1
Preliminaries

This is a collection of more or less elementary topics you will need to be familiar with. Much of it you may know already but look it over anyhow. Munroe [1953] tells most of what you need – and nothing more, which I like.

1.1 Lebesgue measure: an outline

Let's begin with the unit interval $[0, 1]$ and the class $\mathfrak{F}'$ of subsets $A, B, \ldots \subset [0, 1]$ comprised of countable sums of non-overlapping intervals I: open, closed, half and half, I don't care. The measure of $A \in \mathfrak{F}'$ is declared to be the sum of their lengths: $\mathrm{m}(A) = \sum |I|$. Obviously, A can be written as such a sum of intervals in many ways, but I take it as geometrically obvious that $\mathrm{m}(A)$ is not changed thereby. It is plain that

(1) $\mathrm{m}(A) \le \mathrm{m}(B)$ if $A \subset B$ and also
(2) $\mathrm{m}(A \cup B) \le \mathrm{m}(A) + \mathrm{m}(B)$, or more precisely
(3) $\mathrm{m}(A \cap B) + \mathrm{m}(A \cup B) = \mathrm{m}(A) + \mathrm{m}(B)$.

1.1.1 Carathéodory's lemma

Lemma 1.1.1 *If the sets* $A_n \in \mathfrak{F}'$ *decrease to the empty set, then* $\mathrm{m}(A_n) \downarrow 0$.

Proof. If not, then $\mathrm{m}(A_n) \ge c > 0$ for any $n \ge 1$. Choose compact $B_n =$ a finite sum of closed intervals inside A_n with $\mathrm{m}(B_n) > \mathrm{m}(A_n) - c\,2^{-n-1}$, and let $C_n = B_1 \cap \cdots \cap B_n$. These are compact, decreasing, and $\bigcap_1^\infty C_n \subset$

$\bigcap_1^\infty A_n = \varnothing$ so C_n is void from some n on, by Heine–Borel. But this cannot be. In fact,

$$\begin{aligned}\mathrm{m}(C_n) &= \mathrm{m}(C_{n-1} \cap B_n) \\ &= \mathrm{m}(C_{n-1}) + \mathrm{m}(B_n) - \mathrm{m}(C_{n-1} \cup B_n) && \text{by (3)} \\ &> \mathrm{m}(C_{n-1}) + \mathrm{m}(A_n) - c\,2^{-n-1} - \mathrm{m}(A_{n-1})\end{aligned}$$

since $C_{n-1} \cup B_n \subset A_{n-1}$. Now move $\mathrm{m}(C_{n-1})$ to the left and add from $n = 2$ on to produce

$$\begin{aligned}\mathrm{m}(C_n) - \mathrm{m}(C_1) &> \mathrm{m}(A_n) - \mathrm{m}(A_1) - \frac{c}{4} \\ &> c - \left(\mathrm{m}(B_1) + \frac{c}{4}\right) - \frac{c}{4} \\ &= \frac{c}{2} - \mathrm{m}(B_1)\end{aligned}$$

for any $n \geq 2$. This is contradictory since $\mathrm{m}(B_1) = \mathrm{m}(C_1)$ and $\mathrm{m}(C_n) = 0$.

1.1.2 Measurable sets

Any open set $G \subset [0, 1]$ is the sum of a countable number of disjoint open intervals, so $\mathrm{m}(G)$ is defined; the recipe $\mathrm{m}(A) = \inf \mathrm{m}(G)$ for open $G \supset A$ extends on from $A \in \mathfrak{F}'$ to any $A \subset [0, 1]$; and you have $\mathrm{m}(A \cup B) \leq \mathrm{m}(A) + \mathrm{m}(B)$ in general. Declare A to be "measurable" if $\mathrm{m}(A)$ so defined coincides with $\sup \mathrm{m}(K)$ for compact $K \subset A$ and write $A \in \mathfrak{F}''$ in that case. Obviously, $\mathfrak{F}' \subset \mathfrak{F}''$; in particular, any open G belongs to $\mathfrak{F}''$. I claim that any compact K does, too. It is easy to make open G_n decrease to K. Then $G_n \setminus K$ is open and decreases to the empty set, so $\mathrm{m}(G_n \setminus K) \downarrow 0$ by Carathéodory's lemma, and $\mathrm{m}(K) \leq \mathrm{m}(G_n) \leq \mathrm{m}(G_n \setminus K) + \mathrm{m}(K)$ as $n \uparrow \infty$, i.e. $\mathrm{m}(K) = \inf_{G \supset K} \mathrm{m}(G)$, as required. Note, by the way, that G_n may be chosen as the sum of finitely many non-overlapping open intervals, by Heine–Borel, so that $1 - \mathrm{m}(K) = \lim_{n\uparrow\infty}[1 - \mathrm{m}(G_n)] = \lim_{n\uparrow\infty} \mathrm{m}(G_n')$. Here, the compact G_n' is contained in the open complement K', and $K' \setminus G_n' \in \mathfrak{F}'$ decreases to the empty set, with the result that $\mathrm{m}(K') = \mathrm{m}(K' \setminus G_n') + \mathrm{m}(G_n')$ is very nearly $1 - \mathrm{m}(G_n)$, i.e. $\mathrm{m}(K') = 1 - \mathrm{m}(K)$ or, equivalently, $\mathrm{m}(G') = 1 - \mathrm{m}(G)$.

The rest is easy enough.

(1) $\mathfrak{F}''$ is closed under complements, generalizing the fact that $G' = K \in \mathfrak{F}''$: indeed, if $A \in \mathfrak{F}''$, then

$$\inf_{G \supset A'} \mathrm{m}(G) = \inf_{K' \supset A'} \mathrm{m}(K') = 1 - \sup_{K \subset A} \mathrm{m}(K)$$

and likewise

$$\sup_{K\subset A'} \mathrm{m}(K) = \sup_{G'\subset A'} \mathrm{m}(G') = 1 - \inf_{G\supset A} \mathrm{m}(G),$$

and as these match, so $A' \in \mathfrak{F}''$; in particular, $\mathrm{m}(A') = 1 - \mathrm{m}(A)$.

(2) If $A_n \in \mathfrak{F}''$ and $A_i \cap A_j = \varnothing$ for $i \neq j$, then $A = \bigcup_1^\infty A_n \in \mathfrak{F}''$. Here's the proof. Choose $c > 0$ and $G_n \supset A_n \supset K_n$ so that $\mathrm{m}(G_n) - c\,2^{-n} < \mathrm{m}(A_n) < \mathrm{m}(K_n) + c\,2^{-n}$. Then $G = \bigcup_1^\infty G_n \supset A$, and you have

$$\begin{aligned}
\mathrm{m}(A) &\le \mathrm{m}(G) \le \mathrm{m}\big(G \setminus \bigcup_1^n G_j\big) + \mathrm{m}\big(\bigcup_1^n G_j\big) \\
&< \varepsilon + \sum_1^n \mathrm{m}(G_j) \quad \text{for small } \varepsilon \text{ and large } n \text{ by Carathéodory's lemma} \\
&< \varepsilon + \sum_1^n \mathrm{m}(A_j) + c \\
&< \varepsilon + \sum_1^n \mathrm{m}(K_j) + 2c \\
&< \varepsilon + \mathrm{m}\left(\bigcup_1^n K_j\right) + 2c \\
&< \varepsilon + \mathrm{m}(A) + 2c,
\end{aligned}$$

line 5 being justified by the fact that Ks do not abut, being compact and disjoint by inclusion in the As – in short, ε being small and c arbitrary, you have

$$\mathrm{m}(A) \le \sup_{K\subset A} \mathrm{m}(K) \ \text{ by line 5} \quad \text{and} \quad \mathrm{m}(A) \ge \inf_{G\supset A} \mathrm{m}(G) \ \text{ by line 6,}$$

which is to say $A \in \mathfrak{F}''$, and you have also a bonus from the third line:

(3) $\mathrm{m}(A) = \sum_1^\infty \mathrm{m}(A_n)$.

Discussion. In the parlance adopted here, (1) and (2) state $\mathfrak{F}''$ is a "field", unlike $\mathfrak{F}'$, meaning that it is closed under complements and both countable unions and countable intersections. (3) states that m is "countably additive" on $\mathfrak{F}''$; equivalently, if $A_1 \supset A_2 \supset \cdots$ and $\bigcap_1^\infty A_n = A$ then $\mathrm{m}(A_n) \downarrow \mathrm{m}(A)$, generalizing Carathéodory's lemma from $\mathfrak{F}'$ to $\mathfrak{F}''$. Below, I dispense with $\mathfrak{F}''$ and write $\mathfrak{F}$, plain, in place of $\mathfrak{F}''$. The extension of all this from $[0,1]$ to the whole line $\mathbb{R}$ will be obvious, with one proviso: $\mathrm{m}(A_n) \downarrow \mathrm{m}(A)$ as above if $\mathrm{m}(A_1) < \infty$, but not, for example, if $A_n = [n, \infty)$.

1.1.3 The integral

The function $f\colon [0,1] \to \mathbb{R}$ is "measurable" if $(x : a \le f(x) \le b)$ belongs to $\mathfrak{F}$ for every $a < b$, this measurability being assumed for all functions mentioned below. Then if $f \ge 0$, the sums

$$I_n = \sum_1^\infty (k-1)2^{-n}\,\mathrm{m}(0 \le x \le 1 : (k-1)2^{-n} \le f(x) < k2^{-n})$$

increase with n, and you declare the integral $I = \int_0^1 f(x)\,\mathrm{d}x$ to be $\lim_{n\uparrow\infty} I_n \le \infty$; f is "summable" if $I < +\infty$. The integral for $f = f^+ - f^-$ of indefinite signature is $\int f = \int f^+ - \int f^-$ provided either $\int f^+ < \infty$ or $\int f^- < \infty$.

The integral for $f \ge 0$ has a pleasing geometrical interpretation. The 1-dimensional measure $\mathrm{m} = \mathrm{m}_1$ on $\mathbb{R}$ may be imitated by the 2-dimensional measure m_2 on $\mathbb{R}^2$. The plan is the same, only now you start with rectangles $I \times J$ with measure (area) $\mathrm{m}_2(I \times J) = |I| \times |J|$. Then the figure

$$F_n = \bigcup_1^\infty (0 \le x \le 1 : (k-1)2^{-n} \le f(x) < k2^{-n}) \times [0, (k-1)2^{-n}) \subset \mathbb{R}^2$$

increases to the region "under the graph of f", meaning the 2-dimensional figure

$$F = [(x,y) : 0 \le x \le 1 \ \text{ and } \ 0 \le y < f(x)],$$

and the sum $I_n = \mathrm{m}_2(F_n)$ increases to $I = \mathrm{m}_2(F) = \int_0^1 f(x)\,\mathrm{d}x$. From this point of view, Lebesgue's theorem of monotone convergence is obvious: if $0 \le f_n \uparrow f$, then $\int f_n \uparrow \int f$, which is nothing but the fact that $F(f_n) \uparrow F(f)$ implies $\mathrm{m}_2\, F(f_n) \uparrow \mathrm{m}_2\, F(f)$. The rest of Lebesgue's theorems follow: if $\lim f_n(x) = f(x)$ except perhaps on a set of measure 0, then $\lim \int f_n = \int f$ provided either:

(1) $f_n \downarrow f$ and $\int f_1 < \infty$; or
(2) f_n is dominated by a summable function h, i.e. $|f_n| \le h$ and $\int h < \infty$; or, more simply,
(3) $|f_n|$ is bounded, independently of n.

This stuff extends to $\mathbb{R}$, but once more look out: (2) is fine, but (3) may be insufficient because $\mathrm{m}(\mathbb{R}) = +\infty$. A useful variant is Fatou's lemma: if $f_n \ge 0$, then $\int \liminf f_n \le \liminf \int f_n$ without exception.

Exercise 1.1.2 Prove it.

The story ends with Fubini's theorem: if $0 \le f(x, y)$ is measurable on $\mathbb{R}^2$, then

$$\iint f \, \mathrm{d}x \, \mathrm{d}y = \int f \, \mathrm{dm}_2 = \int \mathrm{d}x \left[\int f(x,y) \, \mathrm{d}y \right] = \int \mathrm{d}y \left[\int f(x,y) \, \mathrm{d}x \right]$$

without exception: either all these integrals are finite and they agree or else they are all $+\infty$, with a self-evident extension to signed functions $f(x, y)$ if either f^+ or f^- is summable.

Exercise 1.1.3 Try to prove it if you want.
Hint: $\mathrm{m}_2(I \times J) = |I| \times |J|$ is the simplest instance. Take it from there.

1.2 Probabilities and expectations

The mathematical set-up imitates Lebesgue measure on the unit interval – only the language changes. There are three parts to it:

(1) A *sample space* $\mathbf{X}$ of 1, 2, 3, or even countably many isolated points $\mathbf{x}$, representing, e.g. the possible outcomes of some experiment; or as it may be $[0, 1]$ itself; or any (even ∞-dimensional) geometrical figure you want with a nice (countable) topology on it.

(2) A *field* $\mathfrak{F}$ of *events* A, B, C, etc. $\subset \mathbf{X}$ which is closed under complements $A' = \mathbf{X} \backslash A$, (countable) unions $A_1 \cup A_2 \cup$etc., and so also under (countable) intersections $B_1 \cap B_2 \cap$etc., and containing every open $G \subset \mathbf{X}$ and so also every compact K.

(3) *Probabilities* $\mathbb{P} \colon \mathfrak{F} \to [0, 1]$, i.e. a non-negative, countably additive measure on $\mathfrak{F}$ of total mass $\mathbb{P}(\mathbf{X}) = 1$, such that

$$\mathbb{P}(A) = \inf_{G \supset A} \mathbb{P}(G) = \sup_{K \subset A} \mathbb{P}(K) \quad \text{for any } A \in \mathfrak{F}.$$

1.2.1 Reduction to Lebesgue measure

This whole scheme looks very general but is not. The context can be quite elaborate, but if that be ignored, then (in principle) you know all about $(\mathbf{X}, \mathfrak{F}, \mathbb{P})$ already. I explain, following Halmos–von Neumann [1942:335–336]. The present account is simplified. To begin with, there may be "lumps" of mass, i.e. "indivisible" events A with $\mathbb{P}(A) > 0$ such that $\mathbb{P}(B) = \mathbb{P}(A)$ or 0 for any $B \subset A$. Obviously, these are countable in number, and if they do not exhaust $\mathbf{X}$, put them aside and look at $\mathbb{P}$ on the rest of $\mathbf{X}$, renormalized to make $\mathbb{P}(\mathbf{X}) = 1$ again. Now suppose, what is natural, that $\mathbf{X}$ has a fixed countable family of open sets $G_n : n \ge 1$

such that each point $\mathbf{x} \in \mathbf{X}$ is the intersection of a sub-collection of these. Then there is a isomorphism between $\mathbf{X}$ and $[0,1]$, mapping $\mathbb{P}$ to the Lebesgue measure m. This means what it has to mean, as will appear presently.

Discussion. Let e_n be the indicator of G_n and map

$$\mathbf{x} \in \mathbf{X} \to \mathbf{y} = \mathbf{e}_1/2 + \mathbf{e}_2/4 + \mathbf{e}_3/8 + \cdots \to \mathbf{z} = F(\mathbf{y}),$$

in which $\mathbf{e}_n = e_n(\mathbf{x})$ and $F(y) = \mathbb{P}(\mathbf{y} \le y)$ for $0 \le y \le 1$. The first map $\mathbf{x} \to \mathbf{y}$ of $\mathbf{X}$ into $[0,1]$ is $1:1$ – in fact, any ambiguity such as $\mathbf{y} = \frac{1}{4} + \frac{1}{8} + \cdots = \frac{1}{2}$ cannot appear since $\bigcap_{n\ge1} G_n$ is then void. (Why?). Besides, F is jump-free: any jump would come from a single point $\mathbf{x}$ and $\mathbb{P}(\mathbf{x}) = 0$ since no lumps are left. Now the second map $\mathbf{x} \to \mathbf{z}$ of $\mathbf{X}$ into $[0,1]$ is not $1:1$ if F has flat places, but who cares? The totality of these come from a null set inside $\mathbf{X}$ and this may be removed. Then the map inverse to $\mathbf{x} \to \mathbf{z}$ induces a faithful pairing, not of points, but of events, placing the event $(\mathbf{z} \in A)$ in correspondence with $B = (\mathbf{x} : \mathbf{y} \in F^{-1}(A))$ via the inverse function $F^{-1}(z) = \max\{y : F(y) \le z\}$, modulo null sets fore and aft. Now

$$\mathbb{P}(\mathbf{z} \le z) = \mathbb{P}(\mathbf{y} \le F^{-1}(z)) = F(F^{-1}(z)) = z,$$

so that $\mathbf{z}$ is distributed by the Lebesgue measure m and $\mathbb{P}(B) = \mathrm{m}(A)$. In this way, $\mathbf{X}$ and $\mathbb{P}$ may be identified with $[0,1]$ and m – in short, the Lebesgue space is all you need to think about from a merely technical point of view.

But not so fast. The technical aspect is one thing, the context is another, and context is what it's all about, really, so this is *not* the way to compute anything of practical interest.

1.2.2 Expectation

A function $f : \mathbf{X} \to \mathbb{R}$, measurable over $\mathfrak{F}$, i.e. with $(\mathbf{x} : a \le f(\mathbf{x}) < b) \in \mathfrak{F}$ for any $a < b$ is now a *random variable*, so-called, and the integral $\mathbb{E}(f) = \int_X f \, d\,\mathbb{P}$ is its mean value or expectation, as in

$$\mathbb{E}(f) = \lim_{n\uparrow\infty} \sum_{k=1}^{\infty} \frac{k}{2^n}\,\mathbb{P}\left(\mathbf{x} : \frac{k-1}{2^n} \le f(\mathbf{x}) < \frac{k}{2^n}\right) \le \infty \quad \text{if } f \ge 0.$$

If $f = f^+ - f^-$ is capable of both signatures, then its expectation $\mathbb{E}(f) = \mathbb{E}(f^+) - \mathbb{E}(f^-)$ is defined only if either $\mathbb{E}(f^+) < \infty$ or $\mathbb{E}(f^-) < \infty$. Obviously, the whole of §1.1 now applies unchanged in view of the reduction just explained.

1.2.3 Independence

This notion is peculiar to probability. Its importance for all our business cannot be over-emphasized. Events A and B are *independent* if $\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B)$. Then A and B' are also independent, as per

$$\mathbb{P}(A \cap B') = \mathbb{P}(A) - \mathbb{P}(A \cap B) = \mathbb{P}(A)\big(1 - \mathbb{P}(B)\big) = \mathbb{P}(A)\,\mathbb{P}(B'),$$

and likewise A' and B, and A' and B' – in other words, any event from the little field $\mathfrak{A} = [\varnothing, A, A', \mathbf{X}]$ is independent of any event from $\mathfrak{B} = [\varnothing, B, B', \mathbf{X}]$. For the independence of three events A, B, C, you ask that $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\,\mathbb{P}(B)\,\mathbb{P}(C)$ and likewise for any triple from the associated little fields. More generally, two fields $\mathfrak{A}$ and $\mathfrak{B}$ are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\,\mathbb{P}(B)$ for any $A \in \mathfrak{A}$ and $B \in \mathfrak{B}$, and similarly for the independence of three fields or more. Likewise for random variables f: associate to f the smallest field containing all the events $(\mathbf{x} : a \le f(\mathbf{x}) < b)$. Then two or more variables are independent if their fields are such. This leads to an important principle: namely, $\mathbb{E}(f_1 f_2) = \mathbb{E}(f_1) \times \mathbb{E}(f_2)$ if f_1 and f_2 are independent, assuming the expectations make sense.

Exercise 1.2.1 Check this rule from scratch when the functions are non-negative.

Caution! In general, the independence of three events taken two at a time is *not* the same as their full independence as per the next exercise.

Exercise 1.2.2 Take independent variables $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ with values $=$ 0 or 1 and probabilities $\mathbb{P}(\mathbf{e} = 1) = \mathbb{P}(\mathbf{e} = 0) = \frac{1}{2}$. Check that the events $A = (\mathbf{e}_1 + \mathbf{e}_2 \text{ even})$, $B = (\mathbf{e}_2 + \mathbf{e}_3 \text{ even})$, $C = (\mathbf{e}_3 + \mathbf{e}_1 \text{ odd})$ are independent taken two at a time but that $0 = \mathbb{P}(A \cap B \cap C) \neq \mathbb{P}(A)\,\mathbb{P}(B)\,\mathbb{P}(C) = \frac{1}{8}$.

Example 1.2.3 $\mathbf{X}$ is comprised of the 2^n strings $\mathbf{x} = (\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n)$ of 0s or 1s representing n tosses of a coin (1 for heads, 0 for tails). $\mathfrak{F}$ is the class of all subsets of $\mathbf{X}$, and the $\mathbf{e}$s have the common distribution $\mathbb{P}(\mathbf{e} = 1) = p$ and $\mathbb{P}(\mathbf{e} = 0) = 1 - p = q$ with $0 < p < 1$. The coin is honest if $p = 1/2$. Here, it is natural to suppose that the individual $\mathbf{e}$s are statistically independent so that probabilities multiply as per $\mathbb{P}(\mathbf{x}) = p^{\#} q^{n-\#}$, $\#$ being the number of successes $(\mathbf{e} = 1)$ in the string $\mathbf{x} = (\mathbf{e}_1, \ldots, \mathbf{e}_n)$.

Then $\mathbb{P}(\# = k) = \binom{n}{k} p^k q^{n-k}$, in which $\binom{n}{k} = n!/k!(n-k)!$ – pronounced "n choose k" – is the number of ways of selecting $0 \le k \le n$ objects out of n like objects. These are the Bernoulli trials, so-called, to which Chapter 2 is devoted, with special attention to their behavior for $n \uparrow \infty$.

Exercise 1.2.4 If the interpretation "n choose k" is not quite clear in your mind, think it over. Justify $\binom{2n}{n} = \sum_{k=0}^{n} \binom{n}{k}^2$ as a check that you've understood. Don't compute. Just think what it says.

Exercise 1.2.5 Compute $\mathbb{E}(\#) = np$ and $\mathbb{E}(\# - np)^2 = npq$ using the binomial theorem.

Example 1.2.6 $\mathbf{X}$ is now the unit interval $0 \le \mathbf{x} \le 1$, $\mathfrak{F}$ is the familiar field including all subintervals, $\mathbb{P}$ is the standard Lebesgue (or should I say Borel) measure defined on it. You may expand $2\mathbf{x}-1$ as $\mathbf{e}_1/2+\mathbf{e}_2/4+\mathbf{e}_3/8+\cdots$. The present $\mathbf{e}$ imitates the old $2\mathbf{e}-1$ of Example 1.2.3 for $p = 1/2$. They are independent, with common distribution $\mathbb{P}(\mathbf{e} = \pm 1) = 1/2$. Now replace $\mathbf{x} \in [0,1]$ by the equivalent *sample path* $\mathbf{x}\colon n \ge 0 \to \mathbb{Z}$ defined by $\mathbf{x}(0) = 0$ and $\mathbf{x}(n) = \mathbf{e}_1 + \cdots + \mathbf{e}_n$, $n \ge 1$, stepping right or left with equal probabilities at times $n = 1, 2, 3, \ldots$. This is the standard random walk RW(1). It occupies Chapter 3.

Example 1.2.7 $\mathbf{X}$ is now $\mathrm{C}[0,\infty)$ = the space of (continuous) sample paths $\mathbf{x}\colon t \in [0,\infty) \to \mathbf{x}(t) \in \mathbb{R}$ with $\mathbf{x}(0) = 0$, say; $\mathfrak{F}$ is the smallest field containing all events

$$Z = \left[\mathbf{x} : \bigcap_{i=1}^{n} (a_i \le \mathbf{x}(t_i) < b_i)\right]$$

for any $n \ge 1$, any $0 < t_1 < \cdots < t_n$, and any as and bs; finally,

$$\mathbb{P}(Z) = \int_{\bigcap_{i=1}^{n}(a_i \le x_i \le b_i)} \frac{e^{-x_1^2/2t_1}}{\sqrt{2\pi t_1}} \frac{e^{-(x_2-x_1)^2/2(t_2-t_1)}}{\sqrt{2\pi(t_2-t_1)}} \cdots \frac{e^{-(x_n-x_{n-1})^2/2(t_n-t_{n-1})}}{\sqrt{2\pi(t_n-t_{n-1})}} \, d^n x.$$

This set-up is the standard Brownian motion BM(1) to be introduced in Chapter 6, not so much for itself, though it is endlessly fascinating, but as an aid to proving limit theorems. BM(1) is what comes out when RW(1) is speeded up (jump time $1/n$) and scaled back (jump size $1/\sqrt{n}$) and n is taken to $+\infty$, as will be proved in §6.4. By the way, it is not obvious that the probabilities displayed above can be extended to the whole of $\mathfrak{F}$ in a countably additive way: the continuity of the sample path $\mathbf{x}$ is descriptive of an *uncountable* number of observations, and $\mathfrak{F}$, with its countable additivity only, does not permit you to speak about such things. Wiener [1923] adapted Carathéodory's lemma to overcome this technical difficulty; see §6.2 for this and for P. Lévy's [1948] beautiful, more economical method.

Exercise 1.2.8 Check that, for any $c > 0$, $c\mathbf{x}(t/c^2)$ is a copy of BM(1), and likewise $t\mathbf{x}(1/t)$. This means that BM(1) has lots of internal "symmetries", of which much more later on.

1.3 Conditional probabilities and expectations

The naïve version of conditional probabilities is the familiar $\mathbb{P}(A|B) = \mathbb{P}(A\cap B)/\,\mathbb{P}(B)$, signifying the probability of A when B is known to have occurred. I need a more subtle and flexible version of this idea.

1.3.1 Signed measures

Let Q be the difference $Q^+ - Q^-$ of two non-negative, countably additive measures $Q^\pm$ on a field $\mathfrak{F}$ of subsets A, B, etc. of a space $\mathbf{X}$ and suppose $Q^\pm(\mathbf{X}) < \infty$. Then it is possible to refine the splitting so that Q^+ and Q^- live on disjoint parts of $\mathbf{X}$. This is obvious if $Q(A) = \int_A f \,\mathrm{d}\,\mathbb{P}$ for some summable function f on $\mathbf{X}$ and some probabilities $\mathbb{P}$ on $\mathfrak{F}$: just take $Q^+(A) = \int_A f^+ \,\mathrm{d}\,\mathbb{P}$ with the positive part f^+ of f. More generally, define

$$Q^+(A) = \sup_{B\subset A} Q(B) \quad \text{and} \quad Q^-(A) = -\inf_{B\subset A} Q(B).$$

Claim 1 *Q^+ and Q^- are countably additive.*

Proof. Take $A_i \cap A_j = \varnothing$ if $i \neq j$ and $B \subset A = \bigcup_1^\infty A_n$ so that $Q(B) > Q^+(A) - \varepsilon$, and so also

$$\sum_{n=1}^{\infty} Q^+(A_n) \geq \sum_{n=1}^{\infty} Q(B \cap A_n) = Q(B) > Q^+(A) - \varepsilon.$$

Similarly, if $B_n \subset A_n$ is taken so that $Q(B_n) > Q^+(A_n) - \varepsilon\, 2^{-n}$, then with $B = \bigcup_1^\infty B_n$, you find

$$\sum_{n=1}^{\infty} Q^+(A_n) - \varepsilon < \sum_{n=1}^{\infty} Q(B_n) = Q(B) \leq Q^+(A),$$

and so on. The same applies to Q^-.

Claim 2 $Q^+ - Q^- = Q$.

Proof. You have

$$\begin{aligned} Q(A) + Q^-(A) &= Q(A) - \inf_{B \subset A} Q(A) \\ &= \sup_{B \subset A} [Q(A) - Q(B)] \\ &= \sup_{B' \subset A} Q(B') \qquad \text{with } B' = A \setminus B \text{ for the moment} \\ &= Q^+(A). \end{aligned}$$

Claim 3 *Q^+ and Q^- live on disjoint parts of $\mathbf{X}$.*

Proof. $Q^-(B') + Q^+(B) = Q^-(\mathbf{X}) + Q(B)$ by Claim 2, so $Q^-(B'_n) + Q^+(B_n) < 2^{-n}$ by choice of B_n in view of $Q^-(\mathbf{X}) = -\inf Q(B)$. Then with

$$C = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} B_k,$$

you have

$$\begin{aligned} Q^-(C') + Q^+(C) &= Q^-\left(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} B'_k \right) + Q^+\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} B_k \right) \\ &\le Q^-\left(\bigcup_{k=n}^{\infty} B'_k \right) + Q^+\left(\bigcup_{k=n}^{\infty} B_k \right) \qquad \text{for any } n \ge 1 \\ &\le \sum_{k=n}^{\infty} [Q^-(B'_k) + Q^+(B_k)] \\ &< 2^{-n} \downarrow 0 \quad \text{as } n \uparrow \infty, \end{aligned}$$

i.e. Q^- lives on C and Q^+ on its complement C'.

1.3.2 Radon–Nikodym

Let Q and P be non-negative, countably additive set functions on $\mathfrak{F}$, of finite total mass, and let $Q(A)$ vanish whenever $P(A) = 0$. Then there is a non-negative function f, measurable over the field $\mathfrak{F}$, such that $Q(A) = \int_A f \, dP$ for any $A \in \mathfrak{F}$. This function is the so-called *Radon–Nikodym derivative* of Q with respect to P: symbolically, $f = dQ/dP$.

Proof. Fix $n \ge 1$ and let $k \ge 1$ vary. $Q - k\, 2^{-n} P$ is negative on some set A_k and positive on its complement A'_k, as per the decomposition