3. Distributions

Recall that a random variable on a probability space $(\Omega, \mathcal{F}, P)$ is a function $X : \Omega \to \mathbb{R}$ that is measurable with respect to the Borel sets.

Every random variable induces a probability measure μ on $\mathbb{R}$ (called its *distribution*) by

$$\mu(A) = P\left(X^{-1}(A)\right)$$

for all $A \in \mathcal{B}$.

To check that μ is a probability measure, note that since X is a function, if $A_1, A_2, ... \in \mathcal{B}$ are disjoint, then so are $\{X \in A_1\}, \{X \in A_2\}, ... \in \mathcal{F}$, hence

$$\mu\left(\bigcup_i A_i\right) = P\left(\{X \in \bigcup_i A_i\}\right) = P\left(\bigcup_i \{X \in A_i\}\right) = \sum_i P\left(\{X \in A_i\}\right) = \sum_i \mu(A_i).$$

The distribution of a random variable X is usually described in terms of its *distribution function*

$$F(x) = P(X \le x) = \mu\left((-\infty, x]\right).$$

In cases where confusion may arise, we will emphasize dependence on the random variable using subscripts - i.e. μ_X, F_X.

Theorem 3.1. *If F is the distribution function of a random variable X, then*

(i) F is nondecreasing
(ii) F is right-continuous (i.e. $\lim_{x \to a^+} F(x) = F(a)$ for all $a \in \mathbb{R}$)
(iii) $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$
(iv) If $F(x^-) = \lim_{y \to x^-} F(y)$, then $F(x^-) = P(X < x)$
(v) $P(X = x) = F(x) - F(x^-)$

Proof.

For (i), note that if $x \le y$, then $\{X \le x\} \subseteq \{X \le y\}$, so $F(x) = P(X \le x) \le P(X \le y) = F(y)$ by monotonicity.

For (ii), observe that if $x \searrow a$, then $\{X \le x\} \searrow \{X \le a\}$, and apply continuity from above.

For (iii), we have $\{X \le x\} \searrow \emptyset$ as $x \searrow -\infty$ and $\{X \le x\} \nearrow \mathbb{R}$ as $x \nearrow \infty$.

For (iv), $\{X \le y\} \nearrow \{X < x\}$ as $y \nearrow x$. (Note that the limit exists since F is monotone.)

For (v), $\{X = x\} = \{X \le x\} \setminus \{X < x\}$. □

In fact, the first three properties in Theorem 3.1 are sufficient to characterize a distribution function.

Theorem 3.2. *If $F : \mathbb{R} \to \mathbb{R}$ satisfies properties (i), (ii), and (iii) from Theorem 3.1, then it is the distribution function of some random variable.*

Proof. (Draw Picture)

Let $\Omega = (0, 1)$, $\mathcal{F} = \mathcal{B}_{(0,1)}$, $P =$ Lebesgue measure, and define $X : (0, 1) \to \mathbb{R}$ by

$$X(\omega) = F^{-1}(\omega) := \inf\{y \in \mathbb{R} : F(y) \ge \omega\}.$$

Note that properties (i) and (iii) ensure that X is well-defined.

To see that F is indeed the distribution function of X, it suffices to show that

$$\{\omega : X(\omega) \leq x\} = \{\omega : \omega \leq F(x)\}$$

for all $x \in \mathbb{R}$, as this implies

$$P(X \leq x) = P(\{\omega : X(\omega) \leq x\}) = P\left(\{\omega : \omega \leq F(x)\}\right) = F(x)$$

where the final equality uses the definition of Lebesgue measure and the fact that $F(x) \in [0, 1]$.

Now if $\omega \leq F(x)$, then $x \in \{y \in \mathbb{R} : F(y) \geq \omega\}$, so $X(\omega) = \inf\{y \in \mathbb{R} : F(y) \geq \omega\} \leq x$.
Consequently, $\{\omega : \omega \leq F(x)\} \subseteq \{\omega : X(\omega) \leq x\}$.

To establish the reverse inclusion, note that if $\omega > F(x)$, then properties (i) and (ii) imply that there is an $\varepsilon > 0$ such that $F(x) \leq F(x + \varepsilon) < \omega$.
Since F is nondecreasing, it follows that $x+\varepsilon$ is a lower bound for $\{y \in \mathbb{R} : F(y) \geq \omega\}$, hence $X(\omega) \geq x+\varepsilon > x$.
Therefore, $\{\omega : \omega \leq F(x)\}^C \subseteq \{\omega : X(\omega) \leq x\}^C$ and thus $\{\omega : X(\omega) \leq x\} \subseteq \{\omega : \omega \leq F(x)\}$. □

Theorem 3.2 shows that any function satisfying properties (i) - (iii) gives rise to a random variable X, and thus to a probability measure μ, the distribution of X. The following result shows that the measure is uniquely determined.

Theorem 3.3. *If F is function satisfying (i)-(iii) in Theorem 3.1, then there is a unique probability measure μ on $(\mathbb{R}, \mathcal{B})$ with $\mu\left((-\infty, x]\right) = F(x)$ for all $x \in \mathbb{R}$.*

Proof. Theorem 3.2 gives the existence of a random variable X with distribution function F. The measure it induces is the desired μ.
To establish uniqueness, suppose that μ and ν both have distribution function F. Define

$$\mathcal{P} = \{(-\infty, a] : a \in \mathbb{R}\}$$
$$\mathcal{L} = \{A \in \mathcal{B} : \mu(A) = \nu(A)\}.$$

Observe that for any $a \in \mathbb{R}$, $\mu\left((-\infty, a]\right) = F(a) = \nu\left((-\infty, a]\right)$, so $\mathcal{P} \subseteq \mathcal{L}$.
Also, for any $a, b \in \mathbb{R}$, $(-\infty, a] \cap (-\infty, b] = (-\infty, a \wedge b] \in \mathcal{P}$, hence $\mathcal{P}$ is a π-system.
Finally, $\mathcal{L}$ is a λ-system since

(1) $\mu(\Omega) = 1 = \nu(\Omega)$, so $\Omega \in \mathcal{L}$.
(2) For any $A, B \in \mathcal{L}$ with $A \subseteq B$, we have

$$\mu(B \setminus A) = \mu(B) - \mu(A) = \nu(B) - \nu(A) = \nu(B \setminus A)$$

(by countable additivity and the definition of $\mathcal{L}$), so $B \setminus A \in \mathcal{L}$.
(3) If $A_n \in \mathcal{L}$ with $A_n \nearrow A$, then

$$\mu(A) = \lim_{n\to\infty} \mu(A_n) = \lim_{n\to\infty} \nu(A_n) = \nu(A)$$

(by continuity from below and the definition of $\mathcal{L}$), so $A \in \mathcal{L}$.

Since the closed rays generate the Borel sets, the π-λ Theorem implies that $\mathcal{B} = \sigma(\mathcal{P}) \subseteq \mathcal{L}$ and thus $\mu(E) = \nu(E)$ for all $E \in \mathcal{B}$.

□

To summarize, every random variable induces a probability measure on $(\mathbb{R}, \mathcal{B})$, every probability measure defines a function satisfying properties (i)-(iii) in Theorem 3.1, and every such function uniquely determines a probability measure.

Consequently, it is equivalent to give the distribution or the distribution function of a random variable.

However, one should be aware that distributions/distribution functions do not determine random variables, even neglecting differences on null sets.

For example, if X is uniform on $[-1, 1]$ (so that $\mu_X = \frac{1}{2}m|_{[-1,1]}$), then $-X$ also has distribution μ_X, but $-X \neq X$ almost surely.

When two random variables X and Y have the same distribution function, we say that they are *equal in distribution* and write $X =_d Y$.

Note that random variables can be equal in distribution even if they are defined on different probability spaces.

Constructing Measures on $\mathbb{R}$. (Brief Review)

It is worth mentioning that we kind of cheated in Theorem 3.2 since we assumed the existence of Lebesgue measure. In fact, the standard derivation of Lebesgue measure in terms of Stieltjes measure functions implies the results in Theorems 3.2 and 3.3. Presumably everyone has seen this argument before and since it is fairly long, we will content ourselves with a brief outline.

Recall that an *algebra of sets* on S is a non-empty collection $\mathcal{A} \subseteq 2^S$ which is closed under complements and finite unions.

A *premeasure* μ_0 on $\mathcal{A}$ is a function $\mu_0 : \mathcal{A} \to [0, \infty]$ such that

1) $\mu_0(\emptyset) = 0$
2) If $A_1, A_2, ...$ is a sequence of disjoint sets in $\mathcal{A}$ whose union also belongs to $\mathcal{A}$, then $\mu_0\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu_0(A_i)$.

(If $\mu_0(S) < \infty$, then 2) implies that $\mu_0(\emptyset) = \mu_0(\emptyset) + \mu_0(\emptyset)$, which implies 1).)

An *outer measure* μ^* on S is a function $\mu^* : 2^S \to [0, \infty]$ such that

i) $\mu^*(\emptyset) = 0$
ii) $\mu^*(A) \le \mu^*(B)$ if $A \subseteq B$.
iii) $\mu^*\left(\bigcup_{i=1}^{\infty} A_i\right) \le \sum_{i=1}^{\infty} \mu^*(A_i)$,

and a set $A \subseteq S$ is said to be μ^**-measurable* if

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^C)$$

for all $E \subseteq S$.

It can be shown that if μ_0 is a premeasure on the algebra $\mathcal{A}$, then the set function defined by

$$\mu^*(E) = \inf\left\{\sum_{i=1}^{\infty} \mu_0(A_i) : A_i \in \mathcal{A},\ E \subseteq \bigcup_{i=1}^{\infty} A_i\right\}$$

is an outer measure satisfying

a) $\mu^*|_{\mathcal{A}} = \mu_0$
b) Every set in $\mathcal{A}$ is μ^*-measurable.

To obtain a measure, one then appeals to the Carathéodory Extension Theorem:

Theorem 3.4 (Carathéodory)**.** *If μ^* is an outer measure on S, then the collection $\mathcal{M}$ of μ^*-measurable sets is a σ-algebra, and the restriction of μ^* to $\mathcal{M}$ is a (complete) measure.*

Finally, one can show that if μ_0 is σ-finite, then the measure $\mu = \mu^*|_{\mathcal{M}}$ is the unique extension of μ_0 to $\mathcal{M}$.

Using these ideas, one can construct Borel measures on $\mathbb{R}$ by taking any nondecreasing, right-continuous function F (called a *Lebesgue-Stieltjes measure function*) and defining a premeasure on the algebra

$$\mathcal{A} = \left\{ \bigcup_{i=1}^{n} (a_i, b_i] : -\infty \le a_i \le b_i \le \infty,\ (a_i, b_i] \cap (a_j, b_j] = \emptyset,\ n \in \mathbb{N} \right\}$$

by

$$\mu_0(\emptyset) = 0,$$
$$\mu_0\left(\bigsqcup_{i=1}^{n} (a_i, b_i] \right) = \sum_{i=1}^{n} \left[F(b_i) - F(a_i) \right].$$

Lebesgue measure is the special case $F(x) = x$.

The above construction is typical of how one builds premeasures on algebras from more elementary objects:

A *semialgebra* $\mathcal{S}$ is a nonempty collection of sets satisfying

i) $A, B \in \mathcal{S}$ implies $A \cap B \in \mathcal{S}$

ii) $A \in \mathcal{S}$ implies there exists a finite collection of disjoint sets $A_1, ..., A_n \in \mathcal{S}$ with $A^C = \bigsqcup_{i=1}^{n} A_i$.

(Some authors require that $S \in \mathcal{S}$ as well and call the above a semiring. We will not worry about this distinction as we are ultimately only concerned with the algebra $\mathcal{S}$ generates.)

An important example of a semialgebra on $\mathbb{R}$ is the collection of *h-intervals* - that is, sets of the form $(a, b]$ or (a, ∞) or $\emptyset$ with $-\infty \le a < b < \infty$.
On $\mathbb{R}^d$, the collection of products of h-intervals - e.g. $(a_1, b_1] \times \cdots \times (a_d, b_d]$ - is a semialgebra.

If $\mathcal{S}$ is a semialgebra, then one readily verifies that $\overline{\mathcal{S}} = \{\text{finite disjoint unions of sets in } \mathcal{S}\}$ is an algebra (called the *algebra generated by* $\mathcal{S}$). Note that this construction ensures that $\sigma(\mathcal{S}) = \sigma\left(\overline{\mathcal{S}}\right)$.

Given a semialgebra $\mathcal{S}$ and a function $\nu : \mathcal{S} \to [0, \infty)$ such that if $A \in \mathcal{S}$ is the disjoint union of $A_1, ..., A_n \in \mathcal{S}$, then $\nu(A) = \sum_{i=1}^{n} \nu(A_i)$, define $\overline{\nu} : \overline{\mathcal{S}} \to [0, \infty)$ by $\overline{\nu}\left(\bigsqcup_{i=1}^{m} B_i\right) = \sum_{i=1}^{m} \nu(B_i)$.

It is easy to check that $\overline{\nu}$ is well-defined, finite, and finitely additive on $\overline{\mathcal{S}}$.

To verify countable additivity (so that $\overline{\nu}$ is a premeasure on $\overline{\mathcal{S}}$), it suffices to show that if $\{B_n\}_{n=1}^{\infty}$ is a sequence of sets in $\overline{\mathcal{S}}$ with $B_n \searrow \emptyset$, then $\overline{\nu}(B_n) \searrow 0$.

Indeed if $\{A_i\}_{i=1}^{\infty}$ is a countable collection of disjoint sets in $\overline{\mathcal{S}}$ such that $A = \bigcup_{i=1}^{\infty} A_i \in \overline{\mathcal{S}}$, then for any $n \in \mathbb{N}$, $B_n = \bigcup_{i=n}^{\infty} A_i = A \setminus \bigcup_{i=1}^{n-1} A_i$ belongs to the algebra $\overline{\mathcal{S}}$, so finite additivity implies that $\overline{\nu}(A) = \sum_{i=1}^{n-1} \overline{\nu}(A_i) + \overline{\nu}(B_n)$.

Alternatively, one can show that $\overline{\nu}$ is countably additive on $\overline{\mathcal{S}}$ if ν is countably subadditive on $\mathcal{S}$ - that is, for every countable disjoint collection $\{A_i\}_{i \in I} \subseteq \mathcal{S}$ such that $\bigcup_{i \in I} A_i \in \mathcal{S}$, one has $\nu\left(\bigsqcup_{i \in I} A_i\right) \le \sum_{i \in I} \nu(A_i)$. (The implication is immediate if ν is countably additive on $\mathcal{S}$.)

Thus if one takes a finitely additive $[0, \infty)$-valued function ν on a semialgebra $\mathcal{S}$, extends it in the obvious way to the function $\overline{\nu}$ on the $\overline{\mathcal{S}}$, and then checks that $\overline{\nu}$ is countably additive, then the Carathéodory construction guarantees the existence of a unique measure μ on $\sigma(\mathcal{S})$ which agrees with ν on $\mathcal{S}$.

Classifying Distributions on $\mathbb{R}$.

At this point, we recall the following definitions from measure theory:

Definition. If μ and ν are measures on $(S, \mathcal{G})$, then we say that ν is *absolutely continuous* with respect to μ (and write $\nu \ll \mu$) if $\nu(A) = 0$ for all $A \in \mathcal{G}$ with $\mu(A) = 0$.

Definition. If μ and ν are measures on $(S, \mathcal{G})$, then we say that μ and ν are *mutually singular* (and write $\mu \perp \nu$) if there exist $E, F \in \mathcal{G}$ such that

i) $E \cap F = \emptyset$

ii) $E \cup F = S$

iii) $\mu(F) = 0 = \nu(E)$.

A fundamental result in measure theory is the Lebesgue-Radon-Nikodym Theorem (which we state only for positive measures).

Theorem 3.5 (Lebesgue-Radon-Nikodym)**.** *If μ and ν are σ-finite measures on $(S, \mathcal{G})$, then there exist unique σ-finite measures λ, ρ on $(S, \mathcal{G})$ such that*

$$\lambda \perp \mu, \qquad \rho \ll \mu, \qquad \nu = \lambda + \rho.$$

Moreover, there is a measurable function $f : S \to [0, \infty)$ such that $\rho(E) = \int_E f d\mu$ for all $E \in \mathcal{G}$.

The function f from Theorem 3.5 is called the *Radon-Nikodym derivative* of ρ with respect to μ, and one writes $f = \frac{d\rho}{d\mu}$ (or $d\rho = f d\mu$).

If ν is a finite measure, then λ and ρ are finite, so f is μ-integrable.

If a random variable X has distribution μ which is absolutely continuous with respect to Lebesgue measure, then we say that (the distribution of) X has *density function* $f = \frac{d\mu}{dm}$.

Thus for all $E \in \mathcal{B}$, $P(X \in E) = \mu(E) = \int_E f(x)dx$.

In particular, the distribution function of X can be written as

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)dt.$$

Accordingly, F is an absolutely continuous function and is m-almost everywhere differentiable with $F' = f$.

Conversely, if g is a nonnegative measurable function with $\int_{\mathbb{R}} g(x)dx = 1$, then $G(x) = \int_{-\infty}^{x} g(t)dt$ satisfies (i)-(iii) in Theorem 3.1, so Theorem 3.2 gives a random variable with density g.

In undergraduate probability, such an X is called continuous. This is actually somewhat of a misnomer. Rather, we have

Definition. If the distribution of X has a density, then we say that X is *absolutely continuous*.

The other class of random variables discussed in undergraduate probability are discrete random variables.

Definition. A measure μ is said to be *discrete* if there is a countable set S with $\mu\left(S^C\right) = 0$. A random variable is called discrete if its distribution is.

Note that if X is discrete, then $\mu \perp m$.

An example of a discrete distribution is the point mass at a: $P(X = a) = 1$, $F(x) = 1_{[a,\infty)}(x)$.

More generally, given any countable set $S \subset \mathbb{R}$ and any sequence of nonnegative numbers $p_1, p_2, ...$ with $\sum_{i=1}^{\infty} p_i = 1$, if we enumerate S by $S = \{s_1, s_2, ...\}$, then the random variable X with $P(X = s_i) = p_i$, $F(x) = \sum_{i=1}^{\infty} p_i 1_{[s_i,\infty)}(x)$ is discrete, and indeed all discrete random variables are of this form. (Countable additivity implies that μ is determined by its values on singleton subsets of S.)

In the case $S = \mathbb{Q}$ and $p_i > 0$ for all i, we have a discrete random variable whose distribution function is discontinuous on a dense set.

If we think of summation as integration with respect to counting measure, then just as the absolutely continuous random variables correspond to densities ($f \geq 0$ with $\int f dm = 1$), we see that the discrete random variables correspond to mass functions ($p \geq 0$ with $\int p\, dc = 1$).

There is also a third fundamental class of random variables, which we almost never have to deal with, but mention for the sake of completeness. To describe it, we need another definition.

Definition. A measure μ is called *continuous* if $\mu(\{x\}) = 0$ for all $x \in \mathbb{R}$.

By countable additivity, a discrete probability measure is not continuous and vice versa.

Absolutely continuous distributions are continuous, but it is possible for a continuous distribution to be singular with respect to Lebesgue measure.

Definition. A random variable X with continuous distribution $\mu \perp m$ is called *singular continuous*.

An example is given by the "uniform distribution on the Cantor set" formed by taking $[0, 1]$ and successively removing the open middle third of all remaining intervals. The distribution function is the Cantor function given by $F(x) = \frac{1}{2}$ for $x \in [\frac{1}{3}, \frac{2}{3}]$, $F(x) = \frac{1}{4}$ for $x \in [\frac{1}{9}, \frac{2}{9}]$, $F(x) = \frac{3}{4}$ for $x \in [\frac{7}{9}, \frac{8}{9}]$, etc...

Analogous to the singular/absolutely continuous decomposition in the Theorem 3.5, we have the following result for finite Borel measures on $\mathbb{R}$.

Theorem 3.6. *Any finite Borel measure can be uniquely written as*

$$\mu = \mu_d + \mu_c$$

where μ_d is discrete and μ_c is continuous.

Proof. Let $E = \{x \in \mathbb{R} : \mu(\{x\}) > 0\}$.
For any countable $F \subseteq E$, $\sum_{x \in F} \mu(\{x\}) = \mu(F) < \infty$ by countable additivity and finiteness.
It follows that $E_k = \{x \in \mathbb{R} : \mu(\{x\}) > k^{-1}\}$ is finite for all $k \in \mathbb{N}$.
Consequently, $E = \bigcup_{k=1}^{\infty} E_k$ is a countable union of finite sets and thus is countable.
The result follows by defining $\mu_d(A) = \mu(A \cap E)$, $\mu_c(A) = \mu(A \cap E^C)$. □

(The proof is easily modified to accommodate σ-finite measures.)

Thus if μ is a probability distribution, then it follows from the Radon-Nikodym Theorem that $\mu = \mu_{ac} + \mu_s$ where $\mu_{ac} \ll m$ and $\mu_s \perp m$. By Theorem 3.6, $\mu_s = \mu_d + \mu_{sc}$ where μ_d is discrete and μ_{sc} is singular continuous. Since μ is a probability measure, each of $\mu_{ac}, \mu_d, \mu_{sc}$ is finite and thus is identically zero or a multiple of a probability measure. Accordingly, we have

Theorem 3.7. *Every distribution is a convex combination of an absolutely continuous distribution, a discrete distribution, and an absolutely singular distribution.*

Remark. Theorem 3.7 is not especially useful in practice. Rather, we mention these facts because so many introductory texts make a big deal about distinguishing between discrete and continuous random variables. There are certainly important practical differences between the two, and it is worth knowing that more pathological examples exist as well. However, one of the advantages of the measure theoretic approach is a more unified perspective, and excessive focus on differences in detail can sometimes obscure the bigger picture.