國立清華大學
NATIONAL TSING HUA UNIVERSITY

EE 306001
Probability

Lecture 13: continuous random variable

李祈均

Probability density functions

Cumulative distribution functions

Normal random variables

Summary of PDF properties

Let X be a continuous random variable with PDF $f_X(x)$

- $f_X(x) \geq 0$ for all x
- $\int_{-\infty}^{\infty} f_X(x)dx = 1$
- If δ is small, then $P([x, x+\delta]) \approx f_X(x)\delta$
- For any subset ***B*** on the real line

$$P(X \in B) = \int_B f_X(x)dx$$

Expectation properties for continuous random variable

Let X be a continuous random variable with PDF $f_X(x)$

- The expectation of X is defined by

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

- The expected value rule for a function g(X) has the form

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx$$

- The variance of X is defined by

$$Var(X) = E[(X - E[X])^2] = \int_{-\infty}^{\infty} (x - E[X])^2 f_X(x) dx$$

- So we have,

$$0 \leq Var(X) = E[X^2] - E[X]^2$$

- If $Y = aX + b$, where a and b are scalars, then

$$E[Y] = aE[X] + b$$

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

Summary of CDF properties

- The CDF of a random variable X is defined by

$$F_X(x) = P(X \le x), \text{for all x}$$

And has the following properties:

- $F_X(x)$ is monotonically non-decreasing
  - If $x \le y$, then $F_X(x) \le F_Y(y)$
- $F_X(x)$ tends to 0 as $x \to -\infty$ and tends to 1 as $x \to \infty$
- If X is discrete, $F_X(x)$is piece-wise constant function
- If X is continuous, $F_X(x)$is a continuous function

- If X is discrete and takes integer values, the PMF and the CDF can be obtained from each other by summing or differencing:

$$F_X(x) = \sum_{i=-\infty} p_X(i)$$

$$p_X(i) = P(X \leq k) - P(X \leq k-1) = F_X(k) - F_X(k-1)$$

- If X is continuous, the PDF and the CDF can be obtained from each other by integration and differentiation

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt$$

$$f_X(x) = \frac{dF_X}{dx}(x)$$

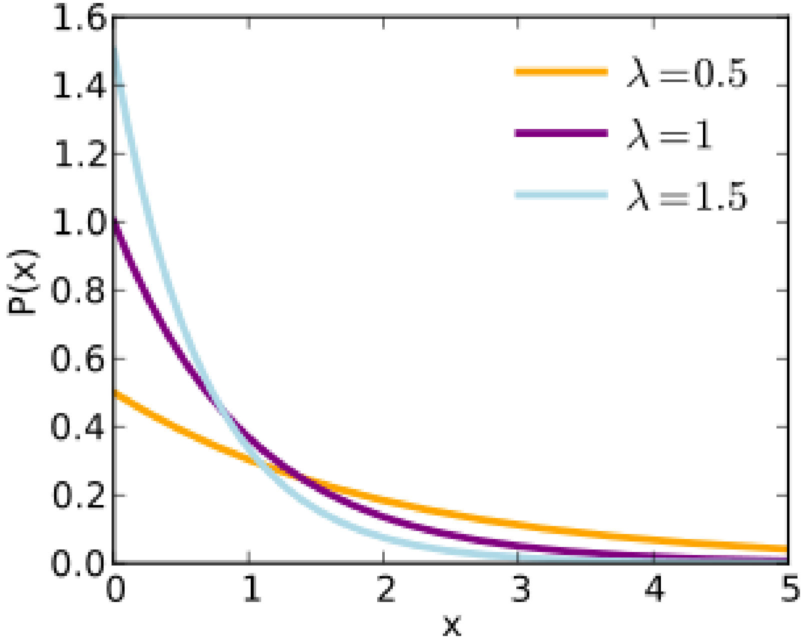
this is valid for those x at which the PDF is continuous

Exponential distribution

- Definition

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

  - λ is a positive parameter characterizing the PDF



- Legitimate PDF?

$$\int_{-\infty}^{\infty} f_X(x)dx = \int_0^{\infty} \lambda e^{-\lambda x} dx = -e^{-\lambda x}|_0^{\infty} = 1$$

- Exponential PDF good for?
  - Amount of time until an incident of interest takes place
    - Message arrival
    - Equipment breakdown
    - Light bulb burning out
    - …
- E[X]= $1/\lambda$
- Var[X]=$\frac{1}{\lambda^2}$

Geometric and Exponential CDFs

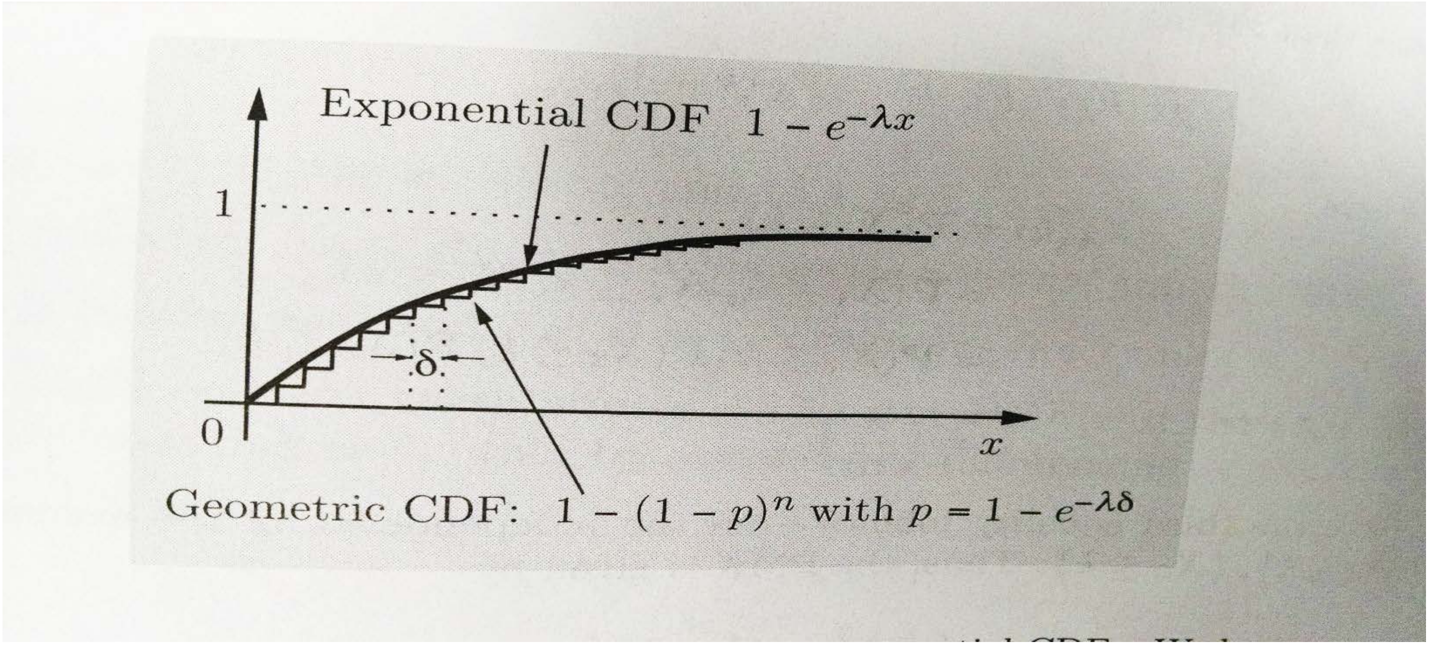
$$\sum_{k=0}^{n} r^k = \frac{1-r^{n+1}}{1-r}$$

- Geometric PMF
  - Number of trials until the first success
  - CDF?

$$F_{geo}(n) = \sum_{k=1}^{n} p(1-p)^{k-1} = p\,\frac{1-(1-p)^n}{1-(1-p)}$$
$$= 1-(1-p)^n, \text{for n} = 1,2,\ldots$$

- Exponential PDF

$$F_{exp}(x) = \int_0^x \lambda e^{-\lambda t}dt = -e^{-\lambda t}\Big|_0^x = 1-e^{-\lambda x}$$



If we set:

$$\delta = -\ln(1-p)/\lambda$$

Then:

$$e^{-\lambda\delta} = 1 - p$$

We see that

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

So, if we toss very quickly (every δ seconds, where $\delta << 1$ with a small probability of getting a head ($p = 1 - e^{-\lambda\delta}$), the first time to get a head can then be closely approximated by exponential distribution

p2

The time until a small meteorite first lands anywhere in the Sahara desert is modeled as an exponential random variable with a mean of 10 days. The time is currently midnight. What is the probability that a meteorite first lands some time between 6am and 6pm of the first day?

- Let X be the time elapsed until the event of interest, measured in days. Then X is exponential, with mean $\frac{1}{\lambda} = 10$ $\left(\lambda = \frac{1}{10}\right)$. The desired probability is:

$$P\left(\frac{1}{4} \leq X \leq \frac{3}{4}\right) = P\left(X \geq \frac{1}{4}\right) - P\left(X > \frac{3}{4}\right)$$

$$= e^{-1/40} + e^{-3/40} = 0.0476$$

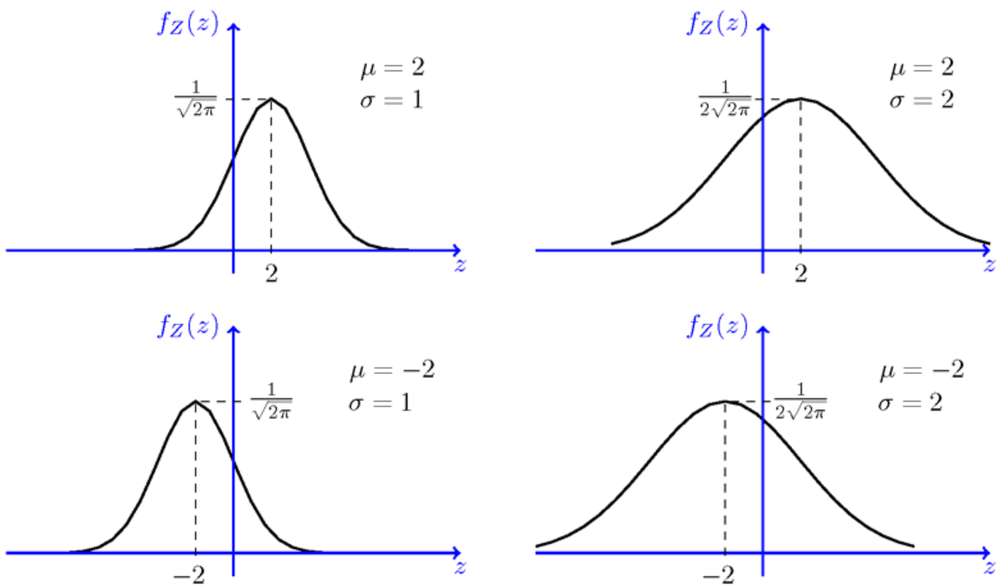
$$F_{exp}(x) = 1 - e^{-\lambda x}$$

NORMAL RANDOM VARIABLE

One of the most used random variable (parameter: μ,σ)

Reason will come clear when we get to limit theorem

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



How to computer probability? (find the area, how?) use CDF

CDF calculation for a normal random variable

- No close form solution for normal CDF
- Two step procedure in calculating the CDF of a normal random variable

Say a normal random variable X with μ and variance σ^2:

a) "Standardize" X: subtract μ and divide by σ to obtain a standard normal random variable Y
   a) Think about what happen to Expected value and Variance after standardization?

b) Read the CDF values from the standard normal table

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

	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817

Just look up table $P(X<x)$

p3

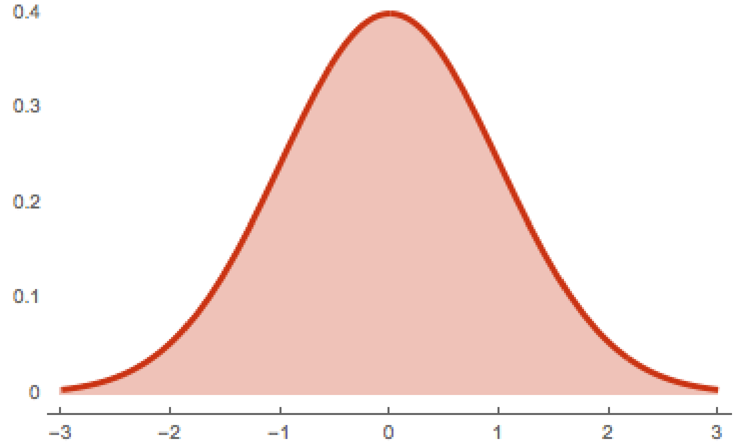
Let X and Y be Gaussian random variables, with $X \sim N(0,1)$ and $Y \sim N(1,4)$

a) Find $P(X \leq 1.5)$and $P(X \leq -1)$

Just go through the drill (look at the table):

$$P(X \leq 1.5) = \Phi(1.5) = 0.9332$$

$$P(X \leq -1) = 1 - P(X \leq 1) = 1 - \Phi(1) = 1 - 0.8413$$
$$= 0.1587$$



b) What is the distribution of $\frac{Y-1}{2}$

Linear transformation of Gaussian random variable is also a random variable (will prove this next section) -> a very nice property that we will keep using in the future

$$Y \sim N(1,4)$$

$$E\left[\frac{Y-1}{2}\right] = \frac{1}{2}(E[Y] - 1)$$

$$Var\left(\frac{Y-1}{2}\right) = Var\left(\frac{Y}{2}\right) = \frac{1}{4}Var(Y) = 1$$

$$\frac{Y-1}{2} \sim N(0,1)$$

Standardization of normal random variable

c) Find $P(-1 \leq Y \leq 1)$

standardize

$$P(-1 \leq Y \leq 1) = P\left(\frac{-1-1}{2} \leq \frac{Y-1}{2} \leq \frac{1-1}{2}\right)$$

$$= \Phi(0) - \Phi(-1) = \Phi(0) - \big(1 - \Phi(1)\big) = 0.3413$$

Look up table

Signal detection

- A binary message is transmitted as a signal *s*, which is either -1 or +1. The communication channel corrupts the transmission with additive normal noise with mean μ=0 and variances σ^2.

This actually has a proof, we will do it in later sections

The receiver that the signal -1 (or 1) was transmitted if the value received is <0 (or $\geq$0)

Now what is the probability of error?

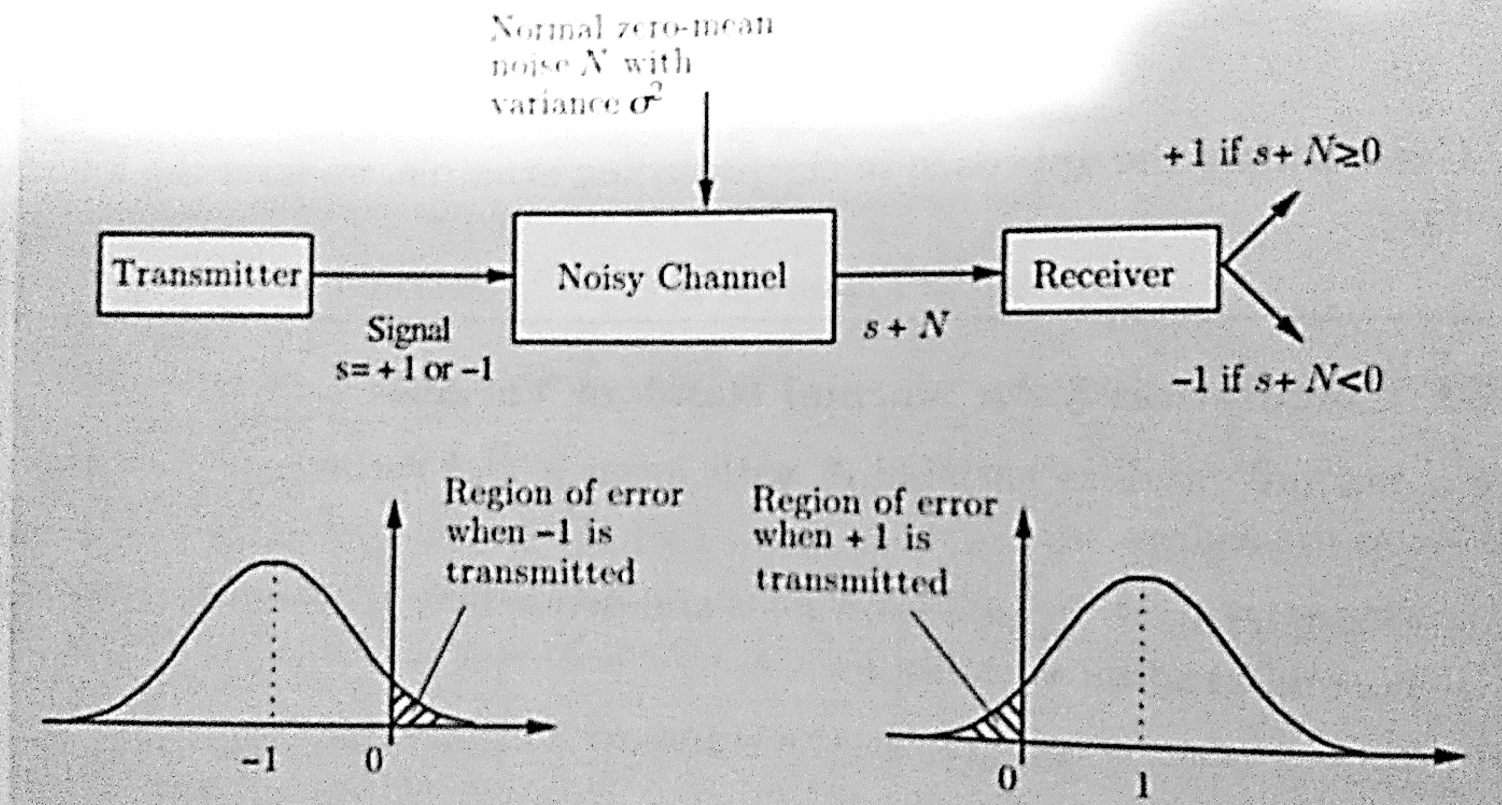


Figure 3.11: The signal detection sche...

- An error occurs whenever -1 is transmitted and the noise N is at least 1 so that s+N=-1+N$\geq$0

- Or whenever 1 is transmitted and the noise N is smaller than -1 so that s+N=1+N<0

$$P(N \geq 1) = 1 - P(N < 1) = 1 - P\left(\frac{N-\mu}{\sigma} < \frac{1-\mu}{\sigma}\right)$$

$$= 1 - \Phi\left(\frac{1-\mu}{\sigma}\right) = 1 - \Phi\left(\frac{1}{\sigma}\right)$$

By symmetry $P(N \leq -1)$ is the same

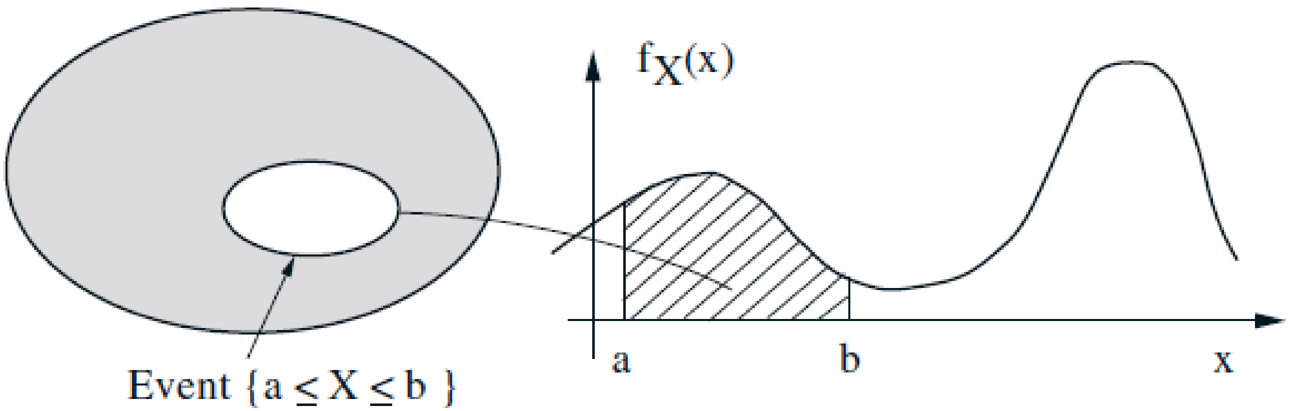
If σ = 1, probability of error = 1 - Φ(1) = 1 – 0.8413 = 0.1587

Readings: Section 3.4 – 3.5

Lecture outline

- Multiple random variables
  - Conditioning
  - Independence
- Examples

Continuous r.v.'s and pdf's



$$P(a \le X \le b) = \int_a^b f_X(x)dx$$

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

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

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

  - 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 ($\mathrm{X} \in A$)

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

- 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 and for PMF
- Two random variables are independent iff their joint PDF factors into product

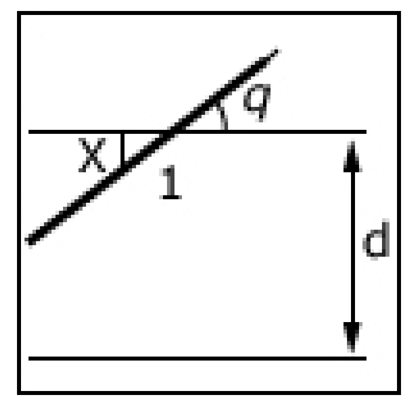
$$f_{X,Y}(x,y) = f_X(x) * f_Y(y)$$

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

Buffon's needle

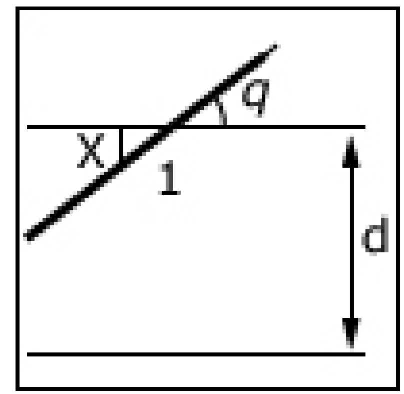
Say a surface is ruled with parallel lines, which are at distance *d* from each other. Suppose that we throw a needle of length *l* on the surface at random, it can happen either intersecting or non-intersection,

What is the probability that the needle will intersect one of the lines?



- Now we suppose $l < d$, so that the needle can not intersect two lines simultaneously

- Let X be the vertical distance form the midpoint of the needle to the nearest of the parallel lines
  - The midpoint of the needle can only locate between the two parallel lines



- Let θ be the acute angle formed by the axis of the needle and the parallel line

- Now we can model the pair of the random variable $(X, \theta)$with a uniform joint PDF over the rectangular set $\left\{(x, \theta) | 0 \leq x \leq \frac{d}{2}, 0 \leq \theta \leq \frac{\pi}{2}\right\}$

- Now, with the idea of this is a uniform probability
  - The density of X is uniform over d/2

$$f_X(x) = \frac{2}{d}, \text{if } 0 \le \text{x} \le \frac{d}{2}$$

  - The density of Y is uniform over (0, π/2)

$$f_\theta(\theta) = \frac{2}{\pi}, \text{if } 0 \le \theta \le \frac{\pi}{2}$$

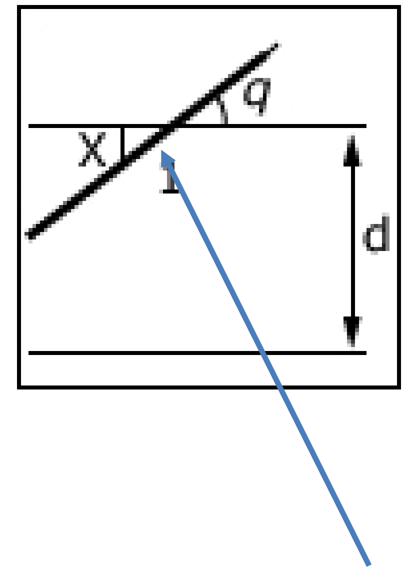
- With that setup:
  - Assuming that the orientation and the distance to the parallel lines are independent

$$f_{X,\theta}(x,\theta) = \begin{cases} \frac{4}{\pi d}, if\ x \in \left[0, \frac{d}{2}\right], \theta \in [0, \frac{\pi}{2}] \\ 0, otherwise \end{cases}$$

- Now, we will go ahead and try to identify the event of interest:

- The length of the segment between the mid point of the needle and the point of intersection of the axis of the needle with the closest parallel line is $x / \sin\theta$

- So the needle will intersect with the parallel line iff this length is less than $\frac{l}{2}$



- Once we identify the event of interest, the only thing left is integration:

Event of interest: $X \leq \frac{l}{2}\sin\theta$

$$P\left(X \leq \left(\frac{l}{2}\right)\sin\theta\right) = \iint_{x \leq \left(\frac{l}{2}\right)\sin\theta} f_{X,\theta}(x,\theta)\,dx\,d\theta$$

$$= \frac{4}{\pi d}\int_0^{\pi/2}\int_0^{l/2\sin\theta} dx\,d\theta = \frac{4}{\pi d}\int_0^{\pi/2}\frac{l}{2}\sin\theta\,d\theta = \frac{2l}{\pi d}(-\cos\theta)\Big|_0^{\frac{\pi}{2}}$$

$$= \frac{2l}{\pi d}$$

It has been used to empirically evaluate the number of π

- More of a side note:
  - This way of evaluating is called Monte-Carlo method
  - Why?
    - When you have a function that's extremely difficult to compute
    - You can generate random samples over and over
    - Then estimate the probability, which is then equals to that function of the number (irrational number)
    - Especially useful for physicists, statisticians, and even computer scientists nowadays

Conditioning

Recall:

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

- Density gives us probabilities of little intervals

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

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

  – 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:

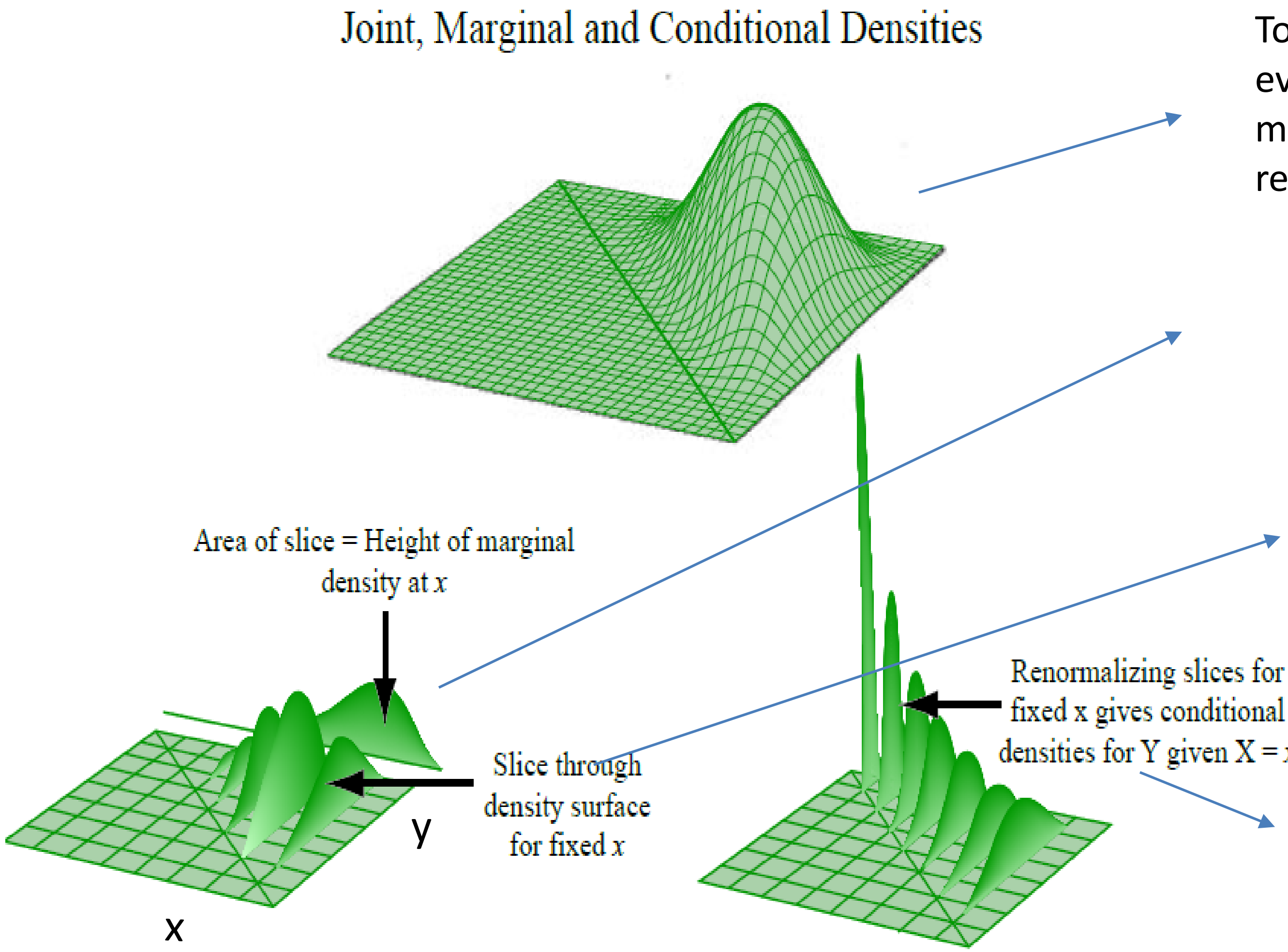
$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}, \text{if } f_Y(y) > 0$$

- 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
  - But you need to normalize that slice

Independence

- Given that we know something about conditional PDF

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



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

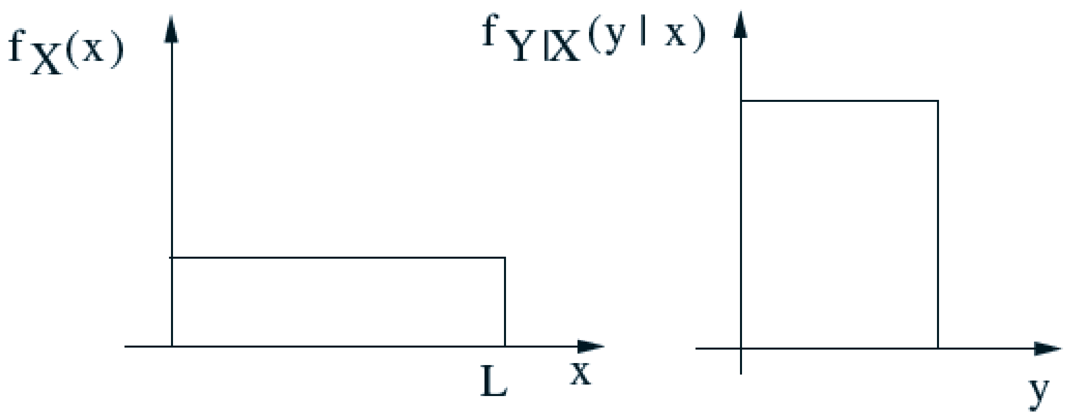
Marginal PDF:
Integrate over all Y=y's

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

Conditional PDF:
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 $[0, X]$

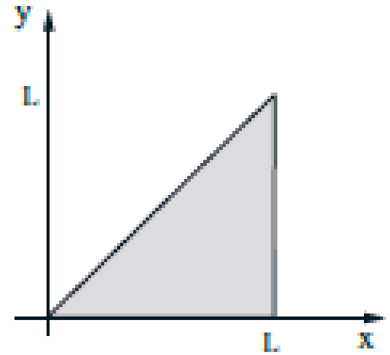


- What is the joint PDF between X, Y

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

- Essentially the multiplication that we know of

  - So what is ? $f_X(x) = 1/l$
  - What is? $f_{Y|X}(y|x) = \frac{1}{x}$
  - 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

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

- 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!