Intro to probability concepts

Math 10A

October 31, 2017

This year's Serge Lang Undergraduate Lecture will be given by Keith Devlin of our main athletic rival. The title is

> *When the precision of mathematics meets the*
> *messiness of the world of people.*

The lecture will be given at 4:10PM on Thursday, November 2 (this week!) in 60 Evans.

Please visit the Facebook event page for the lecture as well.

Breakfast

Next breakfast: Tomorrow at 9AM. There are still some places available. Please send me email to sign up.

Foothill DC dinner, Friday (Nov. 3) at 6:30PM.

The fundamental concepts

- Probability space
- Random variable
- Probability density function (PDF)
- Cumulative distribution function (CDF)

Flipping a coin ten times

We take a generic ("fair") coin and flip it ten times in a row. We record the result as a binary string, writing "0" for tails and "1" for heads. For example, we might write

$$01\ 01\ 11\ 00\ 10;$$

spaces have been inserted for readability.

This slide has "complex text"

Let Ω be the set of all possible "outcomes"—the possible strings that we could write down. Then Ω has $2 \times 2 \times \cdots \times 2 = 2^{10} = 1024$ elements.

Since the coin is fair, each outcome is as likely as every other outcome. The probability of getting any particular outcome is $1/1024$.

Because Ω is equipped with a probability function, it's a *probability space*.

We say that the probability on Ω is *uniform* because all outcomes are assigned the same probability.

Random variable

A random variable is a function on a probability space. As an example, consider the function

$$X : \Omega \to \{\, 0, 1, 2, \ldots, 9, 10 \,\}$$

that takes each string to the number of 1s in the string. For example,
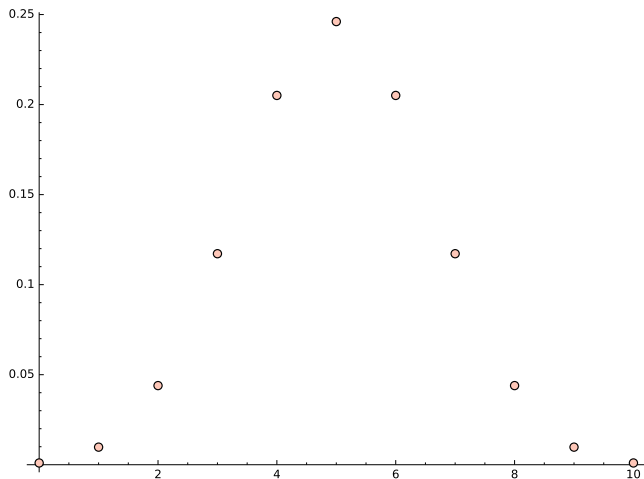
$$X(01\,01\,11\,00\,10) = 5.$$

What is the probability that $X = 5$? The answer is

(# of strings with five 1s)$/1024 = 252/1024 \approx 0.246$.

The number 252 is a *binomial coefficient*; you'll see lots of binomial coefficients in Math 10B.

Probability density function

A natural activity is to graph the function $i \mapsto P(X = i)$:

This function is the *PDF* attached to *X*.

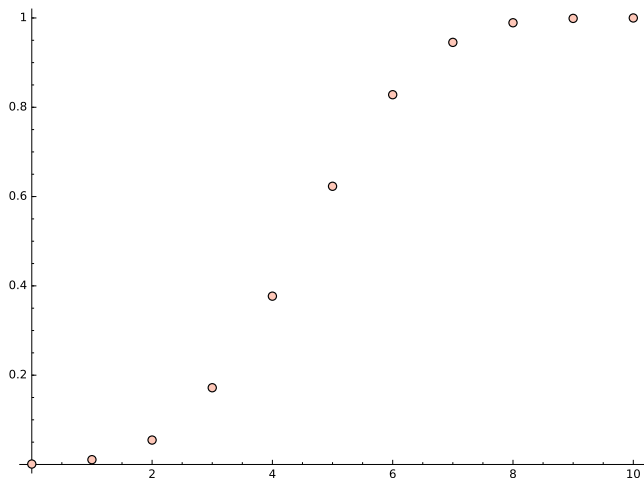*What is the probability that X is between 3 and 7?*

We add the relevant values of the function (i.e., the PDF) on the previous slide. Call the function f. Then the desired probability is

$$f(3) + f(4) + f(5) + f(6) + f(7) = 57/64 \approx 0.89.$$

The probability that X lies in a given range is obtained by summing up appropriate values of f.

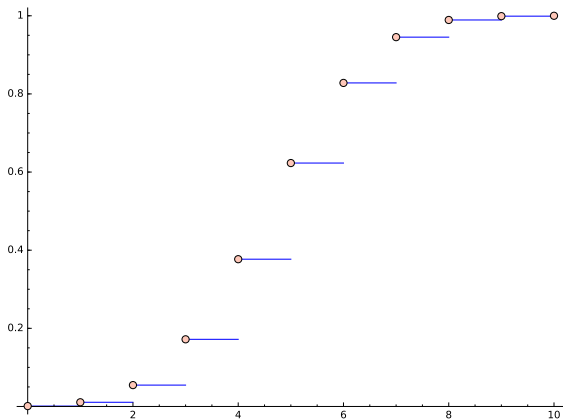# Cumulative distribution function

Consider the function $i \mapsto P(X \leq i)$:

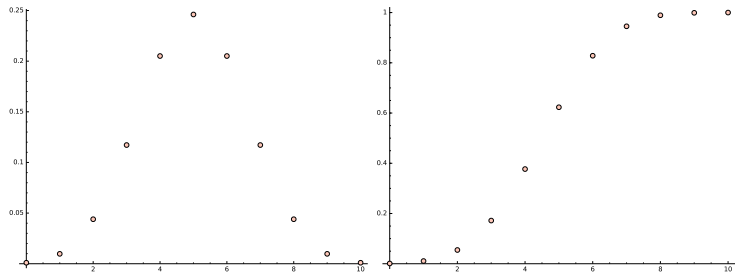This is the *CDF* attached to *X*. It might look logistic to you.

I lied, slightly

A CDF is supposed to be defined for all real numbers. The CDF in this case looks like this:

It has the value 0 to the left of 0 and the value 1 to the right of 10.

PDF and CDF

Nothing new here—I'm just showing you the CDF and the PDF side by side.

PDF and CDF

Let $f =$ the PDF, $F =$ the CDF. Then

- the PDF totals to 1: $\displaystyle\sum_i f(i) = 1$;

- the CDF is an increasing function: $F(i) \leq F(i+1)$;

- you get from the PDF to the CDF by integrating (summing):
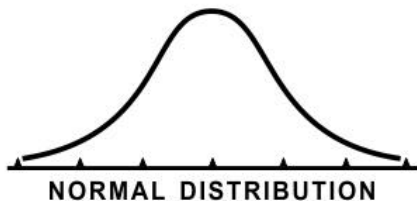
$$F(i) = \sum_{j \leq i} f(j);$$

- you get from the CDF to the PDF by differentiating (taking a difference):

$$f(i) = F(i) - F(i-1).$$

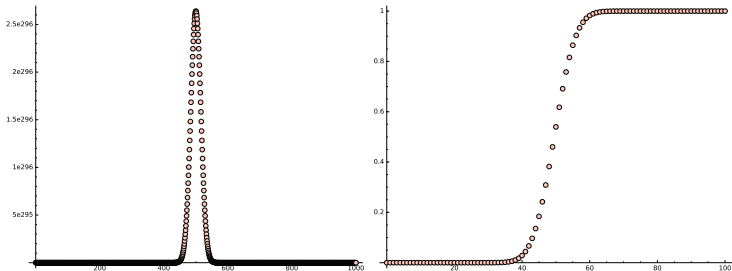Notice that the CDF and the PDF determine each other. They are different ways to package the same information.

We've had a taste of the 10B view of the world.

But we're 10A!

More tosses

Here's what happens if we toss the coin 100 times instead of 10 times:

The picture begins to look continuous. More importantly, the probability of any particular number of tosses is approaching 0. The probability of getting exactly 50 heads in 100 tosses is about 0.08. Getting 500 heads in 1000 tosses has probability about 0.025.

"Don't let students off the hook. They must face challenging reading head-on."

For us 10A folk, the PDF will be a calculus-friendly function $f(x)$ on $(-\infty, +\infty)$ with these properties:

- it's non-negative (≥ 0);
- the total area under $y = f(x)$ is 1:

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

Examples: $f(x) = \dfrac{1}{\sqrt{\pi}} e^{-x^2}, \quad f(x) = \dfrac{1}{\pi} \cdot \dfrac{1}{1 + x^2}.$

The point of a PDF is that integrals of *f* give probabilities:

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

The probability that *X* has any particular value is negligible (and 0 in the limit), but the probability that *X* lies in a range can be computed (and is non-zero when we care about it).

A "real world" example

Let Ω be the set of students who have ever taken at least one course at a UC, over the entire history of the University. Effectively, Ω is infinite.

Let

$$X : \Omega \to [0, 5]$$

be the function that assigns to each student that student's GPA, expressed as a 10-digit decimal. (I go beyond 4 because of $A+$ grades.)

The probability that X lies in $[2, 3]$ is substantial, but the probability that X has value 2.1234567890 is negligible.

The associated PDF $f(x)$ is 0 outside $[0, 5]$. What might it look like on that interval? This is an interesting thought exercise.

If *X* is a random variable, the CDF of *X* is the function

$$F(t) = P(X \leq t) = \int_{-\infty}^{t} f(x)\,dx.$$

Then:

- $F(t) \to 0$ for $t \to -\infty$;
- $F(t) \to 1$ for $t \to +\infty$;
- $F(t)$ is an increasing function of t;
- $F'(t) = f(t)$ (Fundamental theorem!).

Example
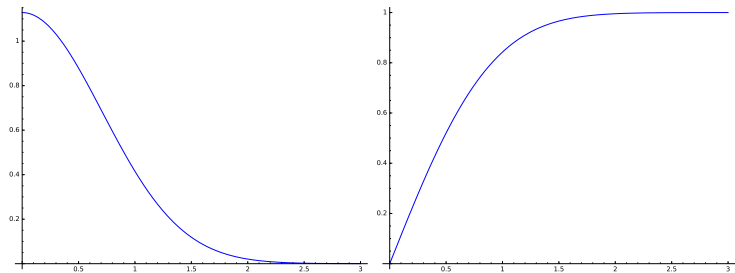
Let

$$f(x) = \begin{cases} 0 & \text{if } x < 0, \\ \dfrac{2}{\sqrt{\pi}} e^{-x^2} & \text{if } x \geq 0. \end{cases}$$

Then $f(x)$ is a PDF that is "concentrated" on the set of non-negative real numbers. It corresponds to a "Gaussian" random variable X that takes no negative values. The corresponding CDF is
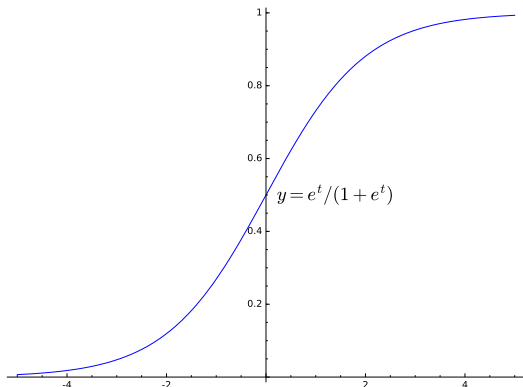
$$F(x) = \begin{cases} 0 & \text{if } x < 0, \\ \displaystyle\int_0^x \dfrac{2}{\sqrt{\pi}} e^{-t^2} \, dt & \text{if } x \geq 0. \end{cases}$$
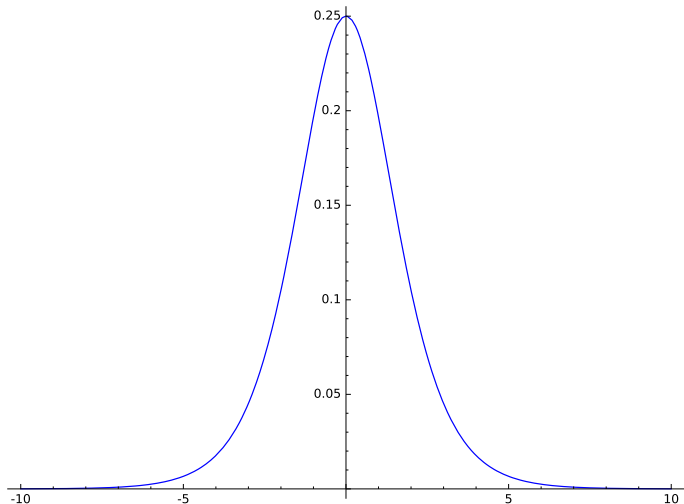
On the left side, there's the right-half of a Gaussian curve. On the right side, there's the corresponding CDF $F(x)$. The plots stop at 3. To the left of the y-axis, both functions are identically 0.

A familiar CDF

This is the graph of $y = F(x) = \dfrac{e^x}{1 + e^x}$, the logistic function.
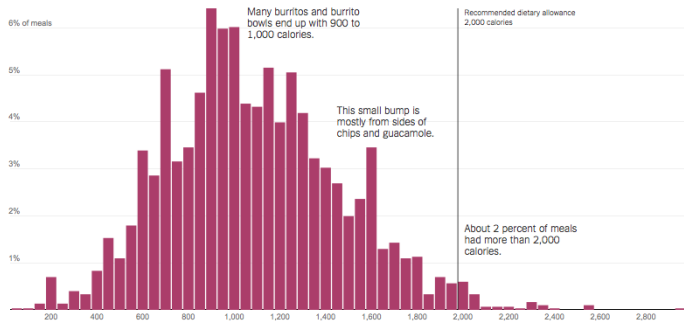We've seen it before.

The corresponding PDF is $F'(x) = F(x)(1 - F(x))$.

It's clear visually that the graph is symmetric about the vertical axis. Can you check that algebraically?

Most meals have more than 1,000 calories and almost a full day's worth of sodium.

Many burritos and burrito bowls end up with 900 to 1,000 calories.

This small bump is mostly from sides of chips and guacamole.

Recommended dietary allowance 2,000 calories

About 2 percent of meals had more than 2,000 calories.

I projected this image one year ago in Math 10A. It's from the fake @nytimes article At Chipotle, How Many Calories Do People Really Eat? If you have a meal at Chipotle, what is the probability that you will eat 400–450 calories? 2000–2050 calories? You see that $P(X > 2000) = 0.02$.

The NY Times provides a CDF

ELECTION 2016

22.5 million votes

Number of early votes cast

87,817 votes

40 days before the election

8
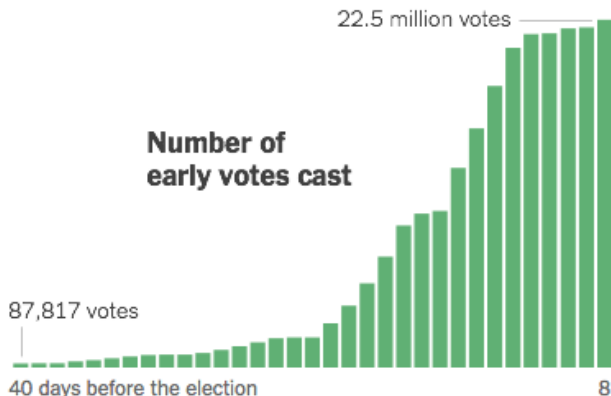
To make a true CDF, you need to scale the vertical axis so that the last bar has top at $y = 1$ rather than at $y = 22,500,000$.

Pareto

Fix $p > 1$ and note that

$$\int_1^\infty \frac{1}{x^p}\,dx = -\frac{1}{p-1}\,\frac{1}{x^{p-1}}\Big]_1^\infty = \frac{1}{p-1}.$$
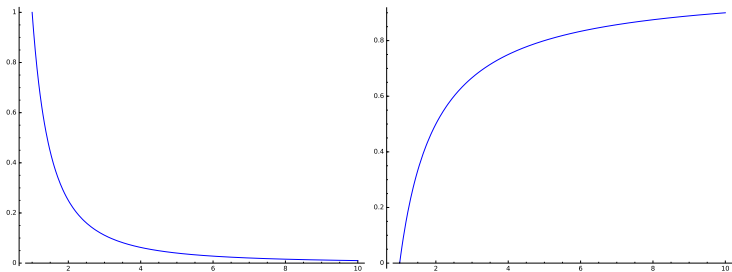
The function

$$f(x) = \begin{cases} 0 & \text{if } x < 1, \\ \dfrac{p-1}{x^p} & \text{if } x \geq 1 \end{cases}$$

is then a PDF. The CDF is

$$F(t) = (p-1)\int_1^t \frac{1}{x^p}\,dx = -\frac{1}{x^{p-1}}\Big]_1^t = 1 - \frac{1}{t^{p-1}}$$

for $t \geq 1$.

What these look like

Here $p = 2$, so the PDF is the graph of $y = \dfrac{1}{x^2}$ and the CDF is the graph of $y = 1 - \dfrac{1}{x}$. Both functions are 0 to the left of $x = 1$.

Pareto

If *X* has *f* as its PDF, then *X* satisfies the Pareto distribution.
Wikipedia:

> *The Pareto distribution, named after the Italian civil
> engineer, economist, and sociologist Vilfredo Pareto,
> is a power law probability distribution that is used in
> description of social, scientific, geophysical, actuarial,
> and many other types of observable phenomena."*