3.6 Cumulative distribution functions

Unit 3: Discrete Random Variables

Adapted from Blitzstein-Hwang Chapter 3.

Another function that describes the distribution of an r.v. is the *cumulative distribution function* (CDF). Unlike the PMF, which only discrete r.v.s possess, the CDF is defined for *all* r.v.s.

> DEFINITION 3.6.1.
>
> The *cumulative distribution function* (CDF) of an r.v. X is the function F_X given by $F_X(x) = P(X \leq x)$. When there is no risk of ambiguity, we sometimes drop the subscript and just write F (or some other letter) for a CDF.

The next example demonstrates that for discrete r.v.s, we can freely convert between CDF and PMF.

Example 3.6.2.

Let $X \sim \text{Bin}(4, 1/2)$. Figure 3.6.3 shows the PMF and CDF of X.

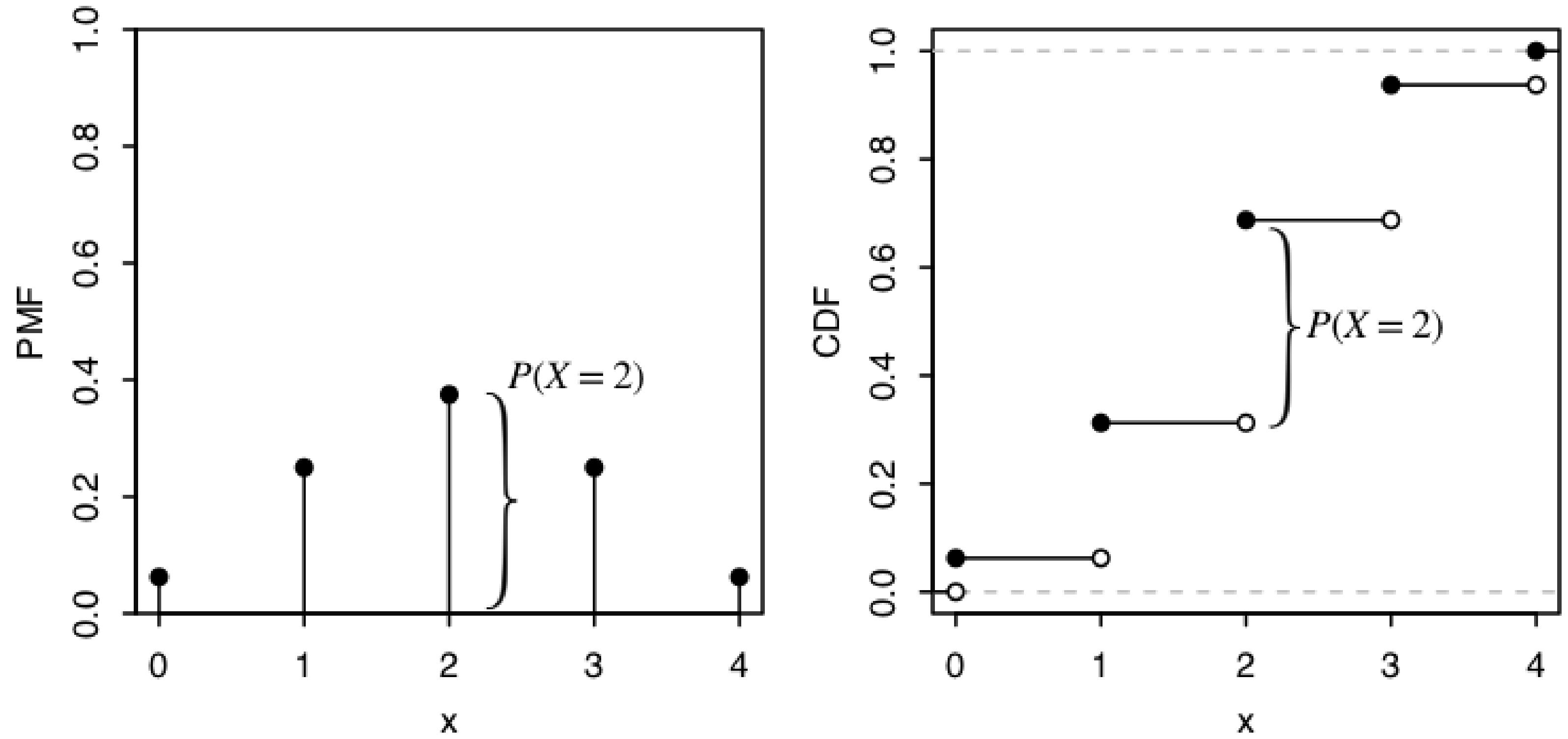

Figure 3.6.3: $\mathrm{Bin}(4, 1/2)$ PMF and CDF. The height of the vertical bar $P(X = 2)$ in the PMF is also the height of the jump in the CDF at 2.

View Larger Image

Image Description

- *From PMF to CDF*: To find $P(X \le 1.5)$, which is the CDF evaluated at 1.5, we sum the PMF over all values of the support that are less than or equal to 1.5:

  $$P(X \le 1.5) = P(X = 0) + P(X = 1) = \left(\frac{1}{2}\right)^4 + 4\left(\frac{1}{2}\right)^4 = \frac{5}{16}.$$

  Similarly, the value of the CDF at an arbitrary point x is the sum of the heights of the vertical bars of the PMF at values less than or equal to x.

- *From CDF to PMF*: The CDF of a discrete r.v. consists of jumps and flat regions. The height of a jump in the CDF at x is equal to the value of the PMF at x. For example, in Figure 3.6.3, the height of the jump in the CDF at 2 is the same as the height of the corresponding vertical bar in the PMF; this is indicated in the figure with curly braces. The flat regions of the CDF correspond to values outside the support of X, so the PMF is equal to 0 in those regions.

Valid CDFs satisfy the following criteria.

Theorem 3.6.4 (Valid CDFs).

Any CDF F has the following properties.

- Increasing: If $x_1 \le x_2$, then $F(x_1) \le F(x_2)$.

- Right-continuous: As in Figure 3.6.3, the CDF is continuous except possibly for having some jumps. Wherever there is a jump, the CDF is continuous from the right. That is, for any a, we have

$$F(a) = \lim_{x \to a^+} F(x).$$

- Convergence to 0 and 1 in the limits:

$$\lim_{x \to -\infty} F(x) = 0 \text{ and } \lim_{x \to \infty} F(x) = 1.$$

The converse is true too: it turns out that given any function F meeting these criteria, we can construct a random variable whose CDF is F. To recap, we have now seen three equivalent ways of expressing the distribution of a random variable. Two of these are the PMF and the CDF: we know these two functions contain the same information, since we can always figure out the CDF from the PMF and vice versa. Generally the PMF is easier to work with for discrete r.v.s, since evaluating the CDF requires a summation. A third way to describe a distribution is with a story that explains (in a precise way) how the distribution can arise. We used the stories of the Binomial and Hypergeometric distributions to derive the corresponding PMFs. Thus the story and the PMF also contain the same information, though we can often achieve more intuitive proofs with the story than with PMF calculations.