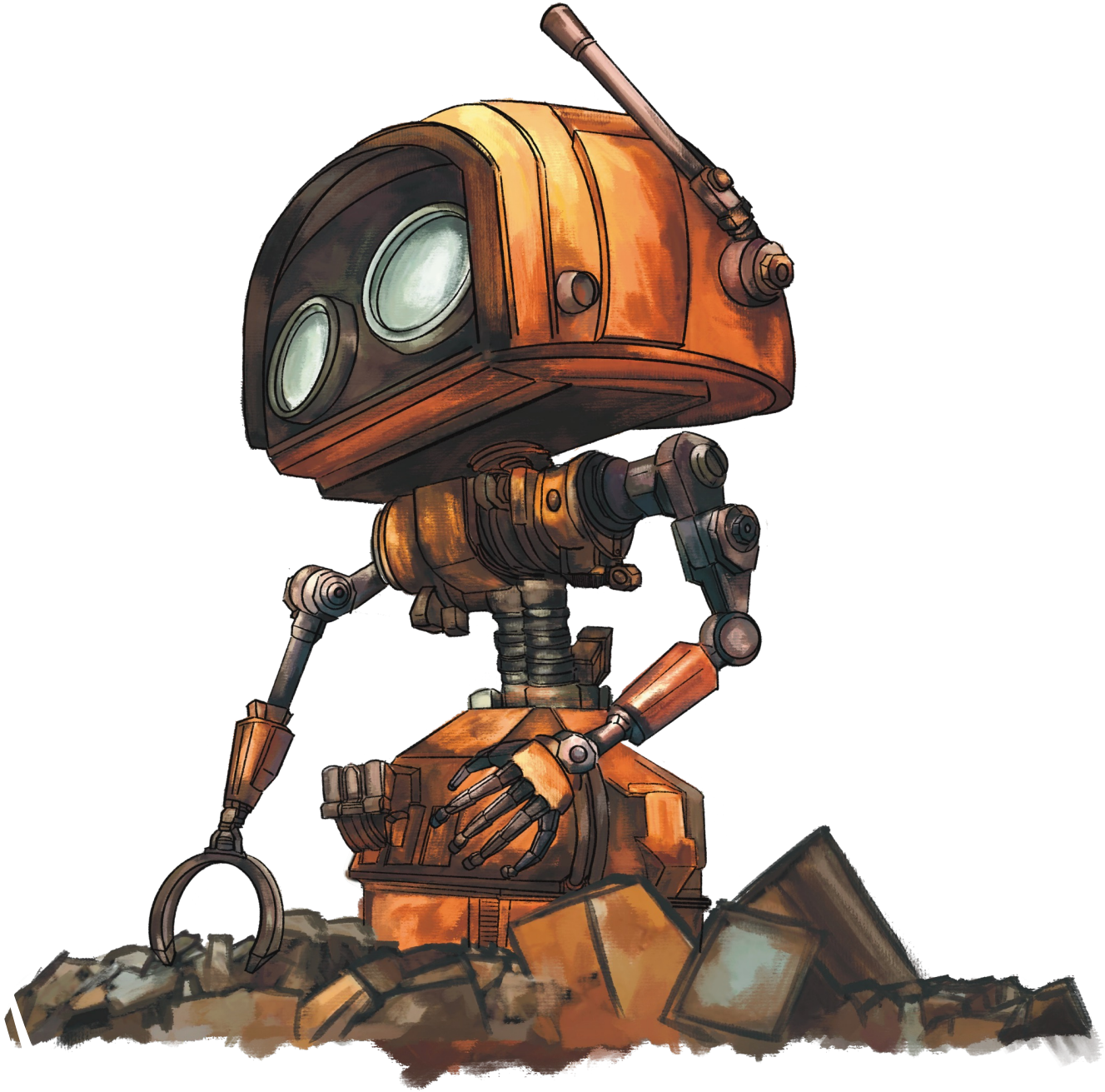
CS 3630, Fall 2025

Lecture 4:

A Trash Sorting Robot: Sensing



Sensing

For our trash sorting robot, we'll consider three sensors:

- Conductivity: A binary sensor that outputs *True* or *False*, based on measurement of electrical conductivity.
- Camera w/detection algorithms: This sensor outputs *bottle*, *cardboard*, or *paper*, based on a detection algorithm (note: it cannot detect scrap metal or cans).
- Scale: Outputs a continuous value that denotes the measured *weight in kg* of the object.

These three kinds of measurements are each treated using distinct probabilistic models.

Binary Sensors

Binary Sensors

- Consider a simple conductivity sensor.
- In an ideal world, the sensor would return the value *True* when the object category is either scrap metal or can, and the value *False* for paper, cardboard and bottle.
- In the real world, metal cans can be dirty causing the sensor to return *False*, even though metal cans conduct electricity.
- There are numerous reasons that a binary sensor could return the wrong value for any of the five categories, but what is *more interesting than the cause of the error is the probability associated to the error*.
- What is the probability that the sensor will return *True* for a metal can? *False* for a piece of cardboard? *True* for a bottle? Etc….

➢ ***If we know these probabilities, we can reason about the object category based on sensor reading!***

Conditional probability revisited

- Conditional probabilities quantify the probabilities associated with correct/incorrect sensor readings.
- For each category, we estimate the probability of True and False.
- We collect these values into a conditional probability table (CPT):

Category (C)	P(False|C)	P(True|C)
Cardboard	0.99	0.01
Paper	0.99	0.01
Cans	0.1	0.9
Scrap Metal	0.15	0.85
Bottle	0.95	0.05

Given that the object is cardboard, the probability of False is 0.99.

Given that the object is scrap metal, the probability of True is 0.85.

Each entry in the table is a conditional probability value. The conditioning event is the category.

Some things to remember

- For a fixed category C, $P(Conductivity|C)$ is itself a probability!
- Therefore:

$$P(True|C) = 1 - P(False|C)$$

➢ Because of this fact, each row in the table sums to one!

Category (C)	P(False|C)	P(True|C)
Cardboard	0.99	0.01
Paper	0.99	0.01
Cans	0.1	0.9
Scrap Metal	0.15	0.85
Bottle	0.95	0.05

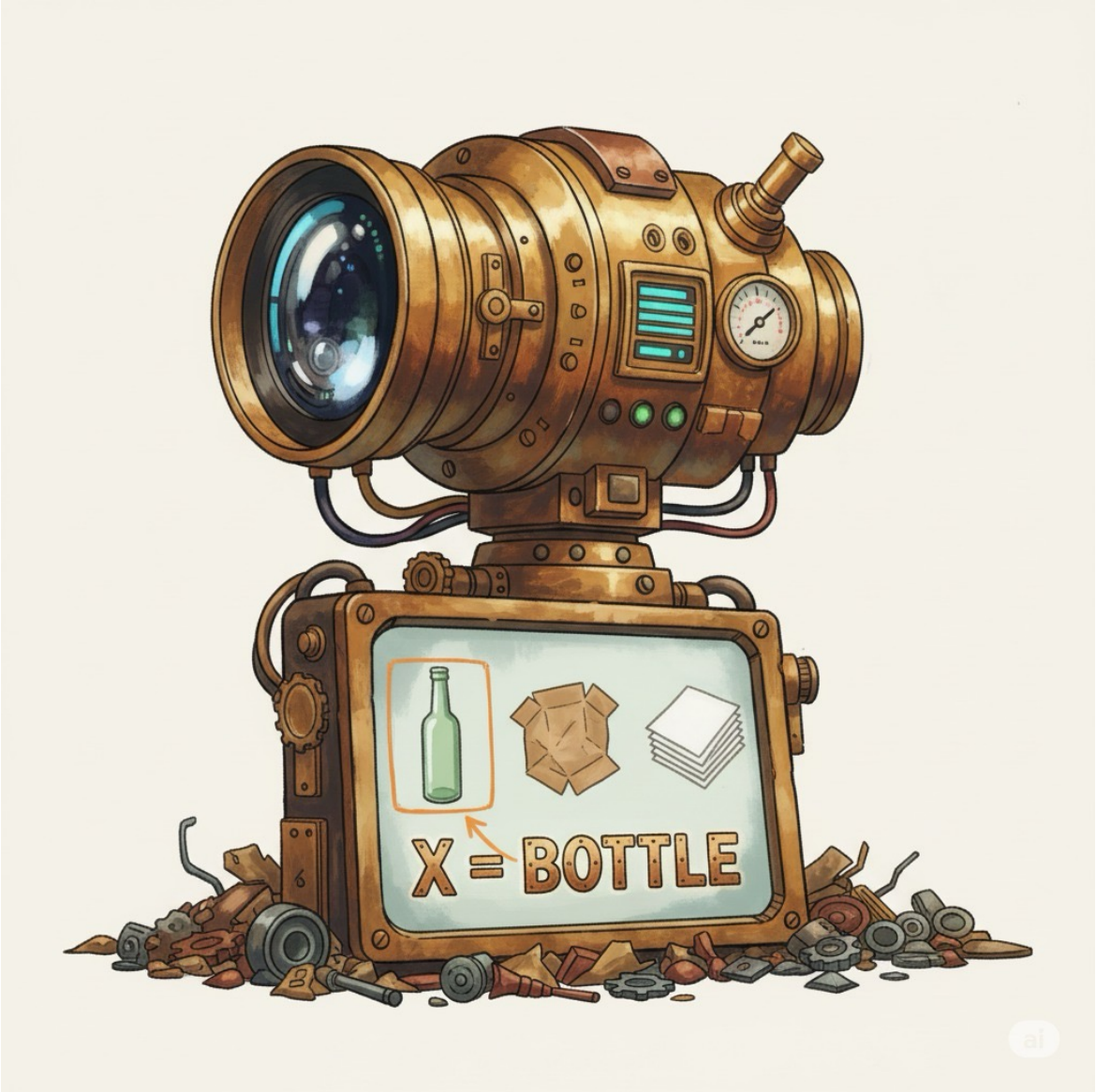
- *We can think of the category C as defining a particular context.*
- *The conditional probability for an outcome tells us the probability of that outcome in a specific context.*
- *If the context is "a piece of cardboard is in the work cell," then the probability of False is 0.99.*

➢ ***If we think of $f(C) = P(Cond|C)$ as a function of C, then $f(C)$ is <u>NOT</u> a probability.***
➢ ***Note that the columns do <u>Not</u> sum to one. (more about this soon…)***

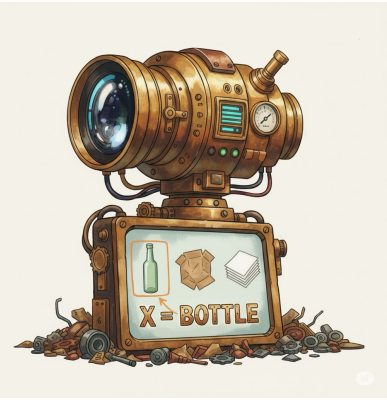
We must estimate the conditional probabilities!

- In practice, it is not possible to ***know*** the conditional probabilities.
- It may even be the case that these probabilities change over time.
- We can determine the conditional probability values by:
  - A. Reasoning about the physics of the sensor, combining intuition with physical laws to arrive to reasonable guesses for these values
  - B. Gathering lots of data, and estimating the conditional probabilities using relative frequency (aka histograms):
    1. Collect N conductivity measurements on pieces of cardboard.
    2. Let N_{true} be the number of times the sensor returns true.
    3. $P(True|cardboard) = \frac{N_{true}}{N}$, $P(False|cardboard) = \frac{N - N_{true}}{N}$
    4. Repeat for each category.
  - C. Reading the data sheet that was shipped with the sensor (in this case, the manufacturer used either A or B).

Multi-valued sensors



Multi-valued sensors



- We could consider a binary sensor as a device that returns a value from a set

$$X \in \{x_1, x_2\} = \{True, False\}.$$

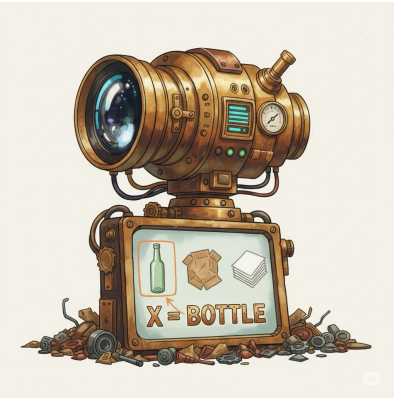
- If we take this view, it's a simple matter to extend our approach to any set of discrete outcomes: $X \in \{x_1, \ldots, x_n\}$.
- For our trash sorting robot, we have a computer vision sensor that returns a value
  - $\mathrm{X} \in \{bottle, cardboard, paper\}$.

- Each possible outcome gives rise to one column in our CPT for the sensor:

Category	bottle	cardboard	paper
cardboard	0.02	0.88	0.1
paper	0.02	0.2	0.78
can	0.333333	0.333333	0.333333
scrap metal	0.333333	0.333333	0.333333
bottle	0.95	0.02	0.03

- As before, each entry is $P(DetectorReading \mid Category)$.
- Do not confuse the Category and the DetectorReading, ***even if they share the same name****!*
- Note that each row still sums to one.
- For cans and scrap metal, this detection sensor becomes confused, and returns one of the three values at random, each with probability of $\frac{1}{3}$.

The value of multiple sensors



- The Conductivity sensor
  - does a good job of discriminating between the events $\{bottle, cardboard, paper\}$ and $\{scrap\ metal, can\}$,
  - but is unable to resolve ambiguity in either of these events.
- The computer vision sensor
  - does a good job of discriminating between $\{bottle, cardboard, paper\}$,
  - but is useless for $\{scrap\ metal, can\}$.

Category (C)	P(False|C)	P(True|C)
Cardboard	0.99	0.01
Paper	0.99	0.01
Cans	0.1	0.9
Scrap Metal	0.15	0.85
Bottle	0.95	0.05

Conductivity Sensor

Category	bottle	cardboard	paper
cardboard	0.02	0.88	0.1
paper	0.02	0.2	0.78
can	0.333333	0.333333	0.333333
scrap metal	0.333333	0.333333	0.333333
bottle	0.95	0.02	0.03

Computer Vision Sensor

We'll see how to combine information from different sensors soon.

The weight sensor (aka scale)

Continuous random variables

- Recall the cumulative distribution function (CDF):

$$F_X(\alpha) = P(X \leq \alpha)$$

- If F_X is continuous everywhere, then X is a ***continuous random variable***.

- If X is a continuous random variable with CDF $F_X(\alpha)$, then the ***probability density function (pdf)*** for X is given by

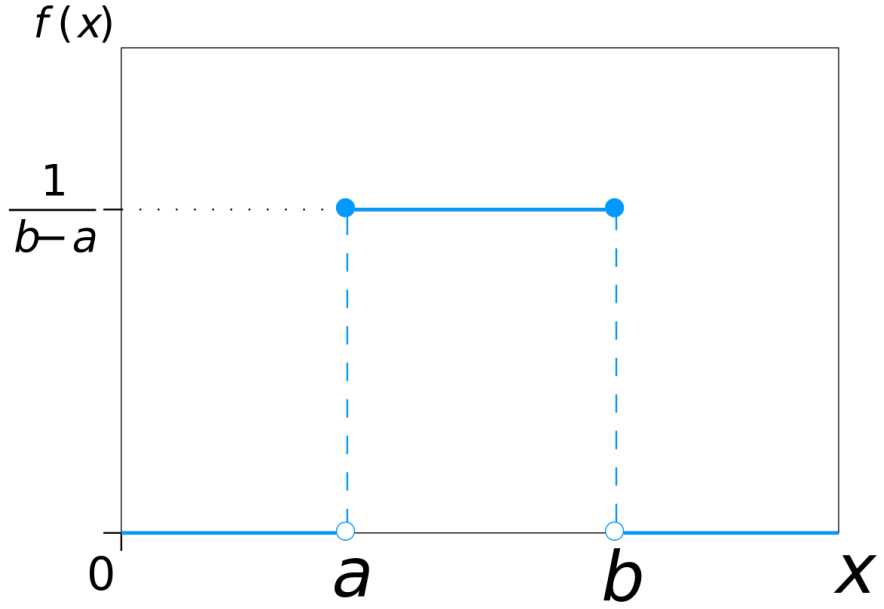
$$f_X(x) = \frac{d}{dx} F_X(x)$$

If we think of $F_X(\alpha)$ as probability mass for event $X \leq \alpha$, we can think of the derivative of mass as density.

The uniform distribution

- The uniformly distribution is the simplest example of a continuous random variable.
- We saw this distribution in our sampling algorithm.
- We use the notation $X \sim U(a, b)$ to denote that X is a continuous random variable with uniform distribution on the interval $[a, b]$.
- The pdf for such an RV is given by:

pdf for the uniform distribution



Computing probabilities

Applying the fundamental theorem of calculus, we obtain:

$$\int_{\alpha}^{\beta} f_X(u)du = F_X(\beta) - F_X(\alpha)$$

$$= P(X \leq \beta) - P(X \leq \alpha)$$

$$= P(\alpha \leq X \leq \beta)$$

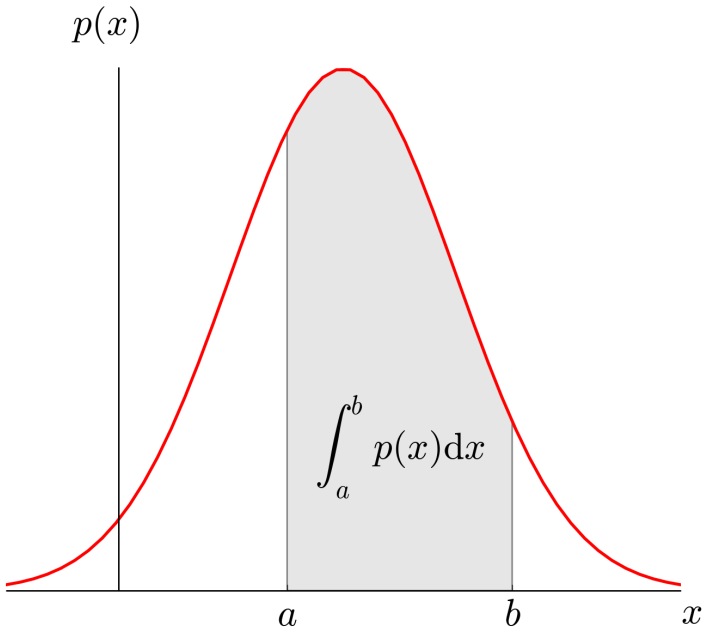
which gives

$$\boldsymbol{P(\alpha \leq X \leq \beta) = \int_{\alpha}^{\beta} f_X(u)du}$$

The probability that $\alpha \leq X \leq \beta$ is equal to the area under the pdf f_X between α and β.

In pictures:

- X takes on values in the continuum.
- $p(x)$, is a probability density function.



What happens when $a = b$?

Since f is continuous

$$\int_a^a f(u)du = 0$$

This leads to the possibly surprising result:

$$P(X = a) = 0$$

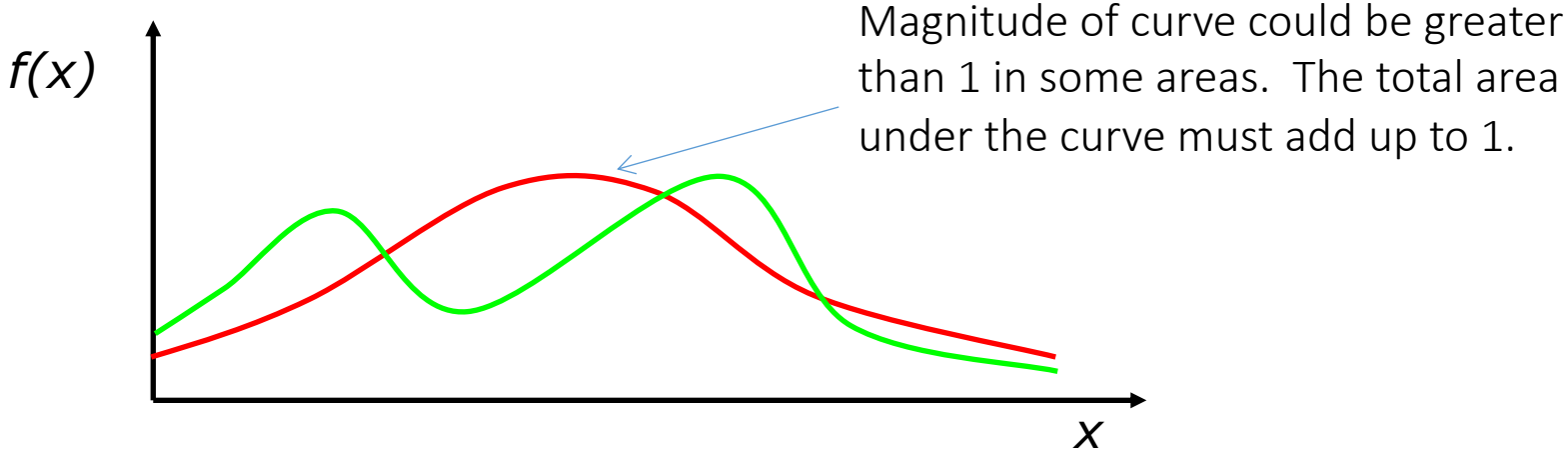
for any scalar a.

More about pdf's

The total area under a pdf equals 1, always, for every pdf.

$$\int_{-\infty}^{\infty} f_X(u)du = F_X(\infty) - F_X(-\infty) = 1 - 0 = 1$$

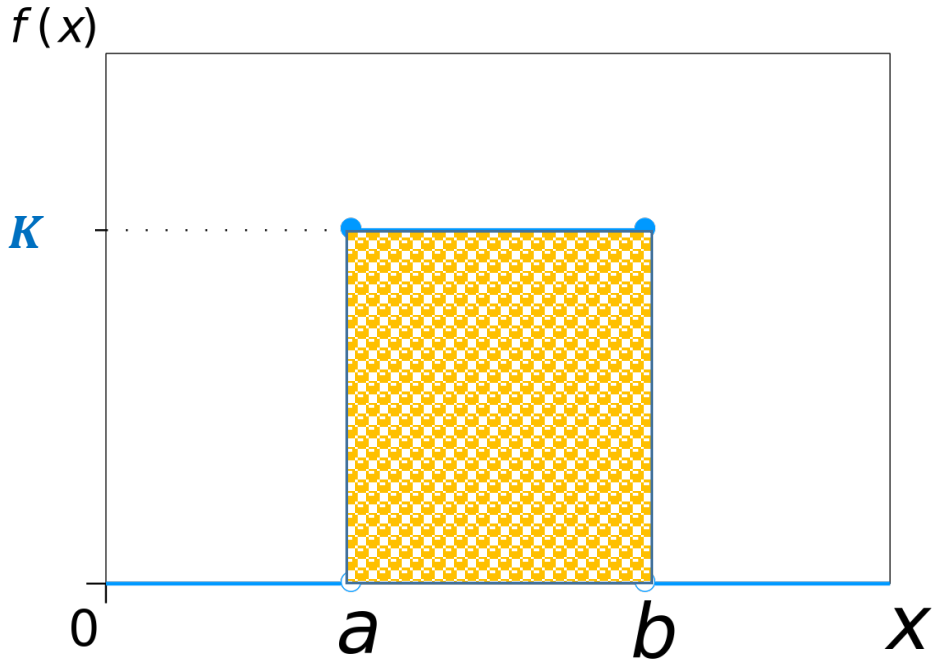
But the magnitude of $f_X(u)$ can take any non-negative value – so long as the total area under the curve integrates to one!



The uniform distribution (again)

- It is now easy to understand why the "height" of the pdf is $\frac{1}{b-a}$:

$$\mathbf{1} = \boldsymbol{P(a \le X \le b)} = \int_{\boldsymbol{a}}^{\boldsymbol{b}} \boldsymbol{K}\, \boldsymbol{du} = \boldsymbol{Kb - Ka} \rightarrow \boldsymbol{K} = \frac{\mathbf{1}}{\boldsymbol{b-a}}$$



In this case, the geometry of rectangles is enough to tell us the answer:

$$Area = K\ (b - a)$$

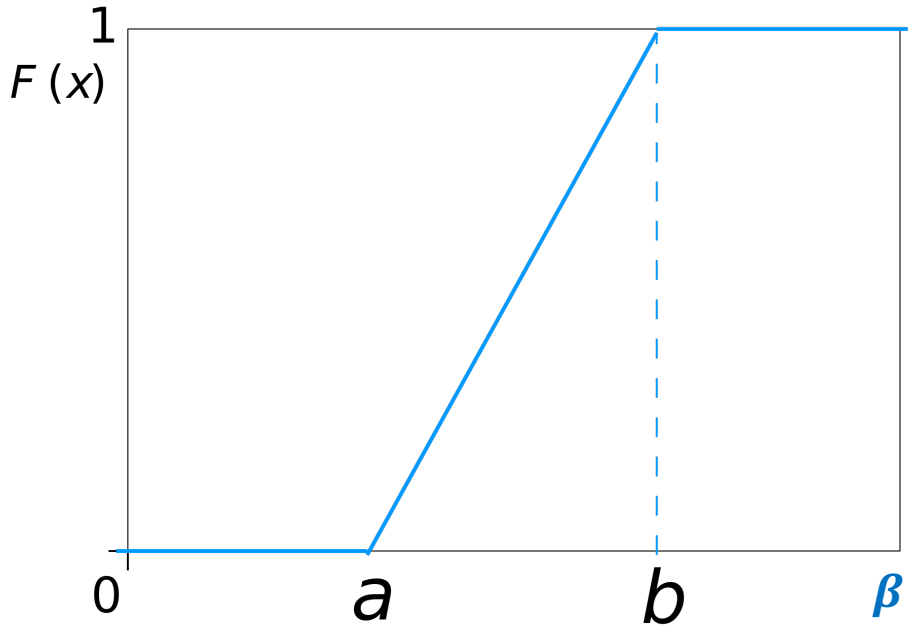
So, if $Area = 1$, then we must have

$$\boldsymbol{K} = \frac{\mathbf{1}}{\boldsymbol{b-a}}$$

The uniform distribution's CDF

- It's easy to compute the CDF for the uniform distribution given its pdf.
- For $(a \leq \beta \leq b)$, simply evaluate the integral

$$F_X(\beta) = P(X \leq \beta) = \int_a^{\beta} \frac{1}{b-a} du$$

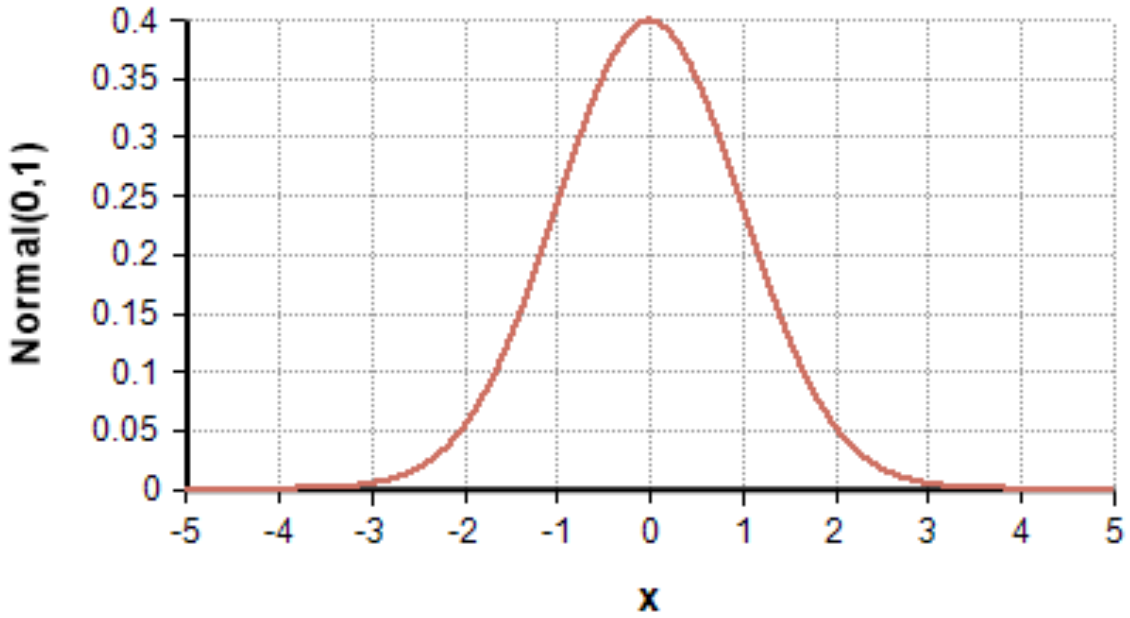


1. For $(a \leq \beta \leq b)$ the integral evaluates to $\frac{\beta - a}{b-a}$.
2. For $(\beta \leq a)$, we have $F_X(\beta) = 0$.
3. For $(b \leq \beta)$, we have $F_X(\beta) = 1$.

Notice that the CDF is continuous everywhere, even though the pdf has discontinuities at a and b.

The Gaussian(aka normal) distribution

The Gaussian distribution is the most famous of all probability distributions, so famous that the Germans put Gauss and his pdf on their money!



Even if you haven't seen this in a probability theory or statistics class, you have likely seen the famous Bell Curve.

The Gaussian distribution

- The Gaussian has two defining parameters.
- The mean, μ
  - Defines the “location” of the pdf.
  - The pdf is symmetric about the mean.
- The variance, σ^2
  - Defines the “spread” of the pdf.
  - Can be specified also in terms of standard deviation, σ.
- The defining equation is given by:

$$f_X(x) = \frac{\mathbf{1}}{\boldsymbol{\sigma\sqrt{2\pi}}} \boldsymbol{e}^{-\frac{\boldsymbol{(x-\mu)^2}}{\boldsymbol{2\sigma^2}}} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

The Gaussian distribution

Let's take a closer look:

- The leading term, $\frac{1}{\sigma\sqrt{2\pi}}$, is a normalizing term, so that $\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1.$

- So, let's simplify notation by writing

$$f_X(x) = Ke^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

The Gaussian distribution

Let's take a closer look:

- The leading term, $\frac{1}{\sigma\sqrt{2\pi}}$, is a normalizing term, so that $\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1.$

- So, let's simplify notation by writing

$$f_X(x) = Ke^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

f_X is a decreasing exponential function.

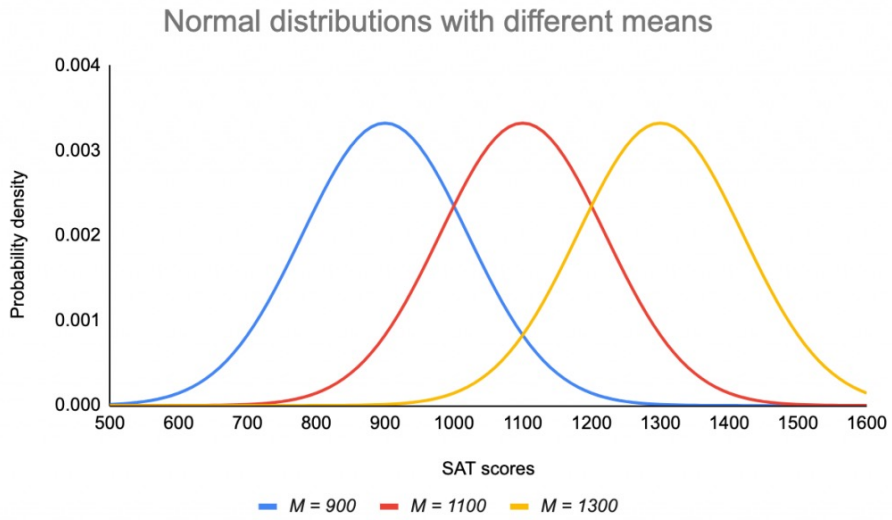
The Gaussian distribution

Let's take a closer look:

- The leading term, $\frac{1}{\sigma\sqrt{2\pi}}$, is a normalizing term, so that $\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1.$

- So, let's simplify notation by writing

$$f_X(x) = Ke^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$



f_X decreases exponentially with the square of the distance to the mean.

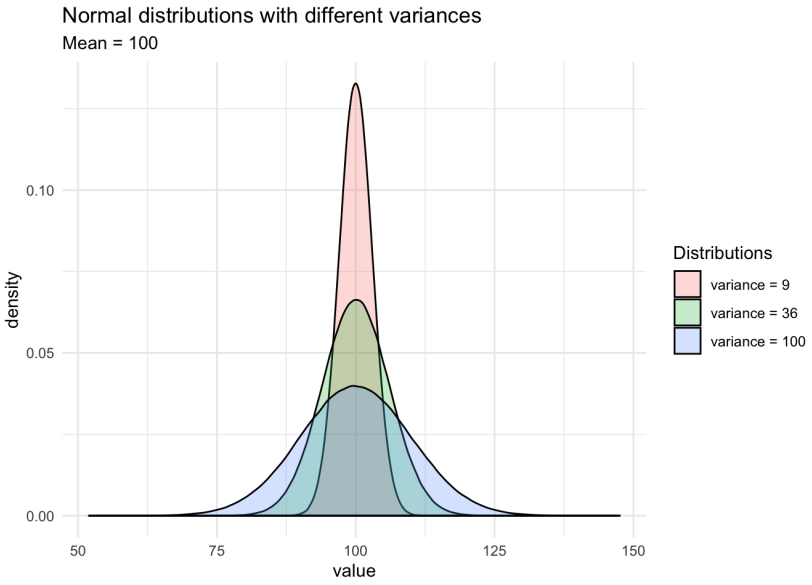
The Gaussian distribution

Let's take a closer look:

- The leading term, $\frac{1}{\sigma\sqrt{2\pi}}$, is a normalizing term, so that $\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1.$

- So, let's simplify notation by writing

$$f_X(x) = Ke^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$



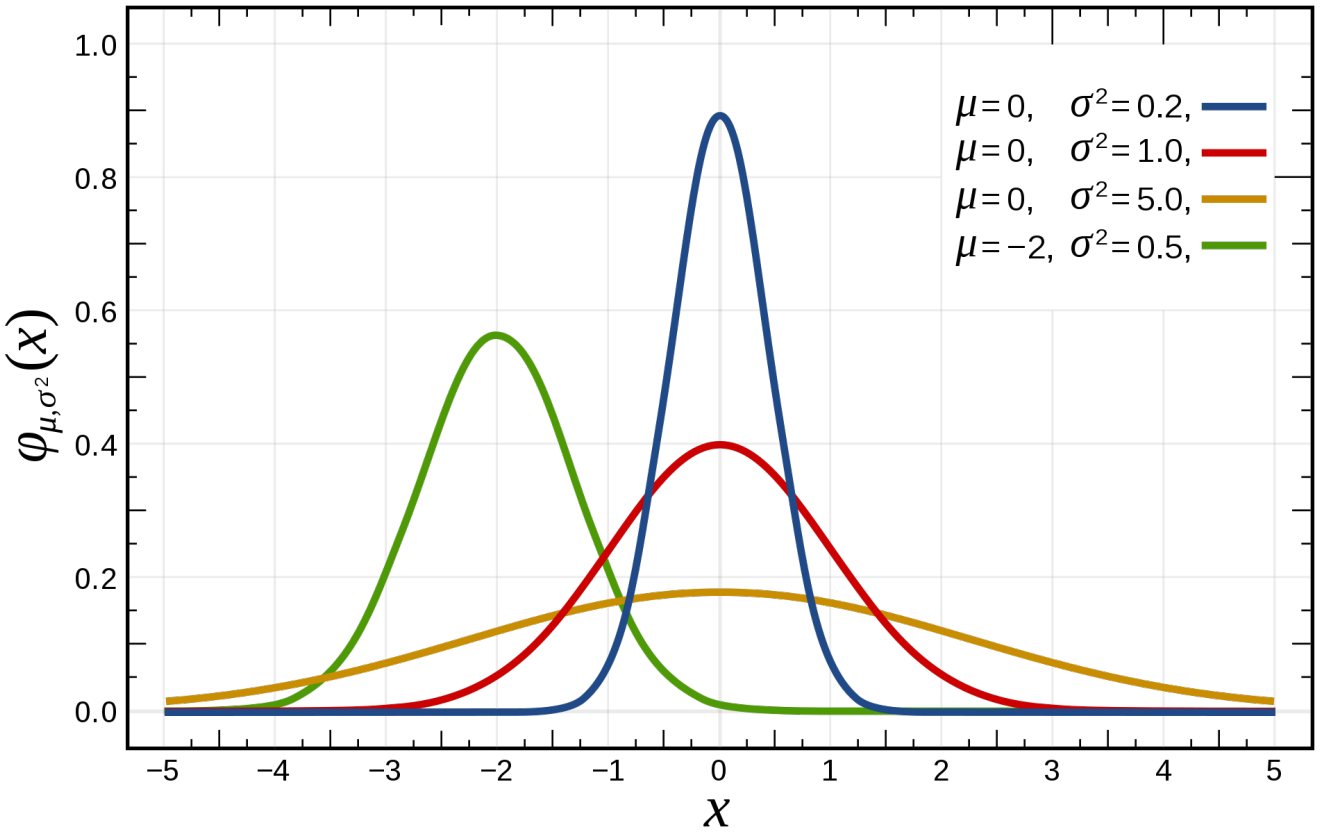
The "rate" of decrease depends on σ^2:

- ***If σ^2 is very large, f_X decreases slowly, thus, a wide spread.***
- ***If σ^2 is very small, f_X decreases quickly, thus, a narrow peak.***

The Gaussian Distribution

- Since the Gaussian is parameterized by its mean and variance, we often write $N(\mu, \sigma^2)$ to denote the Gaussian distribution.
- The special case when $\boldsymbol{\mu = 0, \sigma^2 = 1}$ is called the ***standard normal distribution*** (the red curve in the figure).

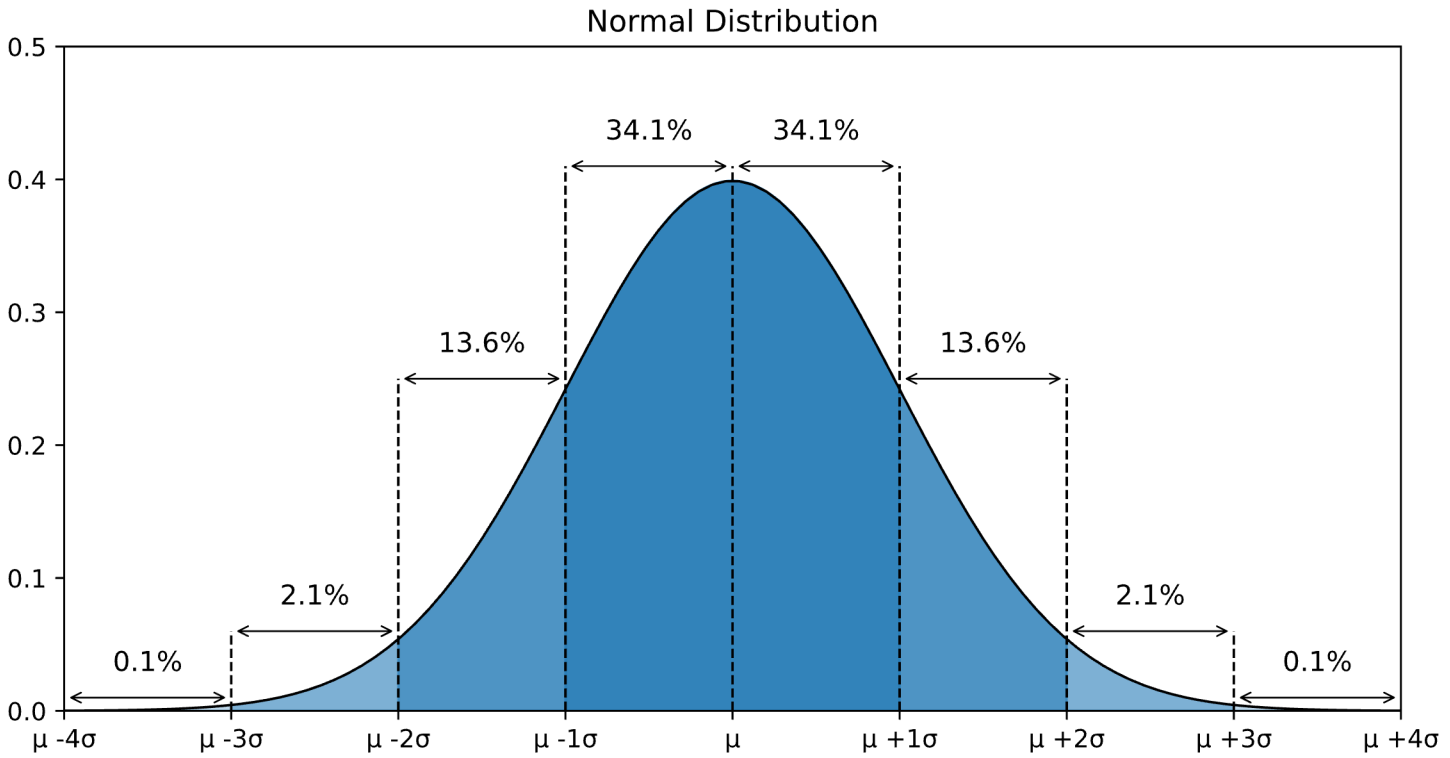
$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



The Gaussian distribution

The standard deviation is a handy way to characterize probabilities.

- 68% of probability mass lies within one standard deviation of μ.
- 99.99966% of the probability mass lies within 6 standard deviations of the mean (for business majors, six sigma is a big thing).



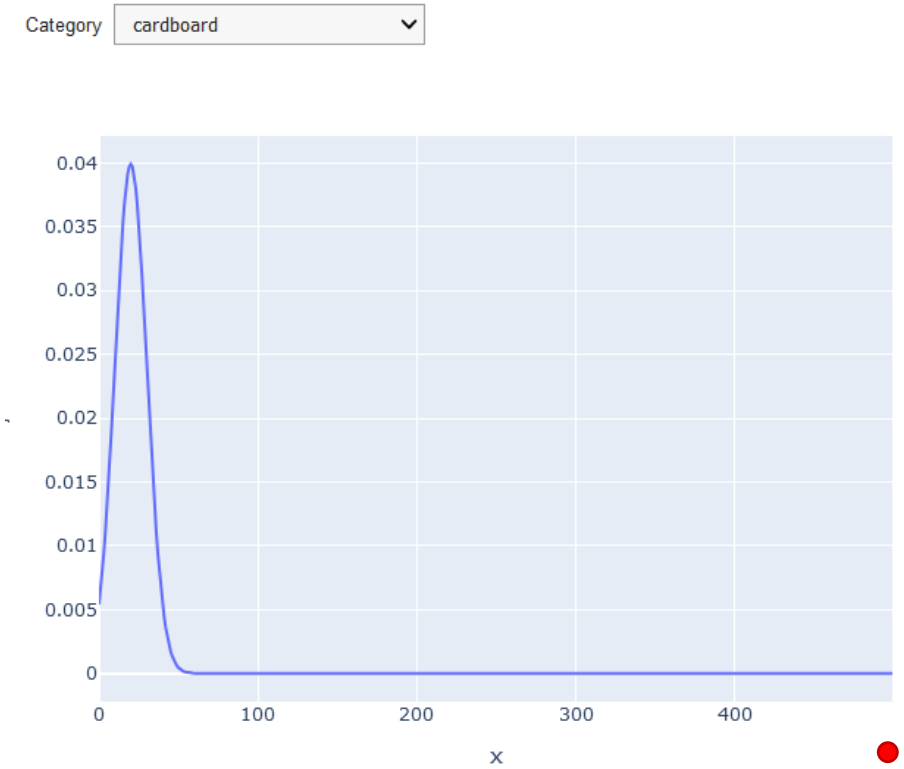
The weight sensor (aka scale)

- Weight of an object = a continuous random variable.
- We'll use the Gaussian distribution to model weight.
  - Annoyance: actually: weight can never be less than zero. Still.
- Each object has its own Gaussian distribution:

Category	**Mean μ**	**Std dev σ**
Cardboard	20	10
Paper	5	5
Can	15	5
Scrap metal	150	100
Bottle	300	200

Cardboard:

- Distribution centered at $\mu = 20$.
- Very narrow distribution.
- Notice truncation at zero.



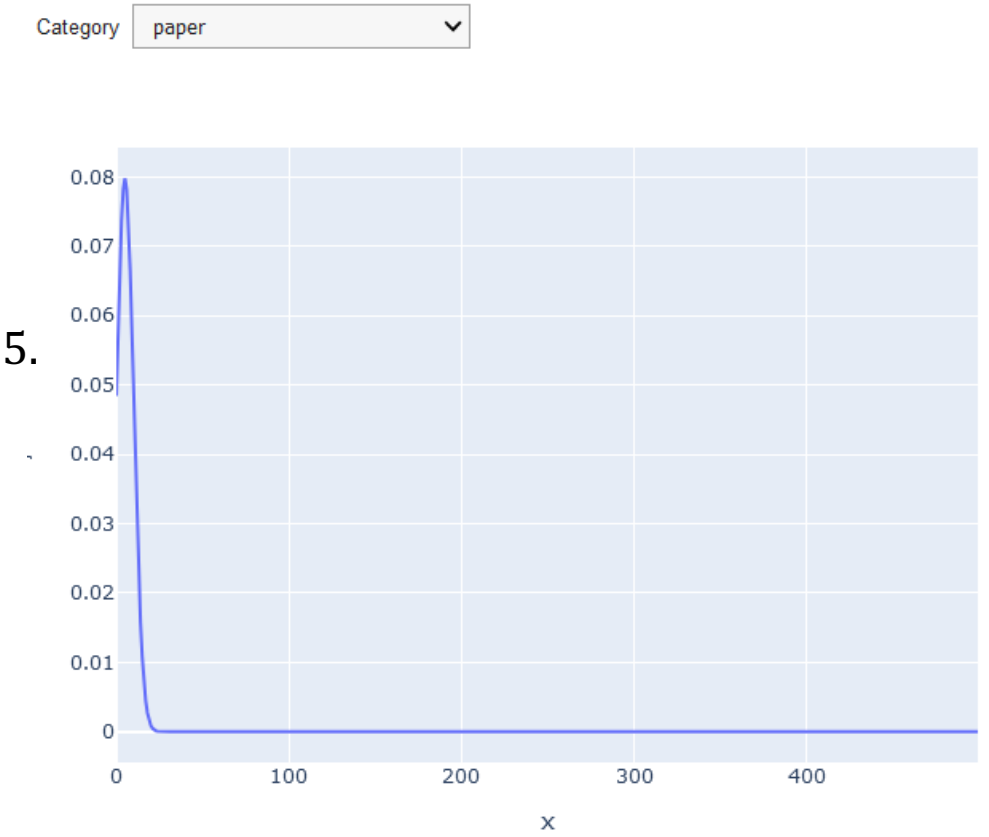
The weight sensor (aka scale)

- Weight of an object = a continuous random variable.
- We'll use the Gaussian distribution to model weight.
  - Annoyance: actually: weight can never be less than zero. Still.
- Each object has its own Gaussian distribution:

Category	Mean μ	Std dev σ
Cardboard	20	10
Paper	5	5
Can	15	5
Scrap metal	150	100
Bottle	300	200

Paper:

- Distribution centered at $\mu = 5$.
- *Very* narrow distribution.
- Notice truncation at zero.



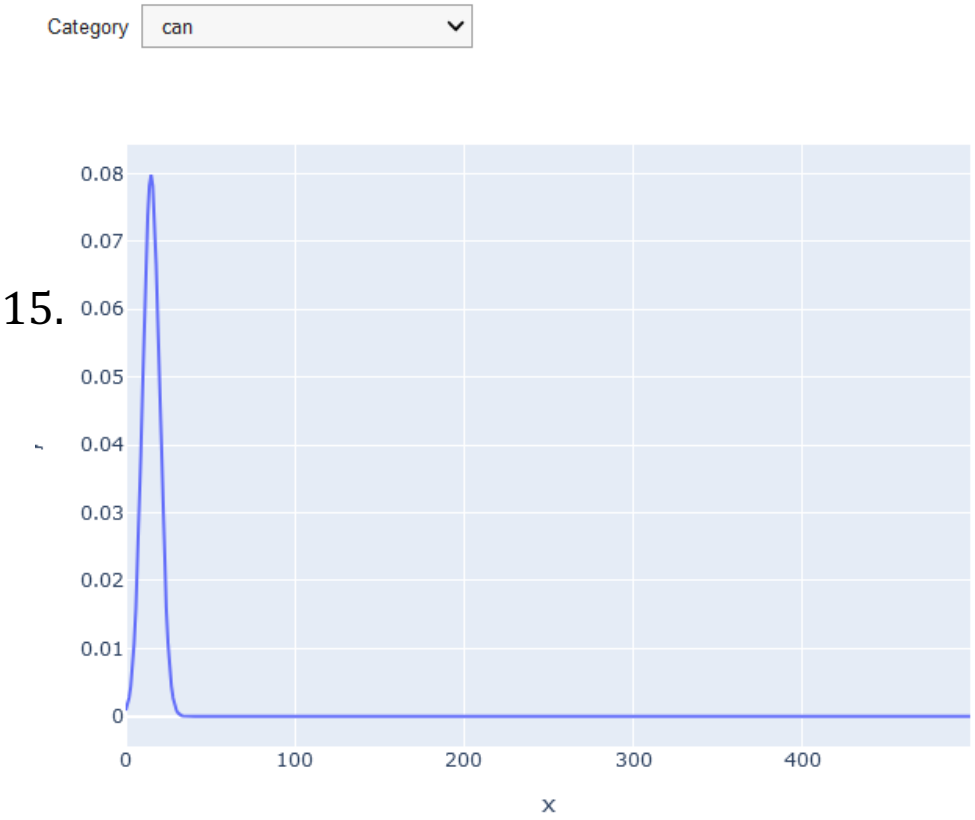
The weight sensor (aka scale)

- Weight of an object = a continuous random variable.
- We'll use the Gaussian distribution to model weight.
  - Annoyance: actually: weight can never be less than zero. Still.
- Each object has its own Gaussian distribution:

Category	Mean μ	Std dev σ
Cardboard	20	10
Paper	5	5
Can	15	5
Scrap metal	150	100
Bottle	300	200

Can:

- Distribution centered at $\mu = 15$.
- Very narrow distribution.
- Notice truncation at zero doesn't really chop off much probability mass.



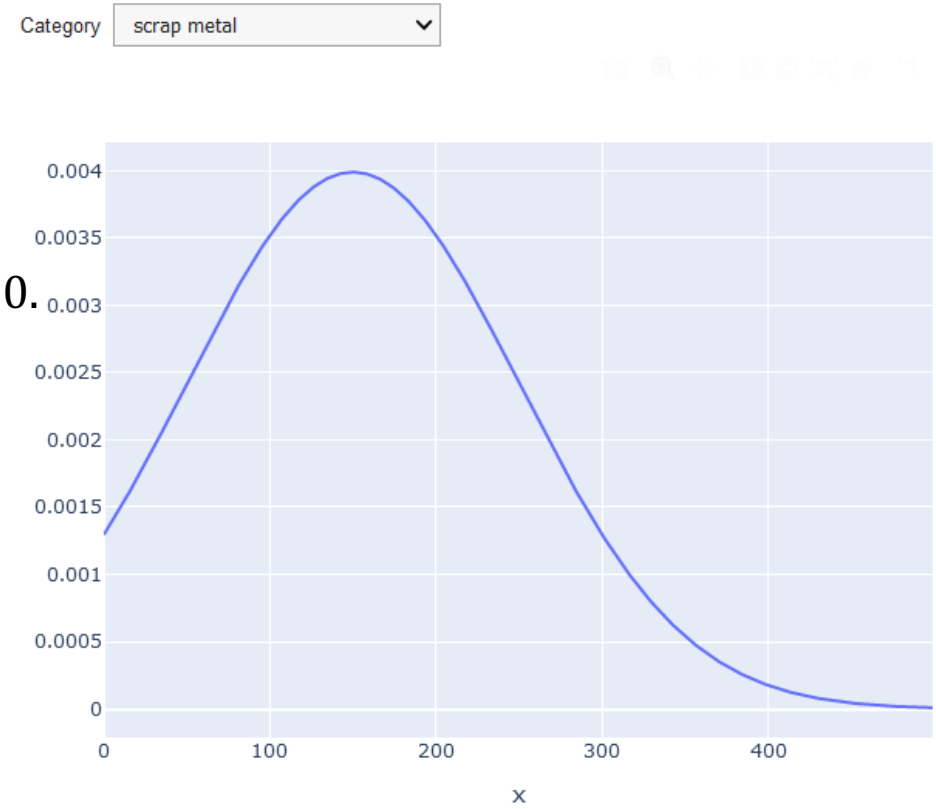
The weight sensor (aka scale)

- Weight of an object = a continuous random variable.
- We'll use the Gaussian distribution to model weight.
  - Annoyance: actually: weight can never be less than zero. Still.
- Each object has its own Gaussian distribution:

Category	Mean μ	Std dev σ
Cardboard	20	10
Paper	5	5
Can	15	5
Scrap metal	150	100
Bottle	300	200

Scrap Metal:

- Distribution centered at $\mu = 150$.
- Wide distribution.
- Notice truncation at zero chops off significant probability mass.



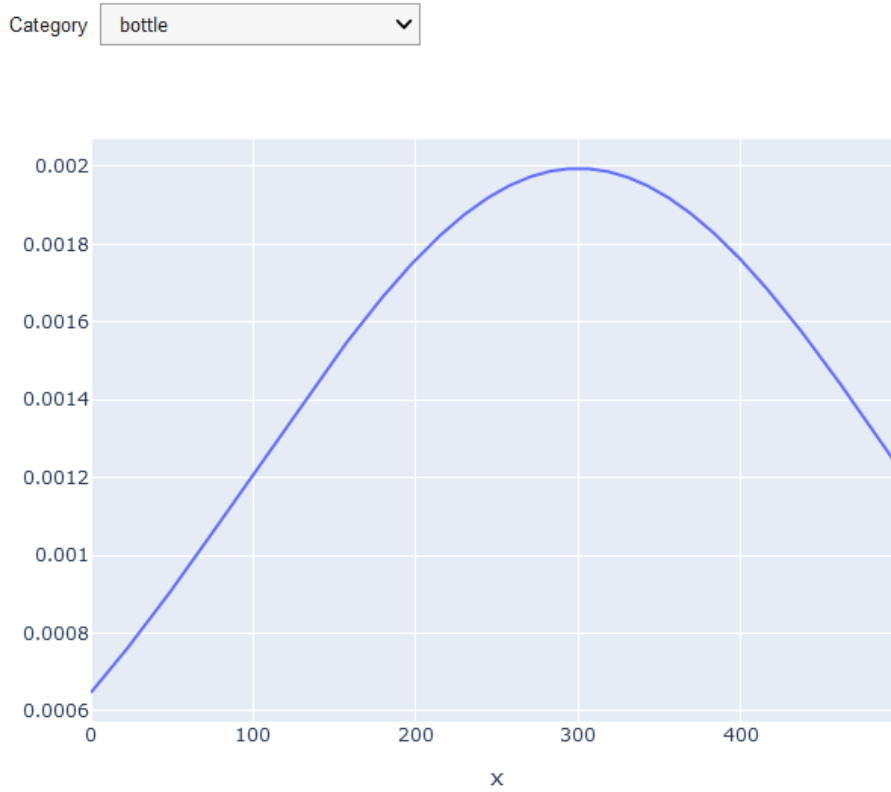
The weight sensor (aka scale)

- Weight of an object = a continuous random variable.
- We'll use the Gaussian distribution to model weight.
  - Annoyance: actually: weight can never be less than zero. Still.
- Each object has its own Gaussian distribution:

Category	Mean μ	Std dev σ
Cardboard	20	10
Paper	5	5
Can	15	5
Scrap metal	150	100
Bottle	300	200

Bottle:

- Distribution centered at $\mu = 300$.
- Wide distribution.
- Notice truncation at zero doesn't exclude much probability mass.
- Truncation on the right is merely an artifact of the display. This pdf continues all the way to $+\infty$.



Conditional distributions

- Instead of thinking about five individual pdfs for the different objects, we can think of weight as a random variable characterized by conditional probability distributions:

Category	Mean μ	Std dev σ
Cardboard	20	10
Paper	5	5
Can	15	5
Scrap metal	150	100
Bottle	300	200

Category (C)	$f_{W\|C}(W\|C)$
Cardboard	$N(20, \sigma = 10)$
Paper	$N(5, \sigma = 5)$
Can	$N(15, \sigma = 5)$
Scrap metal	$N(150, \sigma = 100)$
Bottle	$N(300, \sigma = 200)$

$$f_{X|C}(x|C = Scrap\ Metal) = \frac{1}{10\sqrt{2\pi}} e^{-\frac{(x-150)^2}{2\ 100^2}}$$

Simulation by sampling

- We can simulate the sensor readings that will occur during operation of our trash sorting robot.
- The idea is a simple extension of the sampling algorithm we developed in Section 2.1.
    1. Generate a sample category $c \sim P(C)$ using the algorithm from Section 2.1.
    2. Generate a sample sensor value by sampling the conditional distribution $s \sim f_{X|C}(x|C = c)$, where $f_{X|C}$ is the conditional density (or pmf) associated to the desired sensor.

Next Lecture: Perception, Planning, and Learning

- Bayes Theorem
  - Allows us to “invert” sensor models to obtain probabilities about the world state (and robot state).
- Maximum Likelihood Estimation
  - How to use the sensor model directly to estimate the world (or robot) state
  - A good choice if we have no prior knowledge about the world
- MAP Estimation
  - Incorporates prior knowledge
  - Provides a posterior probability distribution over world (or robot) states that takes into account both evidence (sensors) and prior knowledge.
- Decision Theory:
  - Simple risk minimization
- Learning:
  - Estimating parameters of the Gaussian distribution

Bonus: Three concepts from probability theory

In this lecture, we used a lot of ***conditional distributions***. In probability theory the following three concepts are closely related:

- Joint Distributions
- Conditional Probability
- Independence

The following slides should deepen your understanding of conditional probabilities.

Joint Probability

Consider two events, $X, Y \subset \Omega$. The joint probability of $X \; and \; Y$ is the probability that both events occur.

- When we talk about a joint probability distribution, we use the notation $\mathrm{P}(X, Y)$, indicating that $X \; and \; Y$ are random events.

- When we talk about the joint probability for two specific events, we write

$$P(X = x \text{ and } Y = y) = \; P(x, y)$$

✓ Recall, upper case denotes a random event, and lower case denotes a specific value.

An Example

Roll two dice, observe x_1 and x_2.

We know that there are 36 possible outcomes, all of which are equally likely (assuming the dice are *fair*).

It's easy to compute probabilities by simply counting outcomes:

- Probability $x_1 = 6$:

$$(6,1), (6,2), (6,3), (6,4), (6,5), (6,6) \rightarrow P = \frac{6}{36} = \frac{1}{6}$$

- Probability x_1 is even:

$$\begin{matrix}(2,1), (2,2), (2,3), (2,4), (2,5), (2,6) \\ (4,1), (4,2), (4,3), (4,4), (4,5), (4,6) \\ (6,1), (6,2), (6,3), (6,4), (6,5), (6,6)\end{matrix} \quad \rightarrow \quad P = \frac{18}{36} = \frac{1}{2}$$

An Example

Roll two dice, observe x_1 and x_2.

Now suppose we want to know the probability that two events occur.

Again, we can compute probabilities simply by counting outcomes (since all outcomes are equally probable).

- Probability $x_1 = 6$ ***and*** x_2 is even:

$$(6,2), (6,4), (6,6) \rightarrow P = \frac{3}{36} = \frac{1}{12}$$

- Probability x_1 is even and $x_1 > 3$:

$$\begin{matrix}(4,1), (4,2), (4,3), (4,4), (4,5), (4,6) \\ (6,1), (6,2), (6,3), (6,4), (6,5), (6,6)\end{matrix} \quad \rightarrow \quad P = \frac{12}{36} = \frac{1}{3}$$

Conditional Probability

- When two events are related to one another, observing the occurrence of one of the events can influence what we believe about the other.
- In this case, we talk about the conditional probability of x ***given*** y, denoted $P(x \mid y)$.

- This conditional probability is defined in terms of the joint probability of $x \; and \; y$:

$$P(x \mid y) = \frac{P(x, y)}{P(y)}$$

Assuming $P(y) \neq 0$

- *We can rewrite this expression as:*

$$P(x, y) = P(x \mid y) \, P(y)$$

This form will come in handy a bit later in the class

Independence

- If X and Y are independent, then

$$P(x, y) = P(x)P(y)$$

Definition of Independence

- If X and Y are independent, then

$$P(x \mid y) = \frac{P(x,y)}{P(y)} = \frac{P(x)P(y)}{P(y)} = P(x)$$

- *Sensors are useful because their measurements depend on the world state.*
- *However, if we have multiple sensors, quite often there are independence properties for various combinations of sensors. For example, a color sensor might give a measurement that is independent of the measurement given by a scale.*

Let's apply rules of conditional and joint probabilities:

From the previous page, we easily compute the following:

$$P(x_1\ even) = \frac{1}{2}, \qquad P(x_1 == 6) = \frac{1}{6}, \qquad P(x_2\ even) = \frac{1}{2}.$$

Let's look at some combinations of events:

- $P(x_1\ even, x_1 == 6) = \frac{1}{6} \neq P(x_1\ even)P(x_1 == 6) = \frac{1}{6} \times \frac{1}{2} = \frac{1}{12} \rightarrow$ ***NOT independent***

- $P(x_1\ even, x_2\ even) = \frac{9}{36} = P(x_1\ even)P(x_2\ even) = \frac{1}{2} \times \frac{1}{2} \rightarrow$ ***independent***

Let's apply rules of conditional and joint probabilities:

$$P(x_1 == 6 | x_1\ even) = \frac{P(x_1\ even, x_1==6)}{P(x_1\ even)} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$$

This agrees with our intuition, since $x_1 = 6$ in one third of the cases of x_1 being even:

(2,1), (2,2), (2,3), (2,4), (2,5), (2,6)
(4,1), (4,2), (4,3), (4,4), (4,5), (4,6)
(6,1), (6,2), (6,3), (6,4), (6,5), (6,6)