STA 131A: Introduction to Probability Theory

Lecture 12: Continuous Random Variables

Dogyoon Song

Spring 2026, UC Davis

Announcements

Midterm 1 is being graded

- Exam and solutions are posted on the course webpage
- Once grading is complete, scores will be released through Canvas
- You will have a chance to review graded exams in discussion section this Thursday

Post-midterm review

- Homework 3 will include selected Midterm 1-style problems with modified numbers
- The TA will go over Midterm 1 problems in discussion section on Thursday

Mid-course survey

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

Agenda

Brief post-Midterm 1 recap

- Probability basics, conditioning, Bayes' rule, and independence
- Counting principle
- Discrete random variables & Joint, marginal, and conditional PMFs

Continuous random variables

- Motivation: why point probabilities no longer suffice
- Probability density functions
- Expectation and variance for continuous random variables

Review: Probability basics

A **probabilistic model** consists of a sample space Ω and a probability law P.

Set operations and identities

- Events are sets: use $\cup, \cap, {}^c$, and partitions carefully
- Complement, union and intersection:

$$P(A^c) = 1 - P(A), \qquad P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

- De Morgan's laws:

$$(A \cup B)^c = A^c \cap B^c, \qquad (A \cap B)^c = A^c \cup B^c.$$

Guiding principle: Translate words into events before computing probabilities

Review: Conditioning, Bayes, and independence

Conditional probability:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \qquad P(B) > 0.$$

Law of total probability and Bayes' rule: If $\{A_i\}$ is a partition, then

$$P(B) = \sum_i P(B \mid A_i)P(A_i), \qquad \text{(Law of total probability)}$$

and

$$P(A_j \mid B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(B \mid A_j)P(A_j)}{\sum_i P(B \mid A_i)P(A_i)}. \qquad \text{(Bayes' rule)}$$

Independence:

$$A \text{ and } B \text{ independent} \iff P(A \cap B) = P(A)P(B).$$

- Independence is not the same as disjointness.
- Pairwise independence is not the same as mutual independence.

Review: Counting outcomes

For a finite sample space with equally likely outcomes,

$$P(A) = \frac{|A|}{|\Omega|}.$$

Counting tools

- **Product rule:** break a process into stages and multiply the number of choices.
- **Permutations:** ordered selections,

$$\frac{n!}{(n-k)!} = n \times (n-1) \times \cdots \times (n-k+1).$$

- **Combinations:** unordered selections,

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

- **Multinomial coefficients:** labeled groups of sizes $n_1, \ldots, n_r$,

$$\binom{n}{n_1, \ldots, n_r} = \frac{n!}{n_1! \cdots n_r!}.$$

Review: Discrete random variables and PMFs

A **random variable** is a function

$$X : \Omega \to \mathbb{R}.$$

For a discrete random variable, the **PMF** is a function such that

$$p_X(x) = P(X = x).$$

- $p_X(x) \geq 0$, and $\sum_x p_X(x) = 1$.
- For a function $Y = g(X)$,

$$p_Y(y) = \sum_{x:g(x)=y} p_X(x).$$

- Common models:

$\mathrm{Bernoulli}(p)$:	one success/failure trial,
$\mathrm{Binomial}(n, p)$:	number of successes in n independent Bernoulli trials,
$\mathrm{Geometric}(p)$:	trial number of the first success,
$\mathrm{Poisson}(\lambda)$:	count in a fixed time/space window.

Note: probabilities of random-variable statements are indeed probabilities of events in Ω.

Review: Joint, marginal, and conditional PMFs

For discrete random variables X, Y, the **joint PMF** is

$$p_{X,Y}(x,y) = P(X = x, Y = y).$$

Marginal PMFs

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

Conditional PMF

$$p_{X|Y}(x \mid y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}, \qquad p_Y(y) > 0.$$

Law of total expectation

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

Independence of random variables

$$X, Y \text{ independent} \quad \Longleftrightarrow \quad p_{X,Y}(x,y) = p_X(x)p_Y(y) \quad \text{for all } x, y.$$

Continuous random variables: From sums to areas

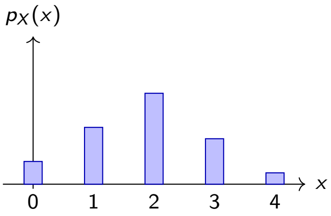
For a discrete random variable, probabilities are given as a sum of point masses:

$$P(X \in A) = \sum_{x \in A} p_X(x).$$

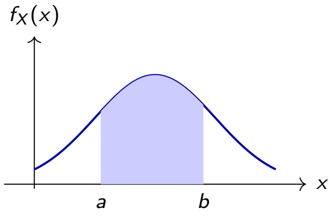
For a continuous random variable, probabilities are described by **areas under a density curve**:

$$P(X \in A) = \int_A f_X(x)\, dx.$$

Discrete



Continuous



Continuous random variable and PDF

Definition (Continuous random variable)

A random variable X is **continuous** if there exists a nonnegative function f_X, called the **probability density function** or **PDF**, such that[a]

$$P(X \in B) = \int_B f_X(x)\,dx \qquad \text{for every subset } B \subseteq \mathbb{R}.$$

[a] A rigorous formulation requires measure theory; for this course, you may think of B as an interval in $\mathbb{R}$.

A valid PDF satisfies

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

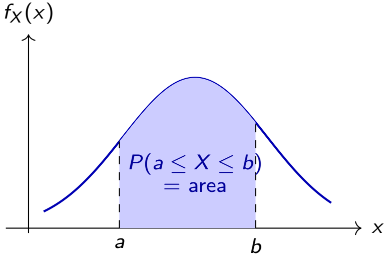
In particular, for any $a, b \in \mathbb{R}$ with $a \leq b$,

$$P(X = a) = \int_a^a f_X(x)\,dx = 0, \qquad P(a \leq X \leq b) = \int_a^b f_X(x)\,dx.$$

Visual interpretation of a PDF

A PDF assigns probability through **area**:

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



For a very small interval,

$$P(x \leq X \leq x + \Delta x) \approx f_X(x)\,\Delta x.$$

Thus, a larger density means more probability mass *near* that point, but $f_X(x)$ itself is not a probability.

Consequence: For a continuous random variable,

$$P(X = x) = 0$$

for every single point x, even though intervals can have positive probability.

Example: Uniform random variable

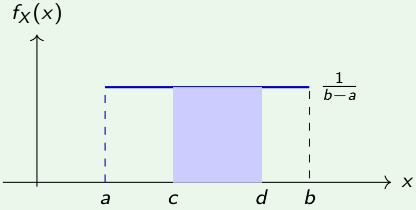
Example (Uniform random variable on an interval)

A random variable X is **uniform** on $[a, b]$, written

$$X \sim \mathrm{Uniform}(a, b),$$

if its PDF is

$$f_X(x) = \begin{cases} \dfrac{1}{b-a}, & a \le x \le b, \\ 0, & \text{otherwise.} \end{cases}$$



For any subinterval $[c, d] \subseteq [a, b]$,

$$P(c \le X \le d) = \int_c^d \frac{1}{b-a}\, dx = \frac{d-c}{b-a}.$$

So under a uniform model, probability is proportional to interval length.

Example: Piecewise-constant PDF (1/2)

Example (Piecewise constant PDF)

Let

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

Find c, and compute $P(0.5 \leq X \leq 1.5)$.
Because the total area must be 1,

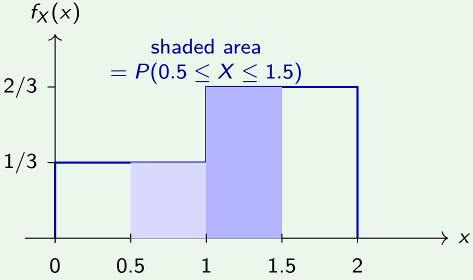
$$\int_{-\infty}^{\infty} f_X(x)\,dx = c(1-0) + 2c(2-1) = 3c = 1, \qquad \Longrightarrow \qquad c = \frac{1}{3}.$$

Then

$$P(0.5 \leq X \leq 1.5) = \int_{0.5}^{1} \frac{1}{3}\,dx + \int_{1}^{1.5} \frac{2}{3}\,dx = \frac{1}{6} + \frac{1}{3} = \frac{1}{2}.$$

Example: Piecewise-constant PDF (2/2)

Example (Piecewise constant PDF)



$$P(0.5 \leq X \leq 1.5) = \int_{0.5}^{1} \frac{1}{3}\,dx + \int_{1}^{1.5} \frac{2}{3}\,dx = \underbrace{0.5 \cdot \frac{1}{3}}_{1/6} + \underbrace{0.5 \cdot \frac{2}{3}}_{1/3} = \frac{1}{2}.$$

Expectation and variance: Continuous case

For a continuous random variable with PDF f_X, the **expectation** of X is

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\, dx.$$

For any function $g : \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx.$$

The **variance** of X is

$$\mathrm{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

Discrete vs continuous analogy

$$\sum_x \quad \longleftrightarrow \quad \int dx, \qquad p_X(x) \quad \longleftrightarrow \quad f_X(x).$$

Example: Mean and variance of a uniform random variable

Example (Mean and variance of a uniform random variable)

Let $X \sim \mathrm{Uniform}(a, b)$. Then

$$f_X(x) = \frac{1}{b-a}, \qquad a \le x \le b.$$

The mean is

$$\mathbb{E}[X] = \int_a^b x \frac{1}{b-a}\,dx = \frac{1}{b-a} \cdot \frac{x^2}{2}\Big|_a^b = \frac{1}{b-a} \cdot \frac{b^2 - a^2}{2} = \frac{a+b}{2}.$$

Also,

$$\mathbb{E}[X^2] = \int_a^b x^2 \frac{1}{b-a}\,dx = \frac{1}{b-a} \cdot \frac{x^3}{3}\Big|_a^b = \frac{1}{b-a} \cdot \frac{b^3 - a^3}{3} = \frac{a^2 + ab + b^2}{3}.$$

Therefore,

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{(b-a)^2}{12}.$$

Example: Exponential random variable

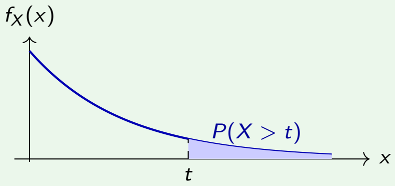
Example (Exponential random variable)

A random variable X has an **exponential distribution** with rate $\lambda > 0$, written

$$X \sim \mathrm{Exponential}(\lambda),$$

if

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



First check normalization:

$$\int_0^\infty \lambda e^{-\lambda x}\, dx = 1.$$

For $t \geq 0$,

$$P(X > t) = \int_t^\infty \lambda e^{-\lambda x}\, dx = e^{-\lambda t}.$$

Note: Exponential RVs are commonly used to model waiting times until the next event.

Example: Mean of an exponential random variable

Example (Mean of an exponential random variable)

If $X \sim \mathrm{Exponential}(\lambda)$, then

$$\mathbb{E}[X] = \int_0^\infty x\lambda e^{-\lambda x}\, dx.$$

Using integration by parts,

$$\int_0^\infty x\lambda e^{-\lambda x}\, dx = -xe^{-\lambda x}\Big|_0^\infty + \int_0^\infty e^{-\lambda x}\, dx = 0 + \frac{1}{\lambda} = \frac{1}{\lambda}.$$

Thus,

$$\mathbb{E}[X] = \frac{1}{\lambda}.$$

Example: If events occur at average rate $\lambda = 2$ per hour, then the average waiting time is

$$\frac{1}{2} \text{ hour} = 30 \text{ minutes}.$$

Wrap-up

Continuous random variables are described by **densities**, not point probabilities.

- The probability of an interval is an area:

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

- For continuous X, $P(X = x) = 0$ for each individual point x.

A valid **PDF** satisfies $f_X(x) \geq 0$ and

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

- A PDF value $f_X(x)$ is not a probability; probability comes from integrating the PDF.

Means and variances are the same as before, with sums replaced by integrals:

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\,dx.$$

Suggested reading: [BT08, Ch. 3.1]

References

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