Why the Cumulative Distribution Function is Enough

Richard Chapling

v1.0 20 April 2021

The purpose of this handout is twofold: to explain the somewhat odd condition $X^{-1}\big((-\infty, a]\big) \in \mathcal{F}$ for a continuous random variable, and to explain why cumulative distribution functions (and their specialised relatives the mass function and density function) seem to be the only thing we are interested in for the probability distribution of random variables.

1 Preliminaries: σ-algebras, π-systems and measures

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. We also need to worry about the following measurable space:

Definition 1. The *Borel σ-algebra* on $\mathbb{R}$, written $\mathcal{B}$, is the smallest σ-algebra on $\mathbb{R}$ containing all the open intervals.

We say that $\mathcal{B}$ is *generated* by the open intervals. It is easier (and, it turns out, equivalent) to think of the σ-algebra generated by a collection $\mathcal{A} \in \mathcal{P}(\Omega)$ as formed by repeated application of the σ-algebra operations (complement and countable union) to the elements of $\mathcal{A}$. We write $\sigma(\mathcal{A})$ for the σ-algebra generated by $\mathcal{A}$.

It is easy to show that

Result 2. *$\mathcal{B}$ is also generated by $\mathcal{I} = \{(-\infty, x] : x \in \mathbb{R}\}$.*

(Show that every open interval can be formed from such intervals by σ-algebra operations.) It is this that will motivate the idea of the cumulative distribution function. To clarify this, we have a definition and a very useful application of it:

Definition 3. A system of subsets $\mathcal{A}$ is called a *π-system* if it is closed under finite intersections.

In comparison to the general objects in the σ-algebra, which turn out to be rather complicated, large in number, and heterogeneous, π-systems are simple to understand and can have a small number of similar-looking sets in them.

For example,

1. $\mathcal{I}$ is a π-system.
2. The set of open intervals of $\mathbb{R}$ is too, provided we include $\emptyset$.
3. For a countable sample space Ω, the set of subsets containing at most one element is a π-system, that generates the whole of the discrete σ-algebra $\mathcal{P}(\Omega)$.

It turns out that a π-system is the "right" sort of generating set for a σ-algebra, for the following reason:

Result 4 (Uniqueness of extension). *Let $\mathcal{A}$ be a π-system, and let μ_1 and μ_2 be measures[1] on $\sigma(\mathcal{A})$. If $\mu_1(A) = \mu_2(A)$ for every $A \in \mathcal{A}$, then $\mu_1(S) = \mu_2(S)$ for every $S \in \sigma(\mathcal{A})$.*

[1] A *measure* is a more general concept than a probability measure, but the change in definition is slight: we replace the condition $\mathbb{P}(\Omega) = 1$ with $\mu(\emptyset) = 0$, which allows μ to assign any nonnegative value to elements of $\mathcal{F}$, rather than just those in $[0, 1]$. (We keep countable additivity as-is.)

(This is a significant result, and a proof is well beyond our scope in both length and difficulty.)

Thus specifying a measure on the (easy, small) π-system is enough to specify it on the (big, complicated) σ-algebra it generates.

With the preliminaries over, we can move on to defining random variables and so on.

2 Random variables

Definition 5. Let $(\Omega, \mathcal{F})$ be a measurable space. A *random variable* is a function $X\colon \Omega \to \mathbb{R}$ with

$$(\forall B \in \mathcal{B}) \quad X^{-1}(B) \in \mathcal{F}. \tag{1}$$

A more practical condition equivalent to (1)[2] is

$$(\forall x \in \mathbb{R}) \quad X^{-1}\big((-\infty, x]\big) \in \mathcal{F},$$

and this is the one usually given when defining a random variable.

Such conditions probably looks rather annoying and superfluous, but as soon as we try to associate probabilities to X, they are absolutely essential, as we will see next.[3]

3 Probability distribution

Definition 6. A (univariate) *probability distribution D* is a particular assignment of probability to certain subsets of the real numbers. That is, it is a probability measure $\mathbb{P}_D$ on $\mathbb{R}$, with a particular σ-algebra, which in this course we can always take (implicitly) to be $\mathcal{B}$.

You already know plenty of examples of such things:

- The Bernoulli, Binomial, Poisson and Geometric distributions all assign positive probability only to a countable set. Such distributions are called *discrete*.
- The Exponential and Normal distributions are very different, in that they do not assign probability to single points. They are examples of *continuous* distributions.[4]

3.1 Distribution of a random variable

Given a probability measure $\mathbb{P}$ and a random variable X, precisely because of the condition (1), the inverse images under X of sets in $\mathcal{B}$ are elements of $\mathcal{F}$, and so can be used as inputs for $\mathbb{P}$. Hence we can obtain a new probability measure on $\mathbb{R}$ using the σ-algebra $\mathcal{B}$, by composition:

$$\begin{aligned} p_X &\colon \mathcal{B} \to [0,1] \\ p_X &= \mathbb{P} \circ X^{-1}. \end{aligned}$$

p_X is called the *law* of X, or the *pushforward measure* associated to X.

[2]Because the set of half-infinite intervals is a π-system that generates $\mathcal{B}$.

[3]Although if we are in the discrete σ-algebra $\mathcal{P}(\Omega)$, note that the condition is automatically satisfied.

[4]The actual definition of a continuous probability distribution is that it does not assign positive probability to any subset of $\mathbb{R}$ of Lebesgue measure 0, but to explain this properly is rather far from the point of this handout. We shall give a simpler criterion for a distribution to be continuous after we define the CDF.

We can sum this up with the following diagram:

$$\begin{array}{ccc} \Omega & \xrightarrow{X} & \mathbb{R} \\ \mathcal{F} & \xleftarrow[X^{-1}]{} & \mathcal{B} \\ \mathbb{P}\big\downarrow & \swarrow_{p_X} & \\ [0,1] & & \end{array}$$

If $p_X = \mathbb{P}_D$ for some probability distribution D, we write $X \sim D$.[5]

3.2 The Cumulative Distribution Function

Since the Borel σ-algebra $\mathcal{B}$ is generated by the intervals $(-\infty, x]$, we make the following definition:

Definition 7 (Cumulative distribution function[6]). Let $X : \Omega \to \mathbb{R}$ be a random variable. The *cumulative distribution function* (CDF) of X is the function $F : \mathbb{R} \to [0, 1]$ given by

$$F_X(x) = \mathbb{P}(X \leqslant x).$$

That is, $F_X = \mathbb{P}\big(X^{-1}\big((-\infty, x]\big)\big)$. Because $\mathcal{B}$ is generated by the π-system $\mathcal{I}$, on which p_X and F_X agree, Theorem 4 implies that the CDF of X determines the law of X, so *it is the only thing we need to understand everything about the probability distribution of X*.

Lastly, we prove the main properties of the CDF.

Proposition 8 (Properties of CDF). *Let F_X be the CDF of a random variable $X : \Omega \to \mathbb{R}$. Then*

1. *F_X is nondecreasing.*
2. *F_X is* right-continuous, *in that for every $x \in \mathbb{R}$, $\lim_{y \downarrow x} F_X(x) = F_X(x)$.*
3. $\lim_{x \to +\infty} F_X(x) = 1$.
4. $\lim_{x \to -\infty} F_X(x) = 0$.

Proof.

1. This is a consequence of the monotonicity of the probability measure: if $y > x$,

$$\mathbb{P}(X \leqslant y) = \mathbb{P}(X \leqslant x) + \mathbb{P}(x < X \leqslant y) \geqslant \mathbb{P}(X \leqslant x).$$

2. This, and the other two parts, are a consequence of the continuity of probability measure. Since $(x, \infty) = \bigcup_{n=1}^{\infty} (x + 1/n, \infty)$ is an increasing union, we have

$$1 - F_X(x + 1/n) = \mathbb{P}(X > x + 1/n) \to \mathbb{P}(X > x) = 1 - F_X(x).$$

3. We have $\mathbb{R} = \bigcup_{n=1}^{\infty} (-\infty, n]$, so $1 = \mathbb{P}(\mathbb{R}) = \lim_{n\to\infty} \mathbb{P}((-\infty, n]) = \lim_{n\to\infty} F_X(n)$, and the result follows from monotonicity.
4. Similarly to the previous result, but using $1 - F_X$ and $\mathbb{P}(X > -n)$. □

While the CDF is the most useful function associated to the distribution theoretically, for many distributions, the CDF is not expressible in a closed form: a simple example of this is that of the binomial distribution, which requires a hypergeometric function to express. Therefore, it is helpful to have some ancillary functions that, while less general, are more useful in specific cases.

[5] Notice that unlike other uses of this symbol, the objects on each side of this relation are different: a random variable on the left and a probability distribution on the right.

[6] We prefer this to the common synonym "distribution function" to make it more distinct from other functions associated to the distribution of the random variable.

3.3 Discrete distributions

For discrete distributions, there is a countable subset S on which $m_X(x) = F_X(x) - \lim_{y \uparrow x} F_X(y) > 0$. The *probability mass function* (PMF) of X is then $m \colon S \longrightarrow [0,1]$, and for $A \subseteq S$, we can write

$$\mathbb{P}(X \in A) = \sum_{x \in A} m_X(x).$$

3.4 Continuous distributions

For continuous distributions, the relationship between the CDF and the probability density function is more complicated. We say a probability distribution X is *continuous* if there is a function f_X so that $\mathbb{P}(X \in B) = \int_B f_X$ for any $B \in \mathcal{B}$; this function is called the *probability density function* of X. Therefore,

$$F_X(x) = \mathbb{P}(X \leqslant x) = \int_{-\infty}^{x} f_X,$$

and so the CDF is continuous. If f_X is continuous, the Fundamental Theorem of Calculus implies that $F_X' = f_X$, but it is easy to give examples of continuous distributions where f_X is not continuous: one need only consider the uniform distribution for a counterexample. Hence, much as we would like to, we cannot say that $F_X' = f_X$ for any continuous distribution. Moreover, there are continuous F_X satisfying the conditions to be a CDF that are not CDFs of continuous probability distributions. The counterexamples here are rather more complicated: the Cantor Function[7] is the usual example. The issue here is not resolvable using the Riemann integral, and the solution uses the full power of the Lebesgue integral: an early triumph of Lebesgue's theory was to obtain the following characterisation of functions satisfying both parts of the Fundamental Theorem of Integral Calculus:

Result 9 (Fundamental Theorems of Integral Calculus for the Lebesgue Integral).

1. *Let f be Lebesgue-integrable. Then $F(x) := \int_a^x f$ is absolutely continuous, and is differentiable for almost every x, with $F'(x) = f(x)$.*
2. *Let F be an absolutely continuous function. Then $F'(x)$ exists for almost every x, and $F(b) - F(a) = \int_a^b F'$.*

There are two terms that need to be explained here.

A property holds *almost everywhere* if it holds on a set whose complement has (Lebesgue-)measure zero.

A simple characterisation of sets of measure zero is that they can be covered by a countable collection of intervals of arbitrarily small length. For example, the rationals have measure zero, but so does the Cantor set.

A function f is called *absolutely continuous* on an interval I if for any $\varepsilon > 0$, we can find a $\delta > 0$ so that for any finite sequence of disjoint intervals $(x_i, y_i) \subseteq I$ with total length smaller than δ (i.e. $\sum_i (y_i - x_i) < \delta$), then $\sum_i |f(y_i) - f(x_i)| < \varepsilon$.

This is a precise way of saying that f changes by little over *collections* of intervals of small total length: it is stronger than ordinary continuity, and also stronger than uniform continuity.

With this, we can characterise the CDFs that are the CDFs of continuous probability distributions: they are precisely those that are absolutely continuous.

[7] See e.g. https://encyclopediaofmath.org/wiki/Cantor_ternary_function