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 *discrete distribution*, or equivalently Y is a *discrete* rv, if the set of (distinct) possible values of Y is finite or at most countably infinite; rvs for which the set of possible values is uncountable are called *continuous random variables*.

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 $\overset{\{\text{yes, no}\}}{\textit{dichotomous}}$ or $\overset{\{1, 0\}}{\textit{binary}}$.

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 $Y = \left(\begin{array}{c} \text{the actual (true)} \\ \text{weight of the package} \end{array} \right)$ 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 *conceptually* continuous (i.e., no limit in principle to the precision of measurement) as continuous.

Definition Given a *discrete* rv Y, the *probability (mass) function* (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 *support* of (the distribution of) Y.

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

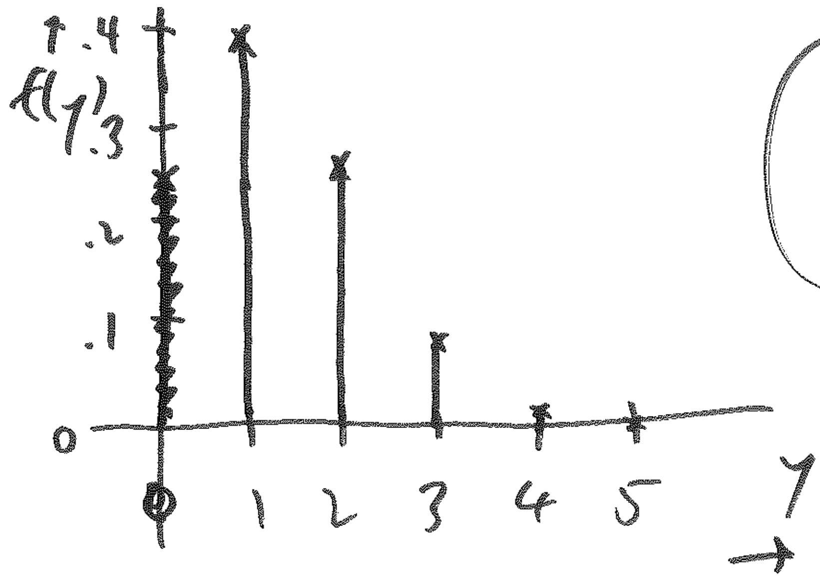
Example In the powerball lottery (see homework 1 problem 2) 5 white numbered balls are drawn at random without replacement from a bin with balls numbered $\{1, 2, \dots, 69\}$.

Let $\overline{\underline{W}}_i$ = # on ith drawn white ball.

$$\text{Clearly} \quad p(\overline{\underline{W}}_1 = w_1) = \left\{ \begin{array}{ll} \frac{1}{69} & \text{for } w_1 = 1, 2, \dots, 69 \\ 0 & \text{otherwise} \end{array} \right\}.$$

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

Definition For any two integers $a \overset{(<)}{\leq} b$, a rv $\overline{\underline{Y}}$ that's equally likely to be any of the values $\{a, a+1, \dots, b\}$ has the <u>uniform distribution</u> 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) —

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 $\leftrightarrow$ $Y \overset{\text{is distributed as}}{\sim} \text{Bernoulli}(p)$

or

$(Y \mid p)$

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

$Y \sim \text{Uniform}\{a, b\} \leftrightarrow Y$ 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 has the Binomial distribution

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

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

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

independent.

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

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, $B_i \overset{\text{IID}}{\sim} \text{Bernoulli}(p)$ 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 sum of a bunch of IID rvs, a topic we'll examine in detail later.

Continuous random variables

Example

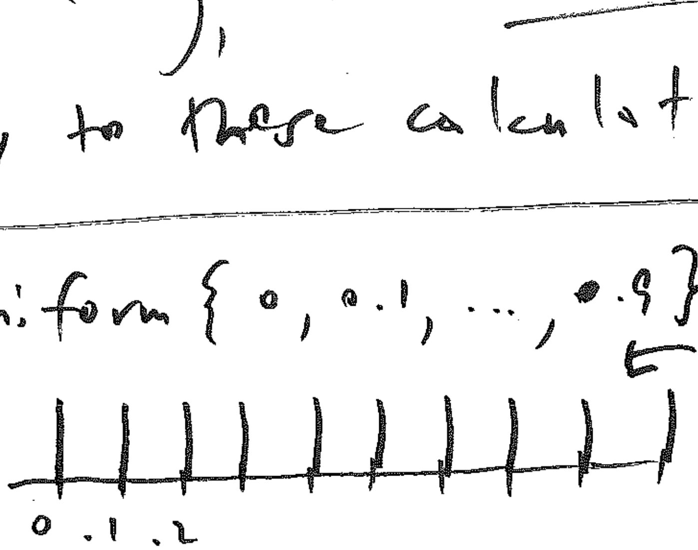
(Round-off error in computer science)

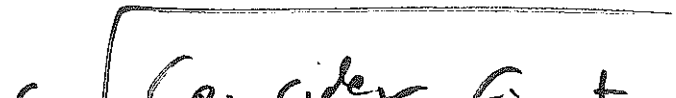
Single-precision decimal floating point numbers carry about 7 sigfigs of accuracy,

3.141592|653589
π 3 ~.04 error

leading to 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

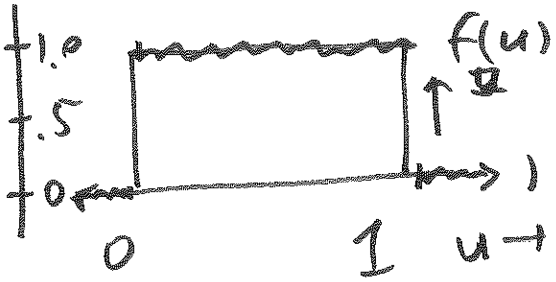
Uniform{0, 0.1, ..., 0.9} and then {0, .01, .02, ..., .99}
← discrete pf →





In the limit with more & more sig figs

this should go to



the

<u>continuous</u> <u>uniform</u> 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 <u>integration</u>

$$f_{\underline{U}}(u) = \left\{ \begin{array}{ll} 1 & \text{for } 0 \le u \le 1 \\ 0 & \text{else} \end{array} \right\}.$$

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 *probability density function* (pdf) of Y. The set $\{y : f_Y(y) > 0\}$ is called the *support* of (the distribution of) Y.

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

You'll recall from calculus that if f_Y is continuous

What about individual points — *singletons* — $\{y\}$ on $\mathbb{R}$?

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,

$$\boxed{P(Y = y) = 0 \text{ for all } -\infty < y < \infty}$$

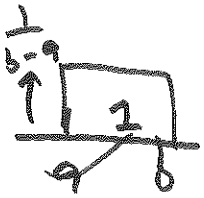
weirdly, this doesn't mean that the value y of Y is impossible, as it does with discrete rv; it just means that singletons have to have 0 probability (otherwise $\int_{-\infty}^{\infty} f_Y(y)\,dy = +\infty$ not 1).

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

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

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

$$f_{\underline{Y}}(y) = \left\{ \begin{array}{cc} \frac{1}{b-a} & \text{for } a \leq y \leq b \\ 0 & \text{else} \end{array} \right\}.$$



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

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

People sometimes also write (with A a set)

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

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 $[a,b]$

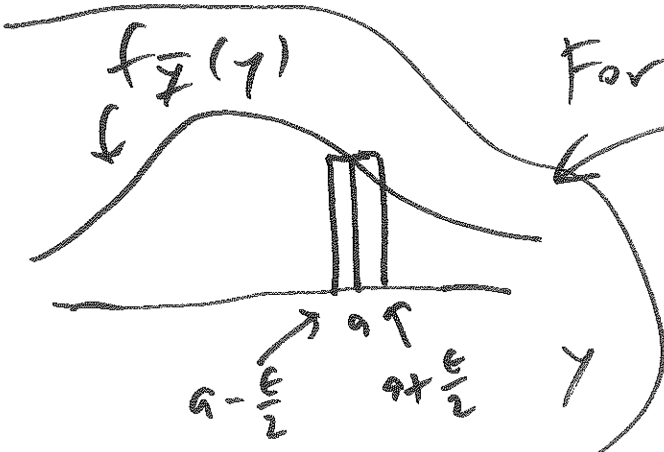
versus

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

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

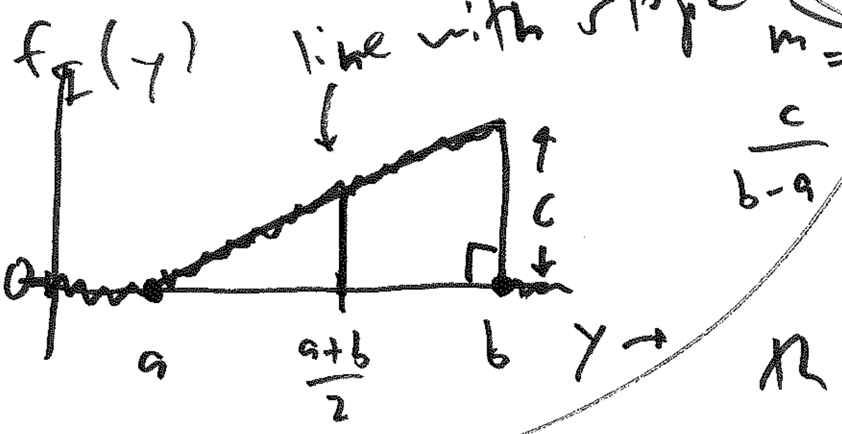
$$P\left(a - \frac{\epsilon}{2} \le Y \le a + \frac{\epsilon}{2}\right) = \int_{a-\frac{\epsilon}{2}}^{a+\frac{\epsilon}{2}} f_Y(y)\,dy$$

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

(Connect to histograms Sec. 1.3)

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.

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

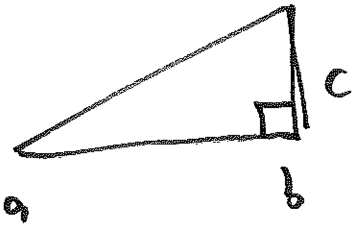
and passes through the point $\overset{(x_1,\, y_1) =}{\wedge}$ $(a, 0)$, so

the equation of the line is [illegible] $f_X(y)$

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

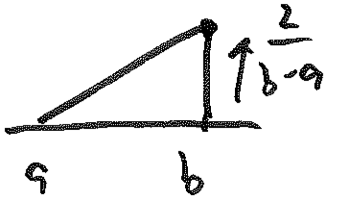
Densities have to integrate to 1,

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

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



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

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

.

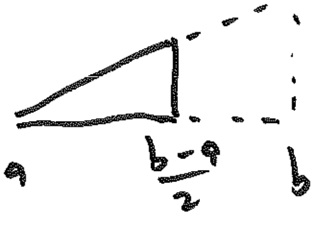
You can see that calculating probabilities with continuous rvs requires you to dust off your integral calculus.

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

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 = ...

(10.45)

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 extremely large values. DS give as an example the pdf

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

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

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

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 mixed (discrete/continuous) distribution

$\boxed{\text{Definition.}}$

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):

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

(R Boy example) (see p. ⑩ of doc. cam notes)

Unifying idea connecting discrete & continuous rvs

Discrete ↔ pf (pmf)

Continuous ↔ pdf

Mixed ↔ (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)

and it's also clear that a cdf $F_Y(y)$ has to be a non-decreasing function of y: if $y_1 < y_2$ then $F_Y(y_1) \le F_Y(y_2)$.

Furthermore, $\lim_{y \to -\infty} F_Y(y) = 0$ and $\lim_{y \to +\infty} F_Y(y) = 1$.

CDFs F_Y can be continuous on all of $\mathbb{R}$ (when Y is continuous) but certainly don't have to be (see the CDF of the Bernoulli (p) distribution).

Technical fact:

Def: $F_Y(y^-) \overset{\triangle}{=} \lim_{\substack{y^* \to y \\ y^* < y}} F_Y(y^*) \overset{\triangle}{=} \lim_{y^* \uparrow y} F_Y(y^*)$

limit from the left

(y^* goes to y from below)

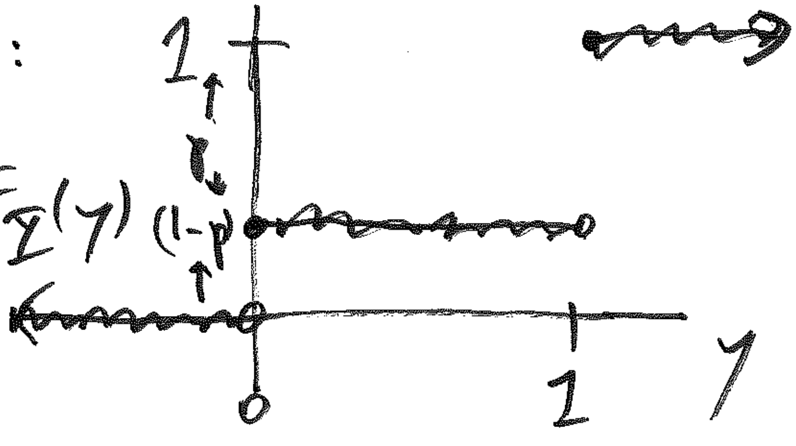
Example: $Y \sim$ Bernoulli(p)

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

Notice that there's a clever way to write this pf:

$$P(Y=y) = p^{y}(1-p)^{1-y} I_{\{0,1\}}(y)$$

The cdf of Y is 0 for $y<0$; at $Y=0$ it jumps up to $(1-p)$ and stays there for $0 \le y < 1$; and at $Y=1$ it jumps up to 1 & stays there for $y \ge 1$:



You can see that in general

$$0 \le F_Y(y) \le 1$$

Def. $$F_Y(y^+) \triangleq \lim_{\substack{y^* \to y \\ y^* > y}} F_Y(y^*) \triangleq \lim_{y^* \downarrow y} F_Y(y^*)$$

limit from right

(y^* going to y from above)

techical fact: $F_Y(y) = F_Y(y^+)$ for all $-\infty < y < \infty$

people call this <u>continuity from the right</u>

or <u>continuity from above</u>

Consequences of the CDF definition

① $P(Y > y) = 1 - F_Y(y)$

② For all y_1, y_2 with $y_1 < y_2$

$P(y_1 < Y \leq y_2) = F_Y(y_2) - F_Y(y_1)$.

If $F_Y(y^-) = F_Y(y^+) = F_Y(y)$ then F_Y is continuous at y

Consequence ② means that if Y is continuous, there's an intimate connection between $\underset{(\text{cdf})}{F_Y(y)}$ and $\underset{(\text{pdf})}{f_Y(y)}$:

Y continuous: $y_1 < y_2$

$$P(y_1 < Y \leq y_2) = \boxed{\begin{aligned} & F_Y(y_2) - F_Y(y_1) \\ = & \int_{y_1}^{y_2} f_Y(y)\,dy \end{aligned}}$$

and thus

Theorem If Y is a continuous rv, with pdf $f_Y(y)$ and CDF $F_Y(y)$ then

at all continuity points of f

$$F_Y(y) = \int_{-\infty}^{y} f_Y(t)\,dt \quad \text{and} \quad \frac{d}{dy} F_Y(y) = f_Y(y).$$

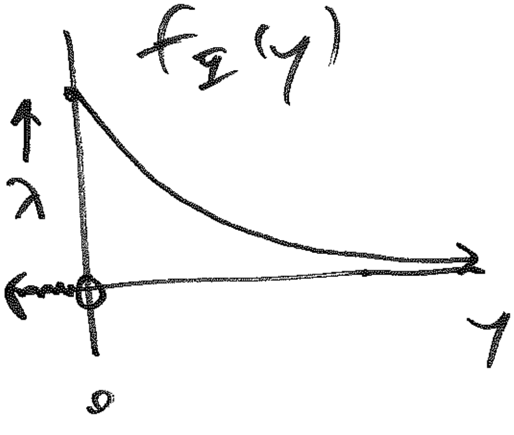
In other words

Y continuous $\leftrightarrow$ the derivative of $\underset{\text{CDF}}{F_Y(y)}$ is $\overset{\text{pdf}}{f_Y(y)}$ (and $F_Y(y)$ is an anti-derivative (integral)) of $f_Y(y)$.

Definition

Y follows an exponential distribution $\mathcal{E}(\lambda)$ with (rate) parameter $\lambda > 0$

$$\leftrightarrow \underset{\text{pdf}}{f_Y(y)} = \begin{cases} \lambda e^{-\lambda y} & y > 0 \\ 0 & y \leq 0 \end{cases}$$



The exponential dist. has a fundamental connection to the Poisson distribution in Poisson processes that we'll explore later.

It's easy to calculate the CDF of an exponential distribution:

Notation

(Y exponentially distributed with parameter $\lambda > 0$) $\leftrightarrow$ $\overset{(Y \mid \lambda)}{\overset{\downarrow}{Y}} \sim \text{Exponential}(\lambda)$

$\updownarrow$

$Y \sim \mathcal{E}(\lambda)$

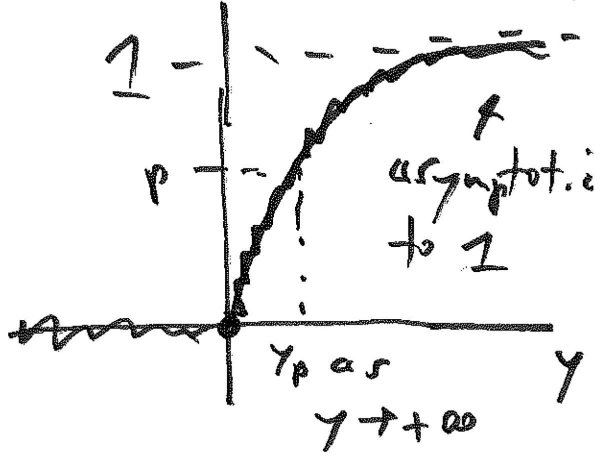
for $y > 0$

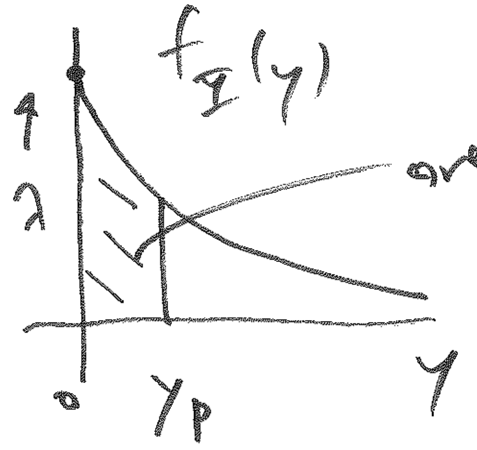
$$F_Y(y) = \int_{-\infty}^{y} f_Y(t)\,dt = \int_0^y \lambda e^{-\lambda t}\,dt$$

$$= \lambda \frac{e^{-\lambda t}}{-\lambda}\Bigg|_0^y = -1\left(e^{-\lambda y} - 1\right) = 1 - e^{-\lambda y}$$

So

$$F_Y(y) = \begin{cases} 0 & \text{for } y \leq 0 \\ 1 - e^{-\lambda y} & y > 0 \end{cases}$$





Q: what's the place y_p on the positive part of $\mathbb{R}$ where $P(0 \le Y \le y_p) = p$?

for ($y_p > 0$)

well, $\underset{\wedge}{P}(0 \le Y \le y_p) = \boxed{F_Y(y_p) = p}$

$= \boxed{1 - e^{-\lambda y_p} = p}$ so $y_p = F_Y^{-1}(p)$

$1 - p = e^{-\lambda y_p}$

$\log(1-p) = -\lambda y_p$

$y_p = -\frac{\log(1-p)}{\lambda} = F_Y^{-1}(p)$

Def. y_p is called the p^{th} quantile or the $100p^{\text{th}}$ percentile of (the distribution of) Y.

Some care is required when Y is discrete or mixed.

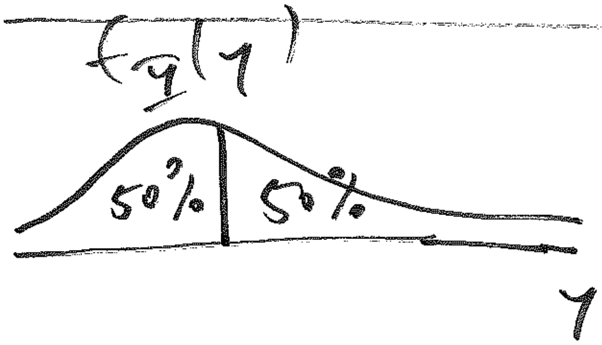
General definition

Y rv with CDF $F_Y(y)$.

For all $0 < p < 1$ define

$$F_Y^{-1}(p) = \text{the } \underline{\text{smallest}} \text{ } y \text{ value such that } F_Y(y) \geq p$$

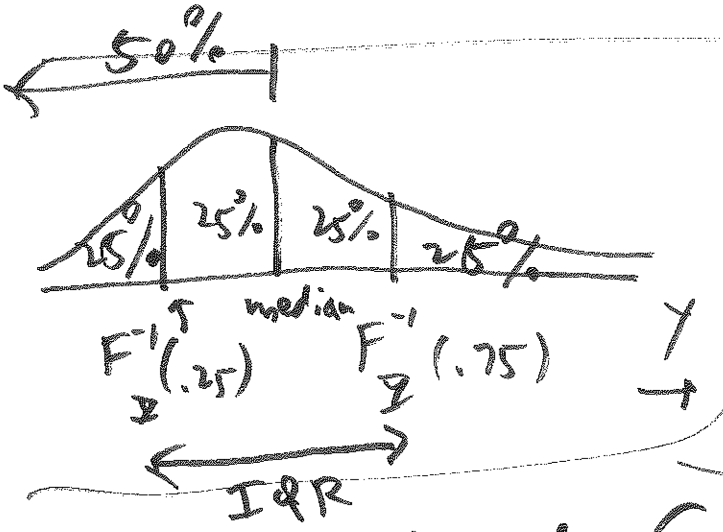
Then $F_Y^{-1}(p)$ is the p^{th} quantile (= $100p^{th}$ percentile) of Y and F_Y^{-1} is the <u>quantile function</u>.



Measures of <u>center</u> for the distribution of a rv Y

One way to define the <u>center</u> of a distribution is to find the 50^{th} percentile.

Definition The $\overset{(0.5)}{\frac{1}{2}}$ quantile ≜ the 50th percentile of a distribution is called the median of the dist.

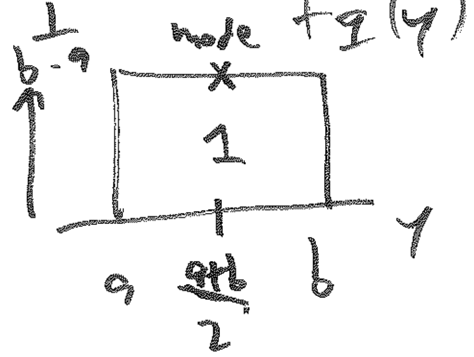


measures of spread for the distribution of a rv X

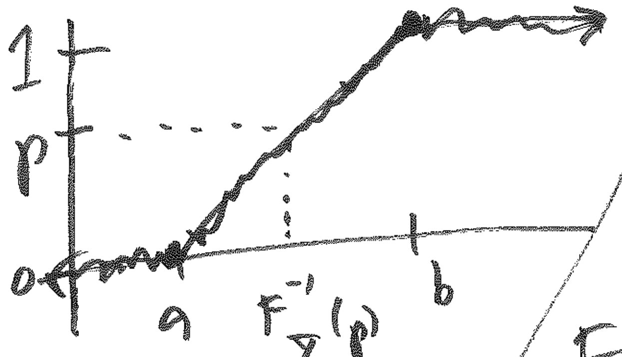
One way to define the spread of a dist. is to see how far apart its 75th and 25th percentiles are.

Definition The $\overset{(0.25)}{\frac{1}{4}}$ quantile = $\overset{F_X^{-1}(.25)}{\text{the 25th percentile}}$ is the lower quartile; the $\overset{(0.75)}{\frac{3}{4}}$ quantile $\overset{F_X^{-1}(.75)}{=}$ the 75th percentile is the upper quartile, and $[F_X^{-1}(.75) - F_X^{-1}(.25)]$ = interquartile range (IQR)

Example $Y \sim \text{Uniform}(a, b)$; then



$$F_Y(y) = \begin{cases} 0 & \text{for } y \le a \\ \frac{y-a}{b-a} & a \le y \le b \\ 1 & y \ge b \end{cases}$$



Easy to invert F_Y:

$$F_Y^{-1}(p) = (1-p)a + pb \text{ for } 0 < p < 1$$

And (no surprise) the median is $\frac{a+b}{2}$.

Studying two random variables at a time

Def. X, Y rvs: the <u>joint</u> (or <u>bivariate</u>) distribution of (X, Y) is the collection $P[(X, Y) \in C]$ of all probabilities for all sets $C \in \mathbb{R}^2$ such that $(X, Y) \in C$ isn't weird.

Case 1 (X and Y both discrete)

Def. X, Y r.v. → If there are only finitely or countably infinitely many possible values (x,y) for (X,Y), X and Y have a discrete joint dist.

Def The joint probability (mass) function (joint pmf) of (X,Y) discrete is the function $f_{X,Y}(x,y) = P(X=x \overset{\text{and}}{,} Y=y)$.

the set $\{(x,y) : f_{X,Y}(x,y) > 0\}$ is the support of f_{XY}

Consequences ① $$\sum_{\text{all }(x,y)} f_{\underset{\nwarrow \text{ can (omit comma)}}{XY}}(x,y) = 1.$$

② For any (non-weird) set C of ordered pairs (x,y), $$P[(X,Y) \in C] = \sum_{(x,y)\in C} f_{XY}(x,y)$$

Def Two rv X and Y have a continuous joint distribution if you can find a nonnegative function $f_{X,Y}(x,y)$ defined for all $(x,y) \in \mathbb{R}^2$ (the real plane)

Case 2: X, Y both continuous

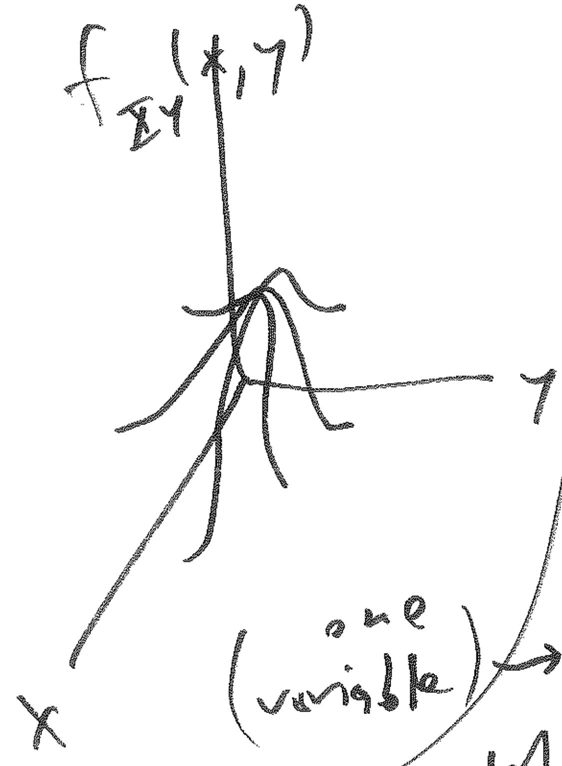
such that for every (non-weird) subset C of the plane $P[(X,Y) \in C] = \iint_C f_{XY}(x,y)\,dx\,dy$.

$f_{XY}(x,y)$ is the joint pdf of (X,Y).

the set $\{(x,y) : f_{XY}(x,y) > 0\}$ is the support of (the dist. of) (X,Y).

Immediate Consequences ① For all (x,y) in $\mathbb{R}^2$,

$f_{XY}(x,y) \geq 0$, and $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy = 1$.



② If (X, Y) have a continuous joint distribution, then X and Y each have a continuous (marginal) (one variable) → univariate distribution when considered separately.

③ For all continuous pdfs $f_{XY}(x,y)$,

(a) Every individual point, and every countably infinite sequence (or set) of points in $\mathbb{R}^2$, has probability 0 under f_{XY}. (b) If g is a continuous function of one real variable defined on (a, b), then the sets $\{(x,y): y = g(x), a<x<b\}$ and $\{(x,y): x = g(y), a<y<b\}$ also have probability 0.

④ This means that the converse of ② is (unfortunately) not true: If X has a continuous distribution on $\mathbb{R} = \mathbb{R}^1$ and $Y \triangleq X$, then both X and Y have continuous distributions but $P\left[\begin{array}{l}(X, Y) \text{ lies on the} \\ \text{line } y = x\end{array}\right] = \overset{1}{\mathbf{0}}$, so (X, Y) can't have a continuous joint distribution on $\mathbb{R}^2$.

Example

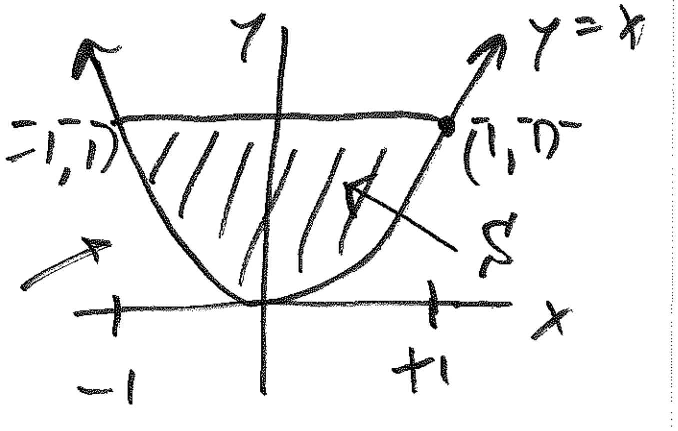
Joint distributions can lead to tricky integrals

Suppose that (X, Y) have joint pdf

$$f_{XY}(x, y) = \begin{cases} cx^2y & \text{for } 0 \le x^2 \le y \le 1 \\ 0 & \text{else} \end{cases}$$

Let's work out the normalizing constant.

The support of f_{XY} is the shaded region S here



$$1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\, dx\, dy$$

$$= \iint_{S} f_{XY}(x,y)\, dy\, dx$$

$$= \int_{-1}^{+1}\int_{x^2}^{1} c\, x^2 y\, dy\, dx$$

$$= \int_{-1}^{1} c x^2 \left(\int_{x^2}^{1} y\, dy\right) dx$$

$$= \int_{-1}^{1} c x^2 \left(\frac{y^2}{2} \Big|_{x^2}^{1} \right) dx$$

$$= \int_{-1}^{1} c x^2 \left(\frac{1}{2} - \frac{x^4}{2} \right) dx$$

$$= \frac{1}{2} c \int_{-1}^{1} x^2\, dx - \frac{1}{2} c \int_{-1}^{1} \frac{x^6}{2}\, dx$$

$$= \frac{1}{2} c \left(\frac{x^3}{3} \Big|_{-1}^{1} \right) - \frac{c}{2} \left(\frac{x^7}{14} \Big|_{-1}^{1} \right) = \frac{4}{21} c = 1$$

So $\boxed{c = \frac{21}{4}}$

The other way to parameterize the support $\mathcal{S}$ is to let y go from 0 to 1 while x goes from $-\sqrt{y}$ to $\sqrt{y}$:

$$1 = \iint_{\mathcal{S}} f_{\mathbb{X}\mathbb{Y}}(x,y)\,dx\,dy$$

$$= \int_0^1 \int_{-\sqrt{y}}^{\sqrt{y}} c x^2 y \,dx\,dy$$

$$= \int_0^1 c y \left(\int_{-\sqrt{y}}^{\sqrt{y}} x^2 \,dx \right) dy$$

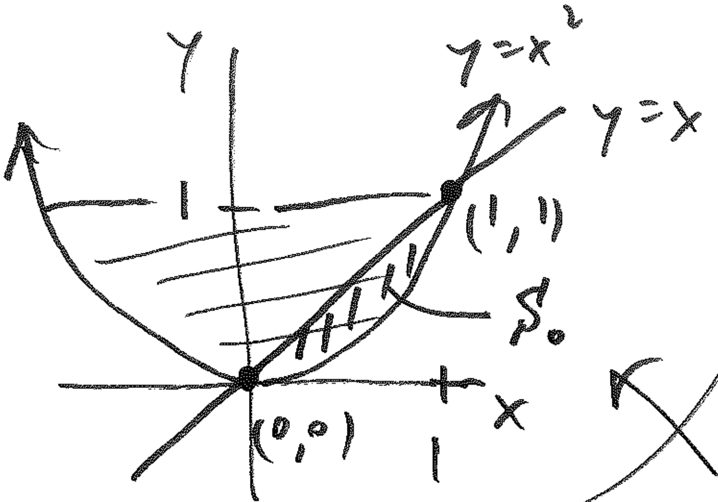
$$= \int_0^1 c y \left(\frac{x^3}{3} \Bigg|_{-\sqrt{y}}^{\sqrt{y}} \right) dy$$

$$= c \int_0^1 y \cdot \frac{1}{3} \left(y^{3/2} - -y^{3/2} \right) dy$$

$$= \frac{c}{3}\int_0^1 2y^{5/2}\,dy = \frac{2c}{3}\left(\frac{y^{\frac{7}{2}}}{7/2}\Bigg|_0^1\right)$$

$= \frac{4}{21}c$ as before ($\iint dx\,dy$ and $\iint dy\,dx$ always have to agree, of course).

Example, continued | Let's compute $P(X \geq Y)$



The relevant part S_0 of S where $x \geq y$ is sketched

here, S_0

$$P(X \geq Y) = \iint_{S_0} f_{XY}(x,y)\,dy\,dx$$

$$= \int_0^1 \int_{x^2}^{x} \frac{21}{4}x^2 y\,dy\,dx \underset{(\dots)}{=} \frac{3}{20}.$$

You can have bivariate distributions in which one of (X, Y) is discrete and the other is continuous.

Definition mixed bivariate distribution

(X, Y) rv such that X is discrete and Y is continuous → suppose you can find a function $f_{XY}(x, y)$ defined on $\mathbb{R}^2$ such that for every pair of (non-weird) subsets A and B of $\mathbb{R}$ (assume integral exists)

$$P(X \in A \text{ and } Y \in B) = \int_B \sum_{x \in A} f_{XY}(x, y)\, dy.$$

Then f_{XY} is the joint pmf/pdf of (X, Y)

Immediate consequence

If X takes on values $x_1, x_2, \ldots,$ then

$$\int_{-\infty}^{\infty} \sum_{i=1}^{\infty} f_{XY}(x_i, y)\, dy = 1.$$

Example Randomized controlled (clinical) trial; patients in Ⓣ get a treatment, patients in Ⓒ get a placebo. Outcome is success (e.g., cancer goes into remission) or failure; let

$$X_i = \begin{cases} 1 & \text{if patient } i \text{ in Ⓣ is a success} \\ 0 & \text{else} \end{cases}$$

and let θ ← (unknown) be the proportion of patients in the *population* of all patients who might get the treatment who would have no relapse if they had been in the study. Then our uncertainty about θ is continuous on $(0, 1)$ and (X_i, θ) has a *mixed* bivariate distribution.

If you model $(X \mid \theta)$ as Bernoulli(θ) and $\theta \sim$ Uniform$(0,1)$ the joint pf/pdf of (X, θ) would be

$$f_{X,\Theta}(x,\theta) = \left\{ \begin{array}{ll} \theta^{x}(1-\theta)^{1-x} & \text{for} \left(\begin{array}{c} x=0,1 \\ 0<\theta<1 \end{array} \right) \\ 0 & \text{else} \end{array} \right\}$$

pf/pdf $\nearrow$

Then (e.g.) $P(X=1) = P(X=1$ and θ is anything between 0 and 1$)$

$$= \int_0^1 \theta^1 (1-\theta)^{1-1}\, d\theta = \int_0^1 \theta\, d\theta = \frac{1}{2}.$$

Bivariate CDFs

Def. The joint CDF of two rvs X and Y is the function $F_{XY}(x,y)$

satisfying $F_{XY}(x,y) = P(X \le x \text{ and } Y \le y)$

for all $-\infty < X < \infty$ and $-\infty < y < \infty$

Consequences of this definition

① If (X, Y) has the joint CDF $F_{XY}(x,y)$, you can obtain the marginal CDF $F_X(x)$ from the joint CDF as $F_X(x) = \lim_{y\to\infty} F_{XY}(x,y)$, and similarly the marginal CDF $F_Y(y)$ is just $F_Y(y) = \lim_{x\to\infty} F_{XY}(x,y)$

---

② The joint pdf and joint CDF are related in a manner similar to their relationship in the univariate (one rv at a time) case:

If (X, Y) have a joint pdf $f_{XY}(x,y)$

then $F_{XY}(x,y) = \int_{-\infty}^{y}\int_{-\infty}^{x} f_{XY}(r,s)\,dr\,ds$

$$= \int_{-\infty}^{x}\int_{-\infty}^{y} f_{XY}(r,s)\,ds\,dr$$

and $f_{XY}(x,y) = \frac{\partial^2}{\partial x\,\partial y} F_{XY}(x,y) = \frac{\partial^2}{\partial y\,\partial x} F_{XY}(x,y)$

(at every (x,y) where the partial derivatives exist).

~~Consequences~~ continued

③ If (X, Y) have a discrete joint distribution with joint pmf $f_{XY}(x,y)$, then the marginal pmf $f_X(x)$ of X is $\boxed{f_X(x) = \sum_{y} f_{XY}(x,y)}$

(and similarly for $f_Y(y)$).

The idea behind marginal distributions is that it's harder to visualize a joint (2-dimensional) distribution than it is to visualize each of its 1-dimensional marginal distributions.

④ If (X, Y) have a continuous joint distribution with joint pdf $f_{XY}(x, y)$, the <u>marginal</u> pdf $f_X(x)$ of X is

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x, y) \overset{\text{(marginalizing out } Y)}{\overset{\downarrow}{dy}} \quad (\text{for all } -\infty < x < \infty)$$

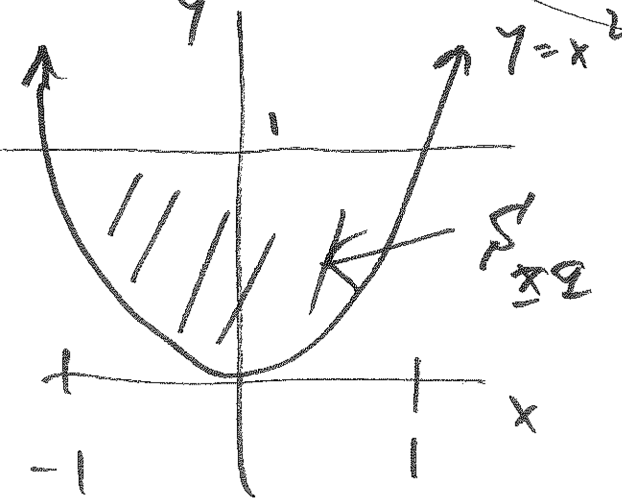
and the marginal pdf $f_Y(y)$ of Y is

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x, y)\, dx \quad \left(\text{for all } -\infty < y < \infty\right).$$

Earlier example, continued

$(\underline{X},\underline{Y})$ have joint pdf

$$f_{\underline{X}\underline{Y}}(x,y) = \begin{cases} \frac{21}{4}x^2y, & \overset{S_{\underline{X}\underline{Y}}}{\boxed{x^2 \le y \le 1}} \\ 0 & \text{else} \end{cases}$$

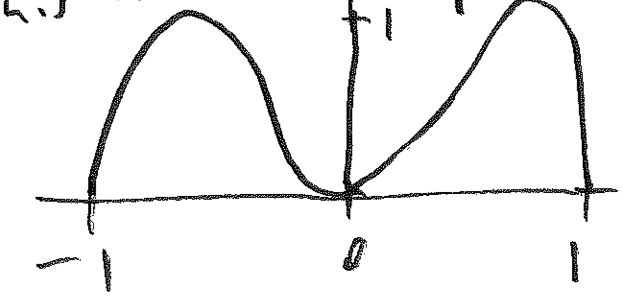


You can see from the sketch of the support $S_{\underline{X}\underline{Y}}$ of $f_{\underline{X}\underline{Y}}(x,y)$ that $-1 \le \underline{X} \le 1$, so the support $\overset{S_{\underline{X}}}{\wedge}$ of $\underline{X}$ is $(-1, 1)$, and its marginal pdf is

$$f_{\underline{X}}(x) = \int_{-\infty}^{\infty} f_{\underline{X}\underline{Y}}(x,y)\,dy = \int_{x^2}^{1} \frac{21}{4}x^2y\,dy$$

$$= \begin{cases} \frac{21}{8}x^2(1-x^4) & -1<x<1 \\ 0 & \text{else} \end{cases}$$

This is a weird pdf



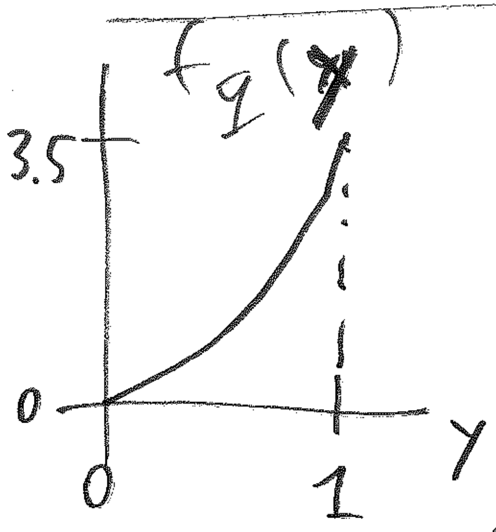
(supposed to be symmetric) & bimodal

Similarly, the support S_Y of Y is $(0, 1)$, and its marginal pdf is

$$f_Y(y) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dx = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{21}{4} x^2 y \, dx$$

$$= \left\{ \begin{array}{ll} \frac{7}{2} y^{\frac{5}{2}} & \text{for } 0 < y < 1 \\ 0 & \text{else} \end{array} \right\}$$

(4 Aug 17)



Consequences, continued ⑤

If you have the joint dist. $f_{XY}(x,y)$, you can reconstruct the marginals $f_X(x)$ and $f_Y(y)$, <u>but not the other way around</u>: if all you have is the marginals, they do not uniquely determine the joint.