25.10.2021

YZV 231E

Probability Theory & Stats

Gü.

Recap: Bayes

cause $\qquad$ result / effect

$B \longrightarrow A \qquad P(A|B)$

$B \longleftarrow A. \qquad \underbrace{P(B|A)}_{\text{posterior.}}$

Ex: $B = \{\text{person is sick}\}$ $\qquad p(B|A)$: posterior prob. that the person is sick given that his/her test is true.

(ex 4.26) (Kay)

$A = \{\text{test is positive}\}$.

Bayes Thm: $\displaystyle p(B|A) = \frac{\overbrace{P(A|B)}^{\text{likelihood}}\, p(B)}{\cancel{p(A)}\; p(A|B)\,p(B) + p(A|B^c)\,p(B^c)}$ → prior prob. $\quad$ → use Total prob. Law.

$p(B|A)\,p(A) = p(A|B)\,P(B)$

$P(B \cap A) = p(A \cap B)$

Given, $\boxed{0.001\,\%}$ of the general population has a certain disease.

It is known that $\quad P(A|B) = 0.99 \quad (TP)$

$P(A|B^c) = 0.2 \quad (FP)$ $\quad \Rightarrow p(B|A) \Rightarrow$

$p(B) = 10^{-5}$

$$P(B|A) = \frac{P(A|B)\,P(B)}{P(A|B)P(B) + P(A|B^c)\,P(B^c)}$$

$$= \frac{(0.99)\,(10^{-5})}{0.99\,(10^{-5}) + 0.2\,(0.99999)} = \underbrace{4.95 \cdot 10^{-5}}$$

quite a low prob. → we can reject the hypothesis

→ For instance, set prior prob. $\underline{P(B) = 0.5}$;

compare!

Recalculate $P(B|A) = \frac{0.99\,(0.5)}{0.99\,(0.5) + 0.2\,(0.5)} = \underline{0.83}$!

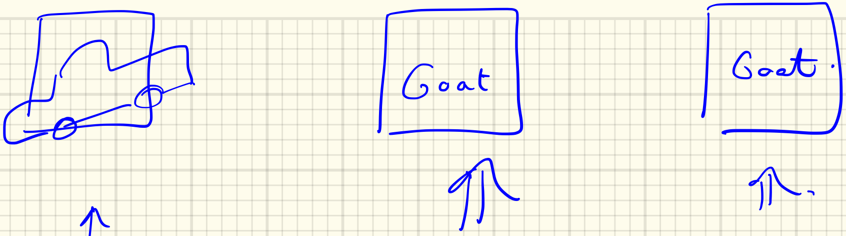
Think about the point of this example &

note the subtle part in the prior prob. assumption.

Check Ex 4.23 "Monty Hall" problem →

EX 4.23

Monty Hall.



$\frac{1}{3} \longrightarrow \frac{2}{3}$ switching your initial door choice.

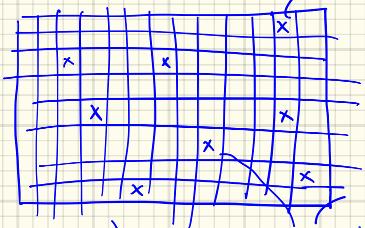
↳ sticking to your initial choice.

Chapter 4 : Cluster Recognition

Cluster = Gang

A = {observed crime data}

B = {cluster (gang) exists}

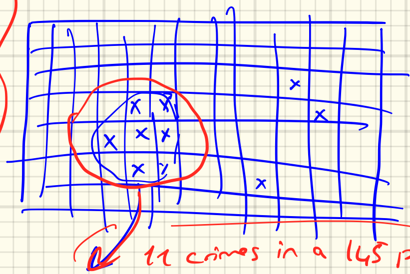


1 ; crime is present.

$P_{\text{cluster}}^{\text{no}} = 0.01$ observed

$P_{\text{cluster}} = 0.1$

from the data.



11 crimes in a 145 pixel area.

$P(B|A) = ?$

↑ hypothesis

$$\frac{P(B^c|A)}{P(B|A)} = \text{odds ratio}$$

$$P(A|B^c) = P(k=11 \mid \text{no cluster exists})$$

$$P(A|B) = P(\underbrace{k=11}_{\text{in 145 pixels}} \mid \text{cluster exists}) = \binom{145}{11} p_c^{11} (1-p_c)^{145-11}$$

You need $\Rightarrow \underbrace{P(B)}_{=10^{-6}}$: prior assumed.

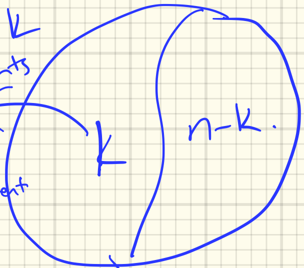
Multinomial Coefficient

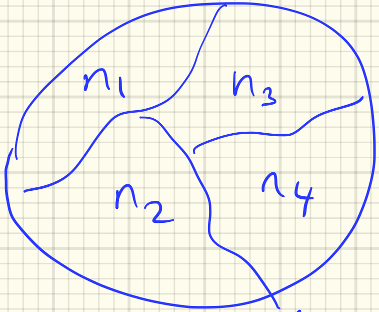
Recall Binomial:

Partitions. → Now you have:

$$\sum_{i=1} n_i = n \quad \text{\# partitions}$$

Choose k elements out of n element set





ex: $n_1 + n_2 + n_3 + n_4 = n$

partitioning into 4 sets w/ n_1 people, n_2 people, n_3 ", n_4 people — Total of n people

$\hookrightarrow \binom{n}{k}$ Binomial coeff. → multinomial coef

Q: In how many ways can we do this partition?

$$\binom{n}{n_1\ n_2\ n_3\ n_4} = \frac{n!}{n_1!\,n_2!\,n_3!\,n_4!}$$

(see Ex 1.33 Bertsekas) 52-card deck → deal to 4 players.

$P(\underbrace{\text{each gets an ace}}_{\text{Event } A}) = ?$ 13 cards/player

$$\frac{4!}{1!\,1!\,1!\,1!}$$

prob. way of distributing 1 ace to each

also count how many ways remaining 48 cards are distributed to the 4 people:

$$\frac{48!}{12!\,12!\,12!\,12!}$$

Unordered.

$$\rightarrow P(A) = \frac{|A|}{|\Omega|} = \frac{4!\,48!/(12!)^4}{52!/(13!)^4}$$

Exercise: Think about the ordered dealing!

Multiple (Independent) Experiments

$A = \bigcap A_i$

When subexperiments are independent,

$$\rightarrow P(A) = p(A_1) \cdot p(A_2) \cdot p(A_3) \cdots P(A_n)$$

When we don't have independence:

$$\Rightarrow p(A) = P(A_1) P(A_2 | A_1) P(A_3 | A_2 \cap A_1) \cdots P\left(A_n \,\Big|\, \bigcap_{i=1}^{n-1} A_i\right)$$

Special Dependence Case :

Ex: Dependent Bernoulli Trial : Say 2 coins :

1 fair $(p = 0.5)$; 1 unfair coin $(p = 0.25)$

Rule of the experiment : Choose at random 1 coin :

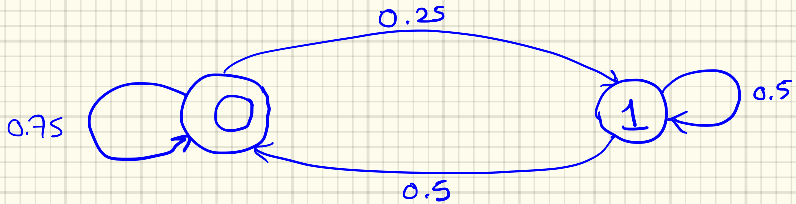
Get a tail $\Rightarrow$ switch to unfair coin

Get a head $\Rightarrow$ switch to fair coin.

Event : Getting 10 Tails in succession. $\rightarrow A = \{\underbrace{0, 0, \ldots, 0}_{10}\}$

$$\rightarrow P(\underbrace{\text{tail on the } i^{th} \text{ toss}}_{\text{Tail} \equiv 0} \mid \text{tail on the } (i-1)^{st} \text{ toss}) = P(0|0)$$

We can draw a diagram



Markov State Probability Diagram.

<u>Def:</u> This type of <u>Bernoulli sequence</u>, where the probabilities for trial i depend <u>only</u> on the outcome of the previous trial, is called a

Markov Sequence.

$$P(A_i | A_{i-1}, A_{i-1}, \dots, A_1) = P(A_i | A_{i-1}) \rightarrow$$ due Markov sequence property

now $P(A)$ becomes $= P(A_1)\, P(A_2|A_1)\, P(A_3|A_2)\, P(A_4|A_3) \cdots P(A_n|A_{n-1})$

"TTT ... T"

$A_i = \{0\}$; $P(A) = \underbrace{P(A_1)}\, \prod_{i=2}^{10} \underbrace{P(A_i|A_{i-1})}$ → $P(A_i|A_{i-1}) = P(0|0)$

$\underline{A_1 = \{0\}}$

we need to calculate these two prob.

$= 0.75$

$i = 2, \dots, 10$

$\Rightarrow P(A_1) \Longrightarrow P(\text{Tail} \mid \text{Fair})\,P(\text{Fair}) + P(\text{Tail} \mid \text{Unfair})\,P(\text{Unfair})$

this is total prob. law.

Recall

$$P(A_1) = 0.5 \cdot 0.5 + 0.75 \cdot 0.5 = \frac{5}{8}$$

$$\Rightarrow P(A) = P(A_1) \prod_{i=2}^{10} P(A_i \mid A_{i-1}) = \frac{5}{8}(0.75)^9 = \boxed{0.0469}$$

much larger.

Compare this prob. to independent & fair coins:

$$P(A) = \left(\frac{1}{2}\right)^{10} = 0.00097$$

$\Rightarrow$ This example is a simple case of a Markov chain.

Random Variables (r.v.)

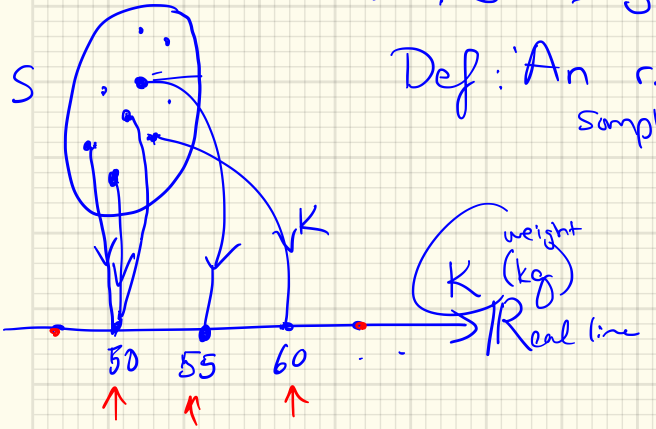
(Sample S, Events, Axioms of Prob.) ✓

↳ R.v's are derived.

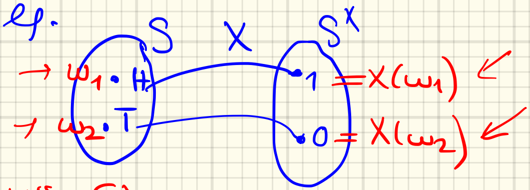
$X : S \to S^X$ ($\mathbb{R}$) ; X is a fn.

Def: An r.v. is a mapping (function) from the sample space S to a subset of the real line $\mathbb{R} = \{x : x \in (-\infty, \infty)\}$

S^X.



e.g.

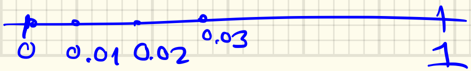


Discrete r.v. vs Continuous r.v.

finite, countably # ∞ of values

$S^X = \{0 : 0.01 : 1\}$

$|S^X| = 101$



uncountably ∞.

$S^X = [0, 1]$

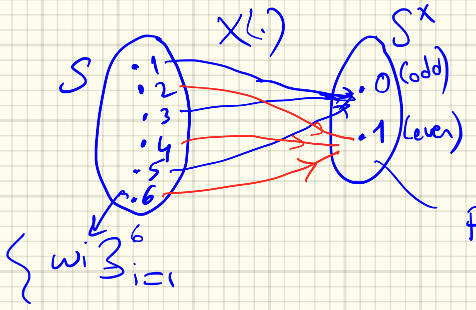
$|S^X| = \infty$.

Note: $X(\cdot)$ could be one-to-one or many-to-one mapping

e.g. Dice roll : r.v. X.

→

Ex: Random experiment: A Fair Dice Roll : R.v. X → even X odd.



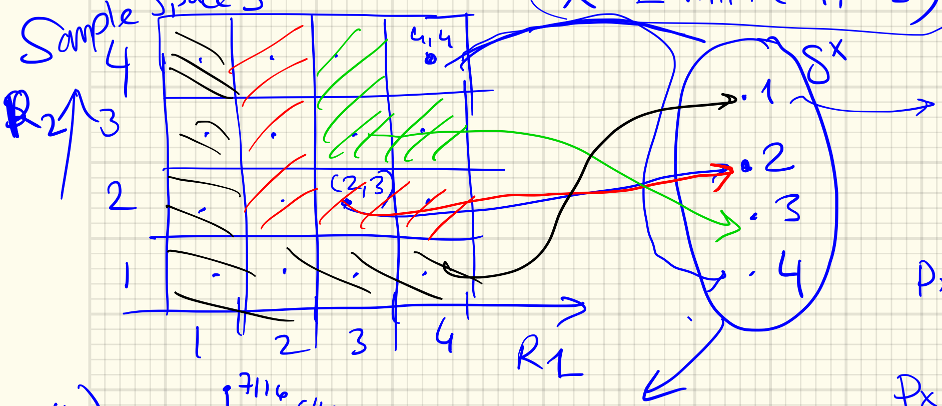
$$X(\omega_i) = \begin{cases} 0, & i = 1, 3, 5 \\ 1, & i = 2, 4, 6 \end{cases}$$

$$P[X(\omega_i) = x] = ?$$

Ex: Two independent rolls of a fair tetrahedral die (die w/ 4 faces)

$$X = \min(R_1, R_2)$$

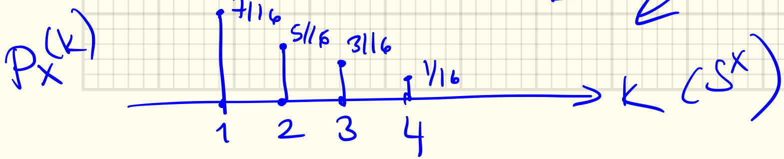
r.v. maps outcomes to numerical values.



Calculate $P_X(1)$, $P_X(2)$, $P_X(3)$, $P_X(4)$

$P_X(2) = \frac{5}{16}$, $P_X(3) = \frac{3}{16}$

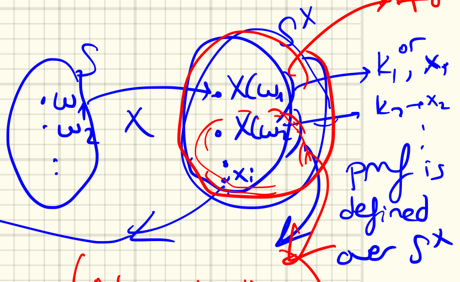
$P_X(1) = \frac{7}{16}$, $P_X(4) = \frac{1}{16}$



Def: Probability Mass Fn (pmf) $p_X[x] = P_X[X(\omega)]$

little p: pmf ; capital X : r.v.

$\Rightarrow$ $p_X[\cdot] : S^X \to [0,1]$



Properties: (1) $0 \le p_X[k] \le 1$

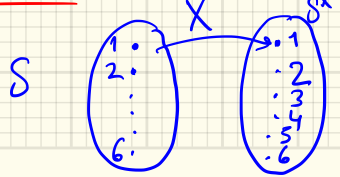
(2) $\sum_{i=1}^{M} p_X[i] = 1$; $\sum_{i=1}^{\infty} p_X[i] = 1$ (Normalization property)

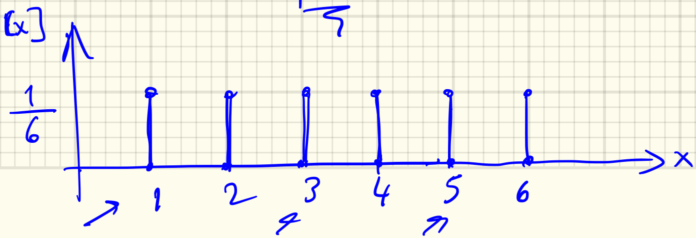
$|S^X| = M$

(3) $p_X(x \in A) = \sum_{\{i : x_i \in A\}} p_X[\underbrace{X(\omega_i)}_{x_i}]$

event A defined in S^X

Ex: Discrete Uniform R.V. eg. 1 fair die roll.

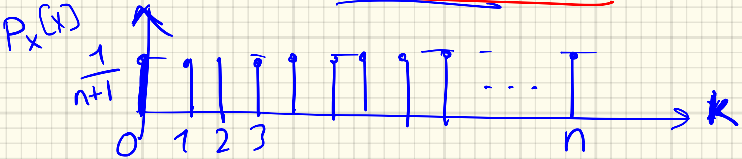




Important pmfs

1) Discrete Uniform r.v.

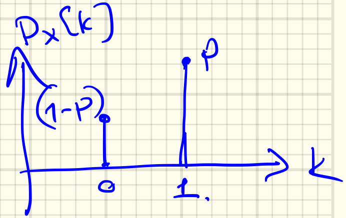
$X(\cdot) \to \{0, 1, \ldots, n\}$



$$p_X[k] = \begin{cases} \frac{1}{n+1}, & k = 0, 1, \ldots, n \\ 0, & o/w. \end{cases}$$

2) Bernouilli r.v.

$$p_X[k] = \begin{cases} (1-p), & k = 0 \\ p, & k = 1 \end{cases}$$

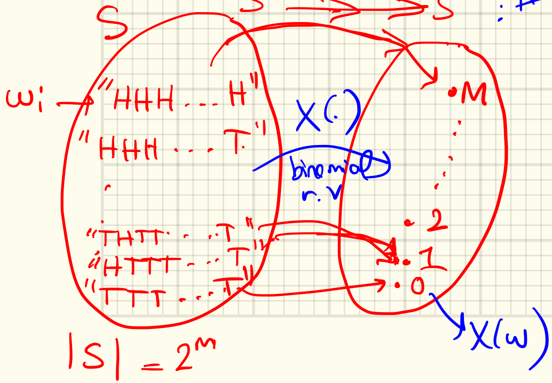


3) Binomial r.v.

k successes in M trials

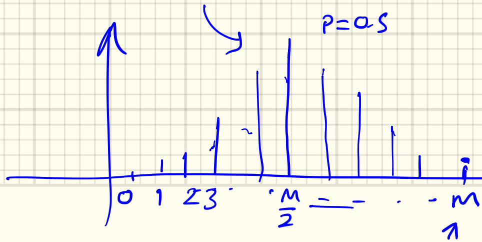
$p_X[k]$

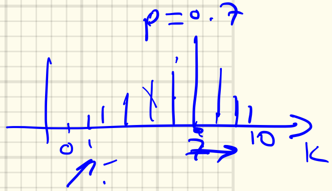
$S \xrightarrow{X} S^X$: # heads (successes) in m (Bernouilli) trials, $k = 0, 1, \ldots, M$



$|S| = 2^m$

$$p_X[k] = \binom{M}{k} p^k (1-p)^{M-k}$$





→ Sec 3.10 (Kay). Real world problem Quality Control Engineer.

→ memory chips to be shipped to computer companies. →

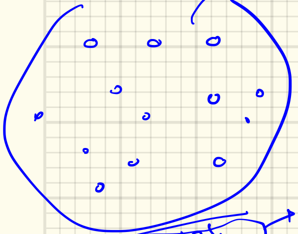
Only a proportion of the chips can be tested.

Criterion for Acceptance (to-be-shipped) of a Batch of Chips is that

At least 95 % of the chips tested are "good".

→ I will test only 100 chips.

Success: 95 or more are "good".



draw a chip; you set a probability for drawing a "good" chip: $p = 0.94$. (an estimate you come up w/ from experience)

100 chips. → a "good batch"

What is the probability law?

→ Binomial Law.

→ Need to calculate prob. of more than 95 successes out of 100 draws:

0, 1, 2, ..., 95, 96, ..., 100

$$P(k \geq 95) = P_x[k=95] + P_x[k=96] + \dots + P_x[k=100]$$

$$= \sum_{k=95}^{100} \binom{100}{k} p^k (1-p)^{100-k} \quad \text{w/ } p = 0.94$$

≈ 0.45 → This is too high!

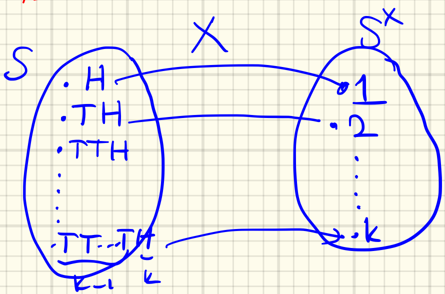
To reduce this probability →

Quality Control engineer changes the strategy to:
ship a batch only if 98 or more of the chips are "good".

$$P(k \geq 98) = \sum_{k=98}^{100} \binom{100}{k} p^k (1-p)^{(100-k)} \approx 0.05 \quad w/\ p = 0.94.$$

4) Geometric pmf:

prob of 1^{st} success at k^{th} trial.

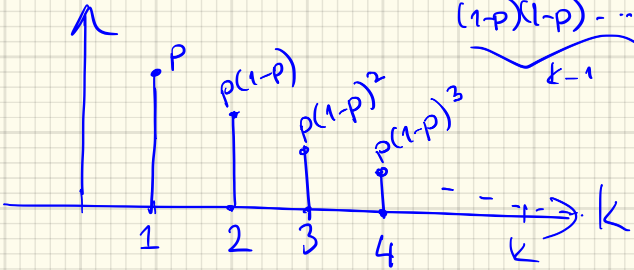


$$P_X(k) = (1-p)^{k-1} \cdot p. \qquad k = 1,$$

H.

$$\underbrace{TTTT \ldots T}_{k-1}\,.\,\underbrace{H}_{k}.$$

$$\underbrace{(1-p)(1-p)\cdots(1-p)}_{k-1} \cdot p$$



(5) Poisson R.v. used a lot in Queueing problems / Allocation of resources.

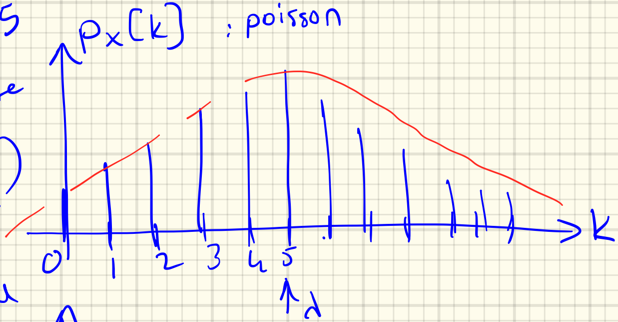
$$P_X[k] = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k \geq 0, \quad \lambda > 0 \text{ (real)}$$

X : no of arrivals (requests) → λ : Avg. # arrivals/unit time. (events)

X : no of events

Binomial pmf → Poisson pmf
($M \to \infty$, $p \to 0$)
$M.p \to \lambda$ (constant)

eg. $\lambda = 5$
(check the figure in the book Kay)



Binomial