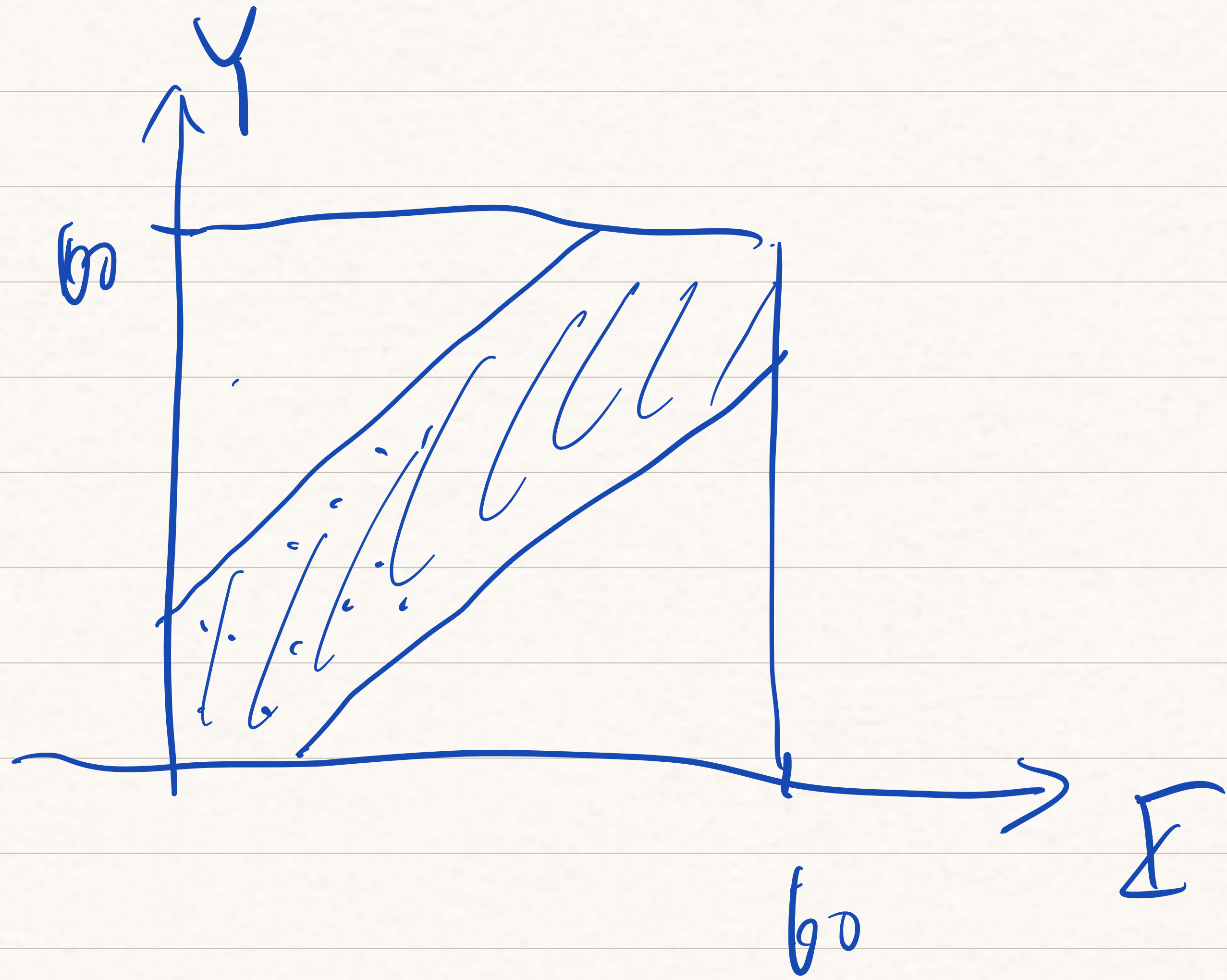
Continuous Random Variable III

Aug 4, 2022

- LLSE next Monday
- Conditional probability
- yesterday's R & J example



$$\boxed{\mathbb{P}(|X-Y| \le 15)}$$

$f_X(x)$ $f_Y(y)$ $f_{XY}(x,y)$

$Z = |X-Y|$ pdf ?

Joint distribution: Joint PDF

- A joint density function for two continuous random variables X, Y is a function $f: \mathbb{R}^2 \to \mathbb{R}$, such that
  - f is nonnegative, $f_{X,Y}(x, y) \geq 0, \forall x, y \in \mathbb{R}$
  - Total integral is 1, $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx\, dy = 1$

- The joint distribution of two continuous random variables X, Y is given by, $\forall a \leq b, c \leq d$

$$\mathbb{P}(a \leq X \leq b, c \leq Y \leq d) = \int_c^d \int_a^b f_{X,Y}(x, y) dx\, dy\ .$$

Normal random variable

(normal distribution, Gaussian distribution)

- A continuous random variable $\boldsymbol{X}$ is normal or Gaussian if the PDF is in the form

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\,\sigma^2}}$$

- $\mathbb{E}(X) = \mu, Var(X) = \sigma^2$ $= \mathbb{E}(X^2) - \mathbb{E}(X)^2$

- $\boldsymbol{X} \sim \mathcal{N}(\mu, \sigma^2)$

mean $\sigma = SD$

Useful integral $\int_{-\infty}^{\infty} e^{-x^{\wedge}2}\, dx = \sqrt{\pi}$

Normal random variable

(normal distribution, Gaussian distribution)

- A continuous random variable $\boldsymbol{X} \sim \mathcal{N}(\mu, \sigma^2), a, b \neq 0, \boldsymbol{Y} = a\boldsymbol{X} + b.$ Then $\boldsymbol{Y} \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$

- Further if $\boldsymbol{Y} = \frac{\boldsymbol{X} - \mu}{\sigma}$, then $\boldsymbol{Y} \sim \mathcal{N}(0,1)$

Sum of i.i.d. Normal

- Let $X \sim \mathcal{N}(0,1), Y \sim \mathcal{N}(0,1), X \perp Y$. Let $a, b \in \mathbb{R}$ be constant. Then $Z = aX + bY \sim \mathcal{N}(0, a^2 + b^2)$

- A general case, let $X \sim \mathcal{N}(\mu_1, \sigma_1^2), Y \sim \mathcal{N}(\mu_2, \sigma_2^2), X \perp Y$. Let $a, b \in \mathbb{R}$ be constant. Then $Z = aX + bY \sim \mathcal{N}(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$

CDF of standard normal

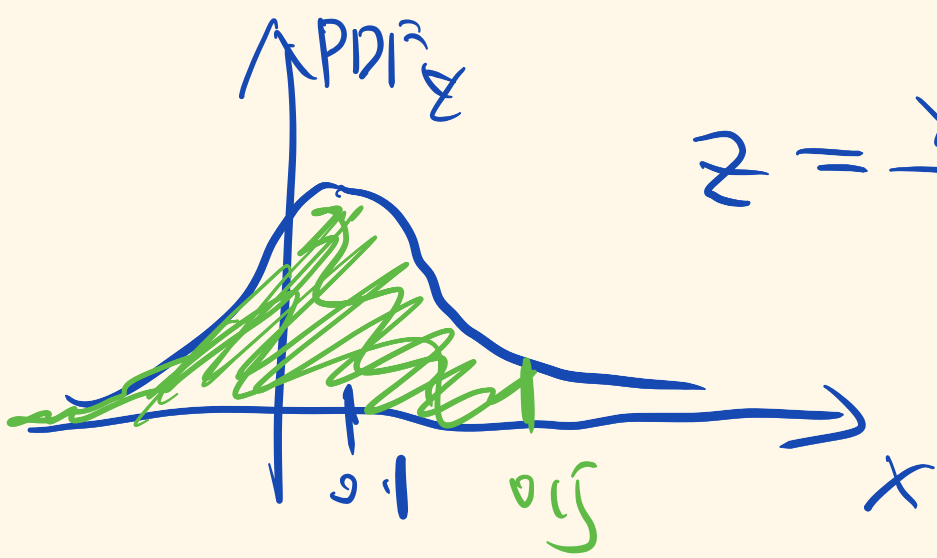
- CDF of $\mathcal{N}(0,1)$ standard normal is denote by Φ

$$\Phi(y) = \mathbb{P}(Y \le y) = \mathbb{P}(Y < y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{t^2}{2}} dt$$

$0.1 \quad 0.01$

- CDF for $\boldsymbol{X} \sim \mathcal{N}(\mu, \sigma^2)$ calculation
  1. standardize X by defining a new normal r.v. $Y = \frac{\boldsymbol{X}-\mu}{\sigma}$
  2. $\mathbb{P}(X \le x) = \mathbb{P}\left(\frac{\boldsymbol{X}-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right) = \mathbb{P}\left(Y \le \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)$

$$CDF_Z(0.5) = \Phi\left(\frac{x-\mu}{\sigma}\right) = \Phi(4)$$



$$z = \frac{x-\mu}{\sigma} = \frac{0.5-0.1}{0.1} = 4.$$

$$\Phi(4)$$

Sample mean: $M_n = \frac{S_n}{n}$

Central Limit Theorem (CLT)

$Var(\ ) = n \cdot \sigma^2$

Let $X_1, X_2, \dots, X_n$ be a sequence of iid random variables with $\mathbb{E}(X_i) = \mu$, $Var(X_i) = \sigma^2$

$S_n = \sum_{i=1}^{n} X_i$ $\quad \mathbb{E}(S_n) = n\mu$

mean 0

$Var(aX) = a^2 Var(X)$

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}$$

constant

constant

$\mathbb{E}(Z_n) = \frac{\mathbb{E}(S_n) - n\mu}{\sigma\sqrt{n}}$

$$\mathbb{E}(Z_n) = 0, Var(Z_n) = \frac{n\sigma^2}{\sigma^2 n} = 1$$

The CDF of Z_n converge to standard normal CDF

CDF $Z_n(z)$

$$\lim_{n\to\infty} \mathbb{P}(Z_n \le z) = \Phi(z), \forall z$$

$$\mathrm{Var}(Z_n) = \left(\frac{1}{\sigma\sqrt{n}}\right)^2 \mathrm{Var}(S_n - n\mu)$$

$$= \frac{1}{n\sigma^2} \mathrm{Var}(S_n)$$

$$= \frac{n\sigma^2}{n\sigma^2} = 1$$

Normal approximation based on CLT

Let $X_1, X_2, \ldots, X_n$ be a sequence of iid random variables with $\mathbb{E}(X_i) = \mu$, $Var(X_i) = \sigma^2$. If n is large, $\mathbb{P}(S_n \leq c)$ can be approximated by treating S_n as if it were normal:

1. Calculate the mean $n\mu$ and the variance $n\sigma^2$ of S_n
2. calculate the normalization value $z = \frac{c-n\mu}{\sigma\sqrt{n}}$ (z-score)
3. Use approximation $\mathbb{P}(S_n \leq c) \approx \Phi(z) =$

   where $\Phi(z)$ is available from standard normal CDF table.

$M_n = \frac{X_1 + \dots + X_n}{n}$ $X_i \sim \text{Bernoulli}(p)$

Example 1. Polling

We want to find out the value p representing the fraction of people supporting candidate A in a city. How many people we need to interview if we wish to estimate within accuracy of 0.01 with 95% probability.

Want find out what n makes

$$\mathbb{P}(|M_n - p| \geq 0.01) \leq 0.05$$

sample mean.

assume worst variance.

$$\mathbb{E}(M_n) = p, \quad \text{Var}(M_n) = \frac{1}{n^2} n \cdot \text{Var}(X_1) = \frac{p(1-p)}{n} \leq \frac{1}{4n}$$

the variance takes max value when $p = \frac{1}{2}$.

$$\mathbb{E}(M_n - p) = 0, \quad Var(M_n - p) = Var(M_n) \leq \frac{1}{4n}$$

$$z = \frac{0.01}{1/\sqrt{4n}} = 2n \cdot 0.01$$

$$\mathbb{P}(|M_n - p| \geq 0.01) \doteq 2\mathbb{P}(M_n - p \geq 0.01)$$
$$= 2 - 2\mathbb{P}(M_n - p \leq 0.01)$$
$$= 2 - 2\Phi(2n\, 0.01) \leq 0.05$$

$$\Phi(2n\, 0.01) \geq 0.975$$

$$\Phi(1.96) = 0.975 \qquad \text{so } n \geq \frac{1.96^2}{4 \cdot (0.01)^2} = 9604$$

Example 2.

$S_{100} = \sum_{1}^{100} X_i$

X_i

We load on a plane 100 packages, weight of each package is independent random variable follows uniform distribution between 5-50 kg. What is the probability that the total weight will exceed 3000 kg?

$$\mathbb{P}\left(\sum_{i=1}^{100} X_i > 3000\right) = \mathbb{P}(S_{100} > 3000) = 1 - \mathbb{P}(S_{100} \le 3000)$$

$$\mathbb{E}(X_i) = \frac{5+50}{2} = 27.5 = \mu. \quad \mathbb{E}(S_{100}) = n\mu = 2750$$

$$\mathrm{Var}(X_i) = \frac{(50-5)^2}{12} = 168.75 = \sigma^2 \quad \mathrm{var}(S_{100}) = n\sigma^2$$

$$SD(S_{100}) = \sqrt{n}\,\sigma$$

$$z = \frac{c - \mu n}{\sigma\sqrt{n}} = \frac{3000 - 27.5 \times 100}{\sqrt{168.75 \times 100}} = 1.92$$

$$\mathbb{P}(S_{100} > 3000) = 1 - \mathbb{P}(S_{100} \le 3000) \overset{CLT}{\approx} 1 - \Phi(1.92)$$

$$= 1 - 0.9726 = 0.0274.$$

Conditioning on an event

Conditional PDF of a continuous random variable X, given an event A with $\mathbb{P}(A) > 0$, is defined as a nonnegative function $f_{X|A}$ that satisfies

$$\mathbb{P}(X \in B|A) = \int_B f_{X|A}(x)dx$$

for any subset B of the real line.

$$\int_{-\infty}^{\infty} f_{X|A}(x)\,dx = 1.$$

Conditioning on an event

event

- If the event we condition on takes form of $\{X \in A\}$ and $\mathbb{P}(X \in A) > 0$ then

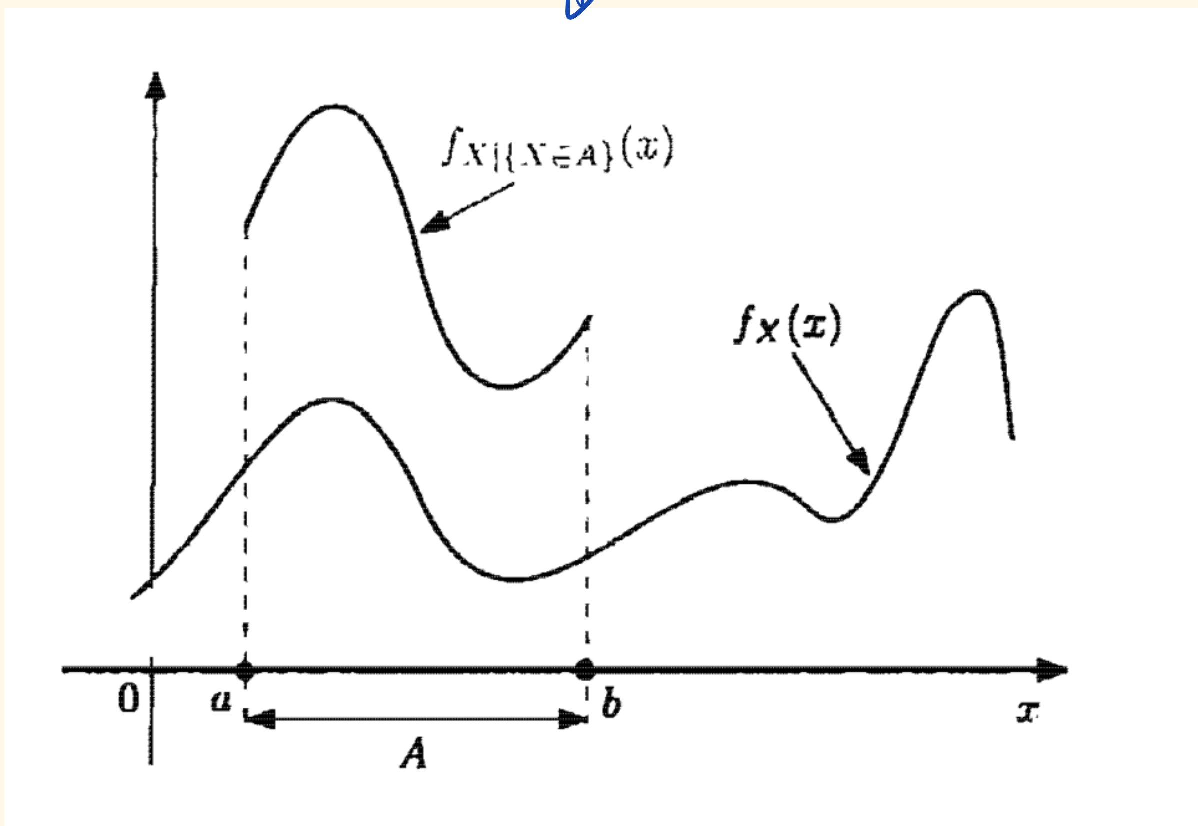
$$\mathbb{P}(X \in B | X \in A) = \frac{\mathbb{P}(X \in B, X \in A)}{\mathbb{P}(X \in A)} = \frac{\int_{A \cap B} f_X(x)dx}{\mathbb{P}(X \in A)}$$

- And $f_{X|\{X \in A\}}(x) = \begin{cases} \frac{f_X(x)}{\mathbb{P}(X \in A)} & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases}$

conditional PDF

integrate to 1.

conditional PDF



$T > t + x$ $\qquad T = t + X$ $\qquad \mathbb{P}(X > x) = e^{-\lambda x}$

Example 3. Exponential is memoryless

The time T till a new light bulb burns out is an exponential random variable with parameter λ. Alice turns the light on, leaves the room and when she returns, t time later, finds that the light bulb is still on, which correspond to event $A = \{T > t\}$. Let X be the additional time till the light bulb burns out. What is the conditional CDF of X. Given the event A?

$$\mathbb{P}(X \le x \mid A) = \mathbb{P}(X \le x \mid T > t) = 1 - \mathbb{P}(X > x \mid T > t)$$

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

$$= \frac{\mathbb{P}(T > t + x)}{\mathbb{P}(T > t)} = \frac{e^{-\lambda(t+x)}}{e^{-\lambda t}} = e^{-\lambda x}$$

$$\mathbb{P}(X \le x \mid A) = 1 - e^{-\lambda x} \qquad X \mid A \sim \exp(\lambda)$$

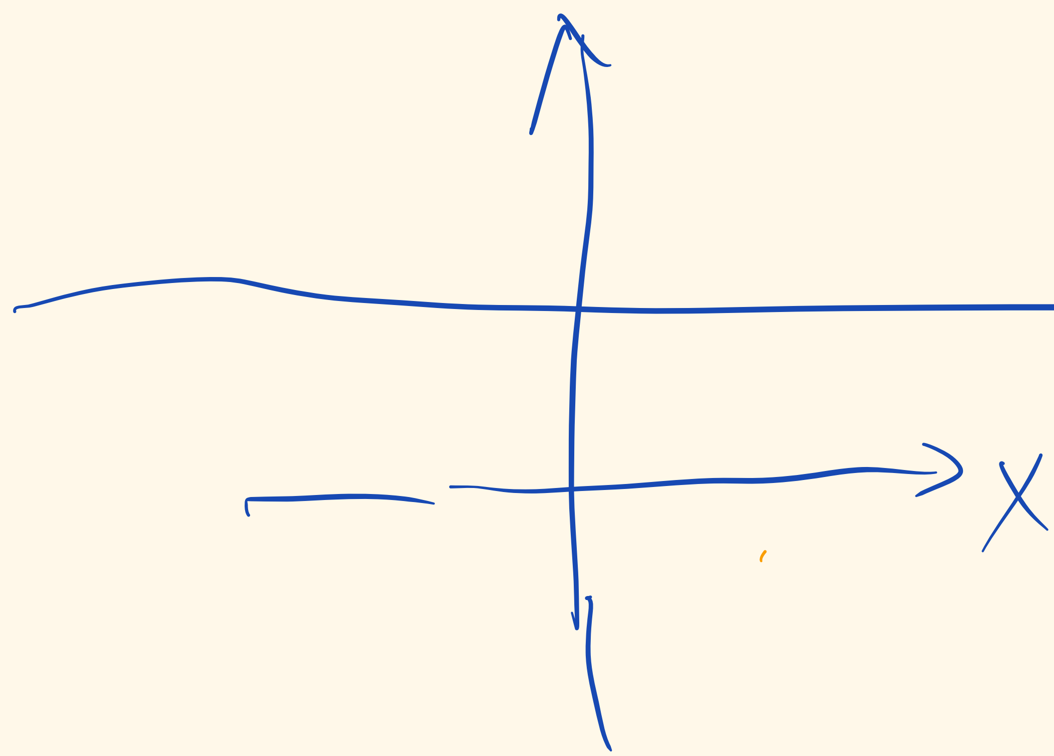
Conditioning on another random variable

- Two random variables X, Y with joint PDF $f_{X,Y}$. For any fixed y with $f_Y(y) > 0$ the conditional PDF of X given $Y = y$ is defined by

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

$\forall\, x, y.$

slice at $Y=y$



c·

Example 4.

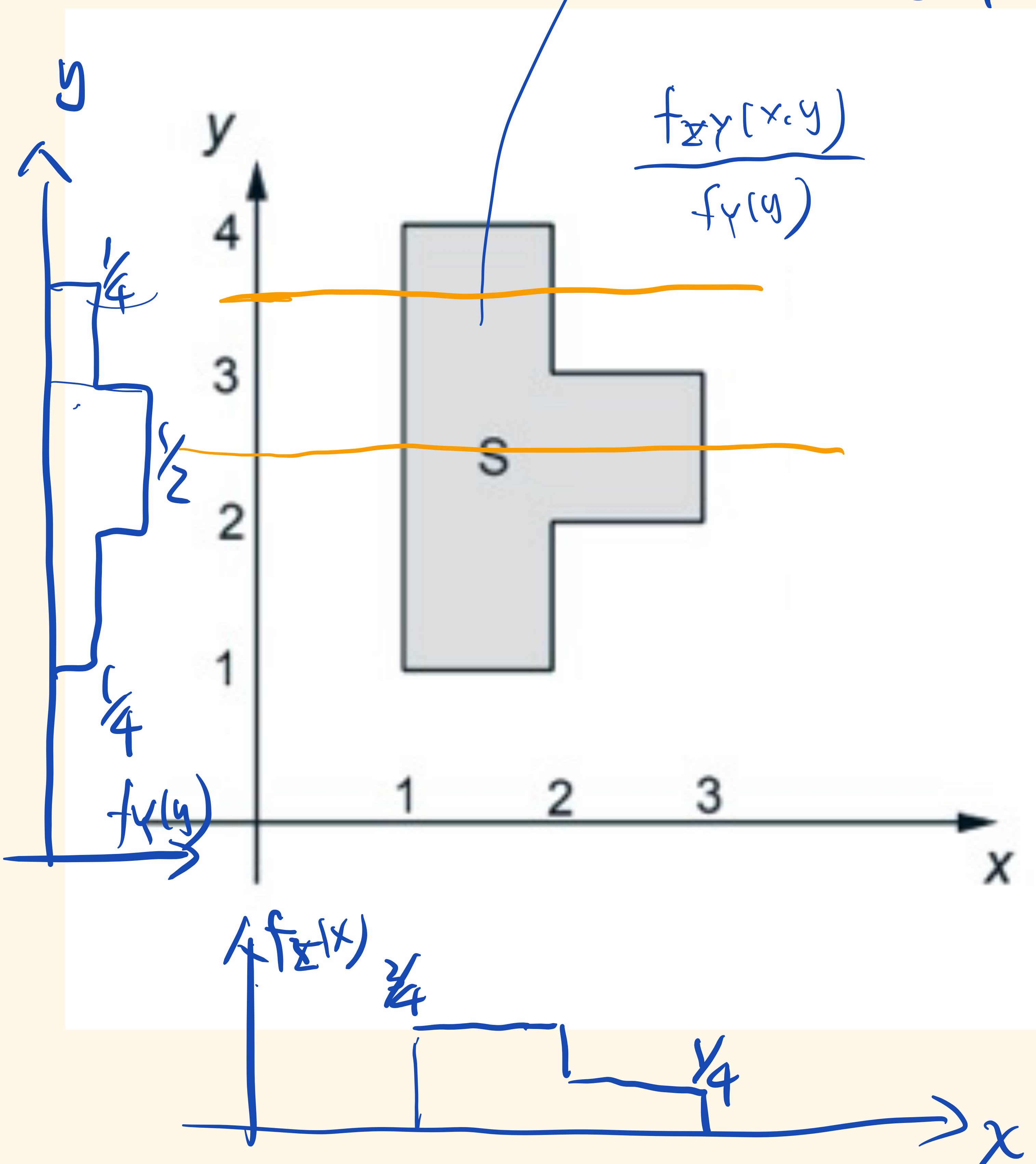
$$\iint_{-\infty}^{\infty} c\,dx\,dy = 1$$

$$= c \cdot \text{Area}(S)$$

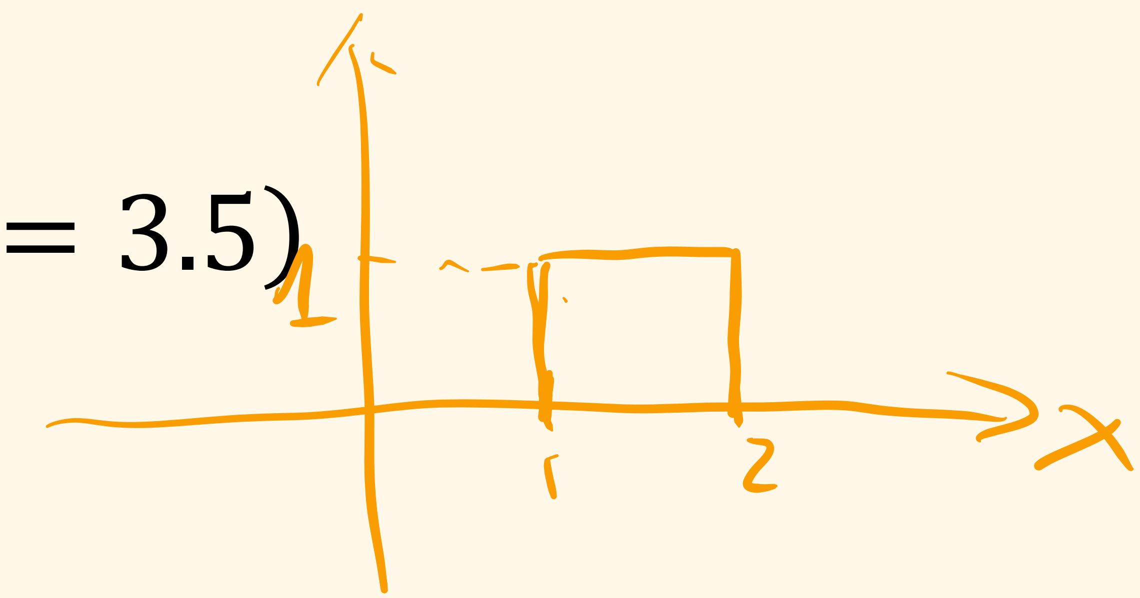
$$= c \cdot 4 = 1$$

$\frac{1}{4}$

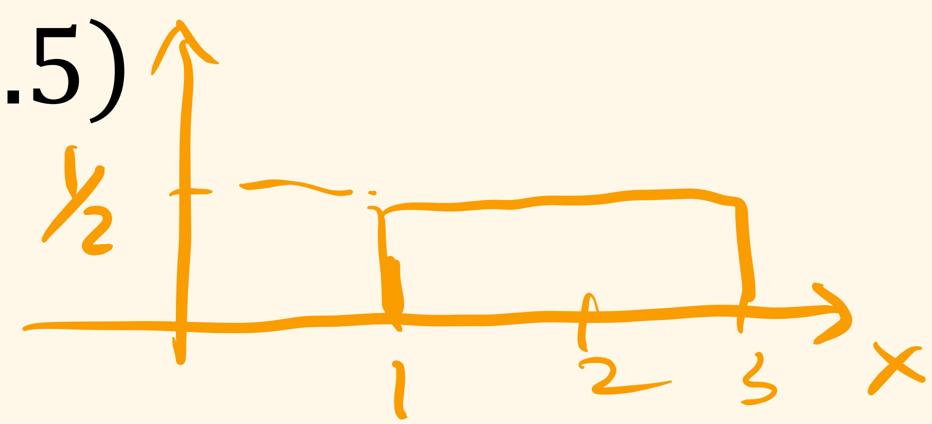
$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$



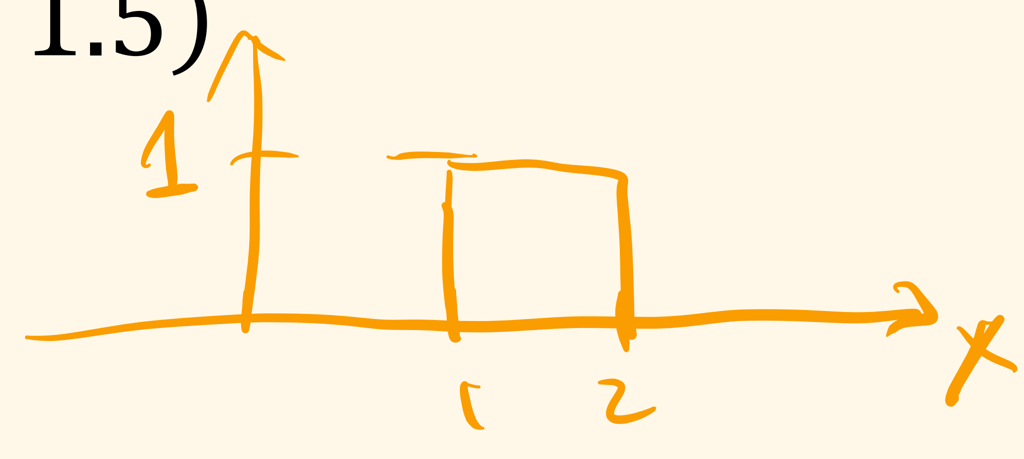
$f_{X|Y}(x|Y = 3.5)$



$f_{X|Y}(x|Y = 2.5)$



$f_{X|Y}(x|Y = 1.5)$



Conditioning on another random variable

- Two random variables X, Y with joint PDF $f_{X,Y}$. The joint, marginal and conditional PDFs are

$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}$

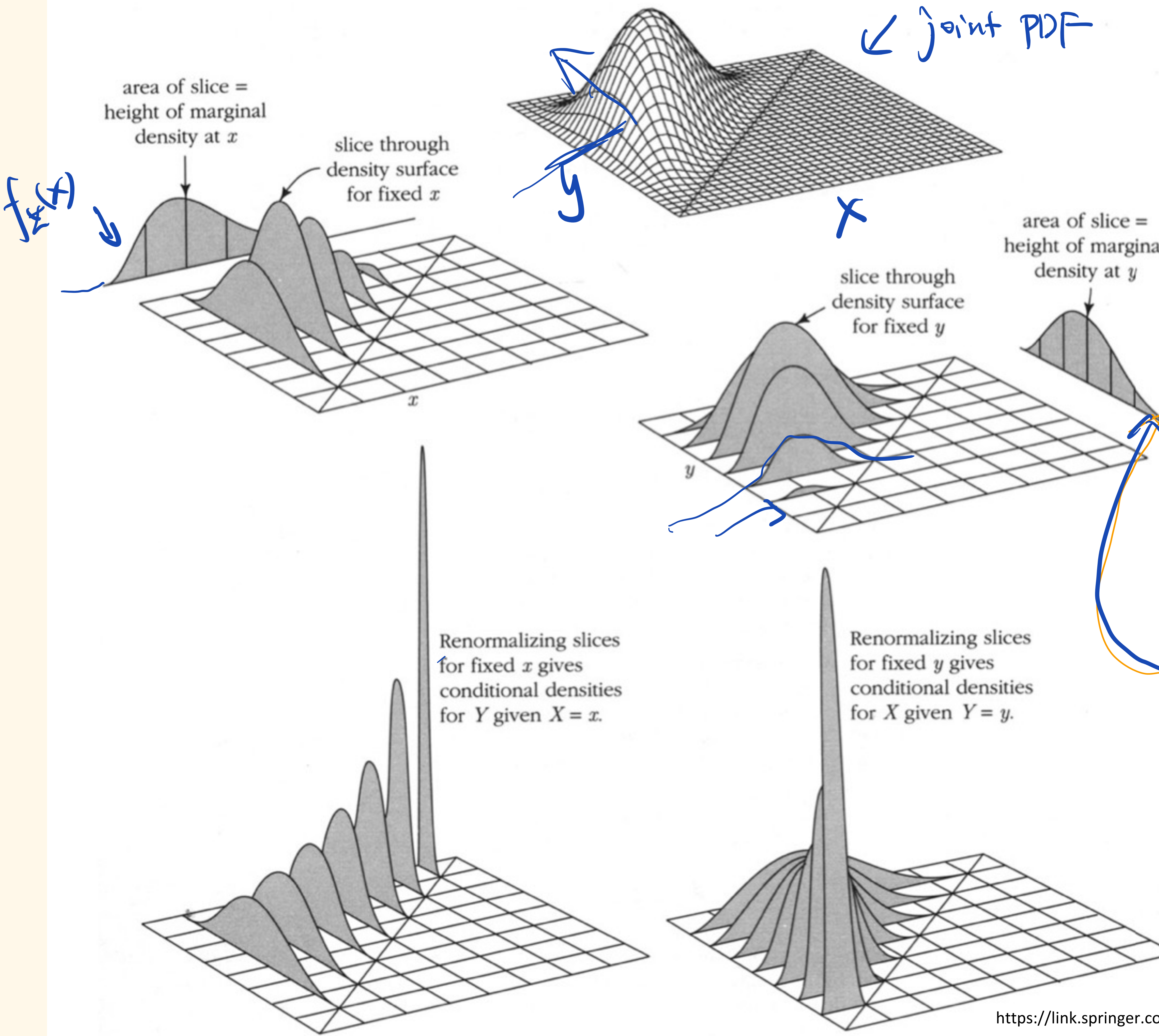
$$f_{X,Y}(x,y) = f_{X|Y}(x|y)f_Y(y)$$

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

The conditional PDF $f_{X|Y}(x|y)$ is defined for those y for which $f_Y(y) > 0$

- $f_{X|Y}(x|y)$ is a legit PDF, we can use it to calculate probability

$$\mathbb{P}(X \in \mathrm{A}|Y = y) = \int_A f_{X|Y}(x|y)dx$$



Marginals:

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

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

Conditional:

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

https://link.springer.com/content/pdf/10.1007%2F978-1-4612-4374-8_6.pdf

Conditional expectation

- Let X, Y be jointly continuous random variables, and let A be an event with $\mathbb{P}(A) > 0$

The conditional expectation of X given the event A is

$$\mathbb{E}(X|A) = \int_{-\infty}^{\infty} x f_{X|A}(x)dx$$

$$\mathbb{E}(g(X)|A) = \int_{-\infty}^{\infty} g(x) f_{X|A}(x)dx$$

The conditional expectation of X given $Y = y$ is

a function of y. ←

$$\mathbb{E}(X|Y = y) = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx$$

$$\mathbb{E}(g(X)|Y=y) = \int_{-\infty}^{\infty} g(x) f_{X|Y}(x|y)dx$$

Law of ~~iterative~~ expectation

iterated

Let $A_1, A_2, \dots, A_n$ be disjoint events that form a partition of the sample space, and assume that $\mathbb{P}(A_i) > 0, \forall i$,

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

Similarly

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} \mathbb{E}(X|Y = y) f_Y(y)\, dy$$

Overall,

$$\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X|Y))$$

$$\mathbb{E}\big(\underbrace{\mathbb{E}(X|Y)}_{\text{function of } Y}\big) = \int_{-\infty}^{\infty} \mathbb{E}(X|Y=y)\, f_Y(y)\, dy$$

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

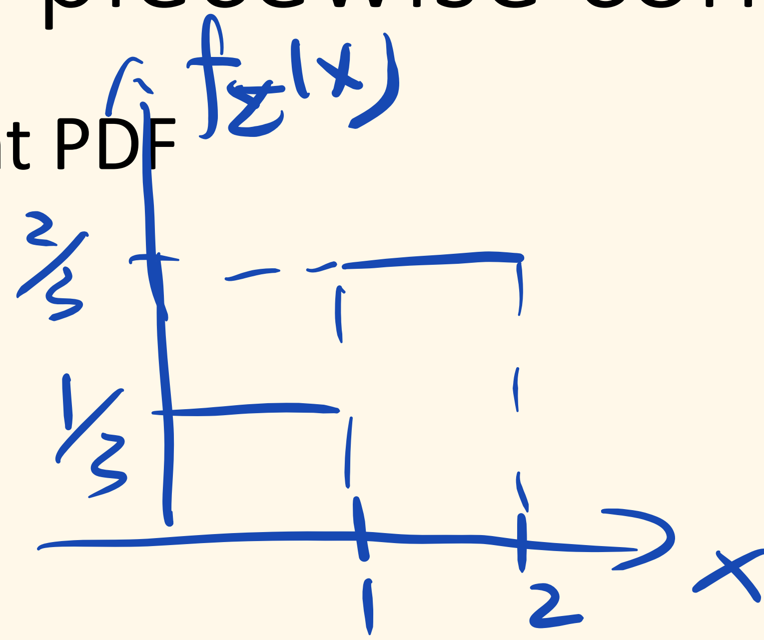
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x \, \underline{f_{XY}(x,y)}\, dx\, dy$$

$$= \int_{-\infty}^{\infty} x\, f_X(x)\, dx = \mathbb{E}(X)$$

Example 5. Mean and variance of a piecewise constant PDF

Consider a random variable X has piecewise constant PDF

- $f_X(x) = \begin{cases} \frac{1}{3} & \text{if } 0 \le x \le 1 \\ \frac{2}{3} & \text{if } 1 < x \le 2 \\ 0 & \text{otherwise} \end{cases}$



consider events

- $A_1 = \{X \text{ lies in the first interval } [0,1]\}$
- $A_2 = \{X \text{ lies in the second interval } (1,2]\}$

$$\mathbb{P}(A_1) = \int_0^1 f_X(x)\,dx = \frac{1}{3}. \qquad \mathbb{P}(A_2) = \int_1^2 f_X(x)\,dx = \frac{2}{3}$$

$$f_{X|A_1}(x|A_1) = \begin{cases} \frac{f_X(x)}{\mathbb{P}(A_1)} & x \in [0,1) \\ 0 & \text{o.w.} \end{cases} = \begin{cases} 1 & x \in [0,1] \\ 0 & \text{o.w.} \end{cases}$$

$$X|A_1 \sim \underline{U[0,1]}$$

$$f_{X|A_2}(x|A_2) = \begin{cases} 1 & x \in (1,2] \\ 0 & \text{o.w.} \end{cases}$$

$$Z \sim U(a,b) \quad \mathbb{E}(Z) = \frac{b+a}{2}$$

$$\mathbb{E}(Z^2) = \frac{a^2+ab+b^2}{3}$$

$$\mathrm{Var}(Z) = \frac{b^2-a^2}{12}$$

$$\mathbb{E}(X|A_1) = \frac{1}{2} \qquad \mathbb{E}(X|A_2) = \frac{3}{2}$$

$$\mathbb{E}(X^2|A_1) = \frac{1}{3} \qquad \mathbb{E}(X^2|A_2) = \frac{7}{3}$$

$$\mathbb{E}(X) = \mathbb{P}(A_1)\mathbb{E}(X|A_1) + \mathbb{P}(A_2)\mathbb{E}(X|A_2) = \frac{1}{3}\cdot\frac{1}{2} + \frac{2}{3}\cdot\frac{3}{2} = \frac{7}{6}$$

$$\mathbb{E}(X^2) = \mathbb{P}(A_1)\mathbb{E}(X^2|A_1) + \mathbb{P}(A_2)\mathbb{E}(X^2|A_2) = \frac{1}{3}\frac{1}{3} + \frac{2}{3}\cdot\frac{7}{3} = \frac{15}{9}$$

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \frac{15}{\ldots} - \frac{49}{\ldots} \quad 11$$

$$\mathrm{var}(X) = E(X^2) - (E(X))^2 = \frac{\text{[illegible]}}{9} - \frac{\text{[illegible]}}{36} = \frac{\text{[illegible]}}{36}$$

for piecewise constant PDF, it is easier to find mean and variance using this way, by partitioning intervals, and each conditional PDF will be uniform.