Basic Probability Notes

Terminology:

The probability

1. of A, denoted as $P(A)$ is a real number between 0 and 1 inclusive.

   $P(A) = 0$ means that A is false.
   $P(A) = 1$ means that A is true.
   $0 < P(A) < 1$ correspond to varying degrees of certainty.

2. The joint probability of A and B, denoted as $P(A, B)$ is the prob that both A and B are true.

   $$P(A,B) = P(B,A)$$

3. The conditional probability of A given B, denoted as $P(A|B)$ is the prob we would assign to A if we knew B to be true.

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

   Rearranging the above eqn, we get:
   $P(A,B) = P(A|B) \cdot P(B) \leftarrow$ Product Rule

   Similarly, we have

   $$P(B|A) = \frac{P(B,A)}{P(A)} \longrightarrow P(B,A) = P(B|A) \cdot P(A)$$

   $$\longrightarrow P(A,B) = P(B|A) \cdot P(A)$$

4. Sum Rule: $P(A) + P(\bar{A}) = 1$
The prob of a statement being true and the prob of a statement being false sum to 1.

Now, suppose we have a set of mutually exclusive statements, A_i, exactly one of which must be true, we have:

$$\sum_i P(A_i) = 1$$

5. Conditioning Rule: $P(A|B) + P(\bar{A}|B) = 1$

More Probability Formulas:

1. $P(A|B)P(B) + P(\bar{A}|B)P(B) = P(B)$

Proof:

$$\begin{aligned} LHS &= P(A|B)P(B) + P(\bar{A}|B)P(B) \\ &= P(B)\big(\underbrace{P(A|B) + P(\bar{A}|B)}_{=1 \text{ By conditioning rule}}\big) \\ &= P(B) \\ &= RHS \end{aligned}$$

2. $P(A, B) + P(\bar{A}, B) = P(B)$

Proof:

$$\begin{aligned} LHS &= P(A, B) + P(\bar{A}, B) \\ &= P(A|B)P(B) + P(\bar{A}|B)P(B) \\ &= P(B) \quad \text{(See above eqn)} \\ &= RHS \end{aligned}$$

3. $\sum_i P(A_i | C) = 1$

Here, the A_i's are a mutually exclusive set, exactly one of which must be true.

4. $P(A, B | C) = P(A | B, C) \cdot P(B | C)$

5. $P(B) = \sum_i P(A_i | B) \leftarrow$ Marginalization

Independence:

- 2 statements, A and B are independent iff $P(A, B) = P(A) \cdot P(B)$.

- Furthermore, if A and B are independent, then $P(A|B) = P(A)$.

Proof:

$$
\begin{aligned}
LHS &= P(A|B) \\
&= \frac{P(A, B)}{P(B)} \\
&= \frac{P(A) \cdot \cancel{P(B)}}{\cancel{P(B)}} \\
&= P(A) \\
&= RHS
\end{aligned}
$$

Random Variable:

- A random variable (r.v.) is a variable taking on numerical values determined by the outcome of a random phenomenon.

- A discrete random variable has a countable number of possible values.

- A continuous random variable takes on all the values in some interval of numbers.

- Discrete random variables use Probability Mass Function (PMF) to describe their distributions.

The notation $P_X(x)$ refers to the PMF of the r.v. X.
$P_X(x) = P(X=x)$

Properties of PMFs:
1. $0 \le P_X(x) \le 1$ (PMFs are always btwn 0 and 1, inclusive)
2. $\sum_{-\infty}^{\infty} P_X(x) = \sum_{x \in X} P_X(x) = 1$

- Continuous r.v. use Probability Density Function (PDF) to describe their distributions.

- We use the notation $f_X(x)$ to refer to the PDF of a r.v. X.

Properties of PDFs:
1. $0 \le f_X(x)$
2. $\int_{-\infty}^{\infty} f_X(x)\,dx = \int_{x \in X} f_X(x)\,dx = 1$
3. $P(a \le x \le b) = \int_a^b f_X(x)\,dx$

- For discrete random variables, the expected value, denoted as $E(x)$, is:

$$E(x) = \sum_i P(r_i)\, x_i$$

where r_i is the outcome of x_i.

The variance denoted as $Var(x)$ is:

$$Var(x) = \sum_{i=1}^{n} P(r_i) \cdot (x_i - \mu)^2$$

where μ is the expected value
I.e. $\mu = \sum_i P(r_i)\, x_i$

The standard deviation, denoted as σ, is:

$$\sigma = \sqrt{Var(x)}$$

- For continuous random variables:

$E(x) = \int x p(x)\, dx$ where $p(x)$ is the PDF.

$Var(x) = \int x^2 p(x)\, dx - \mu$ where $p(x)$ is the PDF and μ is the expected value.

$$\sigma = \sqrt{Var(x)}$$