10.10.2022

YZV 231E

Probability Theory & Stats

Week 4

Gü.

Recap: Conditional Probability $\Rightarrow$ revise my beliefs / prob. given some event B occured.

$$P(A|B) = \frac{P(A\cap B)}{P(B)}$$

Bayes Theorem: make inferences based on new partial information available.

$B^{(\text{cause})} \longrightarrow A^{(\text{effect})}$: $\overbrace{P(A|B)}^{\text{likelihood}}$

inference: $B \longleftarrow A$ Given this effect : $\underbrace{P(B|A)}_{\text{posterior}}$

how likely was the cause B?

Ex : $B = \{$person is sick from a certain disease$\}$

(ex 4.26 [Skay]) $A = \{$test is positive$\}$

Given 0.001% of the general population has the disease.

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

$P(B|A)$ = posterior prob.

$P(A|B)$: likelihood

$P(B) \Rightarrow$ prior

$P(A|B) = 0.99$ (TP)

$P(A|B^c) = 0.2$ (FP)

$\Longrightarrow P(B|A) = ?$

Use Bayes Thm: $P(B|A) = \dfrac{P(A|B)\,P(B)}{\underbrace{P(A)}_{P(A|B)\,p(B) + P(A|B^c)\,P(B^c)}}$ } using Total Prob. Law.

$$= \frac{(0.99)(10^{-5})}{0.99(10^{-5}) + 0.2(1-10^{-5})} = 4.95 \cdot 10^{-5} \approx \boxed{5 \cdot 10^{-5}}$$

$= 0.005\%$

quite small prob.

$\therefore$ we can reject the hypothesis.

compare

For instance, set prior prob. $P(B) = 0.5$

Recalculate: $p(B|A) = \dfrac{0.99(0.5)}{0.99(0.5) + 0.2(0.5)} = \boxed{0.83}$!

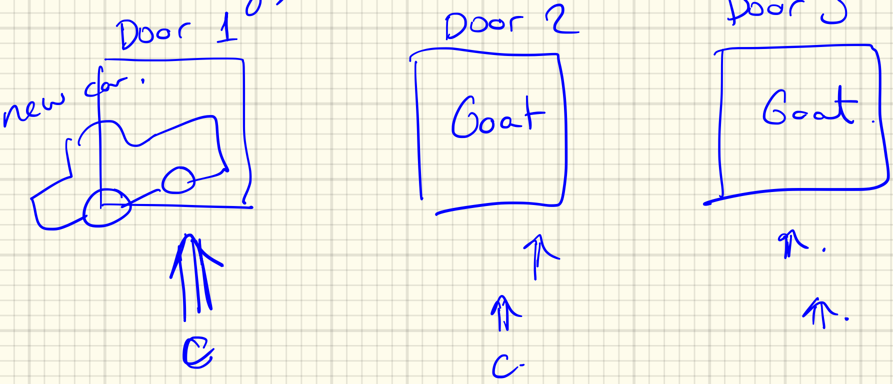
Note / Think about the point of this example

Note the subtlety in prior probability assumption.

side Note: $1 - P(A|B) \cancel{\neq P(A|B^c)}$ Don't make this mistake!

$\boxed{1 - P(A|B) = P(A^c|B)}$ → valid ✓

Ex 1.12 Bertsekas "Monty Hall" ; Host of the show.
4.4 (S.Kay)



- If the contestant sticks to his/her initial choice:

$$P(\text{winning}) = \frac{1}{3}$$

- If the contestant switches his/her initial choise

$$P(\text{winning}) = \frac{2}{3}$$

increases after Monty revealed one other door.

exercise : Study this ex.

Chapter 4 (Clust section) : Cluster Recognition : Real world example.

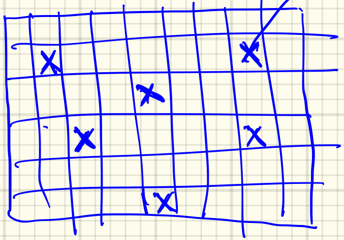
Skay Book

Crime Analysis : Gang is present

$B = \{\text{Gang (Cluster) exists}\}$

$A = \{\text{observed crime data}\}$

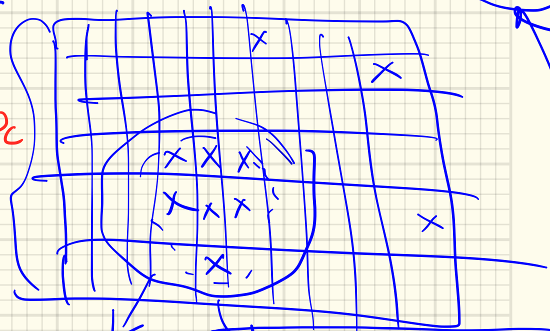
Fig 4.10



50×50

$P_{\substack{cluster\\ no\_gang}} = \frac{0.01}{5} = P_{nc}$

$P_{cluster} = 0.1 = P_c$



possible cluster area

11 crimes in a 145 pixel area.

No cluster region

$P(B|A) = ?$

$P(A|B) = P(k = 11 \mid \text{crime cluster exists}) = \binom{145}{11} P_c^{11} (1-P_c)^{145-11}$

$P(A|B^c) = P(k = 11 \mid \text{no cluster exists}) = \binom{145}{11} P_{nc}^{11} (1-P_{nc})^{134}$

11 crimes in 145 cells

$\boxed{P(B) = 10^{-6}}$

You need prior assumption

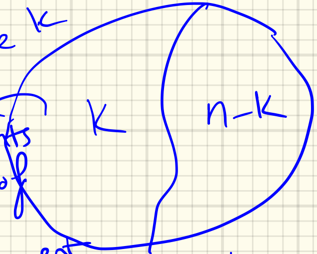
Odds Ratio against the hypothesis $= \frac{P(B^c|A)}{P(B|A)} = \frac{P(A|B^c)P(B^c)}{P(A|B)P(B)} = \frac{(0.01)^{11}(0.99)^{134}(1-10^{-6})}{(0.1)^{11}(0.9)^{134}(10^{-6})}$

$\approx 3.52 \rightarrow$? Not strong odds !!

Reject the hypothesis!

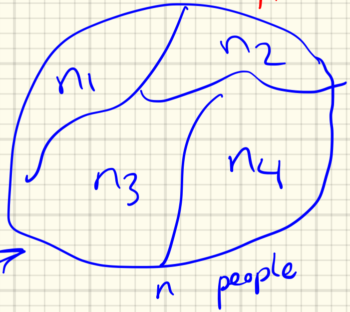
(Recall Binomial coef : Multinomial coefficient

choose k elements out of an n-element set.



$$\binom{n}{k}.$$

Binary partition



Q. In how many ways can we do this partition? # partitions?

$$\sum_{i=1} n_i = n$$

$$\underbrace{n_1}_{\text{people}} + \underbrace{n_2}_{\text{people}} + \underbrace{n_3}_{\text{people}} + \underbrace{n_4}_{\text{people}} = n$$; total of n people.

Multinomial coefficient

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

see Ex 1.33 (Bertsekas)

52 card deck → deal to 4 players (13 cards/player)

P(each gets an ace) = ? Event A : how many 4 aces are distributed to 4 people

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

Also count in how many ways remaining 48 cards are distributed to 4 people: $\dfrac{48!}{12!\,12!\,12!\,12!}$

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

Multiple (Independent) Random Experiment:

When sub-experiments are independent: $A_1, \dots, A_n$ are independent.

Not general

$$(1) \rightarrow \boxed{P(A) = P(A_1)P(A_2)P(A_3)\dots P(A_n)}$$

$P(A_1 \cap A_2 \dots \cap A_n)$

$A = \bigcap A_i$ ✓ : A joint event.

When we don't have independence: (no independence assumption)

most general

$$(2) \rightarrow \boxed{P(A) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)\dots P(A_n | \bigcap_{i=1}^{n-1} A_i)}$$

$P(A_n | A_1, A_2, A_3 \dots A_{n-1})$

Special Dependence Case:

Ex: Dependent Bernoulli Trial: Say 2 coins:

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

Rule of the experiment: Choose at random 1 coin:

Get a Tail $\rightarrow$ Switch to unfair coin

Geat a Head $\rightarrow$ Switch to fair coin.

Event: Getting 10 Tails in succession:

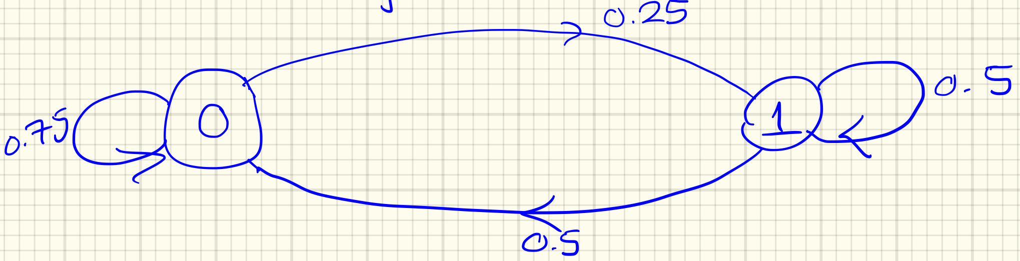
the joint event $\rightarrow$ $A = \{0, 0, \dots, 0\}$

$\{Tail, Tail, \dots, Tail\}$

Tail $\equiv 0$

$\rightarrow P(\text{Tail on the } \underset{\sim}{\underline{i\text{th}}} \text{ toss} \mid \text{Tail on the } \underset{\sim}{\underline{(i-1)^{st}}} \text{ toss}) = \underbrace{P(0 \mid 0)}$

We can draw a diagram



Markov State Probability diagram.

$P(1 \mid 0)$
$P(1 \mid 1)$
$P(0 \mid 1)$

Def: This type of Bernoulli sequence, where the probability for trial i (in the sequence) depends only on the outcome of the previous trial, is called a Markov Sequence.

$$P(A_i \mid A_{i-1}, A_{i-2}, \ldots, A_2, A_1) = P(A_i \mid A_{i-1})$$

due to Markov sequence property

3 $$\boxed{P(A) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_2)P(A_4 \mid A_3) \ldots P(A_n \mid A_{n-1})}$$

$A_1 = \{0\}$
$A_i = \{0\}$

$$P(A) = P(A_1) \prod_{i=2}^{10} P(A_i \mid A_{i-1}) \quad ; \quad P(A_i \mid A_{i-1}) = P(0 \mid 0) = ? \; 0.75$$

$i = 2, \ldots, 10$.

$$P(A_1) = P(\text{Tail} \mid \text{Fair}) \underbrace{P(\text{Fair})} + P(\text{Tail} \mid \text{Unfair}) \underbrace{P(\text{Unfair})}$$

This is Total Prob. Law. (Recall)

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

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

$\approx 5\%$

Compare this probability to the case we have independent & fair coins:

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

much larger probability

$$A = \{TT \cdots T\}$$

$\frac{1}{2} \cdot \frac{1}{2} \cdots$

$\rightarrow$ This example is a simple case of a Markov chain

Random Variables (r.v.)

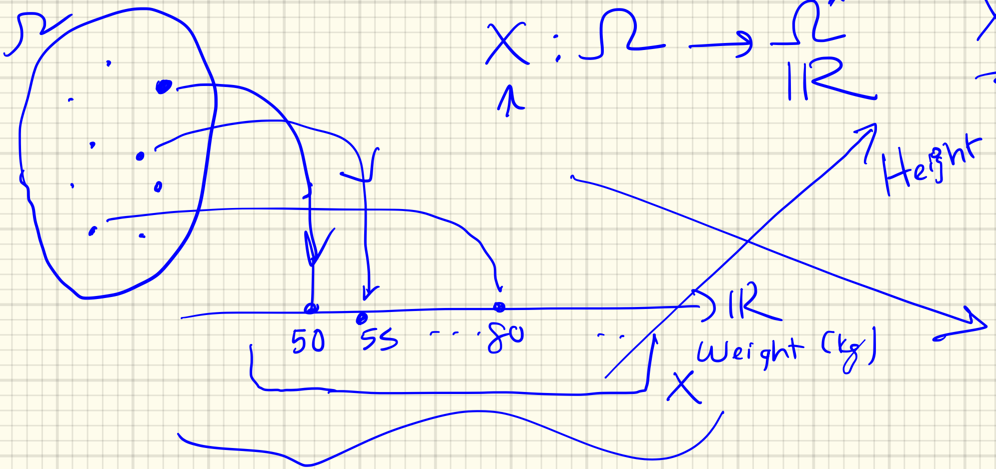
AoP ✓

R.v.'s derived quantities.

$X : \Omega \to \Omega^X$

$\mathbb{R}$

X is a function.



Def: An r.v. is a mapping (function) from the sample space Ω to a subset of the real line

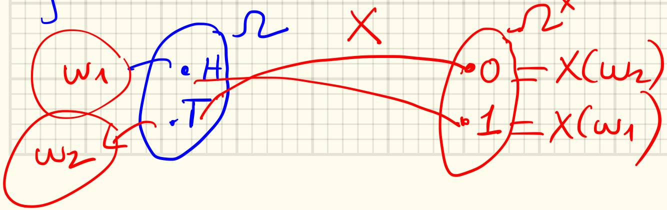
$\mathbb{R} = \{x : x \in (-\infty, \infty)\}$

Continuous r.v

$\Omega^X = [0,1]$

$|\Omega^X| = \infty$.

eg. Flip a coin

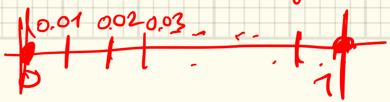


Discrete R.V. vs $\Omega^X = [0,1]$ Continuous r.V.

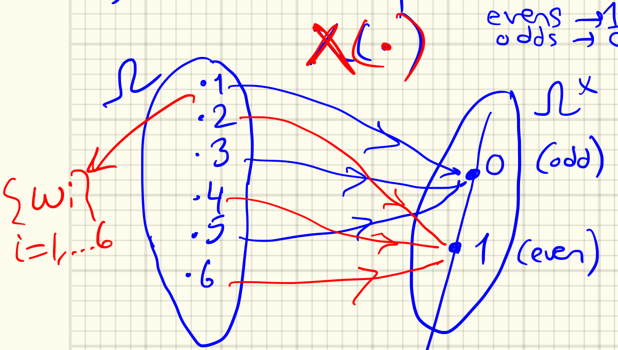
finite, countably ∞ # of values

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

$|\Omega^X| = 101$



Ex: Random experiment: fair dice roll.

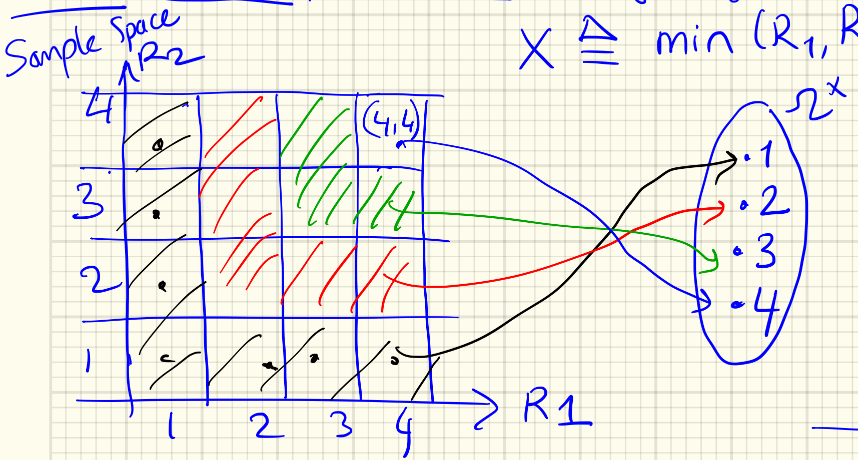


$$X(\omega_i) = \begin{cases} 0, & i = 1, 3, 5 \\ 1, & i = 2, 4, 6 \end{cases}$$

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

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

r.v. maps outcomes to numerical values.



Calculate

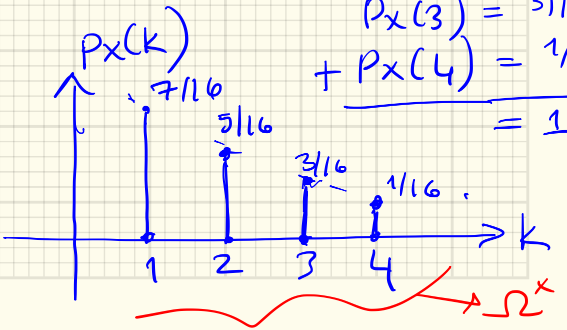
$P_X(1) = 7/16$

$P_X(2) = 5/16$

$P_X(3) = 3/16$

$+ P_X(4) = 1/16$

$= 1.$



Probability Mass Function: (pmf)

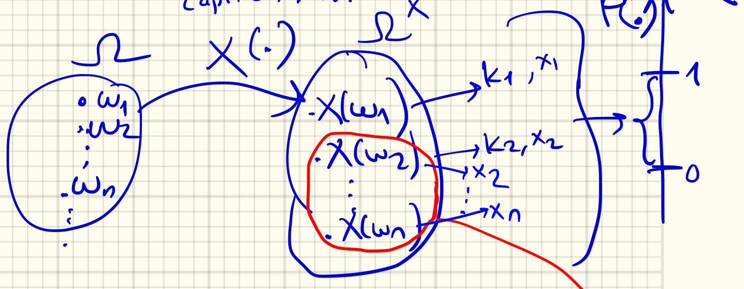
$p_X[x] = P_X[X(\omega)]$

little p → pmf

capital X: r.v.

$$p_X[\cdot] : \Omega^X \longrightarrow [0,1]$$

$\in \mathbb{R}$.



A: an event.

Properties of the pmf:

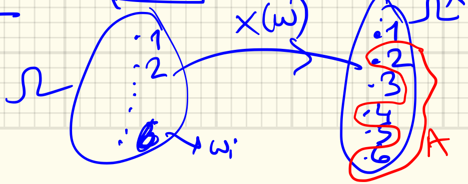
(1) $0 \leq p_X[k] \leq 1$

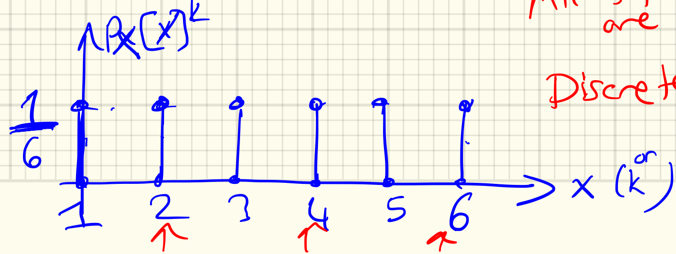
(2) $\sum_{i=1}^{M} p_X[i] = 1$; $\sum_{i=1}^{\infty} p_X[i] = 1$ (Normalization property)

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

event A defined in Ω^X

ex: 1 $\boxed{\text{fair}}$ die roll:





All 3 properties are satisfied.

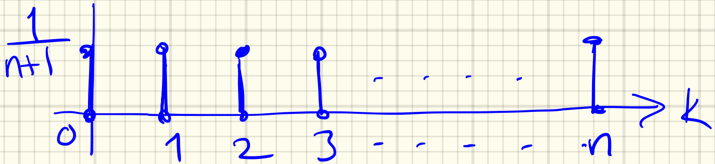
Discrete uniform r.v.

Important pmfs:

1) Discrete Uniform r.v.

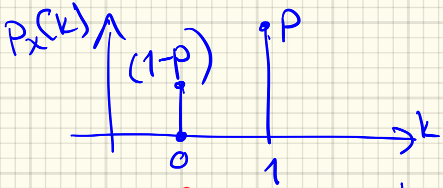
$X(\cdot) \rightarrow \{0, 1, \dots, n\}$

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



2) Bernoulli r.v.

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



✓ (3 properties of pmf)

3) Binomial r.v.

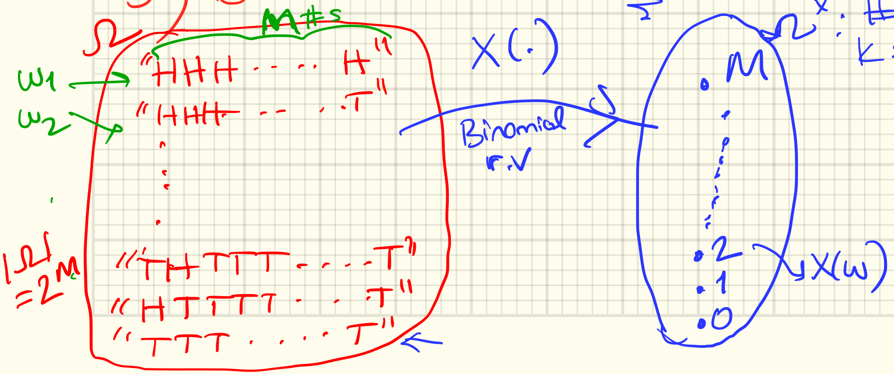
$\underline{\underline{k}}$ successes in $\underline{M \text{ trials}}$: $p_X[k]$

X: # heads (successes) in M (Bernoulli) trials

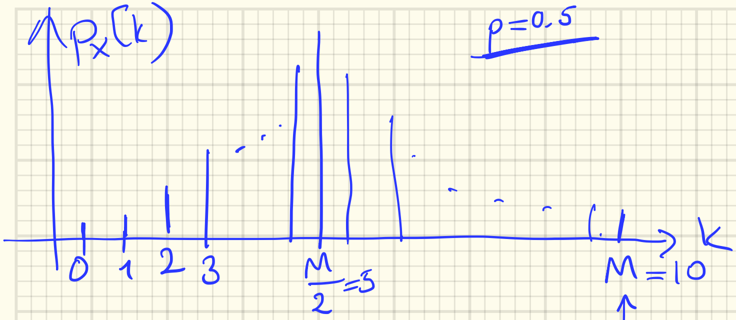
$k = 0, 1, \dots, M$;

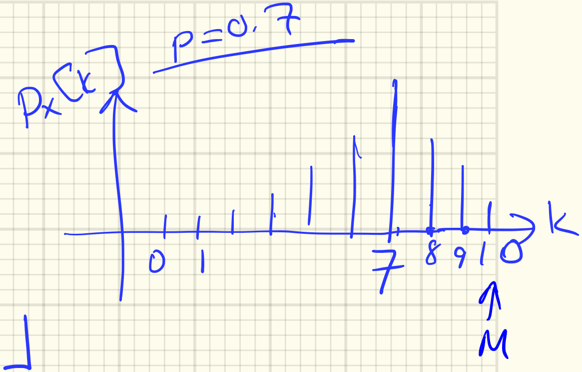
$$p_X[k] = \binom{M}{k} p^k (1-p)^{M-k}$$

✓



Binomial(M, p) : M & p are the parameters of the pmf.





Location of the max of the pmf: $\lfloor (M+1)p \rfloor$

$\lfloor x \rfloor$: largest integer $\leq x$

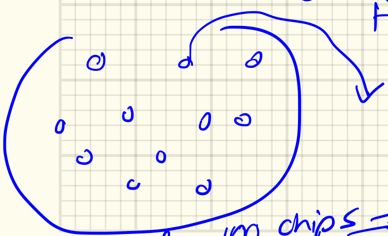
Real World problem: [Sec 3.10 Skay]. Quality Control Engineer

Test only a proportion of the chips to be shipped.

Criterion for acceptance of a Batch (to-be-shipped)

At least 95% of the tested chips are "good" (non-defective)

95 or more are good : Success event.



$$P_X[k \geq 95] = P_X[k=95] + P_X[k=96] + \dots + P_X[k=100]$$

$$= \sum_{k=95}^{100} \underbrace{\binom{100}{k} p^k (1-p)^{100-k}}_{P_X[k]} \approx 0.45 \, !$$

Test only 100 chips →

$\boxed{p = 0.94 \text{ : proportion of good chips in a batch}}$

To reduce this probability : Quality control engineer changes the strategy to :

Ship a batch only if 98 or more of the chips are good.

$$P_X[k \geq 98] = \sum_{k=98}^{100} \binom{100}{k} p^k (1-p)^{100-k} \approx 0.05 \quad \text{w/ } p = 0.94$$

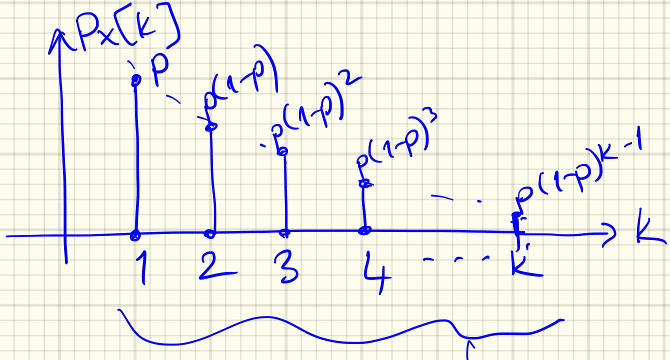
✓ an acceptable proportion of defective batches to be shipped.

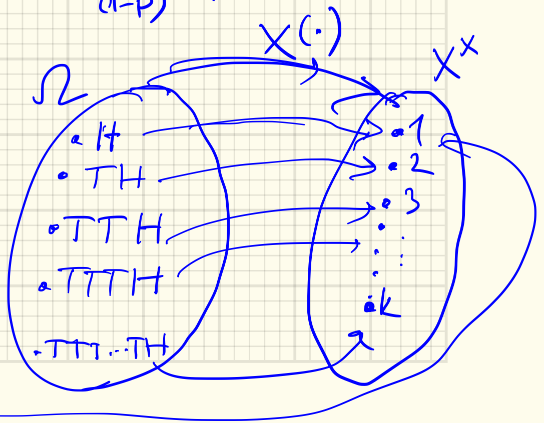
4) Geometric pmf : Prob. of <u>success</u> at the kth trial.

T T T T T ... H

$\underbrace{\text{T T T T T}}_{k-1 \text{ fails}}$ → $(1-p)^{k-1}$; H ↑ kth is a success

$$P_X[k] = (1-p)^{k-1} \cdot p \; .$$

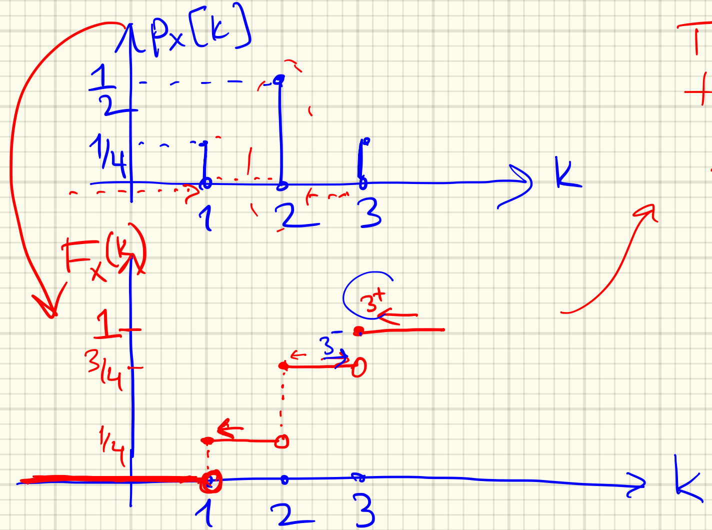




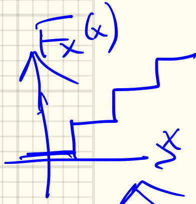
Cumulative Distribution Function (CDF)

Alternative means of summarizing the probabilities of an r.v.

Def: $F_X(x) = P[X \le x]$



The CDF is a running sum that adds up probabilities of the pmf, starting at $-\infty$, ending at $+\infty$.



★ For a discrete pmf, the cdf $F_X(x)$ are always staircase like.

→ pmf from cdf: $p_X[k] = F_X(k^+) - F_X(k^-)$ } size of the jumps.

$$p_X[3] = 1 - \frac{3}{4} = \frac{1}{4}$$