ELEMENTS OF PROBABILITY THEORY

- A collection of subsets of a set Ω is called a σ–algebra if it contains Ω and is closed under the operations of taking complements and countable unions of its elements.
- A sub-σ–algebra is a collection of subsets of a σ–algebra which satisfies the axioms of a σ–algebra.
- A *measurable space* is a pair $(\Omega, \mathcal{F})$ where Ω is a set and $\mathcal{F}$ is a σ–algebra of subsets of Ω.
- Let $(\Omega, \mathcal{F})$ and $(E, \mathcal{G})$ be two measurable spaces. A function $X : \Omega \mapsto E$ such that the *event*

$$\{\omega \in \Omega : X(\omega) \in A\} =: \{X \in A\}$$

belongs to $\mathcal{F}$ for arbitrary $A \in \mathcal{G}$ is called a *measurable function* or *random variable*.

- Let $(\Omega, \mathcal{F})$ be a measurable space. A function $\mu : \mathcal{F} \mapsto [0, 1]$ is called a *probability measure* if $\mu(\emptyset) = 1$, $\mu(\Omega) = 1$ and $\mu(\cup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} \mu(A_k)$ for all sequences of pairwise disjoint sets $\{A_k\}_{k=1}^{\infty} \in \mathcal{F}$.

- The triplet $(\Omega, \mathcal{F}, \mu)$ is called a *probability space.*

- Let X be a random variable (measurable function) from $(\Omega, \mathcal{F}, \mu)$ to $(E, \mathcal{G})$. If E is a metric space then we may define *expectation* with respect to the measure μ by

$$\mathbb{E}[X] = \int_{\Omega} X(\omega) \, d\mu(\omega).$$

- More generally, let $f : E \mapsto \mathbb{R}$ be $\mathcal{G}$–measurable. Then,

$$\mathbb{E}[f(X)] = \int_{\Omega} f(X(\omega)) \, d\mu(\omega).$$

- Let U be a topological space. We will use the notation $\mathcal{B}(U)$ to denote the Borel σ–algebra of U: the smallest σ–algebra containing all open sets of U. Every random variable from a probability space $(\Omega, \mathcal{F}, \mu)$ to a measurable space $(E, \mathcal{B}(E))$ induces a probability measure on E:

$$\mu_X(B) = \mathbb{P}X^{-1}(B) = \mu(\omega \in \Omega; X(\omega) \in B), \quad B \in \mathcal{B}(E).$$

The measure μ_X is called the *distribution* (or sometimes the *law*) of X.

Example 1 *Let $\mathcal{I}$ denote a subset of the positive integers. A vector $\rho_0 = \{\rho_{0,i},\, i \in \mathcal{I}\}$ is a distribution on $\mathcal{I}$ if it has nonnegative entries and its total mass equals 1: $\sum_{i \in \mathcal{I}} \rho_{0,i} = 1$.*

- We can use the distribution of a random variable to compute expectations and probabilities:

$$\mathbb{E}[f(X)] = \int_S f(x)\, d\mu_X(x)$$

and

$$\mathbb{P}[X \in G] = \int_G d\mu_X(x), \quad G \in \mathcal{B}(E).$$

- When $E = \mathbb{R}^d$ and we can write $d\mu_X(x) = \rho(x)\, dx$, then we refer to $\rho(x)$ as the *probability density function* (pdf), or *density with respect to Lebesque measure* for X.

- When $E = \mathbb{R}^d$ then by $L^p(\Omega; \mathbb{R}^d)$, or sometimes $L^p(\Omega; \mu)$ or even simply $L^p(\mu)$, we mean the Banach space of measurable functions on Ω with norm

$$\|X\|_{L^p} = \Big(\mathbb{E}|X|^p\Big)^{1/p}.$$

Example 2 *i) Consider the random variable $X : \Omega \mapsto \mathbb{R}$ with pdf*

$$\gamma_{\sigma,m}(x) := (2\pi\sigma)^{-\frac{1}{2}} \exp\left(-\frac{(x-m)^2}{2\sigma}\right).$$

Such an X is termed a **Gaussian** *or* **normal** *random variable.*
The mean is

$$\mathbb{E}X = \int_{\mathbb{R}} x\gamma_{\sigma,m}(x)\,dx = m$$

and the variance is

$$\mathbb{E}(X-m)^2 = \int_{\mathbb{R}} (x-m)^2\gamma_{\sigma,m}(x)\,dx = \sigma.$$

Since the mean and variance specify completely a Gaussian random variable on $\mathbb{R}$, the Gaussian is commonly denoted by $\mathcal{N}(m,\sigma)$. The **standard normal** *random variable is $\mathcal{N}(0,1)$.*

ii) Let $m \in \mathbb{R}^d$ *and* $\Sigma \in \mathbb{R}^{d \times d}$ *be symmetric and positive definite. The random variable* $X : \Omega \mapsto \mathbb{R}^d$ *with pdf*

$$\gamma_{\Sigma,m}(x) := \left((2\pi)^d det\Sigma\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\langle\Sigma^{-1}(x-m),(x-m)\rangle\right)$$

is termed a **multivariate Gaussian** *or* **normal** *random variable. The mean is*

$$\mathbb{E}(X) = m \tag{1}$$

and the covariance matrix is

$$\mathbb{E}\Big((X-m)\otimes(X-m)\Big) = \Sigma. \tag{2}$$

Since the mean and covariance matrix completely specify a Gaussian random variable on $\mathbb{R}^d$*, the Gaussian is commonly denoted by* $\mathcal{N}(m,\Sigma)$.

Example 3 *An exponential random variable* $T : \Omega \to \mathbb{R}^+$ *with rate* $\lambda > 0$ *satisfies*

$$\mathbb{P}(T > t) = e^{-\lambda t}, \quad \forall t \geqslant 0.$$

We write $T \sim \exp(\lambda)$. *The related pdf is*

$$f_T(t) = \begin{cases} \lambda e^{-\lambda t}, & t \geqslant 0, \\ 0, & t < 0. \end{cases} \tag{3}$$

Notice that

$$\mathbb{E}T = \int_{-\infty}^{\infty} t f_T(t) dt = \frac{1}{\lambda} \int_0^{\infty} (\lambda t) e^{-\lambda t} d(\lambda t) = \frac{1}{\lambda}.$$

If the times $\tau_n = t_{n+1} - t_n$ *are i.i.d random variables with* $\tau_0 \sim \exp(\lambda)$ *then, for* $t_0 = 0$,

$$t_n = \sum_{k=0}^{n-1} \tau_k$$

and it is possible to show that

$$\mathbb{P}(0 \leqslant t_k \leqslant t < t_{k+1}) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}. \tag{4}$$

- Assume that $\mathbb{E}|X| < \infty$ and let $\mathcal{G}$ be a sub–σ–algebra of $\mathcal{F}$. The **conditional expectation** of X with respect to $\mathcal{G}$ is defined to be the function $\mathbb{E}[X|\mathcal{G}] : \Omega \mapsto E$ which is $\mathcal{G}$–measurable and satisfies

$$\int_G \mathbb{E}[X|\mathcal{G}] \, d\mu = \int_G X \, d\mu \quad \forall \, G \in \mathcal{G}.$$

- We can define $\mathbb{E}[f(X)|\mathcal{G}]$ and the conditional probability $\mathbb{P}[X \in F|\mathcal{G}] = \mathbb{E}[I_F(X)|\mathcal{G}]$, where I_F is the indicator function of F, in a similar manner.

ELEMENTS OF THE THEORY OF STOCHASTIC PROCESSES

Definition of a Stochastic Process

- Let T be an ordered set. A **stochastic process** is a collection of random variables $X = \{X_t; t \in T\}$ where, for each fixed $t \in T$, X_t is a random variable from $(\Omega, \mathcal{F})$ to $(E, \mathcal{G})$.
- The measurable space $\{\Omega, \mathcal{F}\}$ is called the **sample space**. The space $(E, \mathcal{G})$ is called the **state space** .
- In this course we will take the set T to be $[0, +\infty)$.
- The state space E will usually be $\mathbb{R}^d$ equipped with the σ–algebra of Borel sets.
- A stochastic process X may be viewed as a function of both $t \in T$ and $\omega \in \Omega$. We will sometimes write $X(t), X(t, \omega)$ or $X_t(\omega)$ instead of X_t. For a fixed sample point $\omega \in \Omega$, the function $X_t(\omega) : T \mapsto E$ is called a **sample path** (realization, trajectory) of the process X.

Definition of a Stochastic Process

- The **finite dimensional distributions** (fdd) of a stochastic process are the E^k–valued random variables $(X(t_1), X(t_2), \ldots, X(t_k))$ for arbitrary positive integer k and arbitrary times $t_i \in T$, $i \in \{1, \ldots, k\}$.
- We will say that two processes X_t and Y_t are equivalent if they have same finite dimensional distributions.
- From experiments or numerical simulations we can only obtain information about the (fdd) of a process.

Stationary Processes

- A process is called **(strictly) stationary** if all fdd are invariant under are time translation: for any integer k and times $t_i \in T$, the distribution of $(X(t_1), X(t_2), \ldots, X(t_k))$ is equal to that of $(X(s+t_1), X(s+t_2), \ldots, X(s+t_k))$ for any s such that $s + t_i \in T$ for all $i \in \{1, \ldots, k\}$.
- Let X_t be a stationary stochastic process with finite second moment (i.e. $X_t \in L^2$). Stationarity implies that $\mathbb{E}X_t = \mu$, $\mathbb{E}((X_t - \mu)(X_s - \mu)) = C(t-s)$. The converse is not true.
- A stochastic process $X_t \in L^2$ is called **second-order stationary** (or stationary in the wide sense) if the first moment $\mathbb{E}X_t$ is a constant and the second moment depends only on the difference $t - s$:

$$\mathbb{E}X_t = \mu, \quad \mathbb{E}((X_t - \mu)(X_s - \mu)) = C(t-s).$$

Stationary Processes

- The function $C(t)$ is called the **correlation** (or **covariance**) function of X_t.
- Let $X_t \in L^2$ be a mean zero second order stationary process on $\mathbb{R}$ which is **mean square continuous**, i.e.

$$\lim_{t \to s} \mathbb{E}|X_t - X_s|^2 = 0.$$

- Then the correlation function admits the representation

$$C(t) = \int_{-\infty}^{\infty} e^{itx} f(x)\, dx, \quad t \in \mathbb{R}.$$

- the function $f(x)$ is called the **spectral density** of the process X_t.
- In many cases, the experimentally measured quantity is the spectral density (or power spectrum) of the stochastic process.

Stationary Processes

- Given the correlation function of X_t, and assuming that $C(t) \in L^1(\mathbb{R})$, we can calculate the spectral density through its Fourier transform:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} C(t)\, dt.$$

- The correlation function of a second order stationary process enables us to associate a time scale to X_t, the **correlation time** τ_{cor}:

$$\tau_{cor} = \frac{1}{C(0)} \int_0^{\infty} C(\tau)\, d\tau = \int_0^{\infty} \mathbb{E}(X_\tau X_0)/\mathbb{E}(X_0^2)\, d\tau.$$

- The slower the decay of the correlation function, the larger the correlation time is. We have to assume sufficiently fast decay of correlations so that the correlation time is finite.

Example 4 *Consider a second stationary process with correlation function*

$$C(t) = C(0)e^{-\gamma|t|}.$$

The spectral density of this process is

$$\begin{aligned} f(x) &= \frac{1}{2\pi}C(0)\int_{-\infty}^{\infty} e^{-itx}e^{-\gamma|t|}\,dt \\ &= C(0)\frac{1}{\pi}\frac{\gamma}{\gamma^2+x^2}. \end{aligned}$$

The correlation time is

$$\tau_{cor} = \int_0^{\infty} e^{-\gamma t}\,dt = \gamma^{-1}.$$

Gaussian Processes

- The most important class of stochastic processes is that of Gaussian processes:

Definition 5 *A* **Gaussian process** *is one for which $E = \mathbb{R}^d$ and all the finite dimensional distributions are Gaussian.*

- A Gaussian process $x(t)$ is characterized by its mean

$$m(t) := \mathbb{E}x(t)$$

  and the covariance function

$$C(t,s) = \mathbb{E}\Big(\big(x(t) - m(t)\big) \otimes \big(x(s) - m(s)\big)\Big).$$

- Thus, the first two moments of a Gaussian process are sufficient for a complete characterization of the process.

- A corollary of this is that a second order stationary Gaussian process is also a stationary process.

Brownian Motion

- The most important continuous time stochastic process is **Brownian motion**. Brownian motion is a mean zero, continuous (i.e. it has continuous sample paths: for a.e $\omega \in \Omega$ the function X_t is a continuous function of time) process with independent Gaussian increments.
- A process X_t has **independent increments** if for every sequence $t_0 < t_1 ... t_n$ the random variables

$$X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$$

are independent.
- If, furthermore, for any t_1, t_2 and Borel set $B \subset \mathbb{R}$

$$\mathbb{P}(X_{t_2+s} - X_{t_1+s} \in B)$$

is independent of s, then the process X_t has **stationary independent increments**.

Definition 6 *i) A one dimensional standard* Brownian motion $W(t) : \mathbb{R}^+ \to \mathbb{R}$ *is a real valued stochastic process with the following properties:*

(a) $W(0) = 0$;

(b) $W(t)$ *is continuous;*

(c) $W(t)$ *has independent increments.*

(d) For every $t > s \geqslant 0$ $W(t) - W(s)$ *has a Gaussian distribution with mean* 0 *and variance* $t - s$*. That is, the density of the random variable* $W(t) - W(s)$ *is*

$$g(x; t, s) = \Big(2\pi(t-s)\Big)^{-\frac{1}{2}} \exp\left(-\frac{x^2}{2(t-s)}\right); \tag{5}$$

ii) A d–dimensional standard Brownian motion $W(t) : \mathbb{R}^+ \to \mathbb{R}^d$ *is a collection of d independent one dimensional Brownian motions:*

$$W(t) = (W_1(t), \ldots, W_d(t)),$$

where $W_i(t), i = 1, \ldots, d$ *are independent one dimensional Brownian motions. The density of the Gaussian random vector* $W(t) - W(s)$ *is thus*

$$g(x; t, s) = \Big(2\pi(t - s)\Big)^{-d/2} \exp\left(-\frac{\|x\|^2}{2(t - s)}\right).$$

Brownian motion is sometimes referred to as the **Wiener process** .

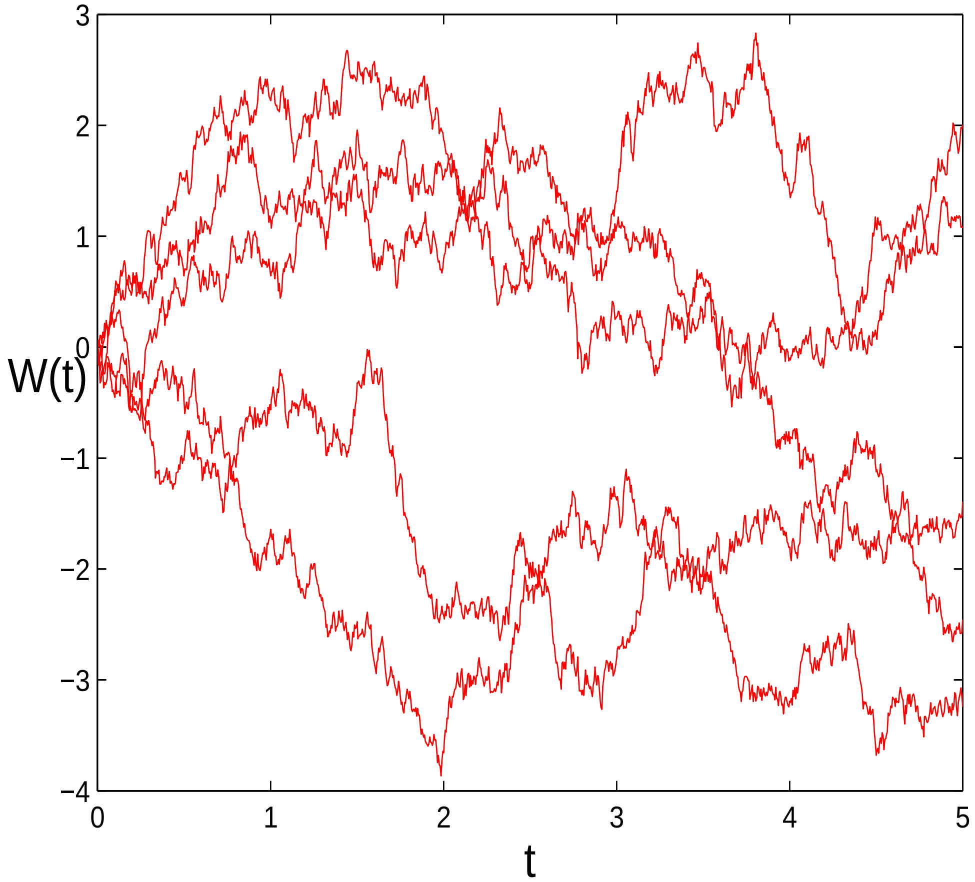

Figure 1: Brownian sample paths

It is possible to prove rigorously the existence of the Wiener process (Brownian motion):

Theorem 1 *(Wiener) There exists an almost-surely continuous process* W_t *with independent increments such and* $W_0 = 0$*, such that for each* $t \geqslant$ *the random variable* W_t *is* $\mathcal{N}(0, t)$*. Furthermore,* W_t *is almost surely locally Hölder continuous with exponent* α *for any* $\alpha \in (0, \frac{1}{2})$*.*

Notice that Brownian paths are not differentiable.

Brownian motion is a Gaussian process. For the d–dimensional Brownian motion, and for I the $d \times d$ dimensional identity, we have (see (1) and (2))

$$\mathbb{E}W(t) = 0 \quad \forall t \geqslant 0$$

and

$$\mathbb{E}\Big((W(t) - W(s)) \otimes (W(t) - W(s))\Big) = (t - s)I. \tag{6}$$

Moreover,

$$\mathbb{E}\Big(W(t) \otimes W(s)\Big) = \min(t, s)I. \tag{7}$$

- From the formula for the Gaussian density $g(x, t-s)$, eqn. (5), we immediately conclude that $W(t) - W(s)$ and $W(t+u) - W(s+u)$ have the same pdf. Consequently, Brownian motion has stationary increments.
- Notice, however, that Brownian motion itself **is not** a stationary process.
- Since $W(t) = W(t) - W(0)$, the pdf of $W(t)$ is

$$g(x,t) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t}.$$

- We can easily calculate all moments of the Brownian motion:

$$\begin{aligned} \mathbb{E}(x^n(t)) &= \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{+\infty} x^n e^{-x^2/2t}\, dx \\ &= \begin{cases} 1.3\ldots(n-1)t^{n/2}, & n \,\text{even}, \\ 0, & n \,\text{odd}. \end{cases} \end{aligned}$$

The Poisson Process

- Another fundamental continuous time process is the **Poisson process** :

  Definition 7 *The Poisson process with intensity* λ*, denoted by* $N(t)$*, is an integer-valued, continuous time, stochastic process with independent increments satisfying*

  $$\mathbb{P}[(N(t) - N(s)) = k] = \frac{e^{-\lambda(t-s)}\big(\lambda(t-s)\big)^k}{k!}, \quad t > s \geqslant 0,\, k \in \mathbb{N}.$$

- Notice the connection to exponential random variables via (4).

- Both Brownian motion and the Poisson process are **homogeneous** (or **time-homogeneous**): the increments between successive times s and t depend only on $t - s$.

The Path Space

- Let $(\Omega, \mathcal{F}, \mu)$ be a probability space, (E, ρ) a metric space and let $T = [0, \infty)$. Let $\{X_t\}$ be a stochastic process from $(\Omega, \mathcal{F}, \mu)$ to (E, ρ) with continuous sample paths.
- The above means that for every $\omega \in \Omega$ we have that $X_t \in C_E := C([0, \infty); E)$.
- The space of continuous functions C_E is called the *path space* of the stochastic process.
- We can put a metric on E as follows:

$$\rho_E(X^1, X^2) := \sum_{n=1}^{\infty} \frac{1}{2^n} \max_{0 \leqslant t \leqslant n} \min\big(\rho(X_t^1, X_t^2), 1\big).$$

- We can then define the Borel sets on C_E, using the topology induced by this metric, and $\{X_t\}$ can be thought of as a random variable on $(\Omega, \mathcal{F}, \mu)$ with state space $(C_E, \mathcal{B}(C_E))$.

The Path Space

- The probability measure $\mathbb{P}X_t^{-1}$ on $(C_E, \mathcal{B}(C_E))$ is called the *law* of $\{X_t\}$.
- The law of a stochastic process is a probability measure on its path space.

  Example 8 *The space of continuous functions C_E is the path space of Brownian motion (the Wiener process). The law of Brownian motion, that is the measure that it induces on $C([0,\infty),\mathbb{R}^d)$, is known as the* **Wiener measure***.*