STA286 Lecture 07

Neil Montgomery

Last edited: 2017-01-25 09:59

cumulative distribution functions

Recall the cdf:

$$F(x) = P(X \leqslant x),$$

which completely decribes the distribution of X.

discrete random variables

CDFs are nice because *all* random variables have one, but they aren't the most natural representations of distributions.

For a discrete random variable X, the function that maps the *individual possible values* to their probabilities is more natural.

discrete random variables

CDFs are nice because *all* random variables have one, but they aren't the most natural representations of distributions.

For a discrete random variable X, the function that maps the *individual possible values* to their probabilities is more natural.

Example: Let X be the number of tosses of a coin until the first `H` appears. This function maps individual possible values to probabilities:

$$p(x) = f(x) = P(X = x) = \left(\frac{1}{2}\right)^x, \quad x \in \{1, 2, 3, \dots\}$$

discrete random variables

CDFs are nice because *all* random variables have one, but they aren't the most natural representations of distributions.

For a discrete random variable X, the function that maps the *individual possible values* to their probabilities is more natural.

Example: Let X be the number of tosses of a coin until the first `H` appears. This function maps individual possible values to probabilities:

$$p(x) = f(x) = P(X = x) = \left(\frac{1}{2}\right)^x, \quad x \in \{1, 2, 3, \ldots\}$$

Such a function is (best) called a *probability mass function* or pmf. I tend to use $p(x)$ notation for pmfs.

discrete random variables

CDFs are nice because *all* random variables have one, but they aren't the most natural representations of distributions.

For a discrete random variable X, the function that maps the *individual possible values* to their probabilities is more natural.

Example: Let X be the number of tosses of a coin until the first $\mathtt{H}$ appears. This function maps individual possible values to probabilities:

$$p(x) = f(x) = P(X = x) = \left(\frac{1}{2}\right)^x, \quad x \in \{1, 2, 3, \ldots\}$$

Such a function is (best) called a *probability mass function* or pmf. I tend to use $p(x)$ notation for pmfs.

Textbook notes: the book uses $f(x)$ notation, which I dislike. It also gives the following (terrible) synonyms:

- "probability function" (name already taken by P!)

discrete random variables

CDFs are nice because *all* random variables have one, but they aren't the most natural representations of distributions.

For a discrete random variable X, the function that maps the *individual possible values* to their probabilities is more natural.

Example: Let X be the number of tosses of a coin until the first `H` appears. This function maps individual possible values to probabilities:

$$p(x) = f(x) = P(X = x) = \left(\frac{1}{2}\right)^x, \quad x \in \{1, 2, 3, \ldots\}$$

Such a function is (best) called a *probability mass function* or pmf. I tend to use $p(x)$ notation for pmfs.

Textbook notes: the book uses $f(x)$ notation, which I dislike. It also gives the following (terrible) synonyms:

- "probability function" (name already taken by P!)
- "probability distribution" (name already being used for a fundamental concept!)

more pmf examples

See if a product is defective: A factory makes a defective item with probability p. Select an item at random from a factory. Let $X = 1$ if the item is defective, and let $X = 0$ otherwise.

more pmf examples

See if a product is defective: A factory makes a defective item with probability p. Select an item at random from a factory. Let $X = 1$ if the item is defective, and let $X = 0$ otherwise.

The pmf of X is:

$$p(x) = P(X = x) = \begin{cases} p & : x = 1 \\ 1 - p & : x = 0 \end{cases}$$

more pmf examples

See if a product is defective: A factory makes a defective item with probability p. Select an item at random from a factory. Let $X = 1$ if the item is defective, and let $X = 0$ otherwise.

The pmf of X is:

$$p(x) = P(X = x) = \begin{cases} p & : x = 1 \\ 1 - p & : x = 0 \end{cases}$$

More compact version: $p(x) = p^x(1-p)^{1-x}$

defining properties of pmf

A function $p(x)$ is a pmf if and only if:

1. $$p(x) \geqslant 0$$

defining properties of pmf

A function $p(x)$ is a pmf if and only if:

1. $$p(x) \geqslant 0$$
2. $$\sum_{\{x \mid P(X=x)>0\}} p(x) = 1$$

checking if a function is a valid pmf

I said this function is a pmf. Is it?

$$p(x) = f(x) = P(X = x) = \left(\frac{1}{2}\right)^x, \quad x \in \{1, 2, 3, \ldots\}$$

Verify:

1. $p(x) \geqslant 0$

checking if a function is a valid pmf

I said this function is a pmf. Is it?

$$p(x) = f(x) = P(X = x) = \left(\frac{1}{2}\right)^x, \quad x \in \{1, 2, 3, \ldots\}$$

Verify:

1. $p(x) \geqslant 0$
2. Fact: $\sum\limits_{x=0}^{\infty} ar^x = \frac{a}{1-r}$ for $0 < r < 1$. So:

$$\sum_{x=1}^{\infty} \left(\frac{1}{2}\right)^x = \sum_{x=0}^{\infty} \frac{1}{2} \left(\frac{1}{2}\right)^x = 1$$

a pmf completely characterizes a discrete distribution

I told you a cdf completely characterizes any distribution, which is a fact you'll have to take on buffy.

a pmf completely characterizes a discrete distribution

I told you a cdf completely characterizes any distribution, which is a fact you'll have to take on buffy.

A discrete random variable has a pmf. Does the pmf characterize the distribtuion?

a pmf completely characterizes a discrete distribution

I told you a cdf completely characterizes any distribution, which is a fact you'll have to take on buffy.

A discrete random variable has a pmf. Does the pmf characterize the distribtuion?

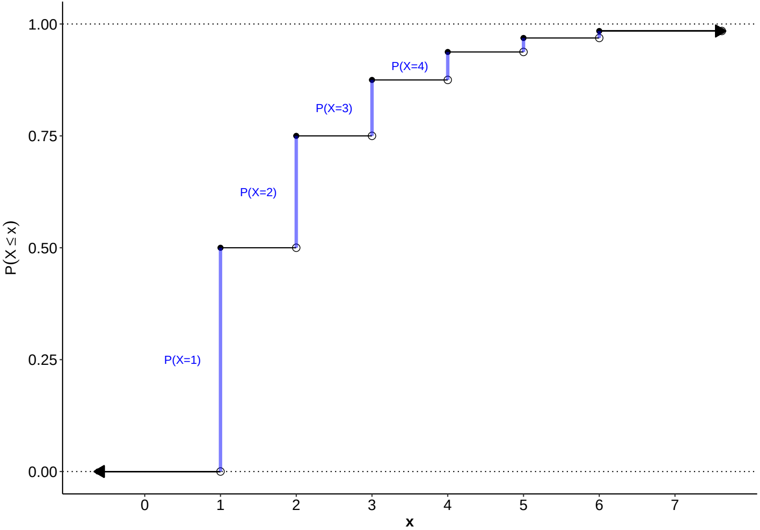
Yes, because you can compute a cdf from a pdf and vice versa. "Obviously:"

$$F(x) = \sum_{y \leqslant x} p(y)$$

For the reverse direction you take the jump points of the cdf and determine the magnitude of the jump.

possibly easier to see than to understand the formal statement

The cdf of X = "toss to first H", with pmf values in blue:



continuous random variables

For a random process taking on values in real intervals, we saw it made sense for $P(X = x) = 0$ for any particular value of x (e.g. bus stop example.)

That condition could be taken as a definition of "continuous random variable".

continuous random variables

For a random process taking on values in real intervals, we saw it made sense for $P(X = x) = 0$ for any particular value of x (e.g. bus stop example.)

That condition could be taken as a definition of "continuous random variable".

We're mainly concerned with probabilities like $P(a < X \leqslant b)$, which could be calculated using $F(b) - F(a)$, but there's another way.

continuous random variables

For a random process taking on values in real intervals, we saw it made sense for $P(X = x) = 0$ for any particular value of x (e.g. bus stop example.)

That condition could be taken as a definition of "continuous random variable".

We're mainly concerned with probabilities like $P(a < X \leqslant b)$, which could be calculated using $F(b) - F(a)$, but there's another way.

If there is a ("Riemann integrable") function f such that:

$$P(a < X \leqslant b) = F(b) - F(a) = \int_a^b f(x)\,dx$$

then we say X is "(absolutely) continuous" and has f as its *probability density function* (or pdf, or just density).

Note: a and b can be $-\infty$ or ∞.

example - bus stop

The bus comes every 10 minutes and you arrive at random “uniformly”. X is the waiting time for the bus.

example - bus stop

The bus comes every 10 minutes and you arrive at random “uniformly”. X is the waiting time for the bus.

Let:

$$f(x) = \begin{cases} \frac{1}{10} & : 0 < x < 10 \\ 0 & : \text{otherwise} \end{cases}$$

example - bus stop

The bus comes every 10 minutes and you arrive at random “uniformly”. X is the waiting time for the bus.

Let:

$$f(x) = \begin{cases} \frac{1}{10} & : 0 < x < 10 \\ 0 & : \text{otherwise} \end{cases}$$

This density gives us all the probabilities such as:

$$P(2 < X \leqslant 4) = \int_2^4 \frac{1}{10}\, dx = \frac{2}{10} \qquad P(X = 2) = \int_2^2 \frac{1}{10}\, dx = 0$$

a density completely characterizes a distribution

Since:

$$F(x) = \int_{-\infty}^{x} f(u)\, du$$

one gets $F'(x) = f(x)$, so F and f contain equivalent information.

a density completely characterizes a distribution

Since:

$$F(x) = \int_{-\infty}^{x} f(u)\, du$$

one gets $F'(x) = f(x)$, so F and f contain equivalent information.

Defining characteristics: A function f is a density as long as $f \geqslant 0$ and $\int_{-\infty}^{\infty} f(x)\, dx = 1$.

another density example

Consider:

$$f(x) = \begin{cases} e^{-x} & : x > 0 \\ 0 & : x \leqslant 0 \end{cases}$$

another density example

Consider:

$$f(x) = \begin{cases} e^{-x} & : x > 0 \\ 0 & : x \leqslant 0 \end{cases}$$

It satisfies the requirements to be a density, since $f \geqslant 0$ and:

$$\int\limits_{-\infty}^{\infty} f(x)\, dx = \int\limits_{0}^{\infty} e^{-x}\, dx = \left[-e^{-x}\right]_0^{\infty}$$

another density example

Consider:

$$f(x) = \begin{cases} e^{-x} & : x > 0 \\ 0 & : x \leqslant 0 \end{cases}$$

It satisfies the requirements to be a density, since $f \geqslant 0$ and:

$$\int_{-\infty}^{\infty} f(x)\, dx = \int_{0}^{\infty} e^{-x}\, dx = \left[-e^{-x}\right]_0^{\infty} = 1$$

Suppose X has this density. Calculate $P(X > 1)$ and determine the cdf of X...

density - meaning and interpretation

Advice: *Always* think of a density as living inside its integral.

density - meaning and interpretation

Advice: *Always* think of a density as living inside its integral.

Heuristic meaning of $f(x)$ can be:

$$f(x)\Delta x \approx \int_x^{x+\Delta x} f(x)\,dx = P(X \in (x, x+\Delta x]).$$

density - meaning and interpretation

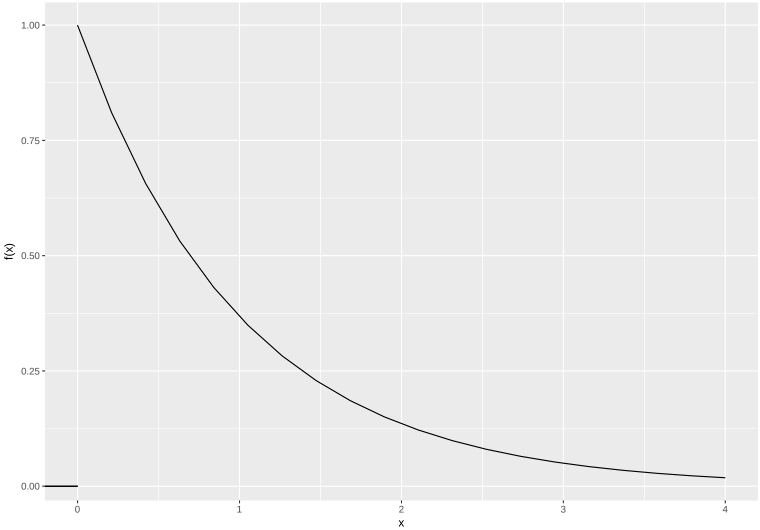
Advice: *Always* think of a density as living inside its integral.

Heuristic meaning of $f(x)$ can be:

$$f(x)\Delta x \approx \int_x^{x+\Delta x} f(x)\,dx = P(X \in (x, x+\Delta x]).$$

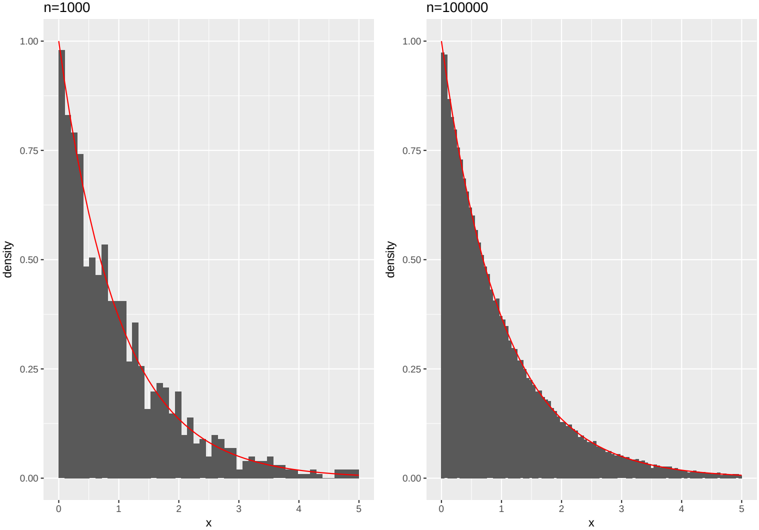
Pictures of densities can be useful, to show relative differences in probabilities.

illustration using e^{-x} density



histogram as "density estimator"

A density can be thought of as the "limit of histograms".



a note on “identically distributed”

The distribution is all we care about.

So if X_1 and X_2 have the same distributions, they are effectively the same (even if they are not the same functions.)

a note on "identically distributed"

The distribution is all we care about.

So if X_1 and X_2 have the same distributions, they are effectively the same (even if they are not the same functions.)

For example, roll a fair die so that $S = \{1, 2, 3, 4, 5, 6\}$.

a note on "identically distributed"

The distribution is all we care about.

So if X_1 and X_2 have the same distributions, they are effectively the same (even if they are not the same functions.)

For example, roll a fair die so that $S = \{1, 2, 3, 4, 5, 6\}$.

Define:

$$X_1 = \begin{cases} 1 & : \text{3 or 4 appears} \\ 0 & : \text{otherwise} \end{cases} \qquad \text{and} \qquad X_2 = \begin{cases} 1 & : \text{5 or 6 appears} \\ 0 & : \text{otherwise} \end{cases}$$

a note on "identically distributed"

The distribution is all we care about.

So if X_1 and X_2 have the same distributions, they are effectively the same (even if they are not the same functions.)

For example, roll a fair die so that $S = \{1, 2, 3, 4, 5, 6\}$.

Define:

$$X_1 = \begin{cases} 1 & : \text{3 or 4 appears} \\ 0 & : \text{otherwise} \end{cases} \qquad \text{and} \qquad X_2 = \begin{cases} 1 & : \text{5 or 6 appears} \\ 0 & : \text{otherwise} \end{cases}$$

X_1 and X_2 are not the same functions. But the have the same p.m.f.:

$$p(x) = \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{1-x}, \ x \in \{0, 1\}.$$

We say X_1 and X_2 are *identically distributed*.