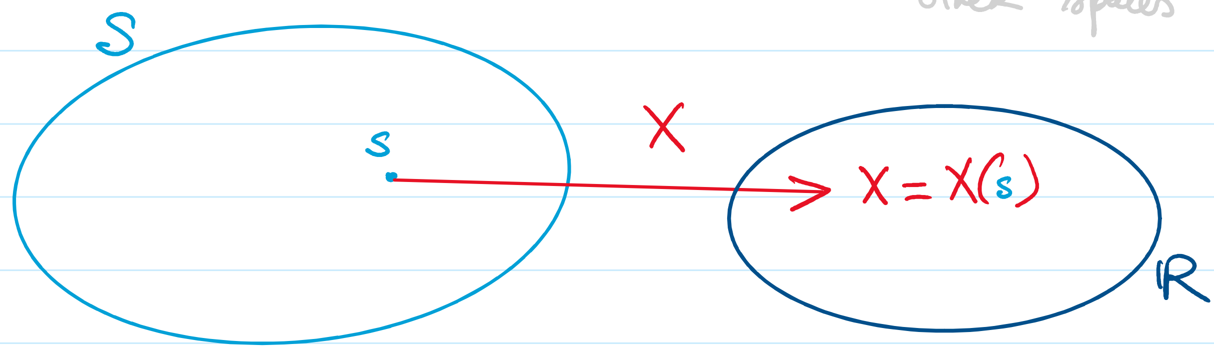
1.4 Random variables

$\boxed{\text{Def.}}$ A random variable is a function from a sample space into the real numbers

may change to other spaces



ω may be unobserved

X observed

Examples

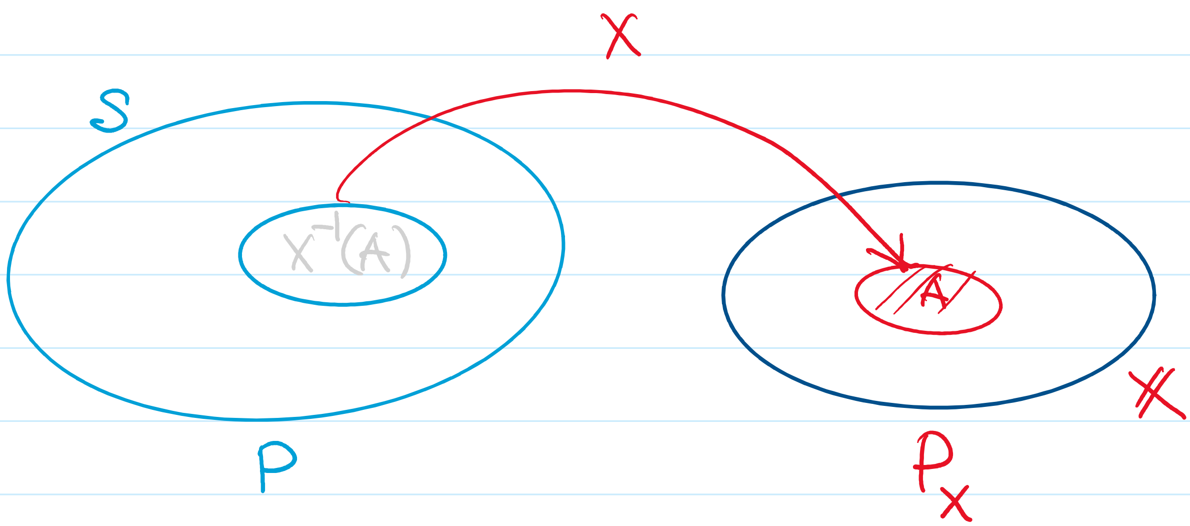
Experiments	Random Variables
1/ Toss 2 dices $S = \{1,\dots,6\} \times \{1,\dots,6\}$	X = sum of the 2 numbers
2/ Toss a coin 25 times $S = \{H, T\}^{25}$	X = # of Heads
3/ Apply different amounts of fertilizer to flower plants	X = yield / acre

PROBABILITY DISTRIBUTION

Def. Suppose the range of random variable X is $\mathcal{X}$. For any subset $A \subset \mathcal{X}$, the **probability distribution** of X is a probability function, denoted P_X, on (a sigma algebra of) subsets of $\mathcal{X}$ s.t. for any such subset $A \subset \mathcal{X}$

$$P_X(X \in A) = P(\{s \in S : X(s) \in A\})$$

$$= P(X^{-1}(A))$$



EXAMPLE: X is the number of heads out of the experiment of tossing coin 3 times

$$s \in S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

s	HHH	HHT	HTH	HTT	THH	THT	TTH	TTT
$X(s)$	3	2	2	1	2	1	1	0

So the Range of (Random) variable X is

$\mathcal{X}$	0	1	2	3
P_X	1/8	3/8	3/8	1/8

EXAMPLE $X = \#$ heads after tossing coin 50 times

$$\mathcal{X} = \{0, \dots, 50\}. \text{ So, } P_X(X = i) = \frac{\binom{50}{i}}{2^{50}} \quad \forall i \in \mathcal{X}$$

REMARK For complicated space (e.g., uncountable), we need to specify probabilities on basic and "simple" sets which generate the associated sigma algebra.

Distribution Functions

$\mathcal{X} = \mathbb{R}$ or $\mathbb{Z}$ or $\mathbb{N}$

Def. The cumulative distribution function (c.d.f.) of a random variable X, denoted by $F_X(x)$, is given by

$$F_X(x) = P_X(X \leq x), \quad \forall x \in \mathcal{X}.$$

notation: upper case   lower case

Theorem The function $F : \mathcal{X} \to \mathbb{R}$ is a cdf if and only if the following conditions hold:

(i) $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$.

(ii) F is a non-decreasing function in x

(iii) F is right-continuous, i.e.

$$\lim_{x \downarrow x_0} F(x) = F(x_0) \quad \forall x_0 \in \mathbb{R}.$$

Proof. "only if": follows directly from AOPs.

"if": quite harder, involving construction of S, $\mathcal{B}$, P and $s \mapsto X(s)$.

Examples

(1) "Tossing coins for a head"

Let $p :=$ probability a coin turns head

$X :=$ # independent coin tosses needed to get a head.

Then $X \in \{1, 2, \dots\}$

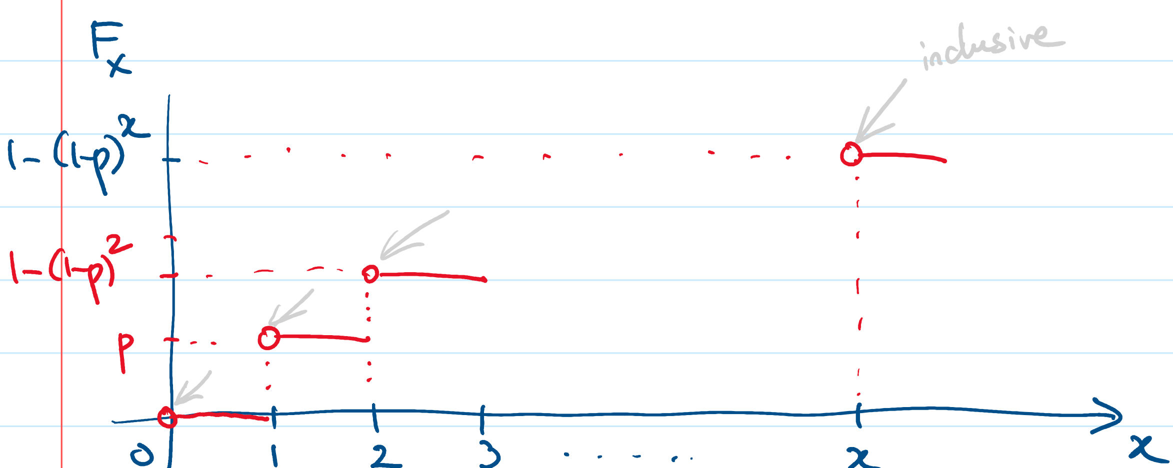
So, $P(X=x) = \begin{cases} (1-p)^{x-1}\, p & \forall x \in \{1,2,\dots\} \\ 0 & \text{otherwise}. \end{cases}$

X has geometric dist

$$\Rightarrow F_X(x) = P(X \leq x), \quad \forall x \geq 0$$

$$= \sum_{i=0}^{x-1} (1-p)^i p$$

$$= p \, \frac{1-(1-p)^x}{1-(1-p)} = 1-(1-p)^x$$



(2) Continuous Cdf

Let $F_X(x) = \dfrac{1}{1+e^{-x}}$

$$\frac{d}{dx} F_X(x) = \frac{e^{-x}}{(1+e^{-x})^2} > 0 \text{, so } F_X \uparrow$$

$$\begin{cases} F_X \downarrow 0 & \text{as } x \to -\infty \\ F_X \uparrow 1 & \text{as } x \to +\infty \end{cases}$$

Def. A Random variable X is continuous if its cdf is a continuous function

X is discrete if its cdf is a step function.

Notes,

- Cdf can also be a "mixture" of continuous segments and jumps

IDENTICAL DISTRIBUTIONS

Def. Two Random variables X and Y are **identically distributed** if

$$P(X \in A) = P(Y \in A) \quad \forall A \in \mathcal{B}.$$

in other words, X and Y have the same distribution.

Remark.
- Sometime we write $X \overset{d}{=} Y$.
- This does NOT say "$X = Y$".

Examples.

① Experiment: Toss a fair coin 3 times

X = # heads

Y = # tails

Then $X \overset{d}{=} Y$ (why?)

② Experiment: Sample the dishes made by a chef on a given day

X = dish tasted by Alice

Y = dish tasted by Bob

Then $X \overset{d}{=} Y$.

if, however, Alice and Bob came in different days (seasons), then we expect $X \overset{d}{\neq} Y$.

$\boxed{\text{Theorem}}$ X and Y are two Random variables associated with sample space S.

$X \stackrel{d}{=} Y$ if and only if $F_X(a) = F_Y(a)$ $\forall a \in \mathbb{R}$.