Chapter 8

Distributions of Random Variables: pmf and cdf

Lecture 8 | Devore 9e: 3.1–3.2

Last time we defined a random variable $X : S \to \mathbb{R}$. Today: how to describe one completely, without ever mentioning S again.

Definition 8.1 (Discrete and continuous)**.** If X takes values in a finite or countable set $x_1, x_2, x_3, \ldots,$ it is a **discrete** random variable. If it takes values filling an interval such as $[a, b]$ or (a, ∞), it is a **continuous** random variable.

Discrete this week; continuous starting in Lecture 13.

8.1 The probability mass function

Definition 8.2 (pmf)**.** The **probability mass function** of a discrete random variable X is

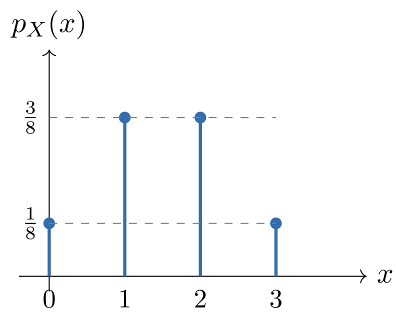
$$p_X(x) := \mathbb{P}\{X = x\} = \mathbb{P}\big(\{\omega \in S : X(\omega) = x\}\big).$$

Example 8.3 (Toss a coin three times)**.** $S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$ and $X = \#$heads, so X takes values $0, 1, 2, 3$. Pull back each value to an event:

$$\begin{aligned} &\{X = 0\} = \{TTT\}, && \{X = 1\} = \{HTT, THT, TTH\}, \\ &\{X = 2\} = \{THH, HTH, HHT\}, && \{X = 3\} = \{HHH\}. \end{aligned}$$

All eight outcomes are equally likely, so

x	0	1	2	3
$p_X(x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$



Theorem 8.4 (Characterization of a pmf)**.** *A function p is the pmf of some discrete random variable if and only if*

$$p(x_i) \geq 0 \quad \textit{for all } i, \qquad \sum_i p(x_i) = 1.$$

This is what makes "find the constant c so that this is a pmf" a well-posed question: the constant is whatever makes the sum equal 1.

Example 8.5 (Shooting at a target — the geometric pmf)**.** Each shot hits with probability $\beta \in (0, 1)$, independently, and you repeat until you hit. Writing H for a hit and M for a miss, the experiment stops at the first H, so the sample space is

$$S = \{H,\ MH,\ MMH,\ MMMH,\ \dots\},$$

which is countably infinite. Let X be the number of shots fired, so $X \in \{1, 2, 3, \dots\}$. Each value pulls back to a single outcome,

$$\{X = 1\} = \{H\}, \qquad \{X = 2\} = \{MH\}, \qquad \{X = 3\} = \{MMH\},\ \dots$$

and by independence

$$\mathbb{P}\{X = 1\} = \beta, \quad \mathbb{P}\{X = 2\} = (1 - \beta)\beta, \quad \mathbb{P}\{X = 3\} = (1 - \beta)^2\beta,\ \dots$$

so

$$\boxed{p_X(n) = (1 - \beta)^{n-1}\beta, \qquad n = 1, 2, 3, \dots}$$

Check: $\sum_{n \geq 1}(1 - \beta)^{n-1}\beta = \beta \cdot \frac{1}{1-(1-\beta)} = 1.$ ✓

This is the **geometric distribution**; we return to it in Lecture 10.

Example 8.6 (Indicator random variable)**.** For an event $A \in \mathcal{F}$ define

$$\mathbf{1}_{\{A\}}(\omega) = \mathbf{1}_A(\omega) := \begin{cases} 1, & \omega \in A, \\ 0, & \omega \notin A. \end{cases}$$

Then $\mathbb{P}\{\mathbf{1}_A = 1\} = \mathbb{P}(A)$ and $\mathbb{P}\{\mathbf{1}_A = 0\} = 1 - \mathbb{P}(A) = \mathbb{P}(A^c)$:

x	0	1
$p(x)$	$\mathbb{P}(A^c)$	$\mathbb{P}(A)$

The indicator is the bridge between events and random variables: every event becomes a $\{0, 1\}$-valued random variable, and we will use this repeatedly.

Remark 8.7. Once you have the pmf you can compute the probability of *any* event about X by adding up. For the shooting example with only 10 bullets available,

$$\mathbb{P}\{X \le 10\} = \mathbb{P}\Big(\bigcup_{k=1}^{10} \{X = k\}\Big) = \sum_{k=1}^{10} (1-\beta)^{k-1}\beta = 1 - (1-\beta)^{10},$$

the events $\{X = k\}$ being disjoint.

8.1.1 The recipe

Every pmf computation in this course follows the same five steps. When a problem stalls, it is almost always because one of them was skipped.

1. Identify the **experiment**: its outcomes and its sample space S.
2. State the **rule** $X : S \to \mathbb{R}$ — what number is being recorded.
3. List the **values** X can take.
4. For each value x, find the **pre-image** $\{X = x\} = \{\omega \in S : X(\omega) = x\}$. This is an event, i.e. a subset of S.
5. Compute $\mathbb{P}$ of each of those events. The resulting table is the **pmf**.

Steps 1–3 are bookkeeping; step 4 is where the thinking happens; step 5 is the probability from the first half of the course. In the three-coin example, step 4 was the list $\{X = 0\} = \{TTT\}$, $\{X = 1\} = \{HTT, THT, TTH\}$, and so on — and once that list was written down, step 5 was immediate.

8.2 The cumulative distribution function

The pmf works only for discrete variables. The cdf works for all of them, which is why it is the more fundamental object.

Definition 8.8 (cdf)**.** The **cumulative distribution function** of a random variable X is

$$F_X(x) := \mathbb{P}\{X \le x\}, \qquad x \in \mathbb{R}.$$

For a discrete X this is computed from the pmf by

$$F_X(x) = \sum_{y\,:\,y \le x} p_X(y).$$

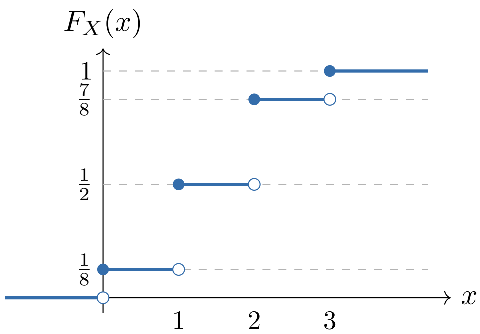
Example 8.9 (cdf of an indicator)**.** With $p(0) = \mathbb{P}(A^c)$ and $p(1) = \mathbb{P}(A)$:

- $x < 0$: no values of X are $\le x$, so $F(x) = 0$;
- $0 \le x < 1$: only the value 0 qualifies, so $F(x) = \mathbb{P}(A^c)$;
- $x \ge 1$: both qualify, so $F(x) = \mathbb{P}(A^c) + \mathbb{P}(A) = 1$.

$$F_{\mathbf{1}_A}(x) = \begin{cases} 0, & x < 0, \\ \mathbb{P}(A^c), & 0 \le x < 1, \\ 1, & x \ge 1. \end{cases}$$

Example 8.10 (cdf of #heads in three tosses)**.** From the pmf $\frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8}$:

$$F_X(x) = \begin{cases} 0, & x < 0, \\ \frac{1}{8}, & 0 \le x < 1, \\ \frac{1}{2}, & 1 \le x < 2, \\ \frac{7}{8}, & 2 \le x < 3, \\ 1, & x \ge 3. \end{cases}$$



The picture shows the two features that characterize every discrete cdf: it is a **step function**, and it is **right continuous** — at a jump the value is the upper dot, because $\{X \le x\}$ uses $\le$.

> **Theorem 8.11** (Classification of cdfs)**.** *Let X be a random variable on $(S, \mathcal{F}, \mathbb{P})$ and $F_X(x) = \mathbb{P}\{X \le x\}$. Then*
>
> *(i) F_X is non-decreasing: $a < b \Rightarrow F_X(a) \le F_X(b)$;*
>
> *(ii) $\lim_{x \to +\infty} F_X(x) = 1$ and $\lim_{x \to -\infty} F_X(x) = 0$;*
>
> *(iii) F_X is right continuous: $\lim_{x \to x_0^+} F_X(x) = F_X(x_0)$.*
>
> *Conversely, if a function F satisfies (i)–(iii), then there exist a probability space and a random variable X on it whose cdf is exactly F.*

Remark 8.12. The converse is what makes distributions first-class objects. From now on we can say "let X be geometric with parameter β" and never mention what S was. The underlying probability space disappears from the course at this point, and only distributions remain.

8.2.1 Reading probabilities off a cdf

For any $a < b$,

$$\mathbb{P}\{a < X \le b\} = F_X(b) - F_X(a), \qquad \mathbb{P}\{X = a\} = F_X(a) - F_X(a^-),$$

where $F_X(a^-) = \lim_{x \to a^-} F_X(x)$. So the jump of the cdf at a *is* the mass at a; where the cdf is continuous, $\mathbb{P}\{X = a\} = 0$. This is the identity that will let continuous random variables make sense in Lecture 13.