Sometimes a joint pmf is not directly specified, but rather given in terms of a conditional pmf and a marginal pmf —
From the definition of conditional pmfs we have

(1) $$p_{X,Y}(x,y) = p_{X|Y}(x|y)\, p_Y(y)$$

and, also,

(2) $$p_{X,Y}(x,y) = p_{Y|X}(y|x)\, p_X(x)$$

These formulas allow us to construct the joint pmf from a marginal pmf and a conditional pmf. For example, in (1) if $p_Y(y)$ and $p_{X|Y}(x|y)$ are known, then (1) not only allows us to compute $p_{X,Y}(x,y)$ but also the marginal of X:

$$p_X(x) = \sum_y p_{X|Y}(x|y)\, p_Y(y)$$

which is a special case of the <u>total probability law</u>.

Here is an example of how one might use these formulas.

Ex. Imagine we play the following game.

In the first stage we roll a balanced 6-sided die until we observe a '6' for the first time. Let Y be the trial on which we observe this first '6'. so that $Y \sim \text{Geometric}(1/6)$.

If $Y=y$, we toss a balanced coin y times and let X be the number of successes tossed. and we win \$X dollars.

We are told here that $X | Y=y \sim \text{binomial}(y, \frac{1}{2})$.

Let's find the joint pmf of X and Y:

$$P_{X,Y}(x,y) = P_{X|Y}(x|y)\, P_Y(y)$$

$$= \binom{y}{x}\left(\frac{1}{2}\right)^x\left(1-\frac{1}{2}\right)^{y-x} \cdot \left(1-\frac{1}{6}\right)^{y-1}\frac{1}{6}.$$

$$= \binom{y}{x}\left(\frac{1}{2}\right)^y\left(\frac{5}{6}\right)^y\left(\frac{5}{6}\right)^{-1}\cdot\frac{1}{6}$$

$$= \frac{1}{5}\binom{y}{x}\left(\frac{5}{12}\right)^y. \quad \text{for } y=1,2,3,4,\ldots$$

and $x=0,1,2,\ldots,y$.

From here we can ask:

What is the probability we win \$0 playing this game?

I.e. we might want to compute $P(X=0)$.

But by the law of total probability

$$P(X=0) = \sum_{y=1}^{\infty} P_{X|Y}(0,y)\, P_Y(y)$$

$$= \sum_{y=1}^{\infty} \frac{1}{5} \binom{y}{0} \left(\frac{5}{12}\right)^y$$

$$= \sum_{y=1}^{\infty} \frac{1}{5}\left(\frac{5}{12}\right)^y \overset{\text{geometric series}}{=} \frac{\frac{1}{5}\left(\frac{5}{12}\right)}{1-\frac{5}{12}} = \frac{\frac{1}{12}}{\frac{7}{12}}$$

$$= \frac{1}{7}.$$

With probability $= 1/7$ we will win No money.

How about the maginal pmf of X: We already saw that

$$p_X(0) = \sum_{y=1}^{\infty} p_{X,Y}(0,y) = \sum_{y=1}^{\infty} \frac{1}{5}\left(\frac{5}{12}\right)^y$$

$$= \frac{\frac{1}{12}}{1-\frac{5}{12}} = \frac{1}{7}.$$

Now,

if $x \geq 1$

$$p_X(x) = \sum_{y=x}^{\infty} p_{X,Y}(x,y) = \sum_{y=x}^{\infty} \frac{1}{5}\binom{y}{x}\left(\frac{5}{12}\right)^y$$

y needs to be at least x

Now we use the fact that

$$\sum_{y=x}^{\infty} \binom{y}{x}\left(\frac{7}{12}\right)^{x+1}\left(\frac{5}{12}\right)^{y-x} = 1.$$

This is the Pascal (or Negative binomial) distribution
The $(x+1)^{st}$ success happens on trial $y+1$

which is equivalent to

$$\sum_{y=x}^{\infty} \binom{y}{x}\left(\frac{5}{12}\right)^y = \left(\frac{5}{12}\right)^x\left(\frac{12}{7}\right)^x \cdot \frac{12}{7} = \frac{12}{7}\left(\frac{5}{7}\right)^x$$

Therefore, when $x \geq 1$

$$\sum_{y=x}^{\infty} \frac{1}{5}\binom{y}{x}\left(\frac{5}{12}\right)^y = \frac{12}{35}\left(\frac{5}{7}\right)^x$$

So we have

$$p_X(x) = \begin{cases} \frac{1}{7} & \text{if } x = 0 \\ \frac{12}{35}\left(\frac{5}{7}\right)^x & \text{if } x \geq 1. \end{cases}$$

Let's check that this is a pmf.

Check:

$$\sum_{x=0}^{\infty} p_X(x) = p_X(0) + \sum_{x=1}^{\infty} p_X(x)$$

$$= \frac{1}{7} + \sum_{x=1}^{\infty} \frac{12}{35}\left(\frac{5}{7}\right)^x$$

$$= \frac{1}{7} + \frac{12}{35}\left\{ \frac{5/7}{1 - 5/7} \right\}$$

$$= \frac{1}{7} + \frac{12}{35}\left\{ \frac{5}{2} \right\} = \frac{1}{7} + \frac{6}{7} = 1.$$

Conditional Expectations

If we have two jointly discrete r.v.s X and Y and we know (say) the conditional pmf of X given $Y=y$, i.e. if we know $p_{X|Y}(x|y)$ then we define the conditional expectation of X given $Y=y$ as

$$E(X|Y=y) = \sum_x x \cdot p_{X|Y}(x|y).$$

That is, the conditional expectation of X is just the weighted average of the values of X weighted by the conditional probabilities instead!

For example, in the dice rolling experiment where $D = |X_1 - X_2|$ and $W = \max\{X_1, X_2\}$ from last lecture:

The joint pmf of D and W is given in this table:

	$D=0$	$D=1$	$D=2$	$D=3$	$D=4$	$D=5$
$W=1$	$\frac{1}{36}$	0	0	0	0	0
$W=2$	$\frac{1}{36}$	$\frac{2}{36}$	0	0	0	0
$W=3$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	0	0	0
$W=4$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	0	0
$W=5$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	0
$W=6$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$

Compute the condition expectation of the maximum of X_1 and X_2 given that $|X_1 - X_2| = 2$, i.e.

Compute $E(W \mid D=2)$.

To compute this we need to know $P_{W|D}(w|2)$

But from the table we can easily find this: it is

w	3	4	5	6
$P_{W\|D}(w\|2)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

and therefore,

$$E(W \mid D=2) = 3\left(\tfrac{1}{4}\right) + 4\left(\tfrac{1}{4}\right) + 5\left(\tfrac{1}{4}\right) + 6\left(\tfrac{1}{4}\right) = 4.5.$$

$\boxtimes$

What about $E(W \mid D=1)$?

w	2	3	4	5	6
$p_{W \mid D}(w \mid 1)$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$

So

$$E(W \mid D=1) = 2\left(\tfrac{1}{5}\right) + 3\left(\tfrac{1}{5}\right) + 4\left(\tfrac{1}{5}\right) + 5\left(\tfrac{1}{5}\right) + 6\left(\tfrac{1}{5}\right)$$

$$= 4.$$

How about $E(W^2 \mid D=1)$?

$$E(W^2 \mid D=1) = 2^2\left(\tfrac{1}{5}\right) + 3^2\left(\tfrac{1}{5}\right) + 4^2\left(\tfrac{1}{5}\right) + 5^2\left(\tfrac{1}{5}\right) + 6^2\left(\tfrac{1}{5}\right)$$

$$= \frac{4+9+16+25+36}{5} = \frac{90}{5} = 18.$$

In fact the general rules of expected value follow for conditional expectation as well, since it is also an expected value just now it is with respect to a conditional pmf instead of (an unconditional) pmf.

Suppose $X \sim$ geometric(p).

Find $P(X > x)$ for each $x = 1, 2, 3, \ldots$

$$P(X > x) = \sum_{j=x+1}^{\infty} p(1-p)^{j-1}$$

$$= \frac{p(1-p)^x}{1-(1-p)} = (1-p)^x.$$

So $P(X > x) = (1-p)^x$ for each $x = 1, 2, 3, \ldots$

We also know $E(X) = \frac{1}{p}$ by earlier work.

What if we are told that $X > 3$, what would be the expected value of X?

That is, what is $E(X \mid \underbrace{X > 3})$

this is now an event of positive probability.

Then

$$E(X \mid X > 3) = \sum_x x \, p_{X \mid (X>3)}(x \mid X > 3)$$

But $$p_{X|(X>3)}(x \mid X>3) = \frac{P(X=x,\ X>3)}{P(X>3)}$$ by the usual formula for conditional probability since the conditioning event has pos. prob.

$$= \frac{P(X=x)}{(1-p)^3} \quad \text{when } x > 3.$$

i.e. $x = 4, 5, 6, \ldots$

So

$$E(X \mid X>3) = \sum_{x=4}^{\infty} x \cdot \frac{P(X=x)}{(1-p)^3}$$

$$= \frac{1}{(1-p)^3} \sum_{x=4}^{\infty} x\, p_X(x).$$

to compute $\sum_{x=4}^{\infty} x\, p_X(x) = 4p(1-p)^3 + 5p(1-p)^4 + 6p(1-p)^5 + \cdots = S$

$$S = 4p(1-p)^3 + 5p(1-p)^4 + 6p(1-p)^5 + \cdots$$
$$(1-p)S = \qquad 4p(1-p)^4 + 5p(1-p)^5 + \cdots$$

$$pS = 4p(1-p)^3 + p(1-p)^4 + p(1-p)^5 + p(1-p)^6 + \cdots$$

$$= 4p(1-p)^3 + \left\{ \frac{p(1-p)^4}{1-(1-p)} \right\} = 4p(1-p)^3 + (1-p)^4$$

So, $S = 4(1-p)^3 + \dfrac{(1-p)^4}{p}$.

So

$$E(X \mid X>3) = \frac{1}{(1-p)^3}\left\{ 4(1-p)^3 + \frac{(1-p)^4}{p} \right\}$$

$$= 4 + \frac{1-p}{p}.$$

$$= 3 + \frac{1}{p}.$$

Law of total probability for Expectations:

$$E(X) = \sum_y E(X|Y=y)\,P(Y=y)$$

Sum is over all possible values of the discrete r.v. Y.

also if $A_1, A_2, A_3, \ldots, A_n$ is a partition of Ω

$$E(X) = \sum_{i=1}^{n} E(X|A_i)\,P(A_i)$$

To see why these are true I will show the first one

We already know

$$p_{X,Y}(x,y) = p_{X|Y}(x|y)\,p_Y(y).$$

equivalently, $P[X=x, Y=y] = P(X=x|Y=y)\,P(Y=y)$

So we "integrate out" the Y variable to capture the marginal of X:

$$p_X(x) = \sum_y p_{X|Y}(x|y)\,p_Y(y).$$

But then

$$E(X) = \sum_x x\,p_X(x) = \sum_x x \sum_y p_{X|Y}(x|y)\,p_Y(y) = \sum_y \left\{ \sum_x x\,p_{X|Y}(x|y) \right\} p_Y(y)$$

$$= \sum_y E(X|Y=y)\,P[Y=y]$$ ∎

Example Suppose $Y \sim \text{geometric}(\frac{1}{6})$ and $X | Y=y \sim \text{binomial}(y, \frac{1}{2})$

Find $E(X)$.

Here, I will use (without proof) that $E(Y) = 6$

and $E(X | Y=y) = \frac{y}{2}$.

Then

$$E(X) = \sum_y E(X | Y=y) P[Y=y]$$

$$= \sum_y \frac{y}{2} \cdot \left(\frac{5}{6}\right)^{y-1} \frac{1}{6} = \frac{1}{2} \sum_y y \left(\frac{5}{6}\right)^{y-1} \cdot \frac{1}{6}$$

$$= \frac{1}{2} \cdot 6 = 3.$$

Recall that from last lecture it took quite some work to show

$$p_X(x) = \begin{cases} \frac{1}{7} & \text{if } x = 0 \\ \frac{12}{35}\left(\frac{5}{7}\right)^x & \text{if } x = 1, 2, 3, \ldots \end{cases}$$

and if we were asked to compute $E(X)$ then computing it directly would require us to find this marginal pmf and then after that comput $\sum_x x \, p_X(x)$.

Independence of random variables.

Recall what it meant for two events A, B to be independent

$$P(A \cap B) = P(A) \cdot P(B).$$

Motivated by this definition suppose X, Y are jointly discrete. Then we will want

$$P[X = x, Y = y] = P[X = x] \cdot P[Y = y]$$

for all possible values (x, y) of the r.v.s.

I.e.,

$$p_{X,Y}(x, y) = p_X(x)\, p_Y(y). \text{ for all possible } x, y$$

the joint pmf factors as the product of its marginal pmfs.

With this definition, if X and Y are independent, then

$$P[X = x \mid Y = y] = \frac{P[X = x, Y = y]}{P(Y = y)} = \frac{P[X = x]\, \cancel{P[Y = y]}}{\cancel{P[Y = y]}} = P[X = x]$$

That is, the information $Y = y$ did not change $P(X = x)$.

Remark

Checking whether or not two discrete r.v.s are independent would require us to check

$$P[X=x, Y=y] = P(X=x] P[Y=y]$$

for all values of x and y. (Tedious for sure)

However, showing two r.v.s are <u>dependent</u> (Not independent) just <u>requires one example</u> of an <u>x and y</u> such that

$$P[X=x, Y=y] \neq P[X=x] P[Y=y]$$

$$p_{X,Y}(x,y) \neq p_X(x)\, p_Y(y)$$

Here is the joint pmf of X, Y

Ex.

$\frac{1}{6} = \frac{1}{3} \times \frac{1}{2}$

$\frac{1}{12} = \frac{1}{3} \times \frac{1}{4}$

X \ Y	0	1	2	
-1	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$	$1/3$
1	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$1/2$
3	$\frac{1}{24}$	$\frac{1}{12}$	$\frac{1}{24}$	$1/6$
	$1/4$	$1/2$	$1/4$	

$\frac{1}{12} = \frac{1}{3} \times \frac{1}{4}$

Are these r.v.s independent? or not?

Notice that the probability mass in each cell is the product of the corresponding marginal masses of the cell.

Therefore, these r.v.s are <u>independent</u>.

Checking whether or not two (or more) r.v.s are independent can be tedious. However, it doesn't diminish the importance of this concept, and very often we will typically ASSUME r.v.s are independent*

* this can usually be guaranteed by a statistical design of an experiment.

The Consequences of independence.

Suppose X and Y are jointly discrete and independent and that $E(X)$ and $E(Y)$ exist

Then

$$E(XY) = E(X)E(Y).$$

used independence here

<u>Proof</u> $E(XY) = \sum_x \sum_y xy\, p_{X,Y}(x,y) = \sum_x \sum_y xy\, p_X(x) p_Y(y)$

$$= \sum_x x\, p_X(x) \sum_y y\, p_Y(y) = E(X)E(Y).$$

Moreover, if X and Y are independent and g and h are any functions such that $E(g(X))$ and $E(h(Y))$ exist then

$$E(g(X)h(Y)) = E(g(X))\,E(h(Y)).$$

An application of this last result is the following.

Suppose X and Y are independent r.v.s.

Compute $Var(X+Y)$.

Solution.

$$Var(X+Y) = E([X+Y]^2) - [E(X+Y)]^2$$

$$= E[X^2 + 2XY + Y^2] - \{E(X) + E(Y)\}^2$$

$$= E(X^2) + 2E(XY) + E(Y^2) - \{[E(X)]^2 + [E(Y)]^2 + 2E(X)E(Y)\}$$

$$= E(X^2) - [E(X)]^2 + E(Y^2) - \{E(Y)\}^2$$

$$= Var(X) + Var(Y)$$

That is, when the r.v.s X and Y are independent we have

$$Var(X+Y) = Var(X) + Var(Y)$$

In words, the Variance of the sum of X and Y is the sum of the Variances of X and Y.

Independence (continued)

random variables X, Y (assumed jointly discrete) are independent provided

(*) $$p_{X,Y}(x,y) = p_X(x)\, p_Y(y) \text{ for all possible } x, y.$$

Consequently, we saw that when g and h are any functions for which $E(g(X))$ and $E(h(Y))$ are finite ~~then~~ if X, Y independent

$$E(g(X)h(Y)) = E(g(X))\, E(h(Y)).$$

It is also true that when X and Y are independent

$$p_{X|Y}(x|y) = p_X(x) \quad \text{for all } x \text{ (and } y\text{)}.$$

This is because

use independence (*)

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)} = \frac{p_X(x)\cancel{p_Y(y)}}{\cancel{p_Y(y)}} = p_X(x).$$

and therefore as a consequence when X and Y are indep.

$$E(X|Y=y) = \sum_x x\, p_{X|Y}(x|y) = \sum_x x\, p_X(x) = E(X).$$

The conditional expectation is the same as the (unconditional) expectation.

We also saw that if X, Y are independent, then

$$(\star) \qquad Var(X+Y) = Var(X) + Var(Y) \quad \text{must follow.}$$

(this assumes, of course, that the variances of each of X and Y exist.)

from $(\star)$ it follows when X, Y, Z are independent

$$Var(X+Y+Z) = Var(X+Y) + Var(Z)$$

since $X+Y$ is independent of Z, and moreover

since X is independent of Y

$$= Var(X) + Var(Y) + Var(Z).$$

$$\boxed{\begin{array}{l}\text{In general, when } X_1, X_2, \ldots, X_n \text{ are independent} \\ Var(X_1 + X_2 + \cdots + X_n) = Var(X_1) + Var(X_2) + \cdots + Var(X_n)\end{array}}$$

we will generalize this result later to the case where $X_1, X_2, \ldots, X_n$ are not independent. when we develop the concept of <u>Covariance</u>.

$\blacksquare$

Continuous Random Variables

We now introduce random variables whose set of possible values form a "continuum" i.e, a subinterval of the real line.

For example, the random variable that measures the amount of time T (in seconds) that a person runs a 100 meter dash. in this case $T \in (0, \infty)$, say.

or the x-coordinate X of a dart thrown at random on a dart board, in this case $X \in (-1, 1)$, say.

Specifically, we call a random variable X continuous if there exists a function $f_X(x)$ called a probability density function, or pdf for short

i.e., a function $f : \mathbb{R} \to \mathbb{R}$ s.t.

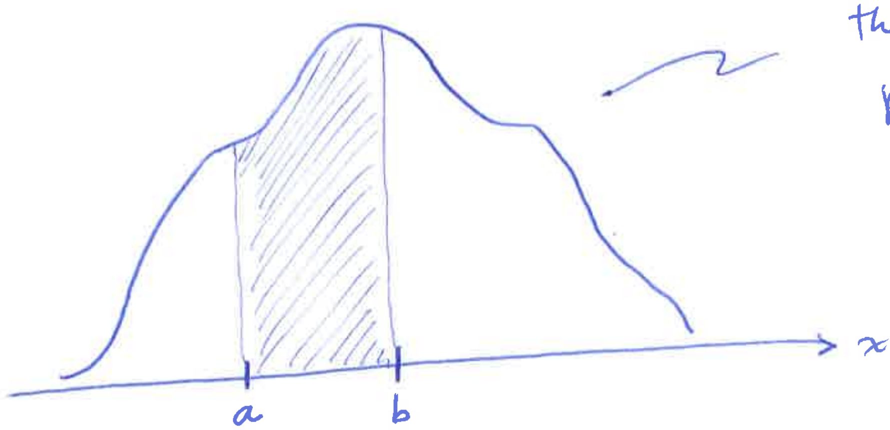
(1) $f(x) \geq 0$ for all x

and

(2) $\displaystyle\int_{-\infty}^{\infty} f(x)\,dx = 1.$

such that

$$P(X \in B) = \int_B f(x)\,dx \quad \text{or} \quad P(a \leq X \leq b) = \int_a^b f(x)\,dx.$$



Suppose this is the graph of the pdf of X.

Then the area under the pdf between a and b is the probability that $X \in [a,b]$.

The function $f(x)$ is returning the probability "denisty" at x that is, probability mass per unit value of the r.v.

So that $\underbrace{f(x)}_{\frac{\text{prob. mass}}{\text{value}}} \underbrace{dx}_{\text{value}}$ ← in units of prob. mass

and $\int_a^b f(x)\,dx$ is in unit of probability mass.

Unlike discrete r.v.s where we add probability masses via the pmf

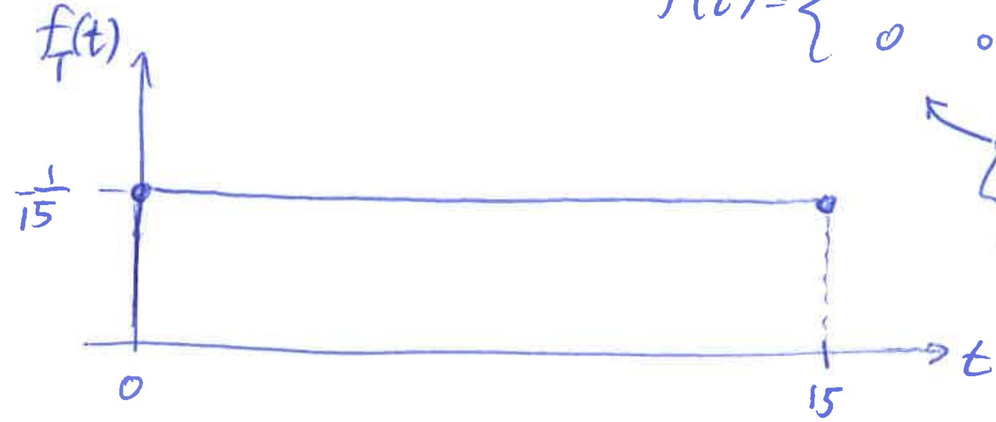
Here, for continuous r.v.s we Integrate probability density via the pdf.

Usually the pdf of a continuous r.v. is modeled by a specific situation.

For example, if we know a bus arrives at a bus stop every 15 minutes, then the time T we need to wait for this bus can be reasonably modeled by the uniform$[0,15]$ pdf:

$$f(t) = \begin{cases} \frac{1}{15} & \text{if } 0 \le t \le 15 \\ 0 & \text{otherwise} \end{cases}$$

a Piece-wise defined function



Notice that $f_T(t) \geq 0$ and $\int_{-\infty}^{\infty} f_T(t)\,dt = 1$.

since

$$\int_{\infty}^{-\infty} f_T(t)\,dt = \int_{-\infty}^{0} f_T(t)\,dt + \int_{0}^{15} f_T(t)\,dt + \int_{15}^{\infty} f_T(t)\,dt$$

$$= \int_{-\infty}^{0} 0\,dt + \int_{0}^{15} \frac{1}{15}\,dt + \int_{15}^{\infty} 0\,dt$$

$$= 0 + \frac{1}{15}\cdot 15 + 0 = 1.$$

Although I showed a lot of detail here, the point is that this is how we should work with piece-wise defined functions.

(Continued) From this uniform$[0,15]$ pdf.

What is the probability we must wait at least 10 minutes for the bus?

$$P(T \geq 10) = P(10 \leq T \leq 15)$$

$$= \int_{10}^{15} \frac{1}{15}\,dt = \frac{15-10}{15} = \frac{1}{3}.$$

Also,

$$P(5 \leq T \leq 8) = \int_{5}^{8} \frac{1}{15}\,dt = \frac{8-5}{15} = \frac{1}{5}.$$

Also,

(*) $$P(T=8) = P(8 \leq T \leq 8) = \int_{8}^{8} \frac{1}{15}\,dt = 0.$$

and we arrive at an interesting feature in working with continuous r.v.s — events that are possible can still get a probability of 0.

Contrast this with the situation of a discrete r.v.: If a discrete r.v. has prob. 0 of taking the value x, say, $P(X=x)=0$ then we know the event $(X=x)$ cannot happen.

Suppose X has the pdf

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0 \end{cases}$$

where $\lambda > 0$ is a fixed constant.

(1) Check that this is a pdf.

(2) Compute $P(X \geq 1)$. Compute $P(1 \leq X \leq 2)$.

Solution

(1) Certainly $f(x) \geq 0$ for all $x \in \mathbb{R}$ since $\lambda > 0$ and exponential functions are always positive (having real exponents) and

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_0^{\infty} f(x)\,dx = \int_0^{\infty} \lambda e^{-\lambda x}\,dx = -e^{-\lambda x}\Big]_{x=0}^{x=\infty} = 0 - (-e^{-\lambda 0})$$

$= 1$. So it is a pdf.

(2) $$P(X \geq 1) = \int_1^{\infty} \lambda e^{-\lambda x}\,dx = -e^{-\lambda x}\Big]_{x=1}^{x=\infty} = 0 - (-e^{-\lambda \cdot 1}) = e^{-\lambda}.$$

$$P(1 \leq X \leq 2) = \int_1^2 \lambda e^{-\lambda x}\,dx = -e^{-\lambda x}\Big]_{x=1}^{x=2} = e^{-\lambda} - e^{-2\lambda}.$$

$\boxtimes$