Lecture 2

Introduction to probability & stats I

- Basic notation:

- Probability density function : $h(x)$
- Cumulative density function : $H(x)$

- Probability of "x" within an interval "dx" is given by "$h(x)\,dx$".
- May drop the "dx" when referring to "probability of measuring x is $p(x)$" and assume it is implicit.

★ **Kolmogorov Axioms :-**

- Given an event A, we can assign a prob. to that event $p(A)$, called "probability of A".

- This $p(A)$ <u>must</u> satisfy three Kolmogorov axioms :

① $p(A) \geq 0$ for each/all A.

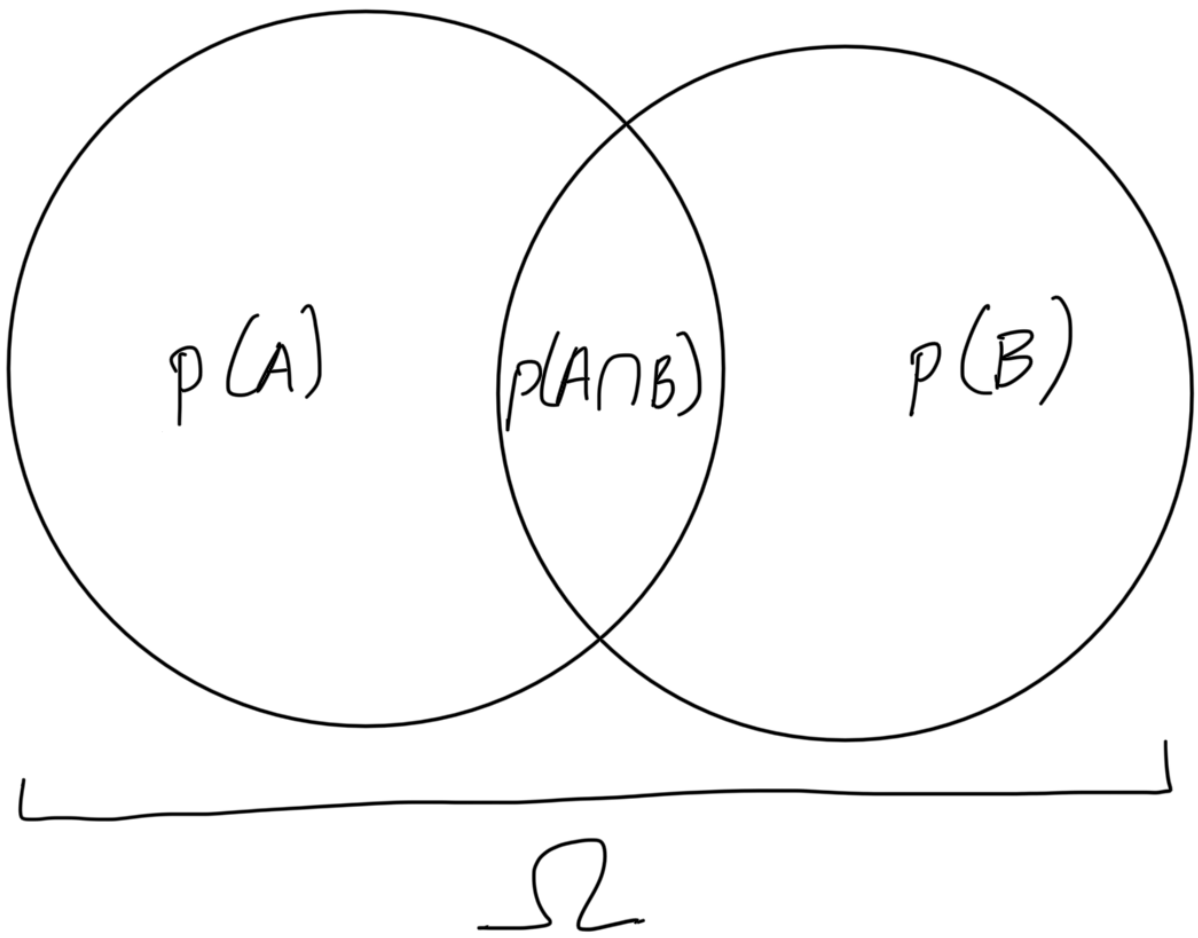
② $p(\Omega) = 1$, where Ω is a set of all possible outcomes.

③

③ If $A_1, A_2, \ldots$ are disjoint events, then,

$$p\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} p(A_i),$$

where $\bigcup$ stands for "union".



- Consequences of Kolmogorov axioms:

• $$\boxed{p(A \cup B) = p(A) + p(B) - p(A \cap B)}$$

→ here $\bigcap$ represents "intersection" of sets of A & B events.

• If the complement of A is $\bar{A}$ (i.e. "not A"), then,

$$\boxed{p(A) + p(\bar{A}) = 1}$$

• The probability that both A & B happen

- The probability that both A & B happen,

$$\boxed{p(A \cap B) = p(A|B)\,p(B) = p(B|A)\,p(A)}$$

- A consequence of the last relation is that if B_i, with $i = 1, 2, \ldots N$ are disjoint events, and their union is the set of all possible outcomes, then,

$$\boxed{\begin{aligned} p(A) &= \sum_i p(A \cap B_i) \\ &= \sum_i p(A|B_i)\,p(B_i) \end{aligned}}$$

• This expression is called as the "law of total probability".

- Alternate interpretation:

$\cup \Leftrightarrow$ "OR" $\qquad \cap \Leftrightarrow$ "AND"

Random variables

- Defined as a variable whose value results from a measurement of a quantity subject to random variations.

• Rather than taking on a single value, RVs take on multiple values associated with a probability.

- Two types:
  1) Discrete: take values from a countable set (e.g. rolls of a die/dice)
  2) Continuous: map onto real numbers. (or complex, or matrices).

- Independent identically distributed (iid) variables are drawn from the same distribution and are independent.

- Two random variables <u>are independent only if</u>:

$$p(x, y) = p(x)\, p(y)$$

probability of obtaining x <u>AND</u> y

for all values of x & y.

- The <u>data</u> are specific ("measured") values of RVs.

- Measured values $\Leftrightarrow x_i$
- Set of all measurements $\Leftrightarrow \{x_i\}$

<u>Bayes' Rule</u>

- When two RVs are <u>not</u> independent,

$$p(x,y) = p(y|x)\,p(x) = p(x|y)\,p(y) \quad -(1)$$

- The marginal probability is defined as,

$$\boxed{p(x) = \int p(x,y)\, dy}$$

- Combining these two equations, we get

$$p(x) = \int p(x|y)\, p(y)\, dy$$

- Re-arranging Eq. (1) & using the above eq^n, we get the Bayes' rule

$$\boxed{p(y|x) = \frac{p(x|y)\, p(y)}{p(x)} = \frac{p(x|y)\, p(y)}{\int p(x|y)\, p(y)\, dy}}$$

- For discrete RVs,

$$p(y_i|x) = \frac{p(x|y_i)\, p(y_i)}{p(x)} = \frac{p(x|y_i)\, p(y_i)}{\sum_{j=1}^{M} p(x|y_j)\, p(y_j)}$$

- The Bayes' rule forms the foundation of Bayesian statistics.

Transformation of random variables

- For a given RV, x, with probability density, $p(x)$,

• if $y = \phi(x)$
such that $x = \phi^{-1}(y)$

then $$p(y) = p\left[\phi^{-1}(y)\right] \frac{d\phi^{-1}(y)}{dy}$$

- For example, if $y = e^{x}$, then

$$x = \phi^{-1}(y) = \ln(y)$$

if $p(x)$ is uniform between $0 \leq x \leq 1$,
then $p(y) = 1 \cdot \frac{1}{y} = \frac{1}{y}$ for $1 \leq y \leq e$

• Thus, a uniform distribution of x is transformed into a non-uniform distribution in y.