Chapter 1: Brownian motion as a Gaussian process

Contents

1 Gaussian random variables

We begin by recalling the basic properties of Gaussian random variables. Brownian motion will shortly be introduced as a particular instance of a Gaussian process, and much of what makes such processes tractable ultimately reduces to elementary facts about Gaussian variables and vectors.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X : (\Omega, \mathcal{F}) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be a random variable. Recall that the law of X, denoted $\mathbb{P}_X$, is the push-forward (image) measure of $\mathbb{P}$ under X:

$$\mathbb{P}_X(dx) = X(\mathbb{P})(dx) = X_{\#}\,\mathbb{P}(dx) = \mathbb{P}(X \in dx),$$

a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

Definition 1. *A random variable X is called a* <u>*standard Gaussian random variable*</u> *(notation $X \sim \mathcal{N}(0,1)$, or equivalently said to be* <u>*centered, with unit variance*</u>*) if its law is absolutely continuous with respect to Lebesgue measure, with density*

$$\mathbb{P}_X(dx) = \frac{1}{\sqrt{2\pi}} \exp\big(-\frac{1}{2}x^2\big)\, dx.$$

More generally, a random variable Y is called <u>Gaussian</u> if there exist $\sigma \geq 0$ and $m \in \mathbb{R}$ such that $Y = m + \sigma X$ for some standard Gaussian X. In this case the law of Y is again absolutely continuous, with density

$$\mathbb{P}_Y(dx) = \frac{1}{\sigma\sqrt{2\pi}} \exp\Big(-\frac{1}{2}\big(\frac{x-m}{\sigma}\big)^2\Big)\, dx.$$

Our first result shows that the class of Gaussian random variables is closed under the natural notion of convergence in probability, and moreover that such convergence automatically upgrades to convergence in every L^p. This result is much deeper than it looks at first sight; it is the essential ingredient needed to contruct Gaussian spaces.

Proposition 2. *Let (Y_n) be a sequence of Gaussian random variables such that $Y_n \xrightarrow{\mathbb{P}} Y$ for some random variable Y. Then Y is itself Gaussian, and moreover $Y_n \xrightarrow{L^p} Y$ for every $p \geq 1$.*

Proof of Proposition 2. Since convergence in probability implies convergence in distribution, we have $Y_n \xrightarrow{d} Y$, and hence, for every $t \in \mathbb{R}$,

$$\mathbb{E}[e^{itY_n}] \longrightarrow \mathbb{E}[e^{itY}].$$

Writing $Y_n = m_n + \sigma_n X_n$ with X_n standard Gaussian, the left-hand side is explicit:

$$\mathbb{E}[e^{itY_n}] = e^{itm_n - \frac{1}{2}t^2\sigma_n^2} =: g_n(t).$$

Since $g_n(t)$ converges (to a limit we temporarily denote $\ell(t)$) for every t, so does its modulus, and therefore

$$e^{-\frac{1}{2}t^2\sigma_n^2} \longrightarrow |\ell(t)|, \qquad \text{i.e.} \qquad \sigma_n^2 \longrightarrow \sigma_\infty^2$$

for some limit $\sigma_\infty^2 \in [0,\infty]$ (the value $\sigma_\infty^2 = 0$ corresponds to the degenerate case where Y is almost surely constant; also, σ_n^2 cannot diverge to ∞, otherwise $\mathbb{E}[e^{itY}]$ would vanish identically, which is incompatible with Y being a genuine random variable). It then follows that $e^{itm_n} \to \ell(t)/|\ell(t)|$ as well, from which we deduce that m_n converges to some $m \in \mathbb{R}$. Indeed, suppose first that (m_n) is bounded, and let m_∞, m'_∞ be two subsequential limits; then $e^{itm_\infty} = e^{itm'_\infty}$ for every $t \in \mathbb{R}$, which forces $m_\infty = m'_\infty$, so that the whole sequence converges. Suppose instead that (m_n) is unbounded, and extract a subsequence $m_{n_k} \to +\infty$. Fix $a \in \mathbb{R}$ and choose $b > a$ such that $F_Y(x) = \mathbb{P}(Y \leq x)$ is continuous at b (possible since F_Y has at most countably many discontinuities). By weak convergence along the subsequence,

$$F_Y(a) \leq F_Y(b) = \lim_{k\to\infty} \mathbb{P}(Y_{n_k} \leq b) \leq \lim_{k\to\infty} \mathbb{P}(Y_{n_k} \leq m_{n_k}) = \tfrac{1}{2},$$

the last equality holding because $(Y_{n_k} - m_{n_k})/\sigma_{n_k}$ is a standard Gaussian variable, symmetric about 0. As this holds for every $a \in \mathbb{R}$, we would obtain $F_Y(a) \leq \frac{1}{2}$ for all a, contradicting $F_Y(a) \to 1$ as $a \to \infty$. This contradiction rules out the unbounded case. We conclude that $m_n \to m$ and $\sigma_n^2 \to \sigma_\infty^2 < \infty$, so that $Y \sim \mathcal{N}(m, \sigma_\infty^2)$ is indeed Gaussian.

It remains to upgrade convergence in probability to convergence in L^p, for every $p \geq 1$. We first record a classical lemma.

Lemma 3. *If (Z_n) is a sequence of random variables such that $Z_n \xrightarrow{\mathbb{P}} 0$ and (Z_n) is uniformly integrable, then $Z_n \xrightarrow{L^1} 0$.*

Recall that a family of (real-valued) random variables $(X_i)_{i\in I}$ is uniformly integrable (u.i.) if

$$\sup_{i\in I} \mathbb{E}\big[|X_i|\, \mathbf{1}_{|X_i|>a}\big] \xrightarrow[a\to\infty]{} 0,$$

and that a simple sufficient criterion is boundedness in L^p for some $p > 1$: indeed, by Hölder's inequality with $\frac{1}{p} + \frac{1}{q} = 1$,

$$\mathbb{E}\big[|X_i|\mathbf{1}_{|X_i|>a}\big] \leq \mathbb{E}[|X_i|^p]^{1/p}\, \mathbb{P}(|X_i| > a)^{1/q} \leq \underbrace{\mathbb{E}[|X_i|^p]^{1/p}}_{\lesssim 1} \cdot \frac{1}{a^{p/q}} \underbrace{\mathbb{E}[|X_i|^p]^{1/q}}_{\lesssim 1} \xrightarrow[a\to\infty]{} 0,$$

uniformly in i.

Proof of Lemma 3. Fix $a > 0$ and split

$$\mathbb{E}[|Z_n|] = \mathbb{E}\big[|Z_n|\mathbf{1}_{|Z_n|<a}\big] + \mathbb{E}\big[|Z_n|\mathbf{1}_{|Z_n|>a}\big].$$

The second term is at most $\varepsilon/2$ for a large enough, uniformly in n, by uniform integrability. For the first term, further split, for any $\delta > 0$,

$$\mathbb{E}\big[|Z_n|\mathbf{1}_{|Z_n|<a}\big] = \mathbb{E}\big[|Z_n|\mathbf{1}_{|Z_n|<a}\mathbf{1}_{|Z_n|>\delta}\big] + \mathbb{E}\big[|Z_n|\mathbf{1}_{|Z_n|<a}\mathbf{1}_{|Z_n|<\delta}\big] \leq a\,\mathbb{P}(|Z_n| > \delta) + \delta.$$

Choosing $\delta = \varepsilon/2$ and letting $n \to \infty$, the first term on the right tends to 0 since $Z_n \xrightarrow{\mathbb{P}} 0$; hence $\mathbb{E}\big[|Z_n|\mathbf{1}_{|Z_n|<a}\big] \le \varepsilon/2$ eventually, for a fixed. Combining both estimates yields $\mathbb{E}[|Z_n|] \le \varepsilon$ for n large, which proves the Lemma. $\square$

Returning to the Proposition, it therefore suffices to check that $\sup_n \mathbb{E}[|Y_n - Y|^{p'}] < \infty$ for every $p' \ge 1$, since the Lemma applied to $(|Y_n - Y|^p)$ shows that $(|Y_n - Y|^p)$ is uniformly integrable (take p' such that $|Y_n - Y|^{p'} = (|Y_n - Y|^p)^{p''}$ with $p'' > 1$) so that convergence in probability to convergence in L^p. We estimate, for a suitable constant c_p,

$$\mathbb{E}\big[|Y_n - Y|^p\big] \le c_p\big(\sup_n \mathbb{E}[|Y_n|^p] + \mathbb{E}[|Y|^p]\big) \le c'_p(|m_n|^p + \sigma_n^{2p}\,\gamma_p) + c''_p,$$

where $\gamma_p = \mathbb{E}[|\xi|^p]$ for $\xi \sim \mathcal{N}(0,1)$, which is finite: indeed $\mathbb{E}[\xi^{2p+1}] = 0$ and $\mathbb{E}[\xi^{2p}] = \dfrac{(2p)!}{2^p p!}$, as follows from the identity $\mathbb{E}[e^{i\xi t}] = e^{-\frac{1}{2}\sigma^2 t^2} e^{imt}$ (with $\sigma = 1, m = 0$), or equivalently, after the substitution $z = it$ and analytic continuation to $z \in \mathbb{C}$, from $\mathbb{E}[e^{z\xi}] = e^{z^2/2}$. Since (m_n) and (σ_n^2) are convergent, hence bounded, the right-hand side is bounded uniformly in n, which completes the proof of Proposition 2 $\square$

2 Gaussian space, Gaussian process

We now introduce the language of Gaussian spaces, which will allow us to manipulate whole families of Gaussian random variables — and eventually stochastic processes — with the same ease as individual ones.

Definition 4. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A Gaussian space is a closed vector subspace of $L^2(\Omega, \mathcal{F}, \mathbb{P})$ all of whose elements are centered Gaussian random variables.*

Definition 5. *Let $(X_t)_{t\in T}$ be a family of random variables. We say that $(X_t)_{t\in T}$ is a Gaussian process if the following two conditions (which turn out to be equivalent) hold:*

(i) $\mathrm{vect}\{X_t,\ t \in T\}$*, the closure in L^2 of the vector space spanned by the X_t's, is a Gaussian space;*

(ii) every finite linear combination of the X_t's is a centered Gaussian random variable.

That (i) implies (ii) is immediate; the converse, that (ii) implies (i), is a consequence of the projection results of Section 1 (a limit in L^2 of centered Gaussian variables is again centered Gaussian, by the Proposition above).

Example 6. *Let $(X_1,\dots,X_n)$ be a Gaussian vector, i.e. suppose $G=\mathrm{vect}\{X_1,\dots,X_n\}$ is a Gaussian space. Writing $X=\sum_{j=1}^n \lambda_j X_j$ for arbitrary $\lambda_1,\dots,\lambda_n\in\mathbb{R}$, we have*

$$\mathbb{E}\Big[\exp\Big(i\sum_{j=1}^n \lambda_j X_j\Big)\Big]=\mathbb{E}\big[e^{iX}\big]=\exp\Big(-\frac{1}{2}\mathbb{E}[X^2]\Big)=\exp\Big(-\frac{1}{2}\sum_{i,j=1}^n \lambda_i\lambda_j\,\mathbb{E}[X_iX_j]\Big),$$

a quadratic form in $(\lambda_1,\dots,\lambda_n)$. Since the characteristic function of $(X_1,\dots,X_n)$ is thus entirely determined by the second-order moments $\mathbb{E}[X_iX_j]$, these second-order characteristics identify the law of the whole vector $(X_1,\dots,X_n)$.

The following proposition is the cornerstone of the whole theory: within a Gaussian space, orthogonality (a second-order, L^2 notion) and independence (a genuinely probabilistic notion) coincide.

Proposition 7. *Let $(G_i)_{i\in I}$ be a family of Gaussian spaces, all contained in a common Gaussian space G. Then the σ-fields $\{\sigma(G_i),\ i\in I\}$ are independent if and only if the spaces G_i are pairwise orthogonal in L^2.*

Proof. Suppose first that the $\sigma(G_i)$ are independent. For $X_i\in G_i$ and $X_j\in G_j$ with $i\neq j$, independence gives $\mathbb{E}[X_iX_j]=\mathbb{E}[X_i]\mathbb{E}[X_j]=0$, since both are centered; hence $G_i\perp G_j$.

Conversely, suppose the G_i are pairwise orthogonal, and let us show that if $X_{i_1}\in G_{i_1},\dots,X_{i_n}\in G_{i_n}$ are drawn from n distinct spaces, then $(X_{i_1},\dots,X_{i_n})$ are independent random variables (from which independence of the σ-fields $\sigma(G_i)$ follows; we leave the passage from finite families to full σ-fields as an exercise). For $\lambda_1,\dots,\lambda_n\in\mathbb{R}$,

$$\begin{aligned}\mathbb{E}\Big[\exp\Big(i\sum_{k=1}^n \lambda_k X_{i_k}\Big)\Big]&=\exp\Big(-\frac{1}{2}\sum_{k,k'=1}^n \lambda_k\lambda_{k'}\,\mathbb{E}[X_{i_k}X_{i_{k'}}]\Big)\\&=\exp\Big(-\frac{1}{2}\sum_{k=1}^n \lambda_k^2\,\mathbb{E}[X_{i_k}^2]\Big)\\&=\prod_{k=1}^n \exp\Big(-\frac{1}{2}\lambda_k^2\,\mathbb{E}[X_{i_k}^2]\Big)=\prod_{k=1}^n \mathbb{E}\big[e^{i\lambda_k X_{i_k}}\big],\end{aligned}$$

where we successively used that $X_{i_k}\in G$, a Gaussian space and the orthogonality of the distinct G_{i_k}. Since the joint characteristic function factorises as the product of the marginal ones, $X_{i_1},\dots,X_{i_n}$ are independent, as claimed. □

An immediate and extremely useful consequence describes conditional expectations within a Gaussian space purely in terms of Hilbert space projections.

Corollary 8. *Let G be a Gaussian space and $G' \subset G$ a closed subspace.*

(i) For every $X \in G$,

$$\mathbb{E}[X \mid \sigma(G')] = \mathrm{proj}_{G'}(X),$$

where $\mathrm{proj}_{G'}(X)$ denotes the orthogonal (Hilbert space) projection of X onto G' in $L^2(\Omega, \sigma(G'), \mathbb{P})$. As an immediate consequence, $X - \mathrm{proj}_{G'}(X)$ is independent of $\sigma(G')$.

(ii) Setting $\sigma^2 = \mathbb{E}\big[(X - \mathrm{proj}_{G'}(X))^2\big]$, the regular conditional law of X given $\sigma(G')$ is again Gaussian: for every Borel set $A \subset \mathbb{R}$,

$$\mathbb{P}\big(X \in A \mid \sigma(G')\big)(\omega) = Q(\omega, A) = \frac{1}{\sigma\sqrt{2\pi}} \int_A \exp\Big(-\frac{1}{2}\Big(\frac{y - \mathrm{proj}_{G'}(X)(\omega)}{\sigma}\Big)^2\Big)\, dy,$$

$\mathbb{P}(d\omega)$-almost surely, with the convention $Q(\omega, A) = \delta_{\mathrm{proj}_{G'}(X)(\omega)}(A)$ when $\sigma = 0$.

Proof. (i) Set $Y = X - \mathrm{proj}_{G'}(X)$. Since Y is by construction orthogonal to G', and both X and G' lie in the Gaussian space G, the Proposition above shows that Y is independent of $\sigma(G')$. Consequently,

$$\mathbb{E}[X \mid \sigma(G')] = \mathbb{E}[\mathrm{proj}_{G'}(X) \mid \sigma(G')] + \mathbb{E}[Y \mid \sigma(G')] = \mathrm{proj}_{G'}(X) + 0,$$

the first term being $\sigma(G')$-measurable and the second vanishing by independence and centering.

(ii) Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a bounded measurable function. Using $X = \mathrm{proj}_{G'}(X) + Y$ and the independence of Y from $\sigma(G')$,

$$\mathbb{E}\big[\varphi(X) \mid \sigma(G')\big] = \mathbb{E}\big[\varphi(\mathrm{proj}_{G'}(X) + Y) \mid \sigma(G')\big] = \int \varphi\big(\mathrm{proj}_{G'}(X) + y\big)\, \mathbb{P}_Y(dy),$$

where $\mathbb{P}_Y = \mathcal{N}(0, \sigma^2)$ by construction, and where the last equality is an instance of the elementary fact that, for Z measurable with respect to a σ-field $\mathcal{G}$ and Y independent of $\mathcal{G}$,

$$\mathbb{E}\big[g(Y, Z) \mid \mathcal{G}\big] = \int g(y, Z)\, \mathbb{P}_Y(dy).$$

This yields precisely the stated Gaussian conditional density. $\square$

Remark 9. *For a general $X \in L^2$, one only has $\mathbb{E}[X \mid \sigma(G')] = \mathrm{proj}_{L^2(\Omega, \sigma(G'), \mathbb{P})}(X)$, the L^2-projection onto the (possibly much larger) space $L^2(\Omega, \sigma(G'), \mathbb{P})$. The remarkable feature of the Gaussian setting is that this conditional expectation coincides with the Hilbert projection onto the much smaller subspace G' itself.*

3 Gaussian measure, Brownian motion

We now construct Gaussian processes indexed by an arbitrary real, separable Hilbert space H, of which Brownian motion will emerge as the canonical example associated with $H = L^2(\mathbb{R}_+, dt)$.

Theorem 10. *Given such an H, there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ carrying a family $(X(h),\ h \in H)$ of centered Gaussian random variables such that*

$$\forall h, k \in H : \qquad \mathbb{E}[X(h)X(k)] = \langle h, k\rangle_H,$$

and such that the map $h \mapsto X(h)$, from H to $L^2(\Omega, \mathcal{F}, \mathbb{P})$, is linear.

Several basic properties follow at once:

(i) $\mathbb{E}\big[(X(h) - X(k))^2\big] = |h|_H^2 + |k|_H^2 - 2\langle h, k\rangle_H = |h - k|_H^2$;

(ii) consequently, the map $X : H \to L^2(\Omega, \mathcal{F}, \mathbb{P})$ is an isometry;

(iii) if $h_n \to h$ in H, then $X(h_n) \to X(h)$ in L^2, since $\mathbb{E}\big[(X(h_n) - X(h))^2\big] = \langle h_n - h, h_n - h\rangle_H \to 0$.

In particular, $(X(h),\ h \in H)$ spans a Gaussian space (which is closed, being the isometric image of the closed space H).

Proof. We treat separately the finite- and infinite-dimensional cases.
Case $\dim H < \infty$. Let $(e_1, \dots, e_n)$ be an orthonormal basis of H and let $(\xi_1, \dots, \xi_n)$ be i.i.d. standard Gaussian random variables on some $(\Omega, \mathcal{F}, \mathbb{P})$. Define, for $h = \sum_{i=1}^n (h, e_i) e_i$,

$$X(h) = \sum_{i=1}^{n} \langle h, e_i\rangle_H\, \xi_i.$$

Then $\mathbb{E}[X(h)X(k)] = \sum_{i,j} \langle h, e_i\rangle_H \langle k, e_j\rangle_H\, \mathbb{E}[\xi_i \xi_j] = \sum_i \langle h, e_i\rangle_H \langle k, e_i\rangle_H = \langle h, k\rangle_H$, as required.
Case $\dim H = \infty$. Take $(\Omega, \mathcal{F}, \mathbb{P})$ rich enough to support a sequence $(\xi_i)_{n\geq 1}$ of independent, centered, standard Gaussian random variables, and let $(e_i)_{n\geq 1}$ be a Hilbert basis of H. For $h = \sum_i \langle h, e_i\rangle_H e_i$, define

$$X(h) = \sum_{i=1}^{\infty} \langle h, e_i\rangle_H\, \xi_i,$$

a series which converges in $L^2(\Omega, \mathcal{F}, \mathbb{P})$: indeed, its partial sums form a Cauchy sequence, since

$$\mathbb{E}\Big[\Big(\sum_{i=n}^{m}\langle h, e_i\rangle_H \xi_i\Big)^2\Big] = \sum_{i=n}^{m}\langle h, e_i\rangle^2 \xrightarrow[n,m\to\infty]{} 0,$$

the tail of a convergent series $\sum_i \langle h, e_i\rangle_H^2 = |h|_H^2$. One checks similarly that $h \mapsto X(h)$ is linear, i.e. that $\mathbb{E}\big[(X(\alpha h + \beta k) - \alpha X(h) - \beta X(k))^2\big] = 0$. $\square$

Example 11. *Take $H = L^2(T, \mathcal{G}, \mu)$ for some σ-finite measure $\mu \geq 0$ on $(T, \mathcal{G})$, and write $X(f)$ for the corresponding family; one calls X a Gaussian measure of intensity μ. In the particular case $f = \mathbf{1}_A$ with $A \in \mathcal{G}$, $\mu(A) < \infty$, one writes $X(A) = X(\mathbf{1}_A)$, so that $X(A) \sim \mathcal{N}(0, \mu(A))$ for every such A; and if $A = \sum_n A_n$ is a countable disjoint union with $\mu(A_n) < \infty$, then $X(A) = \sum_n X(A_n)$, the series converging in $L^2(\Omega, \mathcal{F}, \mathbb{P})$.*

We may now specialise this construction to obtain Brownian motion itself.

Definition 12. *Take $T = \mathbb{R}_+$, $\mathcal{G} = \mathcal{B}(\mathbb{R}_+)$, and $\mu(dt) = dt$ the Lebesgue measure. The process*

$$B_t = X([0,t]) = X(\mathbf{1}_{[0,t]}), \qquad t \in \mathbb{R}_+,$$

is a Gaussian process, and $(B_t,\ t \geq 0)$ is by definition a Brownian motion (BM).

Properties.

(i) For all $A, C \in \mathcal{G}$, we always have

$$\mathbb{E}[X(A)X(C)] = \mathbb{E}[X(\mathbf{1}_A)X(\mathbf{1}_C)] = \langle \mathbf{1}_A, \mathbf{1}_C\rangle_\mu = \int \mathbf{1}_A \mathbf{1}_C \, d\mu = \mu(A \cap C).$$

Applying this with $A = [0,t]$, $C = [0,s]$ immediately gives the celebrated covariance formula

$$\mathbb{E}[B_t B_s] = \int_0^\infty \mathbf{1}_{[0,t]}(x)\, \mathbf{1}_{[0,s]}(x)\, dx = \min(s,t).$$

(ii) For any $t_1 < t_2 < \cdots < t_n$, the increments $B_{t_1}, B_{t_2} - B_{t_1}, \ldots, B_{t_n} - B_{t_{n-1}}$ are centered Gaussian random variables which are, moreover, independent. Indeed, by linearity of $X(\cdot)$,

$$X(\mathbf{1}_{[0,t_2]}) - X(\mathbf{1}_{[0,t_1]}) = X\big(\mathbf{1}_{[0,t_2]} - \mathbf{1}_{[0,t_1]}\big) = X\big(\mathbf{1}_{[t_1,t_2]}\big),$$

and similarly for the other increments, so that the increments listed above coincide with $X([0,t_1]), X([t_1,t_2]), \ldots, X([t_{n-1},t_n])$; since the corresponding indicator functions have disjoint supports, they are orthogonal in $L^2(dx)$, and hence, by the Proposition of Section 2, the corresponding random variables are independent.

More generally, if $h_1, \dots, h_p \in L^2(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+), dt)$ are pairwise orthogonal, then $X(h_1), \dots, X(h_p)$ are independent random variables.

Notation 13. *For $f \in L^2(\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+), dt)$, write*

$$X(f) = \int_0^\infty f(t)\, dB_t,$$

and call this the Wiener integral of f.

Exercise.

(i) If f is continuous with compact support, then

$$\int_0^t f(s)\, dB_s = \lim_{|\Delta| \to 0} \sum_i f(t_i)\big(B_{t_i} - B_{t_{i-1}}\big) \quad \text{in } L^2(\Omega),$$

the limit being taken over subdivisions of $[0, t]$ of mesh tending to 0 (this integral coincides with $\int_0^\infty f(s) \mathbf{1}_{[0,t]}(s)\, dB_s$).

(ii) If moreover $f \in C^1$, an integration by parts gives

$$\int_0^t f(s)\, dB_s = f(t) B_t - \int_0^t f'(s)\, B_s\, ds.$$

(Hint: show that $X(f\mathbf{1}_{[0,t]}) = f(t)X([0,t]) - \int_0^t f'(s)\, X([0,s])\, ds$; see the concluding remark below regarding the measurability, in s, of $\omega \mapsto X([0,s])(\omega)$.)

We collect, in the following proposition, the basic invariance properties enjoyed by Brownian motion; together they express the remarkable richness of the symmetries of this single object.

Proposition 14 (Basic properties of Brownian motion)**.** *Let $\mathcal{F}_t = \sigma(B_s, s \le t)$ denote the natural filtration of B. Then:*

*(i) **(Time-homogeneity)** For every fixed $s \ge 0$, the process $(B_{t+s} - B_s, t \ge 0)$ is again a Brownian motion, independent of $\mathcal{F}_s$. (To be proved later.)*

*(ii) **(Symmetry)** $(-B_t, t \ge 0)$ has the same distribution as $(B_t, t \ge 0)$, and is thus itself a Brownian motion.*

*(iii) **(Scaling)** For every $c > 0$, $(c\, B_{t/c^2}, t \ge 0)$ is a Brownian motion.*

(iv) ***(Time-inversion)*** $(t\,B_{1/t},\, t \geq 0)$*, with the convention that the process takes the value* 0 *at* $t = 0$*, is a Brownian motion.*

(v) ***(Time-reversal)*** $(B_1 - B_{1-t},\, t \leq 1)$ *is a Brownian motion (on* $[0,1]$*).*

The proof is left to the exercise sessions.

Concluding remark. It is worth stressing already at this stage that, for a general $f \in L^2$, the Wiener integral $\int_0^t f(s)\,dB_s$ cannot be defined pathwise, i.e. separately for each fixed ω: "$dB_s(\omega)$" is not a genuine (signed) measure in s, and $\int_0^t f(s)\,dB_s(\omega)$ is therefore not a genuine (Lebesgue-Stieltjes) integral. Rather, the Wiener integral is best understood through the identity $X(f) \sim \mathcal{N}\big(0, \int_0^t f(s)^2\,ds\big)$, together with the picture of a Gaussian process indexed by L^2 functions developed above. This is already, in embryonic form, the beginning of what it means to make sense of a stochastic integral , a theme to which we shall return at length.

4 Brownian motion as a smooth (continuous) process

Given a Brownian motion $(B_t)_{t\geq 0}$, it is tempting to think of it as producing, for each outcome $\omega \in \Omega$, a genuine function (a "path") $t \mapsto B_t(\omega)$ from $\mathbb{R}_+$ to $\mathbb{R}$. This, however, is not entirely straightforward: it is not even clear a priori, and indeed it need not be true, depending on how B was constructed, that $t \mapsto B_t(\omega)$ is measurable for a fixed ω. The purpose of this section is to show that, up to a suitable modification, one may always assume that Brownian motion has continuous paths.

Definition 15. *Let* $(X_t)_{t\in T}$ *and* $(\widetilde{X}_t)_{t\in T}$ *be two processes indexed by the same set* T*, taking values in the same space* E*. We say that* $\widetilde{X}$ *is a modification of* X *if*

$$\forall t \in T: \qquad \mathbb{P}(\widetilde{X}_t = X_t) = 1. \tag{$*$}$$

Two processes related by $(*)$ automatically share the same law, in the sense of having the same finite-dimensional distributions (a notion of "law of a process" that we shall discuss more thoroughly later on). One should be careful, however, that $(*)$ is a statement made separately for each fixed t: the negligible exceptional set on which $\widetilde{X}_t \neq X_t$ is in general allowed to depend on t, and this dependence cannot in general be removed.

This subtlety has a striking consequence: two modifications of one another may have entirely different trajectories. For instance, take $T = \mathbb{R}_+$, $X_t \equiv 0$, and let $S \sim \mathcal{U}_{[0,1]}$ be a uniform random variable; define

$$\widetilde{X}_t = \begin{cases} 1 & \text{if } t = S, \\ 0 & \text{if } t \neq S. \end{cases}$$

Since $\mathbb{P}(\widetilde{X}_t \neq 0) = \mathbb{P}(S = t) = 0$ for every fixed t, $\widetilde{X}$ is indeed a modification of X; yet $\widetilde{X}$ is not continuous (its path has a single spike, at the random time S). One may push this further: setting $\widetilde{X}_t = 1$ if $t - S \in \mathbb{Q}$ and $\widetilde{X}_t = 0$ otherwise still defines a modification of X, and now $\widetilde{X}$ is everywhere discontinuous.

These examples motivate the introduction of a strictly stronger notion of comparison between two processes.

Definition 16. *We say that $\widetilde{X}$ is indistinguishable from X if*

$$\mathbb{P}\big(\forall t \in T :\ X_t = \widetilde{X}_t\big) = 1$$

(one should note that the event $\{\forall t \in T,\ X_t = \widetilde{X}_t\}$ need not be measurable, so that this statement should be read with the appropriate care (the complement set being included in a measurable $\mathbb{P}$-null set for instance).

If two processes are indistinguishable, then in particular one is a modification of the other; the converse, as the examples above show, is false in general: modification does not imply indistinguishability. If two processes are indistinguishable, they moreover have almost surely the same trajectories, as functions of t. However, when $T \subset \mathbb{R}_+$ and both X and $\widetilde{X}$ are known in advance to have almost surely continuous trajectories, the gap between the two notions closes up entirely:

Proposition 17. *If $T \subset \mathbb{R}_+$ and $X, \widetilde{X}$ have almost surely continuous trajectories, then X is a modification of $\widetilde{X}$ if and only if X and $\widetilde{X}$ are indistinguishable.*

Proof. One implication is immediate, as noted above. Conversely, if $\widetilde{X}$ is a modification of X, then almost surely $X_t = \widetilde{X}_t$ for every t in the countable set $T \cap \mathbb{Q}$; by continuity of both paths, this equality then extends almost surely to every $t \in T$, so that X and $\widetilde{X}$ are indistinguishable. (The same argument applies, mutatis mutandis, to càdlàg processes, replacing $\mathbb{Q}$ by rational approximation from the right.) $\square$

It is precisely this proposition that makes the search for a continuous modification of Brownian motion so valuable: once found, it will be essentially unique, unambiguous up to indistinguishability. This is the content of the fundamental theorem we now state and prove.

Theorem 18 (Kolmogorov's continuity criterion)**.** *Let $(X_t)_{t\in I}$, with $I \subset \mathbb{R}$ bounded, be a stochastic process taking values in a complete metric space (E,d). Suppose there exist constants $q, \varepsilon, C > 0$ such that*

$$\forall s,t \in I: \qquad \mathbb{E}\big[d(X_s,X_t)^q\big] \le C\,|t-s|^{1+\varepsilon}.$$

Then there exists a modification $\widetilde{X}$ of X such that, for every $\alpha \in (0,\varepsilon/q)$, the map $t \mapsto \widetilde{X}_t(\omega)$ is α-Hölder continuous for (almost) every ω: more precisely, for every such α there exists a random variable $C_\alpha(\omega)$ such that

$$\forall s,t \in I: \qquad d\big(\widetilde{X}_s(\omega),\widetilde{X}_t(\omega)\big) \le C_\alpha(\omega)\,|t-s|^{\alpha}.$$

In particular $\widetilde{X}$ is a continuous modification of X, unique up to indistinguishability by the previous Proposition; and one may moreover choose C_α so that $\mathbb{E}[C_\alpha^p] < \infty$ for every $p \ge 1$.

Before turning to the proof of this essential result, a few remarks and applications are in order.

(i) The assumption $\varepsilon > 0$ is necessary, and cannot be relaxed to $\varepsilon = 0$. Consider for instance a Poisson process $(N_t)_{t\ge 0}$, $N_t = \sum_{n\ge 1} \mathbf{1}_{T_n \le t}$, where $T_0 = 0$ and the inter-arrival times $T_n - T_{n-1}$ are i.i.d. $\mathrm{Exp}(\lambda)$. Then

$$\mathbb{E}\big[|N_t-N_s|^p\big] = \mathbb{E}[N_{t-s}^p] = \mathbb{E}\big[\mathrm{Poiss}(\lambda(t-s))^p\big] \le C_p\,|t-s|,$$

which is exactly of Kolmogorov type with $\varepsilon = 0$; yet N has (almost surely) no continuous modification, since its paths jump. This shows that the strict positivity of ε cannot be dispensed with.

(ii) If $I = \mathbb{R}_+$ rather than a bounded interval, one simply covers $\mathbb{R}_+ = \bigcup_{k\ge 0}[k,k+1]$ and glues together the locally Hölder modifications obtained on each compact piece $[k,k+1]$.

As a first, and central, application we obtain the celebrated Hölder regularity of Brownian paths.

Corollary 19. *Brownian motion $(B_t)_{t\ge 0}$ admits a continuous modification which is, in addition, locally Hölder continuous of order $\frac{1}{2} - \delta$, for every $\delta \in \big(0,\frac{1}{2}\big)$.*

Proof. For $s < t$, $B_t - B_s \sim \mathcal{N}(0, t-s)$, so that for every $q > 0$,

$$\mathbb{E}\big[|B_t - B_s|^q\big] = (t-s)^{q/2}\,\mathbb{E}[|N|^q] = c_q\,(t-s)^{q/2}.$$

For any $q > 2$, this is precisely the hypothesis of Kolmogorov's criterion with exponent $\varepsilon = \frac{q}{2} - 1 > 0$; we thus obtain a continuous modification of B which is α-Hölder for every $\alpha < \dfrac{q-2}{2q}$. Letting $q \to \infty$, the admissible range of Hölder exponents $\dfrac{q-2}{2q} \uparrow \dfrac{1}{2}$ exhausts every exponent strictly below $\frac{1}{2}$, which is the claimed statement. ■

Exercise. *Show that the resulting continuous version $(t, \omega) \mapsto \widetilde{B}_t(\omega)$ can furthermore be chosen to be jointly $\mathcal{B}(\mathbb{R}_+) \otimes \mathcal{F}$-measurable.*

From now on, Brownian motion will always refer, tacitly, to this continuous modification. We defer the proof of this theorem to the end of the chapter, once the canonical construction of Section 5 is in place.

5 Canonical representation of Brownian motion; the Wiener measure

Having established the existence of a continuous version of Brownian motion, we now describe a particularly convenient way of constructing it once and for all: rather than working with an arbitrary probability space $(\Omega, \mathcal{F}, \mathbb{P})$ carrying a Brownian motion, we transport the whole construction onto the space of continuous functions itself.

Let $\mathcal{C}(\mathbb{R}_+, \mathbb{R})$ denote the space of continuous real-valued functions defined on $\mathbb{R}_+$. Given a continuous Brownian motion B on $(\Omega, \mathcal{F}, \mathbb{P})$, one may regard B as a single random variable taking values in $\mathcal{C}(\mathbb{R}_+, \mathbb{R})$, via the map

$$\Omega \longrightarrow \mathcal{C}(\mathbb{R}_+, \mathbb{R}), \qquad \omega \longmapsto \big(t \mapsto B_t(\omega)\big).$$

To make sense of the measurability of this map, we equip $\mathcal{C}(\mathbb{R}_+, \mathbb{R})$ with the smallest σ-field, denoted $\mathfrak{C}$, making all the coordinate (or canonical) maps

$$X_t : \mathcal{C}(\mathbb{R}_+, \mathbb{R}) \to \mathbb{R}, \qquad X_t(\omega) = \omega(t) \qquad (t \geq 0)$$

measurable; that is, $\mathfrak{C} = \sigma(X_t,\ t \geq 0)$.

Exercise. *Show that $\mathfrak{C}$ coincides with the Borel σ-field on $\mathcal{C}(\mathbb{R}_+,\mathbb{R})$ associated with the topology of uniform convergence on compact sets.*

Definition 20 (Wiener measure)**.** *Let B be a Brownian motion on $(\Omega,\mathcal{F},\mathbb{P})$, and let $\psi:\Omega\to\mathcal{C}(\mathbb{R}_+,\mathbb{R})$, $\psi(\omega)=\big(t\mapsto B_t(\omega)\big)$, be the associated path map, which is $\mathfrak{C}$-measurable[1]. The Wiener measure is the push-forward of $\mathbb{P}$ under ψ:*

$$\mathbb{W}=\psi(\mathbb{P})=\psi_{\#}\mathbb{P},$$

a probability measure on $\big(\mathcal{C}(\mathbb{R}_+,\mathbb{R}),\mathfrak{C}\big)$.

The whole point of this construction is that, under $\mathbb{W}$, the canonical process $(X_t,\ t\geq 0)$, now living directly on $\big(\mathcal{C}(\mathbb{R}_+,\mathbb{R}),\mathfrak{C}\big)$, is itself a Brownian motion. Indeed, for any $0\leq t_0<t_1<\cdots<t_n$ and Borel sets $A_0,\dots,A_n\in\mathcal{B}(\mathbb{R})$,

$$\begin{aligned}
&\mathbb{W}\big(X_{t_0}\in A_0,\ X_{t_1}\in A_1,\dots,X_{t_n}\in A_n\big)\\
&=\mathbb{P}\Big(\psi^{-1}\big(\{\omega(t_0)\in A_0\}\cap\cdots\cap\{\omega(t_n)\in A_n\}\big)\Big)\\
&=\mathbb{P}\big(B_{t_0}\in A_0,\ B_{t_1}\in A_1,\dots,B_{t_n}\in A_n\big) \qquad (*)\\
&=\mathbf{1}_{A_0}(0)\int_{A_1\times\cdots\times A_n}\frac{dx_1\cdots dx_n}{(2\pi)^{n/2}\sqrt{t_1(t_2-t_1)\cdots(t_n-t_{n-1})}}\ \exp\Big(-\frac{1}{2}\sum_{i=1}^{n}\frac{(x_i-x_{i-1})^2}{t_i-t_{i-1}}\Big),
\end{aligned}$$

with the convention $x_0\equiv 0$, this last expression being simply the joint Gaussian density of the increments of B. Crucially, the resulting measure $\mathbb{W}$ does not depend on the particular Brownian motion B used to construct it: it is entirely characterised by the finite-dimensional identity $(*)$, as we now explain.

A probability measure on $\big(\mathcal{C}(\mathbb{R}_+,\mathbb{R}),\mathfrak{C}\big)$ is always characterised by its finite-dimensional distributions, in the following precise sense.

Theorem 21. *Let $\mathbb{P}$ and $\mathbb{Q}$ be two probability measures on $\big(\mathcal{C}(\mathbb{R}_+,\mathbb{R}),\mathfrak{C}\big)$, that are the laws of two continuous processes $Z\sim\mathbb{P}$ and $\widetilde{Z}\sim\mathbb{Q}$. If*

$$(Z_{t_1},\dots,Z_{t_n})\overset{(d)}{=}(\widetilde{Z}_{t_1},\dots,\widetilde{Z}_{t_n})\qquad \textit{for every choice of } t_1,\dots,t_n,$$

then $\mathbb{P}=\mathbb{Q}$.

[1] Indeed, with $(X_t)_{t\geq0}$ denoting the canonical mappings on $\mathcal{C}(\mathbb{R}_+,\mathbb{R})$ we have that $\psi(\omega)=(B_s(\omega))_{s\geq0}$ and also that $X_t((s\mapsto B_s(\omega)))=B_t(\omega)$. Therefore $\psi(\omega)=(X_t(s\mapsto B_s(\omega)))_{t\geq0}=(B_t(\omega))_{t\geq0}$

Proof. We use the classical fact that two probability measures μ, ν on a measurable space $(E, \mathcal{E})$ which agree on a π-system $\mathcal{P}$ (i.e. a family of sets stable under finite intersections) generating $\mathcal{E}$ must coincide on all of $\mathcal{E}$. Take

$$\mathcal{P} = \Big\{ \big\{ \omega \in \mathcal{C}(\mathbb{R}_+, \mathbb{R}) : \ (X_{t_1}(\omega), \ldots, X_{t_n}(\omega)) \in A \big\} : n \geq 1, \ t_1, \ldots, t_n \geq 0, \ A \in \mathcal{B}(\mathbb{R}^n) \Big\}.$$

This family is stable under finite intersections (intersecting two such cylinder events, possibly enlarging the list of times involved, again yields a set of the same form), and it generates $\mathfrak{C}$ by definition of the latter (this iw explained in details in the exercise session). Since, by hypothesis,

$$\begin{aligned} &\mathbb{P}\big(Z \in \{\omega : (X_{t_1}(\omega), \ldots, X_{t_n}(\omega)) \in A\}\big) = \mathbb{P}\big((Z_{t_1}, \ldots, Z_{t_n}) \in A\big) \\ &= \mathbb{P}\big((\widetilde{Z}_{t_1}, \ldots, \widetilde{Z}_{t_n}) \in A\big) = \mathbb{Q}\big(\widetilde{Z} \in \{\omega : \cdots \in A\}\big) \end{aligned}$$

for every such cylinder set, $\mathbb{P}$ and $\mathbb{Q}$ agree on $\mathcal{P}$, and therefore coincide on all of $\mathfrak{C}$. □

Proof of Kolmogorov's continuity criterion

We first record a useful preliminary observation, essentially already established above: if $\widetilde{X}$ is a continuous modification of X, then almost surely $X_t = \widetilde{X}_t$ for every t in the countable set $I \cap \mathbb{Q}$, and hence, by continuity, almost surely for every $t \in I$; consequently X and $\widetilde{X}$ are indistinguishable. In particular, any two continuous modifications of X are automatically indistinguishable, which is what justifies speaking of the continuous version of X.

The proof of the theorem splits in three steps.

Step 1. We work first on $I = [0, 1]$ (the general bounded case being identical up to rescaling), and fix once and for all an exponent $\alpha \in \big(0, \frac{\varepsilon}{p}\big)$, where $p \geq 1$ will be chosen later. For a function $x : [0, 1] \to E$, define the (possibly infinite) quantity

$$K_\alpha(x) = \sup_{n \geq 0} \max_{i=1,\ldots,2^n} 2^{n\alpha} \, d\big(x_{(i-1)2^{-n}}, \ x_{i2^{-n}}\big),$$

which measures the worst-case α-Hölder ratio of x across dyadic subdivisions. Applying our moment hypothesis to each dyadic pair $\big((i-1)2^{-n}, \ i2^{-n}\big)$ and summing,

$$\mathbb{E}\big[K_\alpha(X)^p\big] \leq \sum_{n \geq 0} \sum_{i=1}^{2^n} 2^{n\alpha p} \, \mathbb{E}\big[d(X_{(i-1)2^{-n}}, X_{i2^{-n}})^p\big]$$

$$\le C \sum_{n\ge 0} \sum_{i=1}^{2^n} 2^{n\alpha p}\, 2^{-n(1+\varepsilon)}$$
$$= C \sum_{n\ge 0} 2^{n(\alpha p-\varepsilon)},$$

and this last geometric series converges precisely when $\alpha p < \varepsilon$, a condition we impose from now on (which is compatible with $\alpha < \varepsilon/q$ once we take $p \ge q$). We record: $\mathbb{E}[K_\alpha(X)^p] < \infty$.

Step 2. We now show that control of the dyadic increments, encoded by K_α, in fact controls the Hölder ratio over all pairs of dyadic points, not just adjacent ones. Let D denote the set of dyadic numbers in $[0,1]$, i.e. $t \in D$ iff $t = k2^{-n}$ for integers $k, n \ge 0$ (equivalently, t admits a finite binary expansion). We claim that, for every $x : [0,1] \to E$,

$$\sigma_\alpha(x) = \sup_{\substack{s,t\in D\\ s\neq t}} \frac{d(x_s, x_t)}{|t-s|^\alpha} \;\le\; \frac{2^{1+\alpha}}{1-2^{-\alpha}}\, K_\alpha(x). \tag{$*$}$$

Proof of $(*)$. Fix $s, t \in D$ with $0 \le s < t$, and let n be the unique integer such that $2^{-(n+1)} < t - s \le 2^{-n}$. Set $k = \lfloor 2^n s \rfloor$; since $s, t \in D$, we may write their binary expansions relative to the scale 2^{-n} as

$$s = \frac{k}{2^n} + \frac{\varepsilon_{n+1}}{2^{n+1}} + \cdots + \frac{\varepsilon_{n+\ell}}{2^{n+\ell}}, \qquad t = \frac{k}{2^n} + \frac{\widetilde{\varepsilon}_n}{2^n} + \frac{\widetilde{\varepsilon}_{n+1}}{2^{n+1}} + \cdots + \frac{\widetilde{\varepsilon}_{n+m}}{2^{n+m}},$$

with all $\varepsilon_i, \widetilde{\varepsilon}_i \in \{0,1\}$. Introduce the two finite chains of intermediate dyadic points

$$s_i = \frac{k}{2^n} + \frac{\varepsilon_{n+1}}{2^{n+1}} + \cdots + \frac{\varepsilon_{n+i}}{2^{n+i}}\ (0 \le i \le \ell), \qquad t_j = \frac{k}{2^n} + \frac{\widetilde{\varepsilon}_n}{2^n} + \cdots + \frac{\widetilde{\varepsilon}_{n+j}}{2^{n+j}}\ (0 \le j \le m),$$

so that $s_0 = k/2^n$, $s_\ell = s$, $t_0 \in \{k/2^n, (k+1)/2^n\}$, and $t_m = t$. By repeated use of the triangle inequality,

$$d(x_s, x_t) = d(x_{s_\ell}, x_{t_m}) \;\le\; \sum_{i=1}^{\ell} d(x_{s_i}, x_{s_{i-1}}) \;+\; d(x_{s_0}, x_{t_0}) \;+\; \sum_{j=1}^{m} d(x_{t_{j-1}}, x_{t_j}).$$

Each term on the right is now controlled directly from the definition of K_α: since s_0 and t_0 are two consecutive multiples of 2^{-n},

$$d(x_{s_0}, x_{t_0}) \le \max_{i=1,\dots,2^n} d\big(x_{(i-1)2^{-n}}, x_{i2^{-n}}\big) \le K_\alpha(x)\, 2^{-n\alpha};$$

similarly, since s_i and s_{i-1} are consecutive multiples of $2^{-(n+i)}$,

$$d(x_{s_{i-1}}, x_{s_i}) \leq K_\alpha(x)\, 2^{-(n+i)\alpha}, \qquad \text{and likewise} \qquad d(x_{t_{j-1}}, x_{t_j}) \leq K_\alpha(x)\, 2^{-(n+j)\alpha}.$$

Summing all these estimates,

$$d(x_s, x_t) \ \leq \ K_\alpha(x) \Big[2^{-n\alpha} + \sum_{i=1}^{\ell} 2^{-(n+i)\alpha} + \sum_{j=1}^{m} 2^{-(n+j)\alpha} \Big] \ \leq \ K_\alpha(x)\, 2^{-n\alpha} \cdot \frac{2}{1 - 2^{-\alpha}},$$

the last inequality bounding both partial sums, together with the leading term, by the full geometric series $2 \sum_{i \geq 0} 2^{-i\alpha}$. Finally, since $2^{-(n+1)} < t - s \leq 2^{-n}$ gives $2^{-n} \leq 2|t-s|$, hence $2^{-n\alpha} \leq 2^{\alpha} |t-s|^{\alpha}$, we conclude

$$\frac{d(x_s, x_t)}{|t-s|^{\alpha}} \ \leq \ K_\alpha(x)\, \frac{2^{1+\alpha}}{1 - 2^{-\alpha}},$$

which is exactly $(*)$, since s, t were arbitrary in D. $\square$

Step 3. Combining Steps 1 and 2, $\mathbb{E}\big[\sigma_\alpha(X)^p\big] < \infty$, so that the event $A = \{\omega : \sigma_\alpha(X(\omega)) < \infty\}$ has probability 1. For every $\omega \in A$, the map $t \mapsto X_t(\omega)$ is, by $(*)$, Hölder continuous — hence in particular uniformly continuous — on the dense set $D \subset [0,1]$; since (E, d) is complete, it therefore admits a unique continuous extension to all of $[0,1]$, with the same Hölder constant. Concretely, define

$$\widetilde{X}_t(\omega) = \lim_{\substack{s \to t \\ s \in D}} X_s(\omega) \quad \text{if } \omega \in A, \qquad \widetilde{X}_t(\omega) = x_0 \quad \text{otherwise},$$

for some arbitrarily fixed $x_0 \in E$. By continuity and density of D, one checks readily that

$$C_\alpha = \sup_{\substack{s,t \in [0,1] \\ s \neq t}} \frac{d(\widetilde{X}_s, \widetilde{X}_t)}{|t-s|^{\alpha}} \ = \ \sup_{\substack{s,t \in D \\ s \neq t}} \frac{d(\widetilde{X}_s, \widetilde{X}_t)}{|t-s|^{\alpha}} \ = \ \sup_{\substack{s,t \in D \\ s \neq t}} \frac{d(X_s, X_t)}{|t-s|^{\alpha}} \ = \ \sigma_\alpha(X) \qquad \text{on } A$$

(and $C_\alpha = 0$ on A^c), so that $\widetilde{X}$ is α-Hölder continuous everywhere on $[0,1]$, with $\mathbb{E}[C_\alpha^p] = \mathbb{E}[\sigma_\alpha(X)^p \mathbf{1}_A] < \infty$.

It remains only to check that $\widetilde{X}$ is indeed a modification of X. Fix $t \in [0,1]$; by dominated convergence, using the moment hypothesis once more,

$$\mathbb{E}\big[d(\widetilde{X}_t, X_t)^p \wedge 1\big] = \lim_{\substack{s \to t \\ s \in D}} \mathbb{E}\big[d(X_s, X_t)^p \wedge 1\big] \leq \lim_{\substack{s \to t \\ s \in D}} C\, |t-s|^{1+\varepsilon} = 0,$$

so that $X_s \to \widetilde{X}_t$ in probability as $s \to t$ along D. Since also $X_s \to \widetilde{X}_t$ almost surely (by definition of $\widetilde{X}_t$, for $\omega \in A$) and since limits in probability are almost surely unique, we conclude $\widetilde{X}_t = X_t$ almost surely, for every fixed $t \in [0,1]$. This proves that $\widetilde{X}$ is a modification of X, and completes the proof of Kolmogorov's continuity criterion.