1. Fundamental of Probability (cont'd)

1.4 Independence

Def: Two events E and F are said to be independent, if $P(E \cap F) = P(E) \cdot P(F)$; denoted as $E \perp F$. **This is different rom disjoint.**

Assume $P(F) > 0$, then $E \perp F \Leftrightarrow P(E|F) = P(E)$; intuitively, knowing F does not change the probability of E.

Proof:

$$\begin{aligned} E \perp F &\Leftrightarrow P(E \cap F) = P(E) \cdot P(F) \\ &\Leftrightarrow \frac{P(E \cap F)}{P(F)} = P(E) \\ &\Leftrightarrow P(E|F)) = P(E) \end{aligned}$$

More generally, a sequence of events $E_1, E_2, \ldots$ are called independent if for **any** finite index set I,

$$P(\bigcap_{i \in I} E_i) = \prod_{i \in I} P(E_i)$$

1.5 Bayes's rule and law of total probability

Theorem: Let $F_1, F_2, \ldots$ be disjoint events, and $\bigcap_{i=1}^{\infty} F_i = \Omega$, we say $\{F_u\}_{i=1}^{\infty}$ forms a "partition" of the sample space Ω

Then $P(E) = \sum_{i=1}^{\infty} P(E|F_i) \cdot P(F_i)$

Proof: Exercise

Intuition: Decompose the total probability into different cases.

$$P(E \cap F_2) = P(E|F+2) \cdot P(F_2)$$

Bayes' rule

$$P(F_i|E) = \frac{P(E|F_i) \cdot P(F_i)}{\sum_{h=1}^{\infty} P(E|F_j) \cdot P(F_j)}$$

Bayes' rule tells us how to find conditional probability by switching the role of the event and the condition.

Proof:

$$\begin{aligned} P(F_i|E) &= \frac{P(F_i \cap E)}{P(E)} && \text{definition of condition probability} \\ &= \frac{P(E|F_i)P(F_i)}{P(E)} \\ &= \frac{P(E|F_i)P(F_i)}{\sum_{j=1}^{\infty} P(E|F_j)P(F_j)} && \text{law of total probability} \end{aligned}$$

2 Random variables and distributions

2.1 Random variables

(Ω, ξ, P): Probability space.

Def: A random variable X (or r.v.) is a mapping from Ω to R

$$\begin{aligned} X: &\Omega \to R \\ &\omega \to X(\omega) \end{aligned}$$

A random variable transforms arbitrary "outcomes" into numbers.

X introduces a probability on R. For $A \subseteq R$, define

$$\begin{aligned} P(X \in A) :&= P(\{X(\omega) \in A\}) \\ &= P(\{\omega : X(\omega) \in A\}) \\ &= P(X^{-1}(A)) \end{aligned}$$

From now on, we can often "forget" te original probability space and focus on the random variables and their distributions.

Def: let X be a random variable. The **CDF**(cumulative distribution function) F of X is defined by

$$F(x) = P(X \leq x) = P(X \in (-\infty, x]))$$
$$X: \text{random variable}, x: \text{number}$$

Properties of cdf:

1. F is non-decreasing. $F(x_1) \leq F(x_2), x_1 < x_2$
2. limits
   - $\lim_{x \to -\infty} F(x) = 0$
   - $\lim_{x \to \infty} F(x) = 1$
3. $F(x)$ is right continuous

- $lim_{x \downarrow a} F(x) = F(a) : x$ decreases to a(approaching from the right)
- Hint: $\{x \leq a\} = \bigcap_{i=1}^{\infty} \{X \leq a_i\}$ for $a_i \downarrow a$

2.2 Discrete random variables and distribution

A random variable X is called **discrete** if it only takes values in an **at most countable** set $\{x_1, x_2, ...\}$ (finite or countable).

The distribution of a discrete random variable is fully characterized by its **probability mass function**(p.m.f)

$$p(x) := P(X = x); x = x_1, x_2, ...$$

Properties of pmf:

1. $p(x) \geq 0 \qquad \forall x$
2. $\sum_i p(x_i) = 1$

Q: what does the cdf of a discrete random variable look like?

Examples of discrete distributions

1.Bemoulli distribution

$$p(1) = P(X = 1) = p$$
$$p(c) = P(X = c) = 1 - p$$
$$p(x) = 0 otherwise$$

Denote $X \sim Ber(p)$

2.Binomial distribution

$$p(k) = P(X = k) = (^n_k)p^k(1-p)^{n-k}$$

- $X \sim Bin(n, p)$ to choose k successes.
- Binomial distribution is the distribution of number of successes in n independent trials; each having probability p of suucess.

3.Geometric distribution

$$p(k) = P(X = k) = (1-p)^{k-1}p$$
$$(1-p)^{k-1} \text{ : the first k-1 trials are all failures}, p \text{ : success in kth trial}$$

- $X \sim Geo(p)$

- X is the number of trials needed to get the first success in n independent trials with probability p of success each
- X has the memoryless property $P(X > n + m | X > m) = P(x > n) \qquad n, m = 0, 1, \ldots$