2025-01-2

Random Variables

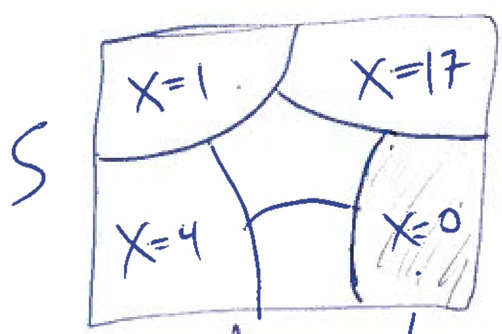
S sample space

P probability function.

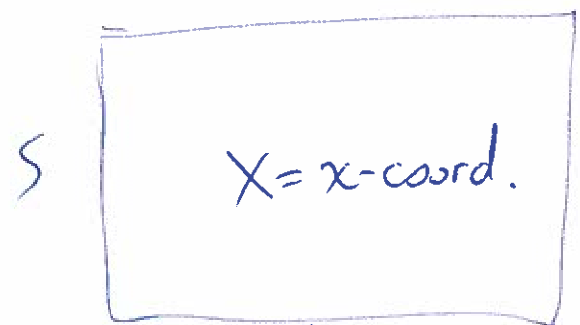
def: A random variable is a function $X : S \to \mathbb{R}$

Usually use X, Y, Z, N to denote R.V.s

Idea: If $s \in S$ is picked using P, then $X(s)$ is a real number. Usually just write X



discrete



continuous

A discrete R.V. is a R.V. that only takes countably many values $\{x_1, x_2, x_3, \ldots\}$

⊛ The probability Mass function (pmf) of X is

$p(a) = P(X = a)$ [only for disc. RV.]

⊛ The commulative distribution func. (CDF) of X is

$F(a) = P(X \leq a)$

Use $p_X(a)$ or $F_X(a)$ to specify the RV.

e.g. sum of 3 dice:

$S = \{(a,b,c) \quad a,b,c \in \{1 \ldots 6\}\}$ $\qquad X = a+b+c.$

here X takes values $3, 4, \ldots 18$.

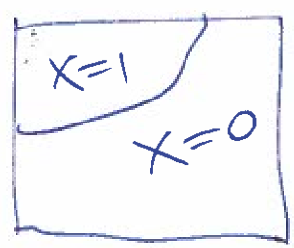
$P(3) = \frac{1}{216} \quad P(4) = \frac{3}{216} \ldots \qquad P(30) = 0 \qquad P(\pi) = 0$ $\quad \boxed{216 = 6}$

$\{X = a\} \stackrel{\text{def}}{=} \{s \in S \text{ s.t. } X(s) = a\}$

$\{X \leq a\} = \{s \in S \text{ s.t. } X(s) \leq a\}$

1) <u>Bernoulli R.V.</u> : X is always 0 or 1

For any $p \in [0,1]$, Bern(p) denotes a R.V. s.t.



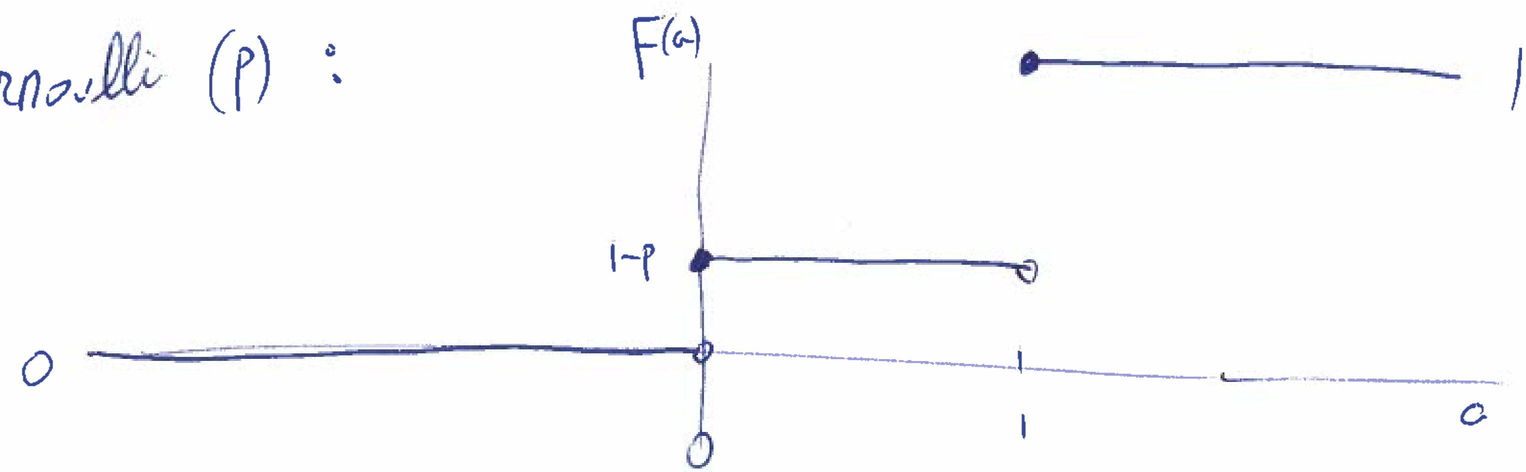
$P(X=1)=p$

$P(X=0)=1-p$

For an event $E \subset S$ let $1_E = \begin{cases} 1 & E \text{ occurs} \\ 0 & E^c \text{ occurs} \end{cases}$

I_E is a Bernoulli R.V. with parameter $P(E)$

CDF of Bernoulli (p) :



2) Geometric R.V. with parameter $q \in (0,1]$ is a R.V. with pmf is $p(n) = (1-q)^{n-1} \cdot q$

For such R.V. we say $X \sim Geom(q)$

possible values of X are $1, 2, 3, \ldots$

<u>Claim</u>: For a discrete R.V., $\sum_i p(x_i) = 1$

if possible values are $\{x_1, x_2 \ldots\}$

In the geom case: $\sum_{n=1}^{\infty} (1-q)^{n-1} q = 1$

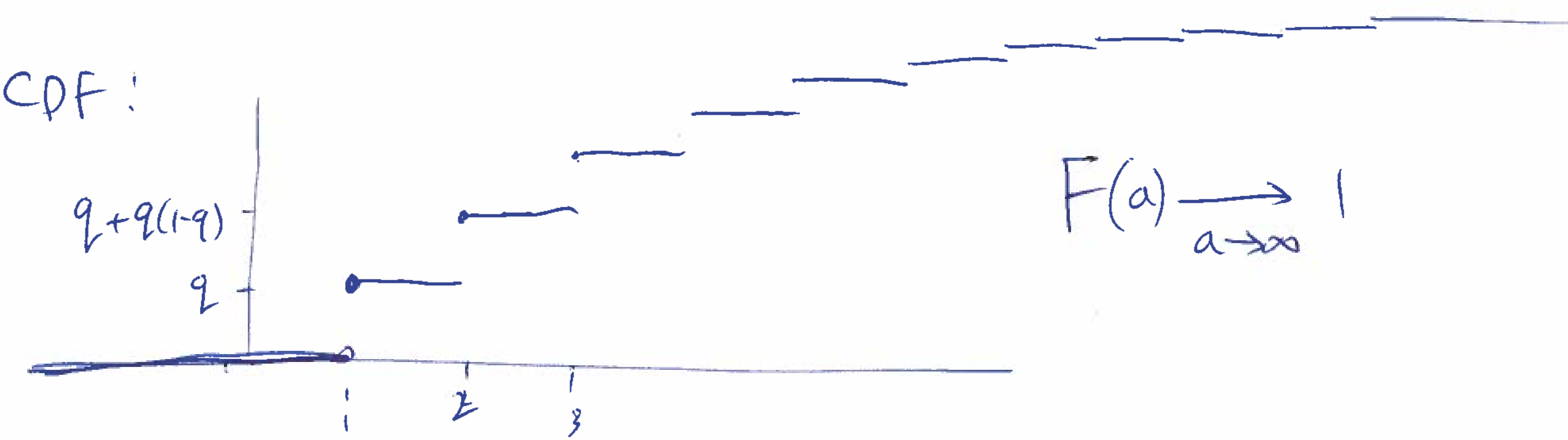
since $$\sum_{n=1}^{\infty} (1-q)^{n-1} \cdot q = q \sum_{n=1}^{\infty} (1-q)^{n-1} = q \sum_{n=0}^{\infty} (1-q)^{n} = q \cdot \frac{1}{1-(1-q)} = 1$$

If an experiment succ. with prob. q each time, independently

Let N = number of tries up to first success.

$$P(N=n) = \underbrace{(1-q)(1-q)\cdots(1-q)}_{n-1}\cdot q = (1-q)^{n-1}\cdot q$$

CDF:



$$F(a) \xrightarrow[a\to\infty]{} 1$$

memoryless property of Geom(q): $P(X>n+m \mid X>n) = P(X>m)$

__Idea__: If the experiment failed n times, then prob. of $> m$ additional fails $=$ prob $> m$ fails at the start.

__Proof__: $P(X>n) = \sum_{a=n+1}^{\infty} p(a) = \sum_{n+1}^{\infty} (1-q)^{a-1} \cdot q = (1-q)^n$

so $P(X>n+m \mid X>n) = \dfrac{P(X>n+m,\ X>n)}{P(X>n)} = \dfrac{(1-q)^{n+m}}{(1-q)^n} = (1-q)^m$