Continuous Random Variable II

Aug 3, 2022

Probability density function (PDF)

- A random variable if called continuous if there is a nonnegative function $f_{\boldsymbol{X}}$ called <u>probability density function</u> (PDF) of $\boldsymbol{X}$ such that

$$\mathbb{P}(\boldsymbol{X} \in B) = \int_B f_{\boldsymbol{X}}(x)dx \quad \text{for every subset } B \subset \mathbb{R}.$$

- The probability that the value of $\boldsymbol{X}$ falls with in an interval is

$$\mathbb{P}(a \leq \boldsymbol{X} \leq b) = \int_b^a f_{\boldsymbol{X}}(x)dx$$

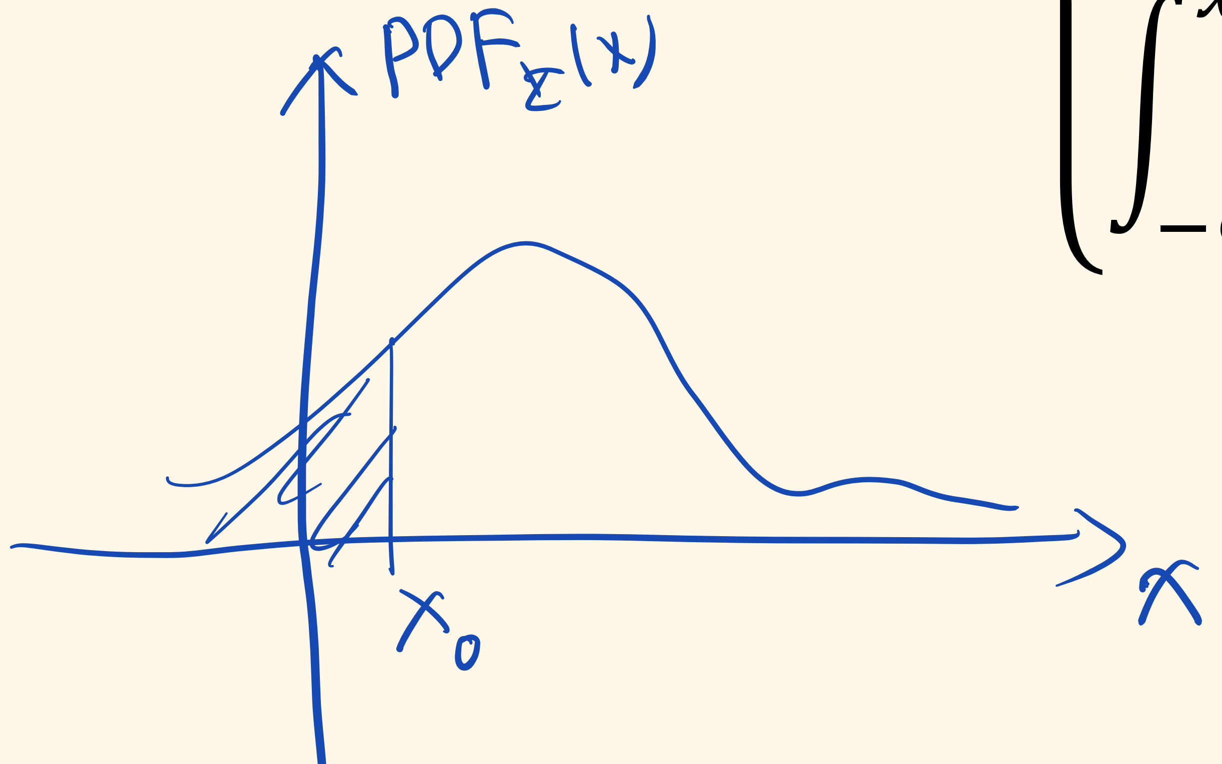
$$\mathbb{P}(X \in (x, x+\delta)) = \int_x^{x+\delta} f_X(t)\,dt = f_X(x)\,\delta$$

Cumulative density function (CDF)

The CDF of a random variable $\boldsymbol{X}$ with PDF $f_{\boldsymbol{X}}$ (or PMF $p_{\boldsymbol{X}}$) is denoted as $F_{\boldsymbol{X}}$

$\forall x,$

$$F_{\boldsymbol{X}}(x) = \mathbb{P}(\boldsymbol{X} \le x) = \begin{cases} \displaystyle\sum_{k \le x} p_{\boldsymbol{X}}(k) & \text{if } \boldsymbol{X} \text{ is discrete} \\ \displaystyle\int_{-\infty}^{x} f_{\boldsymbol{X}}(t)\,dt & \text{if } \boldsymbol{X} \text{ is continuous} \end{cases}$$



$$CDF_X(x_0) = \int_{-\infty}^{x_0} PDF(t)\,dt$$

$$\int e^{-\lambda t}\,dt = -\frac{1}{\lambda} e^{-\lambda t}$$

Geometric and exponential CDFs

Exponential PDF

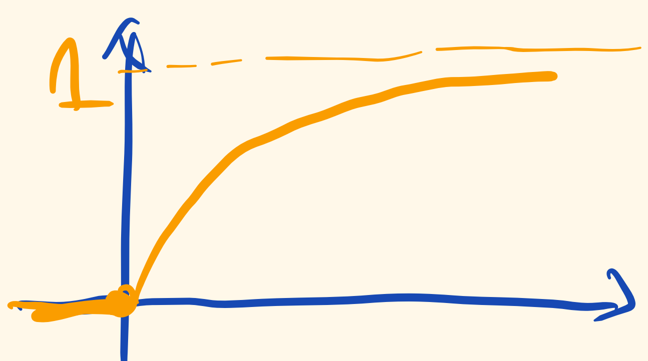
$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Exponential CDF

$$F_X(x) = \mathbb{P}(X \le x) = \int_{-\infty}^{x} f_X(t)\,dt$$

$$= \int_0^x \lambda e^{-\lambda t}\,dt$$

$$= -e^{-\lambda t}\Big|_0^x = \begin{cases} 0 & x \le 0 \\ 1 - e^{-\lambda x} & x > 0 \end{cases}$$



Geometric PMF

$$p_X(k) = (1-p)^{k-1} p$$

Geometric CDF

$$F_X(n) = \mathbb{P}(X \le n) = \sum_{k=1}^{n} p(1-p)^{k-1}$$

$$= p\,\frac{1-(1-p)^n}{1-(1-p)} = 1-(1-p)^n \qquad n = 1, 2, \ldots$$

Geometric and exponential CDFs

$x > 0$

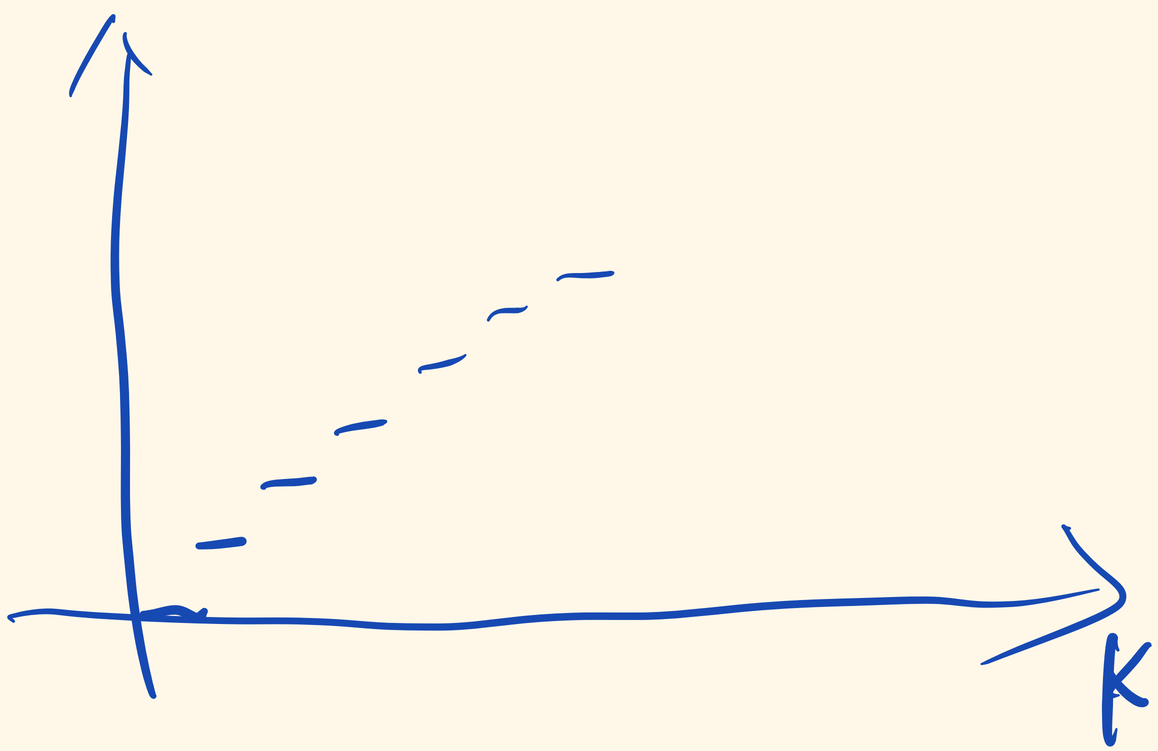
$n = 1, 2, 3, \ldots$

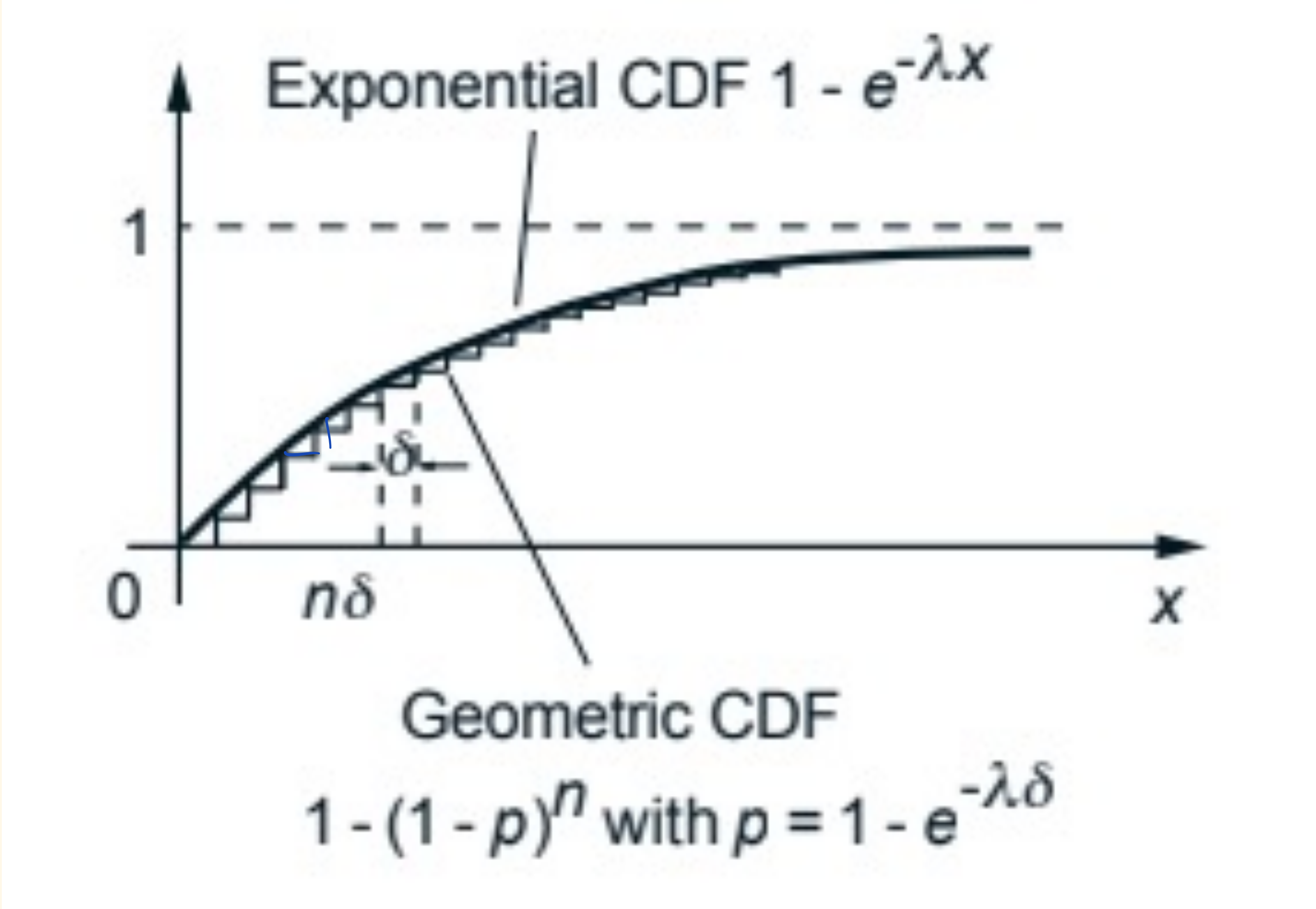
$F_{exp}(x) = 1 - e^{-\lambda x}$, $F_{Geo}(n) = 1 - (1-p)^n$

let $e^{-\lambda\delta} = 1-p$, $\delta = -\log(1-p)/\lambda$

$x = n\delta$ $F_{exp}(n\delta) = F_{geo}(n)$

$e^{-\lambda x} = (1-p)^n$ $x = n\delta$, δ s.t.





Joint distribution: Joint PDF

(x, y)

- 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, $\forall a \leq b, c \leq d$

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

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f_X(x) dx$$

Joint distribution: Marginals

discrete

$$p_X(x) = \sum_y p_{XY}(x,y)$$

$$= \sum_y \mathbb{P}(X, Y=y)$$

- The marginal PDF f_X of $\boldsymbol{X}$ is given by

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

- Similarly

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

Joint distribution: Joint CDFs

- If X, Y are two random variables associated with the same experiment, we define their joint CDF by
$$F_{X,Y}(x, y) = \mathbb{P}(X \leq x, Y \leq y)$$

- The joint PDF of two continuous random variables X, Y is $f_{X,Y}$, then
$$F_{X,Y}(x, y) = \mathbb{P}(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(x, y) dx\, dy\,.$$

Independence

Discrete $\forall\, a,b$.

$\mathbb{P}(X=a, Y=b) = \mathbb{P}(X=a)\,\mathbb{P}(Y=b)$

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

- Two random variables X, Y are independent if the event $a \leq X \leq b$ and $c \leq Y \leq d$ are independent for all $a \leq b, c \leq d$.

$$\mathbb{P}(a \leq X \leq b, c \leq Y \leq d) = \mathbb{P}(a \leq X \leq b)\mathbb{P}(a \leq X \leq b)$$

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

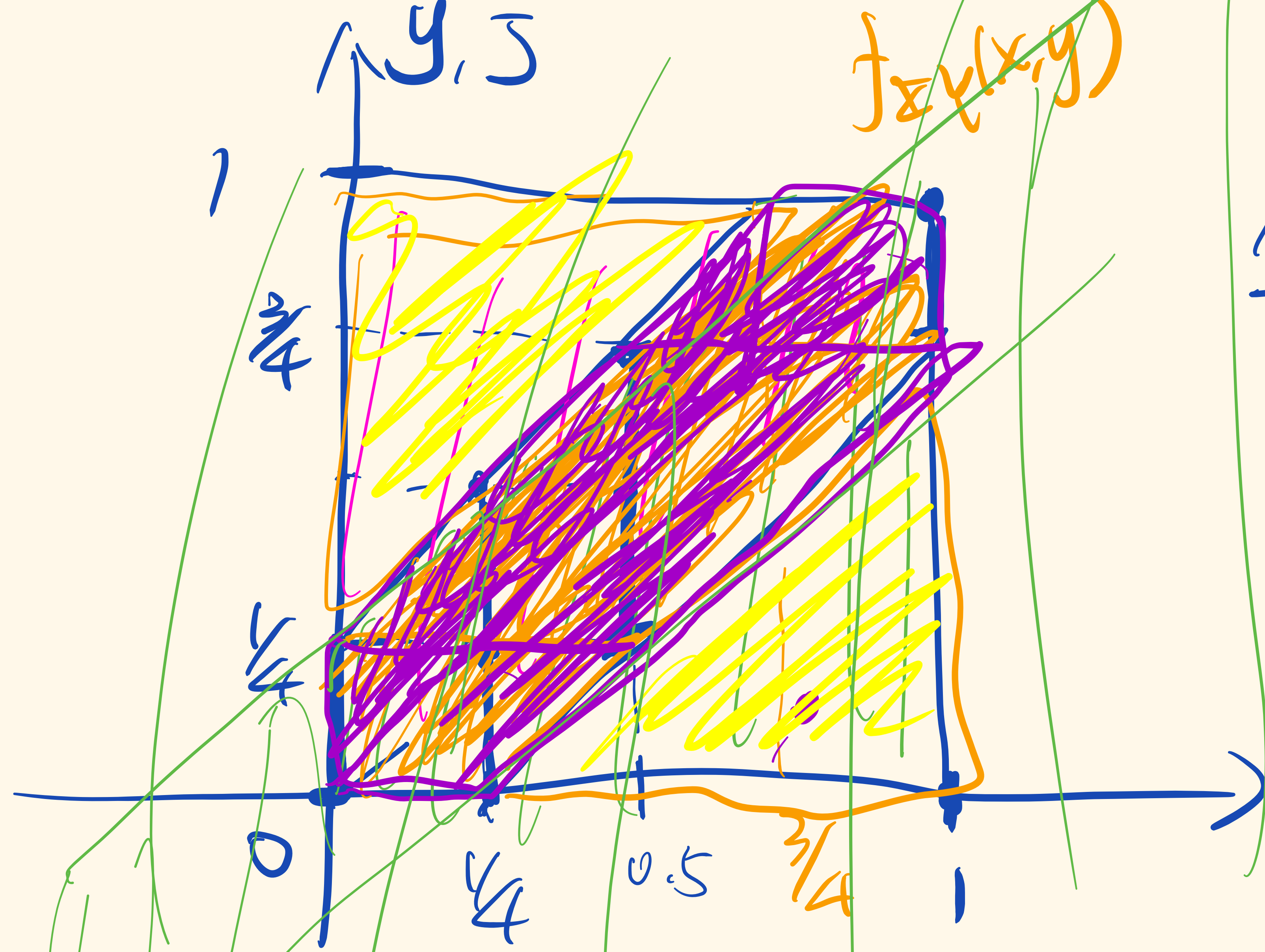
- The joint density of independent random variables X, Y is the product of the marginal densities

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

Example 1. 2D uniform PDF

$\mathbb{P}\left(X \le Y+15 \wedge X \ge Y-15\right)$ $\quad \mathbb{P}(Y-15 \le X \le Y+15)$

Romeo and Juliet have a date at a given time and each will arrive at the meeting place with a delay between 0 and 1 hour. Let X, Y denote the delays of R and J respectively. All pairs of delay (x, y) are equally likely. The first ~~two~~ one arrive will wait 15 min and leave if the other hasn't arrived. What's the probability that they meet.



$$f_{XY}(x,y) = 1$$

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

$$= \int_0^1\int_0^1 c\,dx\,dy = c$$

$$\mathbb{P}(\text{meet}) = \frac{}{1}$$

$$= 7/16.$$

$$1 - \left(\frac{3}{4}\right)^2$$

Example 2.

$$f_X(x) = \begin{cases} 3/4 & 1 \le x \le 2 \\ 1/4 & 2 < x \le 3 \\ 0 & \text{otherwise} \end{cases}$$

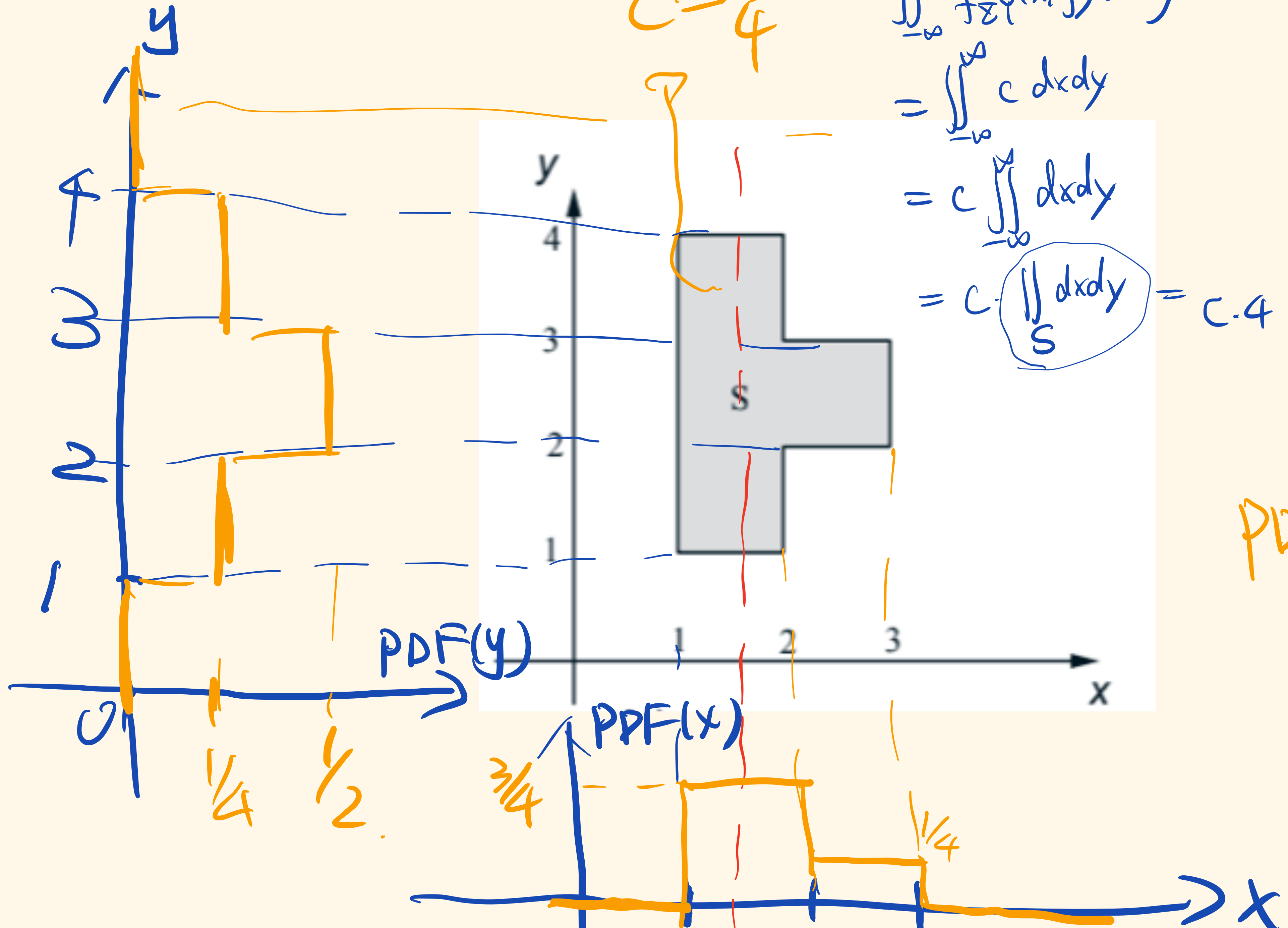
- The joint PDF of random variable *X* and *Y* is a constant *c* on the set *S* in figure, and *0* outside, Find the value of *c* and the marginal PDFs of *X* and *Y*

$c = \frac{1}{4}$

$$\iint_{-\infty}^{\infty} f_{XY}(x,y)\,dxdy = \iint_{-\infty}^{\infty} c\,dxdy = c\iint_{-\infty}^{\infty} dxdy = c\iint_S dxdy = c\cdot 4 = 1 \quad \text{so } c = \frac{1}{4}$$

$f_{XY}(x,y) = c$

$$\iint_{-\infty}^{\infty} = \iint_S + \iint_{\text{not } S}$$



$$\text{PDF}(x) = \int f_{XY}(x,y)\,dy$$

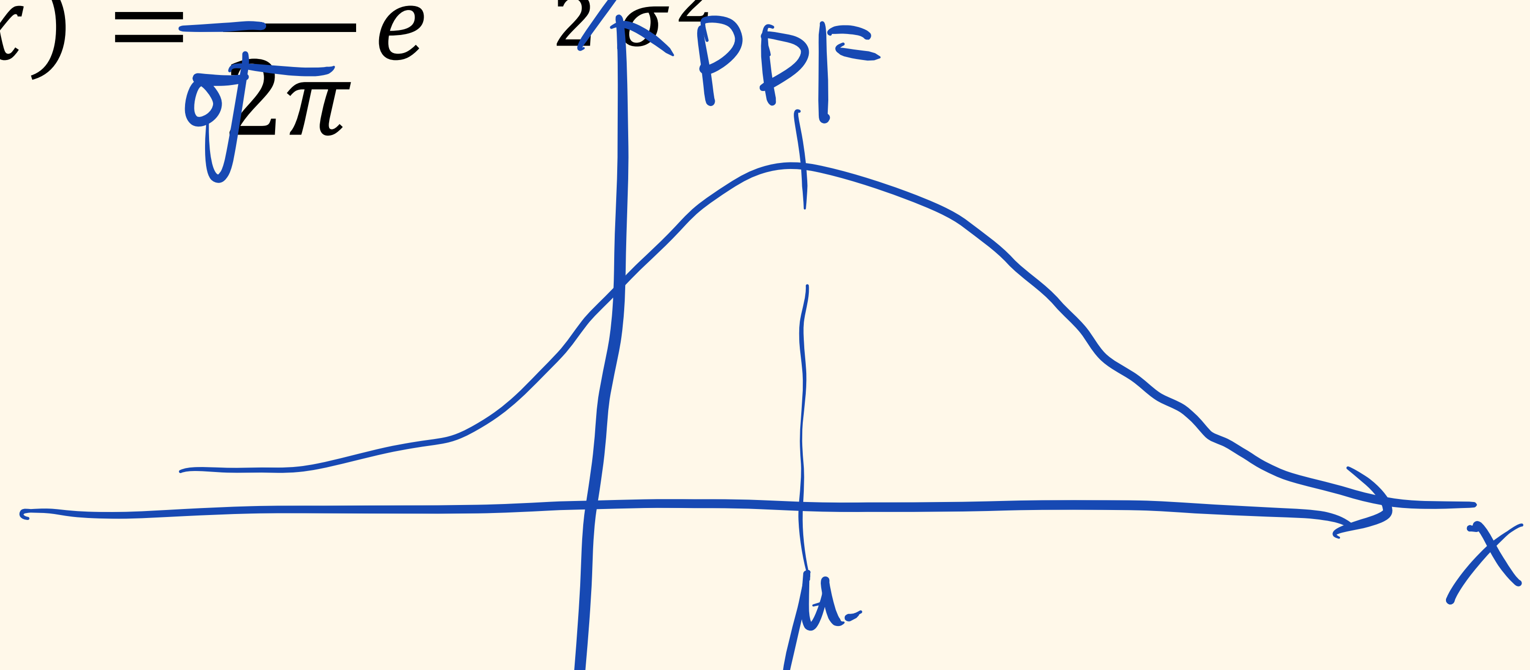
1 2 3

Normal random variable

(normal distribution, Gaussian distribution)

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

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



$$\int_{-\infty}^{\infty} f_X(x)\,dx = 1$$

$$\mathbb{E}(X) = \mu$$

$$\mathrm{var}(X) = \sigma^2$$

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

μ

Normal random variable

(normal distribution, Gaussian distribution)

$X \sim \mathcal{N}(0,1)$

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

$\mathbb{E}(Y) = \mathbb{E}(aX) + b = a\,\mathbb{E}(X) + b = a\mu + b$

$\mathrm{var}(Y) = \mathrm{var}(aX + b) = a^2\,\mathrm{var}(X) = a^2\sigma^2$

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

$X \sim (\mu, \sigma^2)$

$\mathcal{N}(0,1)$ - standard normal

CDF of standard normal

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

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

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

$0.1 \quad (0.1)^2$

1. standardize X by defining a new normal r.v. $\boxed{Y = \frac{X-\mu}{\sigma}} \sim N(0,1)$
2. $\mathbb{P}(X \le x)$ $= \mathbb{P}\left(\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right)$

0.5

Y

$$= \mathbb{P}\left(Y \le \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

z-score

$$\frac{0.5-0.1}{0.1} = 4$$

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 $X \sim N(\mu_1, \sigma_1^2)$ $Y \sim N(\mu_2, \sigma_2^2)$

$Z = aX + bY \sim N(\mu_1 + \mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$

Central Limit Theorem (CLT)

Let $X_1, X_2, \dots, X_n$ be a sequence of iid random variables with

$\mathbb{E}(X_i) = \mu,\ Var(X_i) = \sigma^2$ $\qquad \mathbb{E}\left(\frac{S_n}{n}\right) \to \mu$ as $n \to \infty$

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

$\mathcal{N}(0,1)$ $\quad \mathbb{E}(Z_n) = 0, Var(Z_n) = \dfrac{n\sigma^2}{\sigma^2 n} = 1$

The CDF of Z_n converge to standard normal CDF

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

Normal approximation based on CLT

Let $X_1, X_2, \dots, X_n$ be a sequence of iid random variables with
CDF
$\mathbb{E}(X_i) = \mu,\ Var(X_i) = \sigma^2$. 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 $\mathbb{P}(S_n \leq c) \approx \Phi(z)$

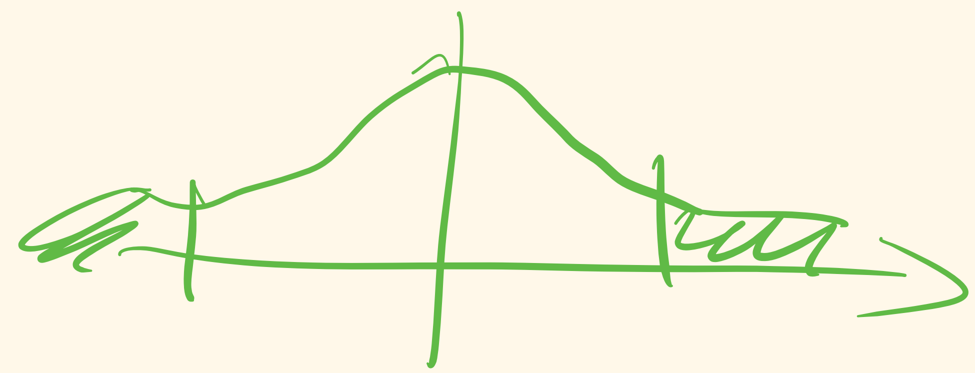
   where $\Phi(z)$ is available from standard normal CDF table.

Example 3. Polling

We want to find out the value p representing the fraction of people supporting candidate A in a city.

$$\mathbb{P}(|M_n - p| \geq \varepsilon) \leq \delta$$

$\varepsilon \downarrow 0.01 \qquad \delta \downarrow 0.05$



$n \geq 50{,}000$

$$\mathbb{P}(|M_n - p| \geq \varepsilon) \doteq 2\,\mathbb{P}(M_n - p \geq \varepsilon) \leq \delta$$

$$\mathbb{P}(M_n - p \geq \varepsilon) \leq 1 - \Phi(z) = 1 - \Phi(2\varepsilon\sqrt{n})$$

$$2 - 2\Phi(2 \cdot 0.01 \cdot \sqrt{n}) \leq 0.05, \quad \Phi(2 \cdot 0.01 \cdot \sqrt{n}) \leq 0.975$$

table.

$$\Phi(1.96) = 0.975$$

Example 3. Polling

How many people we need to interview if we wish to estimate within accuracy of 0.01 with 95% probability.

$$2 \cdot 0.01\sqrt{n} = 1.96$$

$$\Rightarrow n \geq 9604$$

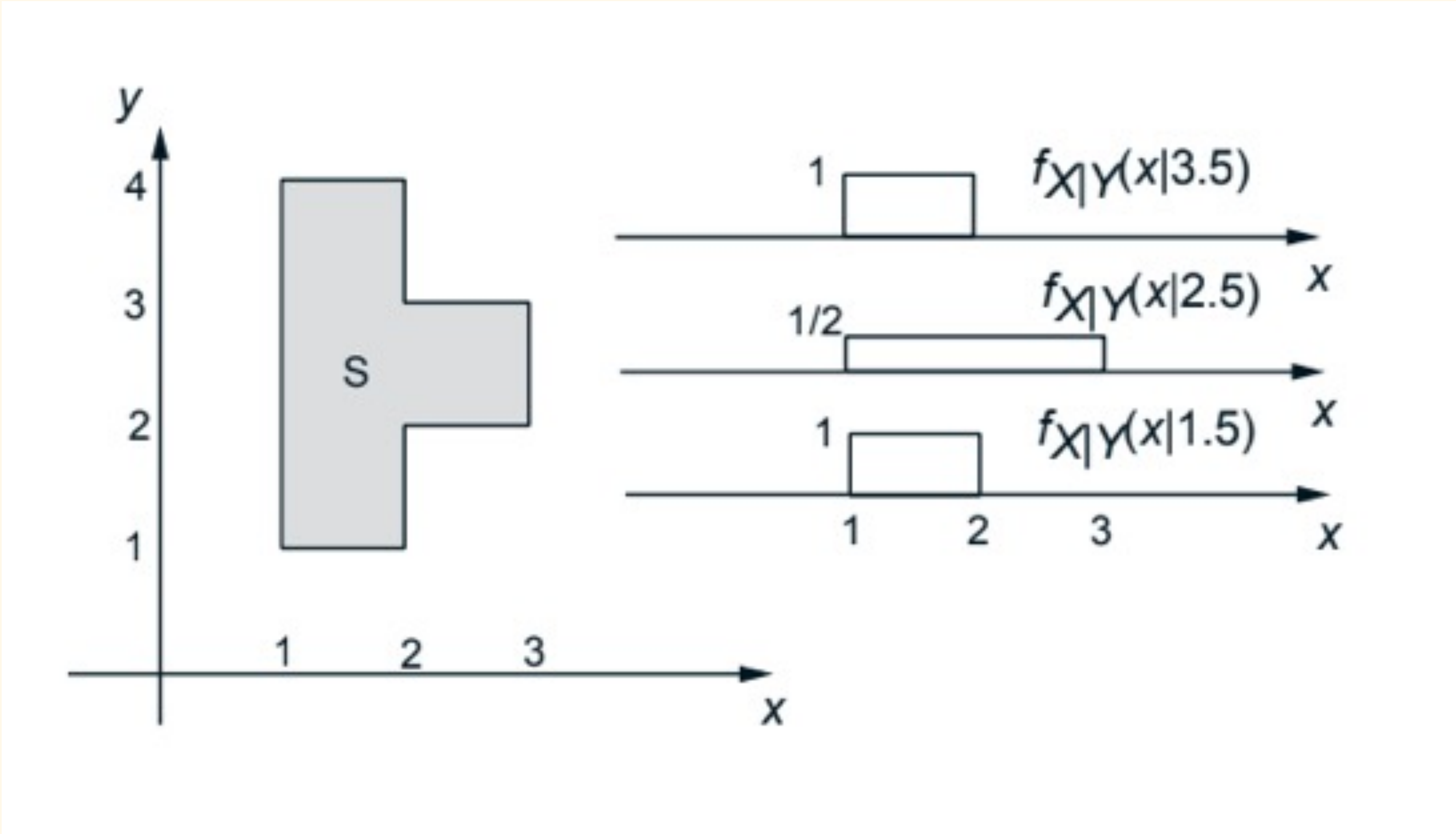
Conditioning

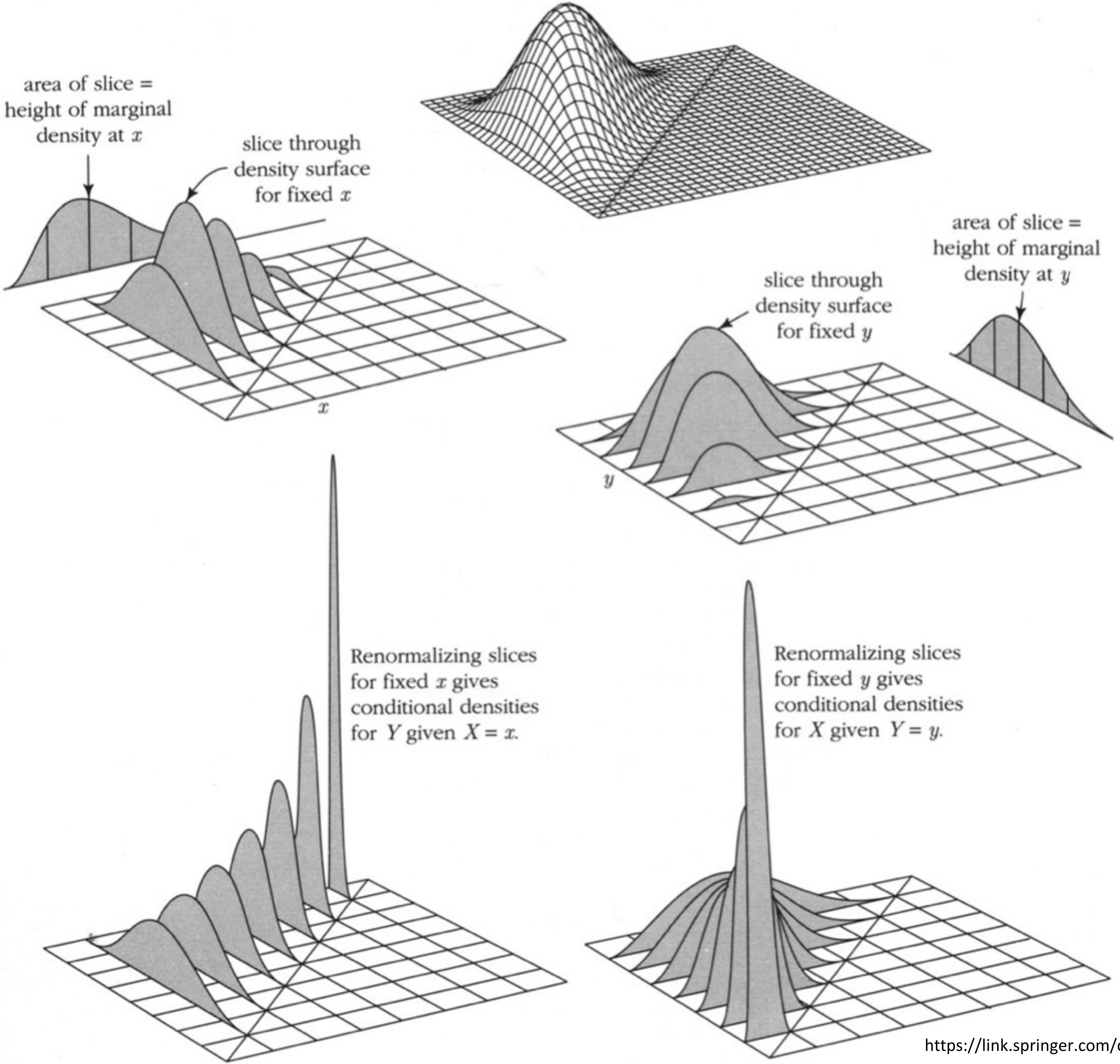
- Two random variables X, Y with join PDF $f_{X,Y}$. For any fixed y with $f_{Y(y)} > 0$ the conditional PDF of X given $Y = y$ is defined by

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

Conditioning

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

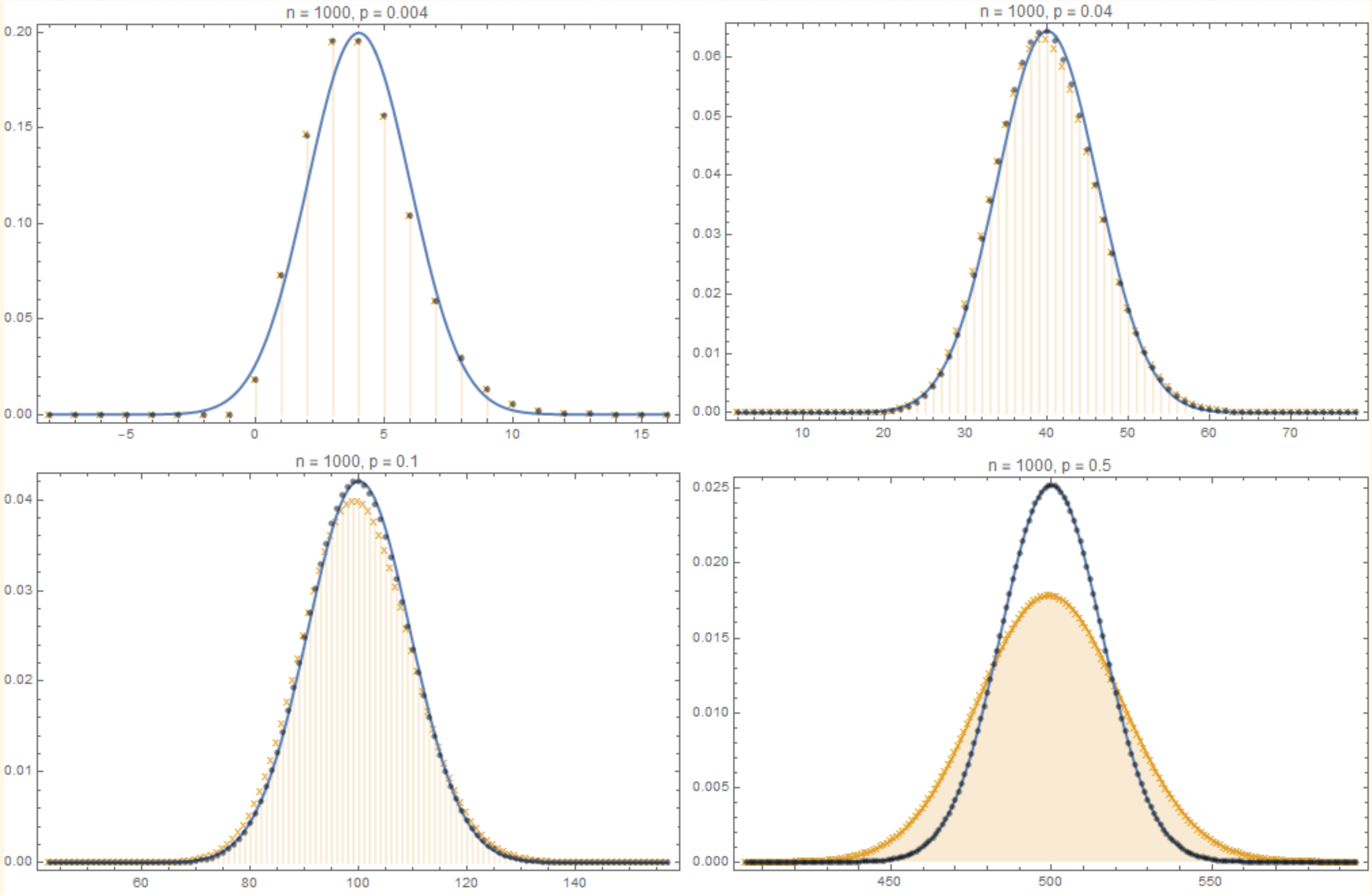



area of slice =
height of marginal
density at x
slice through
density surface
for fixed x
x
area of slice =
height of marginal
density at y
slice through
density surface
for fixed y
y
Renormalizing slices
for fixed x gives
conditional densities
for Y given $X = x$.
Renormalizing slices
for fixed y gives
conditional densities
for X given $Y = y$.

Approximation of binomial

- When p is small, n is large, binomial is best approximated by poisson distribution
- When n is large, p is not very small, binomial is best approximated by normal distribution

- Here is a good illustration: https://math.stackexchange.com/questions/3278070/approximation-of-binomial-distribution-poisson-vs-normal-distribution



https://math.stackexchange.com/questions/3278070/approximation-of-binomial-distribution-poisson-vs-normal-distribution