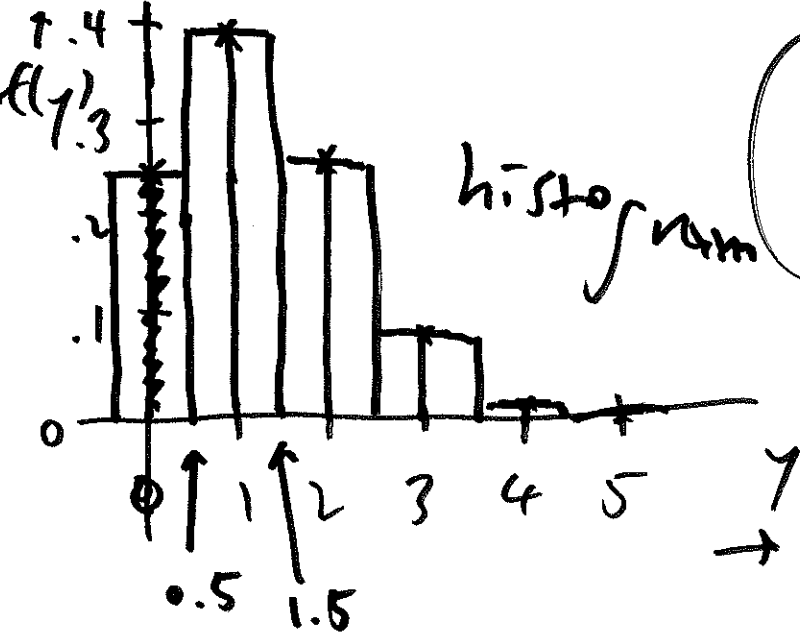
every measurement you ever make is in actuality discrete, but it's useful to regard rvs that are <u>conceptually</u> continuous (i.e., no limit in principle to the precision of measurement) as continuous.

(23 Apr 19) $\boxed{\text{Definition}}$ Given a <u>discrete</u> rv Y, the <u>probability</u> (mass) <u>function</u> (pmf or pf) of Y is the function f_Y that keeps track of the probabilities associated with Y: $f_Y(y) = P(Y = y)$.

The set $\{ y : f_Y(y) > 0 \}$ is called the <u>support</u> of (the distribution of) Y.

(DS is almost unique in using "pf"; nearly everybody talks about the pmf.)



histogram

pf of Y in the T-S case study

Definition

A rv Y that only takes on the values $\{0, 1\}$ — i.e., a <u>binary</u> rv — is said to have a <u>Bernoulli</u> distribution with <u>parameter</u> p — written Bernoulli(p) —

James Bernoulli Swiss (1655–1705)

if $$f_Y(y) = P(Y = y) = \begin{cases} p & \text{for } y = 1 \\ 1-p & y = 0 \\ 0 & \text{else} \end{cases}$$
$$= p^y (1-p)^{1-y}$$

Notation Y follows a Bernoulli(p) distribution $\leftrightarrow$ $Y \sim$ Bernoulli(p) or $(Y \mid p)$

$\sim$: is distributed as

Example In the powerball lottery (see ~~[illegible]~~ take home test problem 2) 5 white numbered balls are drawn at random without replacement from a bin with balls numbered $\{1, 2, \ldots, 69\}$.

Let $\underline{W}_i$ = # on ith drawn white ball ($i = 1, \ldots, 5$).

Clearly
$$p(\underline{W}_1 = w_1) = \begin{cases} \frac{1}{69} & \text{for } w_1 = 1, 2, \ldots, 69 \\ 0 & \text{otherwise} \end{cases}.$$

Less clearly (but true) $\underline{W}_2, \ldots, \underline{W}_5$ follow the same distribution if nothing is known about the previous draws.

Definition For any two integers $a \leq b$ ($<$), a rv $\underline{Y}$ that's equally likely to be any of the values $\{a, a+1, \ldots, b\}$ has the uniform distribution Uniform$\{a, b\}$. Evidently its pf is $f_Y(y) = P(Y = y) = \begin{cases} \dfrac{1}{b-a+1} & \text{for } y = a, \dots, b \\ 0 & \text{else} \end{cases}$

~~"is distributed as"~~

$Y \sim \text{Uniform}\{a, b\} \leftrightarrow Y$ chosen at random from $\{a, a+1, \dots, b\}$.

Definition: n random trials are performed, with each trial recorded as a success S or failure F. If each trial is independent of all the others and the chance p of success is constant across the trials, then $Y = $ # of successes

not sample space

has the Binomial distribution

pf

$$f_Y(y) = P(Y = y) = \begin{cases} \binom{n}{y} p^y (1-p)^{n-y} & \text{for } y = 0, 1, \dots, n \\ 0 & \text{else} \end{cases}$$

(with parameters n and p)

In shorthand $Y \sim \text{Binomial}(n, p)$.
or $(Y|n,p)$

Let $B_i = \begin{cases} 1 & \text{if trial } i \text{ is a success} \\ 0 & \quad . \quad . \quad . \quad \text{failure} \end{cases}$

for $i = 1, \dots, n$; then under these assumptions

$B_i \overset{IID}{\sim} \text{Bernoulli}(p)$ and all the B_i are

independent.

Notation $X_i \overset{IID}{\sim} f_{X_i}(x_i)$

means that all of the rvs $X_1, X_2, \dots$

are independent and identically distributed

draws from the distribution with pf

$f_{X_i}(x_i)$. Thus with the success/failure

trials, $B_i \overset{IID}{\sim} \text{Bernoulli}(p)$ and
$(i = 1, \dots, n)$

$\left(Y = \sum_{i=1}^{n} B_i \right) \sim \text{Binomial}(n, p)$.

This is our first example of the distribution of the <u>sum</u> of a bunch of IID rvs, a topic we'll examine in detail later.

Continuous random variables

Example (Round-off error in computer science)

Single-precision decimal floating point numbers carry about 7 sigfigs of accuracy,

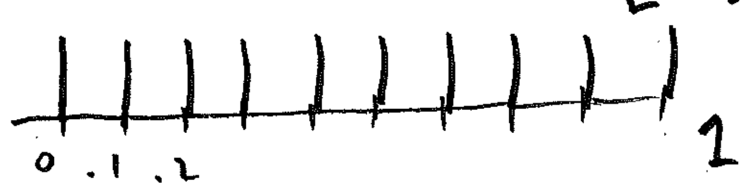
3.141592 | 653589
π 3 ~ .04 error

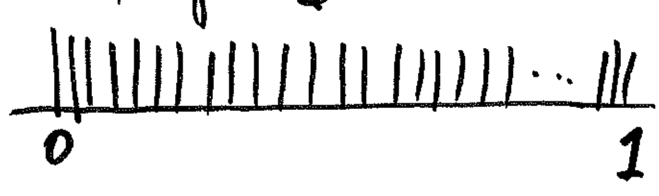
leading to <u>roundoff error</u> in the last digit; it's important to study how these errors accumulate as the number of steps in a calculation increases.

Since there's no reason one decimal digit would be favored over another in rounding, the <u>uniform</u> distribution is key to these calculations.

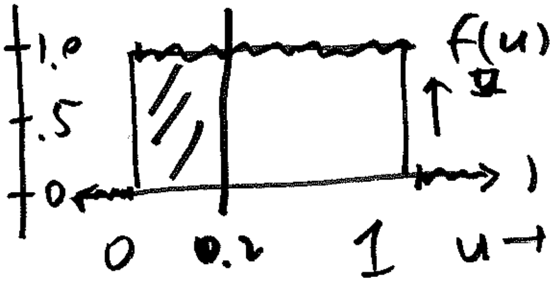
Consider first Uniform$\{0, 0.1, \dots, 0.9\}$ and then $\{0, .01, .02, \dots, .99\}$

← discrete pf →





In the limit with more & more sig figs this should go to



the continuous uniform distribution Uniform $(0,1)$ on the unit interval.

The analogue of the discrete pf in this continuous case is the smooth function

analogue of summation is integration

$$f_{\underline{Y}}(u) = \begin{Bmatrix} 1 & \text{for } 0 \le u \le 1 \\ 0 & \text{else} \end{Bmatrix}.$$

Definition

A random variable

($\underline{Y}$ has a continuous distribution) $\leftrightarrow$ ($\underline{Y}$ is a continuous rv) if there exists a continuous nonnegative function $f_{\underline{Y}}$ defined on $\mathbb{R}$ such that for every interval $[a,b]$, $P(a \le \underline{Y} \le b) = \int_a^b f_{\underline{Y}}(y)\,dy$.

In this definition, a can be $-\infty$ and b can be $+\infty$.

Definition If Y is a continuous rv, the function f_Y in the previous definition is called the <u>probability density function</u> (pdf) (PDF) of Y. The set $\mathcal{Y} = \{y : f_Y(y) > 0\}$ is called the <u>support</u> of (the distribution of) Y.

Clearly (a) $f_Y(y) \geq 0$ for all y and (b) $\int_{-\infty}^{\infty} f_Y(y)\,dy = 1$.

What about individual points — <u>singletons</u> — $\{y\}$ on $\mathbb{R}$?

You'll recall from calculus that if f_Y is continuous on its support, $\int_a^b f_Y(y)\,dy$ can equally well stand for $P(a \leq Y \leq b)$ or $P(a < Y \leq b)$ or $P(a \leq Y < b)$ or $P(a < Y < b)$, because (e.g.) $\int_a^a f_Y(y)\,dy = 0$ if f_Y is continuous at $y = a$. Thus, importantly, $\boxed{P(Y = y) = 0 \text{ for all } -\infty < y < \infty}$

weirdly, this doesn't mean that the value y of Y is impossible, as it does with discrete rv; it just means that singletons have to have 0 probability (otherwise $\int_{-\infty}^{\infty} f_Y(y)\,dy = +\infty$ not 1).

Definition With a and b any two real numbers satisfying $a < b$,

$$Y \underset{\text{is distributed as}}{\sim} \text{Uniform}(a, b) \overset{\text{iff}}{\longleftrightarrow} P\left(\begin{array}{c} Y \text{ is in any subinterval} \\ \text{of } (a, b) \end{array} \right)$$

$$= \text{the length of the subinterval} \overset{\text{iff}}{\longleftrightarrow}$$

$$f_Y(y) = \left\{ \begin{array}{ll} \frac{1}{b-a} & \text{for } a \le y \le b \\ 0 & \text{else} \end{array} \right\}.$$

Definition The <u>indicator function</u> for any (true/false) proposition A is

$$I(A) = \left\{ \begin{array}{ll} 1 & \text{if } A \text{ true} \\ 0 & \text{if false} \end{array} \right\}$$

People sometimes also write (with A a set $\subset \mathbb{R}$)

$$I_A(y) = \left\{ \begin{array}{ll} 1 & \text{if } y \in A \\ 0 & \text{else} \end{array} \right\}.$$

With this definition, $Y \sim \text{Uniform}(a,b)$ (37)

$$\leftrightarrow f_Y(y) = \frac{1}{b-a} I(a \le y \le b) = \frac{I_{[a,b]}(y)}{b-a}.$$

Contrast $Y \sim \text{Uniform}(a,b)$ continuous and uniform on (a,b) or $[a,b)$ or $(a,b]$ or $[a,b]$

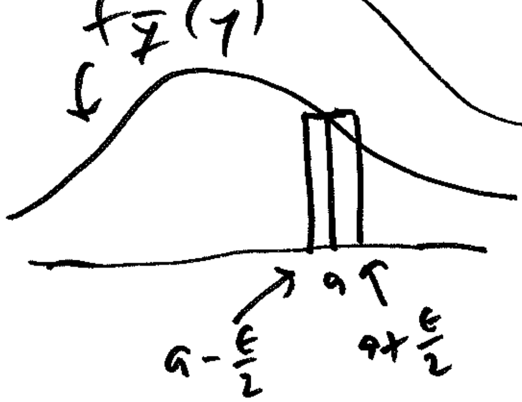
versus

$Y \sim \text{Uniform}\{a,b\}$ for a,b integers with $a<b$ $\leftrightarrow$ Y discrete and uniform on $\{a, a+1, \ldots, b\}$.

Density and probability are not the same thing

Density values $f_Y(y)$ are themselves not probabilities; for example, they can easily be > 1 & can even be $+\infty$, as we'll see later.

Density values define probabilities: $P(a \le Y \le b) = \int_a^b f_Y(y)\,dy$.



For small $\epsilon > 0$ you can see from this sketch that

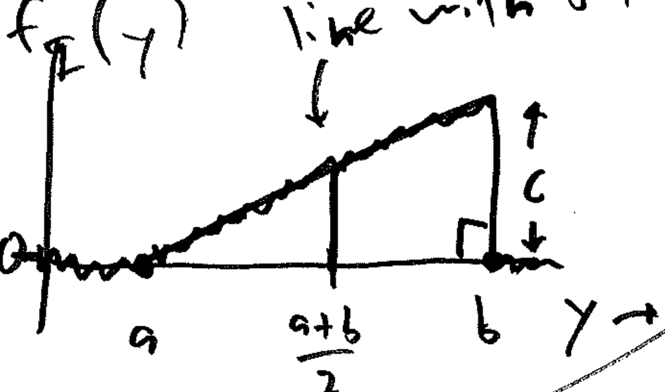
$$P\left(a-\frac{\epsilon}{2} \le Y \le a+\frac{\epsilon}{2}\right) = \int_{a-\frac{\epsilon}{2}}^{a+\frac{\epsilon}{2}} f_Y(y)\,dy$$

(connect to histograms, §3.6.5, §3.3)

$$\doteq \text{area of rectangle}$$

$$= \epsilon \cdot f_Y(a).$$

Example (a triangular distribution)



Can a continuous rv Y have a pdf that looks like a triangle? Let's see what, if any, restrictions would be needed.

The line in the sketch has slope $\frac{c}{b-a} = m$ (29)

and passes through the point $(x_1, y_1) = (a, 0)$, so

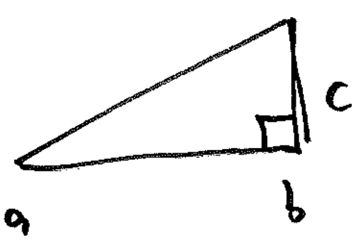
the equation of the line is

$f_Y(y)$

$y - y_1 = m(x - x_1) \leftrightarrow y = \frac{c}{b-a}(x - a) = f_X(x)$

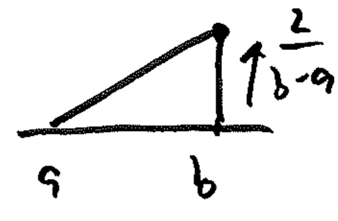
Densities have to integrate to 1,

so $\int_a^b \frac{c}{b-a}(x-a)\,dx = 1 \leftrightarrow c = \frac{2}{b-a}$



Easier way: area of a triangle is $\frac{1}{2}$(base)(height), so

$1 = \frac{1}{2}(b-a)c$ and $c = \frac{2}{b-a}$



(25 Apr 18)