01.11.2021

YZV 231E

Probability Theory & Stats

Gözde ÜNAL

S → Outcomes

→ Sample Space → Events → Probabilities

→ Axioms of Probability: → 3 axioms:

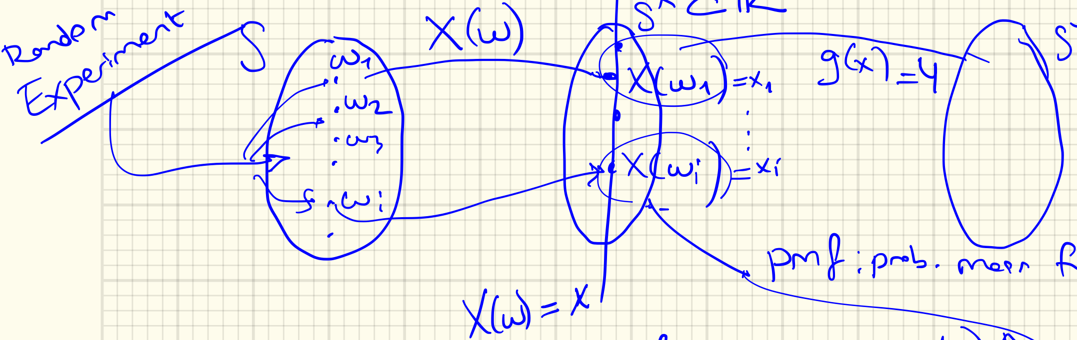
1) $P(A) \geq 0$

2) $P(S) = 1$

3) $P\left(\bigcup_i A_i\right) = \sum_i P(A_i)$ ↳ disjoint

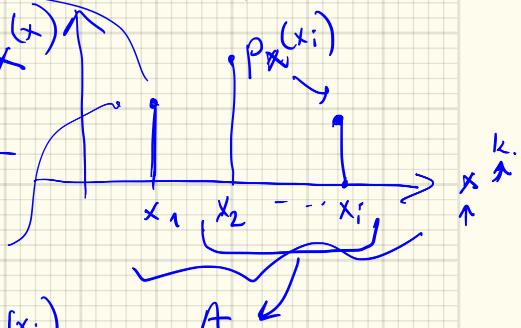
→ Borel field: subsets of S.

Random Variables: (r.v.s) map outcomes to Real line ↳ numerical values



pmf: prob. mass fn. $P_X(x)$ → $X(\omega)$

pmf properties: $P_X(x)$



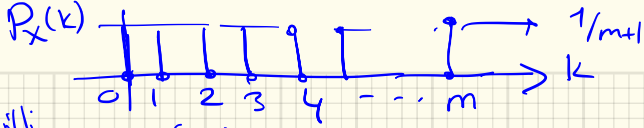
1) $P_X(x) \geq 0$

2) $\sum_{x_i} P_X(x_i) = 1$ Normalization prop.

3) $P(A) = \sum_{i:\, x_i \in A} P_X(x_i)$

Discrete pmfs:

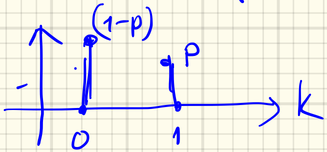
1) Uniform pmf



2) Bernoulli pmf : Bernoulli exp.

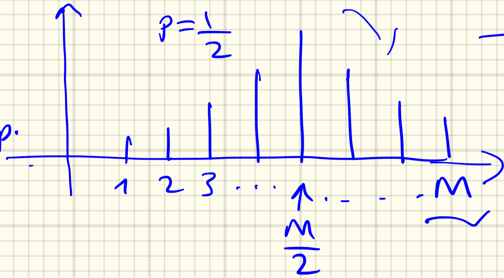
p : success prob.

$(1-p)$: failure prob.



3) Binomial pmf : $\boxed{\text{Bernoulli exp}}$ → indep.

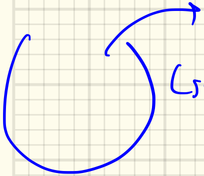
p : success prob.



k successes (heads) out of M coin tosses.

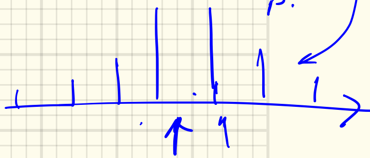
drawing ball

(sampling)



Urn

M even or M odd. p.



$$P_X(X=k) = \binom{M}{k} p^k \cdot (1-p)^{M-k}$$

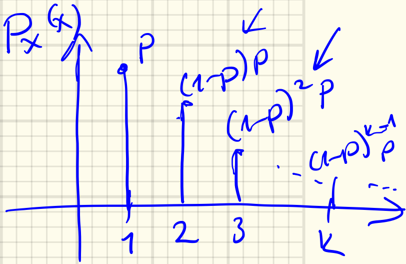
4) Geometric pmf : (Bernoulli experiments) independent.

: Success at the k^{th} trial.

$\underbrace{\underbrace{T\ T\ T \ldots}_{}\ H}_{}$.

$$\underbrace{P_X(X=k)}_{x} = (1-p)^{k-1} \cdot p.$$

$k = 1, 2, \ldots, \infty$



5) Poisson pmf: $p_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}$, $k \geq 0$.

λ : # of events (arrivals, requests) in a unit time.

(Key) Chapter 5 Real World Example: (Servicing Customers)

1 express lane (in 5-6 pm interval) services each customer ~1 min.

Q: Manager wants to know how many extra express lanes should be opened?

Requirements: No more than 1 person waiting in the line 95% of the time

$\equiv$ no more than 2 persons arriving at the express lane in a 1-min. interval.

You model this problem w/ → Poisson pmf: → λ : estimate this: avg arrival rate #customers arriving/min.

	Mon	Tue,	Wed	... Sun
5-6 pm 60 min	71	69	68	72 ...

Note the arrival times in a 60 min interval for 1 week

$\rightarrow \lambda \cong \frac{70 \text{ arrivals}}{60 \text{ min}} = \frac{7}{6}$ average # customers arriving in a 1 min interval.

$P[k \text{ arrivals in } \underbrace{1 \text{ min}}_{\text{in a unit interval}}] = p_X[k] = e^{-\lambda} \frac{\lambda^k}{k!}$, $k = 0, 1, \ldots$

We require $P[X \leq 2] = \sum_{k=0}^{2} p_X[k] = \sum_{k=0}^{2} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda}\left(1+\lambda+\frac{\lambda^2}{2}\right)$

$k \leq 2$

w/ $\lambda = \frac{7}{6} \to \underline{P[X \leq 2]} = \underline{0.88} < 0.95$ X.

Solution: Consider opening a 2nd express lane:

Two lanes; Two sets of arrivals

↳ they are independent. $\left(P\left(\bigcap_i A_i\right) = \prod_i P(A_i)\right)$

Lane 1 Lane 2

1 lane

$\cancel{\lambda = \frac{7}{6}}$

$\lambda^{new} = \frac{7}{12}$

$P[X \leq 2] = P[\text{2 or fewer arrivals at both lanes}]$

$= \underbrace{P[k \leq 2 \text{ in Lane 1}] \cdot P[k \leq 2 \text{ in Lane 2}]}$

$= \left(P[k \leq 2 \text{ in Lane } i]\right)^2$

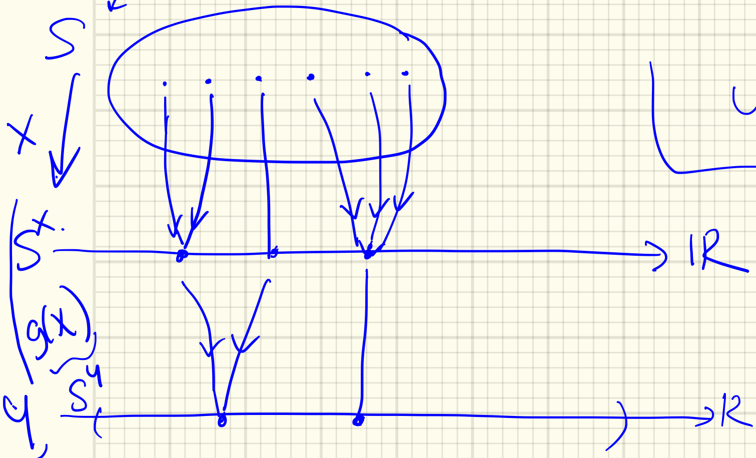
$= \left(\sum_{k=0}^{2} e^{-\lambda} \cdot \frac{\lambda^k}{k!}\right)^2 = \left(e^{-\lambda^{new}}\left(1+\lambda^{new}+\frac{(\lambda^{new})^2}{2}\right)\right)^2$

$= \underline{0.957}$ ✓ meets the requirement

Transformation of r.v.s :

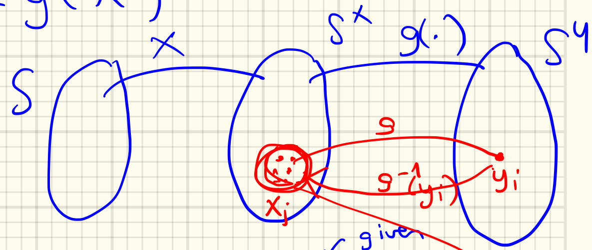
$g : X \longrightarrow Y$: could be many to one

a fn. $\exists$ a single output for any input.



a new random variable

$$Y = g(X)$$



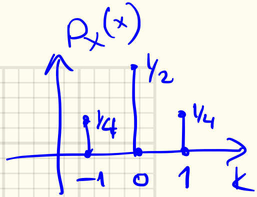
given

$$P_Y[y_i] = \sum_{j : g(x_j) = y_i} P_X[x_j]$$

new xformed pmf

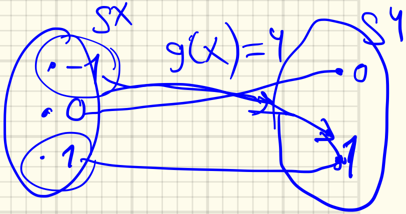
$$= \sum_{x \in g^{-1}(y)} P_X[x]$$

Ex: X r.v. w/ pmf $p_X[k] = \begin{cases} 1/4 & k=-1 \\ 1/2 & k=0 \\ 1/4 & k=1 \end{cases}$



Let $Y = X^2 = g(X)$, defined on the sample space $S^X = \{-1, 0, 1\}$

What is $\underline{S^Y} = \{0, 1\}$

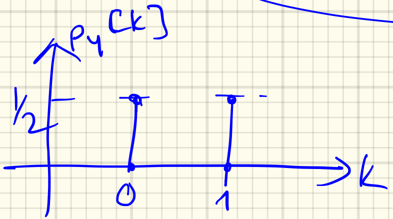


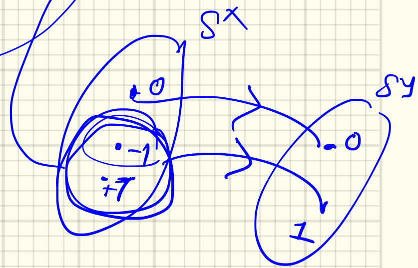
— $g(x_j) = x_j^2 = 0$ y_1 only for $x_j = 0$

$p_Y[0] = p_X[0]$

— $g(x_j) = x_j^2 = 1$ y_2 for $x_j = -1$, $x_j = +1$ $\Big\}$ $p_Y[1] = p_X[-1] + p_X[1] = \frac{1}{2}$

$y_j =$

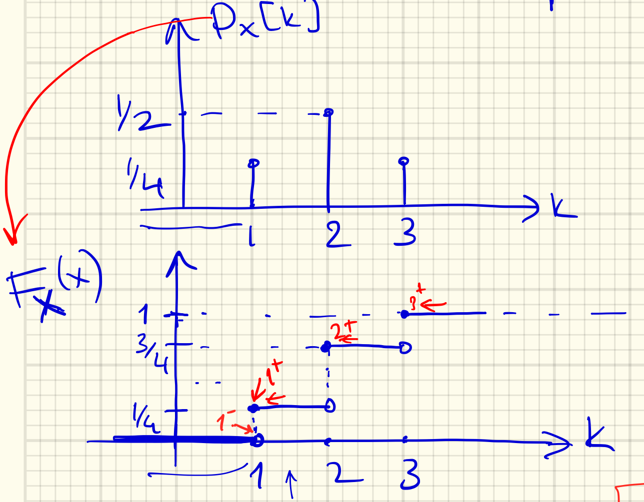




Cumulative Distribution Function (CDF)

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

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

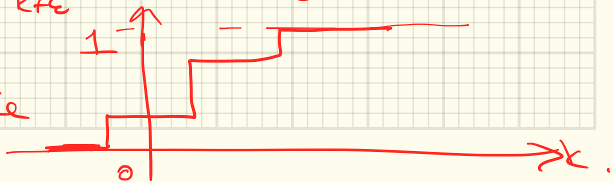


The cdf is a running sum that adds up probabilities of the pmf, starting at $-\infty$, ending at $+\infty$.

— pmf from cdf : $p_X[k] = \boxed{F_X(\underbrace{k^+}_{k+\varepsilon}) - F_X(\underbrace{k^-}_{k-\varepsilon})}$: size of the jumps.

continuous case

For a discrete r.v., the cdf $F_X(x)$ is always stair-case like



Properties of the CDF : (Section 5.8 Kay Book).

1) $0 \leq F_X(x) \leq 1$

2) $\lim_{x \to \infty} F_X(x) = 1$, $\lim_{x \to -\infty} F_X(x) = 0$

$\left(P(\underbrace{X \leq -\infty}_{}) = P(\phi) = 0 \right)$

$\left(P(\underbrace{X \leq \infty}_{S}) = P(S) = 1 \right)$

3) CDF is $\underbrace{\text{right-continuous}}$: as we approach x_0 from the right, the limiting value of the CDF is $F_X(x_0)$

$\lim_{x \to x_0^+} F_X(x) = F_X(x_0)$

4) CDF is monotonically non-decreasing. if $x_1 \leq x_2$.

$\underbrace{F_X(x_1)}_{P(A)} \leq \underbrace{F_X(x_2)}_{P(B)}$

$\to$ show

$A = \{-\infty < X < x_1\}$

$B = \{-\infty < X < x_2\}$

$x_1 \leq x_2$: $A \subset B$

$\to P(A) < P(B)$

5) Intervals : $P[a < X \leq b] = F_X(b) - F_X(a)$

show for $a<b$: $\{-\infty < X \leq b\}$

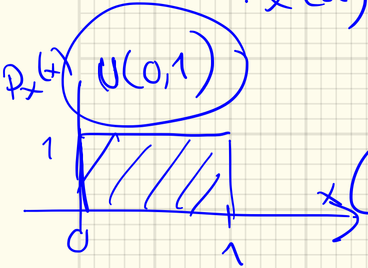
$-\infty$ a b

$(-\infty, a] \cup (a, b]$ — disjoint sets

$(-\infty, b]$

$$\underbrace{P((-\infty, a])}_{F_X(a)} + \underbrace{P(a < X \le b)} = F_X(b)$$

$$\Longrightarrow \quad F_X(b) - F_X(a) = P(a < X \le b)$$



Probability Integral Transformation:

(1) If an r.v. is transformed according to its CDF:

$$\boxed{U = F_X(X)} \rightarrow \text{the transformed r.v. is uniform r.v.}$$

$U \sim U[0,1]$

(2) The transformation $F_X^{-1}(U) \longrightarrow X$ w/ the inverse CDF, ~~$U \to X$~~ produces an r.v. distrib. by $p_X(x)$

$X \xrightarrow[\text{cdf}]{F_X} U$

This is called **Inverse Probability Integral Transformation.**

Prob. model of X:

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

$p_X(x)$

$x_1, x_2, \ldots$ $\qquad u_1, u_2, \ldots$

proof Thm 10.9.1 Kay.

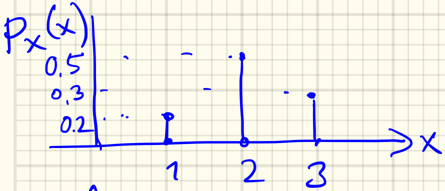
Thm → Given the CDF of X, we want to generate random numbers distributed as X,

$$U \xrightarrow{F_X^{-1}(u)} X.$$

$g = F_X^{-1}$; Instead of using any fn $g(\cdot)$ to transform my r.v. we use $g = F_X^{-1}$ (cdf inverse fn).

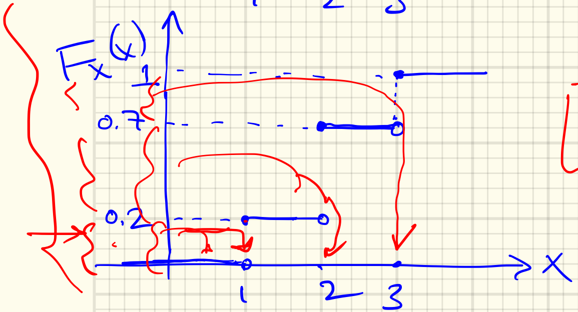
Ex: How to simulate an r.v. on a computer.

Say X takes on values : $S^X = \{1, 2, 3\}$ w/ a pmf.



Q: Write a code that generates $M = 1000$ realizations of X.

→ We have $U[0,1]$ r.v. (uniform), the values you sample are equally likely from interval $(0,1)$



$$\boxed{x = F_X^{-1}(u)} \quad \leftarrow \text{rand}(\cdot) \text{ function.}$$

→ We sample $u_1, u_2, u_3, \dots, u_{M=1000}$

$$\left.\begin{cases} 0 \le u < 0.2 & \to x = 1 \\ 0.2 \le u < 0.7 & \to x = 2 \\ 0.7 \le u \le 1 & \to x = 3 \end{cases}\right\}$$

$x_1, x_2, x_3, \dots, x_{1000}$

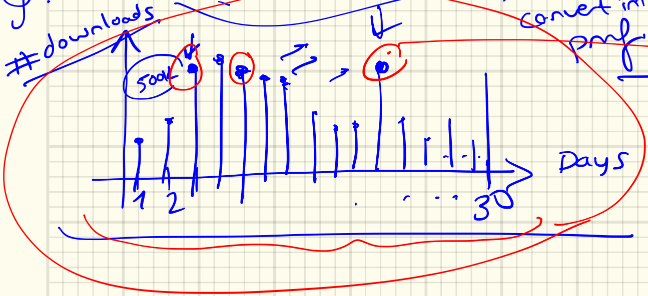
Now, the r.v's we generated are distributed according to $p_X(x)$

$x_1, \dots, x_{1000}$

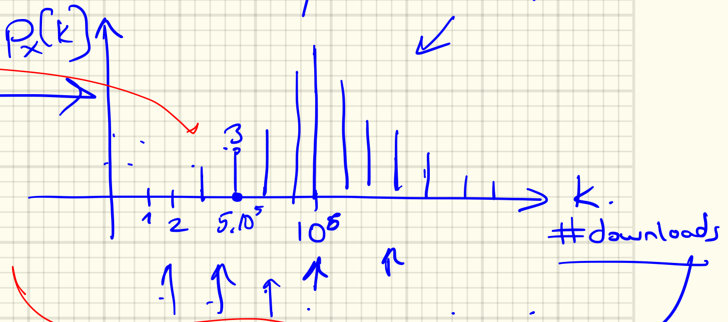
↳ plot its pmf. $\sim \frac{1}{5}$th are $x = 1$

Expectation of an r.v.

eg. You collected data, eg. server download requests in the past month



convert into pmf.



To interpret probability distrib. we may want more compact values from the pmf. → eg. averages

(Note: pmf is a complete description of an r.v. (data collection) to determine probabilities)

Def (Expected Value of an r.v)

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

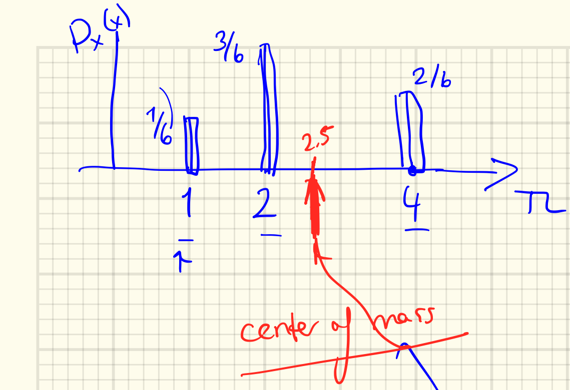
approximation

average value of the outcomes of a large # experimental trials.

Sample mean: An r.v. w/ N realization

$$x_i = 1, 2, \dots, N$$

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$$



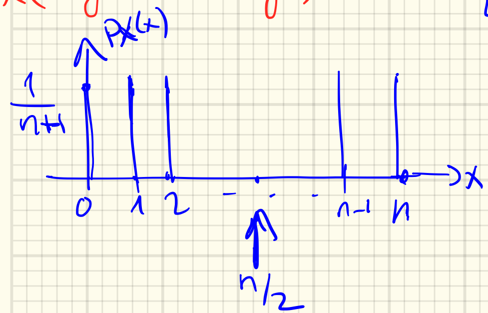
$$1\cdot\frac{1}{6} + 2\cdot\frac{3}{6} + 4\cdot\frac{2}{6} = 2.5\,\pi.$$

$$\left(\sum_k k\cdot P_X[k] \qquad \sum_x x\cdot P_X(x)\right) = E[X]$$

expected value

Discrete

Ex (Uniform Pmf:)

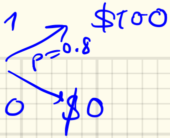


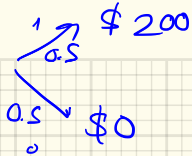
$$E[X] = ? = \sum_{k=0}^{n} k\cdot\frac{1}{n+1} = 0\cdot\frac{1}{n+1} + 1\cdot\frac{1}{n+1} \cdots + n\cdot\frac{1}{n+1}$$

$$\frac{1}{n+1}\sum_k k = \frac{1}{\cancel{n+1}}\cdot\frac{n\,\cancel{(n+1)}}{2} = \frac{n}{2}$$

Note: For a symmetric pmf. around a certain value → that is the expected value of the associated pmf

Ex 2.8 Quiz problem:

Q1 : 

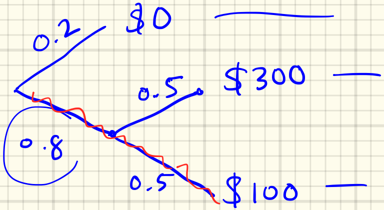
; Q2 

If 1st attempted question answered incorrectly, the quiz terminates.

Q: Which question should be answered first to maximize the expected value of the total prize money you get?

Q1 answered first case:

show in a sequential tree diagram:

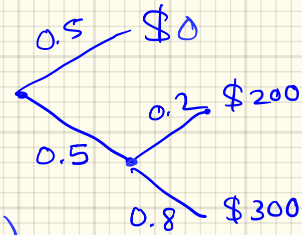


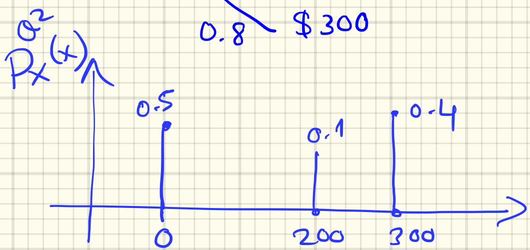
$P_X(100) = (0.8).(0.5) = 0.4$

$E[X] = 0.(0.2) + 100.(0.4) + 300(0.4)$

$= \$160$

Q.2 answered first case:





$E[X] = 0.(0.5) + 200.(0.1) + 300(0.4)$

$= \$140$

Expectation helps us in our decision making / would go w/ Q1 first

Expected Values of Some Important R.V.s (Sec 6.4 Kay)

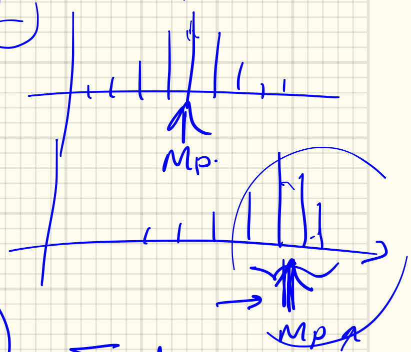
1. Bernoulli: $X \sim Ber(p)$

$$E[X] = \sum_{k=0,1} k \, p_X[k] = 0.(1-p) + 1.p = p.$$

2. Binomial: $X \sim bin(M,p)$:

$$E[X] = \sum_{k=0}^{M} k \binom{M}{k} p^k (1-p)^{M-k} = \underline{M.p}$$

→ check the Kay book for the derivation



3. Geometric: $X \sim geom(p)$

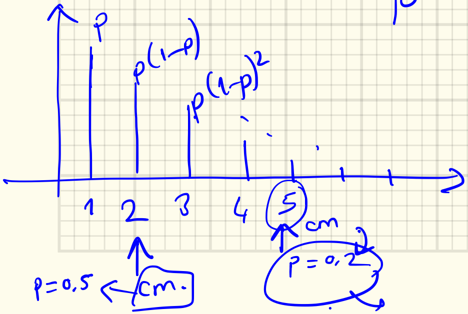
$$E[X] = \sum_{k=1}^{\infty} k.(1-p)^{k-1} .p = p \sum_{k=1}^{\infty} k \underbrace{(1-p)}_{\alpha}{}^{k-1}$$

$$E[X] = p . \frac{1}{\underbrace{(1-(1-p))^2}_{p^2}} = \boxed{\frac{1}{p}}$$

$$\frac{d}{d\alpha}\left(\sum_{k=0}^{\infty} \alpha^k\right) = \sum_k \frac{d}{d\alpha}\alpha^k = \sum_k k.\alpha^{k-1}$$

$$\frac{d}{d\alpha}\left(\frac{1}{1-\alpha}\right), \; |\alpha| < 1$$

$$+\frac{(1)}{(1-\alpha)^2}$$



4) Poisson: $E[X] = \lambda$

$$E[X] = \sum_{k=0}^{\infty} k \cdot e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!}$$

$$\frac{d}{d\lambda}\left(\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}\right) = e^{\lambda}$$

$$\sum_{k=0}^{\infty} k \cdot \frac{\lambda^{k-1}}{k!}$$

$$= e^{-\lambda} \cdot \lambda \cdot e^{\lambda}$$

$$E[X] = \lambda.$$

Variance: eg. second moment: $\sum_x x^2 p_X(x) = E[X^2]$

$$Var(X) \triangleq E\Big[\big(\underbrace{X - E[X]}\big)^2\Big]$$

→ Standard deviation $\triangleq \sqrt{Var(X)}$

$$\sigma_X \triangleq \sqrt{\underbrace{Var(X)}_{\sigma_X^2}}$$

E[X]

cm.

we take distance from the mean X square it X take average → that is variance