CMPUT 267: Machine Learning I · Fall 2026

Random Variables and Distributions

Xi Ye

2026-09-08

Recall: House-Price Prediction

Size (m^2)	**Age (years)**	**Rooms**	**Price ($)**
90	25	3	310,000
140	8	4	510,000
115	40	3	355,000

- $\mathcal{D}$ is the **randomly sampled** dataset we observed.
- Because the dataset is **uncertain**, the predictor produced by the learner is **uncertain** as well.

> Probability gives us a mathematical way to describe uncertainty.

This Lecture: Probability

Today we will review:

1. Outcomes, events, and probability
2. Random variables and distributions
3. Joint and marginal probability
4. Features, datasets, and predictors as random variables

> ⚠ We introduce the probability concepts needed for this course, not in the most mathematically rigourous way.

Experiment, Outcome, and Outcome Space

Probability describes the likelihood of events. Three basic concepts:

- **Experiment**: a process that produces an uncertain result
- **Outcome**: the one result produced by the experiment
- **Outcome Space**: the set of every possible outcome

Example: roll a die

Experiment	Outcome	Outcome Space
Roll a six-sided die once	4	$\mathcal{X} = \{1, 2, 3, 4, 5, 6\}$

The die produces exactly one of the six possible outcomes.

Events

An **event** is a set of outcomes that we want to ask about. It is a subset of the outcome space:

$$E \subseteq \mathcal{X}.$$

The event occurs when the observed outcome belongs to E.

> An event can contain one outcome, several outcomes, all outcomes, or no outcomes.

Examples: Events

Example: roll a six-sided die, outcome space $\{1, 2, 3, 4, 5, 6\}$

Roll a Four

$A = \{4\}$

Roll an Odd Number

$B = \{1, 3, 5\}$

Roll More Than Three

$C = \{4, 5, 6\}$

Intersections mean **and**: $B \cap C = \{5\}$ (odd and more than 3).

Example: draw from a 52-card deck (each card is one outcome)

Draw a Heart

Event H

13 outcomes

Draw a Face Card

Event F

12 outcomes

Draw a Heart Face Card

$H \cap F = \{J\heartsuit, Q\heartsuit, K\heartsuit\}$

3 outcomes

Probability Distribution

A **probability distribution** is a function P that assigns a probability to every event E from the event space $\mathcal{E}$. For a discrete outcome space $\mathcal{X}$, we can take $\mathcal{E}$ to be the power set of $\mathcal{X}$.

$$P : \mathcal{E} \to [0, 1].$$

Domain $\mathcal{E}$: all possible events (subsets of outcomes)

Co-Domain: $[0, 1]$ scalar value

For a fair six-sided die,

$$P(\{1\}) = \frac{1}{6}, \qquad P(\{3, 4, 5\}) = \frac{1}{2}.$$

Properties of Probability Distribution

A probability distribution must satisfy **two properties**:

Entire outcome space. The probability of the entire outcome space $\mathcal{X}$ must be 1. (E.g., the probability that a die produces 1-6 is 1.)

$$P(\mathcal{X}) = 1.$$

Additivity for disjoint events. If A and B are disjoint events ($A \cap B = \varnothing$, the events cannot occur together) then,

$$P(A \cup B) = P(A) + P(B).$$

Are these events disjoint: A. roll a three; B. roll an even number?

Are these events disjoint: A. roll more than three; B. roll an even number?

Complements

Given the two properties, for any event E and its complement E^c (w.r.t. $\mathcal{X}$):

$$P(E) + P(E^c) = 1$$

For example, $P(\text{roll an odd}) + P(\text{roll an even}) = 1$.

Because $P(\mathcal{X}) = 1$, we also have $P(\varnothing) = 0$.

Quick Check: Events

Suppose every review is either positive or negative, and $P(\text{positive}) = 0.30$.

What is $P(\text{negative})$?

- **A** 0
- **B** 0.30
- **C** 0.70
- **D** Cannot be determined

C. “Negative” is the complement of “positive.”

Random Variables

We usually use a **random variable** to represent a value produced by an experiment.

- The value of **random variable** X is unknown before the experiment
- Outcome space $\mathcal{X}$ specifies the set of possible values of X
- The observed outcome is written with a lowercase letter, such as $X = x$

Example: roll a die

Random Variable X	Possible Values	Observed Value
The (unknown) result of a die roll	$X \in \{1, 2, 3, 4, 5, 6\}$	$X = x$, for example $X = 4$

Events Written with Random Variables

Introducing random variables makes it easier to express events. To write events with X:

- **Equality** specifies one value
- **Set membership** specifies a collection of values
- **An inequality** describes values satisfying a numerical condition

Example: a die-roll random variable X

Roll Exactly Four

$P(X = 4)$

Event $\{4\}$

Roll an Odd Number

$P(X \in \{1, 3, 5\})$

Event $\{1, 3, 5\}$

Roll More Than Four

$P(X > 4)$

Event $\{5, 6\}$

Discrete and Continuous Random Variables

Discrete

Values can be listed.

$$X \in \{0, 1\}$$

Examples: class label, word count, die roll

Continuous

Values fill an interval.

$$X \in [0, \infty)$$

Examples: duration, temperature, measurement

Describing a Discrete Distribution

Let $R \in \mathcal{R} = \{1, 2, 3, 4\}$ be the number of rooms in a randomly selected house.

To describe its distribution, assign a probability to **each possible value**:

Number of rooms r	1	2	3	4
$P(R = r)$	0.15	0.30	0.45	0.10

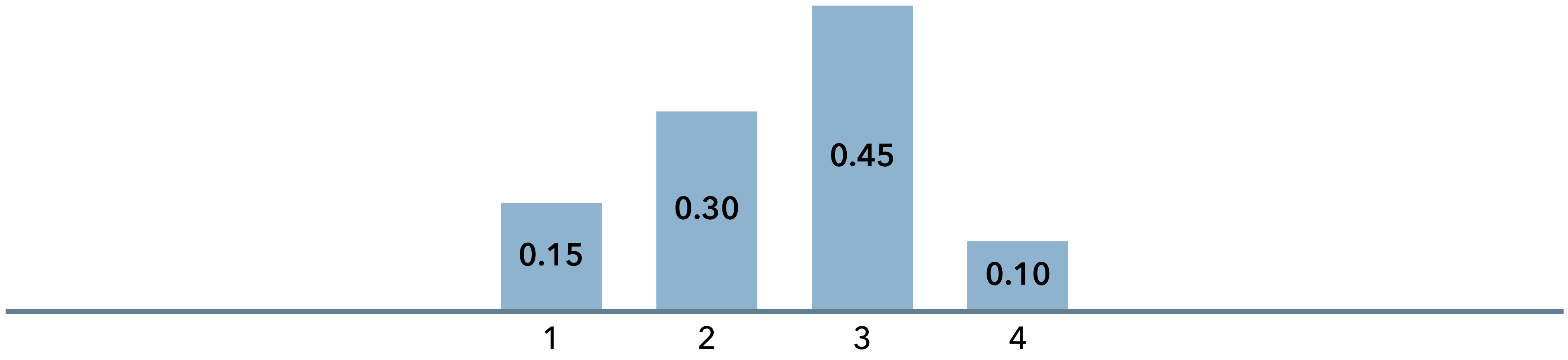
> For a discrete variable, the probabilities of individual values determine the probability of every event.

Probability Mass Function

The function that assigns probability to each possible value of a discrete random variable is called a **probability mass function**, or **PMF**.

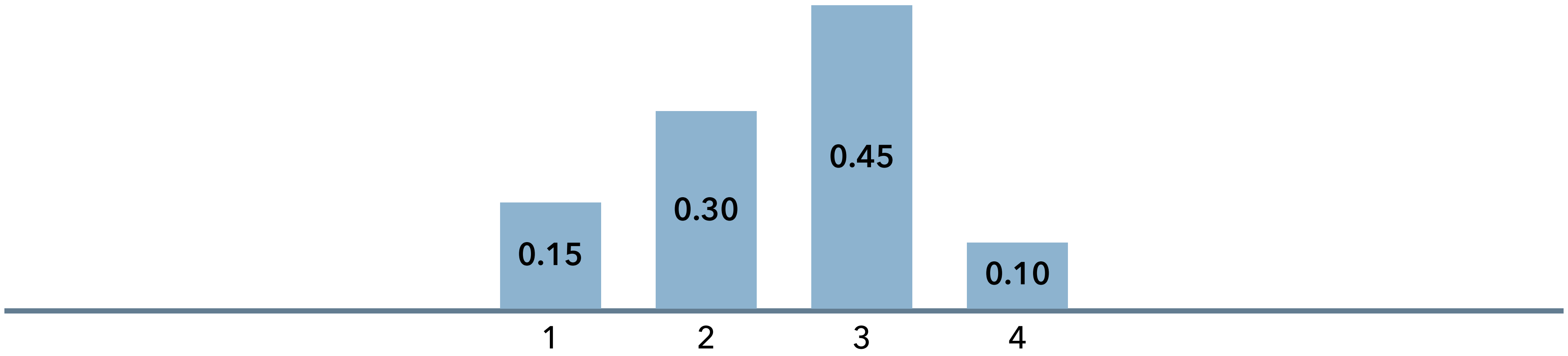
$$p(r) = P(R = r).$$

The value $p(r)$ is the **probability mass** placed on r rooms.



Probability Distribution vs Probability Mass Function

$$P : \mathcal{E} \to [0, 1] \qquad p : \mathcal{X} \to [0, 1]$$



$$p(2) = 0.3$$

$$P(X = 2) = 0.3 \quad essentially \quad P(\{2\}) = 0.3$$

Requirements for a PMF

For every possible number of rooms $r \in \mathcal{R}$,

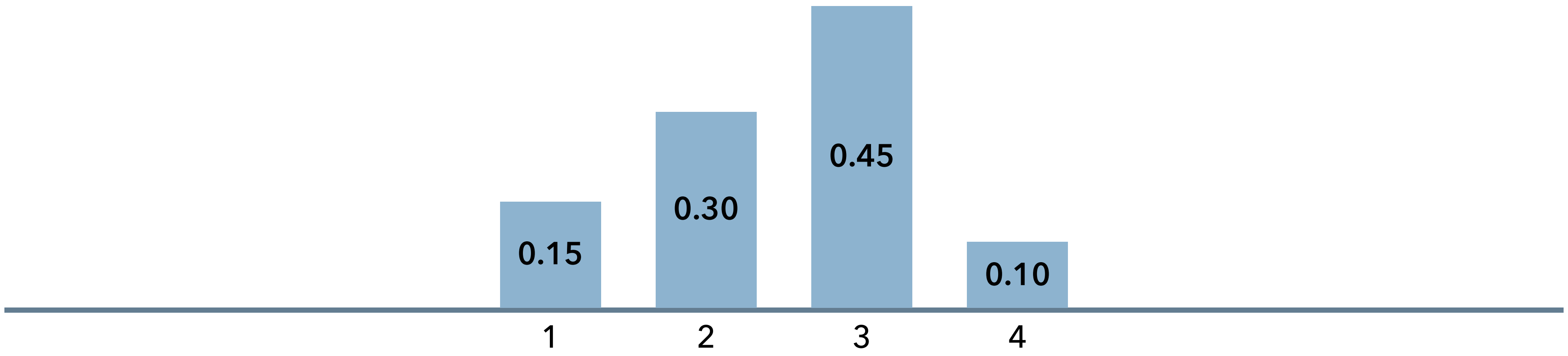
$$p(r) \geq 0,$$

and all masses sum to one:

$$\sum_{r \in \mathcal{R}} p(r) = 1.$$

Event Probabilities Come from the PMF

For an event $E \subseteq \mathcal{R}$, $P(R \in E) = \sum_{r \in E} p(r)$.



$P(R \geq 3) = ?$

$P(R \geq 3) = p(3) + p(4) = 0.45 + 0.10 = 0.55.$

Useful Discrete Distribution: Bernoulli

A Bernoulli distribution is a discrete probability model that describes an experiment with only two possible outcomes.

E.g., let $X = 1$ denote a positive review and $X = 0$ denote a negative review.

$$X \sim \mathrm{Bernoulli}(\alpha)$$

means

$$p(x) = \begin{cases} \alpha & x = 1, \\ 1 - \alpha & x = 0. \end{cases}$$

Continuous Random Variables

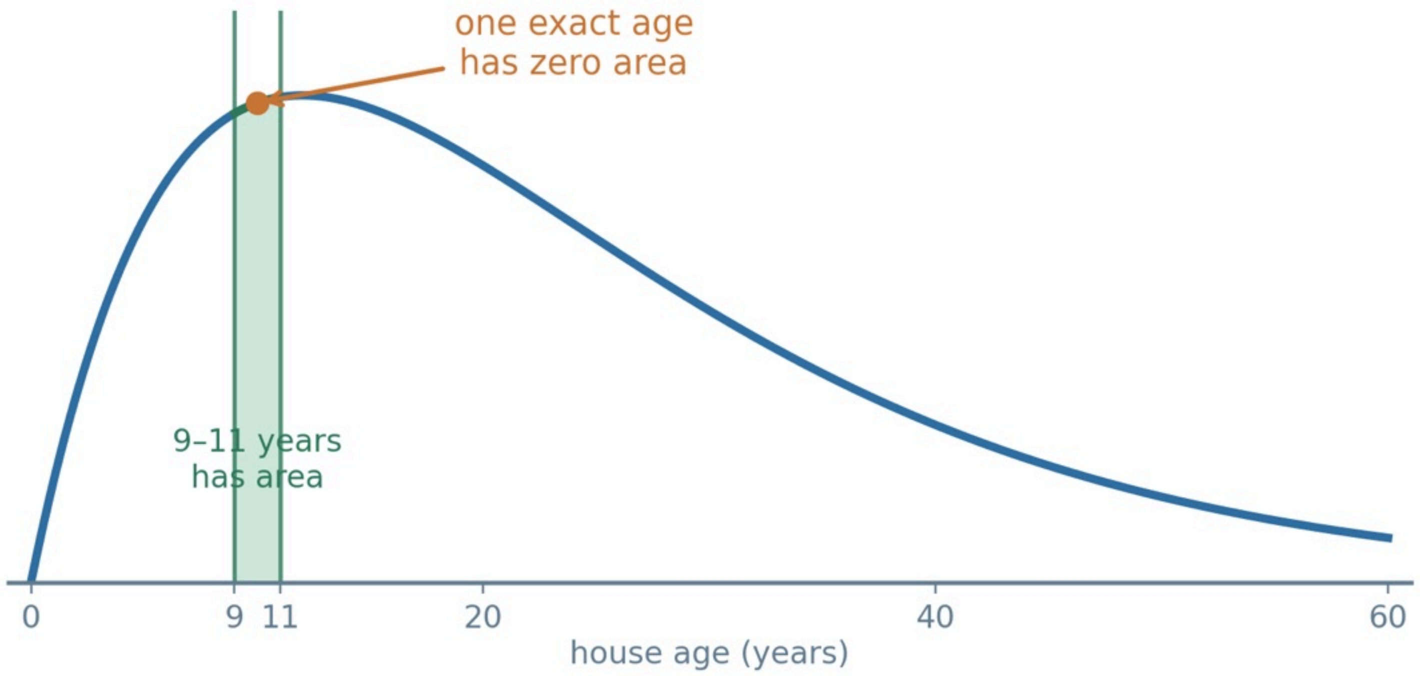
Let $X \in [0, \infty)$ be the exact age of a randomly selected house (it can include decimals).

One exact value

$$P(X = 10) = 0.$$

An interval

$$P(9 \le X \le 11) > 0.$$

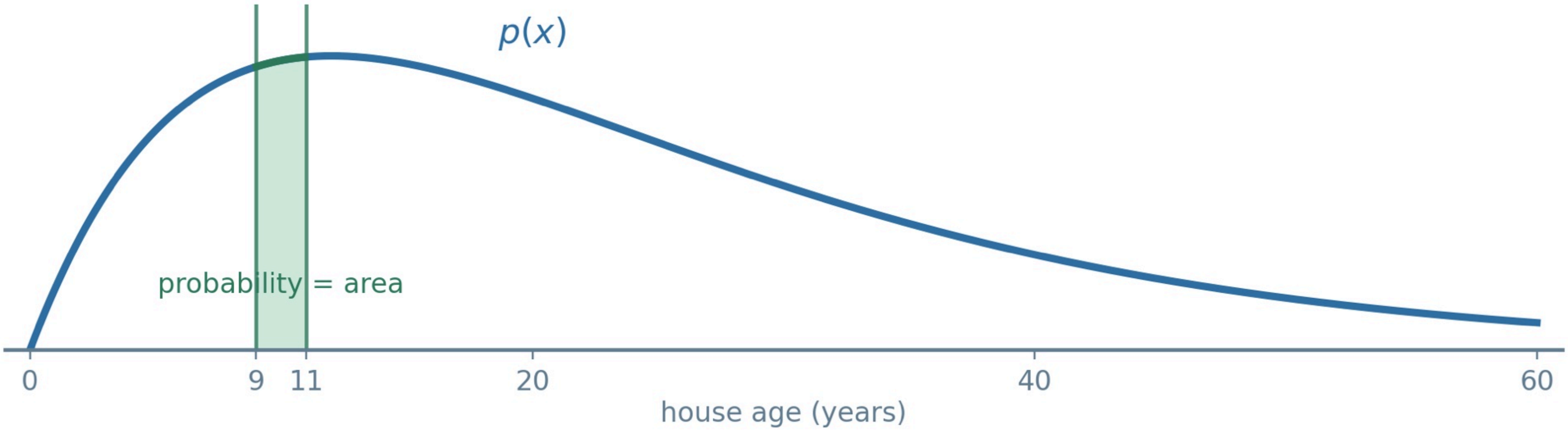


For a continuous variable, probability belongs to **regions**, not individual points.

A PDF Describes Continuous Density

For a continuous random variable X, we describe the likelihood of events using probability density function (PDF), which satisfies

$$p(x) \geq 0, \qquad \int_{\mathcal{X}} p(x)\, dx = 1.$$



Probability (as Area)

For a continuous X,

$$P(a \le X \le b) = \int_a^b p(x)\,dx.$$

Density at a Point

$p(x)$

Can be larger than 1

Probability at a Point

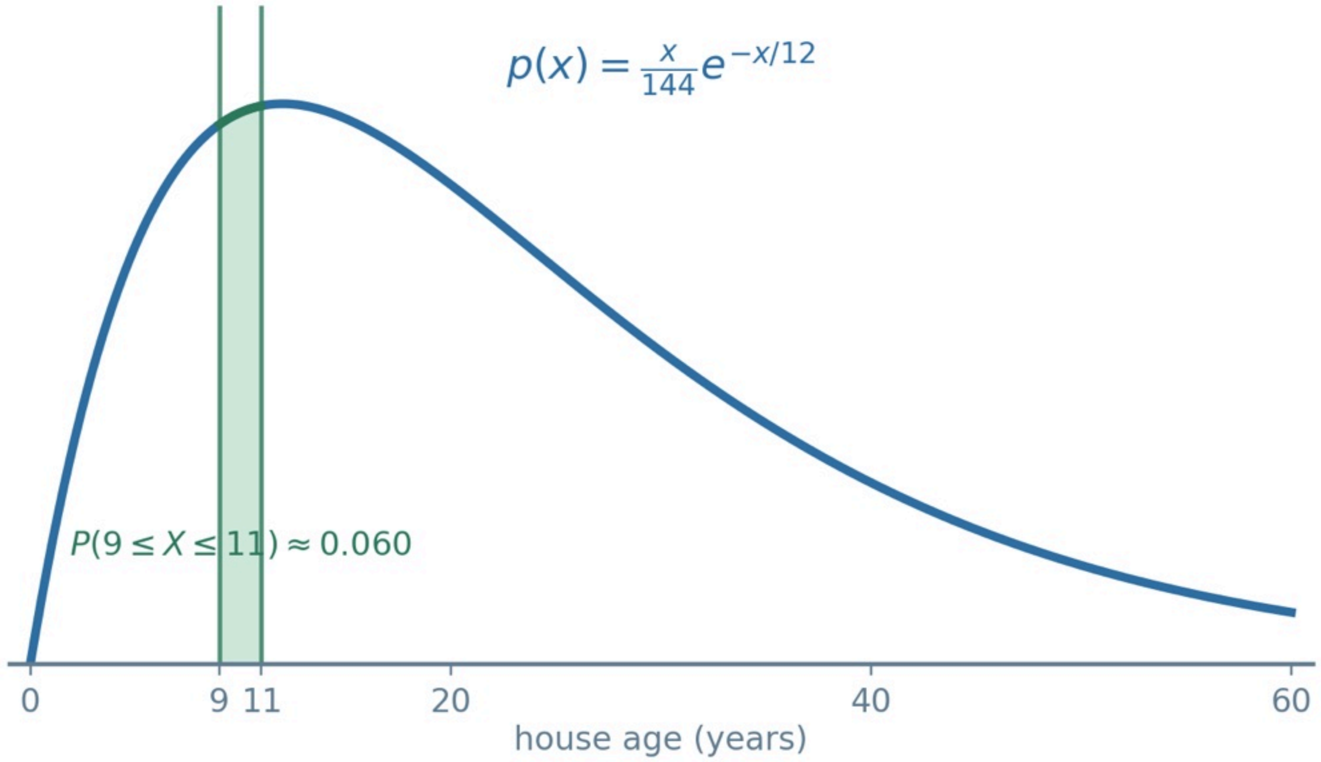
$P(X = x) = 0$

For a continuous variable

Example: House-Age Density

Assuming density function:

$$p(x) = \begin{cases} \dfrac{x}{144} e^{-x/12} & x \geq 0, \\ 0 & x < 0. \end{cases}$$



The shaded interval has probability

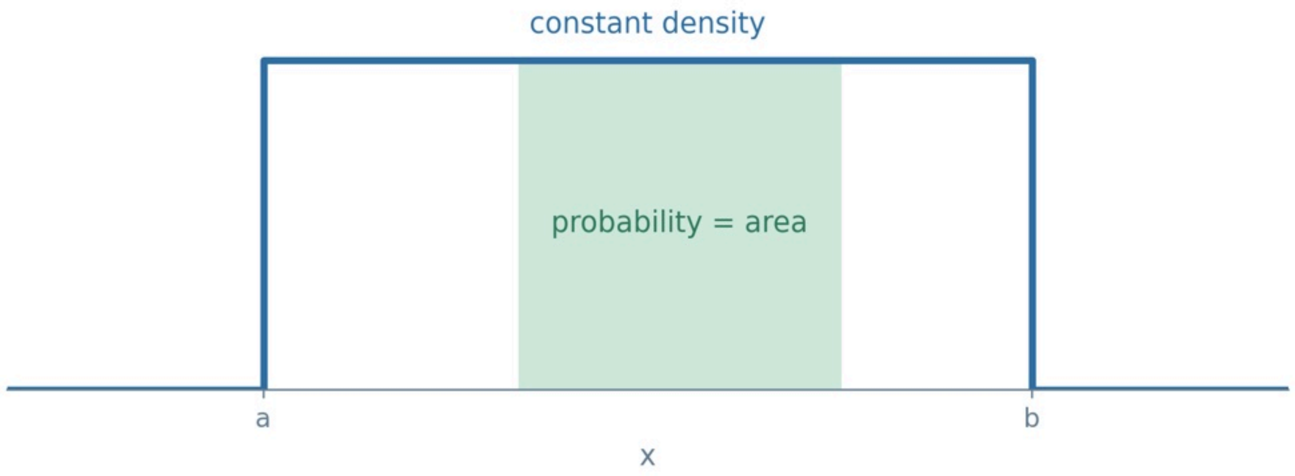
$$P(9 \leq X \leq 11) = \int_9^{11} \frac{x}{144} e^{-x/12} \, dx$$
$$\approx 0.060.$$

Useful Continuous Distribution: Uniform

A **uniform distribution** assigns the same density to every value in an interval $[a, b]$.

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

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



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

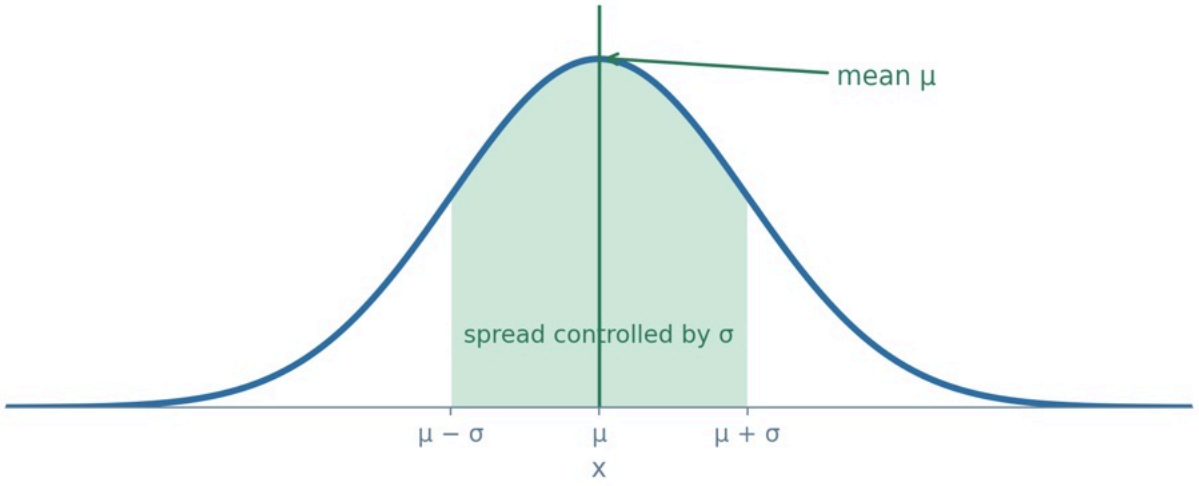
$$P(c \leq X \leq d) = \frac{d-c}{b-a}.$$

Useful Continuous Distribution: Gaussian

A **Gaussian distribution** (also called a normal distribution) is a symmetric, bell-shaped model for values concentrated around a typical value.

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

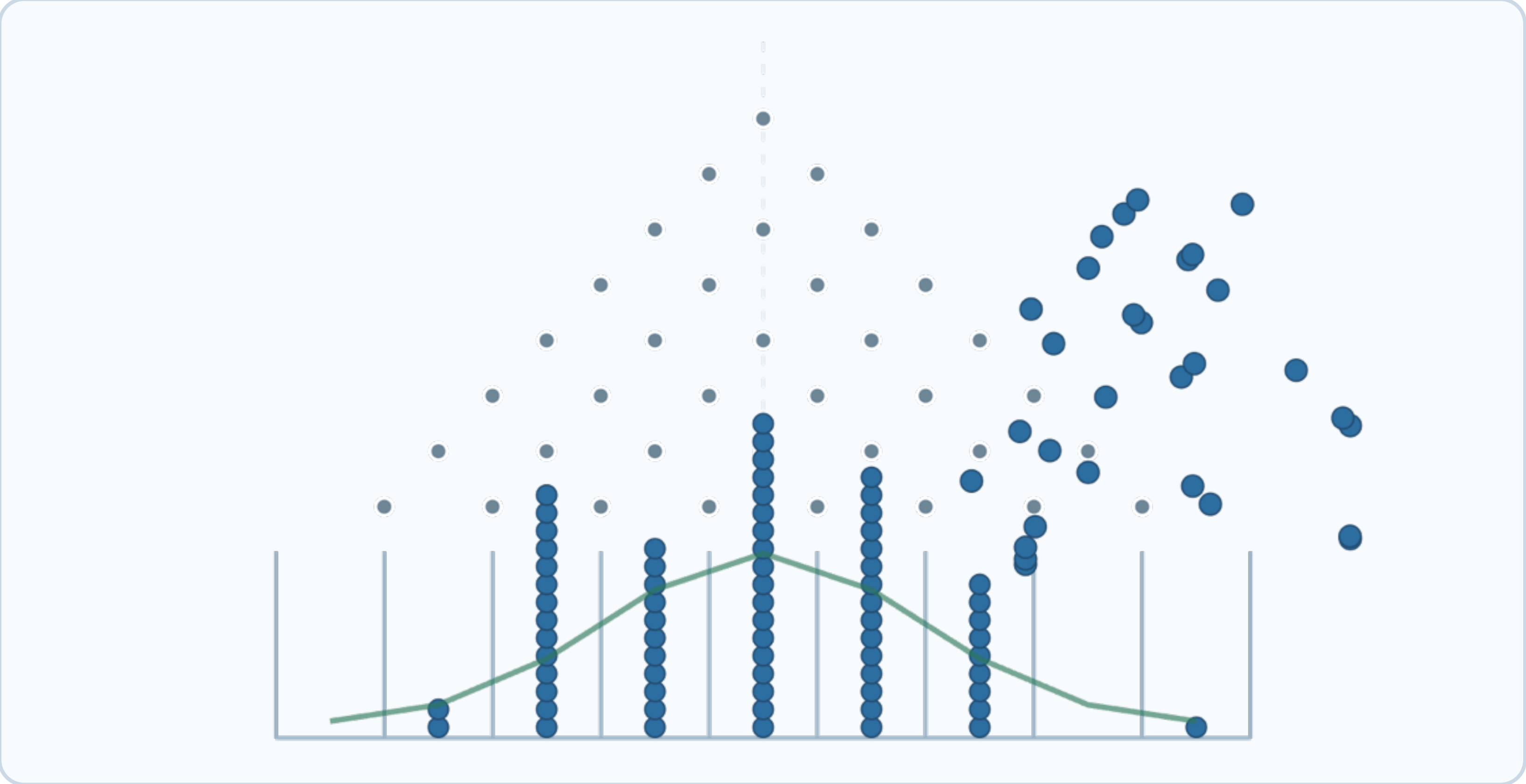


- μ (mean) sets the center.
- σ sets a typical distance from the center; σ^2 is the variance.
- About 68% of values lie within $\mu \pm \sigma$.

Gaussian Distribution (Cont'd)

When **many small, roughly independent effects are added**, their sum often has an approximately Gaussian distribution.

Galton Ball At each peg, a ball moves **left or right with equal probability**.



Outcome: the landing position is the sum of eight steps.

Middle bin	**Far edge**
70 routes	1 route

Multi variable random distributions

One House Has Several Random Quantities

Rooms

$R \in \mathbb{N}$

Discrete feature

Age

$A \in [0, \infty)$

Continuous feature

Price

$Y \in [0, \infty)$

Continuous label

$$(R, A, Y) \in \mathbb{N} \times [0, \infty) \times [0, \infty)$$

Two Variables Form One Joint Variable

Let

$$X = \mathbb{1}\{\text{review contains “great”}\}, \qquad Y = \mathbb{1}\{\text{review has positive sentiment}\}.$$

The pair

$$Z = (X, Y) \in \{0,1\} \times \{0,1\}$$

is a multivariate random variable.

A Joint PMF Describes Combinations

	$Y=0$ Negative	$Y=1$ Positive
$X=0$ No “great”	0.42	0.08
$X=1$ Contains “great”	0.18	0.32

For example,

$p_{X,Y}(1,1) = P(X=1, Y=1) = 0.32$. (The review contains great and the sentiment is positive)

> *i* Every cell is a joint event; all cells together sum to one.

Marginalization

Marginalization allows extracting probabilities of a specific variable X from a joint distribution over X and Y.

To find the probability that $X = 1$ (review contains great), add over every possible value of Y:

$$p_X(1) = \sum_{y \in \{0,1\}} p_{X,Y}(1, y).$$

	$Y = 0$ Negative	$Y = 1$ Positive	**Marginal**
$X = 0$ No "great"	0.42	0.08	0.50
$X = 1$ Contains "great"	0.18	0.32	0.50

Both Marginals Come from the Joint

	$Y=0$ Negative	$Y=1$ Positive	$p_X(x)$
$X=0$ No "great"	0.42	0.08	0.50
$X=1$ Contains "great"	0.18	0.32	0.50
$p_Y(y)$	0.60	0.40	1.00

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

Quick Check: Read the Table

Using the joint PMF,

	$Y = 0$ Negative	$Y = 1$ Positive
$X = 0$ No "great"	0.42	0.08
$X = 1$ Contains "great"	0.18	0.32

What is $P(Y = 1)$?

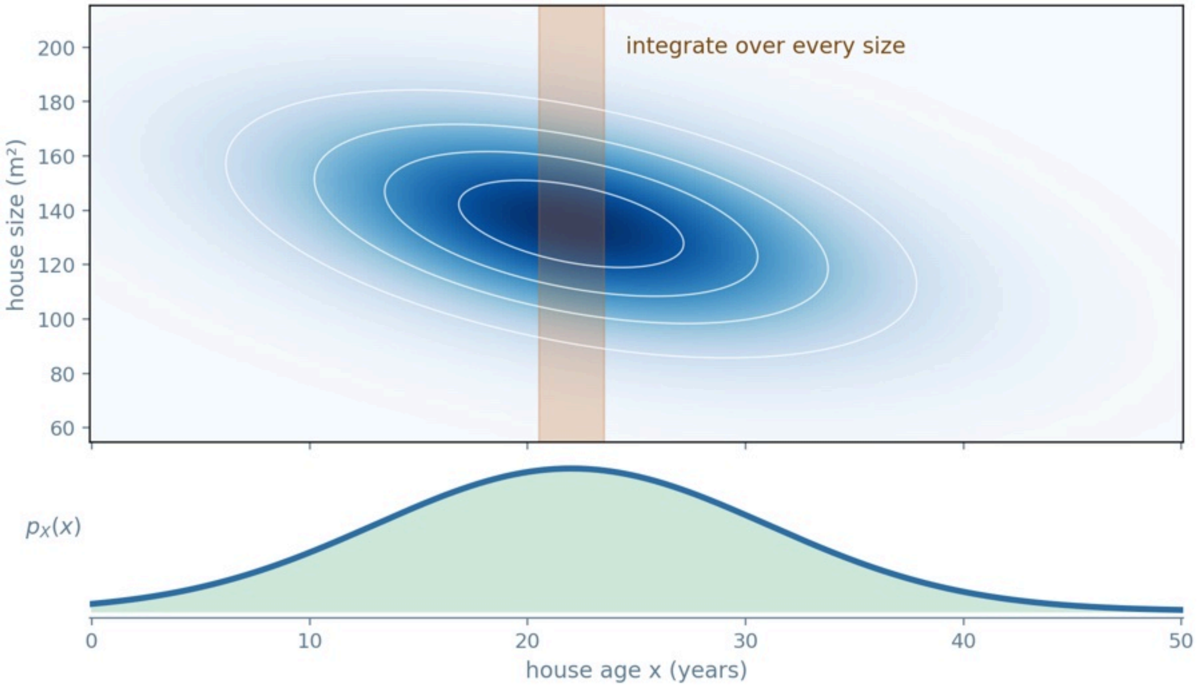
$P(Y = 1) = 0.08 + 0.32 = 0.40.$

Marginalization for Continuous Variables

Let X be house age and Y be house size, with joint PDF $p_{X,Y}(x, y)$.

To ignore house size, integrate over **every possible value** of Y:

$$p_X(x) = \int_{\mathcal{Y}} p_{X,Y}(x, y)\, dy.$$



Similarly,

$$p_Y(y) = \int_{\mathcal{X}} p_{X,Y}(x, y)\, dx.$$

For discrete variables we sum; for continuous variables we integrate.

Functions of Random Variables

If $X \in \mathcal{X}$ is a random variable and $f : \mathcal{X} \to \mathcal{Y}$ is a function, then $f(X)$ is also a random variable (which means you can have PMF / PDF over its output).

Random Input X $\rightarrow$ **Fixed Function** f $\rightarrow$ **Random Output** $f(X)$

Example: A Function of a Die Roll

Let X be a fair die roll and define

$$f(X) = \mathbb{1}\{X \text{ is even}\}.$$

Odd Roll

$X \in \{1, 3, 5\} \Rightarrow f(X) = 0$

Even Roll

$X \in \{2, 4, 6\} \Rightarrow f(X) = 1$

Thus $f(X) \sim \text{Bernoulli}(1/2)$.

Wrap-up

- A random variable takes values in an outcome space.
- A PMF assigns probability mass; a PDF assigns density.
- Uniform and Gaussian distributions are useful continuous models.
- A joint distribution describes combinations of random variables.
- Marginalization sums or integrates out an unobserved variable.
- A function of a random variable is itself a random variable.

Next Lecture

What changes after we observe X?

Conditional probability