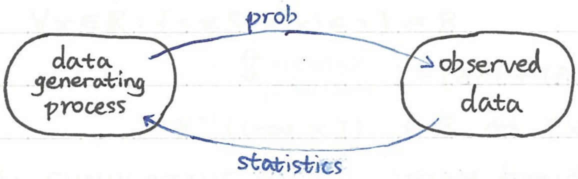
Statistics

Course is based on [CB] Statistical inference.

Probability and Statistics



I Probability

I.1 Probability space

Ingredients

1) $S = \Omega$ is a set called the SAMPLE SPACE set of all outcomes of an experiment
2) Any subset $A \subset S$ is called an EVENT
3) An EVENT SPACE $\mathcal{B} = \mathcal{F}$ is a collection of events satisfying :

$$\left.\begin{array}{l} 1)\ \phi \text{ and } S \in \mathcal{B} \quad \phi \text{ is the empty set} \\ 2)\ \text{If } A \in \mathcal{B} \text{ then } A^c = S \setminus A \ \text{(S minus A or the complement of A in S)} \in \mathcal{B} \\ 3)\ \text{If } A_j \in \mathcal{B} \text{ for } j = 1,2,3,\cdots \ \text{(infinite but countable family)} \text{ then } \bigcup_{j=1}^{\infty} A_j \in \mathcal{B} \end{array}\right\} \overset{\text{sigma}}{\sigma}\text{ - ALGEBRA}$$

Recall : $\left.\begin{array}{l}(A \cup B)^c = A^c \cap B^c \\ (A \cap B)^c = A^c \cup B^c\end{array}\right\}$ De Morgan's Law

Examples

1) If $S = \{a_1, a_2, \cdots, a_n\}$ then $\mathcal{B} = \{\text{all subset of } S\}$ power set of S
2) If $S = \mathbb{R}$, then we choose $\mathcal{B}$ as the family of subsets generated by intervals $(a,b)\ \forall a<b$

$\mathcal{B}$ is called the BOREL ALGEBRA of $\mathbb{R}$ denoted by $\sigma_{\mathcal{B}}$.

(We can do the same for $[a,b]$)

$\Rightarrow \exists$ some subsets of S to which we cannot give a weight

4) a PROBABILITY FUNCTION "WEIGHT" $P = \mathbb{P}$ is a map $P : \mathcal{B} \mapsto \mathbb{R}$ such that (s.t.)

$$\left.\begin{array}{l} 1)\ P(A) \geq 0 \quad \forall A \in \mathcal{B} \\ 2)\ P(S) = 1 \text{ and } P(\phi) = 0 \\ 3)\ \text{If } A_j \cap A_k = \phi \ \forall j,k \text{ then } P\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P(A_j) \end{array}\right\} \text{axioms of probability OR Holmogorov's axioms}$$

Propositions

1) $P(A) \leq 1$ and $P(A^c) = 1 - P(A) \quad \forall A \in \mathcal{B}$

2) $P(A \cup B) = P(A) + P(B) - P(A \cap B) \iff$

$P(A \cap B) = P(A) + P(B) - P(A \cup B) \geq P(A) + P(B) - 1$ Bonfevioni's inequality

3) If $B \subset A$ then $P(B) \leq P(A)$

4) If $C_j \in \mathcal{B}$ s.t. $\bigcup_{j=1}^{\infty} C_j = S$ and $C_j \cap C_k = \phi \ \forall j, k$ partition of S

then $P(A) = \sum_{j=1}^{\infty} P(A \cap C_j) \ \forall A \in \mathcal{B}$

Def. (S sample space , $\mathcal{B}$ event space , P probability function) is called a PROBABILITY SPACE.

Exercise: Find the number of arrangement of n objects from S. p. 14~15 [CB]

	without replacement	with replacement
ordered		
unordered		

Ex. 1.2.20

Def. If $A, B \in \mathcal{B}$ and if $P(B) > 0$

we define the CONDITIONAL PROBABILITY of A given B by

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

Observe that
$$P(A \cap B) = P(A|B)P(B)$$
$$\shortparallel$$
$$P(B \cap A) = P(B|A)P(A)$$
$$\Rightarrow \quad P(A|B) = P(B|A)\frac{P(A)}{P(B)}$$

Example of 2 dice

$A = \{\text{black dice is } 1\}$; $B = \{\text{red dice is } 1\}$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/36}{1/6} = \frac{1}{6}$$

More generally : (Bayes Rule) Exercise: check for its application

If $\{C_j\}_{j=1}^{\infty}$ is a partition of S

$$P(C_j | B) = \frac{P(B|C_j)P(C_j)}{\underbrace{\sum_k \underbrace{P(B|C_k)P(C_k)}_{= P(B \cap C_k)}}_{= P(B)}} \quad \text{if } P(B) \neq 0$$

Def. Let $A, B \in \mathcal{B}$, they are INDEPENDENT if $P(A \cap B) = P(A)P(B)$

Def. A collection of events $A_1, A_2, \cdots \in \mathcal{B}$ are MUTUALLY INDEPENDENT if

$$\forall A_{i_1}, A_{i_2}, \cdots, A_{i_n} : P\left(\bigcap_{j=1}^{n} A_{i_j}\right) = \prod_{j=1}^{n} P(A_{i_j})$$

I.2 Random variable

Def. a RANDOM VARIABLE X on a probability space $(S, \mathcal{B}, P)$

is a function $X : S \mapsto \mathbb{R}$ satisfying

$$\forall x \in \mathbb{R} : \{ s \in S \mid X(s) \leq x \} \in \mathcal{B}$$

$$\Updownarrow$$

$$X^{-1}((-\infty, x]) \in \mathcal{B} \iff \{ X \leq x \} \in \mathcal{B}$$

Def. The CUMULATIVE DISTRIBUTION FUNCTION (CDF) associated to X

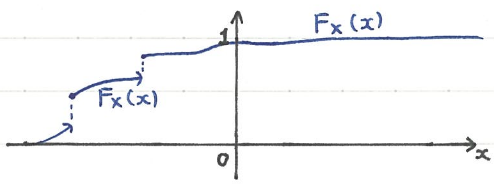
is the function $F_X : \mathbb{R} \mapsto [0,1]$ defined for any $x \in \mathbb{R}$ by

$$F_X(x) := P(\{ X \leq x \})$$

Properties

1) $F_X(x) \leq F_X(y)$ if $x \leq y$ (increasing / monotonic non-decreasing)
2) $\lim\limits_{x \to -\infty} F_X(x) = 0$; $\lim\limits_{x \to \infty} F_X(x) = 1$
3) $\lim\limits_{\varepsilon \searrow 0} F_X(x + \varepsilon) = F_X(x)$ (right continuous)

One has



Thm. Whenever $F : \mathbb{R} \mapsto [0,1]$ satisfies properties 1) ~ 3),

there exists $(S, \mathcal{B}, P)$ and a random variable X such that

$$F_X = F.$$

⚠ non-unique