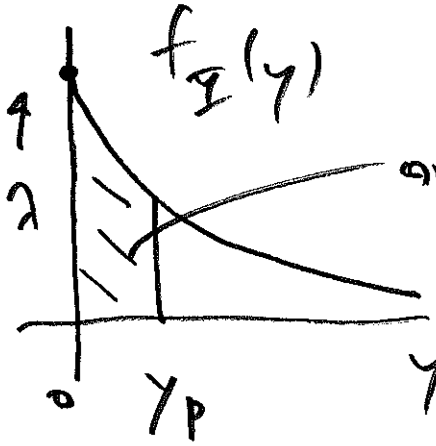


Q: What's the place y_p on the positive part of $\mathbb{R}$ where $P(0 \le Y \le y_p) = p$?

(for $y_p > 0$)

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

$$= \boxed{1 - e^{-\lambda y_p} = p}$$ 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} <u>quantile</u> or the $100p^{th}$ <u>percentile</u> of (the distribution of) Y.

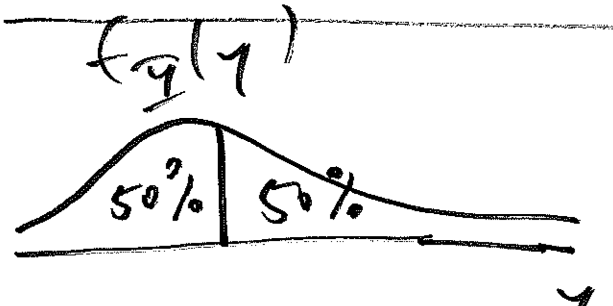
Some care is required when Y is discrete or mixed.

General definition Y rv with CDF $F_Y(y)$.

For all $0 < p < 1$ define

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

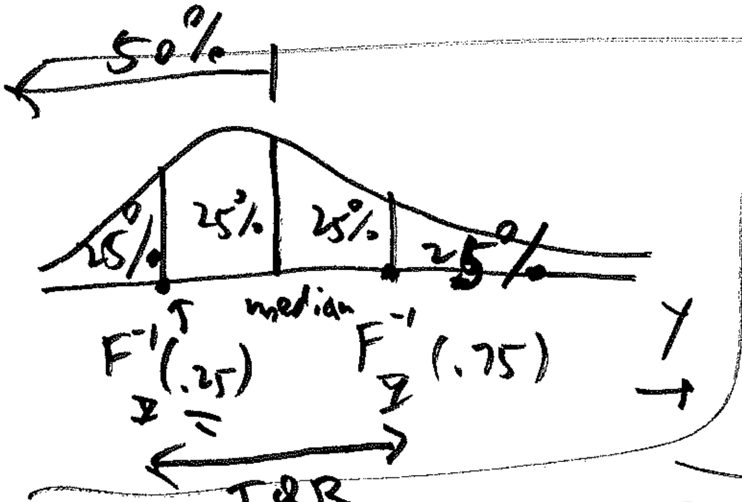
Then $F_Y^{-1}(p)$ is the p^{th} quantile (= $100p^{th}$ percentile) of Y and F_Y^{-1} is the quantile function.



Measures of center for the distribution of a rv Y

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

Definition The $\frac{1}{2}$ (0.5) quantile $\triangleq$ the 50th percentile of a distribution is called the <u>median</u> of the dist.

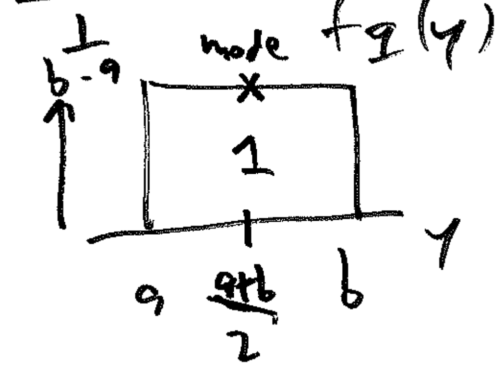


<u>measures of spread</u> for the distribution of a rv X

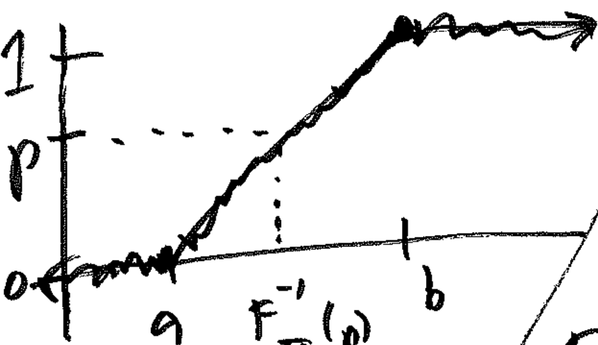
One way to define the <u>spread</u> of a dist. is to see how far apart its 75th and 25th percentiles are.

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

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



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



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) 55

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 <u>discrete joint dist.</u>

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

The set $\{(x,y) : f_{XY}(x,y) > 0\}$ is the <u>support</u> of f_{XY}

Consequences

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

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 $\underline{X}$ and $\underline{Y}$ have a

Case 2: $\underline{X}$, $\underline{Y}$ both continuous

<u>continuous joint distribution</u> if you can find a nonnegative function $f_{X,Y}(x,y)$ defined for all $(x,y) \in \mathbb{R}^2$ (the real plane)

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

<u>support</u> of (the dist. of) (X,Y).

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

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

Wd

plot exp(-x^2 - y^2)

(2) If (X, Y) have a continuous joint distribution, then X and Y each have a continuous (marginal) univariate (one variable) 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 $\mathbb{R}^2$ (0-dimensional), 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 $\{(x,y): x = g(y),\ a < y < b\}$ (1-dimensional) 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

$$P\left[(X, Y) \text{ lies on the line } y = x\right] = 1$$

0 ?

So (X, Y) can't have a continuous joint distribution on $\mathbb{R}^2$.

Example

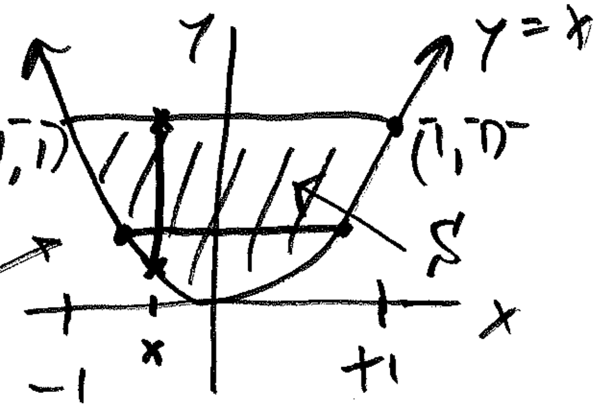
Joint distributions can lead to tricky integrals

Suppose that (X, Y) have joint pdf

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

Let's work out the normalizing constant.

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



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

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

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

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

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

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

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

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

search on "iterated integrals in wolfram alpha" so $\boxed{c = \frac{21}{4}}$

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

$$1 = \iint_S f_{XZ}(x,y)\,dx\,dy$$

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

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

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

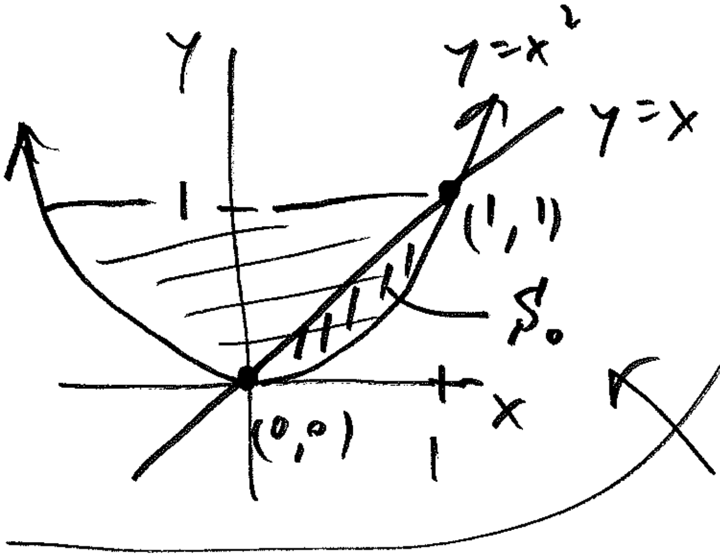
$$= c \int_0^1 y \cdot \frac{1}{3} \left(y^{3/2} - - y^{3/2} \right) dy$$

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

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

102

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

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

integrate $21\,x^2\,y/4\;dy$, $y = x^2 .. x$

then integrate result from $x = 0..1$

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

Definition

Case 3

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 (T) get a treatment, patients in (C) get a placebo. Outcome is success (e.g., cancer goes into remission) or failure; let $X_i = \begin{cases} 1 & \text{if patient } i \text{ in (T) is a success} \\ 0 & \text{else} \end{cases}$

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 pmf/pdf of (X, θ) would be

$$f_{X,\Theta}(x,\theta) = \begin{cases} \theta^{x}(1-\theta)^{1-x} & \text{for } \begin{pmatrix} x = 0, 1 \\ 0 < \theta < 1 \end{pmatrix} \\ 0 & \text{else} \end{cases}$$

pf/pdf

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

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

(2 May 19)

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 $-\infty < y < \infty$