y	$P(Y = y)$
0	0.237
1	0.396
2	0.264
3	0.088
4	0.015
5	0.001

(see p. 27)

You can see that a rv Y is completely specified by two things: the values it can take on, and the probabilities for those values.

Definition: The (probability) distribution of a random variable Y is the collection of all probabilities of the form $P(Y \in A)$ for all sets A of real numbers in the non-weird collection $\mathcal{C}_{\mathbb{R}}$ of subsets of the real number line $\mathbb{R}$.

The rv Y in the T-S case study has a finite set of possible values —

this is true of some, but not all, rvs.

Definition A random variable Y has a <u>discrete distribution</u>, or equivalently Y is a <u>discrete</u> rv, if the set of possible values (distinct) of Y is finite or at most countably infinite; rvs for which the set of possible values is uncountable are called <u>continuous random variables</u>.

Examples (1) The rv $X = \begin{cases} 1 & \text{if } Y > 0 \\ 0 & \text{otherwise} \end{cases}$ (with Y = # T-s babies) is discrete, taking on only the values $\{0, 1\}$ – such rvs are called <u>dichotomous</u> $\{\text{yes, no}\}$ or <u>binary</u> $\{1, 0\}$.

(2) Imagine a scale for weighing things that has a dial you can set to specify how many significant figures (sigfigs) of precision you want. Buy a "1 pound" package of butter at your favorite market and weigh it!

possible weights (ounces)
16
16.0
15.99
15.993
15.9928
⋮

If there's no conceptual limit to the number of sigfigs you could get,

a rv $Y = \begin{pmatrix} \text{the actual (true)} \\ \text{weight of the package} \end{pmatrix}$

should be modeled as continuous, having values (e.g.) on $(0, \infty)$, the positive part of $\mathbb{R}$. Reality check: Infinite precision is impossible in practice;

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.

$\boxed{\text{Definition}}$ Given a <u>discrete</u> rv Y, the <u>probability (mass) 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.)

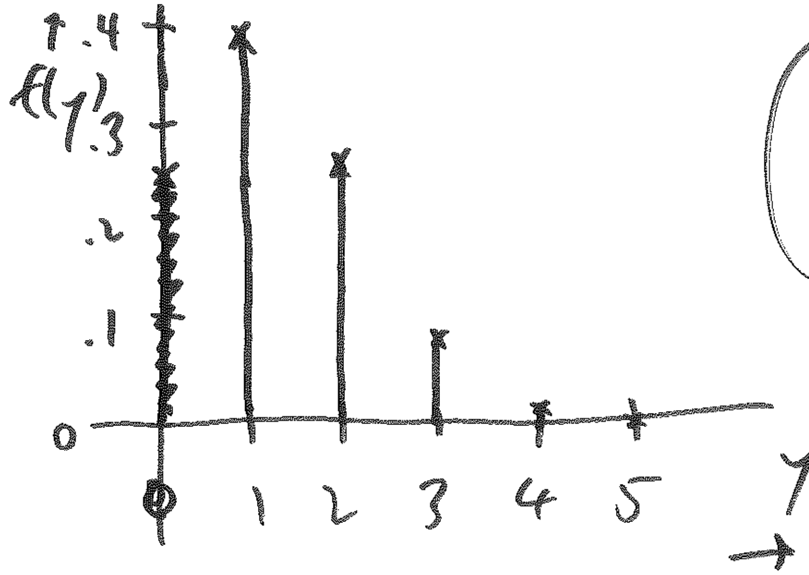
Example | In the powerball lottery (see homework 1 problem 2) 5 white numbered balls are drawn at random without replacement from a bin with balls numbered $\{1, 2, \ldots, 69\}$.

Let W_i = # on i^{th} drawn white ball.

Clearly $p(W_1 = w_1) = \begin{cases} \frac{1}{69} & \text{for } w_1 = 1, 2, \ldots, 69 \\ 0 & \text{otherwise} \end{cases}$.

Less clearly (but true) $W_2, \ldots, W_5$ follow the same distribution if nothing is known about the previous draws.

Definition | For any two integers $a \overset{(<)}{\leq} b$, a rv Y that's equally likely to be any of the values $\{a, a+1, \ldots, b\}$ has the uniform distribution Uniform$\{a, b\}$. Evidently



pf of Y in the T-S case study

Definition

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

James Bernoulli Swiss (1655–1705)

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

$$= p^y (1-p)^{1-y}$$

Notation: Y follows a Bernoulli(p) distribution $\leftrightarrow$ $Y \sim$ Bernoulli(p) (is distributed as) or $(Y \mid p)$

its pf is $f_Y(y) = P(Y=y) = \begin{cases} \dfrac{1}{b-a+1}, & \text{for } y=a,\dots,b \\ 0 & \text{else} \end{cases}$

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

<u>Definition</u>

n random <u>trials</u> 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 has the <u>Binomial distribution</u>

not sample space

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 $(Y \mid n, p)$

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

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

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

independent. $\boxed{\text{Notation}}$ $X_i \overset{IID}{\sim} f_{X_i}(x_i)$

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

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
$(i = 1, \ldots, n)$

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

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)

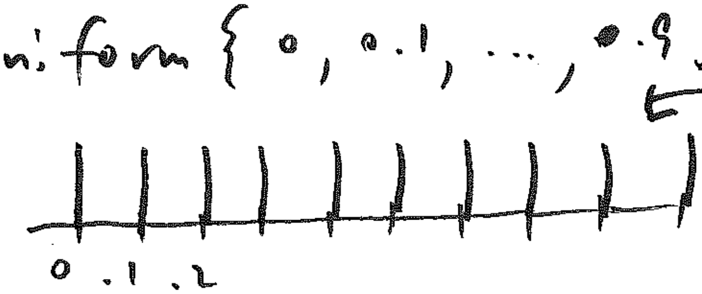
Single-precision floating point numbers carry about 7 decimal sigfigs of accuracy,

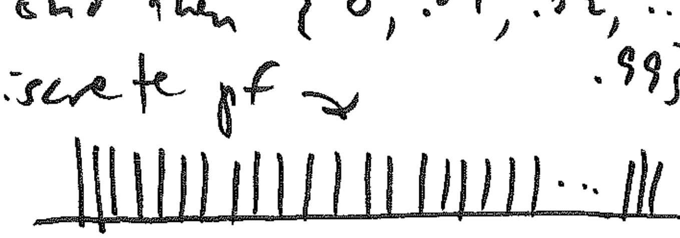
3.141592 | 653589
π 3 ~ 04 error

leading to roundoff error 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 uniform distribution is key to these calculations. Consider first

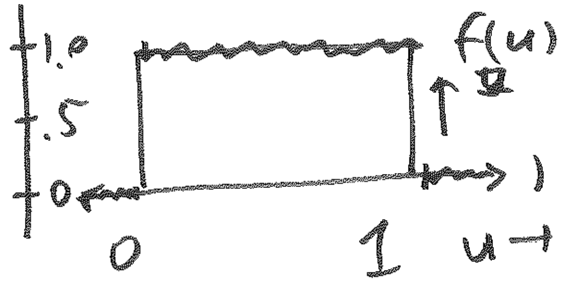
Uniform$\{0, 0.1, \ldots, 0.9\}$ and then $\{0, .01, .02, \ldots, .99\}$

← 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

$$f_U(u) = \begin{cases} 1 & \text{for } 0 \le u \le 1 \\ 0 & \text{else} \end{cases}.$$

Definition

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

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

$\boxed{\text{Definition}}$ If Y is a continuous rv, the function f_Y in the previous definition is called the <u>probability density function</u> (pdf) of Y. The set $\{y : f_Y(y) > 0\}$ 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 $P(a \le Y \le b)$ or $P(a < Y \le b)$ or $P(a \le Y < b)$ 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, $\boxed{P(Y = y) = 0 \text{ for all } -\infty < y < \infty}$

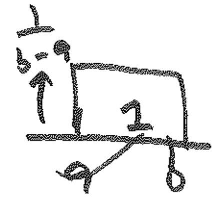
weirdly, this doesn't mean that the value y of Y is impossible, as it does with discrete rv; it just means that singletons have to have 0 probability (otherwise $\int_{-\infty}^{\infty} f_Y(y)\,dy = +\infty$ not 1).

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

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

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

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



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)

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

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

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

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

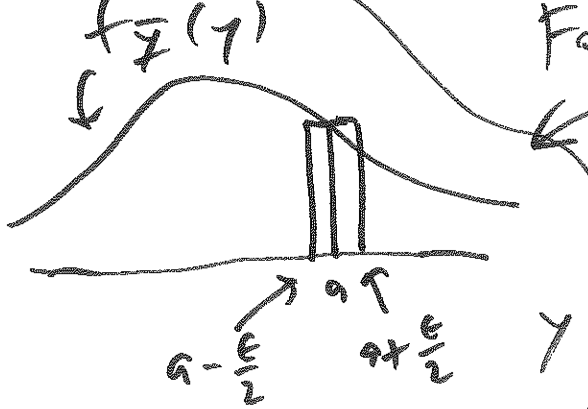
versus

$Y \sim \text{Uniform}\{a, b\}$ 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: $P(a \le Y \le b) = \int_a^b f_Y(y)\,dy$.



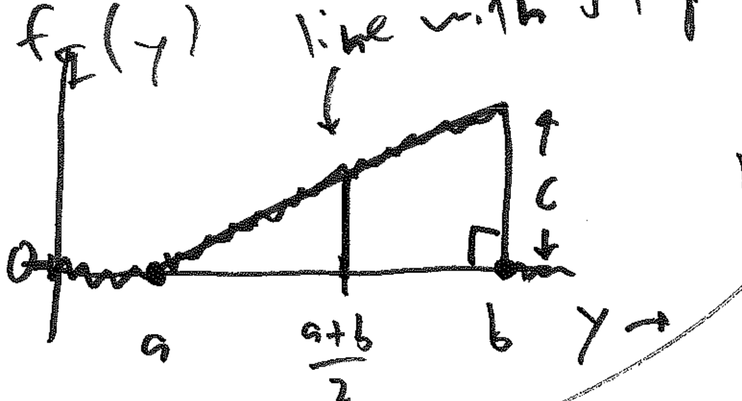
For small $\epsilon > 0$ you can see from this sketch that

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

(Connect to histograms; t.b. 5-13)

$\doteq$ area of rectangle $= \epsilon \cdot f_Y(a)$.

Example (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.

19

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

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

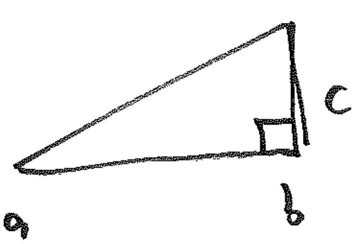
the equation of the line is

$f_X(y)$

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

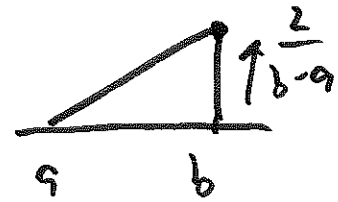
Densities have to integrate to 1,

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



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

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



Thus the triangular distribution that starts at 0 for $y=a$ and rises linearly to $\frac{2}{b-a}$ at $y=b$ has density (pdf)

$$f_Y(y) = \begin{cases} \dfrac{2(y-a)}{(b-a)^2} & a \le y \le b \\ 0 & \text{else} \end{cases} \quad (*)$$

You can see that calculating probabilities with continuous rvs requires you to dust off your integral calculus.

$\boxed{\text{Example}}$ with the triangular distribution (Y) $(*)$ above, what's $P\left(a \le Y \le \frac{b-a}{2}\right)$?

Hard(?) way: $$\int_a^{\frac{b-a}{2}} \frac{2(x-a)}{(b-a)^2}\,dx = \frac{(3a-b)^2}{4(b-a)^2}$$

Easy(?) way: 

area of triangle = ...

Sometimes it's mathematically convenient to work with unbounded continuous rvs, just as was true in the Poisson case study for discrete rvs.

Example (DS p. 105)

Y = voltage in an electrical system: in practice Y cannot be infinite, but you may not know ahead of time what its maximum practical value is, so model it as unbounded but without much probability for extreme large values. DS give as an example the pdf

$$f_Y(y) = \frac{1}{(1+y)^2} I(y > 0).$$

check: $\int_0^\infty \frac{1}{(1+y)^2} dy = 1$

$$= \frac{(1+y)^{-1}}{-1} \Big|_0^\infty = -1(0-1) = 1 \checkmark$$

You can check that $\int_{1000}^{\infty} \frac{1}{(1+x)^2} dx = \frac{1}{1001} \doteq .001$, so the right tail beyond $Y = 1000$ has almost no probability, matching the correct qualitative behavior.

Sometimes a rv will be neither discrete nor continuous; people then say that it has a <u>mixed</u> (discrete/continuous) <u>distribution</u>. Definition.

Example: In medical clinical trials of people with potentially fatal diseases, the outcome variable Y_i for person i in (say) the treatment group might be

Y = survival time in days from the beginning of the trial; however, and a good thing too, some patients may still be alive at the time T_{end} at which the trial finishes. Your probability model for Y_i would then have a continuous part for $0 \le Y \le T_{end}$ and a discrete lump of probability p at $Y = T_{end}$ signifying $(Y > T_{end})$ but we don't know what Y would have been if we could have observed it: (right-censoring)

$$\int_0^{T_{end}} f_{Y_i}(y)\,dy = (1-p) \quad \text{and} \quad P(Y_i > T_{end}) = p.$$

(RBoy example) (see p. 10 of doc. cam notes)

Unifying idea connecting discrete & continuous rvs

Discrete $\leftrightarrow$ pf (pmf)
Continuous $\leftrightarrow$ pdf
~~Mixed~~ $\leftrightarrow$ (pf + pdf)

$\boxed{Q:}$ Is there something that uniquely characterizes the distribution of Y, both when Y is discrete & when it's continuous?

$\boxed{A:}$ Yes, the cumulative distribution function (cdf) $F_Y(y)$

Definition:

The cumulative distribution function (cdf) of a rv Y is defined to be

$$F_Y(y) = P(Y \le y) \text{ for all } -\infty < y < \infty$$

(9 Aug 17)