

Discrete Random Variables

Terminology and notations

- **Definition:** Mathematically, a **random variable** (r.v.) on a sample space S is a **function**¹ from S to the real numbers. More informally, a random variable is a numerical quantity that is “random”, in the sense that its value depends on the outcome of a random experiment.
- **Notation:** One commonly uses capital letters X, X_1, X_2, Y , etc., to denote random variables, and lower case letters x, x_1, x_2, y , etc., to denote the corresponding values (i.e., the real numbers that the random variables map into).
- **Discrete random variable:** A random variable whose set of values is finite or countably infinite, i.e., can be listed as a (finite or infinite) sequence x_1, x_2, \dots . In the formulas below, X is assumed to be a *discrete* r.v., with values denoted by x_1, x_2, \dots .
- **Notational shortcuts:** Notations such as “ $P(X = 2)$ ”, “ $P(X \leq 2)$ ”, “ $P(|X - 2| \leq 1)$ ”, where X is a random variable, are, strictly speaking, mathematically nonsensical (since one applies a probability function $P(\dots)$ to an *equation or inequality* rather than a *set*), but they are commonly used and convenient shortcuts to express what would otherwise require a very clumsy formulation: For example, “ $P(X = 2)$ ” is shorthand for the following: “ $P(A)$, where A denotes the set of outcomes at which the random variable X takes the value 2”; i.e.,

$$P(X = 2) = P(\{\omega \in S : X(\omega) = 2\}).$$

Here are some other examples of such shortcuts, and their interpretation:

$$\begin{aligned} P(|X - 2| \leq 1) &= P(\{\omega \in S : |X(\omega) - 2| \leq 1\}), \\ P(X = Y) &= P(\{\omega \in S : X(\omega) = Y(\omega)\}), \\ P(X = x) &= P(\{\omega \in S : X(\omega) = x\}), \\ P(X \leq x) &= P(\{\omega \in S : X(\omega) \leq x\}). \end{aligned}$$

Usually one can obtain the correct interpretation of such notations by simply reading off the formula aloud, while translating each symbol into words. For example, if X denotes the number of heads in a series of coin tosses, then “ $P(X \leq 2)$ ” translates into “the probability that the number of heads is less than or equal to two.”

Probability mass function (p.m.f.)

- **Definition and notation:** $p(x) = P(X = x)$. (Here $P(X = x)$ is to be interpreted as above. Note that $p(x) = P(X = x) = 0$ if x is *not* among the values x_i of X , so in formulas involving $p(x)$ only values of the form x_i contribute.)
- **Properties:** (1) $0 \leq p(x) \leq 1$ for all x , (2) $\sum_i p(x_i) = 1$.

Cumulative distribution function (c.d.f.)

- **Definition and notation:** $F(x) = P(X \leq x)$ (with $P(X \leq x)$ interpreted as above).
- **Properties:** (1) $F(x)$ is non-decreasing and $0 \leq F(x) \leq 1$ for all x , (2) $\lim_{x \rightarrow -\infty} F(x) = 0$, (3) $\lim_{x \rightarrow \infty} F(x) = 1$.

¹More precisely, a so-called “measurable” function. (You can ignore this issue.)

Expectation (expected value, mean)

- **Definition and notation:** $\mu = E(X) = \sum_i x_i p(x_i)$
- **Properties:** $E(c) = c$ (where c denotes a constant random variable), $E(cX) = cE(X)$, $E(X + Y) = E(X) + E(Y)$
- **Expectation of a function of X :** $E(g(X)) = \sum_i g(x_i)p(x_i)$

Variance

- **Definition and notation:** $\sigma^2 = \text{Var}(X) = E(X^2) - E(X)^2 = E((X - \mu)^2)$ (where $\mu = E(X)$)
- **Properties:** $\text{Var}(c) = 0$, $\text{Var}(cX) = c^2 \text{Var}(X)$, $\text{Var}(X + c) = \text{Var}(X)$.
- **Standard deviation:** $\sigma = \sqrt{\text{Var}(X)}$

Notes and tips

- **Always specify the set of values x of a p.m.f.:** A formula alone, such as $p(x) = p(1-p)^x$, is useless without knowing the set of “legal” values x for this function. You can specify the values, by either listing all of them explicitly (e.g., “ $x = 2, 3, 5, 7, 11$ ”), or with notational shortcuts (e.g., “ $p(i) = p(1-p)^i$, $i = 1, 2, \dots$ ”), if there is a clear pattern in these values.
- **Distinguish between an upper case letter X , denoting an abstract random variable (which, mathematically, is a *function*), and lower case x , denoting the values of such a random variable (which are *real numbers*):** For example, in the notation $E(X)$, the “ X ” refers to the abstract random variable X , and thus should be an upper case letter. By contrast, in the notations $p(x)$ and $F(x)$, the x refers to real numbers, and so needs to be a lower case letter.
- **In contrast to random variables, whose domains are abstract, and possibly very complicated, sample spaces, probability mass functions $p(x)$ and distribution functions $F(x)$ are ordinary functions from the reals to the reals:** This is the reason why these functions are such a useful tool: A p.m.f. (or a c.d.f.) encodes all the *essential* properties of a random variable, but “lives” in a familiar setting, that of one-variable functions of the type one encounters in calculus. One can manipulate these functions just like any ordinary one-variable function, e.g., integrate, differentiate, find maxima, minima, etc.
- **Expectation, variance, etc., are just ordinary nonrandom numbers, not functions:** It is important to keep in mind that $E(X)$, $\text{Var}(X)$, etc., are just ordinary *numbers* associated with a random variable X . Taking the expectation of a random variable X completely “removes” the randomness from X and produces an ordinary number. Despite the function-like notation, $E(X)$ and $\text{Var}(X)$ are not functions in the usual (calculus) sense. (By the same token, the “random variable” X here, despite its name, is not a variable in the usual (calculus) sense, but a function.)
- **Integral rule of thumb for expectations:** An easy way to visualize the meaning of expectation, and also to remember its properties, is with this rule of thumb: Think of a random variable as being a function on a unit interval (for example, the stock price charts that you see in the Wallstreet Journal). Then its expectation represents the integral over this function. This rule of thumb is analogous to the “area rule of thumb” for probabilities.