# Basic Math & Algebra

SM in AI Tutorial - 1

# Expectation

Let $X$ be a random variable with a finite number of finite outcomes $x_1, x_2, \ldots, x_k$ occurring with probabilities $p_1, p_2, \ldots, p_k$, respectively. The **expectation** of $X$ is defined as

$$\mathrm{E}[X] = \sum_{i=1}^{k} x_i\, p_i = x_1 p_1 + x_2 p_2 + \cdots + x_k p_k.$$

Since all probabilities $p_i$ add up to 1 ($p_1 + p_2 + \cdots + p_k = 1$), the expected value is the weighted average, with $p_i$'s being the weights.

# Basic Properties

- **Non-negativity:** If $X \geq 0$ (a.s.), then $\operatorname{E}[X] \geq 0$.
- **Linearity of expectation:** The expected value operator (or **expectation operator**) $\operatorname{E}[\cdot]$ is linear in the sense that, for any random variables $X$ and $Y$, and a constant $a$,

$$\begin{aligned}\operatorname{E}[X+Y] &= \operatorname{E}[X] + \operatorname{E}[Y],\\ \operatorname{E}[aX] &= a\operatorname{E}[X],\end{aligned}$$

whenever the right-hand side is well-defined. This means that the expected value of the sum of any finite number of random variables is the sum of the expected values of the individual random variables, and the expected value scales linearly with a multiplicative constant.

# Variance

The variance of a random variable $X$ is the expected value of the squared deviation from the mean of $X$, $\mu = \mathrm{E}[X]$:

$$\operatorname{Var}(X) = \mathrm{E}\left[(X-\mu)^2\right].$$

This definition encompasses random variables that are generated by processes that are discrete, continuous, neither, or mixed. The variance can also be thought of as the covariance of a random variable with itself:

$$\operatorname{Var}(X) = \operatorname{Cov}(X, X).$$

The variance is also equivalent to the second cumulant of a probability distribution that generates $X$. The variance is typically designated as $\operatorname{Var}(X)$, $\sigma_X^2$, or simply $\sigma^2$ (pronounced "sigma squared"). The expression for the variance can be expanded:

$$\begin{aligned}\operatorname{Var}(X) &= \mathrm{E}\left[(X-\mathrm{E}[X])^2\right]\\ &= \mathrm{E}\left[X^2 - 2X\,\mathrm{E}[X] + \mathrm{E}[X]^2\right]\\ &= \mathrm{E}\left[X^2\right] - 2\,\mathrm{E}[X]\,\mathrm{E}[X] + \mathrm{E}[X]^2\\ &= \mathrm{E}\left[X^2\right] - \mathrm{E}[X]^2\end{aligned}$$

In other words, the variance of $X$ is equal to the mean of the square of $X$ minus the square of the mean of $X$. This equation should not be used for computations using floating point arithmetic because it suffers from catastrophic cancellation if the two components of the equation are similar in magnitude. There exist numerically stable alternatives.

# Derivaties & Chain rule

**Basic Properties and Formulas**

If $f(x)$ and $g(x)$ are differentiable functions (the derivative exists), $c$ and $n$ are any real numbers,

1. $(cf)' = c f'(x)$
2. $(f \pm g)' = f'(x) \pm g'(x)$
3. $(fg)' = f'g + fg'$ – **Product Rule**
4. $\left(\frac{f}{g}\right)' = \frac{f'g - fg'}{g^2}$ – **Quotient Rule**
5. $\frac{d}{dx}(c) = 0$
6. $\frac{d}{dx}(x^n) = n x^{n-1}$ – **Power Rule**
7. $\frac{d}{dx}\left(f\left(g(x)\right)\right) = f'\left(g(x)\right) g'(x)$

   This is the **Chain Rule**

# Higher Order Derivatives

**Higher Order Derivatives**

The Second Derivative is denoted as

$f''(x) = f^{(2)}(x) = \dfrac{d^2 f}{dx^2}$ and is defined as

$f''(x) = \left(f'(x)\right)'$, *i.e.* the derivative of the first derivative, $f'(x)$.

The n$^{\text{th}}$ Derivative is denoted as

$f^{(n)}(x) = \dfrac{d^n f}{dx^n}$ and is defined as

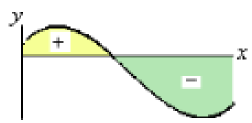
$f^{(n)}(x) = \left(f^{(n-1)}(x)\right)'$, *i.e.* the derivative of the $(n\text{-}1)^{\text{st}}$ derivative, $f^{(n-1)}(x)$.

# Application of Integrals

Calculus Cheat Sheet

## Applications of Integrals

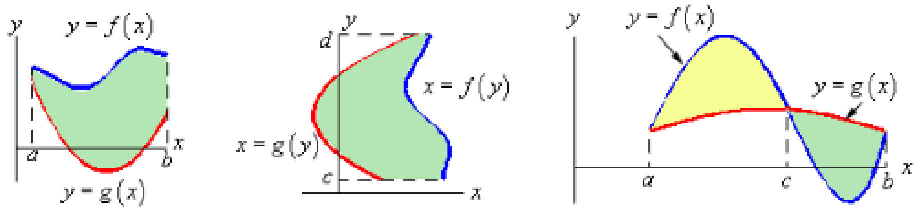
**Net Area :** $\int_a^b f(x)\,dx$ represents the net area between $f(x)$ and the $x$-axis with area above $x$-axis positive and area below $x$-axis negative.




**Area Between Curves :** The general formulas for the two main cases for each are,

$y = f(x) \Rightarrow A = \int_a^b [\text{upper function}] - [\text{lower function}]\,dx$ & $x = f(y) \Rightarrow A = \int_c^d [\text{right function}] - [\text{left function}]\,dy$

If the curves intersect then the area of each portion must be found individually. Here are some sketches of a couple possible situations and formulas for a couple of possible cases.




$A = \int_a^b f(x) - g(x)\,dx$ $\quad$ $A = \int_c^d f(y) - g(y)\,dy$ $\quad$ $A = \int_a^c f(x) - g(x)\,dx + \int_c^b g(x) - f(x)\,dx$

# Practice Questions

- Find expected value of [ 1.0, 2.0, 3.0, 4.0, 5.0, 6.0 ]
- Expected number of trails to get a one head in a coin flip
- Find Eigen values & vectors of

$$\begin{pmatrix} -2 & -4 & 2 \\ -2 & 1 & 2 \\ 4 & 2 & 5 \end{pmatrix}.$$