The random variable $Y$ is defined as a function $f$ of the random variable $X$ for all sets $A$ of real numbers in the real line. The function $f$ is the collection of all probabilities of the form $P_{X}(A)$ for all sets $A$. The probability distribution of a random variable $X$ can take on only the two possible values, $0$ and $1$, at the rational points. You can see this from $\mathbb{Q}$.
This is true of some, but not all, rvs.

Definition: A random variable has a **discrete distribution**, or equivalently, it is a discrete RV, if the set of (distinct) possible values is finite or at most countably infinite; rvs for which the set of possible values is uncountable are called **continuous random variables**.

Example: 1. The RV $X = \begin{cases} 1 & \text{if } z > 0 \\ 0 & \text{otherwise} \end{cases}$ (with $z = \# T-s b a l i n g$) is discrete, taking on only the values $\{0, 1\}$ — such rvs are called dichotomous or binary.
Imagine a scale for weighing things that has a dial you can set to specify how many significant figures of precision you want. Buy a "1 pound" package of butter at your favorite market and weigh it.

If there's no conceptual limit to the number of significant figures you could get, a rv $X = (\text{weight of the package})$ 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 (mass) discrete RV \( I \), the probability function (pmf or pf) of \( I \) is the function \( f_g(y) \) that keeps track of the probability associated with \( I \): \( f_g(y) = P(I = y) \). The set \( \{ y : f_g(y) > 0 \} \) is called the support of (the distribution of) \( I \). (As is almost unique in using “pf,” nearly everybody talks about the pmf.)
Example] In the powerball lottery (see homework 1 problem 2) 5 white balls are drawn at random without replacement from a bin with balls numbered \{1, 2, \ldots, 69\}.

Let \(W_i = \# \text{ of } i\text{th drawn ball.}

Clearly \( p( W_i = w_i ) = \begin{cases} \frac{1}{69} & \text{for } w_i = 1, 2, \ldots, 69 \\ 0 & \text{otherwise} \end{cases} \)

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

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

\[
    f(y) = p(z = y) = \begin{cases} p & \text{for } y = 1 \\ 1-p & y = 0 \\ 0 & \text{else} \end{cases}
\]

Notation:

$Y$ follows a Bernoulli $(p)$ distribution $\iff$ $Y \sim$ Bernoulli $(p)$.
If \( f(y) = P(Y = y) \) for \( y = a, ... , b \)\)
\[
\left\{ \begin{array}{ll}
\frac{1}{b-a+1} & \text{for } y = a, \ldots, b \\
0 & \text{else}
\end{array} \right.
\]

\( Y \sim \text{Uniform} \{a, b\} \iff \text{randomly chosen from} \{a, a+1, \ldots, b\} \)

**Definition** \( n \) 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 \( Y = \# \text{ of successes} \) has the **Binomial distribution**

\[
f_Y(y) = P(Y = y) = \left\{ \begin{array}{ll}
\binom{n}{y} p^y (1-p)^{n-y} & \text{for } y = 0, 1, \ldots, n \\
0 & \text{else}
\end{array} \right.
\]
In shorthand \( Y \sim \text{Binomial}(n, p) \) or \((Y \mid n, p)\)

Let \( B_i = \{1 \text{ if trial } i \text{ is a success}\} \)

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

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

independent.

Notation \( X_i \sim f(x_i) \)

means that all of the \( X_i, X_2, \ldots \)

are independent and identically distributed draws from the distribution with pdf \( f(x_i) \).

Thus with the success/failure trials, \( B_i \sim \text{Bernoulli}(p) \) and

\[(Y = \sum_{i=1}^{n} B_i) \sim \text{Binomial}(n, p)\].
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 floating point decimal numbers carry about 7 sig figs of accuracy.

leading to round off 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, 0.01, 0.02, \ldots, 0.99\}
In the limit with more & more right...

This should go to \( \frac{1}{0.5} \) the continuous uniform distribution

Uniform \((0,1)\) on the unit interval.

The analogue of the discrete \( G \) in this continuous case is the smooth function \( f(u) \),

\[
f(u) = \begin{cases} 1 & \text{for } 0 \leq u \leq 1 \\ 0 & \text{else} \end{cases}
\]

Definition

A random variable \( X \) has a continuous distribution \( \Leftrightarrow \) \( X \) is a continuous rv if there exists a continuous non-negative function \( f_x \) defined on \( \mathbb{R} \) such that for every interval \([a, b] \),

\[
P(a \leq X \leq b) = \int_{a}^{b} f_X(y) \, dy.
\]
In this definition, $a$ can be $-\infty$ and $b$ can be $\infty$.

**Definition:** If $Y$ is a continuous r.v., 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) $Y$.

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

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

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

Weirdly, this doesn't mean that the value \( y \) of \( Z \) is impossible, or it does with discrete rv, it just means that singletons have to have 0 probability (otherwise \( \int_{-\infty}^{\infty} f_z(y) \, dy = \infty \) not 1).
Definition | with \( a \) and \( b \) any two real numbers satisfying \( a < b \), is distributed as 
\[ Y \sim \text{Uniform}(a, b) \iff P(\text{of (a,b)}) \]
\[ = \text{the length of the subinterval} \iff \]
\[ f_Y(y) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq y \leq b \\ 0 & \text{else} \end{cases} \]

Definition | The indicator function

\( \text{true}/\text{false} \)
for any proposition \( A \) is \( I(A) = \begin{cases} 1 & \text{if A is true} \\ 0 & \text{if A is false} \end{cases} \)

People sometimes also write (with \( \subset \)) set
\[ I_A(y) = \begin{cases} 1 & \text{if } y \in A \\ 0 & \text{else} \end{cases} \]
with this definition, \( Y \sim \text{Uniform}(a, b) \)
\[
\Leftrightarrow f_Y(y) = \frac{1}{b-a} I(a \leq y \leq b) = \frac{I_{[a,b]}(y)}{b-a}.
\]

\textbf{Contrast}

\( Y \sim \text{Uniform}(a, b) \) continuous

\( Y \sim \text{Uniform}\{a, b\} \) for \( a, b \) integers

with \( a < b \) \( \iff \) \( Y \) discrete and uniform on \( \{a, a+1, \ldots, b\} \).

\( Y \sim \text{Uniform}\{a, b\} \) for \( a, b \) integers

\( Y \sim \text{Uniform}(a, b) \) continuous

Density and probability are not the same thing

Density value \( f_Y(y) \)

are themselves not probabilities; for example, they can easily be \( \geq 1 \) and even be \( \infty \).
Density values define probability: \( P(a \leq X \leq b) = \int_{a}^{b} f_{X}(y) \, dy \).

For small \( \varepsilon > 0 \) you can see from this sketch that:

\[
P\left(a - \frac{\varepsilon}{2} \leq Y \leq a + \frac{\varepsilon}{2}\right) = \int_{a - \frac{\varepsilon}{2}}^{a + \frac{\varepsilon}{2}} f_{Y}(y) \, dy \]

Example (triangular distribution)

\( f_{X}(y) \) line with slope \( m = \frac{c}{b-a} \)

Can a continuous rv 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 \( \frac{c}{b-a} \), and passes through the point \((a, 0)\), so the equation of the line is

\[
\gamma - \gamma_a = m(x - x_a) \quad \Rightarrow \quad \gamma = \frac{c}{b-a} (x - a) + f_a(\gamma)
\]

Density have to integrate to 1,

\[
\int_a^b \frac{c}{b-a} (x-a) \, dx = 1 \quad \Rightarrow \quad c = \frac{2}{b-a}
\]

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 $y = a$ and rises linearly to

$$f(y) = \begin{cases} \frac{2}{b-a} & a < y < b \\ \frac{(y-a)^2}{(b-a)^2} & b < y < \frac{b+a}{2} \\ 0 & \text{else} \end{cases}$$

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

Example: with the triangular distribution above, what is $P(a < X < \frac{b-a}{2})$?

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

Easy(?): $\frac{1}{2} \cdot \frac{b-a}{2} \cdot \frac{b-a}{2} = \frac{(b-a)^2}{8}$
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. 104)

\[ Y = \text{voltage in an electrical system} \]

In practice, it 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. If we give as an example the pdf

\[
 f_E(y) = \frac{1}{(1+y)^2} I(y > 0) 
\]

\[
 = \frac{1}{(1+y)^2} \bigg|_0^\infty = -1(0 - 1) = 1
\]

Check:

\[
 \int_0^\infty \frac{1}{(1+y)^2} dy = 1 
\]
You can check that \( \int_{100}^{\infty} \frac{1}{100(1+x)} \, dx = 1 - \frac{1}{100} = 0.001 \), so the right tail beyond \( X = 1000 \) 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. 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...
$T =$ survival time in days from beginning of the trial; however, one good thing can happen, some patients may still be alive at the time $T_{\text{end}}$, at which the trial finishes. Your model for $T$ would then have a continuous part for $0 \leq T \leq T_{\text{end}}$ and a discrete lump of probability $p$ at $T = T_{\text{end}}$ signifying $(T > T_{\text{end}})$ but we don't know what $T$ would have been if we could have observed it:

$$\int_0^{T_{\text{end}}} f_T(t) \, dt = 1 - p \quad \text{and} \quad P(T > T_{\text{end}}) = p.$$  

(Example, see p. 10 of doc. with notes)
Unifying idea connecting discrete & continuous rvs

Q: Is there something that uniquely characterizes the distribution of $X$, both when $X$ is discrete & when it's continuous?

A: Yes, the cumulative distribution function $F_X(y)$ is defined to be

$$F_X(y) = P(X \leq y) \quad \text{for all} \quad -\infty < y < \infty$$
and it's also clear that a cdf $F_{\Sigma}(y)$ has to be a non-decreasing function of $y$: if $y_1 < y_2$, then $F_{\Sigma}(y_1) \leq F_{\Sigma}(y_2)$.

Furthermore, $\lim_{y \to -\infty} F_{\Sigma}(y) = 0$ and $\lim_{y \to +\infty} F_{\Sigma}(y) = 1$. CDFs can be continuous or continuous on all of $\mathbb{R}$ but certainly don't have to be (see the CDF of the Bernoulli ($p$) distribution).

Technical fact:

Def. $F_{\Sigma}(y^-) = \lim_{y^* \to y^-} F_{\Sigma}(y^*) =$ limit from the left.
Example: \( I \sim \text{Bernoulli} (p) \)

\[
P(I = y) = \begin{cases} 
  p & \text{for } y = 1 \\
  1-p & \text{else} \\
  0 & \text{otherwise}
\end{cases}
\]

Notice that this is a clever way to write this.

pf: \[ P(I = y) = p^y (1-p)^{1-y} I_{\{0,1\}}(y) \]

The cdf of \( I \) is 0 for \( y < 0 \); at \( I = 0 \) it jumps up to \((1-p)\) and stays there for \( 0 \leq y < 1 \); and at \( I = 1 \) it jumps up to 1 & stays there for \( y \geq 1 \). You can see that in general:

\[ 0 \leq F_I(y) \leq 1 \]
Definition: \( F_\Sigma (y^+) \triangleq \lim_{y^* \to y} F_\Sigma (y^*) = \lim_{y^* \downarrow y} F_\Sigma (y) \)

Technical fact: \( F_\Sigma (y) = F_\Sigma (y^+) \) for all \(-\infty < y < \infty\)

People call this continuity from the right or continuity from above.

Consequences of the CDF definition:

1. \( P(\Sigma > y) = 1 - F_\Sigma (y) \)
2. For all \( y_1, y_2 \) with \( y_1 < y_2 \), \( P(y_1 < \Sigma \leq y_2) = F_\Sigma (y_2) - F_\Sigma (y_1) \)

If \( F_\Sigma (y^-) = F_\Sigma (y^+) = F_\Sigma (y) \) then \( F_\Sigma \) is continuous.

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

$(\text{cdf}) \implies (pdf)$

$Y$ continuous: $\forall y$

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

$= \int_{y_1}^{y_2} f_Y(y) dy$

and thus

**Theorem**

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

$F_Y(y) = \int_{-\infty}^{y} f_Y(t) dt$ and

$\frac{d}{dy} F_Y(y) = f_Y(y)$.
In other words, $F_\mathcal{E}(y)$ is the cumulative distribution function (CDF) of $\mathcal{E}(\gamma)$ (and its derivative). $F_\mathcal{E}(y)$ is an anti-derivative of $f_\mathcal{E}(y)$.

**Definition** $Y$ follows an exponential distribution with parameter $\lambda > 0$ if

$$f_\mathcal{E}(y) = \begin{cases} \lambda e^{-\lambda y} & y > 0 \\ 0 & y \leq 0 \end{cases}$$

The exponential distribution 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.

\( \Gamma \) exponentially distributed with parameter \( \lambda > 0 \) for \( \gamma > 0 \):

\[
\mathbb{E}(\gamma) = \int_{0}^{\infty} \frac{\lambda}{\Gamma(\gamma)} e^{-\lambda t} dt = \int_{0}^{\infty} \frac{\lambda}{\Gamma(\gamma)} e^{-\lambda t} dt
\]

\[
= \frac{\lambda}{\Gamma(\gamma)} \left[ -e^{-\lambda t} \right]_{0}^{\infty} = 1 - e^{-\lambda \gamma}
\]

So:

\[
\mathbb{F}_\gamma(\gamma) = \begin{cases} 0 & \text{for } \gamma \leq 0 \\ 1 - e^{-\lambda \gamma} & \gamma > 0 \end{cases}
\]
What's the place $y_p$ on the positive part of $R$ where $P(0 \leq Y \leq y_p) = p$?

Well, $P(0 \leq Y \leq y_p) = F_Y(y_p) = p$.

So $y_p = F_Y^{-1}(p)$.

Def. $y_p$ is called the $p_{th}$ quantile of the distribution of $Y$. 

$1 - p = e^{-x}$

$\log (1 - p) = -x$

$y_p = -\frac{\log (1 - p)}{\mu} = F_Y^{-1}(p)$ or the $100p^{th}$ percentile of the distribution of $Y$. 
Some care is required when $f$ is discrete or mixed.

**General definition**

$Y$ is rv with CDF $F_Y(y)$. For all $0 < p < 1$ define

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

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

$F_Y(y)$

50% 50%

One way to define the center of a distribution is to find the 50th percentile.
Definition: The $\frac{1}{4}$ quantile = the 25th percentile of a distribution is called the median of the dist.

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

Definition: The $\frac{1}{4}$ quantile = the 25th percentile is the lower quantile; the $\frac{3}{4}$ quantile = the 75th percentile is the upper quantile.

$\left( F_{Y}(0.25), F_{Y}(0.75) \right) = \text{interquartile range (IQR)}$
Example: \( I \sim \text{Uniform}(a, b) \); then,

\[
F_I(y) = \begin{cases} 
0 & \text{for } y < a \\
\frac{y-a}{b-a} & a \leq y \leq b \\
1 & y > b 
\end{cases}
\]

Easy to invert \( F_I \):

\[
F_I^{-1}(p) = (1-p)a + pb \quad \text{for } 0 \leq p < 1
\]

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

Studying two random variables at a time:

Def: \( X, Y \) r.v.'s: the joint (or bivariate) distribution of \((X, Y)\) is the collection \( P[(X, Y) \in C] \) of all probabilities for all sets \( C \subseteq \mathbb{R}^2 \) such that \((X, Y) \in C\) isn't weird.
Case 1: $X$ and $Y$ both discrete

Def. $X$, $Y$ rv. $\rightarrow$ 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 function (joint pmf) of $(X,Y)$ discrete is the function $f(x,y) = P(X=x, Y=y)$.

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

Consequences

1. $\sum f(x,y) = 1$ for all $(x,y) \in \text{support of } f_{xy}$ (non-weird)

2. For any set $C$ of ordered pairs $(x,y)$, $P((X,Y) \in C) = \sum f_{xy}(x,y)$
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) such that for every (non-empty) subset $C$ of the plane $P[\{X,Y\} \in C] = \iint_{\{x,y\}} f_{X,Y}(x,y) \, dx \, dy$.

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

Immediate Consequences

1. For all $(x,y)$ in $\mathbb{R}^2$, $f_{X,Y}(x,y) \geq 0$ and $\iint_{\mathbb{R}^2} f_{X,Y}(x,y) \, dx \, dy = 1$. 
(2) If \((x, y)\) have a \(\text{continuous joint distribution}\), \(x\) and \(y\) each have a \(\text{continuous (marginal) univariate distribution}\) when considered separately.

(3) For all continuous pdfs \(f_{X,Y}(x, y)\),

(a) Every individual point, or every \(\text{(or set)}\) in \(\mathbb{R}^2\), countably infinite sequence of points, has probability 0 under \(f_{X,Y}\).

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

**Example**

Joint distributions \((X, Y)\) have joint pdf that can lead to tricky integrals.

Suppose that \( f_{X,Y}(x,y) = \begin{cases} cx^2 & \text{for } 0 \leq x, y \leq 1 \\ 0 & \text{else} \end{cases} \)

Let's work out the normalizing constant.

The support of \( f_{X,Y} \) is the shaded region here.
\[ 1 = \int \int f(x, y) \, dy \, dx \]

\[ = \int \int f(x, y) \, dy \, dx \]

\[ = \int_{-1}^{1} \int_{1}^{x^2} f(x, y) \, dy \, dx \]

\[ = \int_{-1}^{1} \left( \int_{1}^{x^2} f(x, y) \, dy \right) \, dx \]

\[ = \int_{-1}^{1} x^2 \left( \frac{y^2}{2} \right) \, dy \]

\[ = \int_{-1}^{1} x^2 \left( \frac{y^2}{2} \right) \, dy \]

\[ = \frac{1}{2} \int_{-1}^{1} x^4 \, dx - \frac{1}{2} \int_{-1}^{1} x^6 \, dx \]

\[ = \frac{1}{2} \left( \frac{x^3}{3} \right)_{-1}^{1} - \frac{1}{2} \left( \frac{x^7}{7} \right)_{-1}^{1} \]

\[ = \frac{1}{2} \left( \frac{1}{3} \right) - \frac{1}{2} \left( \frac{1}{7} \right) = \frac{1}{21} \cdot c = 1 \]
so \( c = \frac{21}{4} \)

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 = \iiint_{S} f_{\mathbb{R}^2}(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 \\
= \int_{0}^{1} cy \left( \frac{y^3}{3} - \frac{y^3}{3} \right) \, dy \\
= c \int_{0}^{1} y \cdot \frac{1}{3} (y^{3/2} - y^{3/2}) \, dy
\[ z = \frac{1}{3} \int_0^1 2y^{5/2} \, dy = \frac{2c}{3} \left( \frac{y^{7/2}}{y^{7/2}} \right) \bigg|_0^1 \quad \text{(10)} \]

\[ = \frac{4}{21} c \text{ as before} \]

(always hope to agree, of course).

**Example, continued**

Let's compute \( P(X \geq 2) \)

\[ P(X \geq 2) = \iint_{S} f_{XY}(x, y) \, dy \, dx \]

The relevant part \( S' \) of \( S \) where \( X \geq 2 \) is sketched here, \( S' \)

\[ P(X \geq 2) = \int_0^1 \int_{x^2}^{\min(x, y)} 21x^2y \, dy \, dx = \frac{3}{20} \]

\[ \text{(..)} \]

\[ = \int_0^1 \int_{x^2}^{\min(x, y)} 21x^2y \, dy \, dx = \frac{3}{20} \]
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 $f_{X,Y}(x, y)$ defined on $\mathbb{R}^2$ such that for every pair of (non-void) subsets $A$ and $B$ of $\mathbb{R}$ (assure integral exists)

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

Then $f_{X,Y}$ is the joint **pdf** of $(X, Y)$ if $X$ takes on values $x_1, x_2, \ldots,$

Immediate consequence then

$$\sum_{-\infty}^{\infty} \int f(x_i, y) \, dy = 1.$$
Example randomized controlled (clinical) trial; patients in $\Theta$ get a treatment, patients in $\Theta'$ get a placebo. Outcome is success (e.g., cancer goes into remission) or failure; let $X_i = \{1\text{ if patient } i \text{ is a success, } 0\text{ else}\}$.

$\Theta$ (unknown) and let $\pi$ 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 $\Theta$ is continuous on $(0, 1)$ and $(X_i, \Theta)$ has a mixed bivariate distribution.
If you model \((X|\theta)\) as Bernoulli(\(\theta\)) and \(\theta\) as uniform(0,1), the joint pdf of \((X, \theta)\) would be

\[
f_{X,\theta}(x, \theta) = \begin{cases} \theta^x(1-\theta)^{1-x} & \text{for } (x=0,1) \\ 0 & \text{else} \end{cases}
\]

Then (e.g.) \(P(X = 1) = P(X = 1 \text{ and } \theta \text{ is anything between } 0 \text{ and } 1) \)

\[
= \int_0^1 \theta^1(1-\theta)^{0-1} \, d\theta = \int_0^1 \theta \, d\theta = \frac{1}{2}.
\]

**Bivariate CDFs**

The joint CDF of two rvs \(X\) and \(Y\) is the function \(F_{X,Y}(x,y)\) satisfying \(F_{X,Y}(x,y) = P(X \leq x \text{ and } Y \leq y)\) for all \(-\infty < x < \infty\) and \(-\infty < y < \infty\).
Consequence of this definition:

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

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 \(f_{X,Y}(x,y)\),

then

\[
F_{X,Y}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(r,s) \, dr \, ds
\]

and

\[
f_{X,Y}(x,y) = \frac{d^2}{dx \, dy} F_{X,Y}(x,y) = \frac{d^2}{dy \, dx} F_{X,Y}(x,y)\]

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

**Consequence 3** If \((X, Y)\) have a discrete joint distribution with joint pdf \(f_{X,Y}(x,y)\), then the marginal pdf \(f_X(x)\) of \(X\) is

\[
f_X(x) = \sum_y f_{X,Y}(x,y)
\]

(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.

If \((X, Y)\) have a continuous joint distribution with joint pdf \(f_{X,Y}(x, y)\), the marginal pdf \(f_X(x)\) of \(X\) is (marginalizing out \(Y\))
\[
f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, dy \quad (\text{for all } -\infty < x < \infty)
\]
and the marginal pdf \(f_Y(y)\) of \(Y\) is
\[
f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) \, dx \quad (\text{for all } -\infty < y < \infty).
\]
Earlier example, continued

Let \((X, Y)\) have joint pdf

\[
f_{X,Y}(x, y) = \begin{cases} \frac{21}{4} x^2 y, & x^2 \leq y \leq 1 \\ 0, & \text{else} \end{cases}
\]

You can see from the sketch of the support \(S\) of \(f_{X,Y}(x, y)\) that

\[
-1 \leq x \leq 1, \quad \text{so the support of } X \text{ is } (-1, 1), \quad \text{and its marginal pdf is}
\]

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

This is a weird pdf

(supposed to be symmetric & bimodal)
Similarly, the support of \( F \) is \( \mathbb{R}^n \), and its pdf is

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

\[
= \begin{cases} 
\frac{7}{2} y^3 & \text{for } 0 < y < 1 \\
0 & \text{else}
\end{cases}
\]

(14 May 11)

Consequences: If you have continued 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.