Week 5: 19/2/2020
21/2/2020

1.4 Continuous Random Variables

Recap (Week 3) — *Discrete Random Variables*

- A *random variable* X assigns values to outcomes in the sample space.
- A *discrete probability distribution* gives the values that a discrete random variable X may assume and their probabilities.
- Must satisfy $0 \leq P(X = x_i) \leq 1$ and $\sum P(X = x_i) = 1$.
- The *mean* and *variance* of X are given by

E( ) is "expected" value

$$\boxed{\begin{aligned} E(X) &= \sum x_i \times P(X = x_i) \\ E(X^2) &= \sum (x_i)^2 \times P(X = x_i) \\ \text{var}(X) &= E(X^2) - (E(X))^2 \end{aligned}}$$

E(X), E(X²) and var(X) are numbers

Examples

- Bernoulli distribution (H or T)
- Geometric distribution (num tosses until H)
- Binomial distribution (num H from n tosses)
- Each distribution has corresponding R functions

Week 5 lab Worksheet

  - `dgeom`, `dbinom` (probability "density" $P(X = x)$)
  - `pgeom`, `pbinom` (cumulative probability $P(X \leq x)$)
  - `qgeom`, `qbinom` (quantile, given p find $P(X \leq q) = p$)
  - `rgeom`, `rbinom` (random variate generation)

Image from
http://www.cloudworksmg.com/wp-content/uploads/2014/12/whats-big-idea.png

New idea — A **continuous random variable** is a random variable which can take on *any value* in a given range.

Example 1.25 Heights and weights of people, magnitudes of earthquakes, levels of reservoirs, blood cholesterol levels, etc. □

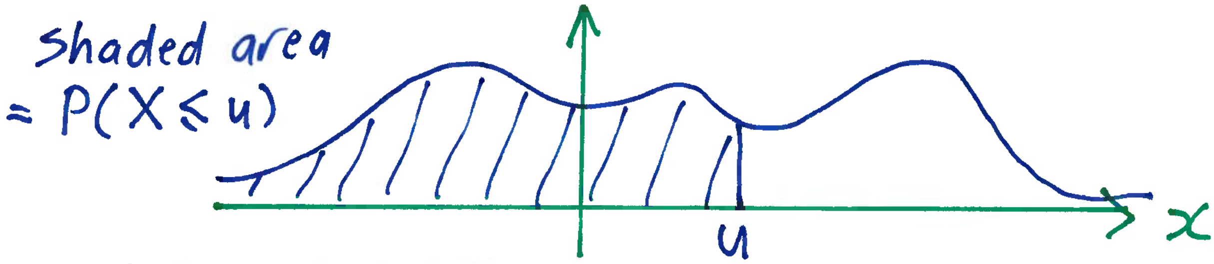
■ We **cannot** define probabilities for a continuous random variable in the same way as for a discrete random variables.

- If $X \sim \text{Binomial}(n, p)$ then X is **discrete** and takes on the values $0, 1, 2, \ldots, n$. Hence we can easily write down the probabilities $P(X = 0)$, $P(X = 1)$, ..., $P(X = n)$.
- Suppose Y is a continuous random variable which can take on any value in the real numbers.
  - Y has an **uncountably infinite** set of possible values.
  - We **cannot list** probabilities which describe the behaviour of Y.
  - Instead it turns out that the best way to describe the behaviour of Y is to look at probabilities of the form

$$* \quad \boxed{P(Y \leq y)} \quad *$$

Warning! $P(Y = y) = 0$

no chance of observing any particular value



1.4.1 Probability Distribution (PDF and CDF)

■ For many continuous random variables X there is a function $f(x)$ such that

on or above x-axis

- $\boxed{f(x) \geq 0}$ for all real numbers x; and
- $P(X \leq u)$ is given by the $\boxed{\textit{area under } f(x)}$ above the x-axis, and to the left of $x = u$.

The function $f(x)$ is called the

probability density function (pdf) of X.

The function $F(u) = P(X \leq u)$ is called the

cumulative distribution function (cdf) of X

$u \longrightarrow \boxed{\text{cdf}} \longrightarrow$ probability $P(X \leq u)$

► If you have studied some calculus, $P(X \leq u)$ is given by a definite integral, i.e.,

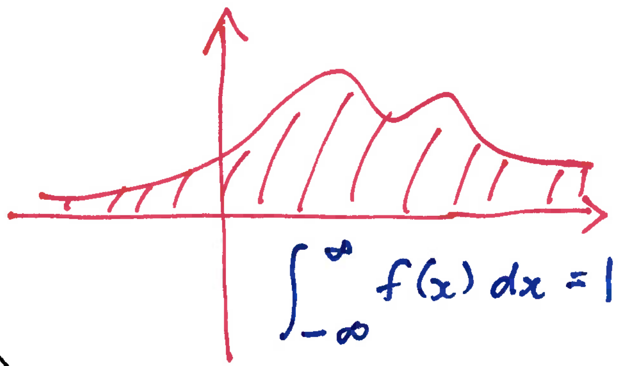
$$P(X \leq u) \;=\; \int_{-\infty}^{u} f(x)\,\mathrm{d}x$$

Not expected to know any calculus.

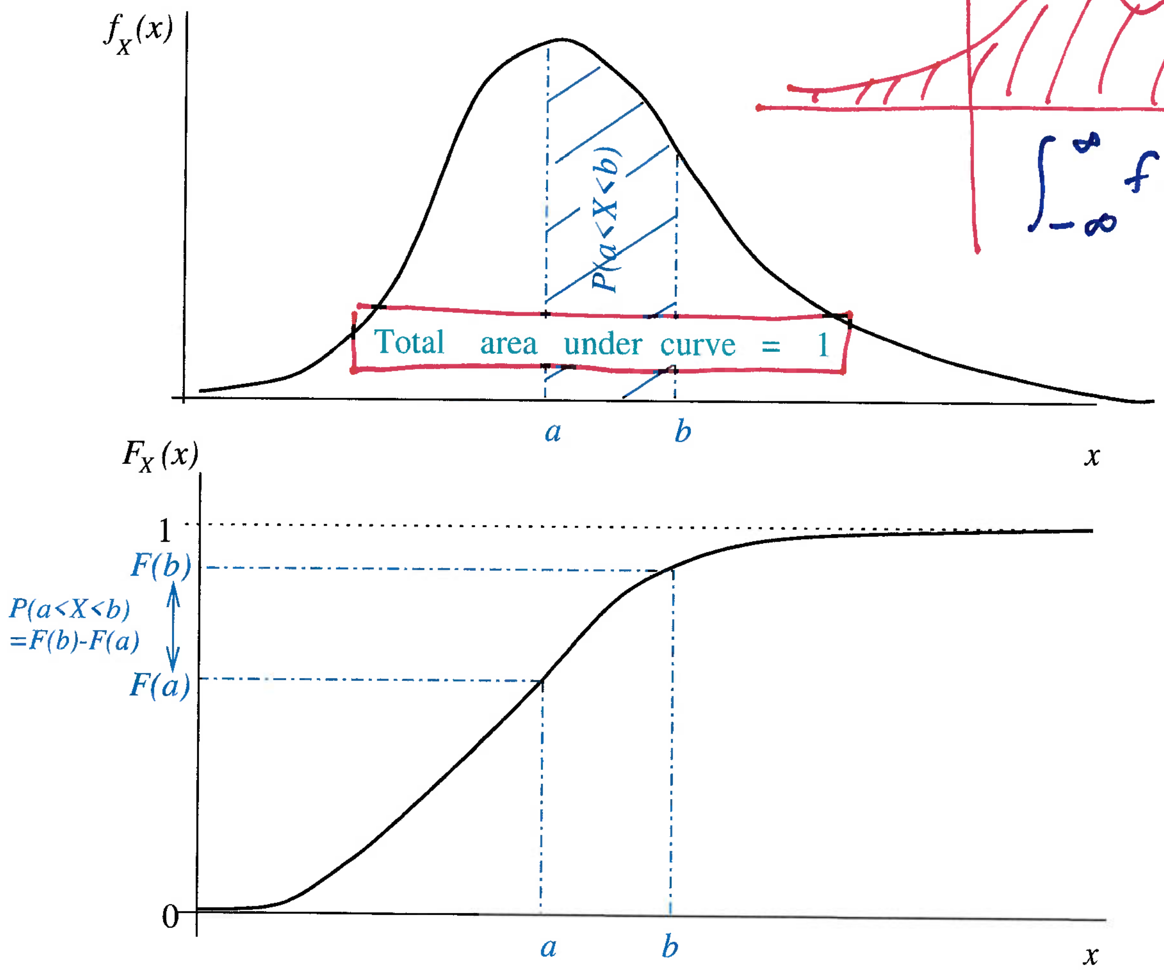
$\int_{-\infty}^{u}$ roughly means area under from $x = -\infty$ to $x = u$

■ For any continuous random variable X with probability density function (pdf) $f(x)$

- The area under $f(x)$ and above the x-axis is 1.
- $P(a < X < b) \;=\; F(b) - F(a)$

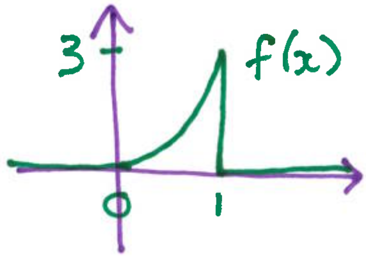


$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$



pdf002

Example 1.26 Let X be a continuous random variable with pdf given by



$$\text{pdf: } f(x) = \begin{cases} 3x^2 & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

so that

$$\text{cdf: } F(u) = \begin{cases} 0 & u < 0 \\ u^3 & 0 \leq u \leq 1 \\ 1 & u > 1 \end{cases}$$

Then probabilities are calculated using the cdf $F(u)$, e.g.,

$$P(X \leq \tfrac{1}{2}) = F(\tfrac{1}{2}) = (\tfrac{1}{2})^3 = \tfrac{1}{8}$$

$$P(X \geq \tfrac{2}{3}) = 1 - F(\tfrac{2}{3}) = 1 - (\tfrac{2}{3})^3 = 1 - \tfrac{8}{27} = \tfrac{19}{27}$$

$$P(X \leq 3) = F(3) = 1$$

Notice that

$$P(X = \tfrac{5}{6}) = P(\tfrac{5}{6} \leq X \leq \tfrac{5}{6}) = F(\tfrac{5}{6}) - F(\tfrac{5}{6}) = 0$$

□

For any x always have

$$P(X = x) = 0$$

when X is a continuous random variable

Checklist

CDF $F(u)$ ✓ is a probability

- **Is a probability**, i.e., $F(u) = P(X \leq u)$.
- So must satisfy $0 \leq F(u) \leq 1$ for all u.
- Whenever $a \leq b$ we have $P(X \leq a) \leq P(X \leq b)$ so

$$F(a) \leq F(b) \qquad \text{whenever } a \leq b$$

  i.e., $F(u)$ is a **non-decreasing** function of u.

- Since $P(X \leq -\infty) = 0$ and $P(X \leq \infty) = 1$ we must have

$$\lim_{u \to -\infty} F(u) = 0 \qquad \text{and} \qquad \lim_{u \to \infty} F(u) = 1$$

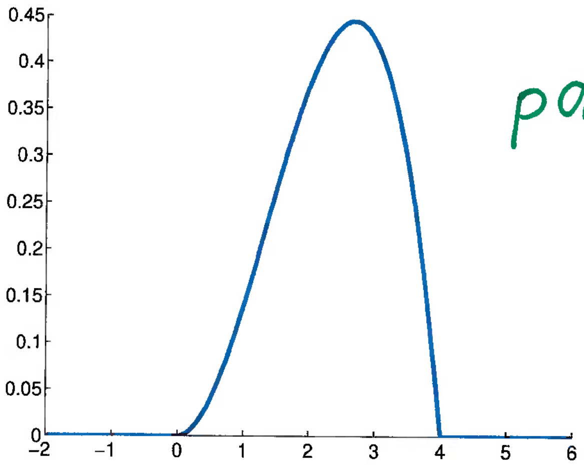
PDF $f(x)$ ✗ not a probability (just a curve)

- Is **not** a probability and so $f(x) > 1$ is possible (and ok).
- Must have $f(x) \geq 0$ for all x.
- Must be a **piece-wise continuous** function.
- The probability $P(a \leq X \leq b)$ is given by the **area under** $f(x)$ and above the x-axis, between $x = a$ and $x = b$
- Since the probability that X lies in $(-\infty, \infty)$ must be 1, the area between $f(x)$ and the x-axis must be 1.

Example 1.27 A company has been monitoring its daily telephone usage. The daily use of time conforms to the following pdf (measured in hours).

$$f(x) = \begin{cases} \frac{3}{64}x^2(4-x) & 0 \le x \le 4 \\ 0 & \text{otherwise} \end{cases}$$

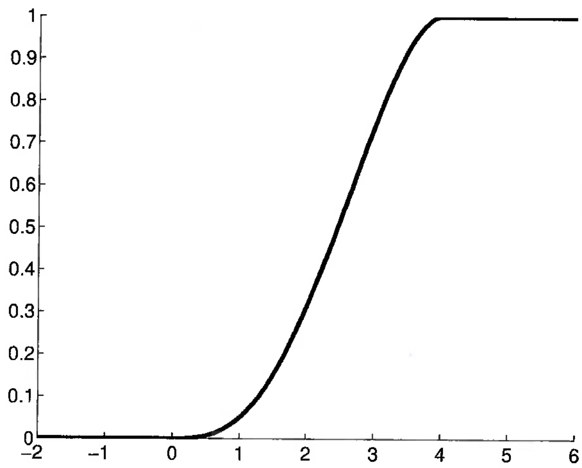
Sketch of $f(x)$



pdf

The cdf for daily telephone usage X is

$$F(u) = \begin{cases} 0 & u < 0 \\ -\frac{3}{256}u^4 + \frac{1}{16}u^3 & 0 \le u \le 4 \\ 1 & u > 4 \end{cases}$$



cdf

Supppose the current budget of the company covers only 3 hours of daily telephone usage. What is the probability that the budgeted figure is exceeded?

$$\begin{aligned} P(X > 3) &= 1 - P(X \leq 3) \\ &= 1 - F(3) \\ &= 1 - (-\tfrac{3}{256}(3)^4 + \tfrac{1}{16}(3)^3) \\ &= \frac{67}{256} \\ &= 0.2617 \quad (4dp) \end{aligned}$$

□

To find a probability

either plug into $F(x)$ [cdf]

or area under $f(x)$ [pdf]

↳ then often need calculus

1.4.2 Mean and Variance

Compare the following definitions with those for a *discrete* random variable (see page 1061).

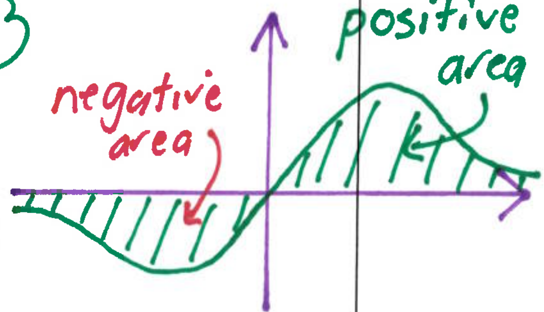
$E(x) =$ value × probability + value × probability + ...

■ For a *continuous* random variable X with pdf $f(x)$ the

expected value of X

(also called the **mean of** X) is

$$E(X) \;=\; \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x$$



This is the $\boxed{\textbf{signed area}}$ between the curve $xf(x)$ and the x-axis, i.e., positive area wherever $xf(x) > 0$ and negative area wherever $xf(x) < 0$.

Notes —

Note → $f(x)$ always ≥ 0

- X is a random variable and $E(X)$ is a number.
- The expected value of a function g of a random variable X is defined to be

$$E(g(X)) \;=\; \int_{-\infty}^{\infty} g(x)\,f(x)\,\mathrm{d}x$$

- For example, using $g(x) = x^2$

$$E(X^2) \;=\; \int_{-\infty}^{\infty} x^2\,f(x)\,\mathrm{d}x$$

= area under $x^2 f(x)$ curve

- If X is a continuous random variable with pdf $f(x)$ then

$$E(aX+b) \;=\; aE(X)+b$$

for any real numbers a and b, i.e., *expectation is a linear operator.*

■ For a continuous random variable X with pdf $f(x)$

$$\text{var}(X) \;=\; E((X-\mu)^2) \;=\; E(X^2)-\mu^2$$

where $\mu = E(X)$.

This is the same as for a discrete random variable. ✓

$$var(X) = E(X^2) - (E(x))^2$$

Example 1.28 Let X be a continuous random variable with pdf given by (see Example 1.26 on page 1065)

$$f(x) \;=\; \begin{cases} 3x^2 & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

If you already know a little bit of calculus then

$$\left.\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x\, f(x)\,\mathrm{d}x \;=\; \int_0^1 x\,3x^2\,\mathrm{d}x \\ &= \int_0^1 3x^3\,\mathrm{d}x \;=\; \left[\tfrac{3}{4}x^4\right]_0^1 \;=\; \tfrac{3}{4} \end{aligned}\right\}$$

If you already know how to integrate

Don't panic

Not expecting you to know any calculus.

$$
\begin{aligned}
E(X^2) &= \int_{-\infty}^{\infty} x^2 f(x)\,\mathrm{d}x \;=\; \int_0^1 x^2\, 3x^2\,\mathrm{d}x \\
&= \int_0^1 3x^4\,\mathrm{d}x \;=\; \left[\tfrac{3}{5}x^5\right]_0^1 \;=\; \tfrac{3}{5} \\
\mathsf{var}(X) &= E(X^2) - (E(X))^2 \\
&= \tfrac{3}{5} - (\tfrac{3}{4})^2 \;=\; \tfrac{3}{80}
\end{aligned}
$$

Since we have access to software . . .

If you are interested

- WolframAlpha —

  `www.wolframalpha.com/calculators/integral-calculator/`

  ```
  integrate x*3*x^2 from 0 to 1
  integrate x^2*3*x^2 from 0 to 1
  ```

- SymPy (Python-based) — $\boxed{\texttt{live.sympy.org}}$ ← website

```
f = 3*(x**2)
E = integrate(x*f, (x,0,1))
E2 = integrate((x**2)*f, (x,0,1))
var = E2 - E**2
```

$\to 3x^2$

$\int_0^1 x\,(3x^2)\,dx$

$\int_0^1 x^2(3x^2)\,dx$

a**2 is a^2 in Python □