Information Theory Basics

(Cover, Thomas)

Ω: set of outcomes of a randomized experiment (sample space)

$P : \mathcal{P}(\Omega) \to [0,1]$ is called a probability distribution if

(1) $P(\emptyset) = 0$

(2) $\forall A \subseteq \Omega$ $\quad P(A) = 1 - P(A^c)$

(3) If $A_1, A_2, \dots$ are disjoint subsets of Ω, then $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum P(A_i)$

If Ω is discrete (i.e. finite or countably ∞), then (Ω, P) is called a **discrete probability space**.

e.g. 1. $\Omega = \{H, T\}$ $\quad P(\{H\}) = 0.5, \quad P(\{T\}) = 0.5$

2. $\Omega = \mathbb{N}$. $\quad \forall n \in \mathbb{N} \quad P(\{n\}) = \frac{1}{2^n}$.

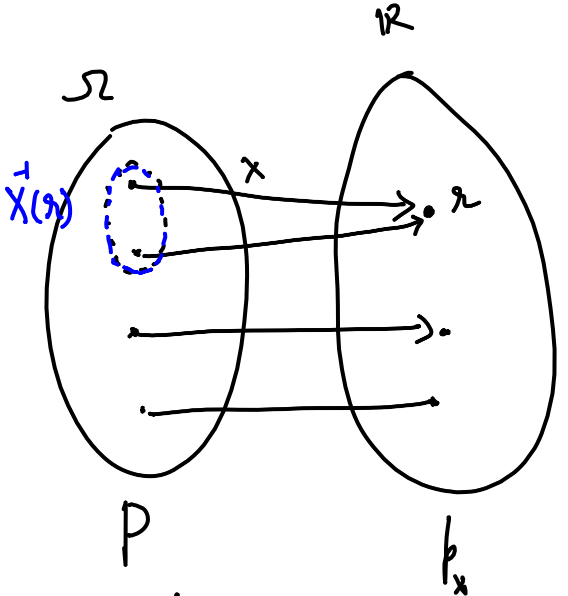
3. $\Omega = \mathbb{N}$ $\quad \forall n \in \mathbb{N} \quad P(\{n\}) = \frac{6}{\pi^2} \frac{1}{n^2}$, since $\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$

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

Instead of talking about probabilites of events (i.e. subsets of Ω), we can define probability induced by a random variable.

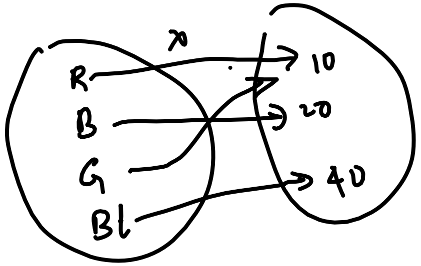
The probability distribution induced by X on $\mathbb{R}$, denoted p_X is defined for every $x \in \mathbb{R}$ as

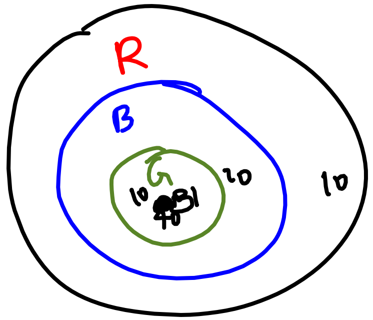
$$p_X(x) = P\left(\{\omega \in \Omega \mid X(\omega) = x\}\right) = P\left(X^{-1}(x)\right)$$



E.g. Game of darts with concentric rings with scores 10, 20, 10, 40 respectively for red, blue, green, black regions.

Let





$$p_X(10) = P(\{\text{Red}, \text{Green}\}) = 1/2.$$

$p_X(20) = P(\{\text{Green}\}) = 1/4$ & $p_X(40) = P(\{\text{Black}\}) = 1/4$

Let $X : \Omega_1 \to \mathbb{R}$ and $Y : \Omega_2 \to \mathbb{R}$ be 2 random variables. The joint distribution $p_{X,Y} : \Omega_1 \times \Omega_2 \to \mathbb{R}$ is defined as:

$$p_{X,Y}(r_1, r_2) = F\left(X^{-1}(r_1), Y^{-1}(r_2)\right)$$

where F is the joint distribution on $\Omega_1 \times \Omega_2$.

Example

$\Omega_1 , \Omega_2 = \{H, T\}$

$\Omega_1 \times \Omega_2 = \{ (H,H), (H,T), (T,H), (T,T) \}.$

F is defined by the following table:

$\Omega_1 \backslash \Omega_2$	H	T	
H	0.3	0.4	0.7
T	0.1	0.2	0.3
	0.4	0.6	

marginal

$= P_1(H)$

$= P_1(T)$

Conditional

$P_{\Omega_2|\Omega_1=H} \left(\frac{0.3}{0.7}, \frac{0.4}{0.7}\right)$

$P_{\Omega_2|\Omega_1=T} \left(\frac{0.1}{0.3}, \frac{0.2}{0.3}\right)$

marginal $\xrightarrow{P_2}$ $P_2(H)$ $P_2(T)$

Conditional $\rightarrow$ $(0.3/0.4,\ 0.1/0.4)$ $P_{\Omega_1|\Omega_2=H}$ $(0.4/0.6,\ 0.2/0.6)$ $P_{\Omega_1|\Omega_2=T}$