We have not yet shown the necessity for σ-fields. Restrict attention to $([0,1],\mathcal{F},\mathbf{m})$ where $\mathcal{F}$ is either (i) $\mathcal{B}$, the Borel σ-algebra or (ii) $\overline{\mathcal{B}}$ the possibly larger σ-algebra of Lebesgue measurable sets (as defined by Caratheodary). This consists of two distinct issues.

(1) Showing that $\overline{\mathcal{B}}$ (hence $\mathcal{B}$) does not contain all subsets of $[0,1]$.
(2) Showing that it is not possible at all to define a p.m. $\mathbf{P}$ on the σ-field of all subsets so that $\mathbf{P}[a,b] = b-a$ for all $0 \le a \le b \le 1$. In other words, one cannot consistently extend $\mathbf{m}$ from $\overline{\mathcal{B}}$ (on which it is uniquely determined by the condition $\mathbf{m}[a,b] = b-a$) to a p.m. $\mathbf{P}$ on the σ-algebra of all subsets.

(1) **$\overline{\mathcal{B}}$ does not contain all subsets of $[0,1]$:** We shall need the following 'translation invariance property' of $\mathbf{m}$ on $\overline{\mathcal{B}}$.

Exercise 20. For any $A \subset [0,1]$ and any $x \in [0,1]$, $\mathbf{m}(A+x) = \mathbf{m}(A)$, where $A+x := \{y+x(\text{mod } 1) \,:\, y \in A\}$ (eg: $[0.4,0.9]+0.2 = [0,0.1]\cup[0.6,1]$). Show that for any $A \in \overline{\mathcal{B}}$ and $x \in [0,1]$ that $A+x \in \overline{\mathcal{B}}$ and that $\mathbf{m}(A+x) = \mathbf{m}(A)$.

Now we construct a subset $A \subset [0,1]$ and countably (infinitely) many $x_k \in [0,1]$ such that the sets $A+x_k$ are pairwise disjoint and $\cup_k(A+x_k)$ is the whole of $[0,1]$. Then, if A were in $\overline{\mathcal{B}}$, by the exercise $A+x_k$ would have the same probability as A. But $\sum \mathbf{m}(A+x_k)$ must be equal to $\mathbf{m}[0,1] = 1$, which is impossible! Hence $A \notin \overline{\mathcal{B}}$.

How to construct such a set A and $\{x_k\}$? Define an equivalence relation on $[0,1]$ by $x \sim y$ if $x-y \in \mathbb{Q}$ (check that this is indeed an equivalence relation). Then, $[0,1]$ splits into pairwise disjoint equivalence classes whose union is the whole of $[0,1]$.

Invoke *axiom of choice* to get a set A that has exactly one point from each equivalence class. Consider $A+r$, $r \in \mathbb{Q}\cap[0,1)$. If $A+r$ and $A+s$ intersect then we get an $x \in [0,1]$ such that $x = y+r = z+s$ (mod 1) for some $y,z \in A$. This implies that $y-z = r-s$ (mod 1) and hence that $y \sim z$. So we must have $y = z$ (as A has only one element from each equivalence class) and that forces $r = s$ (why?). Thus $A+r$, $r \in \mathbb{Q}\cap[0,1)$ are pairwise disjoint. Further given $x \in [0,1]$, there is a $y \in A$ belonging to the $[[x]]$. Therefore $x \in A+r$ where $r = y-x$ or $y-x+1$. Thus we have constructed the set A whose countably many translates $A+r$, $r \in \mathbb{Q}\cap[0,1)$ are pairwise disjoint and exhaustive! This answers question (1).

Remark 21. There is a theorem to the effect that the axiom of choice is necessary to show the existence of a non-measurable set (as an aside, we should perhaps not have used the word 'construct' given that we invoke the axiom of choice).

(2) **m does not extend to all subsets:** The proof above shows in fact that $\mathbf{m}$ cannot be extended to a *translation invariant* p.m. on all subsets. If we do not require translation invariance for the extended measure, the question becomes more difficult.

Note that there do exist probability measures on the σ-algebra of all subsets of $[0,1]$, so one cannot say that there are no measures on all subsets. For example, define $\mathbf{Q}(A) = 1$ if $0.4 \in A$ and $\mathbf{Q}(A) = 0$ otherwise. Then $\mathbf{Q}$ is a p.m. on the space of all subsets of $[0,1]$. $\mathbf{Q}$ is a discrete p.m. in hiding! If we exclude such measures, then it is true that some subsets have to be omitted to define a p.m. You may find the proof for the following general theorem in Billingsley, p. 46 (uses *axiom of choice* and *continuum hypothesis*).

Fact 22. There is no p.m. on the σ-algebra of all subsets of $[0,1]$ that gives zero probability to singletons.

Say that x is an atom of $\mathbf{P}$ if $\mathbf{P}(\{x\}) > 0$ and that $\mathbf{P}$ is purely atomic if $\sum_{\text{atoms}} \mathbf{P}(\{x\}) = 1$. The above fact says that if $\mathbf{P}$ is defined on the σ-algebra of all subsets of $[0,1]$, then $\mathbf{P}$ must be have atoms. It is not hard to see that in fact $\mathbf{P}$ must be purely atomic. To see this let $\mathbf{Q}(A) = \mathbf{P}(A) - \sum_{x\in A} \mathbf{P}(\{x\})$. Then $\mathbf{Q}$ is a non-negative measure without atoms. If $\mathbf{Q}$ is not identically zero, then with $c = \mathbf{Q}([0,1])^{-1}$, we see that $c\mathbf{Q}$ is a p.m. without atoms, and defined on all subsets of $[0,1]$, contradicting the stated fact.

Remark 23. This last manipulation is often useful and shows that we can write any probability measure as a convex combination of a purely atomic p.m. and a completely nonatomic p.m.

(3) **Finitely additive measures** If we relax countable additivity, strange things happen. For example, there does exist a *translation invariant* ($\mu(A+x) = \mu(A)$ for all $A \subset [0,1]$, $x \in [0,1]$, in particular, $\mu(I) = |I|$) *finitely additive* ($\mu(A\cup B) = \mu(A)+\mu(B)$ for all A,B disjoint) p.m. defined on all subsets of $[0,1]$! In higher dimensions, even this fails, as shown by the mind-boggling

Banach-Tarski "paradox": The unit ball in $\mathbb{R}^3$ can be divided into finitely many (five, in fact) disjoint pieces and rearranged (only translating and rotating each piece) into a ball of twice the original radius!!

5. Random variables

Definition 24. Let $(\Omega_i, \mathcal{F}_i, \mathbf{P}_i)$, $i = 1,2$, be two probability spaces. A function $T : \Omega_1 \to \Omega_2$ is called an Ω_2-valued random variable if $T^{-1}A \in \mathcal{F}_1$ for any $A \in \mathcal{F}_2$. Here $T^{-1}(A) := \{\omega \in \Omega_1 \,:\, T(\omega) \in A\}$ for any $A \subset \Omega_2$.

Important cases are when $\Omega_2 = \mathbb{R}$ and $\mathcal{F}_2 = \mathcal{B}(\mathbb{R})$ (we just say "random variable") or $\Omega_2 = \mathbb{R}^d$ and $\mathcal{F}_2 = \mathcal{B}(\mathbb{R}^d)$ ("random vector"). When $\Omega_2 = C[0,1]$ with $\mathcal{F}_2$ its Borel sigma algebra (under the sup-norm metric), T is called a "stochastic process". When Ω_2 is itself the space of all locally finite countable subsets of $\mathbb{R}^d$ (with Borel sigma algebra in an appropriate metric) , we call T a "point process". In genetics or population biology one looks at genealogies, and then we have tree-valued random variables etc. etc.

Remark 25. Some remarks.

1. If $T : \Omega_1 \to \Omega_2$ is any function, then given a σ-algebra $\mathcal{G}$ on Ω_2, the "pull-back" $\{T^{-1}A \,:\, A \in \mathcal{G}\}$ is the smallest σ-algebra on Ω_1 w.r.t. which T is measurable (if we fix $\mathcal{G}$ on Ω_2) . Conversely, given a σ-algebra $\mathcal{F}$ on Ω_1, the "push-forward" $\{A \subset \Omega_2 \,:\, T^{-1}A \in \mathcal{F}\}$ is the largest σ-algebra on Ω_2 w.r.t. which T is measurable (if we fix $\mathcal{F}$ on Ω_1). These properties are simple consequences of the fact that $T^{-1}(A)^c = T^{-1}(A^c)$ and $T^{-1}(\cup A_n) = \cup_n T^{-1}(A_n)$.
2. If S generates $\mathcal{F}_2$, i.e., $\sigma(S) = \mathcal{F}_2$, then it suffices to check that $T^{-1}A \in \mathcal{F}_1$ for any $A \in S$.

Example 26. Consider $([0,1], \mathcal{B})$. Any continuous function $T : [0,1] \to \mathbb{R}$ is a random variable. This is because $T^{-1}(\text{open}) = \text{open}$ and open sets generate $\mathcal{B}(\mathbb{R})$. **Exercise:** Show that T is measurable if it is any of the following. (a) Lower semicontinuous, (b) Right continuous, (c) Non-decreasing, (d) Linear combination of measurable functions, (e) $\limsup$ of a countable sequence of measurable functions. (a) supremum of a countable family of measurable functions.

Push forward of a measure: If $T : \Omega_1 \to \Omega_2$ is a random variable, and $\mathbf{P}$ is a p.m. on $(\Omega_1, \mathcal{F}_1)$, then defining $\mathbf{Q}(A) = \mathbf{P}(T^{-1}A)$, we get a p.m $\mathbf{Q}$, on $(\Omega_2, \mathcal{F}_2)$. $\mathbf{Q}$, often denoted $\mathbf{P}T^{-1}$ is called the push-forward of $\mathbf{P}$ under T.

The reason why $\mathbf{Q}$ is a measure is that if A_n are pairwise disjoint, then $T^{-1}A_n$ are pairwise disjoint. However, note that if B_n are pairwise disjoint in Ω_1, then $T(B_n)$ are in general not disjoint. This is why there is no "pull-back measure" in general (unless T is one-one, in which case the pull-back is just the push-forward under T^{-1}!)

6. Borel Probability measures on Euclidean spaces

Given a Borel p.m. μ on $\mathbb{R}^d$, we define its cumulative distribution functions (CDF) to be $F_\mu(x_1, \ldots, x_d) = \mu((-\infty, x_1] \times \ldots \times (-\infty, x_d])$. Then, by basic properties of probability measures, $F_\mu : \mathbb{R}^d \to [0,1]$ (i) is non-decreasing in each co-ordinate, (ii) $F_\mu(x) \to 0$ if $\max_i x_i \to -\infty$, $F_\mu(x) \to 1$ if $\min_i x_i \to +\infty$, (iii) F_μ is right continuous in each co-ordinate.

Two natural questions. Given an $F : \mathbb{R}^d \to [0,1]$ satisfying (i)-(iii), is there necessarily a Borel p.m. with F as its CDF? If yes, is it unique?

If μ and ν both have CDF F, then for any rectangle $R = (a_1, b_1] \times \ldots \times (a_d, b_d]$, $\mu(R) = \nu(R)$ because they are both determined by F. Since these rectangles form a π-system that generate the Borel σ-algebra, $\mu = \nu$ on $\mathcal{B}$.

What about existence of a p.m. with CDF equal to F? For simplicity take $d = 1$. One boring way is to define $\mu(a,b] = F(b) - F(a)$ and then go through Caratheodary construction. But all the hard work has been done in construction of Lebesgue measure, so no need to repeat it!

Consider the probability space $((0,1), \mathcal{B}, m)$ and define the function $T : (0,1) \to \mathbb{R}$ by $T(u) := \inf\{x \,:\, F(x) \geq u\}$. When F is strictly increasing and continuous, T is just the inverse of F. In general, T is non-decreasing, left continuous. Most importantly, $T(u) \leq x$ if and only if $F(x) \geq u$. Let $\mu := m\, T^{-1}$ be the push-forward of the Lebesgue measure under T. Then,

$$\mu(-\infty, x] = m\{u \,:\, T(u) \leq x\} = m\{u \,:\, F(x) \geq u\} = m(0, F(x)] = F(x).$$

Thus, we have produced a p.m. μ with CDF equal to F. Thus p.m.s on the line are in bijective correspondence with functions satisfying (i)-(iii). Distribution functions (CDFs) are a useful but dispensable tool to study measures on the line, because we have better intuition in working with functions than with measures.

Exercise 27. Do the same for Borel probability measures on $\mathbb{R}^d$.