

## Discrete Random Variables, I

### Terminology

Informally, a **random variable** is a quantity  $X$  whose value depends on some random event. The **space (or range) of  $X$**  is the set  $S$  of possible values of  $X$ . If this set  $S$  is finite or countable (i.e., can be listed as a sequence  $x_1, x_2, \dots$ ), the random variable is called **discrete**.

### General formulas

- **Probability mass function (p.m.f.) (also called a “discrete density function”, or, somewhat less precisely, a “discrete distribution”):**

- **Definition and notation:**  $f(x) = P(X = x)$  for all  $x \in S$ .

- **Properties:** (1)  $f(x) \geq 0$ , (2)  $\sum_{x \in S} f(x) = 1$

- **Uniform distribution on a set  $S$ :** Each of the values  $x \in S$  has the same probability, i.e.,  $f(x) = 1/n$  for each value  $x$ , where  $n$  is the number of values.

- **Expectation (mean):**

- **Definition and notation:**  $\mu = E(X) = \sum_{x \in S} xf(x)$

- **Properties:**  $E(c) = c$ ,  $E(cX) = cE(X)$ ,  $E(X + Y) = E(X) + E(Y)$

- **Expectation of a function of  $X$ :**  $E(u(X)) = \sum_{x \in S} u(x)f(x)$

- **Variance:**

- **Definition and notation:**  $\sigma^2 = \text{Var}(X) = E(X^2) - E(X)^2$

- **Alternate definition:**  $\text{Var}(X) = E((X - \mu)^2)$

- **Properties:**  $\text{Var}(c) = 0$ ,  $\text{Var}(cX) = c^2 \text{Var}(X)$ ,  $\text{Var}(X + c) = \text{Var}(X)$ .

- **Standard deviation:**  $\sigma = \sqrt{\text{Var}(X)}$

- **Moment-generating function:**

- **Definition and notation:**  $M(t) = E(e^{tX}) = \sum_{x \in S} e^{tx} f(x)$

- **Properties:** The derivatives of  $M(t)$  at 0 generate the “moments” of  $X$ :  $M'(0) = E(X)$ ,  $M''(0) = E(X^2)$ ,  $M'''(0) = E(X^3)$ , etc.

## Notes and tips

- Always specify the set of values  $x$  of a p.m.f.:** A formula alone, such as  $f(x) = p(1-p)^x$ , 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., “ $x = 0, 1, \dots, n-1, n$ ”, or “ $n = 1, 2, 3, \dots$ ”) if there is a clear pattern in these values. In the case of standard distributions (e.g., the geometric distribution given by the formula  $f(x) = p(1-p)^{x-1}$ ), make sure to memorize the exact range of values  $x$ , along with the formula for  $f(x)$ , and write it down whenever you use the formula. This helps avoid making mistakes, such as having a summation run from 1 to infinity instead of 0 to infinity, or vice versa, down the line.
- Distinguish between capital  $X$ , denoting an abstract random variable, and lower case  $x$ , denoting the values of such a random variable:** In a formula like  $E(X) = \sum_{x \in S} xf(x)$ , the  $X$  on the left refers to the abstract random variable  $X$  (an expectation is associated with a *random variable*), whereas the  $x$  on the right refers to the values of this random variable (the p.m.f.  $f(x)$  is a function of the *values* of a random variable). **A notation like  $E(x)$  or  $f(X)$  is mathematically nonsensical, you may get penalized in exams for using such notation!**
- 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.
- Integral rule of thumb for expectations:** An easy way to visualize the meaning of expectation, and 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.
- Short-hand notations:** Notations such as “ $P(X = 2)$ ”, “ $P(X \geq 2)$ ”, “ $P(|X - 2| \leq 1)$ ”, 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 event that the random variable  $X$  is equal to 2”. In the latter form the notation makes perfect sense, since the event  $A$  defined there corresponds to a subset of the sample space and thus has a well-defined associated probability  $P(A)$ .

An easy way to obtain the correct interpretation of such notations is to simply read 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 “ $P(X \geq 2)$ ” translates into “the probability that the number of heads is greater or equal to two.”

To compute such probabilities, simply add up corresponding values of the p.m.f. For example, if the values of  $X$  are  $0, 1, \dots, 4$ , then the event “ $|X - 2| \leq 1$ ” corresponds to those values in the above list that satisfy the given restriction, namely  $x = 1, 2, 3$ . Thus,  $P(|X - 2| \leq 1) = f(1) + f(2) + f(3)$ .