by

$$Y(1) = 0, \quad Y(2) = Y(3) = Y(4) = Y(5) = 1.$$

Find the σ-algebra $\mathcal{H}_Y$ generated by Y.

2.19. Let $(\Omega,\mathcal{F},\mu)$ be a probability space and let $p \in [1,\infty]$. A sequence $\{f_n\}_{n=1}^{\infty}$ of functions $f_n \in L^p(\mu)$ is called a *Cauchy sequence* if

$$\|f_n - f_m\|_p \to 0 \qquad \text{as } n,m \to \infty.$$

The sequence is called *convergent* if there exists $f \in L^p(\mu)$ such that $f_n \to f$ in $L^p(\mu)$.

Prove that every convergent sequence is a Cauchy sequence.

A fundamental theorem in measure theory states that the converse is also true: *Every Cauchy sequence in $L^p(\mu)$ is convergent.* A normed linear space with this property is called *complete*. Thus the $L^p(\mu)$ spaces are complete.

2.20. Let B_t be 1-dimensional Brownian motion, $\sigma \in \mathbf{R}$ be constant and $0 \le s < t$. Use (2.2.2) to prove that

$$E\big[\exp(\sigma(B_s - B_t))\big] = \exp\big(\tfrac{1}{2}\sigma^2(t-s)\big). \tag{2.2.18}$$