國立清華大學
NATIONAL TSING HUA UNIVERSITY

EE 306001
Probability

Lecture 14: continuous random variable

李祈均

Readings: Section 3.4 – 3.5

Lecture outline

Continuous random variables

- Multiple random variables
  - Conditioning
  - Independence
- Examples

Joint PDF $f_{X,Y}(x, y)$

$$P\big((X, Y \in S)\big) = \iint_S f_{X,Y}(x, y)dxdy$$

- $f_{X.Y}(x, y)$ is non-negative
- We can imagine *S* is a two-dimensional plane

- For any given subset, B, say a rectangular region:

$$P(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2) = \int_{x_1}^{x_2} \int_{y_1}^{y_2} f_{X,Y}(x, y)dxdy$$

- For the entire two-dimensional, S, plane:

$$P\big((X, Y \in S)\big) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y)dxdy = 1$$

- Intuitive interpretation:
  - Let δ be a small positive number, and consider the probability of a small rectangle

$$P(x_1 \leq X \leq x_1 + \delta, y_1 \leq Y \leq y_1 + \delta) = \int_{x_1}^{x_1+\delta} \int_{y_1}^{y_1+\delta} f_{X,Y}(x, y)dxdy$$

$$\approx f_{X,Y}(x_1, y_1) * \delta^2$$

  - Probability per unit area
  - Joint PDF contains all relevant probabilistic information between X, Y and their statistical dependencies
  - It can be used to calculate probability of event that are defined by both event

- Marginal probability
  - It can also be used to calculate event involving only one of the variable
  - Say, let A be a subset on the real line ($\mathrm{X} \in A$)

$$P(X \in A) = P\big(X \in A, Y \in (-\infty, \infty)\big) = \int_A \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dy\,dx$$

- From this we can see that:

Marginal PDF

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dy$$

$$f_Y(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dx$$

Independence

- The same goes for PDF then for PMF
- Two random variables are independent iff their joint PDF factors into product

$$f_{X,Y}(x, y) = f_X(x) * f_Y(y)$$

- Intuitively it is the same thing as PMF (knowing one variable does not matter in calculating the other one)

Joint CDF

- Having two random variables, we can also define their joint CDF:

$$F_{X,Y}(x,y) = P(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(x,y)dxdy$$

- The advantage of working with joint CDF is the same as for single random variable

$$f_{X,Y}(x,y) = \frac{d^2 F_{X,Y}}{dxdy}(x,y)$$

Trivial example

- If X, Y are both uniform random variable defined over the unit square, we can see that joint CDF is the following form:

$$F_{X,Y}(x, y) = xy, \text{for } 0 \ \leq x \leq 1, 0 \leq y \leq 1$$

To get the PDF from this joint CDF:

$$\frac{d^2 F_{X,Y}}{dxdy}(x, y) = 1$$

$$= f_{X,Y}(x, y), \text{for all (x,y) in the unit square}$$

Conditioning

Recall:

$$P(x \le X \le x + \delta) \approx f_X(x) * \delta$$

- Density gives us probabilities accumulation rate from little intervals

- In the conditional world, we would like to define conditional density as well:

$$P(x \le X \le x + \delta | Y \approx y) \approx f_{X|Y}(x|y) * \delta$$

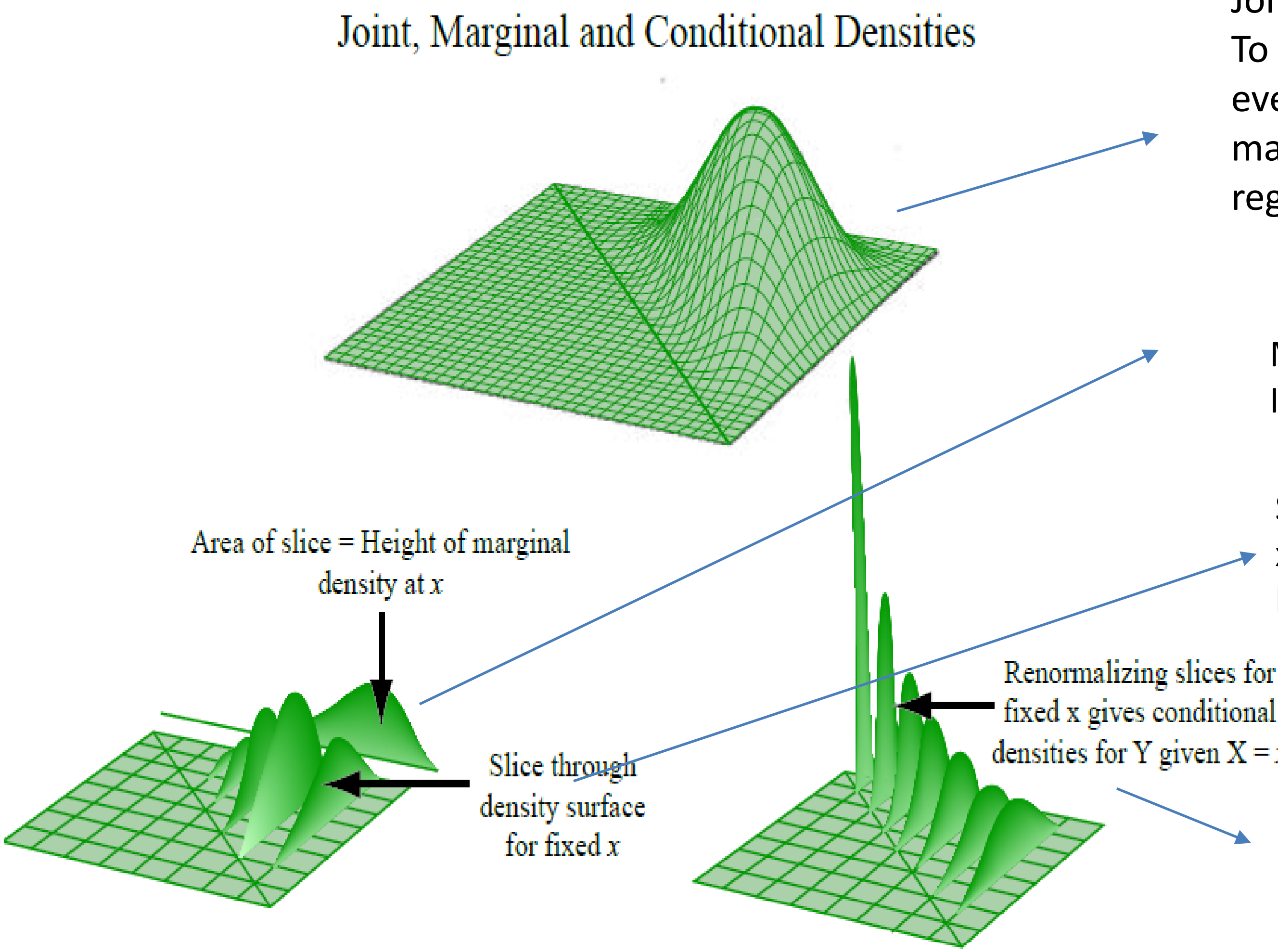
  - Why approximation?
  - Conditional probability is undefined when you condition on an event that has 0 probability
  - So instead of saying Y=y, we say Y is very close to y

- In practice, you may not care, however, to be rigorous, you should realize that it's "in the limit" as Y goes y, not Y equals to y

- This leads to the definition of conditional density:

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}, \text{if } f_Y(y) > 0$$

- Interpretation:
  - Say, I told you what Y is, given that, tell me what X's density looks like (reverse is true too)
  - In essence, you can imagine conditional PDF is just a 'slice' of a PDF, and normalize that slice



Joint PDF:
To calculate Probability of an event – look at how much mass is sitting on top of the region

Marginal PDF of x:
Integrate over all Y=y's

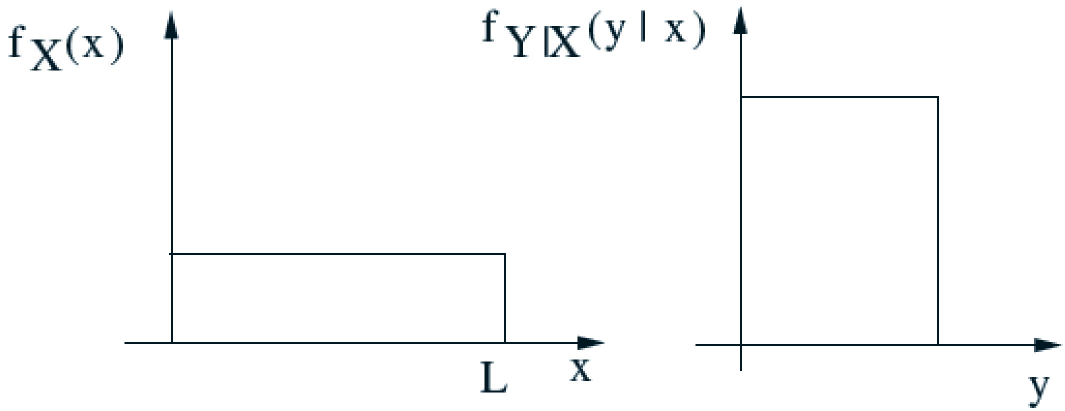
Slicing the joint for different x's and look at how y behaves for a fixed x

Conditional PDF (y given x):
Slicing -> renormalizing

Example: stick-breaking

- We have stick of length l
- Let's break it twice:
  - Break at X: uniform in $[0, l]$
  - Break again at Y: uniform in $[0, X]$

Note the upper case and lower case

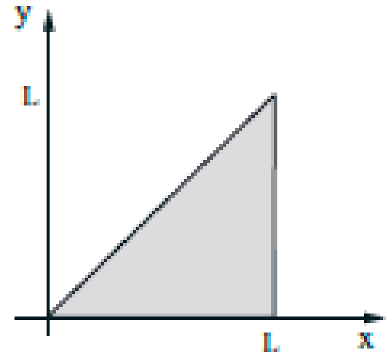


- What is the joint PDF between X, Y

$$f_{X,Y}(x,y) = f_X(x) f_{Y|X}(y|x)$$

- Essentially the multiplication that we know of

  - So what is ? $f_X(x) = 1/l$
  - What is? $f_{Y|X}(y|x) = \frac{1}{x}$ ← Note the lower case
  - The joint should be: $\frac{1}{l}\left(\frac{1}{x}\right)$
  - Over what range?
    - X can range anywhere from 0 to l
    - Y can only be smaller than x



- $f_{X,Y}(x,y) = \frac{1}{lx}, 0 \leq y \leq x \leq l$

- $E[Y|X=x] = \int y f_{Y|X}(y|X=x)dy$
  - $f_{Y|X}(y|X=x) = \frac{1}{\mathrm{x}}$
  - $\int_0^x y\left(\frac{1}{x}\right)dy = \frac{x}{2}$
  - It should be intuitively satisfying, it's just the expected value of Y given the new universe where X has been realized
  - Since Y is uniform, the expected value has to the midpoint x/2

- Marginal PDF of Y: after breaking it twice, how big is the little piece that I am left with

$$f_Y(y) = \int f_{X,Y}(x,y)dx = \int_y^l \frac{1}{lx} dx = \frac{1}{l}\log(\frac{l}{y}), 0 \le y \le l$$

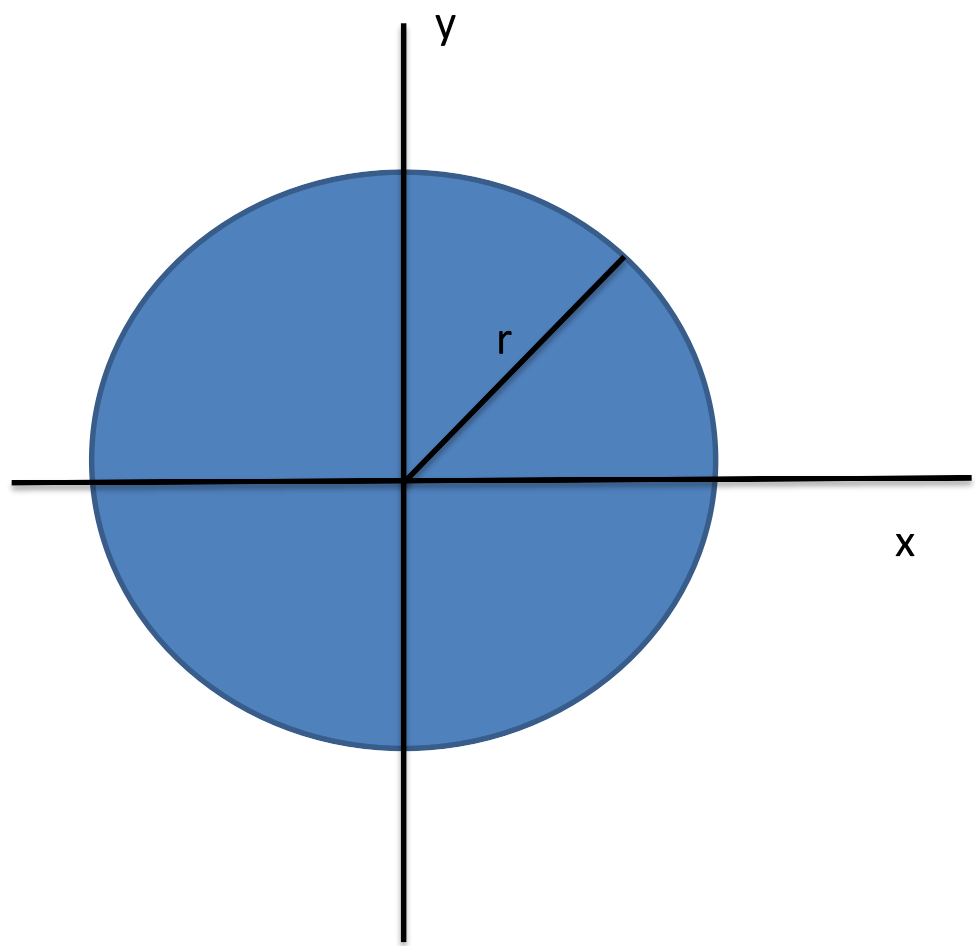
$$f_{X,Y}(x,y) = \frac{1}{lx}, 0 \le y \le x \le l$$

- Note, it is not enough to know the formula, you have to check the 'range' of integration, make sure you are integrating over the range where joint density does not equal to 0

- $E[Y] = \int_0^l y f_Y(y) dy = \int_0^l y(\frac{1}{l}) \log\left(\frac{l}{y}\right) dy$
  - Looks a bit like 'integration by part' problem? ylogy
  - Answer = $\frac{l}{4}$
  - Seems obvious at the moment, you break once, ½, break again another ½,
    - It turns out to be okay in this case, but not in general!

Example 2: circular uniform

You are throwing a dart at a circle with radius r, assuming you hit it within the circle every time, and all points of impact are equally-likely, so joint PDF of X, Y are uniform



- Joint PDF

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{1}{\text{area of the circle}}, \text{if (x,y) is in the circle} \\ 0, \text{otherwise} \end{cases}$$

$$= \begin{cases} \dfrac{1}{\pi r^2}, \text{if } x^2 + y^2 \leq r^2 \\ 0, \text{otherwise} \end{cases}$$

That's the area

- To calculate the conditional PDF $f_{X|Y}(x|y)$, we need to first find the marginal PDF $f_Y(y)$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx = \frac{1}{\pi r^2}\int_{x^2+y^2\leq r^2} dx$$

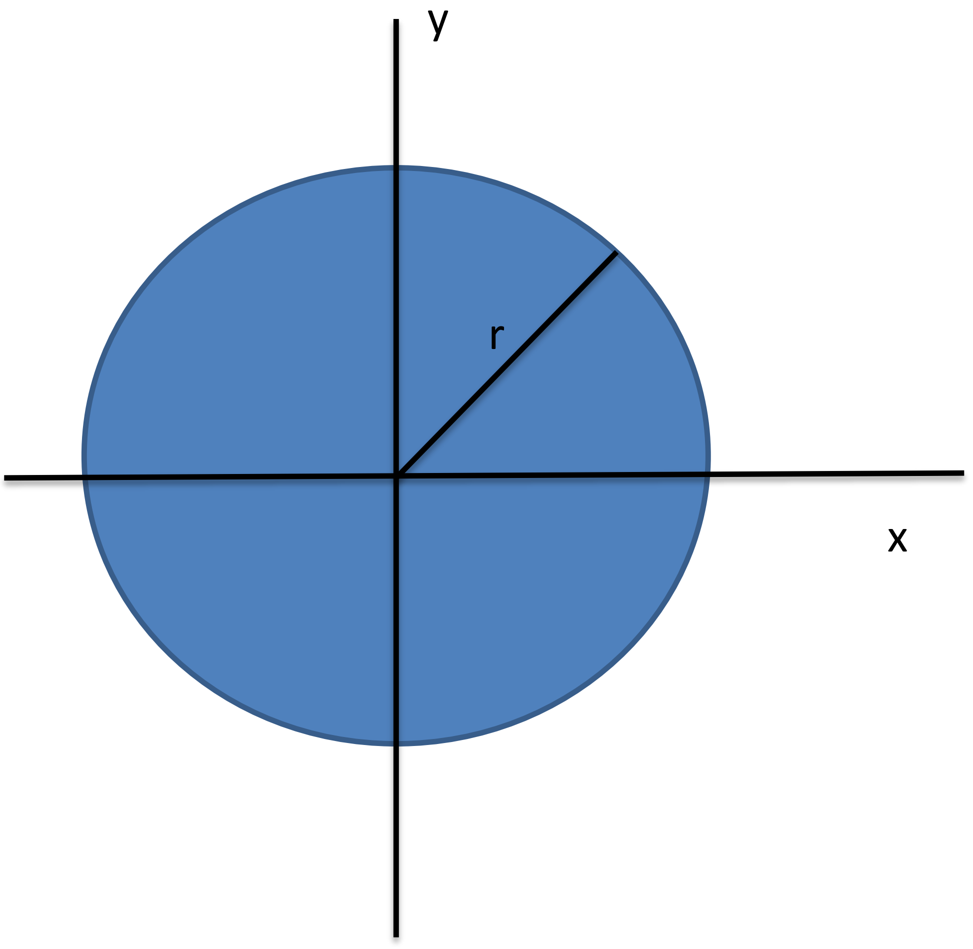
$$= \frac{1}{\pi r^2}\int_{-\sqrt{r^2-y^2}}^{\sqrt{r^2-y^2}} dx = \frac{2}{\pi r^2}\sqrt{r^2-y^2}, \text{if } |y| \leq r$$

*note that f_Y is not uniform

- The conditional PDF is hence,

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{1}{2\sqrt{r^2-y^2}}, \text{if} x^2+y^2 \leq r^2$$

* Given y, $f_{X|Y}$ is uniform



- $f_Y(y)$
  - Integrate out X
  - Not uniform

- $f_{X|Y}(x|y)$
  - Given Y, what is X
  - A slice - uniform

Example 3: exponential random variable is memoryless

- The time T until a new light bulb is burnt out is an exponential random variable with parameter, λ.

- You turn the light on, leave the room, and when you return, t, time after, you find the light bulb is still on; this is the event $A \in \{T > t\}$

- Let X be the additional time until the light bulb burns out. What is the conditional CDF of X, given the event A

- $P(X > x|A) = P(T > t + x|T > t)$

$$= \frac{P(T > t + x \text{ and } T > t)}{P(T > t)} = \frac{P(T > t + x)}{P(T > t)}$$

$$= \frac{e^{-\lambda(t+x)}}{e^{-\lambda t}} = e^{-\lambda x}$$

1- Cdf of exponential: CDF P(t<T)

This demonstrates the memoryless property of exponential, in general if we model a process of certain operation by an exponential, as long as the operation has not been completed, the remaining time up to completion has the same exponential CDF, no matter when the operation started

Example 4

An absent minded professor schedules two student appointment for the same time. The appointment duration are independent and exponentially-distributed with mean 30 minutes.

The first student arrives on time, but the second student arrives five minutes late. What is the expected value of the time between the arrival of the first student and the departure of the second student?

- There are two different ways that the total duration can happen
  - 1. first student stay no more than 5 minutes, second student does not need to wait (because he was late!)
  - 2. first student stays more than 5 minutes, second student gota wait in line

$E[\text{Total Time}]$
$= 5 + E[\text{stay of 2nd student}] * P(1st\text{ stays no more than 5 minutes}) + (E[\text{stay of 1st}|\text{stay of 1st} \geq 5] + E[\text{stay of 2nd}]) * P(1st\text{ stays more than 5 minutes})$

- $E[\text{stay of 2nd student}]=30$
- $E[\text{stay of 1st}|\text{stay of 1}st \geq 5]=5+E[\text{stay of 1}st]=5+30=35$
  - Memoryless property

P(T<t=5)

- $P(\text{1st stays no more than 5 minutes}) = 1 - e^{-\frac{5}{30}}$
- $P(\text{1st stays more than 5 minutes}) = e^{-5/30}$

Plug in and compute:

$$E[\text{Time}] = 35 * \left(1 - e^{-\frac{5}{30}}\right) + 65 * \left(e^{-\frac{5}{30}}\right)$$
$$= 60.394$$

Example 5

- Let X be an exponential random variable with parameter, λ>0. Calculate the probability that X belongs to one of the intervals $[n, n+1]$ with n odd

- Let's do the problem in 2 steps:
  - Compute the probability that X is in interval [n, n+1] for non-negative n
  - Add the probabilities for all odd positive integer values of n

- Let's use the CDF of exponential distribution

$$P(n \le X \le n+1) = F_X(n+1) - F_X(n)$$
$$= \left(1 - e^{-\lambda(n+1)}\right) - \left(1 - e^{-\lambda(n)}\right) = e^{-\lambda n}\left(1 - e^{-\lambda}\right)$$

- Now, onto step two, since the intervals are disjoint, we can simply sum this probability for all odd integers of n to find the probability of interest:

$$P(\{X \in [n, n+1] \text{ for some odd } n\})$$

$$= \sum_{n\ odd} e^{-\lambda n}\left(1 - e^{-\lambda}\right) = \left(1 - e^{-\lambda}\right)\sum_{k=0}^{\infty} e^{-\lambda(2k+1)}$$

$$= \left(1 - e^{-\lambda}\right)e^{-\lambda}\sum_{k=0}^{\infty} e^{-2\lambda k} \quad = \left(1 - e^{-\lambda}\right)e^{-\lambda}\frac{1}{1 - e^{-2\lambda}}$$

$$= \left(1 - e^{-\lambda}\right)e^{-\lambda}\frac{1}{(1 - e^{-\lambda})(1 + e^{-\lambda})} = \frac{e^{-\lambda}}{1 + e^{-\lambda}}$$

Example 6

The joint density of X and Y is given by

$$f_{X,Y}(x,y) = \begin{cases} xe^{-(x+y)}, x > 0, y > 0 \\ 0, \text{otherwise} \end{cases}$$

a) Are X and Y independent?

- Yes!
- Why, because "*joint factors*"

$$f_X(x) = xe^{-x}$$

$$f_Y(y) = e^{-y}$$

How about if:

$$f_{X,Y}(x,y) = \begin{cases} 2, 0 < x < y, 0 < y < 1 \\ 0, \text{otherwise} \end{cases}$$

Would X and Y be independent?

- No !
- Because joint does not equal to marginal * marginal

$$f_X(x) = \int_x^1 f_{X,Y}(x,y)dy = 2(1-x)$$

$$f_Y(y) = \int_0^y f_{X,Y}(x,y)dx = 2y$$

Example 5

Leon leaves his office everyday at a uniform random time between 430pm and 5pm. If he leaves *t* minutes past 430, the time it will take him to reach home is a uniform random number between 20 and 20+(2t)/3 minutes

Let Y be the number of minutes past 430 that Leon leaves his office tomorrow and X be the number of minutes it takes him to reach home.

Find the joint probability density function of X and Y?

- Look at the problem to see what has been defined in words, now let's turn it into equations

$$f_{X|Y}(x|y) = \begin{cases} \dfrac{1}{20 + [(2y)/3] - 20} \\ 0 \end{cases}$$

$$= \begin{cases} \dfrac{3}{2y}, 20 < x < 20 + \dfrac{2y}{3} \\ 0, \text{otherwise} \end{cases}$$

- From the problem statement,

We also know that:

$$f_Y(y) = \begin{cases} \dfrac{1}{30}, 0 < y < 30 \\ 0, \text{elsewhere} \end{cases}$$

Now we can easily find the joint density function

$$f_{X,Y}(x,y) = f_{X|Y}(x|y) f_Y(y)$$
$$= \begin{cases} \dfrac{1}{20y}, 20 < x < 20 + \dfrac{2y}{3}, 0 < y < 30 \\ 0, \text{elsewhere} \end{cases}$$