24.10.2022

YZV 231E

Probability Theory & Stats

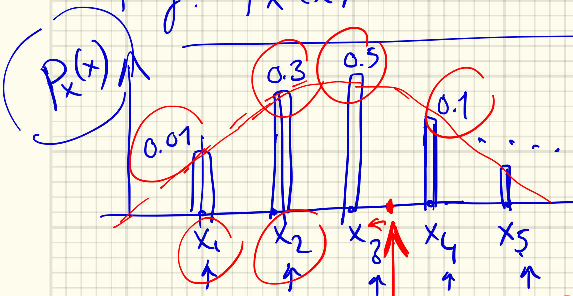
Week 6

Gü.

Recap : r.v.s associate numerical values to outcomes of random experiments.

r.v. $X : \Omega \xrightarrow{X(\cdot)} \Omega^X \subseteq \mathbb{R} \xrightarrow{p_X(x)} [0,1]$

pmf: $p_X(x) : \Omega^X \longrightarrow [0,1]$



pmf properties :

1) $p_X(x) \geq 0$

2) $\sum_{x_i} p_X(x_i) = 1$

Complete characterization :

① Expected Value : (mean)

cm. : $E[X] = \sum_{x_i} x_i \, p_X(x_i) = \sum_k k \, p_X[k]$: 1st moment of the pmf.

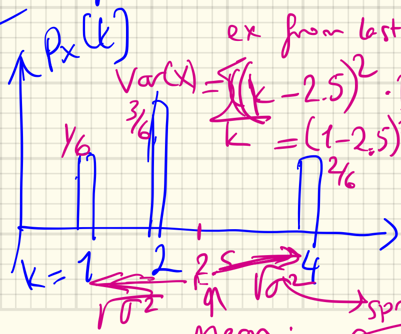
② $E[X^2] = \sum_{x_i} x_i^2 \, p_X(x_i)$: 2nd moment of the pmf.

$= \sum_k k^2 \, p_X[k]$

Variance

$Var(X) = E[(X - \underbrace{E[X]}_{\mu})^2]$

$Var(X) = \sum_k (k - E[X])^2 p_X[k]$



ex from last time

$Var(X) = \sum_k (k - 2.5)^2 \cdot p_X[k]$

$= (1-2.5)^2 \cdot \frac{1}{6} + (2-2.5)^2 \cdot \frac{3}{6} + (4-2.5)^2 \cdot \frac{2}{6}$.

spread/distance from the mean.

* If X is an r.v. $\longrightarrow$ $g(X)$ is a random variable too ; $g(\cdot)$: a function.

$$E[g(X)] = \sum_{x_i} g(x_i)\, p_X(x_i)$$

$$E[X] = \sum_{x_i} x_i \cdot p_X(x_i)$$

When $\underbrace{g(x) = (x-\mu_x)^2}$ $\quad \underbrace{E[(X-\mu_x)^2]}_{Var(X)} = \sum_i \underbrace{(x_i-\mu_x)^2}_{\geq 0}\, \underbrace{p_X(x_i)}_{\geq 0}$

≥ 0.

Properties of $Var(X)$:

1) $Var(X) \geq 0$

2) $Var(\underbrace{\alpha X + \beta}_{\text{Affine function}}) = \sum_i \Big(\underbrace{\alpha x_i + \beta}_{y} - \underbrace{E[\alpha X + \beta]}_{\mu_y}\Big)^2 \cdot p_X(x_i) \Longrightarrow$

$$\begin{bmatrix} x & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

$$\underline{X} \cdot \underline{\alpha}$$

$$E[\alpha X + \beta] = \alpha \mu_x + \beta$$

$$E[\alpha X] + \underbrace{E[\beta]}_{\beta} =$$

$$\alpha E[X] + \beta$$

exercise : derive this.

$$\rightarrow Var(\alpha X + \beta) = \sum_i \left(\alpha x_i + \beta - (\alpha \mu_x + \beta)\right)^2 p_x(x_i)$$

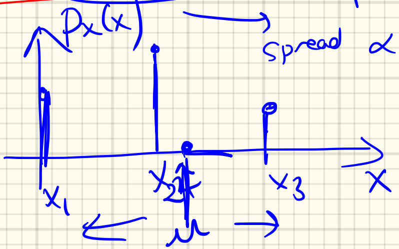
$$= \sum_i \alpha^2 (x_i - \mu_x)^2 \cdot p_x(x_i)$$

$$\boxed{Var(\alpha X + \beta) = \alpha^2 \cdot Var(X)}$$

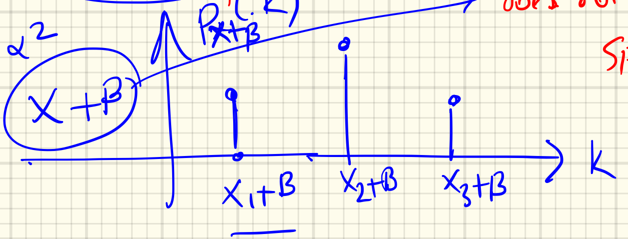
spread $\propto \alpha^2$

★ adding an offset to the r.v. does not affect the spread (variance)

X :



X + β



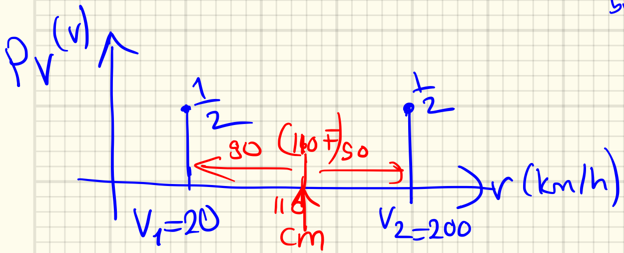
— σ_x : Standard deviation ; same units as the r.v. itself

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

★ But your variance scales by α^2 when you scale your r.v. X by α

Q. $E[g(x)] \overset{?}{=} g(E[X])$

Ex: Traverse a $d = 1000$ km distance at a constant speed V, but V is random in a way s.t. you flip a coin to choose to go by train or bike.



$E[V] = 110$ km/h $= \frac{1}{2} \cdot 200 + \frac{1}{2} \cdot 20$

$Var(V) = \frac{1}{2}(20-110)^2 + \frac{1}{2} \cdot (200-110)^2$

$\sigma^2 = Var(V) = (90)^2 = 8100$

std. $\sigma = 90$ tells us how spread our distrib. is from the mean.

Average Time vs Average Speed:

Q. How much time will it take you get there on average?

Time $T = \frac{d}{V} = \boxed{g(V) = \frac{1000}{V}}$: T is also an r.v.

$$E[T] = E[T(V)] = \sum_{i=1,2} T(v_i)\, Pr(v_i) = \frac{1}{2} \cdot \overset{50 \text{ hours}}{\frac{1000}{20}} + \frac{1}{2} \cdot \overset{5 \text{ hours}}{\frac{1000}{200}}$$

$= 27.5$ hours

Is $E[\underbrace{T(v)}_{g}] \stackrel{?}{=} T(E[v])$ $\qquad \boxed{T(v) = \frac{d}{v}} \propto \frac{1}{v}$

$\underbrace{\qquad}_{27.5} \qquad \underbrace{\overset{g}{T}(110)} = \frac{1000}{110} \approx 9.\ldots \approx 10.$

$27.5 \neq 10.$

$\Rightarrow \quad E[g(x)] \neq g(E[x])$: do not interchange expected values & functions.

Here $g(\cdot)$ (or $T(\cdot)$) was a nonlinear fn. ✗

Exceptions to this rule are linear functions. & affine fns

$E[\underbrace{\alpha X}_{g(x) = \alpha X}] = \alpha E[X]$

$E[\underbrace{\alpha X + \beta}_{g(x) = \alpha X + \beta}] = \alpha E[X] + \beta$: affine functions.

$$\mathbb{E}[\overbrace{T \cdot V}^{\text{distance}}] \overset{?}{=} \mathbb{E}[T] \cdot \mathbb{E}[V]$$

$$\mathbb{E}(1000) \;\cancel{=}\; \underbrace{27.5}_{} \times \underbrace{110}_{}$$

JOINT PMFs

: We have multiple r.v.s; say 2 r.v.s: X & Y

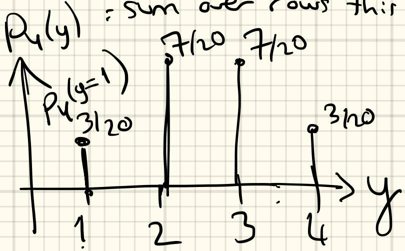
$$p_{X,Y}(x,y) = P(X=x,\ Y=y)$$

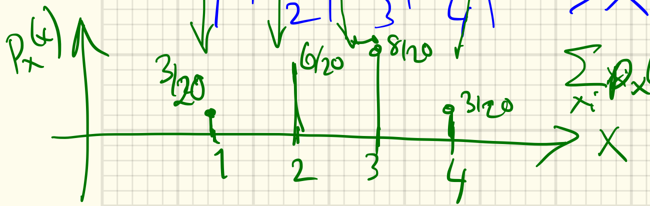
$p_{X,Y}(x,y)$

$y \backslash x$	1	2	3	4
4	0	1/20	1/20	1/20
3	1/20	2/20	3/20	1/20
2	1/20	2/20	3/20	1/20
1	1/20	1/20	1/20	0

$$p_{X,Y}(x=3,\ y=2) = 3/20$$

$p_Y(y)$: sum over rows this time





$$\sum_{x_i} p_X(x_i) = 1$$

Properties of joint pmf:

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

2) $\sum_{x,y} p_{X,Y}(x,y) = 1$

3) Marginalization: joint pmf → marginal pmfs.

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

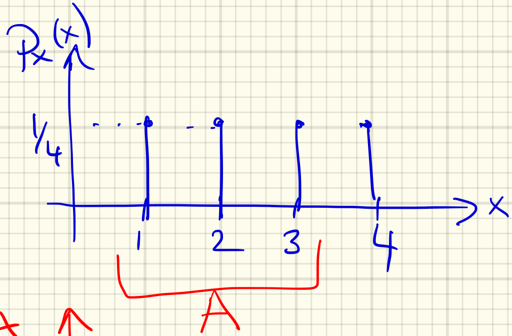
$$p_Y(y) = \sum_x p_{X,Y}(x,y)$$

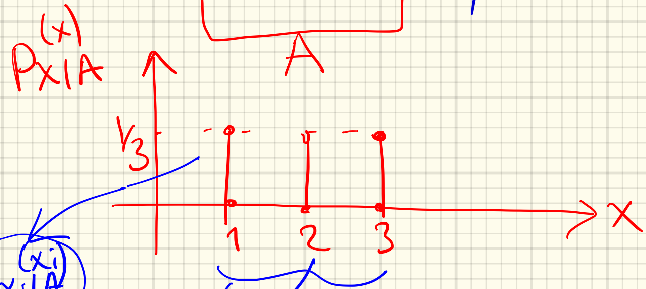
Conditional pmf:

$$P_X(X = x \mid A)$$

$$A = \{X \leq 3\}$$

$P(A) = 3/4$





conditional pmf: is a standard pmf

$$\sum_x P_{X|A} = 1$$

$$E[X \mid A] = \sum_{x_i} x_i \, P_{X_i|A}(x_i)$$

$$= 1 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3} + 3 \cdot \frac{1}{3} = 2.$$

Generalize: $E[g(X) \mid A] = \sum_x g(x) \, P_{X|A}(x)$

Recall $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)} \Longrightarrow$

$$P_{X|Y}(x|y) = \underbrace{P_{X|Y}(X=x \mid Y=y)}_{} = \boxed{\frac{P_{X,Y}(x,y)}{P_Y(y)}}$$

eg. $P_{X|Y}(X \mid \underbrace{y=3}_{\text{we fixed } y.})$ → my new universe

$$\underbrace{\boxed{P_{X|Y}(X \mid Y=\underset{\text{fixed}}{y})}}$$ is a function of x.

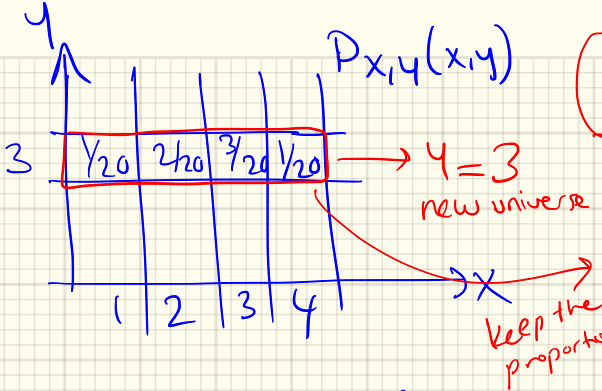
→ like a pmf its.

Properties: 1) $P_{X|Y}(X \mid Y=y) \geq 0$

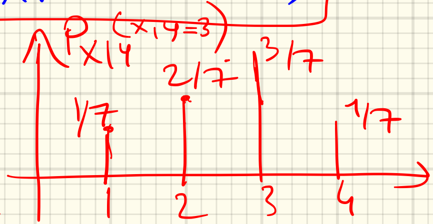
2) $\sum_{x} P_{X|Y}(X \mid Y=y) = 1.$

$$\boxed{\begin{aligned} P_{X,Y}(x,y) &= P_{X|Y}(x|y) \cdot P_Y(y) \\ P_{X,Y}(x,y) &= P_{Y|X}(y|x)\, P_X(x) \end{aligned}}$$

relate joint pmf & the conditional pmfs.



$\boxed{p_{X|Y}(X|Y=3)} \rightarrow$



Q. What happens if we have 3 r.v.s?

(It's clear how to generalize to more than 3 r.v.s)

$p_{X,Y,Z}(x,y,z)$: joint pmf.

marginal pmfs

$$p_X(x) = \sum_{y,z} p_{X,Y,Z}(x,y,z)$$

$$p_Y(y) = \sum_{x,z} p_{X,Y,Z}(x,y,z)$$

$$p_Z(z) = \sum_{x,y} p_{X,Y,Z}(x,y,z)$$

Marginalization

Recall

Multiplication Rule: $p(A_1 \cap A_2 \ldots A_n)$

$$= p(A_1)\, p(A_2 | A_1)\, p(A_3 | A_2 \cap A_1) \ldots p\left(A_n \,\middle|\, \bigcap_{i=1}^{n-1} A_i\right)$$

we Use this also for pmfs:

$$P_{X,Y,Z}(x,y,z) = P_X(x)\, P_{Y|X}(y|x)\, P_{Z|X,Y}(z|x,y)$$

Independence: 3 r.v.s X, Y, Z are independent if:

$$\boxed{P_{X,Y,Z}(x,y,z) = P_X(x) \cdot P_Y(y) \cdot P_Z(z) \quad \forall x,y,z}$$

∗ 3 (or n) r.v.s are independent if their joint pmfs factor out into individual (marginal) pmfs.

∗ Independence for conditional pmfs translates to:

$$P_{X|Y}(x|y) = \frac{P_{X,Y}(x,y)}{P_Y(y)} \overset{\text{use independence}}{=} \frac{P_X(x) \cdot P_Y(y)}{P_Y(y)} = P_X(x)$$

$p_Y(y) \neq 0 \;\; \forall y$

→ If X & Y are independent r.v.s $\quad P_{X|Y}(x|y) = \underbrace{P_X(x)}_{\text{marginal}}$.

(if $p_Y(y) > 0$)

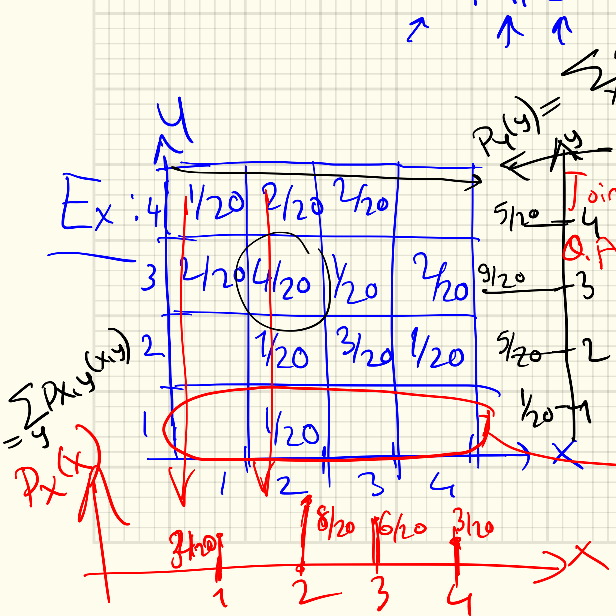
Recall $P(A|B) = P(A)$ if A & B are independent events.

→ If multiple r.v.s (ep. 3) are independent:

$$P_{X|Y,Z}(x|y,z) = P_X(x) \quad \text{: marginal of } x.$$

$$= \frac{P_{X,Y,Z}(x,y,z)}{P_{Y,Z}(y,z)} \quad (P_{Y,Z}(y,z) > 0)$$

Ex:



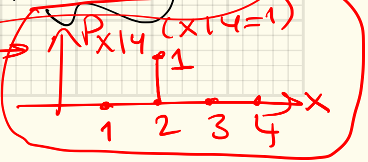
Joint pmf of X & Y are given: $P_{X,Y}(x,y)$

Q. Are X & Y independent? No!

$$P_{X,Y}(x=2, y=3) \overset{?}{=} P_X(x=2) \cdot P_Y(y=3)$$

$$\frac{4}{20} \neq \frac{8}{20} \cdot \frac{9}{20}$$

Check $y=1$



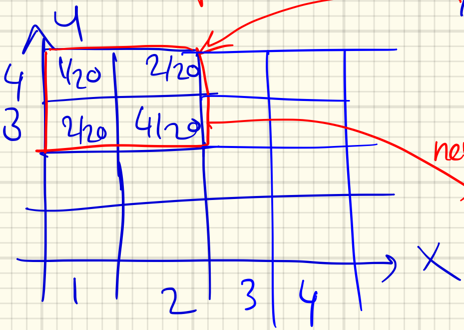
No! $P_{X|Y}(x|y) \neq P_X(x)$

Conditional Independence:

$A = \{X \le 2,\ Y \ge 3\}$

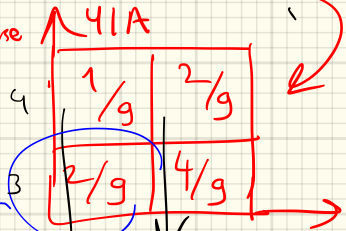
same joint pmf

from prev. page



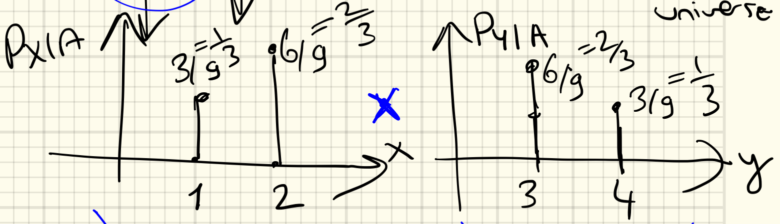
new universe

$P_{X,Y|A} = ?$



$P_{X,Y|A}$

Q. Are X & Y independent? (in this conditional universe?)



$\forall x, y$

$P_{X,Y}(x,y) = P_X(x) \cdot P_Y(y)$

$\therefore$ X|A & Y|A are independent

$\underbrace{P(X=1, Y=3 \mid A)}_{\frac{2}{9}} = \frac{1}{3} \cdot \frac{2}{3} = \frac{2}{9}$ ✓

EXPECTATIONS:

In general: $E_{X,Y}[g(X,Y)] = \sum_{x,y} \underbrace{g(x,y)}_{x,y} \, P_{X,Y}(x,y)$ (★)

In general $\boxed{E[g(X,Y)] \neq g(E[X], E[Y])}$

Recall the exception to this rule.

$E[\alpha X + \beta] = \alpha E(X) + \beta$: affine fns.

For multiple r.v.s:

$$E[X+Y+Z] = E[X] + E[Y] + E[Z]$$

→ Expectation behaves linearly.

You can derive this from (★)

exercise:

Another exception: In case of INDEPENDENT r.v.s.

$\boxed{E[\underbrace{X.Y}_{g(X,Y)}] = E[X].E[Y]}$ for independent r.v.s

$= \sum_{x,y} x.y.\underbrace{p_{X,Y}(x,y)}_{X \& Y \text{ indep r.v.s.}} = \sum_{x,y} x.y.p_X(x)\,p_Y(y) = \underbrace{\sum_x x p_X(x)}_{E[X]}.\underbrace{\sum_y y.p_Y(y)}_{E[Y]}.$

Not valid in general $E[X \cdot Y] \neq E[X] \cdot E[Y]$

What about $E[g(X) \cdot h(Y)] \overset{?}{\cancel{=}} E[g(X)] \cdot E[h(Y)]$

exception
When X & Y are independent $=$ is satisfied.

$\underbrace{g(X)}_{\text{is an r.v.}}$ & $\underbrace{h(Y)}_{\text{is an r.v.}}$ are also independent

exercise Show for independent X & Y

$$E[g(X)\, h(Y)] = E[g(X)] \cdot E[h(Y)]$$
$$= \sum_x \sum_y g(x)\, h(y)\, p_{X,Y}(x,y) \ldots$$

$\boxed{Ex}$: If $X = Y$, $\underline{Var(X+Y)} = Var(2X) = 4\,Var(X)$

Are X & Y independent?

r.v.s are extremely dependent.

consider the sum of the individual variances
$$Var(X) + Var(Y) = 2\,Var(X) \neq.$$

Ex: If $X = -Y$; $Var(\underbrace{X+Y}) = Var(0) = 0$

$: \underbrace{Var(X)} + \underbrace{Var(-Y)}_{\Downarrow \; Var(Y)} = 2Var(X)$

$\Rightarrow$ Variance of the sum r.v. $\neq$ Sum of the individual variances.

Ex: If X & Y are indep, $Z = X - 3Y$ & you are given $Var(X)$ & $Var(Y)$

$\searrow$ X & $-3Y$ are also independent

$Var(Z) = ?$

$= Var(X) + \underbrace{Var(-3Y)}$

$Var(X) + 9 \cdot Var(Y)$

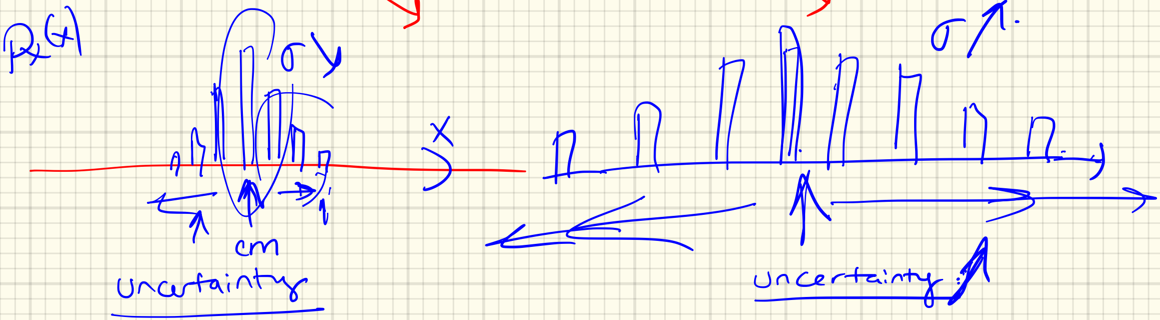
$\boxed{\text{Exercise: If } X \text{ \& } Y \text{ are independent; } \quad Var(\underbrace{X+Y}) = Var(X) + Var(Y)}$

Show this

Note: Variance as a measure of uncertainty.

Var(x) $\uparrow$ uncertainty $\uparrow$

$\downarrow$ $\downarrow$



Memoryless property: Geometric pmf :

$$\underbrace{T\ T\ T\ T\ T\ T}_{k-1}\ \underset{\uparrow}{H} \qquad p_X(k) = (1-p)^{k-1}\cdot p \quad , k \geq 1$$

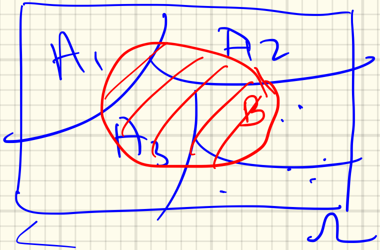
p: success prob.

$$\underbrace{p(1^{st}\text{ success at }(\underline{m+l})^{th}\text{ trial}}\ \Big|\ \underbrace{\underline{m}\text{ failures}}) = p(1^{st}\text{ success at }l^{th}\text{ trial})$$

$$= \frac{P(\overbrace{TTT\dots T}^{m+l-1}H)}{P(\underbrace{TT\dots T}_{m})} = \frac{(1-p)^{m+l-1}\cdot p}{(1-p)^{m}} = \underbrace{(1-p)^{l-1}\cdot p}.$$

Memoryless property: whatever happens in the future is independent from what happened in the past.

Total Expectation Theorem:



Partition Ω into disjoint events $\{A_i\}_{i=1}^{n}$

Recall Total Probability Law:

$$P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \cdots + P(A_n)P(B|A_n)$$

We can use the same idea for pmfs:

$$p_X(x) = P(A_1)\,p_{X|A_1}(x) + \cdots + P(A_n)\,p_{X|A_n}(x)$$

take expectation

$$\sum_i x_i\, p_X(x_i) = \sum_x x_i \Big(\cdot \Big)$$

$$E[X] = P(A_1)E[X|A_1] + \cdots + P(A_n)E[X|A_n]$$

Total Expectation Thm: useful way to calculate expectations, using divide & conquer.

Ex: $\underline{\text{Geometric r.v.}}$ $\nearrow$ 1st case : 1st toss is Heads : $A_1 = \{X_1 = 1\}$

\# coin tosses until 1st Head.

Complementary case $\longrightarrow$: $A_2 = \{X_1 > 1\}$

$$E[X] = \underbrace{P(X=1)}_{p} \cdot \underbrace{E[X \mid X=1]}_{1} + \underbrace{P(X>1)}_{(1-p)} \cdot \underbrace{E[X \mid X>1]}_{= E[X \mid X-1>0]}$$

$$= E[X-1 \mid X-1>0] + 1$$

from memoryless property $\quad = E[X] + 1.$

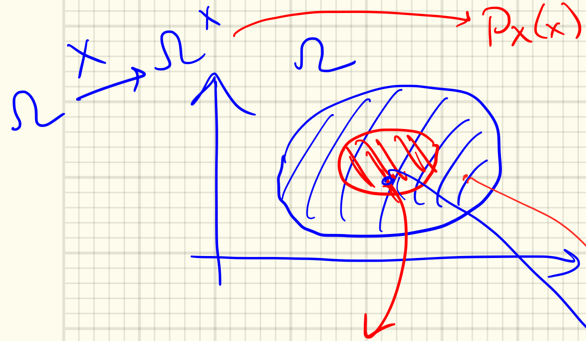
$$\Rightarrow E[X] = p + (1-p)(E[X]+1)$$

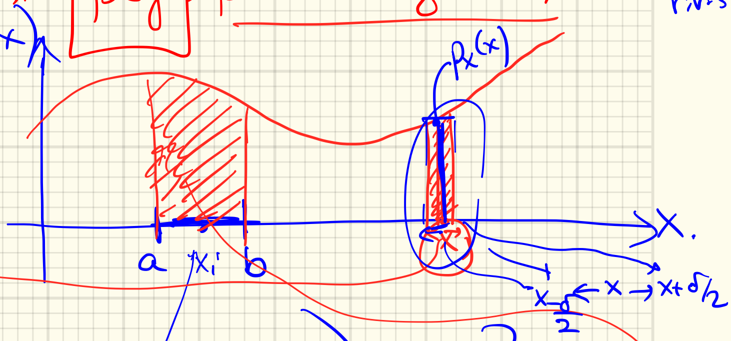
$$\rightarrow E[X] = \frac{1}{p} \quad ; \text{ geometric pmf mean.}$$

Intuitive : $p \searrow \quad E[X] \nearrow$: # flips to get success $\nearrow$!

Continuous Random Variables:

$\Omega \xrightarrow{X} \Omega^X \longrightarrow p_X(x) \to (0,1)$; $\boxed{\text{pdf}}$: Probability density fn's — characterize continuous r.v.s.





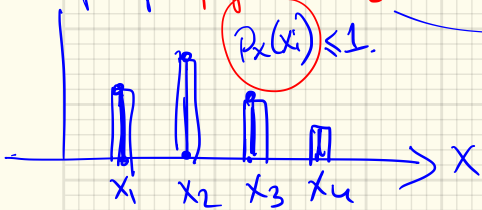
X_i's fall into this region.

$$\underbrace{P(a \le X \le b)}_{\text{event.}} = ? = \int_a^b \underbrace{p_X(x)}_{\text{pdf}}\, dx.$$

Recall pmf

pmf → probability

$p_X(x_i) \le 1.$



$$P\left(x - \frac{\delta}{2} \le X \le x + \frac{\delta}{2}\right) = \int_{x-\frac{\delta}{2}}^{x+\frac{\delta}{2}} p_X(x)\, dx \approx p_X(x) \cdot \delta$$

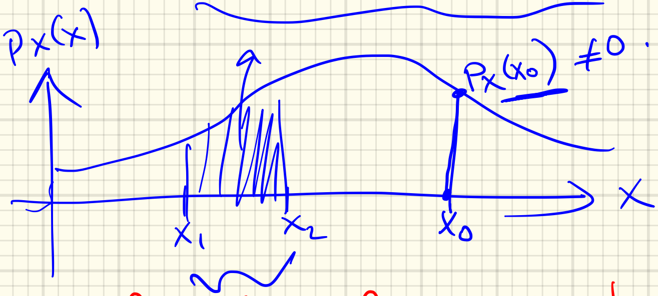
Density × Length

→ Probability / Length × Length.

δ : very small

∴ pdf (density) is not really probability.

$$\int_{x_1}^{x_2} p_x(x)\,dx = \underbrace{p(x_1 \le X \le x_2)}$$



$$\int_{x_0}^{x_0} p_x(x_0)\,dx = 0$$

Properties of pdf:

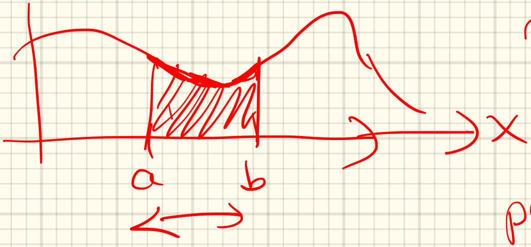
(pdf $\neq$ prob. but integrals of pdf's $\Rightarrow$ prob.

1) $p_x(x) \geq 0$

2) $\underbrace{\int_{-\infty}^{\infty} p_x(x)\,dx = 1}_{p(-\infty < X < \infty) = 1}$

Any function satisfying these 2 properties is a probability density functions.

Note: Density (pdf) is Probability per unit length.



Prob. of interval (a,b) = integral of the pdf in (a,b) interval

$p(a)=0$ ✓ or $(a,b]$

$p(b)=0$. ✓ or $[a,b]$ or $[a,b)$.

3) $$p(X\in B) = \int_B p_X(x)\,dx.$$

B: "nice" sets (measurable) → prob. measure theory.

* pdf is a complete description of a continuous r.v.

Means & Variances : Carry over the formulas w/ integrals.

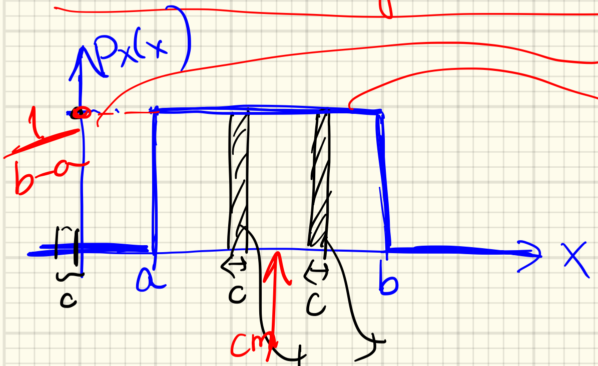
$$E[X] = \int_{-\infty}^{\infty} x \cdot p_X(x)\, dx \quad ; \quad X \text{ is a continuous r.v.}$$

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

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

Also show: $\mathrm{Var}(X) = E[X^2] - (E[X])^2$ ✓

Continuous Uniform r.v.



$$\int_a^b p_x(x)\,dx = 1$$

$$② \quad \int_a^b \frac{1}{b-a}\,dx = 1$$

$$① \quad p_x(x) \geq 0$$

Any interval of the same length in (a,b) have equal probability

Uniform pdf $\quad p_x(x) = \begin{cases} \dfrac{1}{b-a}, & a \leq x \leq b \\ 0, & o/w \end{cases}$

$$E[X] = \int_a^b x \cdot \frac{1}{b-a}\,dx = \frac{a+b}{2} = \mu_x$$

$$Var(X) = \sigma_X^2 = \int_a^b \underbrace{\left(x - \left(\frac{a+b}{2}\right)\right)^2}_{(x-\mu_X)^2} \underbrace{\boxed{\frac{1}{b-a}}}_{p_X(x)} dx$$

$$= \frac{(b-a)^2}{2}$$

$$\sigma_X = \frac{b-a}{\sqrt{2}}$$: std. deviation.

$\sigma_X \propto \underbrace{(b-a)}$: spread of the r.v.