2.2 & 2.3 MATHEMATICAL EXPECTATION:

A very important concept in probability and statistics is that of the mathematical expectation, expected value, or briefly the expectation, of a random variable. Originally, the concept of a mathematical expectation arose in connection with games of chance, and in its simplest form it is the **product** of the amount a player stands to **win** and the **probability** that he or she will win. For instance, if we hold one of $\mathbf{10,000}$ tickets in a raffle for which the grand prize is a trip worth $\$\mathbf{4,800}$, our mathematical expectation is $\mathbf{4,800}.\frac{\mathbf{1}}{\mathbf{10,000}} = \$\mathbf{0.48}$.

This amount will have to be interpreted in the sense of an average - altogether the $\mathbf{10,000}$ tickets pay $\$\mathbf{4,800}$, or on the average $\frac{\$\mathbf{4,800}}{\mathbf{10,000}} = \$\mathbf{0.48}$ per ticket.

Definition: If $\boldsymbol{X}$ is a discrete random variable and $\boldsymbol{f}(\boldsymbol{x})$ is the value of its probability distribution at $\boldsymbol{x}$, the expected value of $\boldsymbol{X}$ is $\boldsymbol{\mu} = \boldsymbol{E}(\boldsymbol{X}) = \sum \boldsymbol{x}\boldsymbol{f}(\boldsymbol{x})$ **or,** $\boldsymbol{E}(\boldsymbol{u}(\boldsymbol{X})) = \sum \boldsymbol{u}(\boldsymbol{x})\boldsymbol{f}(\boldsymbol{x})$

Example: Imagine a game in which, on any play, a player has a $\mathbf{20}\%$ chance of winning $\$\mathbf{3}$ and an $\mathbf{80}\%$ chance of losing $\$\mathbf{1}$. The probability distribution function of the random variable $\boldsymbol{X}$, the amount won or lost on a single play is

$\boldsymbol{x}$	$\$\mathbf{3}$	$-\$\mathbf{1}$
$\boldsymbol{f}(\boldsymbol{x})$	$\mathbf{0.2}$	$\mathbf{0.8}$

and so the average amount won (actually lost, since it is negative) - in the long run – is

$$\boldsymbol{\mu} = \boldsymbol{E}(\boldsymbol{X}) = \sum \boldsymbol{x}\boldsymbol{f}(\boldsymbol{x}) = \mathbf{3} \times \boldsymbol{f}(\mathbf{3}) + (-\mathbf{1}) \times \boldsymbol{f}(-\mathbf{1}) = \mathbf{3} \times \mathbf{0.2} - \mathbf{0.8} = -\mathbf{0.20}$$

What does "in the long run" mean? If you play, are you guaranteed to lose no more than 20 cents?

Solution: If you play and lose, you are guaranteed to lose $\$\mathbf{1}$. An expected loss of $\mathbf{20}$ cents means that if you played the game over and over and over and over…again, the average of your $\$\mathbf{3}$ winnings and your $\$\mathbf{1}$ losses would be a $\mathbf{20}$ cent loss. "In the long run" means that you can't draw conclusions about one or two plays, but rather thousands and thousands of plays.

Example: A lot of $\mathbf{12}$ television sets includes $\mathbf{2}$ with white cords. If ***three*** of the sets are chosen at random for shipment to a hotel, how many sets with **white cords** can the shipper expect to send to the hotel?

Solution: Since $\boldsymbol{x}$ of the two sets with white cords and $(\mathbf{3} - \boldsymbol{x})$ of the $\mathbf{10}$ other sets can be chosen in $\binom{\mathbf{2}}{\boldsymbol{x}}\binom{\mathbf{10}}{\mathbf{3}-\boldsymbol{x}}$ ways, three of the $\mathbf{12}$ sets can be chosen in $\binom{\mathbf{12}}{\mathbf{3}}$ ways, and these $\binom{\mathbf{12}}{\mathbf{3}}$ possibilities are presumably equiprobable, we find that the probability distribution of $\boldsymbol{X}$, the number of sets with white cords shipped to the hotel, is given by

$$\boldsymbol{f}(\boldsymbol{x}) = \frac{\binom{\mathbf{2}}{\boldsymbol{x}}\binom{\mathbf{10}}{\mathbf{3}-\boldsymbol{x}}}{\binom{\mathbf{12}}{\mathbf{3}}}, \text{for } \boldsymbol{x} = \mathbf{0}, \mathbf{1}, \mathbf{2}$$

or, in tabular form,

$\boldsymbol{x}$	$\mathbf{0}$	$\mathbf{1}$	$\mathbf{2}$
$\boldsymbol{f}(\boldsymbol{x})$	$\frac{\mathbf{6}}{\mathbf{11}}$	$\frac{\mathbf{9}}{\mathbf{22}}$	$\frac{\mathbf{1}}{\mathbf{22}}$

Now, $\boldsymbol{E(X) = \sum_{x=0}^{2} xf(x) = 0.\frac{6}{11} + 1.\frac{9}{22} + 2.\frac{1}{22} = \frac{1}{2}}$
and since half a set cannot possibly be shipped, it should be clear that the term "expect" is not used in its colloquial sense. Indeed, it should be interpreted as an average pertaining to repeated shipments made under the given conditions.

Example: A saleswoman has been offered a new job with a fixed salary of $\$\boldsymbol{290}$. Her records from her present job show that her weekly commissions have the following probabilities:

Commission ($\boldsymbol{x}$)	$\boldsymbol{0}$	$\$\boldsymbol{100}$	$\$\boldsymbol{200}$	$\$\boldsymbol{300}$	$\$\boldsymbol{400}$
Probability, $\boldsymbol{f(x)}$	$\boldsymbol{0.05}$	$\boldsymbol{0.15}$	$\boldsymbol{0.25}$	$\boldsymbol{0.45}$	$\boldsymbol{0.1}$

Should she change jobs?

Solution: The expected value (expected income) from her present job is

$$\boldsymbol{E(X) = \sum xf(x) = 0 \times 0.05 + 100 \times 0.15 + 200 \times 0.25 + 300 \times 0.45 + 400 \times 0.1 = \$240}$$

If she is making a decision only on the amount of money earned, then clearly she ought to change jobs.

Properties of Expectation:

- **(i)** If $\boldsymbol{X \geq 0}$, then $\boldsymbol{E(X) \geq 0}$
- **(ii)** For any real numbers $\boldsymbol{a, b}$ & $\boldsymbol{c}$, $\boldsymbol{E(aX + bY + c) = aE(X) + bE(Y) + c}$
- **(iii)** If $\boldsymbol{c}$ is a constant, then $\boldsymbol{E(c) = c}$
- **(iv)** If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are two independent random variables and both $\boldsymbol{E[|X|]}$ and $\boldsymbol{E[|Y|]}$ are finite, then $\boldsymbol{E(XY) = E(X)E(Y)}$

Now, $\boldsymbol{u_i}$ is the distance of that $\boldsymbol{i}$-th point from the origin. In mechanics, the product of a distance and its weight is called a **moment**, so $\boldsymbol{u_i f(u_i)}$ is a moment having a moment arm of length $\boldsymbol{u_i}$. The sum of such products would be the moment of the system of distances and weights. Actually, it is called the **first moment** about the **origin**, since the distances are simply to the first power and the lengths of the arms (distances) are measured from the origin. However, if we compute the first moment about the mean $\boldsymbol{\mu}$, then, since here a moment arm equals $\boldsymbol{(x - \mu)}$, we have $\boldsymbol{\sum(x - \mu)f(x) = E(X - \mu) = E(X) - E(\mu) = \mu - \mu = 0}$

That is, that first moment about $\boldsymbol{\mu}$ is equal to zero. In mechanics $\boldsymbol{\mu}$ is called the centroid.

Variance, $\boldsymbol{\sigma^2}$: The variance of a random variable $\boldsymbol{X}$ refines our knowledge of the probability distribution of $\boldsymbol{X}$ by giving a broad measure of how $\boldsymbol{X}$ is dispersed around its mean. It is the **second moment** about mean $\boldsymbol{\mu}$ because the distances are raised to the second power, and defined as: $\boldsymbol{\sigma^2 = Var(X) = E[(X - \mu)^2] = \sum_{i=1}^{n} (x_i - \mu)^2 f(x_i)}$

Standard Deviation (S.D): The **positive square root** of the variance is called the **standard deviation** of $\boldsymbol{X}$ and is denoted by the Greek letter $\boldsymbol{\sigma}$ (sigma). The standard deviation is a statistic that looks at how far from the mean a group of numbers is, (measures the dispersion of a dataset relative to its mean). If the points are further from the mean, there is a higher deviation within the data; if they are closer to the mean, there is a lower deviation. So the more spread out the group of numbers, the higher standard deviation.

$$\textbf{S.D.}, \boldsymbol{\sigma} = \sqrt{\boldsymbol{Var}(\boldsymbol{X})} = \sqrt{\boldsymbol{E}[(\boldsymbol{X}-\boldsymbol{\mu})^2]} = \sqrt{\sum_{i=1}^{n}(\boldsymbol{x_i}-\boldsymbol{\mu})^2 \boldsymbol{f}(\boldsymbol{x_i})}$$

Properties of Variance:

- **(i)** For any real numbers $\boldsymbol{a}$ & $\boldsymbol{b}$, $\boldsymbol{Var}(\boldsymbol{aX}+\boldsymbol{b}) = \boldsymbol{a}^2\boldsymbol{Var}(\boldsymbol{X})$
- **(ii)** $\boldsymbol{Var}(\boldsymbol{X}) = \boldsymbol{E}[\boldsymbol{X}^2] - \boldsymbol{E}[\boldsymbol{X}]^2 = \boldsymbol{E}[\boldsymbol{X}^2] - \boldsymbol{\mu}^2$
- **(iii)** If $\boldsymbol{X}$ and $\boldsymbol{Y}$ have finite variance and are independent, then

$$\boldsymbol{Var}(\boldsymbol{X}+\boldsymbol{Y}) = \boldsymbol{Var}(\boldsymbol{X}) + \boldsymbol{Var}(\boldsymbol{Y})$$

Examples: 2.2-1 to 2.2-6 (See yourself)
Exercises: 2.2-1 to 2.2-5 (Try yourself)

Geometric distribution: The geometric distribution gives the probability that the first occurrence of success requires $\boldsymbol{x}$ independent trials, each with success probability $\boldsymbol{p}$. If the probability of success on each trial is $\boldsymbol{p}$, then the probability that the $\boldsymbol{x}$-th trial (out of $\boldsymbol{x}$ trials) is the first success is $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{P}(\boldsymbol{X}=\boldsymbol{x}) = (\boldsymbol{1}-\boldsymbol{p})^{\boldsymbol{x}-\boldsymbol{1}}\boldsymbol{p} = \boldsymbol{q}^{\boldsymbol{x}-\boldsymbol{1}}\boldsymbol{p}$, where $\boldsymbol{q} = \boldsymbol{1}-\boldsymbol{p},$ for $\boldsymbol{x} = \boldsymbol{1}, \boldsymbol{2}, \boldsymbol{3}, \dots$

Mean, $\boldsymbol{\mu} = \boldsymbol{E}(\boldsymbol{X}) = \sum \boldsymbol{x}\boldsymbol{f}(\boldsymbol{x}) = \frac{\boldsymbol{1}}{\boldsymbol{p}}$ and Variance, $\boldsymbol{\sigma}^2 = \boldsymbol{Var}(\boldsymbol{X}) = \boldsymbol{E}[\boldsymbol{X}^2] - \boldsymbol{\mu}^2 = \frac{\boldsymbol{q}}{\boldsymbol{p}^2}$

Moment Generating Function:

Let $\boldsymbol{X}$ be a random variable of the discrete type with ***pmf*** $\boldsymbol{f}(\boldsymbol{x})$ and space $\boldsymbol{S}$. If there is a positive number $\boldsymbol{h}$ such that

$$\boldsymbol{E}(\boldsymbol{e}^{\boldsymbol{tX}}) = \sum_{\boldsymbol{x}\in\boldsymbol{S}} \boldsymbol{e}^{\boldsymbol{tx}}\boldsymbol{f}(\boldsymbol{x})$$

exists and is finite for $-\boldsymbol{h} < \boldsymbol{t} < \boldsymbol{h},$ then the function defined by $\boldsymbol{M}(\boldsymbol{t}) = \boldsymbol{E}(\boldsymbol{e}^{\boldsymbol{tX}})$ is called the **moment-generating function of** $\boldsymbol{X}$ (or of the distribution of $\boldsymbol{X}$). This function is often abbreviated as ***mgf***.

Thus, ***mgf*** $\quad \boldsymbol{M}(\boldsymbol{t}) = \boldsymbol{E}(\boldsymbol{e}^{\boldsymbol{tX}}) = \sum_{\boldsymbol{x}\in\boldsymbol{S}} \boldsymbol{e}^{\boldsymbol{tx}}\boldsymbol{f}(\boldsymbol{x})$

In particular, if the moment-generating function exists, then

$\boldsymbol{M}'(\boldsymbol{0}) = \boldsymbol{E}(\boldsymbol{X}) = \boldsymbol{\mu} =$ Mean and $\boldsymbol{M}''(\boldsymbol{0}) - [\boldsymbol{M}'(\boldsymbol{0})]^2 = \boldsymbol{E}[\boldsymbol{X}^2] - \boldsymbol{\mu}^2 = \boldsymbol{\sigma}^2 = \boldsymbol{Var}(\boldsymbol{X})$

Examples: 2.3-1 to 2.3-3 & 2.3-5 to 2.3-7 (See yourself)
Exercises: 2.3-1 to 2.3-17 (Try yourself)