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 probability density function (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

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

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 $F_{\boldsymbol{X}}$

$\forall x,$

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

Geometric and exponential CDFs

Exponential PDF

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

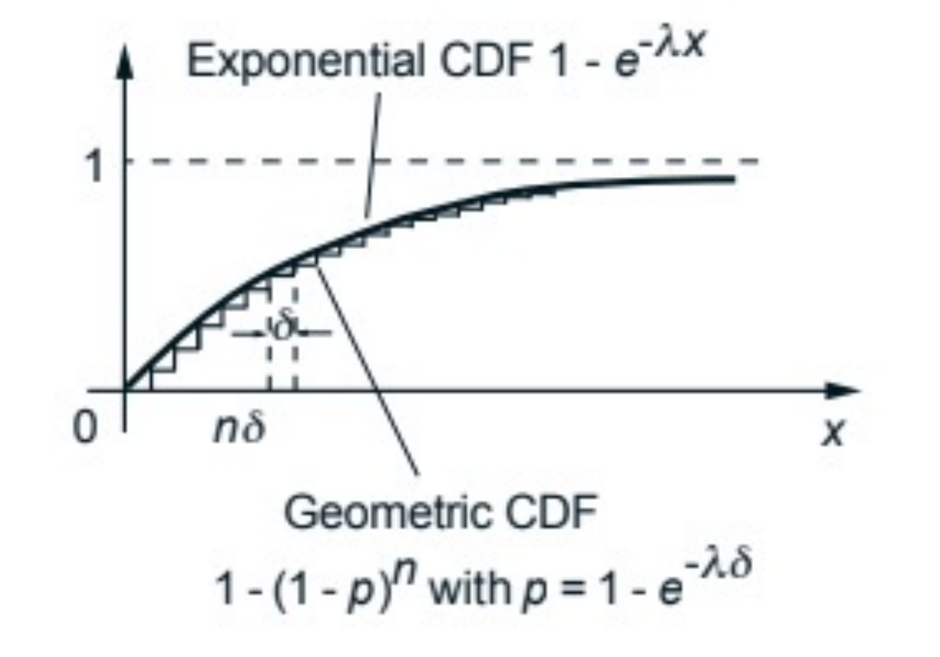
Exponential CDF

Geometric PMF

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

Geometric CDF

Geometric and exponential CDFs



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

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

Joint distribution: Marginals

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

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

- Similarly

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

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

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

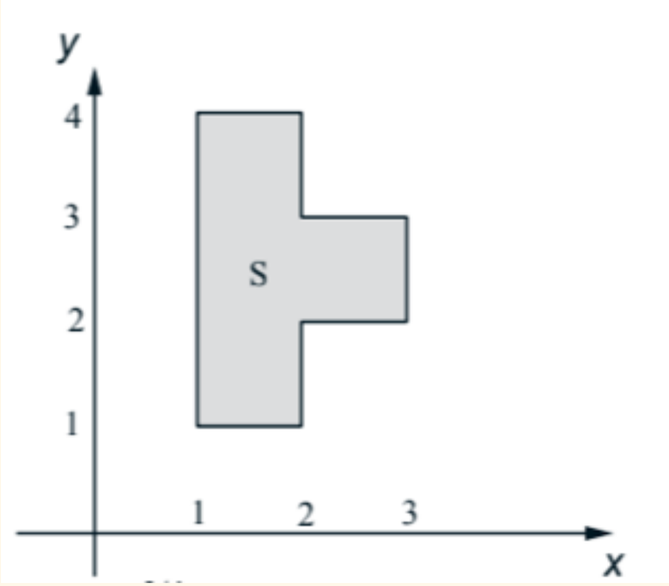
- The joint density of independent random variables X, Y is the product of the marginal densities
$$f_{X,Y}(x, y) = f_X(x) f_Y(y)$$

Example 1. 2D uniform PDF

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 arrive will wait 15 min and leave if the other hasn't arrived. What's the probability that they meet.

Example 2.

- 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*



Normal random variable

(normal distribution, Gaussian distribution)

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

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

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

CDF of standard normal

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

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

- 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. $\mathbb{P}(X \leq x)$

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

Central Limit Theorem (CLT)

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

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

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

The CDF of Z_n converge to standard normal CDF

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

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, $\mathbb{P}(S_n \le c)$ can be approximated by treating S_n as if it were normal:

1. Calculate the mean nμ 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 $\mathbb{P}(S_n \le c) \approx \Phi(z)$

   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.

Example 3. Polling

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

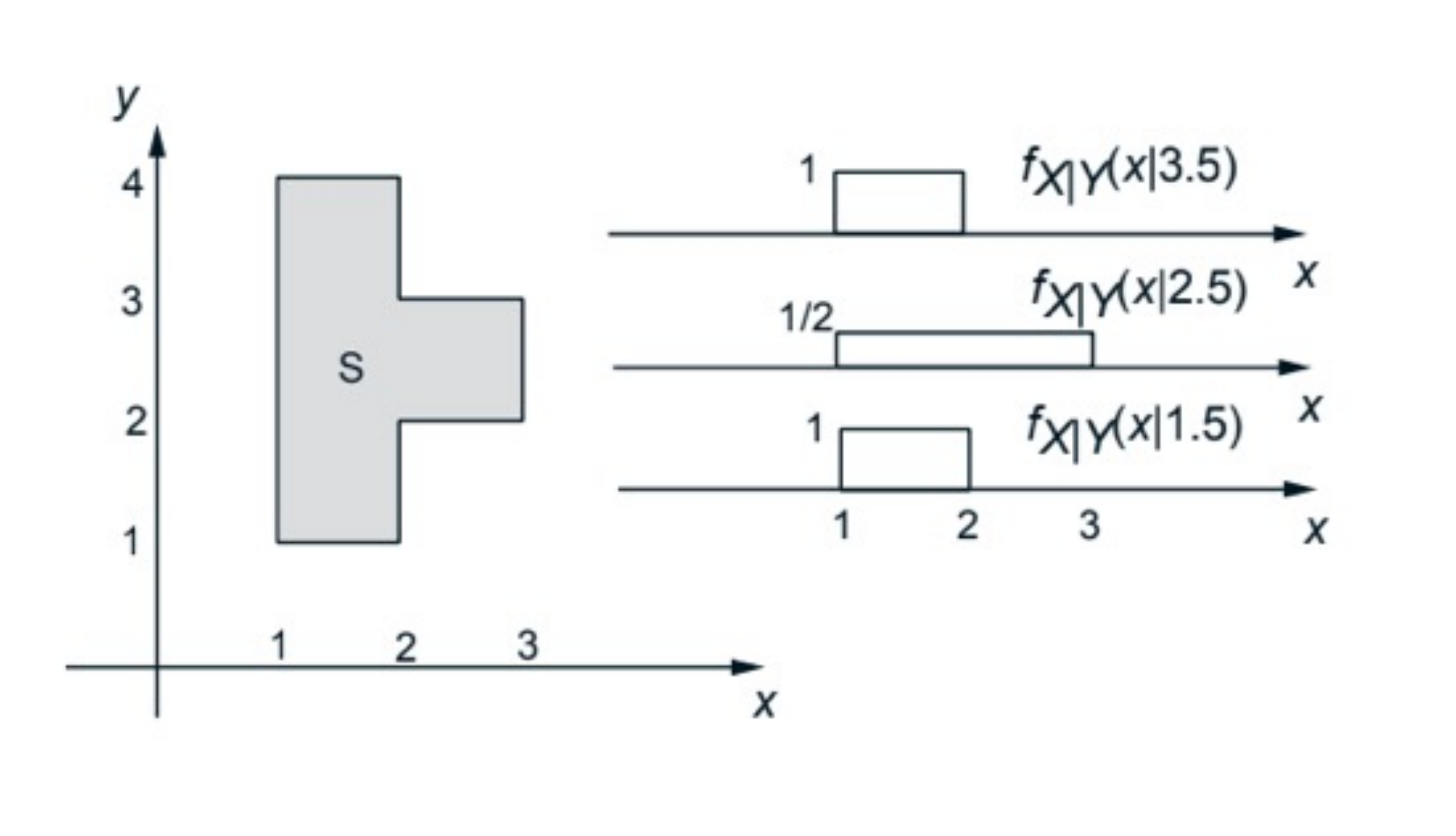
Conditioning

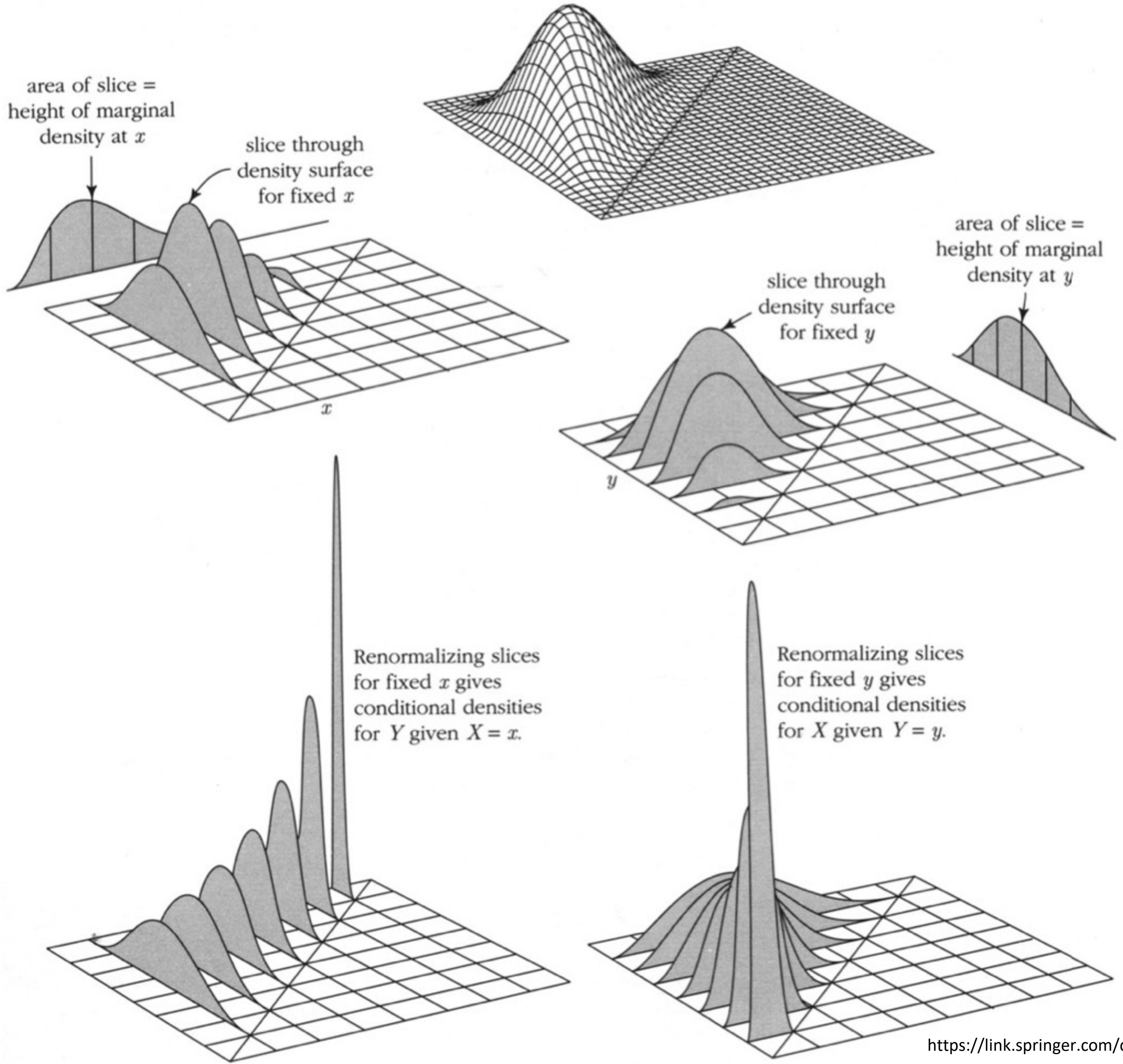
- Two random variables X, Y with join PDF $f_{X,Y}$. For any fixed y with $f_{Y(y)} > 0$ the conditional PDF of X given $Y = y$ is defined by

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

Conditioning

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

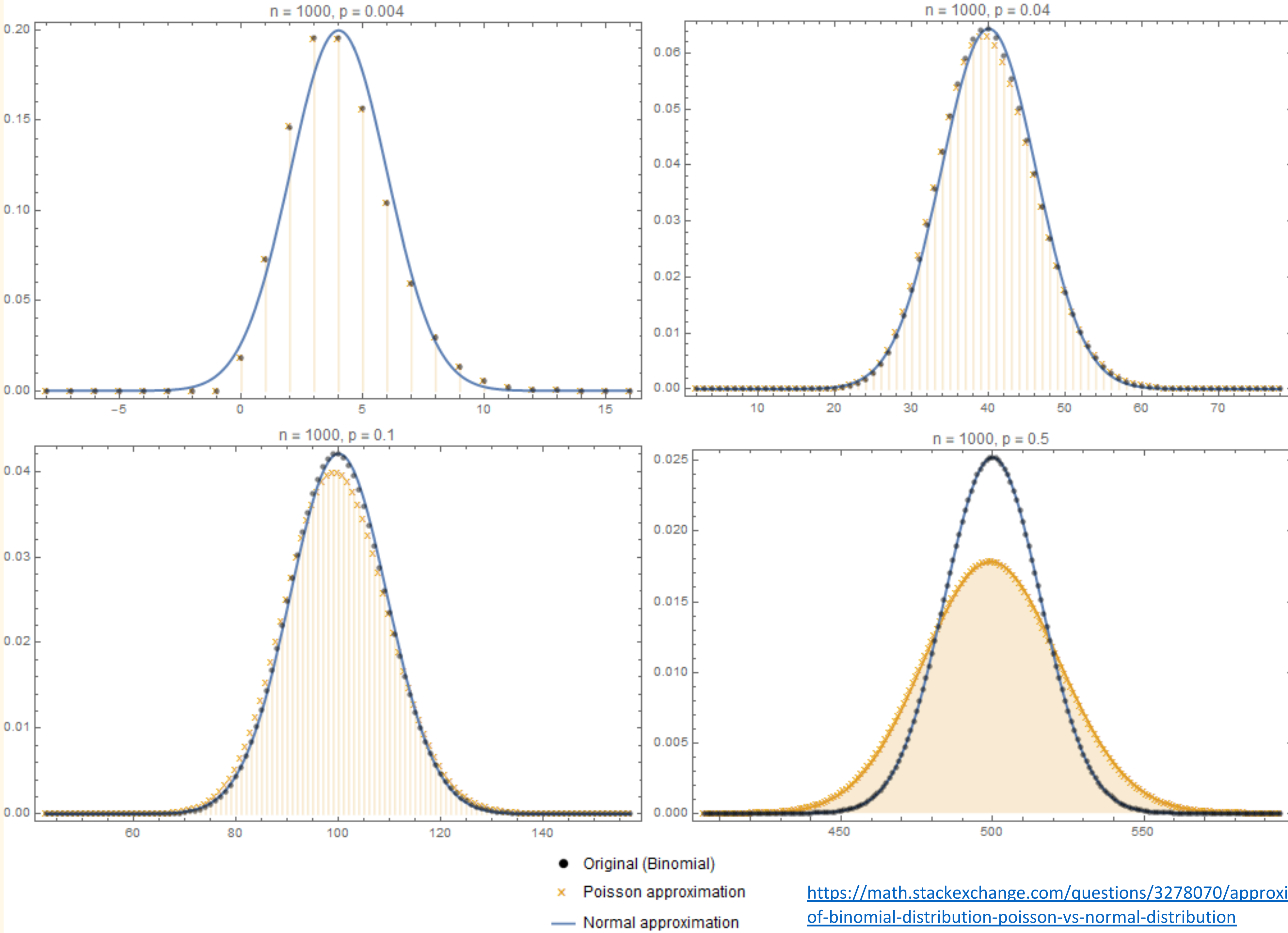



area of slice =
height of marginal
density at x
slice through
density surface
for fixed x
x
area of slice =
height of marginal
density at y
slice through
density surface
for fixed y
y
Renormalizing slices
for fixed x gives
conditional densities
for Y given $X = x$.
Renormalizing slices
for fixed y gives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