Probability

A measure of the likelihood that a particular event will occur, expressed as a number with the symbol P between 0 and 1. The higher the probability, the more likely the event is to occur.

Odds

The ratio of the probability of an event occurring to the probability of the event not occurring. Often expressed as a ratio, such as 3:2, which means the event is three times as likely to occur as not to occur.

Log-odds

The logarithm of the odds, it transform odds into a continuous, unbounded scale. This is useful when working with logistic regression or other machine learning algorithms.

Random Variable

For a given sample space S of some experiment, a random variable (commonly written with X) is any rule that associates a number with each outcome in S. In mathematical language, a random variable is a function whose domain is the sample space and whose range is the set of real numbers. Any random variable whose only possible values are 0 and 1 is called a Bernoulli random variable.

Expectation

- Random variables, by definition, result in different outcomes
- The variation in random variables is captured by their distribution
    - Probability Mass function for Discrete variables
    - Probability Density function for Continuous variables

Expectation gives mean/average/expected value of the random variable given the distribution. It is the weighted average of all possible values of a random variable, with weights being the probabilities of each outcome. It represents the predicted long-term average value of the variable.

- The **expectation**, or **expected value**, of some function $f(x)$ with respect to a probability distribution $P(x)$ is the average value of $f(x)$ when x is drawn from P

Denoted by $\mathbb{E}_{x \sim P}[f(x)]$

- If P is clear from the context $\mathbb{E}_x[f(x)]$
- If x is also clear from the context $\mathbb{E}[f(x)]$
- Sometimes, simply denoted as $\mathbb{E}[f]$

- Mathematically,

$$\mathbb{E}_{x\sim P}[f(x)] = \sum_x P(x)f(x)$$

$$\mathbb{E}_{x\sim P}[f(x)] = \int_x p(x)f(x)dx$$

Multivariate Expectation

- For a multivariate random variable x we can interpret the variable and the expectations by considering each component separately

That is, if $x = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_D \end{bmatrix} \in \mathbb{R}^D$ then

$$\mathbb{E}_x[f(x)] = \begin{bmatrix} \mathbb{E}_{x_1}[f(x_1)] \\ \mathbb{E}_{x_2}[f(x_2)] \\ \dots \\ \mathbb{E}_{x_D}[f(x_D)] \end{bmatrix}$$

Linearity of Expectation

- Important Property of expectation

The Expectation Operator is linear

- Mathematically, if $f(x) = \alpha g(x) + \beta h(x)$ is a multivariate function with $\alpha,\ \beta \in \mathbb{R}$ being scalars, then

$$\mathbb{E}[f] = \alpha\mathbb{E}[g] + \beta\mathbb{E}[h]$$

Variance

- Random variables, by definition, result in different outcomes
- The variation in random variables is captured by their distribution
    - Expectation gives mean/average/expected value of the random variable given the distribution
- Variance gives the *variation from the* expected *value*
- Variance also measures amount of fluctuation of the variable
- Higher variance indicates greater variability in the dataset from the mean or expected value.
- Standard deviation is the square root of the variance, providing a measure of dispersion that is on the same scale as the original data. It is useful for comparing variability across different datasets or distributions.

Univariate variance

- The **variance** and its square root, **standard deviation** of some function $f(x)$ with respect to a probability distribution $P(x)$ measure how much the value of $f(x)$ varies for various samples when x is drawn from P

Denoted by $\mathbb{V}_{x \sim P}[f(x)]$

- If P is clear from the context $\mathbb{V}_x[f(x)]$
- If x is also clear from the context $\mathbb{V}[f(x)]$
- Usually, simply denoted as $\mathbb{V}[f]$ or Var$[f]$

- Mathematically,

$$\mathbb{V}[f(x)] = \mathbb{E}[(f(x) - \mathbb{E}[f(x)])^2] = \mathbb{E}[\left(f(x) - \overline{f(x)}\right)^2]$$

The standard deviation is given as $\sigma[x] = \sqrt{\mathbb{V}[x]}$

Covariance (univariate)

- A measure of the joint variability of two random variables, indicating how they change together. Positive covariance indicates that the variables tend to increase or decrease together, while negative covariance indicates that one variable tends to increase when the other decreases.
- Note that $\mathbb{V}[x] = \mathbb{E}[(x - \overline{x})^2] = \mathbb{E}[(x - \overline{x})(x - \overline{x})]$
- This notion can be generalized to a pair of variables x and y to find the **covariance**
$$\text{Cov}[x, y] = \mathbb{E}[(x - \overline{x})(y - \overline{y})] = \mathbb{E}[(x - \mathbb{E}[x])(y - \mathbb{E}[y])]$$
- Similarly, $\text{Cov}[f(x), g(y)] = \mathbb{E}[(f(x) - \mathbb{E}[f(x)])(g(y) - \mathbb{E}[g(y)]\ )]$
- A related quantity is the correlation, defined as
$$\text{corr}[x, y] = \frac{\text{Cov}[x, y]}{\sqrt{\mathbb{V}[x]\mathbb{V}[y]}}$$
- Measures how linearly correlated the two random variables are
- Note $\text{Cov}[x, x] = \text{Var}[x]$ and $\text{corr}[x, x] = 1$

Interpreting Covariance and correlation

- Positive covariance means when x increases, y is expected to increase too.
- Negative covariance means when x increases, y is expected to decrease
- Correlation close to 1 means strongly, positively correlated
- Correlation close to -1 means strongly, negatively correlated
- Correlation close to 0 means no (linear) correlation
- We can use Pearson's correlation coefficient understand the relations between features (Commonly available in Python)

Covariance Useful Equations

- $\text{Cov}[x, y] = \mathbb{E}[(x - \overline{x})(y - \overline{y})] = \mathbb{E}[(x - \mathbb{E}[x])(y - \mathbb{E}[y])]$
- This can be simplified as $\text{Cov}[x, y] = \overline{xy} - \overline{x}\,\overline{y} = \mathbb{E}[xy] - \mathbb{E}[x]\mathbb{E}[y]$

Independence and covariance

- When x, y are two independent random variables, $\text{Cov}[x, y] = 0$
- When x, y are two independent random variables, $\text{Cov}[x, y] = 0$
- The converse is not true
- That is, $\text{Cov}[x, y] = 0$ need not mean x, y are independent

- That is, covariance is zero even though the variables are not independent
- Zero covariance only means that there is no *linear* relationship
- Independence ⇒ Zero Covariance but Zero Covariance ≠ Independence

Covariance Matrix

- If $\boldsymbol{x} \in \mathbb{R}^n$ is a vector, then it is often useful to know pairwise covariances between all pairs of components
    - Think of the *x* being the components of an input vector such as pixels of an image
- That is, define $\text{Cov}[\boldsymbol{x}, \boldsymbol{x}]_{i,j} = \text{Cov}[\boldsymbol{x}_i, \boldsymbol{x}_j]$

$$\text{Cov}[x, x] = \begin{bmatrix} Cov[x_1, x_1] & \text{Cov}[x_1, x_2] & \cdots & \text{Cov}[x_1, x_n] \\ Cov[x_2, x_1] & \text{Cov}[x_2, x_2] & \dots & \text{Cov}[x_2, x_n] \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}[x_n, x_1] & \text{Cov}[x_n, x_2] & \cdots & \text{Cov}[x_n, x_n] \end{bmatrix} \in \mathbb{R}^{n \times n}$$

- Note that the diagonal elements are simply the variances of individual components, that is, $\text{Cov}[x_i, x_i] = \text{Var}[x_i]$

Independence

- **Independent random variables** – Two random variables ***X*** and ***Y*** are said to be *statistically independent* if and only if $p(x, y) = p(x)p(y)$
    - More precisely, X and *Y* are independent iff
      $p(X = x, Y = y) = p(X = x)p(Y = y) \quad \forall x \in X,\ y \in Y$
    - Examples
        - Independent – X: Throw of a dice, Y: Toss of a coin
        - Not independent – X: Height, Y: Weight
- Independence is equivalent to saying

$p(y|x) = p(y)$ OR $p(x|y) = p(x)$

- Can be seen from product rule $p(x, y) = p(y|x)p(x) = p(x)p(y)$
  $\Rightarrow p(y|x) = p(y)$
- Independence is denoted by $x \perp y$