and it's also clear that a cdf $F_Y(y)$ has to be a <u>non-decreasing function</u> 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.: $F_Y(y^-) \triangleq \lim_{\substack{y^* \to y \\ y^* < y}} F_Y(y^*) \triangleq \lim_{y^* \uparrow y} F_Y(y^*)$

limit from the left

($y^*$ goes to $y$ from below)

**Example:** $Y \sim \text{Bernoulli}(p)$

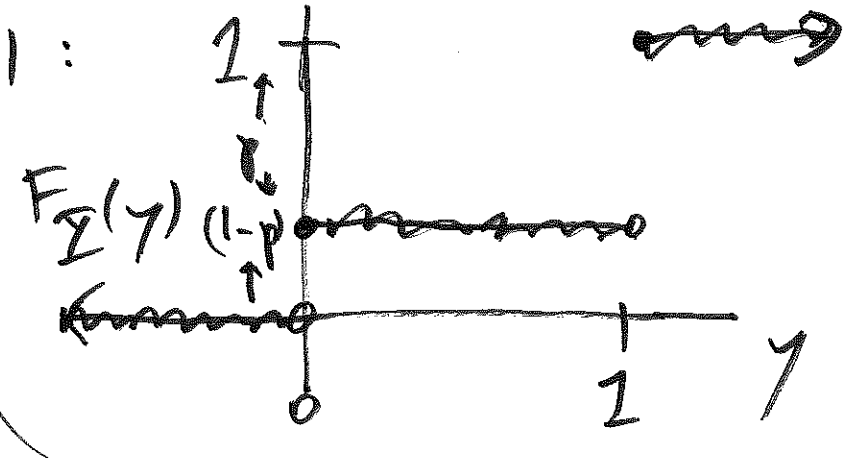
$$p(Y = y) = \begin{cases} p & \text{for } y = 1 \\ 1-p & \phantom{\text{for }} 0 \\ 0 & \text{else} \end{cases}.$$

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

limit from right

$y^* > y$ ($y^*$ going to $y$ from above)

technical 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>

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

---

Consequences of the CDF definition

---

① $P(Y > y) = 1 - F_Y(y)$

---

② For all $y_1, y_2$ with $y_1 < y_2$

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

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

---

$Y$ continuous: $y_1 < y_2$

$$P(y_1 < Y \le y_2) = \boxed{F_Y(y_2) - F_Y(y_1) = \int_{y_1}^{y_2} f_Y(y)\,dy}$$

and thus

---

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

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

at all continuity points of $f$

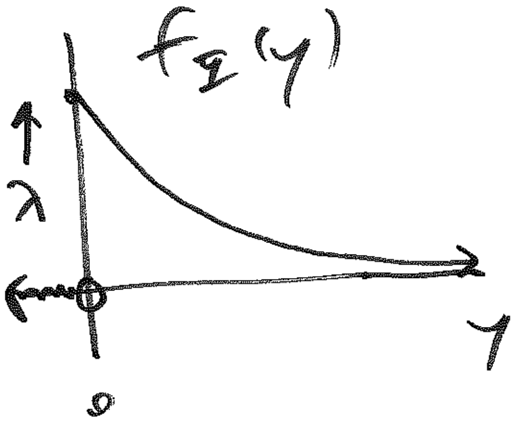
**In other words**

$Y$ continuous $\longleftrightarrow$ 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 <u>exponential distribution</u> $\mathcal{E}(\lambda)$ with parameter (rate) $\lambda > 0$

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




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

$\left( Y \text{ exponentially distributed with parameter } \lambda > 0 \right) \leftrightarrow \underset{(Y \mid \lambda)}{Y} \sim \text{Exponential}(\lambda)$

$\updownarrow$

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

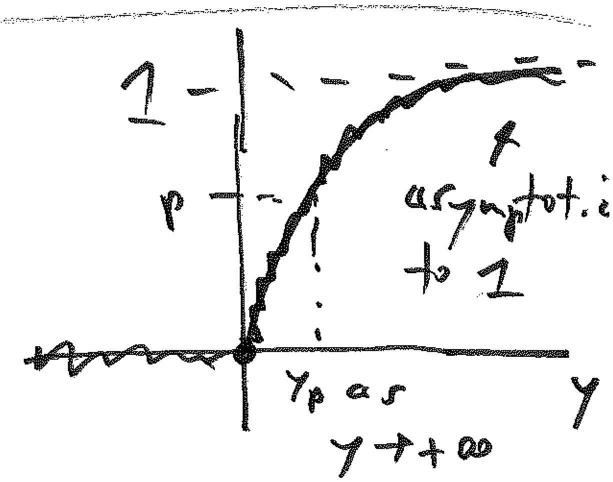
for $y > 0$

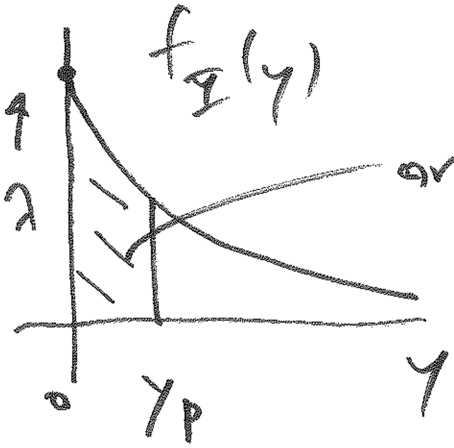
$$F_Y(y) = \int_{-\infty}^{y} f_Y(t)\,dt = \int_0^y \lambda e^{-\lambda t}\,dt$$

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

So

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






for $(y_p > 0)$

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

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

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

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

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

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

$\boxed{\text{Def.}}$ $y_p$ is called the $p^{\underline{th}}$ <u>quantile</u> or the $100p^{\underline{th}}$ <u>percentile</u> of (the distribution of) $Y$.

Some care is required when $Y$ is discrete or mixed.

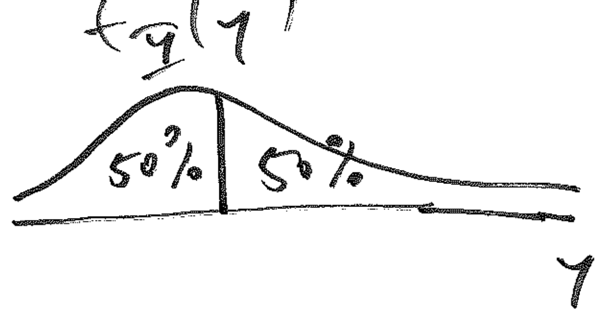
**General definition**

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

For all $0<p<1$ define

$F_Y^{-1}(p)$ = the <u>smallest</u> $y$ value such that $F_Y(y) \ge p$

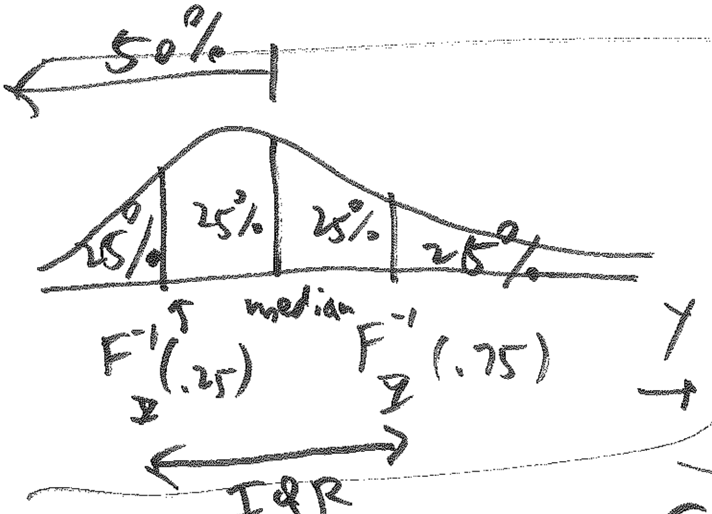
Then $F_Y^{-1}(p)$ is the $p^{th}$ quantile (= $100p^{th}$ percentile) of $Y$ and $F_Y^{-1}$ is the <u>quantile function</u>.




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

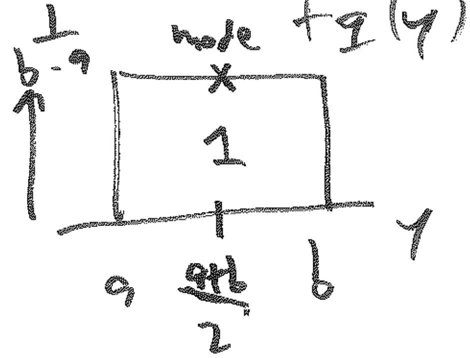



<u>measures of spread</u> for the distribution of a rv $X$

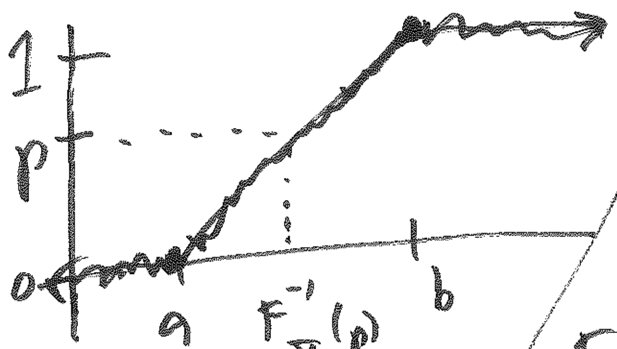
one way to define the <u>spread</u> of a dist. is to see how far apart its $75^{\text{th}}$ and $25^{\text{th}}$ percentiles are.

**Definition** The $\frac{1}{4}$ $(0.25)$ quantile = the $25^{\text{th}}$ percentile $F_X^{-1}(.25)$ is the <u>lower quartile</u>; the $\frac{3}{4}$ $(0.75)$ quantile = the $75^{\text{th}}$ 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 $F_Y$:

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

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

**Studying two random variables at a time**

**Def.** $X, Y$ rvs: the **joint** (or **bivariate**) 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) (95)

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 $f_{X,Y}(x,y) = P(X=x, Y=y)$. (and)

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

Consequences

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

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

96

**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) univariate distribution when considered separately.

(one variable) → univariate

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

(a) Every individual point, and every countably infinite sequence (or set) of points in $R^2$, 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

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

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

Joint distributions can lead to tricky integrals

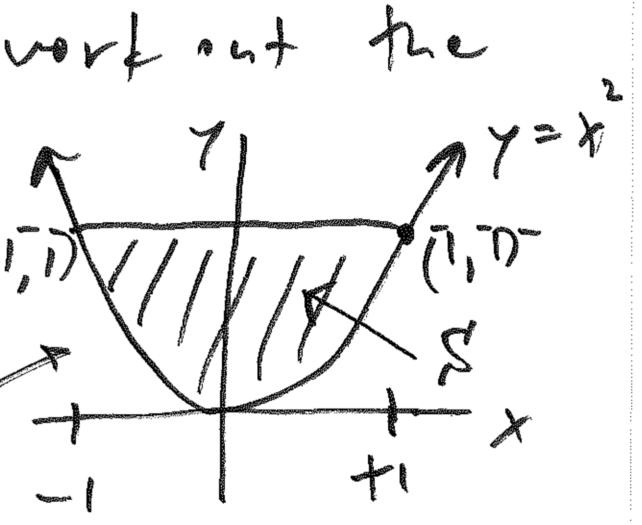
**Example**

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



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

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

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

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

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

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

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

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

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

(100)

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

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

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

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

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

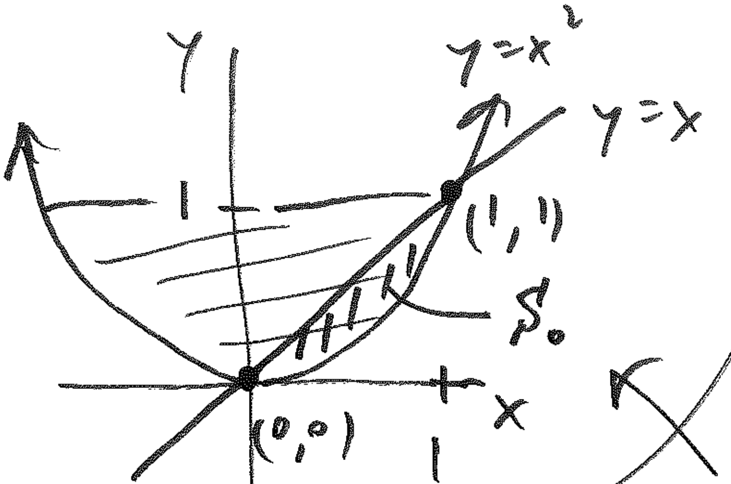
$$= c \int_0^1 y \cdot \frac{1}{3} \left( y^{3/2} - -y^{3/2} \right) dy$$

(101)

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

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

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

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

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 $x_1, x_2, \dots,$ then

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

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

If you model $(X \mid \theta)$ as Bernoulli$(\theta)$
and $\theta \sim$ Uniform$(0,1)$
the joint pf/pdf of $(X, \theta)$ 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 $\theta$ 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 <u>marginal</u> CDF $F_X(x)$ from the joint CDF as $F_X(x) = \lim_{y \to \infty} F_{XY}(x,y)$, and similarly the <u>marginal</u> 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 $f_{XY}(x,y) = \dfrac{\partial^2}{\partial x\,\partial y} F_{XY}(x,y) = \dfrac{\partial^2}{\partial y\,\partial x} F_{XY}(x,y)$

(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 <u>marginal</u> 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

(marginalizing out $Y$)

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

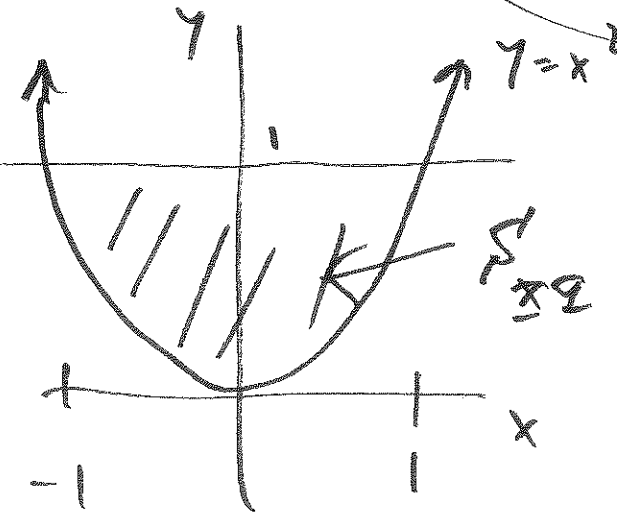
and the marginal pdf $f_Y(y)$ of $Y$ is

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

Earlier example, continued

$(X, Y)$ have joint pdf

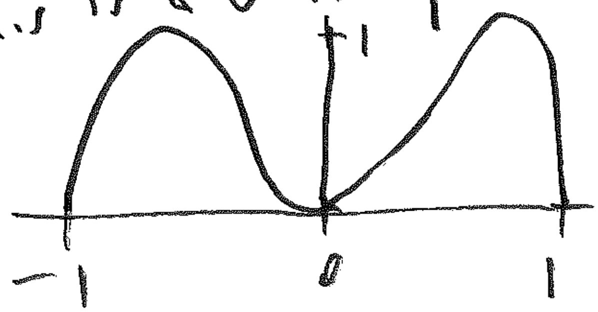
$$f_{XY}(x,y) = \begin{cases} \frac{21}{4}x^2 y, & \underbrace{x^2 \le y \le 1}_{S_{XY}} \\ 0 & \text{else} \end{cases}$$




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

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy = \int_{x^2}^{1} \frac{21}{4}x^2 y\,dy = \begin{cases} \frac{21}{8}x^2(1-x^4) & -1<x<1 \\ 0 & \text{else} \end{cases}$$

this is a weird pdf




(supposed to be symmetric) & bimodal

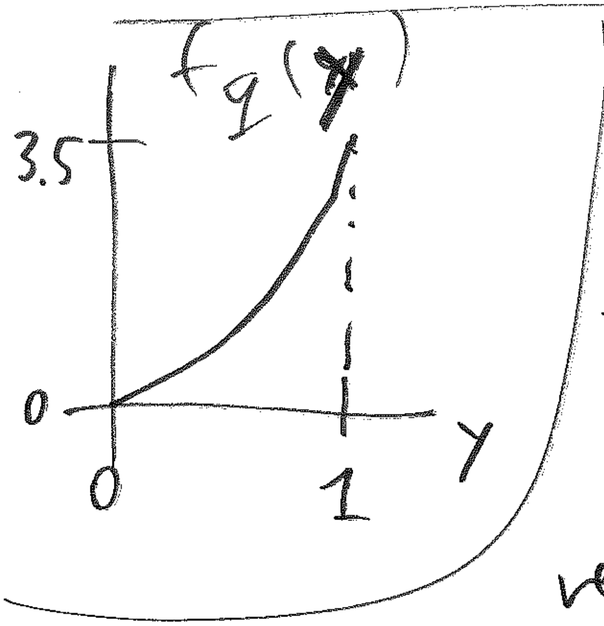
Similarly, the support $S_Y$ of $Y$ is $(0,1)$,

and its marginal pdf is

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

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

(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 $f_Y(y)$, but not the other way around: if all you have is the marginals, they do not uniquely determine the joint.