STA 131A: Introduction to Probability Theory

Lecture 13: Cumulative Distribution Functions

Dogyoon Song

Spring 2026, UC Davis

Announcements

Midterm 1 grades

- Exam and solutions are posted on the course webpage
- Please review the solutions carefully and identify what to revisit
- You can review graded exams in discussion section tomorrow

Homework 4 is posted (Due: Tue, May 5, 11:59 PM)

- Please submit on time and follow the submission instructions
- Please review the homework problems early so you have time to ask questions as needed

Mid-course survey

- Please take 10 minutes to complete the survey on Canvas by Friday, May 1
- All feedback and constructive suggestions/requests are welcome

Office hours today: 2:30–3:30 PM at MSB 4220

Remote lecture next Monday (May 4): More details will be announced on Canvas

Agenda

Last time: continuous random variables

- PDF f_X: density, not probability
- Probabilities as areas:

$$P(a \leq X \leq b) = \int_a^b f_X(x)\,dx$$

- Examples: Uniform and exponential random variables

Today: Cumulative distribution functions (CDFs)

- CDF as a unified description for discrete and continuous random variables
- CDF properties
- Recovering PMFs/PDFs from CDFs
- Example: Geometric and exponential CDFs

Recap: Continuous random variables

For a continuous random variable X, probabilities are described by a PDF:

$$P(X \in B) = \int_B f_X(x)\,dx.$$

A valid PDF satisfies

$$f_X(x) \ge 0, \qquad \int_{-\infty}^{\infty} f_X(x)\,dx = 1.$$

Important points:

- $f_X(x)$ is not a probability; it is a density.
- $P(X = x) = 0$ for every point x.
- Endpoint inclusion does not matter:

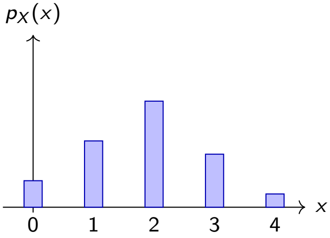
$$P(a \le X \le b) = P(a < X < b).$$

Motivation: Why another object?

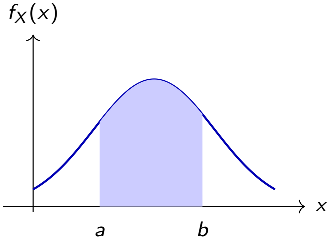
PMFs and PDFs look different:

$$P(X \in A) = \sum_{x \in A} p_X(x) \qquad \text{versus} \qquad P(X \in A) = \int_A f_X(x)\, dx.$$

Discrete



Continuous



Question: Is there one unified way to describe both types of random variables?

Cumulative distribution function

Definition (Cumulative distribution function)

For any random variable X, the **cumulative distribution function** or **CDF** is

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

Interpretation:

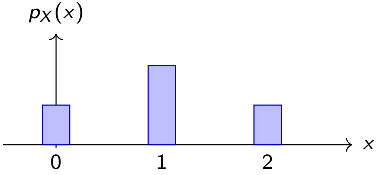
- $F_X(x)$ accumulates probability up to the point x.
- Its value always corresponds to a probability, hence $0 \leq F_X(x) \leq 1$.
- It describes the distribution of X completely.

Useful identity:

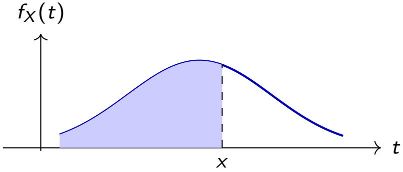
$$P(a < X \leq b) = F_X(b) - F_X(a).$$

Illustration: CDF as accumulated probability

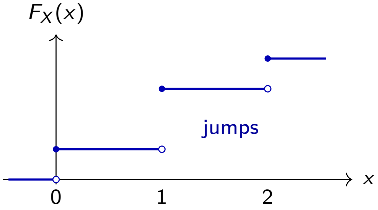
Discrete: PMF



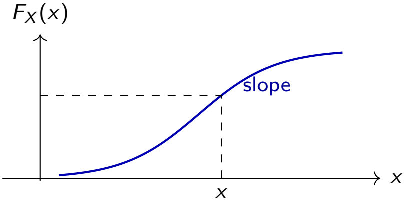
Continuous: PDF



Discrete: CDF



Continuous: CDF



Discrete mass $\leftrightarrow$ jumps in F_X.

Continuous density $\leftrightarrow$ slope of F_X.

CDF from a PMF: Discrete random variables

Example (Discrete random variable)

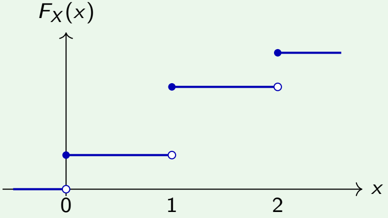
Let X have PMF

$$P(X=0)=\frac{1}{4},\qquad P(X=1)=\frac{1}{2},\qquad P(X=2)=\frac{1}{4}.$$

Then

$$F_X(x)=P(X\le x)=\sum_{x'\le x}p_X(x')$$

$$=\begin{cases}0, & x<0,\\ \frac{1}{4}, & 0\le x<1,\\ \frac{3}{4}, & 1\le x<2,\\ 1, & x\ge 2.\end{cases}$$



Interpretation:

- The CDF accumulates probability.
- The PMF is the jump in the CDF; large probability mass $\leftrightarrow$ larger jump.

CDF from a PDF: Continuous random variables

Example (Continuous random variable)

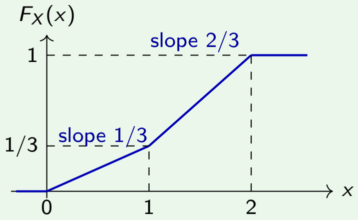
Let X have PDF

$$f_X(x) = \begin{cases} 1/3, & 0 \le x < 1, \\ 2/3, & 1 \le x < 2, \\ 0, & \text{otherwise.} \end{cases}$$

Then

$$F_X(x) = P(X \le x) = \int_{-\infty}^{x} f_X(t)\,dt$$

$$= \begin{cases} 0, & x < 0, \\ \frac{1}{3}x, & 0 \le x < 1, \\ \frac{2}{3}x - \frac{1}{3}, & 1 \le x < 2, \\ 1, & x \ge 2. \end{cases}$$



Interpretation:

- The PDF is the slope of the CDF; large density $\leftrightarrow$ rapidly increasing CDF.

Properties of CDFs

For any random variable X, its CDF $F_X(x) = P(X \leq x)$ satisfies:

- **Monotonicity:**
$$x \leq y \quad \Longrightarrow \quad F_X(x) \leq F_X(y).$$
- **Limits:**
$$\lim_{x \to -\infty} F_X(x) = 0, \qquad \lim_{x \to \infty} F_X(x) = 1.$$
- **Right-continuity:** $F_X(x) = \lim_{x' \to x+} F_X(x')$
- **Interval probabilities:**
$$P(a < X \leq b) = F_X(b) - F_X(a).$$
- **How the CDF behaves:**
  - Discrete RV: piecewise constant with jumps
  - Continuous RV: continuous, with slope given by the PDF when differentiable

Recovering PMFs and PDFs from CDFs

Discrete case: If X has possible values $x_1 < x_2 < \cdots$, then

$$P(X = x_i) = F_X(x_i) - F_X(x_i^-),$$

where $F_X(x_i^-) := \lim_{x' \to x_i -} F_X(x')$ denotes the value just before the jump.

Continuous case: If F_X is differentiable, then

$$f_X(x) = F_X'(x) = \frac{dF_X(x)}{dx}.$$

Upshot:

- Discrete probability mass appears as **jumps** in the CDF.
- Continuous density appears as the **slope** of the CDF.

Example: Deriving PDF from CDF (1/2)

Example (Maximum of two uniform points)

Let U_1, U_2 be independent $\mathrm{Uniform}(0, 1)$, and let

$$M = \max\{U_1, U_2\}.$$

Question: Find the CDF and the PDF of M.

For $0 \le m \le 1$,

$$\begin{aligned} F_M(m) &= P(M \le m) \\ &= P(U_1 \le m,\ U_2 \le m) \\ &= P(U_1 \le m)P(U_2 \le m) \\ &= m^2. \end{aligned}$$

Thus,

$$F_M(m) = \begin{cases} 0, & m < 0, \\ m^2, & 0 \le m \le 1, \\ 1, & m \ge 1. \end{cases}$$

Example: Deriving PDF from CDF (2/2)

Example (Maximum of two uniform points, continued)

Continuing:

$$F_M(m) = \begin{cases} 0, & m < 0, \\ m^2, & 0 \le m \le 1, \\ 1, & m \ge 1. \end{cases}$$

Since F_M is differentiable on $(0, 1)$,

$$f_M(m) = F_M'(m) = 2m, \qquad 0 < m < 1.$$

Therefore,

$$f_M(m) = \begin{cases} 2m, & 0 \le m \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

Takeaway: It is sometimes easier to compute the CDF first and then differentiate.

Pop-up quiz 1: Probability from CDF

Let X have CDF

$$F_X(x) = \begin{cases} 0, & x < 0, \\ x/2, & 0 \le x < 1, \\ 3/4, & 1 \le x < 2, \\ 1, & x \ge 2. \end{cases}$$

Which statement is correct?

A) $P(X = 1) = 0$

B) $P(X = 1) = 1/4$

C) $P(0 < X < 1) = 1$

D) F_X cannot be a CDF because it is flat on $[1, 2)$

Answer: B.
The jump at $x = 1$ is $F_X(1) - F_X(1^-) = 3/4 - 1/2 = 1/4$.

Follow-up: Is this random variable discrete, continuous, or neither? What feature of the CDF tells you that?

Pop-up quiz 2: Computing CDF

Let U_1, U_2 be independent $\mathrm{Uniform}(0, 1)$, and let

$$N = \min\{U_1, U_2\}.$$

For $0 \leq n \leq 1$, which expression equals $F_N(n) = P(N \leq n)$?

A) n^2

B) $1 - (1 - n)^2$

C) $1 - n^2$

D) $(1 - n)^2$

Answer: B.

It is easier to use the complement:

$$P(N \leq n) = 1 - P(N > n) = 1 - P(U_1 > n, U_2 > n) = 1 - (1 - n)^2.$$

Follow-up: How would you obtain the PDF of N from this CDF?

Geometric and exponential CDFs

Geometric

Let $N \sim \mathrm{Geometric}(p)$, where N is the trial number of the first success.

$$P(N > n) = \sum_{k=n+1}^{\infty} (1-p)^{k-1} p = (1-p)^n, \qquad n \in \mathbb{N}$$

So as a function on $\mathbb{R}$,

$$F_N(x) = \begin{cases} 0, & x < 1, \\ 1 - (1-p)^{\lfloor x \rfloor}, & x \geq 1. \end{cases}$$

Exponential

Let $X \sim \mathrm{Exponential}(\lambda)$, where X is the waiting time. Then

$$P(X > x) = \int_x^{\infty} \lambda e^{-\lambda x'}\, dx' = e^{-\lambda x}, \qquad x \geq 0.$$

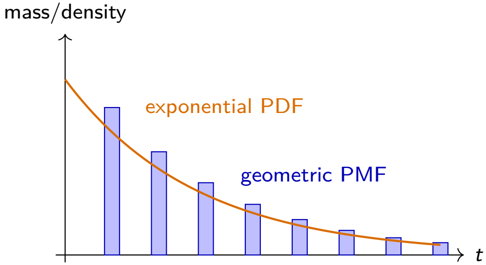
Thus,

$$F_X(x) = \begin{cases} 0, & x < 0, \\ 1 - e^{-\lambda x}, & x \geq 0. \end{cases}$$

Connection: Both random variables describe waiting times. The geometric model is discrete time; the exponential model is continuous time.

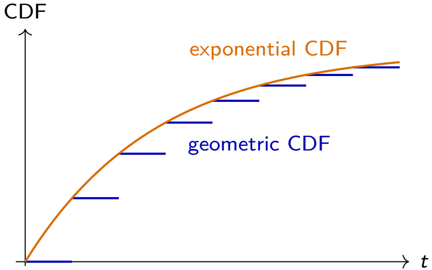
Illustration: Comparing geometric vs. exponential CDFs

PMF/PDF



Use $p = 0.3$ and $\lambda = -\log(1-p)$.

CDF



Step CDF vs. smooth CDF

With $\lambda = -\log(1-p)$, the two survival probabilities agree at integer times:

$$e^{-\lambda n} = (1-p)^n.$$

Reflection on Midterm 1: Problem 2

Theme: conditioning, total probability, and Bayes' rule

- Identify the hidden source/state first: $M_1,\ M_2,\ M_3$.
- Use total probability to compute the overall defect rate:

$$P(D) = \sum_i P(D \mid M_i)P(M_i).$$

- Use Bayes' rule for posterior probabilities:

$$P(M_3 \mid D) = \frac{P(D \mid M_3)P(M_3)}{P(D)}.$$

- For the test result, use the same logic:

$$P(D \mid T) = \frac{P(T \mid D)P(D)}{P(T \mid D)P(D) + P(T \mid D^c)P(D^c)}.$$

Takeaway: The main challenge is identifying the conditioning structure.

Related problems: Practice Midterm B, Problem 1

Reflection on Midterm 1: Problem 3

Theme: counting

- Without replacement:

$$P(X = k) = \frac{\binom{4}{k}\binom{6}{3-k}}{\binom{10}{3}}, \qquad k = 0, 1, 2, 3.$$

- Direct counting can often avoid unnecessary work:

$$P(\text{at least one red}) = 1 - \frac{\binom{6}{3}}{\binom{10}{3}}.$$

- With replacement (now we repeat 3 i.i.d. draws):

$$Z \sim \mathrm{Binomial}\left(3, \frac{2}{5}\right).$$

Takeaway: Always ask: replacement or no replacement? ordered or unordered?

Related problems: Practice Midterm B, Problem 2

Reflection on Midterm 1: Problem 4

Theme: joint, marginal, and conditional PMFs

- Marginals come from row and column sums:

$$p_X(x) = \sum_y p_{X,Y}(x,y), \qquad p_Y(y) = \sum_x p_{X,Y}(x,y).$$

- Conditional PMFs come from slicing and renormalizing:

$$p_{Y|X}(y \mid x) = \frac{p_{X,Y}(x,y)}{p_X(x)}.$$

- Conditional expectations are averages under the conditional PMF:

$$\mathbb{E}[Y \mid X = x] = \sum_y y\, p_{Y|X}(y \mid x).$$

- Independence requires factorization $p_{X,Y}(x,y) = p_X(x)p_Y(y)$ for all cells.

Takeaway: The joint PMF encodes marginals, conditioning, and dependence.

Related problems: Practice Midterm A/B, Problem 4

Wrap-up

The **cumulative distribution function (CDF** of random variable X is

$$F_X(x) = P(X \leq x).$$

- It gives a unified description of discrete and continuous random variables.

For any random variable X,

$$P(a < X \leq b) = F_X(b) - F_X(a).$$

- Discrete mass appears as jumps in F_X.
- Continuous density appears as slope: $f_X = F_X'$, when differentiable.

Two examples of waiting time model:

- Geometric random variable: $F_N(n) = 1 - (1-p)^n$.
- Exponential random variable: $F_X(x) = 1 - e^{-\lambda x}$.

Suggested reading: [BT08, Ch. 3.2]

References

Dimitri Bertsekas and John N Tsitsiklis.
Introduction to probability, volume 1.
Athena Scientific, 2nd edition, 2008.