

# Probability I - Random Variables

Arkaraj Mukherjee

August 31, 2025

For a sample space  $\Omega$ , event space  $\mathcal{F}$  and probability  $P$ , a *random variable* is a function  $X : \Omega \rightarrow \mathbb{R}$ . the set of values the random variable can take, i.e. it's range is called the *support* of the RV (short form of Random Variable) often denoted by

$$\text{supp } X := \{X(\omega) | \omega \in \Omega\}$$

RVs  $X$  for which  $\text{supp } X$  is atmost countable i.e. countably infinite or finite are called *discrete RVs*. The probabilities of the RV taking on some values in its support are governed by the *probability mass function* (*pmf*. in short) often denoted by  $p$ , i.e.

$$\begin{aligned} p : \text{supp } X &\longrightarrow [0, 1] \\ x &\mapsto P(X = x) := P(X^{-1}\{x\}) \end{aligned}$$

In practice we extend this definition of  $p$  on  $\text{supp } X$  all the way to  $\mathbb{R}$  by fixing  $p(x) = 0$  for all  $x \in \mathbb{R} \setminus \text{supp } X$ . For discrete RVs, if we ennumerate  $\text{supp } X = \{x_i | i = 1, 2, \dots\}$  then from the axioms of probability we get that,

$$\sum_{i=1}^{\infty} p(x_i) = \sum_{i=1}^{\infty} \underbrace{P(X^{-1}(\{x_i\}))}_{\text{Why are these three equalities true?}} = P\left(\bigcup_{i=1}^{\infty} X^{-1}(\{x_i\})\right) = P(\Omega) = 1$$

The *cumulative distribution function* (cdf. for short) of a random variable is a function on the real line defined as,

$$\begin{aligned} F : \mathbb{R} &\rightarrow [0, 1] \\ x &\mapsto P(X \leq x) := P(X^{-1}((-\infty, x])) \end{aligned}$$

This is a non-decreasing step-function with jump discontinuities and from the definition we can see that (try proving it)

$$p(x) = F(x) - \lim_{\substack{h \rightarrow x \\ h < x}} F(h)$$

The limit on the right is sometimes known as the *left limit of  $F$  at  $x$* . From the axioms of probability again we can see that

$$F(x) = P(X^{-1}((-\infty, x])) = P\left(\bigcup_{x_i < x} X^{-1}(\{x_i\})\right) = \sum_{x_i < x} P(X^{-1}(\{x_i\})) = \sum_{x_i < x} p(x_i)$$

You should notice that we can also deduce the equation with the left limit from this as well. The *expectation* of a RV is a weighted average of the values it takes, formally this is written as

$$\mathbb{E}[X] := \sum_{i=1}^{\infty} x_i p(x_i)$$

It may not always exist (i.e. converge). Suppose we have a RV  $X$ , then for some function  $f$  on  $\text{supp } X$ ,  $Y = f(X)$  is also a RV i.e. a function  $\Omega \rightarrow \mathbb{R}$  and  $\text{supp } Y = f(\text{supp } X)$ . Also,

$$\begin{aligned}\mathbb{E}[Y] &= \sum_{y \in \text{supp } Y} y P(f(X) = y) = \sum_{y \in \text{supp } Y} \sum_{x \in f^{-1}(y)} y P(X = x) \\ &= \sum_{y \in \text{supp } Y} \sum_{x \in f^{-1}(y)} f(x) P(X = x) = \sum_{x \in \bigcup_{y \in f(\text{supp } X)} f^{-1}(y)} f(x) P(X = x) \\ &= \sum_{x \in \text{supp } X} f(x) P(X = x)\end{aligned}$$

This is a major result. The  $k$ -th moment of a RV is defined as  $\mathbb{E}[X^k] := \mathbb{E}[g(X)]$  where  $g : x \mapsto x^k$ . Let  $X, Y$  be RVs on the same sample space, then we see that when we define  $(X + Y)(\omega) := X(\omega) + Y(\omega)$  then,

$$\begin{aligned}\mathbb{E}[X + Y] &= \sum_{x \in \text{supp } X} \sum_{y \in \text{supp } Y} (x + y) P(X = x \wedge Y = y) \\ &= \sum_{x \in \text{supp } X} x \sum_{y \in \text{supp } Y} P(X = x \wedge Y = y) + \sum_{y \in \text{supp } Y} y \sum_{x \in \text{supp } X} P(X = x \wedge Y = y) \\ &= \sum_{x \in \text{supp } X} x P(X = x) + \sum_{y \in \text{supp } Y} y P(Y = y) = \mathbb{E}[X] + \mathbb{E}[Y]\end{aligned}$$

and clearly for constant  $c$  we have  $\mathbb{E}[cX] = c\mathbb{E}[X]$  thus expectation is linear. For subsets  $A \subseteq \Omega$  we define an *indicator RV*  $I_A$  as the RV that is one on  $A$  and 0 everywhere else, this clearly implies that  $\mathbb{E}[I_A] = P(A)$ . The *mean absolute deviation (MAD)* is defined as  $\mathbb{E}[|X - \mathbb{E}[X]|]$ , the *variance* is defined as  $\text{Var}(X) := \mathbb{E}[(X - \mathbb{E}[X])^2]$  and the *standard deviation* is  $\sqrt{\text{Var}(X)}$ . Clearly  $\text{Var}(X) \geq 0$  and  $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$  thus,  $\mathbb{E}[X]^2 \leq \mathbb{E}[X^2]$ . Also  $\text{Var}(aX + b) = a^2 \text{Var}(X)$ . The *degenerate probability distribution* is where for some  $x_o \in \mathbb{R}$  we have  $\text{supp } X = \{x_o\}$ , for this  $p(x) = \delta_{x, x_o}$  and  $F(x) = 0$  if  $x < x_o$  and 1 otherwise,  $\mathbb{E}[X] = x_o$  and  $\text{Var}(X) = 0$ . The *discrete uniform distribution* is where  $\text{supp } X = \{a, a + 1, \dots, b\}$  for some  $a \leq b$  where we have  $p(x) = 1/|\text{supp } X| = 1/(b - a + 1)$  on its support and zero elsewhere.  $F(x)$  is zero for  $x < a$ ,  $|(-\infty, x] \cap \text{supp } X|/|\text{supp } X| = (\lfloor x \rfloor - a + 1)/(b - a + 1)$  on  $[a, b]$  and 1 elsewhere. Here,  $\mathbb{E}[X] = (a + b)/2$  and  $\text{Var}(X) = \text{Var}(x - (a - 1)) = ((b - a + 1)^2 - 1)/12 \asymp |b - a|^2$ . For some  $p \in [0, 1]$  the *bernoulli distribution*,  $X \sim \text{Ber}(p)$  is where  $\text{supp } X = \{0, 1\}$  with  $p(1) = p$  and  $p(0) = 1 - p$  (usually written as  $q$ ),  $F(x) = 0$  if  $x < 0$ ,  $q$  on  $[0, 1)$  and 1 elsewhere. The *binomial distribution*,  $X \sim \text{Bin}(n, p)$  for  $n \in \mathbb{N}$  and  $p \in [0, 1]$  is where  $\text{supp } X = \{0, 1, \dots, n\}$  and  $p(x) = \binom{n}{x} p^x (1 - p)^{n-x}$  on its support and zero elsewhere but here  $F(x)$  doesn't have a closed form. We say that two RVs  $X, Y$  are independent and write  $X \perp Y$  iff for all  $x, y$   $P(X = x \wedge Y = y) = P(X = x)P(Y = y)$ . Now for discrete RVs  $X, Y$  such that  $X \perp Y$  we can see that for any  $A \subseteq \text{supp } X$  and  $B \subseteq \text{supp } Y$  we have that,

$$P(X \in A \wedge Y \in B) = P\left(\bigcup_{x \in A, y \in B} X^{-1}(\{x\}) \cap Y^{-1}(\{y\})\right) = \sum_{x \in A} \sum_{y \in B} P(X = x \wedge Y = y)$$

$$= \sum_{x \in A} \sum_{y \in B} p_X(x) p_Y(y) = \left( \sum_{x \in A} p_X(x) \right) \cdot \left( \sum_{y \in B} p_Y(y) \right) = P(X \in A) \cdot P(Y \in B)$$

using usual theorems from absolute convergence of  $\sum_{x \in A} p_X(x)$  etc. This tells us that if  $X \perp Y$  then any events defined with  $X, Y$  are also independent. Now we see that if (ignore the fact that the sum of random variables isn't defined for this to work yet, this course sucks)  $X_1, \dots, X_n \sim \text{Ber}(p)$  are independent then  $Y = \sum X_i \sim \text{Bin}(n, p)$ . For a proof, we can see that  $Y$  has support  $\{0, \dots, n\}$  with  $p_Y(k) = P(\text{k of the } X_i\text{'s are 1 and the rest are zero}) = \binom{n}{k} p^k (1-p)^{n-k}$ . Now if  $X \sim \text{Bin}(n, p)$  and  $Y \sim \text{Bin}(m, p)$  are independent then we have that,  $X+Y \sim \text{Bin}(m+n, p)$ . This can be done in two ways, take independent bernoulli RVs  $Z_1, \dots, Z_{m+n} \sim \text{Ber}(p)$  and then,  $X = \sum_{i=1}^n Z_i$  and  $Y = \sum_{i=1}^m Z_i$  giving us that  $X+Y = \sum_{i=1}^{n+m} Z_i \sim \text{Bin}(n+m, p)$ . Apart from this another way is to see that,

$$\begin{aligned} p_{X+Y}(k) &= \sum_{j \geq 0} P(X = j \wedge Y = k - j) = \sum_{j \geq 0} p_X(j) p_Y(k - j) \\ &= \sum_{j \geq 0} \binom{n}{j} \binom{m}{k-j} p^{j+(k-j)} (1-p)^{n-j+(m-(k-j))} \\ &= \sum_{j \geq 0} \binom{n}{j} \binom{m}{k-j} p^k (1-p)^{n+m-k} = [z^k] ((1+z)^n (1+z)^m) p^k (1-p)^{n+m-k} \\ &= \binom{n+m}{k} p^k (1-p)^{n+m-k} \end{aligned}$$

where we have used the convention that  $\binom{a}{b} = 0$  for  $a < b$  and arrive at the same conclusion.