Solutions for $\mathbb{E}$xercise sheet 1 : Gaussian vectors, random walks.

Solution 1 — *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 2 — *Limit in distribution of Gaussian vectors.*
Thanks to the student who found this neat proof of question 1
Let μ_n and σ_n be the parameters of X_n

(1) If we have convergence in distribution, then we have convergence of the characteristic functions to the one of the limit. So there exists a characteristic function $f : \mathbb{R} \to \mathbb{R}$ such that for all $t \in \mathbb{R}$, $f_n(t) = e^{i\mu_n t - \frac{\sigma_n^2}{2}t^2} \to f(t)$. Now taking the modulus then the log yields $\sigma_n^2 \to -\frac{2}{t^2}\log(|f(t)|) = \sigma^2 \geq 0$. We deduce that $|f(t)| = e^{-\frac{\sigma^2}{2}t^2}$. Now
$$e^{i\mu_n t} = e^{\frac{\sigma_n^2}{2}t^2} f_n(t) \to e^{\frac{\sigma^2}{2}t^2} f(t) =: u(t),$$
where u is a continuous function in $\mathbb{C}$ of modulus 1 (with $u(0) =: 1$). Taking ϵ small enough so that the $\mathrm{Re}(u(t)) > 1/2$ for $t \in (0, \varepsilon)$, we can integrate the previous convergence and get
$$\frac{1}{i\mu_n}(e^{i\mu_n\varepsilon} - 1) \to \int_{t=0}^{\varepsilon} u(t)dt \neq 0$$
Upon inverting then multiplying by $(e^{i\mu_n\varepsilon} - 1) \to u(\varepsilon) - 1$, we obtain that μ_n converges to $\frac{u(\varepsilon)-1}{i\int_{t=0}^{\varepsilon} u(t)dt}$. We then have convergence of the parameters, and $X \sim \mathcal{N}(\mu, \sigma^2)$ with $\mu = \lim \mu_n$ and $\sigma^2 = \lim \sigma_n^2$.

(2) By the first question, if there is convergence in probability, there is convergence in distribution, hence μ_n and σ_n converge. Since moments of Gaussian variables are polynomials in (μ_n, σ_n), we have convergence of all moments. Hence $(X_n)_n$ is bounded in every L^p, $p > 1$. Hence $(|X_n|^{p'})$ is uniformly integrable for every p'. By Vitali's theorem, $X_n \to X$ in $L_{p'}$.

Solution 3 — *Some estimates.*
Let $(X_n)_{n\geq 0}$ be the simple symmetric random walk.

(1) $\mathbb{P}(X_{2n} = 0) = \binom{2n}{n}2^{-2n}$. Stirling's formula gives us
$$\mathbb{P}(X_{2n} = 0) \sim \frac{1}{\sqrt{\pi n}}.$$

(2) Using the ballot theorem for the walk between times 0 and $n-1$,
$$\begin{aligned}
&\mathbb{P}(X_n = 0, X_1 \cdots X_{n-1} \neq 0) \\
&= \tfrac{1}{2}\,\mathbb{P}(X_{n-1} = -1, X_1 \cdots X_{n-2} \neq 0) + \tfrac{1}{2}\,\mathbb{P}(X_{n-1} = 1, X_1 \cdots X_{n-2} \neq 0) \\
&= \frac{1}{n-1}(\tfrac{1}{2}\,\mathbb{P}(X_{n-1} = 1) + \tfrac{1}{2}\,\mathbb{P}(X_{n-1} = -1)) \\
&= \frac{1}{n-1}\,\mathbb{P}(X_n = 0)
\end{aligned}$$
Hence $\mathbb{P}(X_n = 0, X_1 \cdots X_{n-1} \neq 0) \sim \mathbb{1}_{n \text{ even}} \sqrt{\frac{2}{\pi}} n^{-3/2}$.

(3) By summation of equivalents,

$$\mathbb{P}(\tau_0 \geq k) = \sum_{n=k}^{\infty} \mathbb{P}(X_n = 0, X_1 \cdots X_{n-1} \neq 0) \sim \sqrt{\frac{2}{\pi}} \sum_{2n \geq k} (2n)^{-3/2} \sim \sqrt{\frac{2}{\pi}} 2^{-3/2} \frac{(k/2)^{-1/2}}{1/2} = \sqrt{\frac{2}{\pi k}}$$

(4) In class, you showed that $\frac{\mathbb{E}[|X_n|]}{n} = \mathbb{P}(\tau_0 > n)$, from which we deduce

$$\mathbb{E}[|X_n|] \sim \sqrt{\frac{2n}{\pi}}.$$

Solution 4 — *Maximum and hitting times.*
Will be completed later.

Solution 5 — *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 6 — *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 $\mathrm{Cov}(X_i, Z) = 0$, $1 \leq i \leq n$. Hence we may continue as in question 1.