Probability

Discrete Random Variable [Mean and Variance]

Prof. Dr. Mostafa Elhosseini

https://youtube.com/drmelhosseini

Sample Mean

- The number of flaws in a 1-inch length of copper wire manufactured by a certain process varies from wire to wire. Overall, **48%** of the wires produced have **no flaws**, **39%** have **one fl**aw, **12%** have **two flaws**, and **1%** have **three flaws**. Let X be the **number of flaws** in a randomly selected piece of wire. Now **imagine** that we have a **sample of 100 wires**, and that the sample follows this distribution perfectly
- What is the **sample me**an?

Sample Mean

$$\text{Mean} = \frac{0(48) + 1(39) + 2(12) + 3(1)}{100} = 0.66$$

This can be rewritten as

$$\text{Mean} = 0(0.48) + 1(0.39) + 2(0.12) + 3(0.01) = 0.66$$

Sample Mean

- The **population mean** of a discrete random variable can be thought of as the **mean of a hypothetical sample** that follows the probability distribution perfectly
- The **mean of a perfect sample** can be obtained by **multiplying each possible value of X by its probability and summing the products**. This is the definition of the **population mean of a discrete random variable**

Population Mean

- The population mean of a random variable X may also be called the **expectation**, or **expected value**, of X, and can be denoted by μ_X, by $E(X)$, or simply by μ.
- Sometimes we will drop the word "population," and simply refer to the **population mean** as the **mean**

Population Mean

- Let X be a discrete random variable with probability mass function $p(x) = P(X = x)$.

The mean of $\boldsymbol{X}$ is given by $\boldsymbol{\mu_X = \sum_x xP(X = x)}$

where the sum is over all possible values of X.

Example - 01

- A certain industrial process is brought down for recalibration whenever the quality of the items produced falls below specifications. Let X **represent the number of times the process is recalibrated during a week and** assume that X has the following probability mass function.

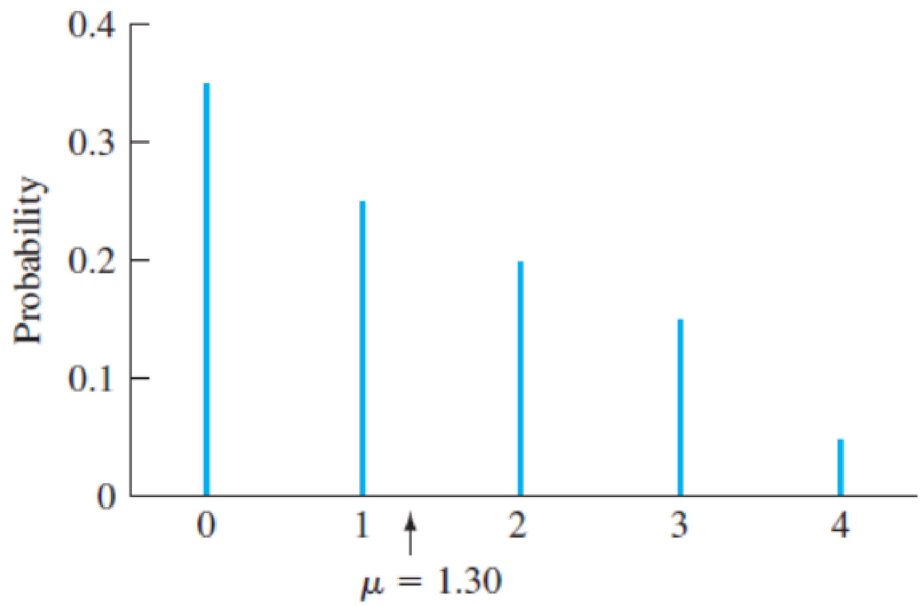
  Find the mean of X.

x	0	1	2	3	4
$p(x)$	0.35	0.25	0.20	0.15	0.05

Example - 01

$$\mu_X = 0(0.35) + 1(0.25) + 2(0.20) + 3(0.15) + 4(0.05) = 1.30$$

The population mean has an important physical interpretation. It is the horizontal component of the center of mass of the probability mass function; that is, it is the point on the horizontal axis at which the graph of the probability mass function would balance if supported there. Figure 2.9 illustrates this property for the probability mass function described in Example 2.36, where the population mean is $\mu = 1.30$.



Recall: Sample variance

- Recall that for a sample $X_1, \ldots, X_n$, the **sample variance** is given by

$$\sum (X_i - \bar{X})^2 / (n+1)$$

- The **sample variance** is thus essentially the average of the squared differences between the sample points and the sample mean (except that we divide by $n - 1$ instead of n).

Population Variance

- By analogy, the **population variance of a discrete random** variable X is a weighted average of the squared differences $(x - \mu_X)^2$ where x ranges through all the possible values of the random variable X. This weighted average is computed by **multiplying each squared difference $(\boldsymbol{x} - \boldsymbol{\mu_X})^2$ by the probability $\boldsymbol{P(X = x)}$ and summing the results**
- The population variance of a random variable X can be denoted by $\boldsymbol{\sigma_X^2}$, by $\boldsymbol{V(X)}$ or simply by $\boldsymbol{\sigma^2}$

Population Variance

- $\boldsymbol{\sigma_X^2} = \sum_x (x - \mu_X)^2 P(X = x)$
- $\boldsymbol{\sigma_X^2} = \sum_x x^2 P(X = x)\, \mu_X^2$
- We also define the **population standard deviation** to be the **square root of the population variance**.
- We denote the population standard deviation of a random variable X by $\boldsymbol{\sigma_X}$ or simply by $\boldsymbol{\sigma}$.
- As with the mean, we will sometimes drop the word "population," and simply refer to the **population variance and population standard deviation** as the **variance** and **standard deviation**, respectively

Example - 02

- Find the **variance and standard deviation** for the random variable X described in Example - 01, representing the number of times a process is recalibrated

x	0	1	2	3	4
$p(x)$	0.35	0.25	0.20	0.15	0.05

References

- Statistics for Engineers and Scientists, Fourth Edition, William Navidi Colorado School of Mines
- Khan Academy