

Lecture 5: Iterated Expectation, Conditional Variance, and Continuous Random Variables

EECS 126 (UC Berkeley)

Spring 2026 — February 3, 2026

Reading: Bertsekas & Tsitsiklis, §4.2–4.3; Ch. 3.

These are my personal notes from Professor Kannan Ramchandran's EECS 126 (Probability and Random Processes) course at UC Berkeley.

We open with a famous game whose expected payout is infinite. We then finish the story of conditional expectation from Lecture 4: the law of iterated expectation (the tower rule), which we use to analyze a random walk, and its companion for variances, the law of total variance. The second half of the lecture moves from discrete to *continuous* random variables, introducing probability density functions (PDFs) and cumulative distribution functions (CDFs).

5.1 The St. Petersburg Paradox

Example 5.1 (St. Petersburg game)

I toss a fair coin until I get a heads. If I toss n times, I give you $\$2^n$. For example,

$$H \mapsto \$2, \quad TH \mapsto \$4, \quad TTH \mapsto \$8, \quad \dots \quad (5.1)$$

Question: how much would you pay to play this game?

Let N be the number of tosses in the game and $W = 2^N$ the amount of money you win. The game ends on the first heads, so $N \sim \text{Geometric}(\frac{1}{2})$:

$$\mathbb{P}(N = n) = \underbrace{\frac{1}{2} \cdot \frac{1}{2} \cdots \frac{1}{2}}_{n \text{ flips}} = 2^{-n}, \quad n = 1, 2, 3, \dots \quad (5.2)$$

(The first $n - 1$ flips are tails and the n -th is heads; each has probability $\frac{1}{2}$.) By LOTUS,

$$\mathbb{E}[W] = \sum_{n=1}^{\infty} 2^n \cdot \mathbb{P}(N = n) = \sum_{n=1}^{\infty} 2^n \cdot \frac{1}{2^n} = \sum_{n=1}^{\infty} 1 = \infty. \quad (5.3)$$

The expected payout is infinite. This is the **St. Petersburg paradox**.

Why is it a paradox? If expected value were the whole story, you should be willing to pay *any* finite price to play. Yet almost nobody would pay more than a few dollars: half the time the game pays \$2, and $\mathbb{P}(W \leq \$8) = \frac{7}{8}$. The infinite mean is driven entirely by astronomically rare, astronomically large payouts (each outcome n contributes exactly \$1 to the mean). Two standard resolutions:

1. **Diminishing utility.** Daniel Bernoulli argued that the *value* of money grows more slowly than the amount, e.g. logarithmically. The expected log-payout is finite: $\mathbb{E}[\log_2 W] = \mathbb{E}[N] = 2$ (the geometric mean from Lecture 4), which corresponds to a “fair price” of about $2^2 = \$4$.

2. **Finite bankroll.** No casino can pay more than some $\$2^M$. If payouts are capped at 2^M , then

$$\mathbb{E}[W] = \sum_{n=1}^M 2^n 2^{-n} + \sum_{n=M+1}^{\infty} 2^M 2^{-n} = M + 2^M \cdot 2^{-M} = M + 1. \quad (5.4)$$

Even a billion-dollar bankroll ($2^{30} \approx 1.07 \times 10^9$) gives an expected payout of only \$31.

5.2 Conditional Expectation as a Random Variable

Recall from Lecture 4 that for each value y with $\mathbb{P}(Y = y) > 0$,

$$\mathbb{E}[X \mid Y = y] = \sum_x x \cdot \mathbb{P}(X = x \mid Y = y) \quad (5.5)$$

is a *number*. Letting $g(y) = \mathbb{E}[X \mid Y = y]$, the conditional expectation $\mathbb{E}[X \mid Y] \triangleq g(Y)$ is a function of Y , and therefore $\mathbb{E}[X \mid Y]$ is a *random variable*.

Example 5.2 (Tossing a fair coin a random number of times)

We toss a fair coin N times, where N is itself a random variable taking values in $\{1, 2, 3, \dots\}$. Let X be the total number of heads. Given $N = n$, the n tosses are independent fair Bernoulli trials, so $X \mid N = n \sim \text{Binomial}(n, \frac{1}{2})$.

$N = 1$: one toss gives 0 or 1 heads, each with probability $\frac{1}{2}$:

$$\mathbb{E}[X \mid N = 1] = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2} \quad (\text{a number}). \quad (5.6)$$

$N = 2$: two tosses give 0, 1, or 2 heads, with probabilities $\frac{1}{4}$ (TT), $\frac{1}{2}$ (HT or TH), and $\frac{1}{4}$ (HH):

$$\mathbb{E}[X \mid N = 2] = 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 1 \quad (\text{a number}). \quad (5.7)$$

General $N = n$: by the binomial mean from Lecture 4,

$$\mathbb{E}[X \mid N = n] = n \cdot \frac{1}{2} = \frac{n}{2} \quad (\text{a number}). \quad (5.8)$$

Sweeping over all values n and replacing n by the random variable N ,

$$\underbrace{\mathbb{E}[X \mid N]}_{\text{r.v.}} = \underbrace{\frac{N}{2}}_{\text{r.v.}}. \quad (5.9)$$

This is an **equality of random variables**: for every outcome of the experiment, the two sides take the same value.

The outcomes are not equally likely. For $N = 2$, averaging the possible head counts as $\frac{0+1+2}{3} = 1$ happens to give the right answer, but only by coincidence: the three outcomes have probabilities $\frac{1}{4}, \frac{1}{2}, \frac{1}{4}$, not $\frac{1}{3}$ each. For $N = 3$ the same shortcut would still give $\frac{3}{2}$ by symmetry, but for a *biased* coin it fails immediately. Always weight by the conditional PMF.

5.3 The Law of Iterated Expectation (Tower Property)

Since $\mathbb{E}[X | Y]$ is a random variable, we can take *its* expectation. The answer is remarkably simple.

Theorem 5.1 (Law of iterated expectation)

For random variables X, Y with $\mathbb{E}[|X|] < \infty$,

$$\mathbb{E}[\mathbb{E}[X | Y]] = \mathbb{E}[X]. \quad (5.10)$$

The inner expectation is over X (with Y held fixed); the outer expectation is over Y .

Proof. Write $\mathbb{E}[X | Y] = g(Y)$ with $g(y) = \mathbb{E}[X | Y = y]$. The outer expectation is the expectation of a function of Y , so by LOTUS (summing over the values y with $\mathbb{P}(Y = y) > 0$):

$$\begin{aligned} \mathbb{E}[\mathbb{E}[X | Y]] &= \sum_y \mathbb{E}[X | Y = y] \cdot \mathbb{P}(Y = y) && \text{(outer expectation, LOTUS)} \\ &= \sum_y \sum_x x \cdot \mathbb{P}(X = x | Y = y) \cdot \mathbb{P}(Y = y) && \text{(definition of } \mathbb{E}[X | Y = y]) \\ &= \sum_y \sum_x x \cdot \mathbb{P}(X = x, Y = y) && \text{(definition of conditional probability)} \\ &= \sum_x x \sum_y \mathbb{P}(X = x, Y = y) && \text{(swap the order of summation)} \\ &= \sum_x x \cdot \mathbb{P}(X = x) && \text{(marginalize out } Y) \\ &= \mathbb{E}[X]. && (5.11) \end{aligned}$$

The third line uses $\mathbb{P}(X = x | Y = y) \mathbb{P}(Y = y) = \mathbb{P}(X = x, Y = y)$, i.e. $p_{X|Y}(x | y) p_Y(y) = p_{X,Y}(x, y)$. Swapping the sums is justified by $\mathbb{E}[|X|] < \infty$ (absolute convergence). $\square$

Intuition: an average of averages. $\mathbb{E}[X | Y = y]$ is the average of X within the “group” where $Y = y$. The tower rule says: average these group averages, weighting each group by how likely it is, and you recover the overall average. The total expectation theorem from Lecture 4, $\mathbb{E}[X] = \sum_i \mathbb{P}(A_i) \mathbb{E}[X | A_i]$, is the special case where Y records which event A_i of a partition occurred.

A useful companion result lets us pull functions of Y through the inner expectation: once we condition on Y , any function of Y is a known constant.

Theorem 5.2 (Taking out what is known)

For any function f with $\mathbb{E}[|f(Y) X|] < \infty$,

$$\mathbb{E}[f(Y) X | Y] = f(Y) \mathbb{E}[X | Y], \quad (5.12)$$

and consequently

$$\mathbb{E}[f(Y) X] = \mathbb{E}[f(Y) \mathbb{E}[X | Y]]. \quad (5.13)$$

Proof. Fix y with $\mathbb{P}(Y = y) > 0$. Given $Y = y$, the random variable $f(Y)$ equals the constant $f(y)$, so it factors out of the conditional expectation:

$$\begin{aligned}\mathbb{E}[f(Y)X \mid Y = y] &= \sum_x f(y)x \mathbb{P}(X = x \mid Y = y) \\ &= f(y) \sum_x x \mathbb{P}(X = x \mid Y = y) = f(y) \mathbb{E}[X \mid Y = y].\end{aligned}\quad (5.14)$$

Replacing y by Y gives $\mathbb{E}[f(Y)X \mid Y] = f(Y) \mathbb{E}[X \mid Y]$. Now apply the tower rule to the random variable $f(Y)X$:

$$\mathbb{E}[f(Y)X] = \mathbb{E}[\mathbb{E}[f(Y)X \mid Y]] = \mathbb{E}[f(Y) \mathbb{E}[X \mid Y]].\quad (5.15)$$

□

Where the brackets go. In (5.13), $f(Y)$ stays *inside* the outer expectation, multiplied by the random variable $\mathbb{E}[X \mid Y]$. It is *not* true in general that $\mathbb{E}[f(Y)X] = \mathbb{E}[f(Y)] \cdot \mathbb{E}[X \mid Y]$; the right-hand side of that would not even be a number. For instance, with $f(Y) = Y^2$: $\mathbb{E}[Y^2X] = \mathbb{E}[Y^2 \mathbb{E}[X \mid Y]]$.

Example 5.3 (Random number of tosses, with $N \sim \text{Geometric}(p)$)

Suppose in Example 5.2 that N is generated by a *first coin* with heads probability p , tossed until its first heads, so that $N \sim \text{Geometric}(p)$. We then toss a *second*, fair coin N times and let X be the number of heads. From (5.9), $\mathbb{E}[X \mid N] = \frac{N}{2} = \frac{1}{2} \cdot N$.

Mean. By the tower rule, linearity, and $\mathbb{E}[N] = \frac{1}{p}$ for $N \sim \text{Geometric}(p)$ (Lecture 4),

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid N]] = \mathbb{E}\left[\frac{N}{2}\right] = \frac{1}{2} \mathbb{E}[N] = \frac{1}{2} \cdot \frac{1}{p} = \frac{1}{2p}.\quad (5.16)$$

Sanity check: if $p = 1$, then $N = 1$ always, and one fair toss gives $\mathbb{E}[X] = \frac{1}{2}$. ✓

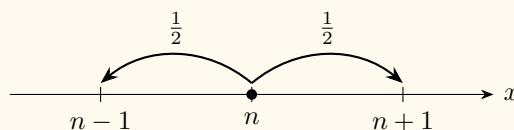
A mixed moment. Taking out what is known with $f(N) = N$, and using $\mathbb{E}[N^2] = \frac{2-p}{p^2}$ from Lecture 4,

$$\mathbb{E}[NX] = \mathbb{E}[N \mathbb{E}[X \mid N]] = \mathbb{E}\left[N \cdot \frac{N}{2}\right] = \frac{1}{2} \mathbb{E}[N^2] = \frac{2-p}{2p^2}.\quad (5.17)$$

5.4 The Symmetric Random Walk

Example 5.4 (Random walk on the integers)

A drunk man is walking along the integer line. He starts at the origin, $X_0 = 0$. At each step, independently of everything before, he moves one unit right with probability $\frac{1}{2}$ or one unit left with probability $\frac{1}{2}$. Let X_n be his location after n steps.



Questions: What is $\mathbb{E}[X_n]$? What is $\mathbb{E}[X_n^2]$, and how far from the origin is he typically?

Write the $(n+1)$ -th step using two indicator random variables,

$$\mathbf{1}_+^{(n+1)} = \begin{cases} 1 & \text{if step } n+1 \text{ is } +1, \\ 0 & \text{otherwise,} \end{cases} \quad \mathbf{1}_-^{(n+1)} = \begin{cases} 1 & \text{if step } n+1 \text{ is } -1, \\ 0 & \text{otherwise.} \end{cases} \quad (5.18)$$

Exactly one of them equals 1, so the step is $\mathbf{1}_+^{(n+1)} - \mathbf{1}_-^{(n+1)} \in \{+1, -1\}$, and

$$X_{n+1} = X_n + \mathbf{1}_+^{(n+1)} - \mathbf{1}_-^{(n+1)}. \quad (5.19)$$

Notation. An indicator only takes the values 0 and 1, so a step of -1 is written by *subtracting* the indicator $\mathbf{1}_-^{(n+1)}$. The superscript $(n+1)$ records that each step is a fresh, independent coin flip. Similarly, the location X_n and the step taken at time n are different random variables, so we keep separate names for them.

The mean: $\mathbb{E}[X_n] = 0$ for all n

Take expectations in (5.19). By linearity, and since $\mathbb{E}[\mathbf{1}_\pm^{(n+1)}] = \mathbb{P}(\text{step is } \pm 1) = \frac{1}{2}$,

$$\mathbb{E}[X_{n+1}] = \mathbb{E}[X_n] + \mathbb{E}[\mathbf{1}_+^{(n+1)}] - \mathbb{E}[\mathbf{1}_-^{(n+1)}] = \mathbb{E}[X_n] + \frac{1}{2} - \frac{1}{2} = \mathbb{E}[X_n]. \quad (5.20)$$

So one step never changes the expected location.

Theorem 5.3 (Mean of the symmetric random walk)

For every $n \geq 0$, $\mathbb{E}[X_n] = 0$.

Proof. By induction on n . **Base case:** $X_0 = 0$ is a constant, so $\mathbb{E}[X_0] = 0$. **Inductive step:** suppose $\mathbb{E}[X_n] = 0$. By (5.20), $\mathbb{E}[X_{n+1}] = \mathbb{E}[X_n] = 0$. Unrolled, this is just a chain of equalities:

$$\mathbb{E}[X_n] = \mathbb{E}[X_{n-1}] = \mathbb{E}[X_{n-2}] = \cdots = \mathbb{E}[X_1] = \mathbb{E}[X_0] = 0. \quad (5.21)$$

□

Two more ways to see it.

1. **Sum of steps.** Let $\xi_i \in \{+1, -1\}$ be the i -th step, so $X_n = \xi_1 + \xi_2 + \cdots + \xi_n$. Each step has mean $\mathbb{E}[\xi_i] = (+1) \cdot \frac{1}{2} + (-1) \cdot \frac{1}{2} = 0$, and by linearity $\mathbb{E}[X_n] = \sum_{i=1}^n \mathbb{E}[\xi_i] = 0$.
2. **Symmetry.** Flipping the direction of every step turns the walk X_n into $-X_n$, and since left and right are equally likely, the flipped walk has exactly the same distribution. So $\mathbb{E}[X_n] = \mathbb{E}[-X_n] = -\mathbb{E}[X_n]$, which forces $\mathbb{E}[X_n] = 0$.

A zero mean does *not* mean the walker stays near the origin. He drifts away, but is equally likely to drift left or right, so the drifts cancel in the average. To measure how far he wanders, we need the second moment.

The second moment: $\mathbb{E}[X_n^2] = n$

Suppose $X_n = k$. The next step is ± 1 with probability $\frac{1}{2}$ each, independent of the past, so

$$\mathbb{P}(X_{n+1}^2 = (k+1)^2 \mid X_n = k) = \frac{1}{2}, \quad \mathbb{P}(X_{n+1}^2 = (k-1)^2 \mid X_n = k) = \frac{1}{2}. \quad (5.22)$$

These are the same two equally likely moves ($k \rightarrow k+1$ and $k \rightarrow k-1$), just with the resulting location squared. Therefore

$$\begin{aligned} \mathbb{E}[X_{n+1}^2 \mid X_n = k] &= \frac{1}{2}(k+1)^2 + \frac{1}{2}(k-1)^2 \\ &= \frac{1}{2}(k^2 + 2k + 1) + \frac{1}{2}(k^2 - 2k + 1) \\ &= k^2 + 1. \end{aligned} \quad (5.23)$$

The edge case $k = 0$. When $k = 0$, $(k+1)^2 = (k-1)^2 = 1$, so the two events above are the *same* event, which has probability 1 (not $\frac{1}{2}$). The computation of $\mathbb{E}[X_{n+1}^2 \mid X_n = 0]$ is still correct, because we are really averaging over the two *moves*: $\frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 1 = 1 = 0^2 + 1$.

We found $g(k) = \mathbb{E}[X_{n+1}^2 \mid X_n = k] = k^2 + 1$. Replacing the value k by the random variable X_n (the “pattern matching” step, exactly as in $\mathbb{E}[X \mid N] = N/2$):

$$\mathbb{E}[X_{n+1}^2 \mid X_n] = X_n^2 + 1. \quad (5.24)$$

By the law of iterated expectation,

$$\begin{aligned} \mathbb{E}[X_{n+1}^2] &= \mathbb{E}[\mathbb{E}[X_{n+1}^2 \mid X_n]] && \text{(law of iterated expectation)} \\ &= \mathbb{E}[X_n^2 + 1] \\ &= 1 + \mathbb{E}[X_n^2]. && \text{(linearity)} \end{aligned} \quad (5.25)$$

Each step adds exactly 1 to the second moment, starting from $\mathbb{E}[X_0^2] = 0$:

$$\mathbb{E}[X_0^2] = 0, \quad \mathbb{E}[X_1^2] = 1, \quad \mathbb{E}[X_2^2] = 2, \quad \dots \implies \mathbb{E}[X_n^2] = n, \quad (5.26)$$

(formally, by induction exactly as for the mean). In particular $\mathbb{E}[X_{n+1}^2] = 1 + n$.

Theorem 5.4 (Spread of the symmetric random walk)

For every $n \geq 0$,

$$\text{Var}(X_n) = \mathbb{E}[X_n^2] - (\mathbb{E}[X_n])^2 = n - 0^2 = n, \quad \text{so} \quad \sigma_{X_n} = \sqrt{n}. \quad (5.27)$$

Cross-check with Lecture 4. With $X_n = \xi_1 + \dots + \xi_n$ and independent steps, each step has $\text{Var}(\xi_i) = \mathbb{E}[\xi_i^2] - 0^2 = 1$ (since $\xi_i^2 = 1$ always). The variance of a sum of independent random variables is the sum of the variances, so $\text{Var}(X_n) = n$. Same answer, no conditioning needed.

Key takeaway. The walker’s expected position is 0 forever, but his typical distance from the origin is $\sqrt{n}$: after 10,000 steps he is typically about 100 units away. This $\sqrt{n}$ growth is the signature of *diffusion*, and it reappears throughout the course.

5.5 Conditional Variance and the Law of Total Variance

Definition 5.1 (Conditional variance)

For y with $\mathbb{P}(Y = y) > 0$, the **conditional variance of X given $Y = y$** is the variance of the conditional distribution of X :

$$\text{Var}(X | Y = y) \triangleq \mathbb{E}[(X - \mathbb{E}[X | Y = y])^2 | Y = y] = \mathbb{E}[X^2 | Y = y] - (\mathbb{E}[X | Y = y])^2. \quad (5.28)$$

This is a number for each y , so $\text{Var}(X | Y)$, obtained by replacing y with Y , is a function of Y , and hence a **random variable**.

(The second form follows exactly as in Lecture 3, applied to the conditional distribution of X given $Y = y$.)

Theorem 5.5 (Law of total variance)

For random variables X, Y with $\mathbb{E}[X^2] < \infty$,

$$\text{Var}(X) = \mathbb{E}[\text{Var}(X | Y)] + \text{Var}(\mathbb{E}[X | Y]). \quad (5.29)$$

Proof. Abbreviate $\mu(Y) = \mathbb{E}[X | Y]$, so that $\text{Var}(X | Y) = \mathbb{E}[X^2 | Y] - \mu(Y)^2$.

Take expectations and use the tower rule on $\mathbb{E}[X^2 | Y]$:

$$\mathbb{E}[\text{Var}(X | Y)] = \mathbb{E}[\mathbb{E}[X^2 | Y]] - \mathbb{E}[\mu(Y)^2] = \mathbb{E}[X^2] - \mathbb{E}[\mu(Y)^2]. \quad (5.30)$$

Next, $\mu(Y)$ is a random variable; apply the variance formula to it, and use the tower rule on its mean, $\mathbb{E}[\mu(Y)] = \mathbb{E}[\mathbb{E}[X | Y]] = \mathbb{E}[X]$:

$$\text{Var}(\mu(Y)) = \mathbb{E}[\mu(Y)^2] - (\mathbb{E}[\mu(Y)])^2 = \mathbb{E}[\mu(Y)^2] - (\mathbb{E}[X])^2. \quad (5.31)$$

Adding the two, the $\mathbb{E}[\mu(Y)^2]$ terms cancel:

$$\mathbb{E}[\text{Var}(X | Y)] + \text{Var}(\mathbb{E}[X | Y]) = \mathbb{E}[X^2] - \mathbb{E}[\mu(Y)^2] + \mathbb{E}[\mu(Y)^2] - (\mathbb{E}[X])^2 = \text{Var}(X). \quad (5.32)$$

□

Interpretation: within-group plus between-group. Group the outcomes by the value of Y . Then $\mathbb{E}[\text{Var}(X | Y)]$ is the average spread *within* each group, and $\text{Var}(\mathbb{E}[X | Y])$ is the spread *between* the group means. Total spread is the sum of the two. Since both terms are nonnegative, $\text{Var}(X) \geq \mathbb{E}[\text{Var}(X | Y)]$: on average, learning Y can only reduce our uncertainty about X .

Example 5.5 (Variance of heads in a random number of tosses)

Continue the example with $N \sim \text{Geometric}(p)$ and $X | N \sim \text{Binomial}(N, \frac{1}{2})$. From the binomial formulas of Lecture 4,

$$\mathbb{E}[X | N] = \frac{N}{2}, \quad \text{Var}(X | N) = N \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{N}{4}. \quad (5.33)$$

Using $\mathbb{E}[N] = \frac{1}{p}$ and $\text{Var}(N) = \frac{1-p}{p^2}$:

$$\mathbb{E}[\text{Var}(X | N)] = \mathbb{E}\left[\frac{N}{4}\right] = \frac{1}{4} \cdot \frac{1}{p} = \frac{1}{4p}, \quad (5.34)$$

$$\text{Var}(\mathbb{E}[X | N]) = \text{Var}\left(\frac{N}{2}\right) = \frac{1}{4} \text{Var}(N) = \frac{1-p}{4p^2}. \quad (5.35)$$

By the law of total variance,

$$\text{Var}(X) = \frac{1}{4p} + \frac{1-p}{4p^2} = \frac{p}{4p^2} + \frac{1-p}{4p^2} = \frac{1}{4p^2}. \quad (5.36)$$

Sanity check: if $p = 1$, then $N = 1$ and $X \sim \text{Bernoulli}(\frac{1}{2})$, whose variance is $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$. ✓

5.6 Continuous Random Variables

Continuous random variables (CRVs) let us measure uncertainty in a principled way for quantities that vary continuously, such as velocity, time, and distance.

Coming up. Important continuous distributions include the *uniform*, *exponential*, and *gamma* distributions, which we develop in upcoming lectures. The exponential is the continuous analog of the geometric (Lecture 4).

Why discrete tools break. Let $X \sim \text{Uniform}[0, 1]$, a number chosen “uniformly at random” from $[0, 1]$. Then:

- $\mathbb{P}(X = x) = 0$ for any specific number x (there are uncountably many equally likely candidates);
- yet $\mathbb{P}(X < \frac{1}{2}) = \frac{1}{2}$.

So we cannot build probabilities by adding up point masses as we did with PMFs. Instead, we assign probabilities directly to *sets* that are “allowable events.”

Definition 5.2 (Allowable events (Borel sets))

The allowable events for a real-valued random variable include all intervals,

$$(-\infty, a), \quad [a, b], \quad (a, b], \quad \dots \quad (5.37)$$

together with everything obtainable from intervals by **countable unions**, **countable intersections**, and **complements**, e.g. $(a, b) \cup (c, d)$. These are the **Borel sets**; they form the Borel σ -algebra from Lecture 1. In practice, any “reasonable” set is allowable.

Recall that for a discrete random variable (DRV) and a set of values B ,

$$\mathbb{P}(X \in B) = \sum_{x \in B} p_X(x). \quad (5.38)$$

For a CRV, the sum over points becomes an integral of a density.

Definition 5.3 (Continuous random variable and PDF)

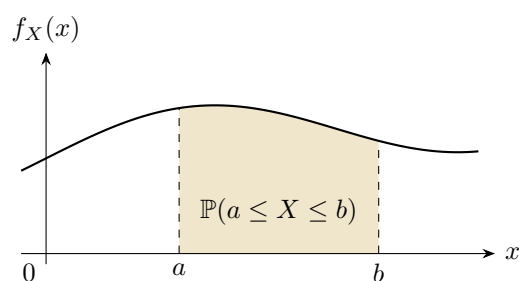
A random variable X is **continuous** if there is a function $f_X : \mathbb{R} \rightarrow [0, \infty)$, called its **probability density function (PDF)**, such that for every allowable set B ,

$$\mathbb{P}(X \in B) = \int_B f_X(x) dx. \quad (5.39)$$

A density must satisfy

$$f_X(x) \geq 0 \text{ for all } x, \quad \text{and} \quad \int_{-\infty}^{\infty} f_X(x) dx = 1. \quad (5.40)$$

In particular, $\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x) dx$ is the area under the density between a and b .

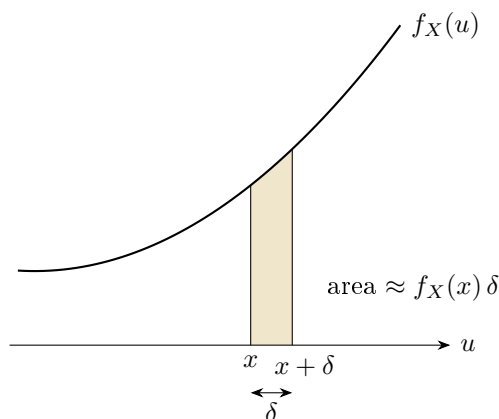
**Theorem 5.6 (Single points have probability zero)**

If X is continuous, then $\mathbb{P}(X = x) = 0$ for every $x \in \mathbb{R}$.

Proof. Fix x and any $\delta > 0$. The event $\{X = x\}$ is contained in $\{x \leq X < x + \delta\}$, so

$$\mathbb{P}(X = x) \leq \mathbb{P}(x \leq X < x + \delta) = \int_x^{x+\delta} f_X(u) du. \quad (5.41)$$

(We use the dummy variable u inside the integral so as not to confuse it with x , which now appears in the limits.) As $\delta \rightarrow 0$, this is the integral of f_X over a shrinking interval, which tends to 0. If f_X is continuous at x , we can see this explicitly: the sliver is approximately a rectangle of height $f_X(x)$ and width δ , so the integral is $\approx f_X(x) \delta \rightarrow 0$. Since $\mathbb{P}(X = x)$ is smaller than a quantity that can be made arbitrarily small, $\mathbb{P}(X = x) = 0$. $\square$



A density is not a probability. $f_X(x)$ has units of *probability per unit length*. Wherever f_X is continuous,

$$f_X(x) = \lim_{\delta \rightarrow 0} \frac{\mathbb{P}(x \leq X < x + \delta)}{\delta}, \quad \text{i.e.} \quad \mathbb{P}(x \leq X < x + \delta) \approx f_X(x) \delta \text{ for small } \delta. \quad (5.42)$$

Don't confuse $f_X(x)$ with a probability: it is a *density*. Only areas under f_X are probabilities, so $f_X(x)$ itself may exceed 1, as the next example shows.

Example 5.6 (An unbounded density)

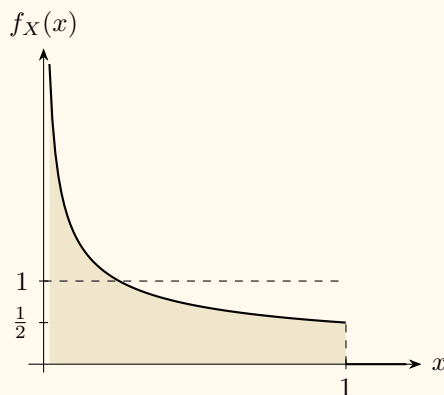
Let

$$f_X(x) = \begin{cases} \frac{1}{2\sqrt{x}} & 0 < x \leq 1, \\ 0 & \text{otherwise.} \end{cases} \quad (5.43)$$

It is a valid PDF. Clearly $f_X \geq 0$. For the total area, write $\frac{1}{2\sqrt{x}} = \frac{1}{2}x^{-1/2}$ and integrate:

$$\int_{-\infty}^{\infty} f_X(x) dx = \int_0^1 \frac{1}{2} x^{-1/2} dx = \frac{1}{2} \left[2x^{1/2} \right]_0^1 = \left[\sqrt{x} \right]_0^1 = 1 - 0 = 1. \quad (5.44)$$

It exceeds 1, and is even unbounded. $f_X(1) = \frac{1}{2}$, but $f_X(0.01) = \frac{1}{2(0.1)} = 5$, and $f_X(x) \rightarrow \infty$ as $x \rightarrow 0^+$. Nevertheless every probability is finite, e.g. $\mathbb{P}(X \leq 0.01) = \int_0^{0.01} \frac{1}{2} x^{-1/2} dx = \sqrt{0.01} = 0.1$.



5.7 The Cumulative Distribution Function

Definition 5.4 (Cumulative distribution function (CDF))

The **cumulative distribution function** of any random variable X (discrete or continuous) is

$$F_X(x) \triangleq \mathbb{P}(X \leq x), \quad x \in \mathbb{R}. \quad (5.45)$$

For a continuous random variable,

$$F_X(x) = \int_{-\infty}^x f_X(u) du. \quad (5.46)$$

Theorem 5.7 (Properties of the CDF)

Every CDF satisfies:

1. **Nondecreasing:** if $x \leq y$ then $F_X(x) \leq F_X(y)$, since $\{X \leq x\} \subseteq \{X \leq y\}$.
2. **Limits:** $\lim_{x \rightarrow -\infty} F_X(x) = 0$ and $\lim_{x \rightarrow +\infty} F_X(x) = 1$.
3. **Intervals:** $\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$, because $\{X \leq b\}$ is the disjoint union of $\{X \leq a\}$ and $\{a < X \leq b\}$.

Theorem 5.8 (Recovering the PMF or PDF from the CDF)

- **CRV:** wherever f_X is continuous,

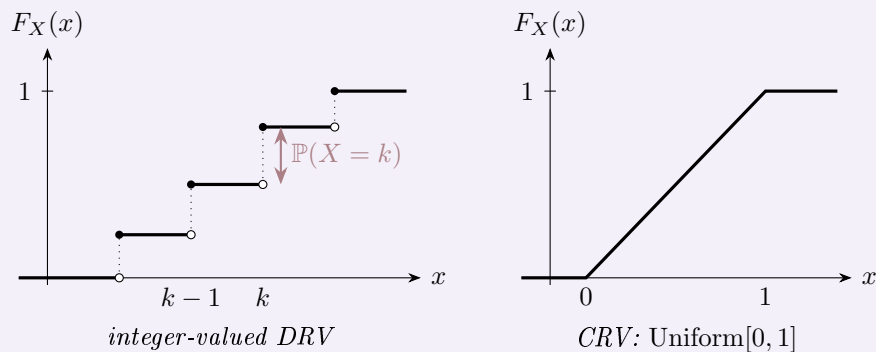
$$f_X(x) = \frac{dF_X(x)}{dx}. \quad (5.47)$$

- **Integer-valued DRV:** for every integer k ,

$$\mathbb{P}(X = k) = F_X(k) - F_X(k - 1). \quad (5.48)$$

Proof. **CRV:** $F_X(x) = \int_{-\infty}^x f_X(u) du$, so this is the fundamental theorem of calculus. **DRV:** because X only takes integer values, there are no possible values strictly between $k - 1$ and k . So $\{X \leq k\}$ is the disjoint union of $\{X \leq k - 1\}$ and $\{X = k\}$, giving $F_X(k) = F_X(k - 1) + \mathbb{P}(X = k)$. (For a discrete random variable taking non-integer values, use the jump of F_X at the value instead.) $\square$

Picturing the two kinds of CDF. For an integer-valued DRV, F_X is a *staircase*. On each open interval $(k - 1, k)$ there are no possible values, so no new probability accumulates and F_X stays flat, holding the value $F_X(k - 1)$. It does not drop back to zero: all the probability collected up through $k - 1$ is still counted. At k itself, F_X jumps up by exactly $\mathbb{P}(X = k)$. For a CRV, single points carry no probability, so F_X rises continuously with no jumps.



Corollary 5.1 (Consequences for continuous random variables)

If X is continuous, then for all $a \leq b$:

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x) dx = F_X(b) - F_X(a), \quad (5.49)$$

$$\mathbb{P}(X = a) = 0, \quad \mathbb{P}(X < a) = \mathbb{P}(X \leq a). \quad (5.50)$$

Because points do not contribute, it does not matter whether the endpoints of an interval are included: $\mathbb{P}(a < X < b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X \leq b)$. This is *false* for discrete random variables, where $\mathbb{P}(X < k)$ and $\mathbb{P}(X \leq k)$ differ by $\mathbb{P}(X = k)$.

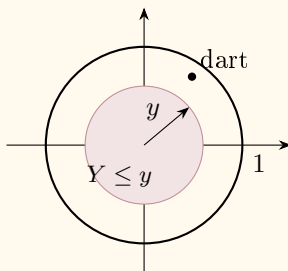
5.8 Example: A Random Dart on a Disk

Example 5.7 (Random dart on the unit disk)

A dart lands uniformly at random on a disk of radius 1 (the unit circle and its interior). “Uniformly” means the probability of landing in any region is proportional to its area:

$$\mathbb{P}(\text{dart lands in region } R) = \frac{\text{Area}(R)}{\text{Area of full disk}} = \frac{\text{Area}(R)}{\pi(1)^2}. \quad (5.51)$$

Let Y be the dart’s distance from the center. Find the PDF $f_Y(y)$.



It is easiest to compute $F_Y(y) = \mathbb{P}(Y \leq y)$ first. For $0 \leq y \leq 1$, the event $\{Y \leq y\}$ means the dart lands in the shaded disk of radius y :

$$F_Y(y) = \mathbb{P}(Y \leq y) = \frac{\text{Area shaded}}{\text{Area full disk}} = \frac{\pi y^2}{\pi(1)^2} = y^2. \quad (5.52)$$

Outside $[0, 1]$: $F_Y(y) = 0$ for $y < 0$ (distance is never negative) and $F_Y(y) = 1$ for $y > 1$ (the dart is always on the disk).

Differentiating,

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{d}{dy}(y^2) = 2y, \quad 0 \leq y \leq 1, \quad (5.53)$$

and $f_Y(y) = 0$ otherwise. Check: $\int_0^1 2y \, dy = [y^2]_0^1 = 1$. ✓

By the CDF,

$$\mathbb{P}(Y \in (0.5, 0.6)) = F_Y(0.6) - F_Y(0.5) = 0.6^2 - 0.5^2 = 0.36 - 0.25 = 0.11. \quad (5.54)$$

Or, equivalently, by the PDF: $\int_{0.5}^{0.6} 2y \, dy = [y^2]_{0.5}^{0.6} = 0.11$.

Three traps in this example.

1. **Y is not uniform on $[0, 1]$.** The dart’s *position* is uniform on the disk, but its *distance* from the center is not: the density $2y$ grows with y , because there is more area at larger radii. The ring $0.5 < Y < 0.6$ has probability 0.11, while the same-width band $0 < Y < 0.1$ near the center has only $0.1^2 = 0.01$.
2. **Don’t stop at the CDF.** y^2 is F_Y , not f_Y ; the PDF requires the extra differentiation step.
3. **$\frac{1}{\pi r^2}$ is a 2-D density.** The constant $\frac{1}{\pi r^2}$ (here $\frac{1}{\pi}$) is the density of the dart’s *position* per unit *area*. It is not f_Y , which is a density per unit *length* along the radius.

The CDF method. This “CDF first, then differentiate” strategy is the standard way to find the density of a function of a random variable.

5.9 Example: Meeting at Random Times

Not finished in lecture. We ran out of time after posing this problem. The full solution is given below.

Example 5.8 (Will A and B meet?)

A and B each arrive at a meeting spot at a uniformly random time between 4:00 pm and 5:00 pm, independently of each other. Whoever arrives first waits 10 minutes and then leaves. What is $\mathbb{P}(A \text{ and } B \text{ meet})$?

Setup. Measure time in minutes after 4:00 pm, so A and B are independent $\text{Uniform}[0, 60]$ random variables. The pair (A, B) is a point in the 60×60 square $[0, 60]^2$. Because A and B are independent and each uniform, the point (A, B) is uniform on the square:

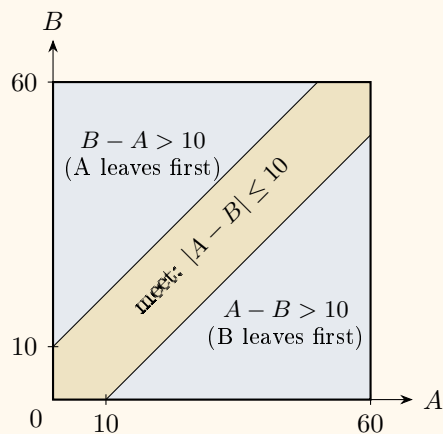
$$\mathbb{P}((A, B) \in R) = \frac{\text{Area}(R)}{60 \times 60} = \frac{\text{Area}(R)}{3600}. \quad (5.55)$$

(Joint distributions of several continuous random variables are developed formally in upcoming lectures; this “probability = area fraction” rule is the two-dimensional version of the uniform distribution, exactly as in the dart example.)

When do they meet? Whoever arrives first waits 10 minutes, so they meet exactly when their arrival times differ by at most 10 minutes:

$$\{A \text{ and } B \text{ meet}\} = \{|A - B| \leq 10\}. \quad (5.56)$$

This is the shaded diagonal band in the figure below.



Compute the complement. The two blue corner regions, where they miss each other, are right triangles:

- $\{A - B > 10\}$: A arrives more than 10 minutes after B , so B has already left. The triangle has vertices $(10, 0)$, $(60, 0)$, $(60, 50)$ and legs of length 50.
- $\{B - A > 10\}$: the mirror image, with vertices $(0, 10)$, $(0, 60)$, $(50, 60)$, also with legs of length 50.

Each triangle has area $\frac{1}{2} \cdot 50 \cdot 50 = 1250$, so together they cover 2500 out of 3600:

$$\mathbb{P}(\text{miss}) = \frac{2 \cdot 1250}{3600} = \frac{2500}{3600} = \frac{25}{36}. \quad (5.57)$$

Therefore

$$\mathbb{P}(A \text{ and } B \text{ meet}) = 1 - \frac{25}{36} = \frac{11}{36} \approx 0.306. \quad (5.58)$$

Equivalently, the band has area $3600 - 2500 = 1100$, and $\frac{1100}{3600} = \frac{11}{36}$.

Checks and generalization.

1. **Boundaries don't matter.** The lines $|A - B| = 10$ have zero area, so it makes no difference whether “waits 10 minutes” includes the endpoint. This is the two-dimensional version of $\mathbb{P}(X < a) = \mathbb{P}(X \leq a)$.
2. **General formula.** With a window of length T and waiting time $w \leq T$, the same picture gives

$$\mathbb{P}(\text{meet}) = 1 - \left(1 - \frac{w}{T}\right)^2. \quad (5.59)$$

With $w = 10$, $T = 60$: $1 - (5/6)^2 = 11/36$. Sanity checks: $w = 0$ gives 0 (they must arrive at the exact same instant), and $w = T$ gives 1.

End of Lecture 5.