GAUSSIAN PROCESSES
EXERCISE SHEET 0: FRAMEWORK OF PROBABILITY AND SOME BASIC CALCULATIONS

Exercise 1.

A probability space is a triple $(\Omega, \Sigma, \mathbb{P})$ consisting of a sample space Ω, a sigma-algebra Σ consisting of subsets of Ω and a probability measure $\mathbb{P}\colon \Sigma \to [0,1]$. The sample space is just any set, but philosophically it can be thought of as the set of all possible outcomes of a random system from which the nature or an oracle chooses a single outcome ω at random. The elements of Σ are called events, and the event $A \in \Sigma$ is considered to have happened if $\omega \in A$. The oracle chooses ω in such a way that if the experiments were repeated infinitely many times, then the number of times when $\omega \in A$ divided by the total number of experiments would converge to $\mathbb{P}[A]$. This framework is very flexible and in particular encapsulates both discrete and continuous random variables.

The collection of events is required to satisfy the following natural axioms that define a sigma algebra:

- We have $\Omega \in \Sigma$.
- If $A \in \Sigma$ then $\Omega \setminus A \in \Sigma$.
- If $A_1, A_2, \dots \in \Sigma$ then $\bigcup_{n=1}^{\infty} A_n \in \Sigma$.

Similarly the probability measure is defined to work nicely with the sigma-algebra Σ:

- We have $\mathbb{P}[\Omega] = 1$.
- If $A_1, A_2, \dots \in \Sigma$ are disjoint, then $\mathbb{P}[\bigcup_{n=1}^{\infty} A_n] = \sum_{n=1}^{\infty} \mathbb{P}[A_n]$.

Random variables are measurable maps $X\colon \Omega \to \mathbb{R}$. Measurability means that $X^{-1}(A) \in \Sigma$ for any A in the Borel sigma-algebra on $\mathbb{R}$.

The Borel sigma-algebra on a topological space is the intersection of all sigma-algebras that contain all open sets. ☐

Exercise 2.

Two random variables X and Y are equal in law if $\mathbb{P}[X \in A] = \mathbb{P}[Y \in A]$ for all Borel sets $A \subset \mathbb{R}$. The random variables are equal almost surely if $\mathbb{P}[X \neq Y] = 0$. An example of random variables that are equal in law but not almost surely can be defined e.g. on the single coin toss probability space $(\Omega, \Sigma, \mathbb{P})$ with $\Omega = \{H, T\}$, $\Sigma = \mathcal{P}(\Omega)$, $\mathbb{P}[H] = \mathbb{P}[T] = 1/2$ by setting $X(H) = 1$, $X(T) = 0$, $Y(H) = 0$, $Y(T) = 1$. ☐

Exercise 3.

We can take as our probability space $\mathbb{R}$ with Borel sigma-algebra and probability measure $d\mathbb{P}(x) :=$

$e^{-x^2/2}/\sqrt{2\pi}\,dx$, where dx is the Lebesgue measure. Defining $X(x) = x$ then gives a standard Gaussian random variable. ☐

Exercise 4.

Let us pick as our probability space the Euclidean space $\mathbb{R}^n$ with the Borel sigma-algebra and probability measure $d\mathbb{P}(x) := e^{-|x|^2/2}/\sqrt{2\pi}\,dx$. Then $X_k(x) = x_k$ for $k = 1, \ldots, n$ define n independent standard Gaussians. ☐

Exercise 5.

A sequence $(X_n)_{n=1}^{\infty}$ of random variables converges almost surely to a random variable X if and only if $X_n(\omega) \to X(\omega)$ outside of set of probability 0. The sequence converges in L^p if $\mathbb{E}|X_n - X|^p \to 0$, and in probability if for all $\varepsilon > 0$ we have $\mathbb{P}[|X_n - X| > \varepsilon] \to 0$ as $n \to \infty$. Convergence in law can be defined in various ways and the following are equivalent:

- $X_n \to X$ in law.
- The CDFs satisfy $F_{X_n}(x) \to F_X(x)$ for all x where $F_X(x)$ is continuous.
- We have $\mathbb{E}f(X_n) \to \mathbb{E}f(X)$ for all bounded and continuous f.
- We have $\mathbb{E}e^{itX_n} \to \mathbb{E}e^{itX}$ for all $t \in \mathbb{R}$.

Letting $A_n = X$ and $B = Y$ where X and Y are as in Exercise 2, we see that $A_n \to B$ in law but not in probability.

Consider next a sequence X_n of independent 0/1 random variables with $\mathbb{P}[X_n = 1] = 1/n$ and $\mathbb{P}[X_n = 0] = 1 - 1/n$. Then $\mathbb{P}[|X_n| > \varepsilon] \to 0$ for any $\varepsilon > 0$, so $X_n \to 0$ in probability. However as $\sum_{n=1}^{\infty} \mathbb{P}[X_n = 1] = \infty$ by Borel–Cantelli lemma we see that almost surely infinitely many of X_n are 1 and hence $X_n \not\to 0$ almost surely. ☐

Exercise 6.

The density of X is given by $p_X(x) = \frac{e^{-\frac{x^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}$. Let us set

$$m_n := \mathbb{E}X^n = \int_{-\infty}^{\infty} x^n p_X(x)\,dx = \int_{-\infty}^{\infty} x^n \frac{e^{-\frac{x^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}\,dx.$$

We clearly have $m_0 = 1$ and $m_1 = 0$. Moreover, by integration by parts we get

$$\begin{aligned} m_{n-2} &= \int_{-\infty}^{\infty} x^{n-2} \frac{e^{-\frac{x^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}\,dx = \Big[\frac{x^{n-1}}{n-1} \frac{e^{-\frac{x^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}\Big]_{-\infty}^{\infty} + \frac{1}{\sigma^2(n-1)} \int_{-\infty}^{\infty} x^n \frac{e^{-\frac{x^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}\,dx \\ &= \frac{1}{\sigma^2(n-1)} m_n. \end{aligned}$$

Hence $m_n = \sigma^2(n-1)m_{n-2}$, and we see by induction that

$$m_n = \begin{cases} \sigma^n (n-1)!!, & n \text{ even} \\ 0, & n \text{ odd} \end{cases}.$$

□

Exercise 7.

To prove the upper bound notice that

$$\mathbb{P}(X>t)=\int_t^\infty \frac{e^{-x^2/2}}{\sqrt{2\pi}}\,dx \le \int_t^\infty \frac{x}{t}\frac{e^{-x^2/2}}{\sqrt{2\pi}}\,dx = \Big[\frac{-e^{-x^2/2}}{t\sqrt{2\pi}}\Big]_t^\infty = \frac{e^{-t^2/2}}{t\sqrt{2\pi}}.$$

For the lower bound we may use a similar trick

$$\mathbb{P}(X>t)\ge \int_t^{t+1} \frac{e^{-x^2/2}}{\sqrt{2\pi}}\,dx \ge \int_t^{t+1} \frac{xe^{-x^2/2}}{(t+1)\sqrt{2\pi}}\,dx = \frac{e^{-t^2/2}-e^{-(t+1)^2/2}}{(t+1)\sqrt{2\pi}} \ge \frac{1-e^{-1/2}}{(t+1)\sqrt{2\pi}}e^{-t^2/2}.$$

□

Exercise 8.

We will show that (4) ⇒ (3) ⇒ (2) ⇒ (1) ⇒ (4).

(4) ⇒ (3): Let $\overline{M}:=\mathbb{E}\overline{X}$ and $C:=(\mathbb{E}(X_i-\overline{M}_i)(X_j-\overline{M}_j))_{i,j=1}^n$. Then we have by linearity that $\mathbb{E}\langle\overline{\lambda},\overline{X}\rangle=\langle\overline{\lambda},\overline{M}\rangle$ and moreover for any $\overline{\lambda}\in\mathbb{R}^n$ we have

$$\mathbb{E}|\langle\overline{\lambda},\overline{X}-\overline{M}\rangle|^2=\mathbb{E}\sum_{i,j=1}^n \overline{\lambda}_i\overline{\lambda}_j(\overline{X}_i-\overline{M}_i)(\overline{X}_j-\overline{M}_j)=\sum_{i,j=1}^n \overline{\lambda}_i C_{i,j}\overline{\lambda_j}=\overline{\lambda}^T C\overline{\lambda}.$$

(Note that this also shows that C is positive definite.) Thus the moment generating function of $\langle\overline{\lambda},\overline{X}\rangle$ is given by $e^{\langle\overline{\lambda},\overline{M}\rangle t+\frac{1}{2}\overline{\lambda}^T C\overline{\lambda}t^2}$ and we take $t=1$.

(3) ⇒ (2): From (3) it follows that the Fourier transform $\hat{p}_X(\overline{\xi})$ of the density p_X of X is given by

$$\hat{p}_X(\overline{\xi})=\mathbb{E}e^{-2\pi i\langle\overline{\xi},\overline{X}\rangle}=e^{-2\pi i\langle\overline{\xi},\overline{M}\rangle-2\pi^2\overline{\xi}^T C\overline{\xi}}.$$

Let us compute the Fourier inverse transform of $\hat{p}_X(\overline{\xi})$:

$$p_X(\overline{x})=\int_{\mathbb{R}^n}\hat{p}_X(\overline{\xi})e^{2\pi i\langle\overline{\xi},\overline{x}\rangle}\,d\overline{\xi}=\int_{\mathbb{R}^n}e^{-2\pi i\langle\overline{\xi},\overline{M}-\overline{x}\rangle-2\pi^2\overline{\xi}^T C\overline{\xi}}\,d\overline{\xi}.$$

Since C is positive definite, the matrix $C^{-1/2}$ is well-defined. We may thus make the change of variables $\overline{\xi}\to C^{-1/2}\overline{\xi}$ to get

$$\begin{aligned}p_X(\overline{x})&=\int_{\mathbb{R}^n}e^{-2\pi i\langle\overline{\xi},C^{-1/2}(\overline{M}-\overline{x})\rangle-2\pi^2|\overline{\xi}|^2}\det(C^{-1/2})d\overline{\xi}\\&=\int_{\mathbb{R}^n}e^{-i\langle\overline{\xi},C^{-1/2}(\overline{M}-\overline{x})\rangle-\frac{1}{2}|\overline{\xi}|^2}\det(C^{-1/2})(2\pi)^{-n}d\overline{\xi}\\&=(2\pi)^{-n/2}\det(C^{-1/2})\prod_{j=1}^n\varphi(y_j),\end{aligned}$$

where $\varphi(t)=\mathbb{E}e^{itN(0,1)}=e^{-\frac{t^2}{2}}$ is the characteristic function of a standard Gaussian and $\overline{y}=C^{-1/2}(\overline{x}-\overline{M})$. Hence we see that

$$p_X(\overline{x})=\frac{e^{-\frac{1}{2}(\overline{x}-\overline{M})^T C^{-1}(\overline{x}-\overline{M})}}{(2\pi)^{-n/2}\sqrt{\det(C)}}$$

and the claim is proven with $D = C^{-1}$.

For an alternative proof, we could simply use the uniqueness theorem for the moment generating function. As the random variables in (2) and (3) have the same moment generating functions, they should have the same distribution.

(2) ⇒ (1): Let us define $\overline{Y} = D^{1/2}(\overline{X} - \overline{M})$. Then we have $\overline{X} = A\overline{Y} + \overline{M}$ with $A = D^{-1/2}$. It remains to check that $\overline{Y}$ is a vector of independent standard Gaussians. We have for any Borel set B that

$$\begin{aligned}\mathbb{P}(\overline{Y} \in B) = \mathbb{P}(\overline{X} \in D^{-1/2}B + \overline{M}) &= \int_{D^{-1/2}B+\overline{M}} \frac{e^{-\frac{1}{2}(\overline{x}-\overline{M})^T D(\overline{x}-\overline{M})}}{\sqrt{(2\pi)^n \det(D)^{-1}}}\, d\overline{x} \\ &= \int_{D^{-1/2}B} \frac{e^{-\frac{1}{2}\overline{x}^T D\overline{x}}}{\sqrt{(2\pi)^n \det(D)^{-1}}}\, d\overline{x} = \int_B \frac{e^{-\frac{1}{2}|\overline{y}|^2}}{\sqrt{(2\pi)^n}}\, d\overline{y},\end{aligned}$$

which shows that Y has the right density.

(1) ⇒ (4): We have

$$\langle \overline{\lambda}, \overline{X} \rangle = \langle \overline{\lambda}, A\overline{Y} \rangle + \langle \overline{\lambda}, \overline{M} \rangle = \langle A^T\overline{\lambda}, \overline{Y} \rangle + \langle \overline{\lambda}, \overline{M} \rangle,$$

and since A has full rank, $A^T\overline{\lambda} \neq 0$. Thus $\langle \overline{\lambda}, \overline{X} \rangle$ is a sum of non-degenerate independent Gaussians and hence a non-degenerate Gaussian.

Let us finally note that unlike matrices C and D, the matrix A is not necessarily unique (in particular it does not have to be symmetric), but one can get C from A by setting $C = AA^T$. □