Some Class

Daniel Yu

September 10, 2024

Contents

1 Probability Formalism

1.1 Definitions and Probability Measures

Remark. Probability is severely counterintuitive and in fact, the following formalism of the field was only defined in the 1930s despite the hundreds of years of history.

Definition 1. Given an experiment with multiple possible outcomes, we set:

$$\Omega = \{ \text{ set of all outcomes} \}.$$

Remark. Assume Ω is at most countable. Dealing with uncountable Ω is where measure-theoretic probability comes in.

Definition 2. An event A is any subset of Ω. $A_i \subseteq \Omega$.

Definition 3. A probability measure, P, is any function $P : \{ \text{ events } \} \to [0, 1]$, such that:

1. $P[\emptyset] = 0$
2. $P[\Omega] = 1$
3. If $\{A_i\}_{i=1}^{\infty}$ (set of different events) is an at most countable **collection of disjoint events** then, $P[\cup_{i=1}^{\infty} A_i] = \sum_{i=1}^{\infty} P[A_i]$.

Note. Example 1
$\Omega = \{H, T\}$

$$\text{set of events } = \{\emptyset, \{H\}, \{T\}, \{H, T\}\}.$$

from the definition of probability measure, property 3:

$$P[\{H\} \cup \{T\}] = P[\{H\}] + P[\{T\}] = P[\Omega] = 1.$$

Note. Example 2

$$\Omega = \{1, 2, 3, 4, 5, 6\} \text{ A six-sided die roll for example.}$$

$A = \{ \text{ outcome is at most } 2\} = \{1, 2\}$
$B = \{ \text{ outcome is at least } 5\} = \{5, 6\}$

$$P[A \cup B] = P[A] + P[B].$$

Remark. Suppose that $P[w]$ is the same for all $w \in \Omega$. So in particular, $P[\{1\}] = \frac{1}{|\Omega|} = \frac{1}{6}$ so for any event A,

$$P[A] = P[\cup_{w \in A}\{w\}] = \sum_{w \in A} P[\{w\}] = \frac{|A|}{|\Omega|}.$$

. This simplifies to counting, very easy, but of course a measure need not be uniform . . .

Note. Example 3
Assume that there is a pile of cards of color, R, G or B and I drew one at random

$$P(R \cup B) = \frac{3}{5}P(R \cup G) = \frac{3}{5}.$$

what is the probability of $P[R]$?
We know that $\Omega = \{R, G, B\}$

$$P[R] + P[B] + P[G] = 1.$$

$$P[R] + P[B] = \frac{3}{5}.$$

$$P[R] + P[G] = \frac{3}{5}.$$

We get $P[R] = \frac{1}{5}$.

Notice that we are solving a **determined** system!

Properties
There are a few consequences of the definition:

1. $P[A^c] = 1 - P[A]$.
   Proof: $P[\Omega] = 1 = P[A \cup A^c] = P[A] + P[A^c]$
2. If $A \subseteq B$, $P[A] \leq P[B]$.
   Proof: $P[B] = P[A \cup \{B \setminus A\}] = P[A] + P[B \setminus A]$. We know that $P[B \setminus A] \geq 0$, so $P[B] \geq P[A]$
3. If we have an arbitrary at most countable collection of events (not just disjoint) $\{B_i\}_{i=1}^{\infty}$, then $P[\cup_{i=1}^{\infty} B_i] \leq \sum_{i=1}^{\infty} P[B_i]$. Intuition is that B_i may overlap events(i.e. outcomes) with B_j and we double count so it provides an upper bound.

Note. Example 4
52 card deck, and we draw 3 cards to form a hand. If all hands have the same possibility, what is the probability of getting three 2's.

$\Omega = \{$ of all possible 3 card combinations without concern for order $\}$. Note that it is not the set of all 52 cards!

$$|\Omega| = \binom{52}{3} = \frac{52 \times 51 \times 50}{6}.$$

and the associated probability measure $P[w] = \frac{1}{|\Omega|}$ for any $w \in \Omega$.

$A = \{\text{hands of three 2's }\} = \{\{2Hearts, 2C, 2D\}, \{2H, 2C, 2S\}, \{2H, 2D, 2S\}, \{2C, 2D, 2S\}\}$.

$$P[\text{ getting 3 2's in hand}] = \frac{|A|}{|\Omega|} = \frac{4}{\text{52 choose 3}}.$$

If we instead, consider the outcomes as cards dealt in order (i.e. order matters), then:

$$|\Omega'| = 52 \times 51 \times 50..$$

i.e. in this case cards are no longer unordered triplets (a set) but ordered triples! So,

$$B \subseteq \Omega' = \{\text{hand of three 2's}\} = 4 \times 3 \times 2.$$

$$P[\text{choose 3 2's in hand}] = \frac{24}{52 \times 51 \times 50}.$$

Note that the two approachs give the same answer!

1.2 Random Variable

Start

> **Definition 4.** A random Variable is any map from Ω to $\mathbb{R}$ or ($\mathbb{R}^n$).

Note. Example $\Omega = \{H, T\}$

$$X = \{1\text{head}, 0\text{tail}\}.$$

Suppose $P[\{H\}] = \frac{1}{3}$. Therefore,

$$P[X = 1] = \frac{1}{3}.$$

$$P[X = 0] = \frac{2}{3}.$$

> **Definition 5.** The **distribution** of X is the $P[X = a]$ for every a where $a \in \mathbb{R}$ and $a \neq 0$. Note that **the distribution of a random variable is the measure it induces on** $\mathbb{R}$

Remark. For any set $S \subseteq \mathbb{R}$, at most countable, we can describe a distribution such that $P[x = a] \geq 0$ $\forall a \in S$ and $P[x = a] = 0$ otherwise. Any function (probability measure) that satisfies:

1. $P[x = a] \geq 0$ for $a \in S$
2. $\sum_{a \in S} P[x = a] = 1$

Remark. Infinite Set
$\Omega = \{1, 2, 3, \ldots\}$, an infinite set.
Assume that $P[z = k]$ forms a geometric series, pick $p \in (0, 1]$ and set $P[z = k] = p \cdot (1 - p)^{k-1}$. This is a distribution:

$$p[1 + (1 - p) + (1 - p)^2 + (1 - p)^3 + \ldots] = p \cdot \frac{1}{1 - \frac{1}{p}} = 1.$$

So this does form a probability distribution. Note that $p \neq 0$ because then this would go to 0.

Remark. Poisson Distribution
$R\sim$ Poisson(λ) means that R is distributed as a poisson distribution.

$$P[R=k]=\frac{e^{-\lambda}\cdot\lambda^k}{k!}\forall k\in\{0,1,2,3,\dots\}.$$

It turns out this is also a distribution:

$$e^{-\lambda}(\sum_{k=0}^{\infty}\frac{\lambda^k}{k!})=1.$$

Note. Example

$$\Omega=\{1,2,3,4,5,6\}. \text{ P is uniform..}$$

$$Y=\{1\text{if w is divisible by }3,0\text{if w is not}\}.$$

The distribution of Y is:

$$P[Y=1]=\frac{1}{3}.$$

$$P[Y=0]=\frac{2}{3}.$$

$$P[Y=\alpha]=0\forall\alpha\notin\{0,1\}.$$

Remark. Even though X and Y are different random variables, they **share the same distribution**

Definition 6. If

$$P[X=1]=p.$$

$$P[X=0]=1-p.$$

We say x is distributed as a **Bernoulli random variable** of parameter p.

Note. Example

$$\Omega=\{\text{outcome of two rolls of a 6-sided die}\}\text{where P is uniform.}$$

M is the random variable represnting the maximum of 2 rolls. What is the distribution of M?

$P[M=x]=\frac{6^2-x^2}{6^2}$

1.3 CDFs and PDFs

Definition 7. Given a random variable X, the **cumulative distribution fuction (cdf)** of X, denoted as

$$F_X(t)=P[X\leq t]$$

. With the properties:

1. $\lim_{t\to-\infty}F_x(t)=\lim_{t\to-\infty}P[x\leq t]=0$
2. $\lim_{t\to\infty}F_x(t)=1$
3. if $s\leq t$, $F_x(s)\leq F_x(t)$ i.e. it is non-decreasing

Note. Thus everything we've done up to this point has been with probability measure, not cdfs, pdfs, etc. yet!

Note. Bernoulli
So if X = Bernoulli distribution then the cdf would be:

$$P[x \leq -17] = 0.$$
$$P[x \leq 0] = P[x = 0] = 1 - p.$$
$$P[x \leq \frac{2}{3}] = P[x = 0] = 1 - p.$$
$$P[x \leq 1] = 1.$$
$$P[x \leq 2] = P[x \leq 1] = 1.$$

Note. In general, an discrete random variable would create a piecewise cdf!

Note. If we pick $U \in [0, 1]$ uniformly, then $F_U(t)$ would be 0 until x = 0, then from x = 0 to x = 1, it would increase along the line $y = x$, then at $x \geq 1$, it would be a straight line with value 1.

Definition 8. The **probability density function (pdf)** of a continuous r.v. is:

$$f_X(t) = \frac{d}{dt} F_X(t).$$

, the first derivative of the cdf. With the two properties:

1. $f_x(t) \geq 0$
2. $\int_{-\infty}^{\infty} f_x(t) = 1$

The way to interpret $f_X(t)$ is that X, the random variable, is a function that maps from outcomes $\Omega \to \mathbb{R}$ and t are values $\in range(X) = \mathbb{R}$. So when we parameterize by t we are letting t vary across the range of values that the random variable X could take.

Note. The pdf of uniform $[0, 1]$ random variable:

$$f_x(t) = \begin{cases} 1, t \in [0, 1] \\ 0, \text{ otherwise} \end{cases} .$$

Proposition 1. If x is a random variable and $a < b$, then

$$P[a < x \leq b] = F_x(b) - F_x(a).$$

If the r.v. is continuous, then:

$$P[a < x \leq b] = \int_a^b f_x(t) dt.$$

Fundamental theorem of calc lol

Proof.

$$F_x(b) - F_x(a) = P[x \leq b] - P[x \leq a].$$

We know that by set theory

$$P[x \leq b] = P[\{x \leq a\} \cup \{a < x \leq b\}] = P[x \leq a] + P[a < x \leq b].$$

Subsitute back to equation 1.

$$F_x(b) - F_x(a) = P[x \leq a] + P[a < x \leq b] - P[x \leq a] = P[a < x \leq b].$$

By the fundamental theorem of calculus this holds for the continuous case: □

Note. Consider

$$f_U(t)dt = \begin{cases} c, t \in [a, b] \\ 0 \text{ otherwise} \end{cases}.$$

To find c:

$$\int_{-\infty}^{\infty} f_U(t)dt = 1.$$

$$\int_a^b cdt == c \cdot (b - a) = 1.$$

Thus the cdf is:

$$\begin{cases} \frac{1}{b-a}, t \in [a, b] \\ 0 \text{ otherwise} \end{cases}.$$

Note. Exponential Distribution
$z \sim \exp(\lambda)$ of it's pdf is:

$$f_z(t) = \begin{cases} e^{-\lambda t}, t \in [0, \infty] \\ 0, \text{otherwise} \end{cases}.$$

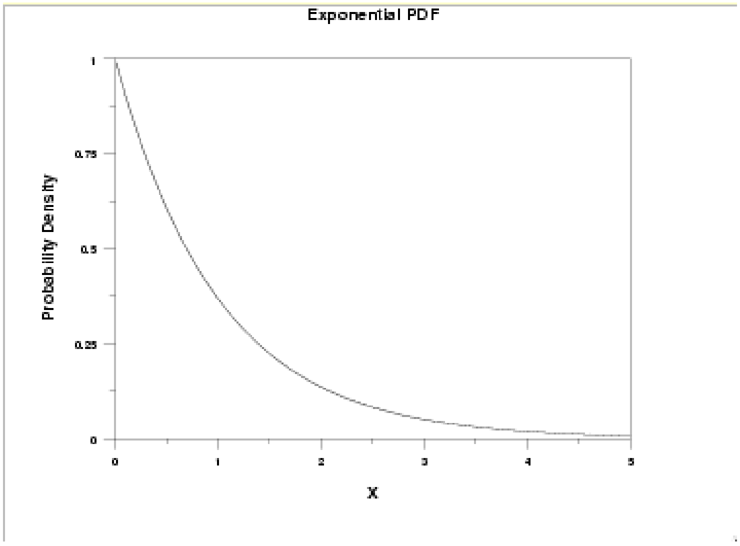

Figure 1: Exponential Distribution PDF

What is it's cdf?

$$F_z(t) = \int_{-\infty}^{t} f_z(s)ds = \int_0^t f_z(s)ds = \int_0^t \lambda e^{-\lambda s} ds$$
$$= -e^{-\lambda t - (-1)} = 1 - e^{-\lambda t}.$$

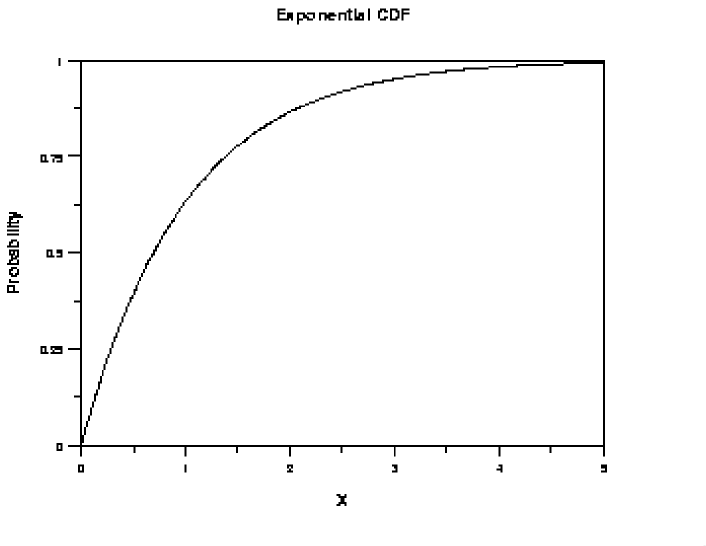

Figure 2: Exponential Distribution CDF

1.3.1 Normal Distribution

We denote the normal distribution as $N \sim$ normal(μ, σ^2).
The pdf of the normal distribution is: $f_N(t) = \frac{1}{\sigma \cdot \sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}}$

Note. Let $X \sim N(0,1)$ and $Y = X^2$. What is the pdf of Y? HARD. So let's look at CDF instead:

$$\begin{aligned} F_Y(t) = P[Y < t] &= P[x^2 \leq t] = P[-\sqrt{t} \leq x \leq \sqrt{t}] \\ &= \int_{-\sqrt{t}}^{\sqrt{t}} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx \text{ the function is even} \\ &= 2 \int_0^{\sqrt{t}} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx. \end{aligned}$$

Now take the pdf of Y by taking first derivative w.r.t. t:

$$\frac{d}{dt} \left(2 \int_0^{\sqrt{t}} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx \right).$$

let $u = \sqrt{t}$, $\frac{d}{dt} = \frac{d}{du} \cdot \frac{du}{dt}$:

$$\begin{aligned} &= \frac{d}{du} \left(2 \int_0^{u} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx \right) \cdot \frac{du}{dt}. \\ &= 2 \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{\frac{-\sqrt{t}^2}{2}} \cdot \frac{d(\sqrt{t})}{dt}. \\ &= \frac{1}{\sqrt{2\pi t}} \cdot e^{\frac{-t}{2}}. \end{aligned}$$

This is known as the chi-squared distribution with 1 degree of freedom.

2 Joint Distributions

We need joint dsitributions to describe the behavior of multiple random variables, we need the joint distributions. The marginal distributions are not enough!

Note. For example, consider the discrete random variables x, y, z over $\Omega = \{1,2,3,4\}$ on a uniform probability measure:

$$X = \begin{cases} 1, w \in \{1,2\} \\ 0, w \in \{3,4\} \end{cases}.$$

$$Y = \begin{cases} 1, w \in \{3,4\} \\ 0, w \in \{1,2\} \end{cases}.$$

$$z = \begin{cases} 1, w \in \{1,3\} \\ 0, w \in \{2,4\} \end{cases}.$$

The marginal distribution of X, Y, Z are bernoulli distributions with $p = .5$. (But they are not the same r.v.). Moreover, $P[X = Y] = 0$ and $P[X = Z] = .5$ despite the fact that X, Y, Z are random variables with same distribution (seeming in contradiction to above)! The marginal distributions give an incomplete idea!

Thus we need the joint distribution!

If X, Y are continuous random variables, we need to understand them via joint cdf or joint pdf.

$$F_{X,Y} = P[X \leq s, Y \leq t].$$

Remark. Example
We are dealing 3 cards out of a 52 card deck with uniform probability.

$$X = \begin{cases} 1 \text{ if 1st card is ace} \\ 0 \text{ otherwise} \end{cases}.$$

$$Y = \begin{cases} 1 \text{ if 2nd card is ace} \\ 0 \text{ otherwise} \end{cases}.$$

$$Z = \begin{cases} 1 \text{ if 3rd card is ace} \\ 0 \text{ otherwise} \end{cases}.$$

clearly $P[X = 1] = \frac{4 \cdot 51 \cdot 50}{52 \cdot 51 \cdot 50} = \frac{1}{13}$

What is the joint probability distribution of X and Y?
$P[X = 0, Y = 0] = \frac{48 \cdot 47 \cdot 50}{52 \cdot 51 \cdot 50}$
$P[X = 0, Y = 1] = \frac{48 \cdot 4 \cdot 50}{52 \cdot 51 \cdot 50}$
$P[X = 1, Y = 0] = \frac{4 \cdot 48 \cdot 50}{52 \cdot 51 \cdot 50}$
$P[X = 1, Y = 1] = \frac{4 \cdot 3 \cdot 50}{52 \cdot 51 \cdot 50}$

We can recover the marginal distribution of Y through this:
$P[X = 0, Y = 0] + P[X = 1, Y = 0] = P[Y = 0] = \frac{48 \cdot 50(4+47)}{52 \cdot 51 \cdot 50} = \frac{48}{52} = \frac{12}{13}$ and thus $P[Y = 1] = \frac{1}{13}$