31.10.2022

YZV 231E

Probability Theory & Stats

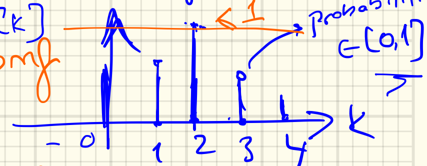
Week 7

Gü

Recap:

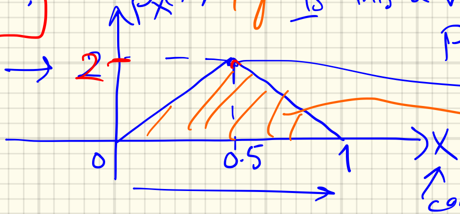
R.V.S : pmf or pdf is a complete characterization of an r.v.

$S \xrightarrow{X(\cdot)} S^X \subseteq \mathbb{R}$; pmf $[0,1]$; $p_X(x)$; $p_X[k]$ pmf



Q. $p_X(x)$: pdf. Is this a valid pdf?

$p_X(0.5) = 2$.



continuous r.v.

Yes ✓ satisfies pdf properties

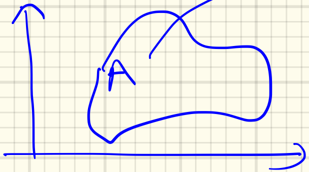
$\int p_X(x)dx = 1$ ✓

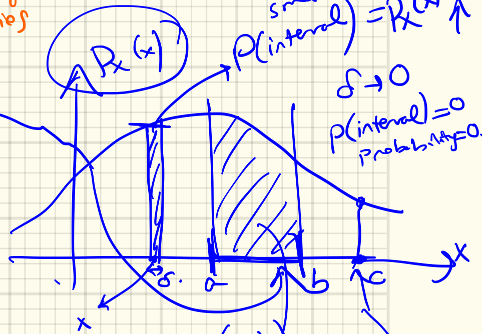
Continuous r.v.s : $\to$ pdf.

Properties pdfs:

1. $p_X(x) \geq 0$ ✓ prob. density but $p_X(x)$ can be larger than 1 $\to$ b/c $p_X(x)$ (pdf) is a prob. density! Not a probability

2. $\int_{-\infty}^{\infty} p_X(x)\,dx = 1$





$P(\text{small interval}) = p_X(x)\cdot\delta$

$\delta \to 0$

$P(\text{interval}) = 0$

probability $= 0$.

$A \to (a,b)$

$P(A) = P(a < X < b) = \int_a^b p_X(x)\,dx.$

3. $P(X = x_0) = \int_{x_0}^{x_0} p_X(x)dx = 0$ b/c $\overset{\delta}{\Delta x} = x_\delta - x_0 = 0$

4. $P(a < X < b) = \int_a^b p_X(x)dx$ $\approx p_X(x + \frac{\delta}{2})\cdot\delta$.

$P(x \leq X \leq x+\delta) = \int_x^{x+\delta} p_X(t)dt \approx p_X(x)\cdot\delta$.

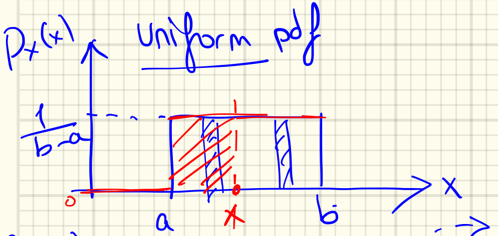
$P(X = c) = 0 = \int_c^c p_X(x)dx = 0$

Cumulative Distribution Function (CDFs):

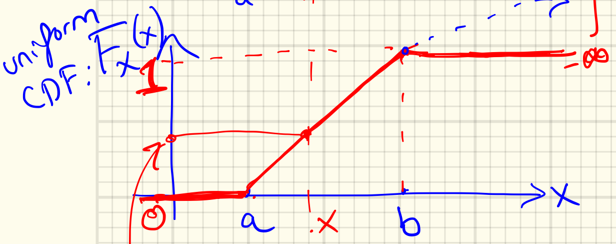
$$F_X(x) \triangleq P(X \le x) = \int_{-\infty}^{x} p_X(x)\,dx$$

- CDF carries the same info as the pdf. ✓
- CDF unifies discrete & continuous r.v.s. b/c the same defn. applies.
- CDFs are well-defined in both domains.

Uniform pdf



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

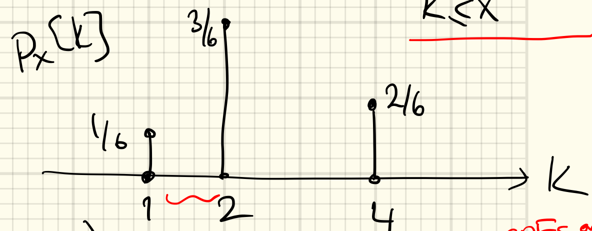


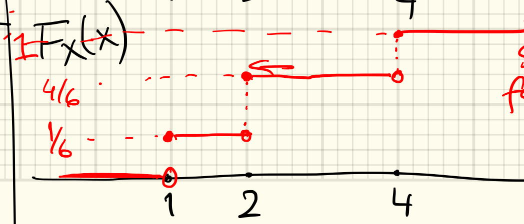
$$\int_{-\infty}^{x} p_X(x)\,dx$$

$*$ $$\frac{d}{dx} F_X(x) = p_X(x)$$

Recall: for discrete r.v.s

$$\boxed{F_X(x) = P(X \le x) = \sum_{k \le x} \underbrace{p_X[k]}_{\text{pmf}}}$$





CDFs are staircase for discrete r.v.s.

$*$ For continuous r.v.s, CDFs are continuous fns. going from 0 to 1.

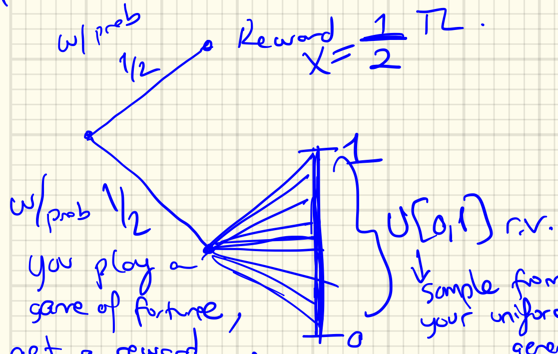
$*$ Monotonically non-decreasing

Mixed r.v.s

. R.v.s that are a mixture of continuous & discrete r.v.s.

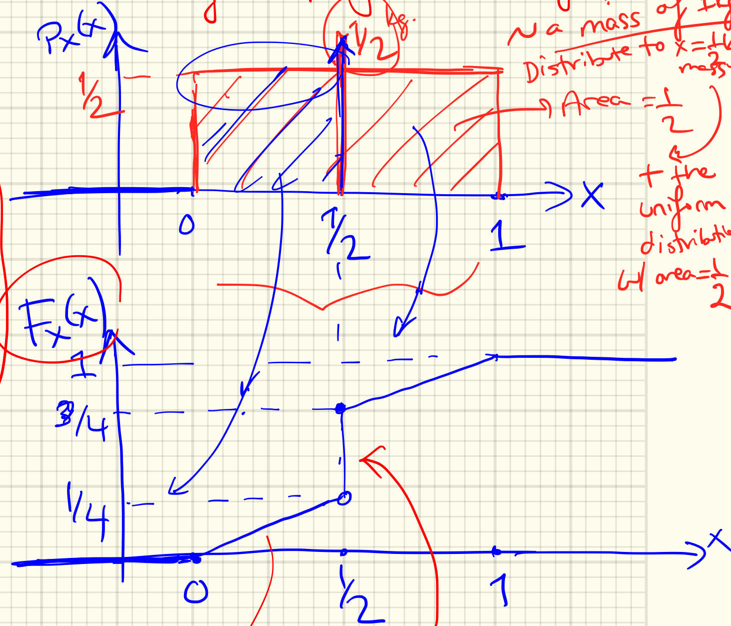
Ex: Play a game: w/a certain prob. you get some award.

mixed r.v. example



w/ prob 1/2 you play a game of fortune, get a reward btw (0,1) → uniform continuous r.v.

A combination of a pdf & a pmf:



~ a mass of 1kg. Distribute to $x = \frac{1}{2}$ kg mass

→ Area $= \frac{1}{2}$

+ the uniform distribution w/ area $= \frac{1}{2}$

$$P\left(X \leq \frac{1}{2}\right) = \text{Prob. of getting a reward of } \tfrac{1}{2}\text{ TL or less}$$

$$= F_X\left(\frac{1}{2}\right) = \frac{3}{4}$$

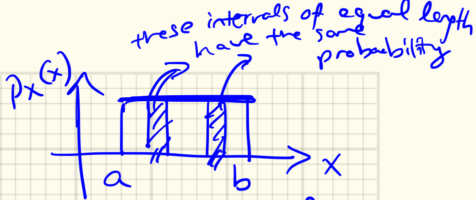
- jumps correspond to discrete r.v. part
- continuous part correspond to continuous r.v. part of this mixed r.v.

* The mass at $x = \frac{1}{2}$ is a impulse fn. Dirac Delta fn.

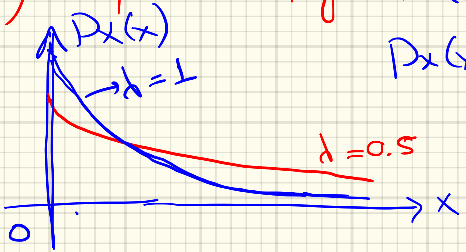
→ Generalized functions.

Important pdfs:

1) Uniform pdf: $\underline{X \sim U[a,b]}$



2) Exponential pdf: $X \sim \exp(\lambda)$ — models lifetime of a product



$$p_X(x) = \begin{cases} \boxed{\lambda e^{-\lambda x}}, & x \geq 0 \\ 0 & \text{o/w} \end{cases}$$

check:

$p_X(x) \geq 0$ ✓

$\int_{-\infty}^{\infty} \lambda e^{-\lambda x} dx = 1$? ✓

λ: event rate

eg. failure rate.

exercise: Calculate $E[X] = \int_0^{\infty} x \cdot \lambda e^{-\lambda x} dx$

$P(X \geq 100)$ = Prob. that the device will last 100 days
= it will fail after 100 days.

X: lifetime (in days) of a device.

$\boxed{\text{Choose } \lambda = 0.01}$: failure/~~d~~death rate

$= 1 - P(X < 100) = 1 - F_X(100) = \int_{100}^{\infty} \lambda e^{-\lambda x} = 1 - \int_{-\infty}^{100} \lambda e^{-\lambda x}$

$= -e^{-\lambda x}\Big|_{100}^{\infty} = 0.367$

If $\lambda = 0.001 \rightarrow P(X \geq 100) = 0.904$

Exercise: find a way to estimate λ. $\rightarrow$ failure rate of the device.

Note: Exp distribution has a memoryless property.

$$\underbrace{P(X > s+t}_{} \mid \underbrace{X > s}_{\text{Given the bulb survived } s \text{ units of time}}) = \underbrace{P(X > t)}_{\text{prob. of the fresh bulb surviving } t \text{ units of time}}$$

X: survival lifetime of a bulb

prob that the bulb survives a further t units of time.

$$P(\underbrace{X > s+t}_{} \mid \underbrace{X > s}_{}) = \frac{P(X > s+t,\ X > s)}{P(X > s)} = \frac{P(X > s+t)}{P(X > s)}$$

$$= \frac{\int_{s+t}^{\infty} \lambda e^{-\lambda x}\,dx}{\int_{s}^{\infty} \lambda e^{-\lambda x}\,dx} = \frac{e^{-\lambda x}\Big|_{s+t}^{\infty}}{e^{-\lambda x}\Big|_{s}^{\infty}} = \frac{(e^{-\lambda(s+t)})}{(e^{-\lambda s})} = e^{-\lambda t} = P(X > t)$$

memoryless property.

3) Gaussian (Normal) pdf : ★★★

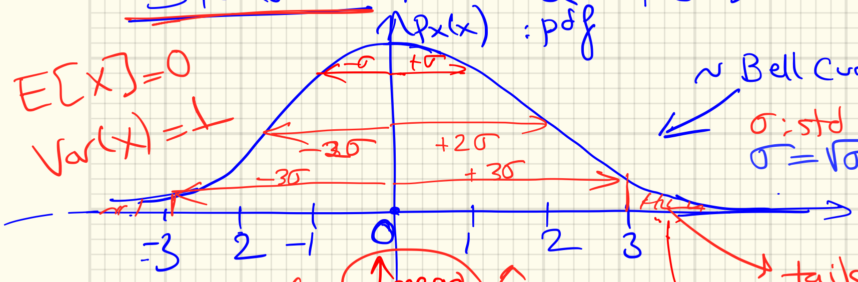
Standard Normal $\sim \mathcal{N}(0, 1) = p_X(x) = \frac{1}{\sqrt{2\pi}} \boxed{e^{-x^2/2}}$

μ, $\sigma^2 = 1$

$\frac{1}{\sqrt{2\pi}}$: normalizing constant. to make the integral = 1.

$E[X] = 0$

$Var(X) = 1$



$\sim$ Bell Curve

σ : std deviation

$\sigma = \sqrt{\sigma^2} = \sqrt{Var(X)}$

exercise : use $p_X(x)$ form, to calculate $E[X] = 0$, $Var(X) = 1$.

↓ IBP.

tails of the distribution.

mean

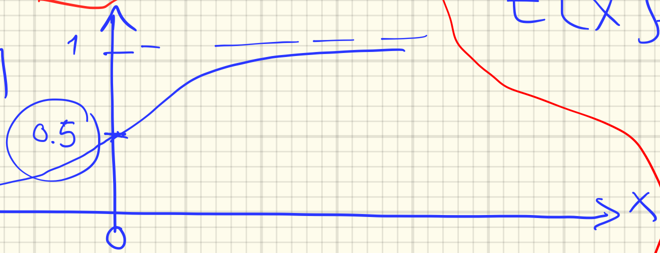
symmetric pdf.

$$E[X] = \int_{-\infty}^{\infty} x \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \, dx$$

$$\underbrace{\qquad}_{\sqrt{\frac{2}{\pi}}\, e^{-x^2/2} \; - \; \cdots}$$

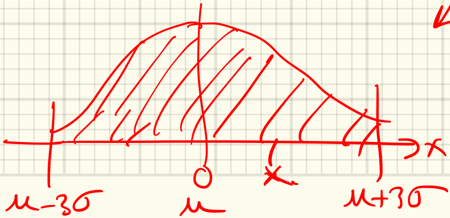
Normal CDF

$F_X(x)$



$\boxed{F_X(x=0) = 0.5}$

Check yourself.



$$P(\mu - 3\sigma \le X \le \mu + 3\sigma)$$

$$= 0.997 \ldots$$

→ General Normal $\sim \mathcal{N}(\mu, \sigma^2)$

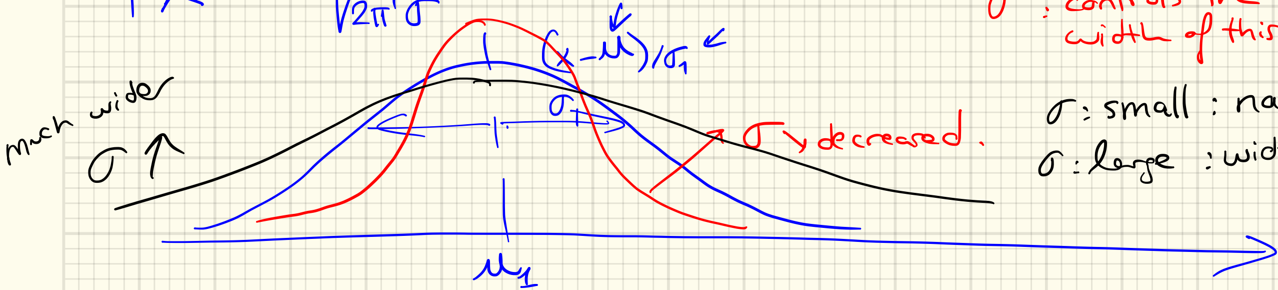
μ: mean, σ^2: var.

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

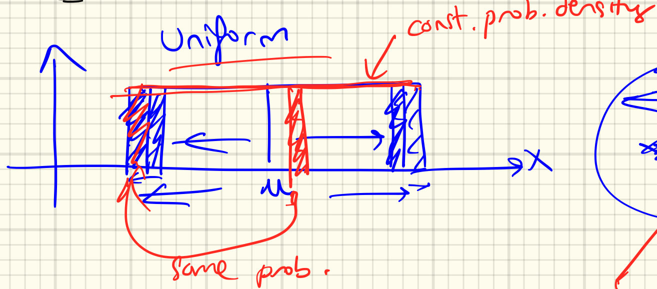
σ : controls the width of this distrib.

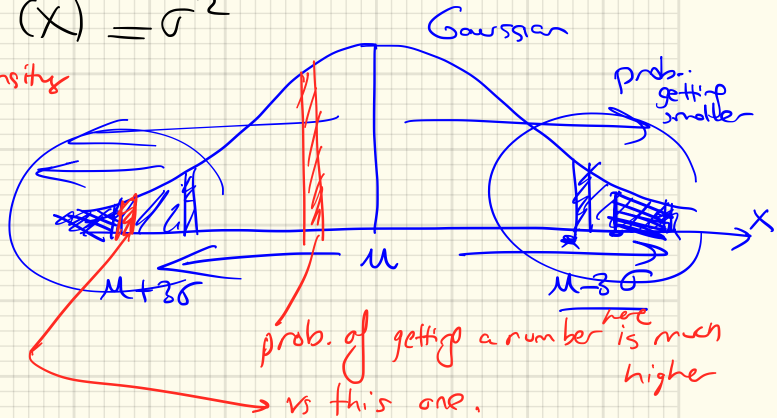
σ : small : narrow pdf

σ : large : wide pdf.



$$E[X] = \mu \quad , \quad Var(X) = \sigma^2$$





prob. of getting a number here is much higher vs this one.

Fact : $\boxed{Y = aX + b}$ (suppose X is normal)

$$E[Y] = E[g(X)] = \int \underbrace{g(x)}_{ax+b} p_X(x)\,dx$$

$$= \int a\,x\,p_X(x)\,dx + \int b\,p_X(x)\,dx = a\,\underbrace{E[X]}\,+ b$$

$$E[X] = \mu \;\rightarrow\; E[Y] = a\mu + b$$

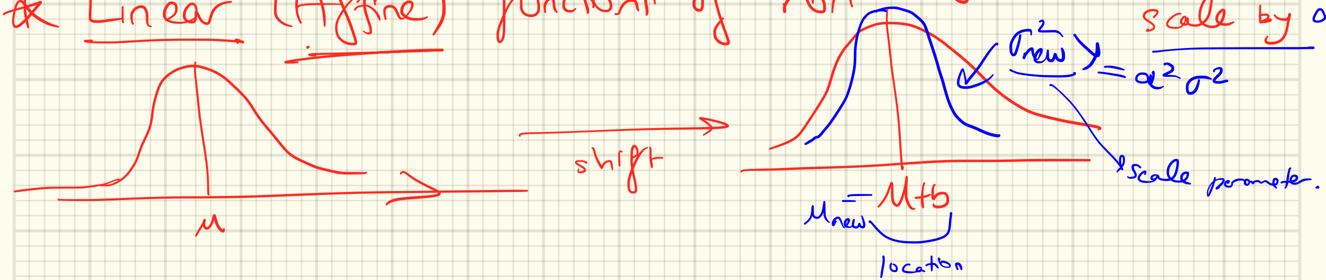
$$Var(Y) = a^2\,Var(X) = a^2\sigma^2$$

$$Y \sim N(a\mu + b,\ \underbrace{a^2\sigma^2}_{\sigma^2_{new}})$$

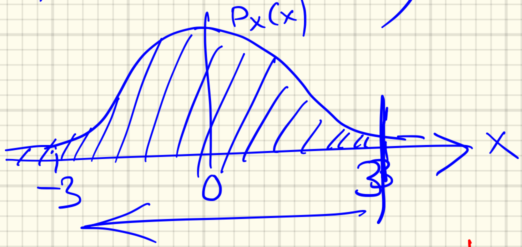
∴ Fact ✓

* Linear (Affine) functions of normal variables are also normal

scale by a



Q: $P(X \le 3) = ?$ $\quad X \sim \mathcal{N}(0,1)$



$$= \int_{-\infty}^{3} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = ?\ !!$$

No closed form formula for this!

Solution: We tabulate this w/ Lookup tables: For the std Normal cdf.

$$\Phi(x) = F_X(x)$$

$$P(X \le 0.63) = \Phi(0.63)$$

$$= F_X(0.63) = 0.7357.$$

	0.00	0.01	0.02	$\boxed{0.03}$	...	0.09
0.0						
0.1						
0.2						
⋮						
0.5						
$\boxed{0.6}$				$\boxed{0.7357}$		
⋮						

Ex: Calculating Normal Probabilities

If $X \sim \mathcal{N}(2, 16)$; $P(X \leq 3) = ?$

Use the LUT (std Normal)

Use: If $X \sim \mathcal{N}(\mu, \sigma^2)$

Then $\boxed{\dfrac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)}$

This is called STANDARDIZATION of an r.v.

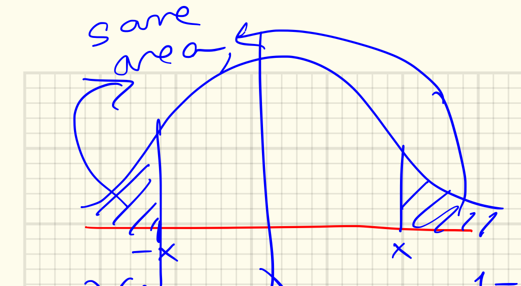
$X \sim \mathcal{N}(2, 16) \rightarrow \mu = 2, \ \sigma = 4$

Event of interest $\{X \leq 3\} \equiv \{\tilde{X} \leq \frac{1}{4}\}$ same event

$P(\underset{\substack{\uparrow \\ \text{non-std Gaussian}}}{X} \leq 3) = P\Big(\underbrace{\dfrac{X-2}{4}}_{\tilde{X}} \leq \dfrac{3-2}{4} = \dfrac{1}{4}\Big) = \underset{\Phi}{\mathrm{CDF}}(0.25) = 0.5987.$

$\tilde{X}$: std Gaussian $\sim \mathcal{N}(0, 1)$

* make sure you know how to do these calculations.

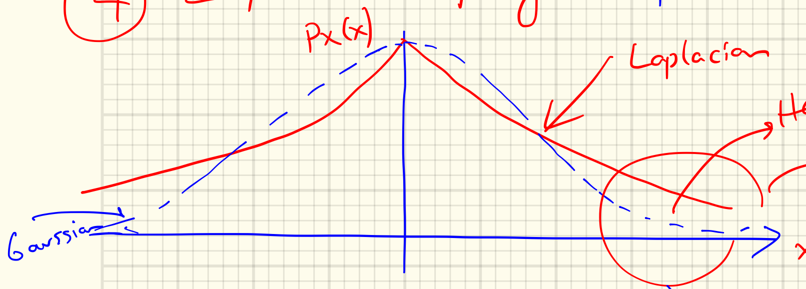


$$p_X(x) \propto e^{-x^2/2\sigma^2}$$

For $P(X \leq -x) \longrightarrow$ use $1 - F_X(x) = 1 - P(X \geq x)$

(4) Laplacian pdf: $p_X(x) \propto e^{-|x|}$

: models phenomena w/ some extreme events.



Heavy-tailed distribution

→ tail probabilities do not go to 0 as fast as that of the Gaussian.

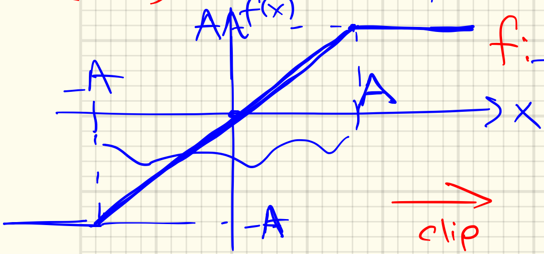
∴ prob. of a large extreme value is far larger w/ Laplace distrib. than a normal.

$$p_X(x) = \frac{1}{\sqrt{2\sigma}} \exp\left(-\frac{2}{\sigma^2}|x|\right), \quad \sigma^2 > 0, \quad E[X] = 0.$$

eg. used in hydrology → use it to model extreme events such as max. daily rainfall / year.

- used to model speech amplitudes.

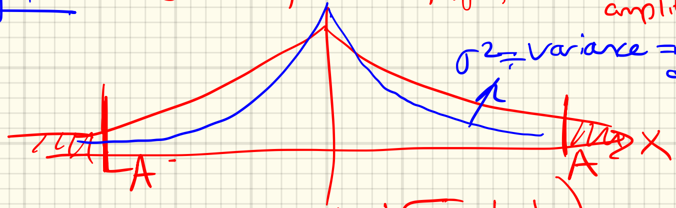
(Skay) Ex 10.10 . Speech analysis. $f(x)$: speech. $|X| \geq A \rightarrow$ clip the speech.



f: Clipping fn.

Use Laplacian pdf for the speech amplitude

σ^2 = variance = measure of spread.



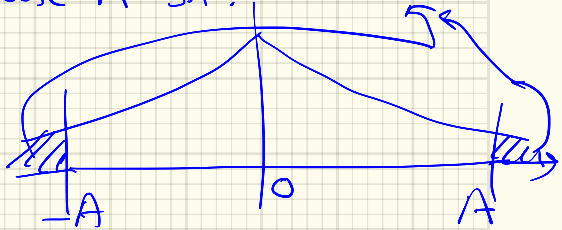
clip.

No-clipping of the signal

$$p_X(x) = \frac{1}{\sqrt{2\sigma^2}} \exp\left(-\sqrt{\frac{2}{\sigma^2}}\,|x|\right), \quad -\infty < x < \infty$$

Design requirement: xmit a speech signal w/o clipping 99% of the time.

Clipping occurs when $|x| > A$: Choose A s.t. $P(\text{clip}) \leq 0.01$.



$$\underbrace{P(\text{Clip})}_{\text{event}} = P(X > A \text{ or } X < -A)$$

$$= 2P(X > A) \quad \text{due to symmetry of the Laplacian around } x = 0$$

$$P_{\text{clip}} = 2\int_A^\infty \frac{1}{\sqrt{2\sigma^2}} \exp\left(-\sqrt{\frac{2}{\sigma^2}}\, x\right) = 2\left(-\frac{1}{2}\exp\left(-\sqrt{\frac{2}{\sigma^2}}\, x\right)\right)\Bigg|_A^\infty = \exp\left(-\sqrt{\frac{2}{\sigma^2}}\, A\right)$$

not $|x|$

$\rightarrow P_{clip} = e^{-\sqrt{\frac{2}{\sigma^2}}A} \leq 0.01 \longrightarrow -\sqrt{\frac{2}{\sigma^2}}\,A \leq \ln 0.01$

$\rightarrow A \geq \sqrt{\frac{\sigma^2}{2}}\,\ln\left(\frac{1}{0.01}\right)$

Let $\boxed{\sigma^2 = 1}$ ($\propto$ speech power)

As $\sigma^2 \nearrow$ A must increase

$A \geq \frac{1}{\sqrt{2}}\ln\left(\frac{1}{0.01}\right)$

Q. How to sample from Laplace distrib?

Generate uniform r.v.s $U[0,1]$

pytorch. random( )

rand(.)

randn(.)

?

$U_1, \ldots, U_N$

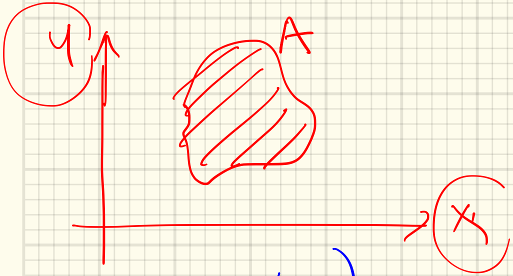
$\downarrow$

$X_1, \ldots X_n$

$X \sim F_X^{-1}(u)$

inverse of the Laplacian CDF

Multiple r.v.s → joint pdfs : $p_{X,Y}(x,y)$

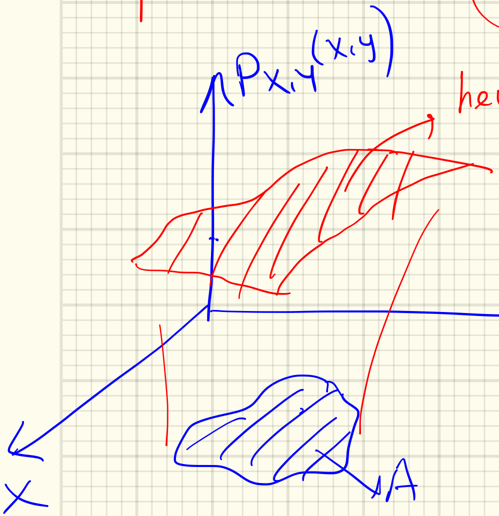


$$P((x,y)\in A) \triangleq \iint_A p_{X,Y}(x,y)\,dx\,dy$$

Properties of $p_{X,Y}(x,y)$

1) $p_{X,Y}(x,y) \geq 0$

2) $\displaystyle\iint_{-\infty}^{\infty} p_{X,Y}(x,y)\,dx\,dy = 1$

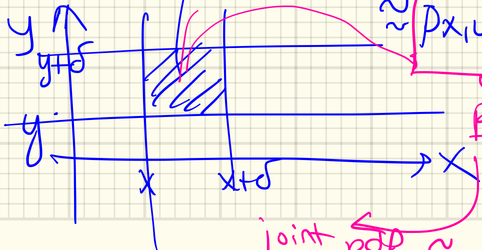


height fn. $p_{X,Y}(x,y)$ = probability per unit area

δ small: $P_{X,Y}(x \le X \le x+\delta,\ y \le Y \le y+\delta) \approx \boxed{p_{X,Y}(x,y)} \cdot \boxed{\delta^2}$

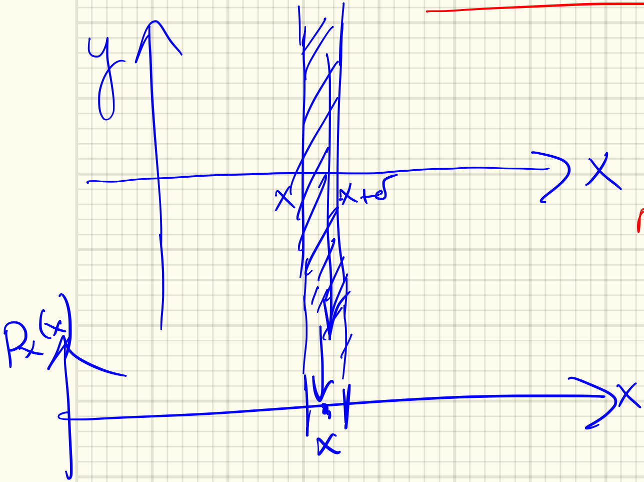
$\boxed{\delta^2}$: Area

$\boxed{p_{X,Y}(x,y)}$: $\dfrac{\text{Probability}}{\text{Area}}$



joint pdf ≈ $\dfrac{\text{Probability}}{\text{Area}}$

→ From the Joint pdf to Marginal pdfs :



$$\underbrace{P_X(x)}_{\text{marginal pdf's}} = \int_{-\infty}^{\infty} P_{X,Y}(x,y)\, dy \quad (y)$$

$$\underbrace{P_Y(y)} = \int_{-\infty}^{\infty} P_{X,Y}(x,y)\, dx \quad (x)$$

* X & Y are <u>independent</u> if $P_{X,Y}(x,y) = \underbrace{P_X(x) \cdot P_Y(y)}_{\text{we factor out the marginals from the joint.}} \quad \forall x,y$

(same as in discrete r.v.s)

* <u>Expectation</u>: $E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\, P_{X,Y}(x,y)\, dx\, dy$

$$\sum \sum$$

Ex: 2 r.v.s X & Y ; their joint pdf :

$$p_{X,Y}(x,y) = \begin{cases} \boxed{k} \cdot (1-|2x-1|) \cdot (1-|2y-1|), & 0 \leq x \leq 1 \\ & 0 \leq y \leq 1 \\ 0, & \text{o/w}. \end{cases}$$

Q. What is k for a valid pdf $p_{X,Y}(x,y)$?

$$\int_0^1 \int_0^1 k \cdot \underbrace{(1-|2x-1|)}\underbrace{(1-|2y-1|)}\, dx\, dy = 1 .$$

$$p_X(x) = \begin{cases} 2(1-|2x-1|), & 0 \leq x \leq 1 \\ 0, & \text{o/w} \end{cases}$$

Q. Are X & Y independent ? $p_{X,Y} \stackrel{?}{=} p_X \cdot p_Y$

Yes ✓

$$\underbrace{k \int_0^1 (1-|2x-1|)\, dx}\ \cdot\ \underbrace{\int_0^1 |1-|2y-1||\, dy} = 1 \rightarrow k = 4.$$

Q. $\underline{P(A)} = ?$ $A = \left\{ \frac{1}{4} \leq X \leq \frac{3}{4},\ \frac{1}{4} \leq Y \leq \frac{3}{4} \right\}$

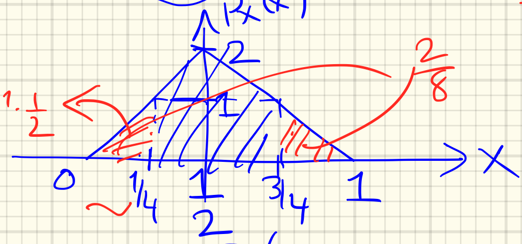
$$P(A) = \int_{1/4}^{3/4}\int_{1/4}^{3/4} p_{x,y}(x,y)\,dx\,dy = \underbrace{\int_{1/4}^{3/4} 2(1-|2x-1|)\,dx}_{6/8} \cdot \underbrace{\int_{1/4}^{3/4} 2\cdot(1-|2y-1|)\,dy}_{6/8}$$

$$= 1 - \left(\frac{2}{8}\right) = \frac{6}{8}.$$

$$P(A) = \left(\frac{6}{8}\right)\left(\frac{6}{8}\right) = \frac{9}{16}$$

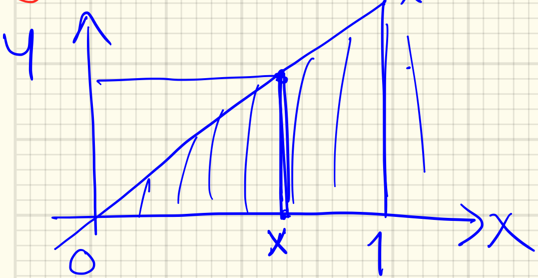
$$\frac{1}{8} = \frac{1}{4}\cdot 1 \cdot \frac{1}{2}$$

$$\frac{2}{8}$$



$$\frac{1}{4} \le X \le \frac{3}{4}$$

$\boxed{Q}$. $P(Y \le X)$



$$= \int_{\substack{0\\(x)}}^{1}\int_{\substack{0\\(y)}}^{x} p_{x,y}(x,y)\,dx\,dy$$

$$= \int_0^1\int_0^x 4\,(1-|2x-1|)(1-|2y-1|)\,dx\,dy.$$

$$= 4\int_0^1 \Big(\quad dx \int_0^x (\quad) \Big)\,dy :$$ tedious integral.

Ex: Needle of $\boxed{\text{Buffon}}$:

2 parallel sticks ; distance d apart. Needle length L

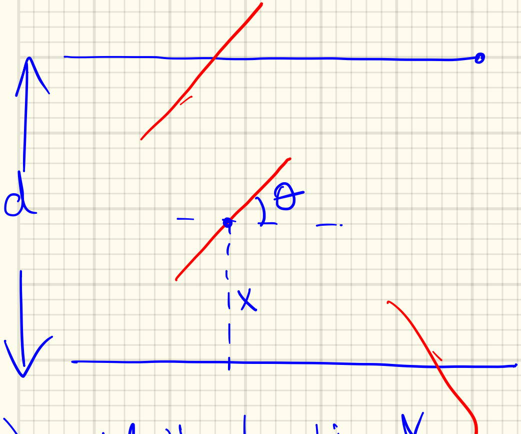
P(needle intersects one of the lines) = ?

$L < d$.

2 possibilities

i) needle does not intersect the stick

ii) needle intersects one of the lines.



Recall Our probabilistic modeling procedure:

1) Set up your sample space Ω
2) Describe a probability law on Ω
3) Identify the event of interest in Ω
4) Calculate its probability.

1) model the location & orientation of the needle

$X \in [0, \frac{d}{2}]$

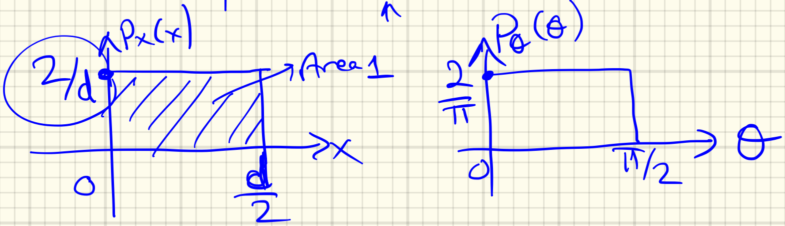
$\Theta \in [0, \frac{\pi}{2}]$

2 r.v.s w/ their space.

2) Prob. model

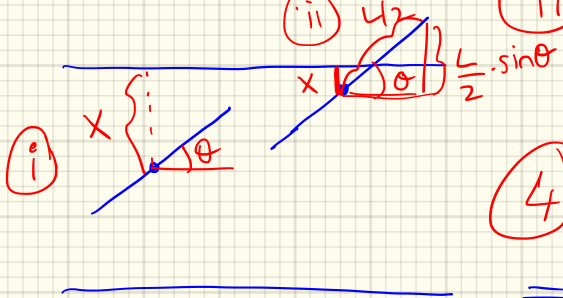
X, Θ : Uniform & independent : an intuitive & simple model

$$P_{X,\theta}(x,\theta) = P_x(x) \cdot P_\theta(\theta), \quad 0 \le x \le \frac{d}{2}, \quad 0 \le \theta \le \frac{\pi}{2}$$



$$P_{X,\theta}(x,\theta) = \left(\frac{2}{d}\right) \cdot \left(\frac{2}{\pi}\right) = \frac{4}{\pi d}, \quad \underbrace{0 \le x \le \frac{d}{2}}, \quad \underbrace{0 \le \theta \le \frac{\pi}{2}}$$

3 2 ways the needle can fall:



ii $P\left(X \le \frac{L}{2} \cdot \sin\theta\right)$

$= A$: we identified the event of interest.

4 $P(A) = P\left(X \le \frac{L}{2} \cdot \sin\theta\right)$

$$= \iint_{\{X \le \frac{L}{2}\sin\theta\}} \underbrace{P_x(x) \cdot P_\theta(\theta)}\, dx\, d\theta$$

4

$$P(A) = \frac{4}{\pi d} \int_0^{\pi/2} \left(\int_0^{\frac{L}{2}\sin\theta} \cdot\, dx \right) d\theta = \frac{4}{\pi d} \int_0^{\pi/2} \frac{L}{2} \sin\theta \, d\theta$$

$$= \frac{4}{\pi d} \cdot \frac{L}{2} \underbrace{\left(-\cos\theta \Big|_0^{\pi/2} \right)}_{1}$$

$$\underbrace{P(\text{needle intersecting a stick})}_{A} = \frac{2L}{\pi d}$$

Historical note: Threw 10000 needles & calculated $P(A)$,

given L & d ✓ : $\pi = \dfrac{2L}{P(A) \cdot d}$

This problem soln was used to calculate an approx. value for π.

$\approx$

→ Note: Instead use **Monte Carlo method** to evaluate integrals. (we'll talk later)