1 Wiener Measure

In order for a mathematician to take A. Einstein’s 1905 article [12] seriously, he should feel obliged to begin by doing what N. Wiener did in his 1923 article [45]. Namely, he should convince himself that there is one, and only one, Borel, probability measure $\mu_{\mathbb{R}}$ on the Frechét space $\mathcal{W}(\mathbb{R}) \equiv \{ w \in C([0, \infty); \mathbb{R}) : w(0) = 0 \}$

with the properties that, for each $t \in [0, \infty)$ and $\tau > 0$, the increment $w \in \mathcal{W}(\mathbb{R}) \mapsto w(t+\tau) - w(t) \in \mathbb{R}$ under $\mu_{\mathbb{R}}$ is a centered Gaussian random variable with variance $\tau$ which is independent of the $\sigma$-algebra $B_t \equiv \sigma(\{ w(s) : s \in [0, t] \})$.\(^2\) Equivalently, for each $n \geq 1$, $0 = t_0 < t_1 < \cdots < t_n$, $\Gamma_1, \ldots, \Gamma_n \in B_{\mathbb{R}}$:\(^3\)

\begin{equation}
\mu_{\mathbb{R}}(\{ w : w(t_1) - w(t_0) \in \Gamma_1, \ldots, w(t_n) - w(t_{n-1}) \in \Gamma_n \}) = \prod_{m=1}^{n} \int_{\Gamma_m} g_{\tau_m}(\eta_m) \, d\eta_m,
\end{equation}

where

\begin{equation}
g_{\tau}(\eta) \equiv \frac{1}{\sqrt{2\pi \tau}} \exp \left( -\frac{y^2}{2\tau} \right), \quad (\tau, y) \in (0, \infty) \times \mathbb{R}
\end{equation}

and $\tau_m = t_m - t_{m-1}$. Actually, even though it requires some effort to produce one such $\mu_{\mathbb{R}}$, it is essentially trivial to check that there is at most one. Indeed, (1) obviously determines $\mu_{\mathbb{R}}$ on $\sigma(\{ w(s) : s \geq 0 \})$, and so the only thing that has to be checked (cf. §3.3 in [35]) is that $B_{\mathcal{W}(\mathbb{R})} = \sigma(\{ w(s) : s \geq 0 \})$.

The reason why the mathematically inclined reader of Einstein should want $\mu_{\mathbb{R}}$ to exist is that, according to Einstein, $\mu_{\mathbb{R}}$-typical paths are the trajectories of Brownian particles.\(^4\) Moreover, even if one ignores its physical implications, the existence of $\mu_{\mathbb{R}}$ raises a question of fundamental mathematical interest: can one put non-trivial, countably additive measures on an infinite dimensional spaces?

---

1 Here we are thinking of the Frechét metric which corresponds to uniform convergence on compact intervals.

2 When $\mathcal{F}$ is a family of functions on some space $\Omega$, we will use $\sigma(\mathcal{F})$ to denote the smallest $\sigma$-algebra over $\Omega$ with respect to which every element of $\mathcal{F}$ is measurable.

3 Given a separable metric space $E$, we use $B_E$ to denote the Borel field over $E$.

4 This term is used rather loosely to describe particles whose motion displays the sort of chaotic behavior observed by the nineteenth century botanist R. Brown, who was looking at pollen under a microscope. Einstein and (more or less simultaneously) M. Smoluchowski accounted for Brown’s observations as a manifestation of the kinetic theory of gases. Thus, unless $\mu_{\mathbb{R}}$ exists, Einstein and Smoluchowski’s whole picture would fall under the shadow of considerable mathematical doubt, a circumstance which might not have particularly disturbed Einstein but would probably have caused Smoluchowski severe distress.
1.1. Deconstructing Brownian Paths. The construction of Wiener measure which we will give in the next subsection is not any one of the three suggested by Wiener in [45]. Instead, it is derived from the ideas of P. Lévy and is based on the intuition explained here.

Assume that $\mu_\mathbb{R}$ exists. It is then clear that there are lots and lots of mutually independent, centered, Gaussian random variables floating around. Indeed, for any $n \geq 1$ and $0 \leq t_0 < \cdots < t_n$, the increments $\{w(t_m) - w(t_{m-1}) : 1 \leq m \leq n\}$ will be mutually independent, centered, Gaussian random variables.

Since one understands how to deal with and construct mutually independent random variables, it is reasonable to seek a clever way to choose a countable set of increments from which the whole path $w$ can be re-constructed via an elementary, deterministic procedure. The point is that, once such a scheme has been found, it should be possible to construct $\mu_\mathbb{R}$. Namely, one will use standard measure theoretic methods to construct an ample supply of mutually independent, centered Gaussian random variables, which one will then plug into the “re-construction” procedure.

Everything which follows relies heavily on the peculiar properties possessed by linear families of centered, Gaussian random variable. Given a probability space $(\Omega, F, P)$, we will say that $\mathfrak{G} \subseteq L^2(P; \mathbb{R})$ is a (centered) Gaussian family if it is a linear subspace each of whose elements is a centered, Gaussian random variable. The essential facts for us are summarized in the following lemma.

**3 Lemma.** Let $(\Omega, F, P)$ be a probability space and $\mathfrak{G}$ be a Gaussian family in $L^2(P; \mathbb{R})$. Then the closure $\bar{\mathfrak{G}}$ of $\mathfrak{G}$ in $L^2(P; \mathbb{R})$ is also a Gaussian family. Moreover, if $K$ is a non-empty subset of $\mathfrak{G}$, then $\sigma(K)$ is independent of $\sigma(K^\perp \cap \mathfrak{G})$. Finally, if $\mathfrak{H}$ is a second Gaussian family in $L^2(P; \mathbb{R})$ and if $\sigma(\mathfrak{G})$ is independent of $\sigma(\mathfrak{H})$, then $\mathfrak{G} + \mathfrak{H}$ is again a Gaussian family.

**Proof:** All of these statements are applications of the fact that $Y \in L^2(P; \mathbb{R})$ is a centered Gaussian random variable if and only if

$$E^P \left[ e^{\sqrt{-1} \xi Y} \right] = \exp \left( -\frac{\xi^2}{2} E^P [Y^2] \right), \quad \xi \in \mathbb{R}. $$

Thus, if $\{Y_n\}_1^\infty \subseteq \mathfrak{G}$ converges in $L^2(P; \mathbb{R})$ to $Y$, then

$$E^P \left[ e^{\sqrt{-1} \xi Y} \right] = \lim_{n \to \infty} E^P \left[ e^{\sqrt{-1} \xi Y_n} \right] = \lim_{n \to \infty} \exp \left( -\frac{\xi^2}{2} E^P [Y_n^2] \right) = \exp \left( -\frac{\xi^2}{2} E^P [Y^2] \right),$$

which means that $Y$ is also a centered, Gaussian random variable. Similarly, to prove the third statement, all that one needs to show is that, for $X \in \mathfrak{G}$

---

5 Below, and throughout, I have adopted the usual probabilistic convention of identifying integrals as expectation values and therefore using $E^P[F]$ to denote $\int F \ dP = \int_{\Omega} F(\omega) \ P(\mathrm{d}\omega)$ when doing integration with respect to a generic probability measure $P$. 

and $Y \in \mathcal{G}$, $X + Y$ is a centered, Gaussian random variable. But, because $X$ and $Y$ are independent and centered,

$$
\mathbb{E}^p \left[ e^{\sqrt{-1} \xi (X+Y)} \right] = \mathbb{E}^p \left[ e^{\sqrt{-1} \xi X} \right] \mathbb{E}^p \left[ e^{\sqrt{-1} \xi Y} \right] = \exp \left( -\frac{\xi^2}{2} \left( \mathbb{E}^p [X^2] + \mathbb{E}^p [Y^2] \right) \right) = \exp \left( -\frac{\xi^2}{2} \mathbb{E}^p [(X + Y)^2] \right).
$$

Finally, to prove the second statement, it suffices to show that for any finite collections $\{X_1, \ldots, X_m\} \subseteq K$ and $Y_1, \ldots, Y_n \subseteq K^\perp \cap \mathcal{G},$

$$
\mathbb{E}^p \left[ \exp \left( \sqrt{-1} \sum_{i=1}^m \xi_i X_i \right) \exp \left( \sqrt{-1} \sum_{j=1}^n \eta_j Y_j \right) \right] = \mathbb{E}^p \left[ \exp \left( \sqrt{-1} \sum_{i=1}^m \xi_i X_i \right) \right] \mathbb{E}^p \left[ \exp \left( \sqrt{-1} \sum_{j=1}^n \eta_j Y_j \right) \right]
$$

for all $(\xi_1, \ldots, \xi_m) \in \mathbb{R}^m$ and $(\eta_1, \ldots, \eta_n) \in \mathbb{R}^n$. In the present situation, this is tantamount to checking that

$$
\mathbb{E}^p \left[ \exp \left( \sqrt{-1} \sum_{i=1}^m \xi_i X_i \right) \exp \left( \sqrt{-1} \sum_{j=1}^n \eta_j Y_j \right) \right] = \exp \left( -\frac{1}{2} \left( \mathbb{E}^p \left[ \left( \sum_{i=1}^m \xi_i X_i \right)^2 \right] + \mathbb{E}^p \left[ \left( \sum_{j=1}^n \eta_j Y_j \right)^2 \right] \right) \right),
$$

which, in turn, follows from the fact that $\sum_{i=1}^m \xi_i X_i$ is orthogonal to $\sum_{j=1}^n \eta_j Y_j$.

The immediate relevance to us of Lemma 3 is that the increments

$$
\{w(t) - w(s) : 0 \leq s < t\}
$$

form a Gaussian family in $L^2(\mu_R; \mathbb{R})$. Armed with this observation, we return to the problem at hand. As a first guess, one might try looking at the increments

$$
\Delta_{m,0}w \equiv w(m+1) - w(m), \quad m \in \mathbb{N},
$$

and using linear interpolation to construct a path $w_0$. That is, take $w_0(t) = w(m) + (t-m)\Delta_{m,0}w$ for $m \geq 0$ and $t \in [m, m+1]$. Of course, $w = w_0$ at

---

6 The reader who wants a quick review of the basic facts about independent random variables might want to take a look at §1.1 in [35]. In particular, Exercise 1.1.13 there is relevant here.
integer times. However, if \( t \) is not an integer, then \( \mu \equiv (w(t) = w_0(t)) = 0 \). For example, because \( w_0((2\ell + 1)2^{-1}) = \frac{1}{2}(w(\ell + 1) - w(\ell)) \),

\[
D_{\ell,1} w \equiv w((2\ell + 1)2^{-1}) - w_0((2\ell + 1)2^{-1}) = \frac{\Delta_{2\ell,1} w - \Delta_{2\ell+1,1} w}{2},
\]

where

\[
\Delta_{m,1} w \equiv w((m + 1)2^{-1}) - w(m2^{-1}).
\]

Since the \( \Delta_{m,1} w \)'s are mutually independent, centered Gaussian random variables with variance \( \frac{1}{2} \), \( D_{\ell,1} w \) is a centered Gaussian random variable with variance \( \frac{1}{4} \). Moreover, each \( D_{\ell,1} w \) is orthogonal to all of the \( \Delta_{m,0} w \)'s, and therefore, by the second part of Lemma 3, the \( \sigma \)-algebras

\[
\sigma \left( \{ D_{\ell,1} w : \ell \in \mathbb{N} \} \right) \quad \text{and} \quad \sigma \left( \{ \Delta_{m,0} w : m \in \mathbb{N} \} \right)
\]

are independent. Thus, there simply is not enough randomness in the \( \Delta_{m,0} w \)'s, and we will have to throw in the \( D_{\ell,1} w \)'s if we are going to have any chance of reconstructing \( w \).

With the preceding in mind, for each \( n \in \mathbb{N} \), we introduce the increments

\[
\Delta_{m,n} w \equiv w((m + 1)2^{-n}) - w(m2^{-n}), \quad m \in \mathbb{N},
\]

construct the polygonal path \( w_n \) so that

\[
w_n(t) = w(m2^{-n}) + (2^n t - m)\Delta_{m,n} w \quad \text{for} \quad m \in \mathbb{N} \quad \text{and} \quad t \in [m2^{-n}, (m + 1)2^{-n}],
\]

take

\[
D_{\ell,n} w \equiv w((2\ell + 1)2^{-n}) - w_n((2\ell + 1)2^{-n}) = \frac{\Delta_{2\ell,n} w - \Delta_{2\ell+1,n} w}{2},
\]

when \( n \geq 1 \), and define the random variables

\[
Y_{m2^{-n}} = \begin{cases} 
\Delta_{m,0} w & \text{if } n = 0 \\
2^{\frac{n+2}{4}} D_{\ell,n} w & \text{if } n \geq 1 \text{ and } m = 2\ell + 1.
\end{cases}
\]

Using Lemma 3 as we did above, one finds that, under \( \mu \),

\[
\{ Y_{m2^{-n}} : (m, n) \in \mathbb{N}^2 \}
\]

constitutes a family of mutually independent, centered Gaussian random variables with variance 1. Moreover, one can easily reconstruct \( w \) from the \( Y_{m2^{-n}} \)'s. Namely, \( w_0(0) = 0 \)

\[
w_0(t) = w_0(m) + (t - m)Y_m w \quad \text{for} \quad m \in \mathbb{N} \quad \text{and} \quad t \in [m, m + 1],
\]

\[
w_n(t) = w_{n-1}(t) + 2^{-\frac{n+1}{2}} (1 - |2^n t - (2m + 1)|) Y_{(2m+1)2^{-n}} w
\]

for \( n \geq 1, m \in \mathbb{N} \), and \( t \in [m2^{-n+1}, (m + 1)2^{-n+1}] \),

and \( w_n \rightarrow w \) uniformly on compacts.
1.2. Lévy’s Construction. It should be now clear how to proceed. First, one constructs a probability space \((\Omega, \mathcal{F}, \mathbb{P})\) which can support a countably infinite family of mutually independent random variables. One should then label these random as \(Y_{m2^{-n}}\) for \((m, n) \in \mathbb{N}^2\). Next, use (5) as the template for constructing polygonal paths \(w_n\):

\[
w_0(0) = 0 \quad \text{and} \quad w_0(t) = w_0(m) + (t - m)Y_m \quad \text{for} \quad m \in \mathbb{N} \& t \in [m, m + 1],
\]

\[
w_n(t) = w_{n-1}(t) + 2^{-\frac{n-1}{2}}(1 - |2^n t - (2m + 1)|)Y_{(2m+1)2^{-n}}
\]

for \(n \geq 1,\ m \in \mathbb{N},\ \text{and} \ t \in [m2^{-n+1}, (m + 1)2^{-n+1}]\).

Notice that \(w_n((2m + 1)2^{-n} \pm 2^{-n}) - w_n((2m + 1)2^{-n})\) equals

\[-2^{-\frac{n+1}{2}}Y_{(2m+1)2^{-n}} \pm \frac{w_{n-1}((m + 1)2^{-n}) - w_{n-1}(m2^{-n+1})}{2}.\]

In particular, with the help of Lemma 3, one can use induction on \(n \in \mathbb{N}\) to check that, for each \(n\),

\[
\{w_n((m + 1)2^{-n}) - w_n(m2^{-n}) : m \in \mathbb{N}\}
\]

is a sequence of mutually independent, centered Gaussian random variables with variance \(2^{-n}\). Thus, if we can show that, as \(n \to \infty\), the sequence \(\{w_n\}_0^\infty\) is \(\mathbb{P}\)-almost surely convergent uniformly on compacts to a limit path \(w\), then we can simply take \(\mu_{\mathbb{P}}\) to be the distribution of a random variable of \(w\) under \(\mathbb{P}\). For this reason, consider \(w_n - w_{n-1}\) and observe that

\[
|w_n(t) - w_{n-1}(t)| \leq 2^{-\frac{n+1}{2}}Y_{(2m+1)2^{-n}} \quad \text{for} \quad t \in [m2^{-n}, (m + 1)2^{-m}].
\]

Hence, for any fixed \(T \in (0, \infty)\),

\[
\sup_{\tau \in [0, T]} |w_n(\tau) - w_{n-1}(\tau)| \leq 2^{-\frac{n+1}{2}} \max_{m \leq 2^{-n+1}T} |Y_{(2m+1)2^{-n}}| \leq 2^{-\frac{n+1}{2}} \left( \sum_{m \leq 2^{-n+1}T} |Y_{(2m+1)2^{-n}}|^4 \right)^{\frac{1}{4}},
\]

and so, by Jensen’s inequality,

\[
\mathbb{E}\mathbb{P}\left[ \sup_{\tau \in [0, T]} |w_n(\tau) - w_{n-1}(\tau)| \right] \leq 2^{-\frac{n+1}{2}} \left( \sum_{m \leq 2^{-n+1}T} \mathbb{E}\mathbb{P}[Y_{(2m+1)2^{-n}}^4] \right)^{\frac{1}{4}} \leq (3(1 + T))^{\frac{1}{4}} 2^{-\frac{n}{4}},
\]

For example, one can (cf. Theorem 1.1.10 in [35]) take \(\Omega = [0, 1)\), \(\mathcal{F} = B_{[0, 1]}\), and \(\mathbb{P}\) to be the restriction of Lebesgue measure to \([0, 1)\). Alternatively, one can use a product construction (cf. Exercise 1.1.14 in [35]).

Given a random variable \(\Xi\) on some probability space \((\Omega, \mathcal{F}, \mathbb{P})\) with values in some measurable space \((E, \mathcal{B})\), probabilists call the pushforward measure \(\Xi_*\mathbb{P} = \mathbb{P}\circ \Xi^{-1}\), given by \(\mathbb{P}\circ \Xi^{-1}(\Gamma) = \mathbb{P}(\Xi \in \Gamma)\), the distribution of \(\Xi\) under \(\mathbb{P}\). In the present setting, we are thinking of \(w\) as being a \(\mathbb{P}(\mathbb{R})\)-valued random variable on \(\Omega\).
which is more than enough to show that
\[
\sum_{n=1}^{\infty} \mathbb{E}^P \left[ \sup_{t \in [0,T]} |w_n(t) - w_{n-1}(t)| \right] < \infty.
\]

Hence, standard measure theoretic machinery now applies and guarantees that, \( \mathbb{P} \)-almost surely, \( \{w_n\} \) is Cauchy convergent in \( \mathfrak{M}(\mathbb{R}) \), and that is exactly what we wanted to show. In other words, if we use \( w \) to denote this limit, then as we said before, we can take \( \mu_w \) to be the distribution of \( w \) under \( \mathbb{P} \).

1.3. Modulus of Continuity. By sharpening the preceding argument a little, one can estimate the modulus of continuity of Brownian paths. To see this, first observe for any \( p \in (1, \infty) \) and \( q \in (1, \infty) \) with \( q \geq \frac{2}{q'} \),

\[
\mathbb{E}^P \left[ \sup_{\tau \in [0,T]} |w(\tau) - w_{n-1}(\tau)|^p \right] \leq C_q (1 + T)^{q' - \frac{2}{q'}}.
\]

where
\[
C_q := \left( \int_{\mathbb{R}} \xi^{2q} g_1(\xi) \, d\xi \right)^{\frac{1}{q'}}
\]

and \( q' = \frac{2}{p-1} \) is the Hölder conjugate of \( q \). Indeed, the reasoning is the same as we used above when \( p = 1 \) and \( q = 2 \). Hence,

\[
(*) \quad \mathbb{E}^P \left[ \sup_{\tau \in [0,T]} |w(\tau) - w_{n-1}(\tau)|^p \right] \leq C_q (1 + T)^{q' - \frac{2}{q'}}.
\]

where \( C_q' = C_q \left( 1 - \frac{2}{q'} \right)^{-1} \).

Next, let \( 0 \leq s < t \leq T \) be given, assume that \( 2^{-n-2} \leq t - s \leq 2^{-n-1} \) for some \( n \in \mathbb{N} \), and choose \( m \in \mathbb{N} \) so that \( m \leq 2^n s < 2^n t \leq m + 1 \). Then (cf. the notation in \( \S 1.1.1 \), and remember that \( w(m2^{-n}) = w_n(m2^{-n}) \) for all \( m \in \mathbb{N} \)),

\[
|w(t) - w(s)| \leq |w_n(t) - w_n(s)| + 2 \sup_{\tau \in [0,T]} |w(\tau) - w_n(\tau)| = 2^n(t-s)|\Delta_{m,n} w| + 2 \sup_{\tau \in [0,T]} |w(\tau) - w_n(\tau)| \leq \frac{1}{2} |\Delta_{m,n} w| + 2 \sup_{\tau \in [0,T]} |w(\tau) - w_n(\tau)|.
\]

and so, for any \( \alpha > 0 \),

\[
\sup_{0 \leq s < t \leq T} \frac{|w(t) - w(s)|}{(t-s)^\alpha} \leq 4^\alpha \sum_{n=0}^{\infty} 2^n \max_{m \leq 2^n (1+T)} |\Delta_{m,n} w| + 2 \sup_{\tau \in [0,T]} |w(\tau) - w_n(\tau)|.
\]

In the hope of avoiding confusion which might arise from an attempt to parse the notation which follows, I point out that I often use \( \mathbb{E}^P[\cdots]^\frac{1}{p} \) when I mean \( (\mathbb{E}^P[\cdots])^{\frac{1}{p}} \).
Now assume that \( \alpha \in (0, \frac{1}{2}) \), and let \( p \in [1, \infty) \) be given. Then,

\[
\mathbb{E}^p \left[ \left( \sup_{0 \leq s < t \leq T} \frac{|w(t) - w(s)|}{(t-s)^\alpha} \right)^p \right]^{1/p} \\
\leq 4^{\alpha - \frac{1}{2}} \sum_{n=0}^{\infty} 2^{\alpha n} \mathbb{E}^p \left[ \max_{m \leq 2^n T} |\Delta_{m,n} w|^p \right]^{1/p} \\
+ 4^{\alpha + \frac{1}{2}} \sum_{n=0}^{\infty} 2^{\alpha n} \mathbb{E}^p \left[ \sup_{\tau \in [0,T]} |w(\tau) - w_n(\tau)|^p \right]^{1/p}.
\]

Finally, choose \( q \in (p, \infty) \) so that \( \frac{1}{q} > 2\alpha \), and observe that

\[
\mathbb{E}^p \left[ \max_{1 \leq m \leq 2^n T} |\Delta_{m,n} w|^p \right]^{1/p} \leq \mathbb{E}^p \left[ \left( \sum_{m=0}^{2^n T} |\Delta_{m,n} w|^{2q} \right)^{\frac{1}{2q}} \right]^{1/p} \\
\leq C_q 2^{-\frac{2}{2q}} (1 + 2^n T)^{\frac{1}{2q}} \leq C_q (1 + T)^{\frac{1}{2q}} 2^{\frac{2}{2q}}.
\]

Hence, in conjunction with the estimate obtained in (*), we now see that, for each \( \alpha \in (0, \frac{1}{2}) \) and \( p \in [1, \infty) \), there is a \( C_p(\alpha) < \infty \) such that

\[
(6) \quad \mathbb{E}^{\mu_R} \left[ \left( \sup_{0 \leq s < t \leq T} \frac{|w(t) - w(s)|}{(t-s)^\alpha} \right)^p \right]^{1/p} \leq C_p(\alpha) (1 + T).
\]

In §1.2, we will see that almost no Brownian path is H"older continuous of any order \( \alpha > \frac{1}{2} \). Finding the exact modulus of continuity of a Brownian path requires more effort. This effort was made by P. Lévy, who showed that, \( \mu_R \)-almost surely,

\[
\lim_{\delta \searrow 0} \sup_{0 \leq t - s \leq \delta} \frac{|w(t) - w(s)|}{\omega(\delta)} = 1, \quad \text{where} \quad \omega(\delta) = \sqrt{2\delta \log \delta^{-1}}.
\]