Solutions for $\mathbb{E}$xercise sheet 1 : Conditional probability distributions, first properties of the Brownian Motion.

Solution 0 — *Gaussian vectors.* (1) The parameters are the mean $\mu \in \mathbb{R}$ and the variance $\sigma^2 \geq 0$. When $\sigma^2 = 0$, the distribution is just the Dirac in μ, and when $\sigma^2 > 0$, it has pdf $f(t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-t^2/(2\sigma^2)}$. In both cases the characteristic function is $\phi(t) = e^{i\mu t - \sigma^2/2t^2}$.

(2) This is immediate to check. By decomposing on the standard Euclidean basis it turns out that $m_i = \mathbb{E}[X_i]$ and $\Sigma_{i,j} = \mathrm{Cov}(X_i, X_j)$. We call those the mean vector and the covariance matrix of X.

(3) We have that $\langle t, X\rangle$ is a Gaussian of mean $\langle t, m\rangle$ and variance $\langle t, \Sigma t\rangle$. So by taking the characteristic function of $\langle t, X\rangle$ at point 1 we get $\mathbb{E}[e^{i\langle t,X\rangle}] = \exp(i\langle t, m\rangle - \frac{1}{2}\langle t, \Sigma t\rangle)$. So the distribution of X is completely characterized by the parameters m and Σ.

(4) Compute $\mathbb{E}[e^{i\langle t,Ax\rangle}] = \mathbb{E}[e^{i\langle {}^\intercal At,x\rangle}] = \exp(i\langle {}^\intercal At, m\rangle - \frac{1}{2}\langle {}^\intercal At, \Sigma {}^\intercal At\rangle) = \exp(i\langle t, Am\rangle - \frac{1}{2}\langle t, A\Sigma {}^\intercal At\rangle)$. Gaussianity and identification of the parameters follows.

(5) If we have the independence condition, then for $t \in V_1$ and $s \in V_2$, we have $\mathrm{Cov}[\langle t, X\rangle, \langle s, X\rangle] = 0$ by Fubini's theorem (justified since everybody is in L^2). But the converse is also true: Suppose that for every $t \in V_1$ and $s \in V_2$, we have $\mathrm{Cov}[\langle t, X\rangle, \langle s, X\rangle] = 0$. Let $f_1, \dots f_m$ be a finite family in V_1 followed by a finite family in V_2. Set $Y = (\langle f_1, X\rangle, \dots, \langle f_m, X\rangle) = (Y_1, Y_2)$. Then, by computing covariances, we see that the covariance matrix of Y is block-diagonal. This means that we have a product decomposition $\mathbb{E}[e^{i(\langle t_1, Y1\rangle + \langle t_2, Y_2\rangle}] = \mathbb{E}[e^{i\langle t_1, Y_1\rangle}]\,\mathbb{E}[e^{i\langle t_2, Y_2\rangle}]$. By injectivity of the characteristic distribution, we have identified the distribution of (Y_1, Y_2) as one of an independent couple of two Gaussian vectors. Now because by definition the σ-algebra spanned by a family of variables is generated by the finite subfamilies, we get the independence of the two σ-algebras.

(6) The classic example : set (X, A) to be an independent couple of a standard Gaussian and a Rademacher variable (uniform on $\{\pm 1\}$). Set $Y = AX$. Then Y is not independent of X ($\mathbb{P}(X > 0, Y > 0) = 0 \neq 1/4$). Yet $\mathrm{Cov}(X, Y) = \mathbb{E}[AX^2] = \mathbb{E}[A]\,\mathbb{E}[X^2] = 0 \times 1 = 0$.

(7) If $X = (X_1, \dots X_n)$ then we compute $\mathbb{E}[e^{i\langle t,X\rangle}] = e^{-\frac{1}{2}\langle t,t\rangle}$. So it's Gaussian. For m a vector and Σ a semi-definite positive matrix, use the spectral theorem to write $\Sigma = {}^\intercal ODO$, and consider $Y = m + {}^\intercal O\sqrt{D}X$. It should have the prescribed parameters.

Solution 1 — *Conditioning and independence.*

- Set $u(x) = \mathbb{E}[f(x, Y)] = \int f(x, y) d\,\mathbb{P}_Y(y)$. According to Fubini's theorem, $u(x)$ is defined $\mathbb{P}_X$-a.e. Let us check that the almost-surely defined random variable $u(X)$ satisfies the universal property required from the conditional expectation $\mathbb{E}[f(X, Y) \mid \mathcal{G}]$.

  Let Z be a $\mathcal{G}$-measurable bounded random variable. Then $Zf(X, Y) \in L^1$, and since Y is independent of (X, Z), which means $\mathbb{P}_{(X,Z,Y)} = \mathbb{P}_{(X,Z)} \otimes \mathbb{P}_Y$.

  We deduce

$$\begin{aligned}
\mathbb{E}[Zf(X, Y)] &= \int zf(x, y) d\,\mathbb{P}_{(X,Z,Y)}(x, z, y) = \int zf(x, y) d(\mathbb{P}_{(X,Z)} \otimes \mathbb{P}_Y)(x, z, y) \\
&= \int z \left(\int f(x, y) d\,\mathbb{P}_Y(y) \right) d\,\mathbb{P}_{(X,Z)}(x, z) \text{ (Fubini)} \\
&= \mathbb{E}[Zu(X)].
\end{aligned}$$

  This proves the claim. I often write this very basic claim about conditional expectations as follows :

$$\mathbb{E}[f(X, Y) \mid \mathcal{G}] = \mathbb{E}[f(x, Y)]_{x=X}.$$

- We may now interpret this as a conditional distribution. Let $\mu(x, \cdot)$ denote the distribution of $f(x, Y)$. Then for every bounded measurable ϕ,

$$\mathbb{E}[\phi(f(X, Y))|\mathcal{G}] = \mathbb{E}[\phi(f(x, Y))]_{x=X} = \left(\int \phi(u)\mu(x, du) \right)_{x=X} = \int \phi(u)\mu(X, du).$$

  This implies that the distribution of $f(X, Y)$ given $\mathcal{G}$ is $\mu(X, \cdot)$. In other words, μ is a conditional probability kernel for $f(X, Y)$ given X.

Solution 2 — *Gaussian conditional distribution and Bayesian statistics 101.* (1) To do this, we project X on $\sigma(Y)$ to write

$$X = \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(Y)} Y + \left(X - \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(Y)} Y \right),$$

the two terms of this sum being uncorrelated hence independent, as they themselves form a Gaussian vector. Writing Z the second term, we end up with

$$X = \frac{\rho}{\sigma_Y^2} Y + Z$$

, where Z is independent of Y. We deduce $\mathrm{Var}(X) = \frac{\rho^2}{\sigma_Y^4} \mathrm{Var}(Y) + \mathrm{Var}(Z)$ (Pythagora's !), and hence $\mathrm{Var}(Z) = \sigma_X^2 - \frac{\rho^2}{\sigma_Y^2}$. Using the previous exercise, we deduce that the conditional probability kernel of X given Y is

$$(y, \cdot) \mapsto \mathbb{P}(\frac{\rho}{\sigma_Y^2} y + Z \in \cdot) = \mathcal{N}(\frac{\rho}{\sigma_Y^2} y, \sigma_X^2 - \frac{\rho^2}{\sigma_Y^2})(\cdot).$$

(3) Applying the previous question, we get that

$$\mathbb{P}_{\theta|\overline{X}=\overline{x}} = \mathcal{N}\left(\frac{\overline{x}}{1+\frac{\sigma^2}{n\tau^2}}, \frac{1}{\frac{n}{\sigma^2}+\frac{1}{\tau^2}}\right)$$

(4) (a) The limit as $\sigma \to \infty$ is $\mathcal{N}(0,\tau^2)$. When the observations are very random, they give no information about θ.
(b) The limit as $\sigma \to 0$ is $\mathcal{N}(\overline{x},0) = \delta_{\overline{x}}$. When the observations are not random, they equal θ almost surely, hence the distribution of θ given the observations is not random.
(c) The limit as $\tau \to \infty$ is $\mathcal{N}(\overline{x},\sigma^2/n)$. The prior distribution of θ is very random hence contains no information. That is why the conditional distribution given $\overline{X}$ is not biased towards 0 anymore. Note that we recover the point of view of *inferential statistics*: when θ is unknown but deterministic, we indeed have $\theta - \overline{x} \sim \mathcal{N}(0,\sigma^2/n)$.
(d) The limit as $\tau \to 0$ is $\mathcal{N}(0,0) = \delta_0$. Indeed since the prior distribution of θ becomes deterministically equal to 0, then the posterior does too.

(5) We may interpret this as follows: a real-world parameter θ must be measured. Prior (theoretical or based on the past) knowledge gives us its *a priori* distribution $\mathcal{N}(0,\tau^2)$. We are also given noisy measurements $X_1,\dots,X_n$ of this parameter, and wonder what the distribution of θ becomes after adding this supplementary information.

(6) It turns out that the conditional distribution of θ given $(X_1,\dots X_n)$ is the same as the one given $\overline{X}$. Indeed if we replay the proof of question 1 and project θ on $\overline{X}$, we get

$$\theta = \frac{n\tau^2}{n\tau^2+\sigma^2}\overline{X} + Z,$$

and it turns out that not only $\mathrm{Cov}(\overline{X},Z)=0$ but also $\mathrm{Cov}(X_i,Z)=0$, $1 \le i \le n$. Hence we may continue as in question 1.

Solution 3 — *Borel-Kolmogorov paradox.*
We start by computing the joint distribution of (θ,ϕ).

$$\begin{aligned}
\mathbb{E}[h(\theta_P,\phi_P)] &= \int h(\theta_p,\phi_p)\,\mathbb{P}_P(dp) \\
&= \frac{3}{4\pi}\int_{B_{\mathbb{R}^3}(0,1)} h(\theta_{p/|p|},\phi_{p/|p|})\mathrm{Leb}_3(dp) \\
&= \frac{3}{4\pi}\int_0^1 r^2 dr \int_{-\pi}^{\pi} d\theta \int_{-\pi/2}^{\pi/2} \cos(\phi)d\phi\, h(\theta,\phi) \\
&= \int_{-\pi}^{\pi}\frac{d\theta}{2\pi}\int_{-\pi/2}^{\pi/2}\frac{\cos(\phi)d\phi}{2}\,h(\theta,\phi),
\end{aligned}$$

where we applied Lebesgue's change of variable theorem in line 3, setting

$$p = (r\cos(\theta)\cos(\phi), r\sin(\theta)\cos(\phi), r\sin(\phi)),$$

which gives

$$\begin{aligned}\mathrm{Leb}_3(dp) &= r^2\cos(\phi)drd\theta d\phi\\ \theta_{p/|p|} &= \theta\\ \phi_{p/|p|} &= \phi.\end{aligned}$$

On the last line, we read that ϕ_P and θ_P are independent, θ has uniform distribution on $[-\pi,\pi]$, while ϕ has density $\cos(\phi)/2$ on $[-\pi/2,\pi/2]$.

(1) With a step further in the computation above, we may deduce that $(\theta_P \mod \pi, \mathrm{sign}(\theta_P), \phi_P)$ are independent random variables whose respective distributions are : uniform in $[0,\pi]$, uniform in $\{-1,1\}$, and with density $\cos(\phi)/2$. From exercise 1, we deduce that conditional on $\theta_P \mod \pi = \theta$, the distribution of P is that of a point of latitude uniformly chosen in $\{\theta, \theta-\pi\}$ and longitude chosen in $[-\pi/2,\pi/2]$ with density $\cos(\phi)/2$.

(2) It comes directly from exercise 1 that conditional on $\phi_P = \phi$, the distribution of P is that of a point with latitude ϕ and uniform longitude.

(3) A conditional probability kernel given some variable Z is only defined up to $\mathbb{P}_Z$-almost-everywhere equality, so it does not really make sense to specialize it at a given point z. However, if there is a continuous representative (i.e. $z \mapsto \mu(z,\cdot)$ is continuous in the space of probability measures), then it is unique. Hence specialization makes sense. This is the case for the two conditional probability kernels defined above.

(4) The paradox is that both procedures yield a probability measure on some great circle of the sphere, that are really different. In one case the measure is the image of the Lebesgue measure in $\mathbb{S}^1$, while in the other case it is not. It comes from the fact that conditioning on negligible events is not well-defined.

Solution 4 — *Transformations.*
We first consider the finite-dimensional marginals of the new process $(X_t)_t$ in these two cases. Remark at first that they still form centered Gaussian vectors, since they are each obtained by a very simple linear transform of some f.d.m. of B. Now we only need to compute covariances.

(1) $\mathrm{Cov}(X_s,X_t) = \mathrm{Cov}(\lambda^{-1/2}B_{\lambda s}, \lambda^{-1/2}B_{\lambda t}) = \lambda^{-1}\mathrm{Cov}(B_{\lambda s}, B_{\lambda t}) = \lambda^{-1}(\lambda s \wedge \lambda t) = s\wedge t.$

(2) For $0 \le s,t \le 1$, $\mathrm{Cov}(X_s,X_t) = \mathrm{Cov}(B_1 - B_{1-s}, B_1 - B_{1-t}) = \mathrm{Cov}(B_1,B_1) - \mathrm{Cov}(B_1,B_{1-t}) - \mathrm{Cov}(B_{1-s},B_1) + \mathrm{Cov}(B_{1-s},B_{1-t}) = 1-(1-t)-(1-s)+(1-s)\wedge(1-t) = 1+(t-1)\wedge(s-1) = t\wedge s.$

Now since those processes are continuous on their domain of definition, they are Brownian motions.

Solution 5 — *A nowhere continuous version of the Brownian motion.*
Let $(X_t)_t$ be a Brownian motion and $(U_i)_i$ be an independent sequence of independent exponential random variables with parameter 1.
Let us show the following property: with probability one, $(U_i)_i$ is dense in $[0, \infty)$. Let $a < b \in \mathbb{Q}$. $\mathbb{P}(U_1 \notin [a, b], \ldots, U_n \notin [a, b]) = \mathbb{P}(U_1 \notin [a, b])^n \to 0$ as $n \to \infty$. So $\mathbb{P}(U_i \notin [a, b] \forall i) = 0$. We have shown $\forall a < b \in \mathbb{Q}^2$, almost surely, $[a, b]$ intersects $(U_i)_i$. Because $\mathbb{Q}^2$ is countable, we can invert $\forall$ and almost surely, and we get that almost surely, $(U_i)_i$ is dense.
Now we define $B_t = X_t + \mathbb{1}_{t \notin \{U_i, i \in \mathbb{N}\}}$. By the previous property, this process is almost surely nowhere continuous, and we can check that the f.d.m's of B and X are equal almost surely (so have the same distribution) because for fixed $t_1, \ldots, t_k$, the probability that $\{t_1, \ldots, t_k\}$ intersects $\{U_i, i \in \mathbb{N}\}$ is 0 (once again by countable union).
Now we modify B on the negligible set where it is still continuous despite all our efforts, by setting $B = \mathbb{1}_{\mathbb{Q}}$, finishing the exercise.

Solution 6 — *Brownian motion is nowhere monotonous.*
Let us fix $a < b \in \mathbb{Q}$ and $a < t_0 < \ldots < t_k < b$. If the Brownian motion is increasing, it implies that $B_{t_i} - B_{t_{i_1}} \geq 0$ for every $1 \leq i \leq k$. So $\{B \text{ increasing on } [a, b]\} \subset \bigcap_{1 \leq i \leq k} \{B_{t_i} - B_{t_{i_1}} \geq 0\}$. This last event has probability 2^{-k} by independence. So $\{B \text{ increasing on } [a, b]\}$ can be included in an event of probability 0. So by countable union

$$\{B \text{ increasing on some interval }\} \subset \cup_{a < b \in \mathbb{Q}} \{B \text{ increasing on } [a, b]\}$$

can be included in an event of probability 0. So the complement property "B is increasing on no nontrivial interval" is almost sure. Same for "decreasing" by symmetry.
Remark: We did not need to show that the property "B is monotonous on no nontrivial interval" is indeed an event (i.e. is a measurable set), because the property of being almost sure or negligible can be defined for non-measurable subsets of Ω. But we can check that it is an event because of the assumption that paths are always continuous.

Solution 7 — *The stationary Ornstein-Uhlenbeck process.*
Firstly, $\mathrm{Cov}(X_t, X_s) = e^{-|t-s|}$. So at each time t, X_t is a standard Gaussian. It does not have independent increments as a quick computation shows.