Lesson 3

Sample, Population, Variance, and Probabilistic Distribution

Population

- A group of data, which includes all the data of interest, For example:
- The population = 1000 when 1000 students gave examination of a computer course, and their results are being analyzed.
- = 100 when 100 Chess games represent the results of matches

Sample

- Represents a subset of the population
- Represents the population (dataset of interest) for uses, such as analysis
- Sample (subset) consists of randomly selected members from the dataset members at the population
- Results from the sample considered as the results of the population

Random Variable

- A variable whose possible values are outcomes of a random phenomenon
- Is a function that maps the outcomes of unpredictable processes to numerical quantities
- Also called stochastic variable or random quantity

Randomness

- Can be around some expected mean value or outcome, and with some normal deviation

Variance

- Variance measure is the sum of squares of the difference in values of a variable with respect to the expected value.
- The variance is also a measure of dispersion with respect to the expected value.

Large and Small Variances

- Large Variance—
- Data points varying greatly from the mean value in a dataset
- Small variance when small variations around the mean, expected or centroid average

Standard Deviation, σ

- A measure of variance σ given by Equations 6.1 a and b
- Find by taking the sum of squares of the difference in values of a variable with respect to the expected value
- Divide by N− 1 (N equals number of Sample or population values or data points) and find square root
-

Standard Error, σ_{est}

- The standard error estimate is a measure of the accuracy of predictions from a relationship.
- Sum of squares of the difference in values of a variable with respect to the expected value divided by N
- Standard error is square root of the sum (Equation 6.2)

σ_{est}

- Small means most of the observed values (y) dots are fairly close to the fitting line in the scatter plot, and better is the estimate based on the equation of the line..

σ_{est}

- Large, many of the observed values are far away from the line.
- Zero, then no variation exists corresponding to the computed line for predictions
- 0 means the correlation between the observed and estimation is perfect.

Probability

- Chance or occurrence of some event, condition, or value
- Chance of winning the match P_w= 78% for a grandmaster in Chess, losing P_l = 12% and P_{draw} = 10%.
- P (x) is discrete three valued function, P_w , P_l and P_{draw}
-

Probability

- Chance of a student scoring grade pointer, 9.0 out of 10 is the $P_{9.0}$.
- P_{GP} , the grade point can vary, continuously between two end points.
- Assume it varies such that maximum P is say at 8.0 and distribution may be between 9.9 and 6.0.

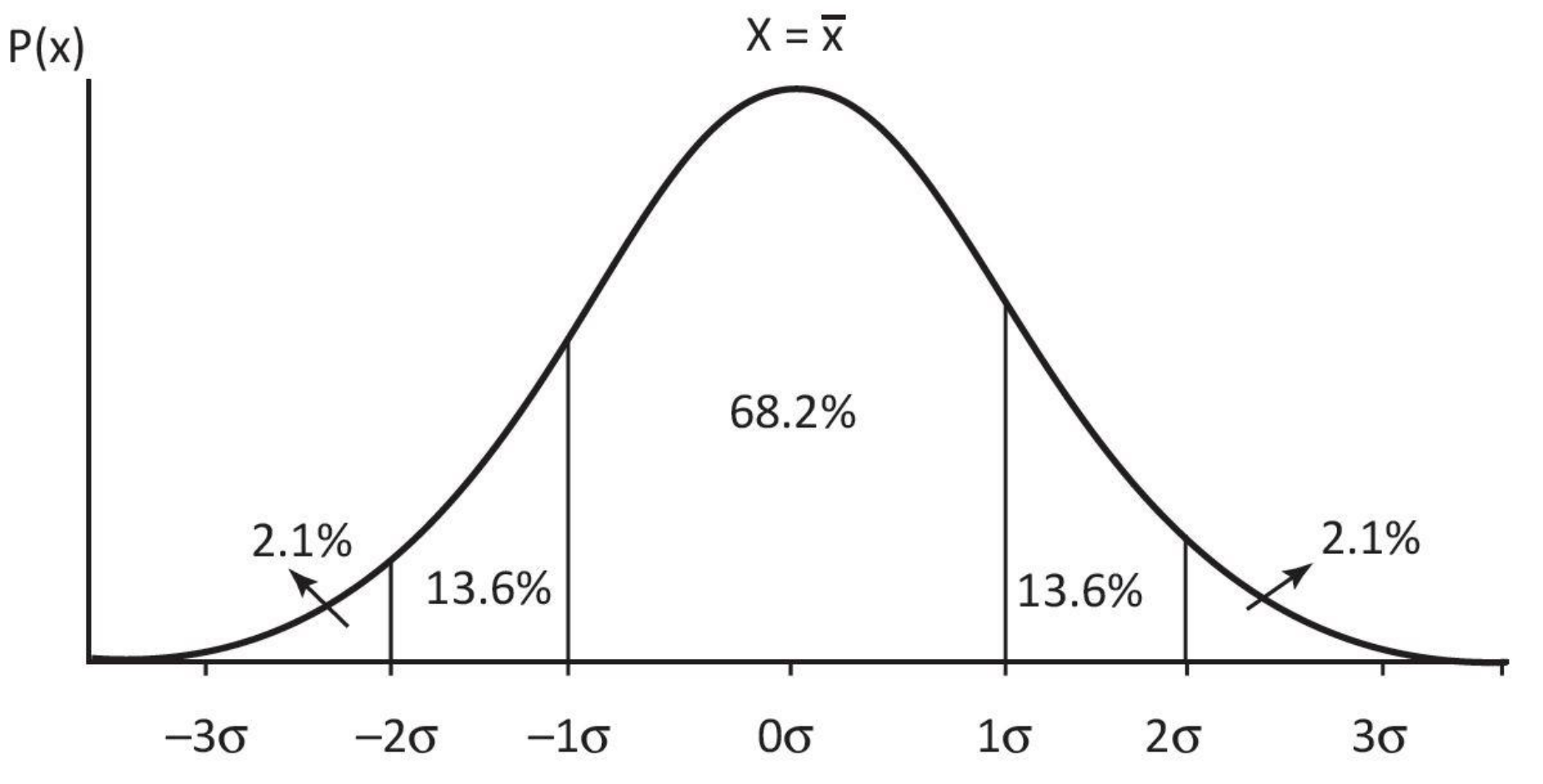
Probability Distribution Function P(x)

- Total sum P(x) for all possible randomly distributed values of x = 1.
- For example, Sum of P_{GP} for all possible values say between 6.0 and 9.9 is $\sum P_{GP(x)} = 1$
- Total Probability, P distributes as a function P(x) depending on of independent variable, x

Normal Probability Distribution

- Normal distribution, P(x) given by Gaussian function which shows a bell shape curve (equation 6.3)
- Maximum P(x) is at value $x_{0,}$ the most probable , mean, expected or average of all possible x values

Figure 6.3 Probability distribution function as a function of x assuming normal distribution around x = $\bar{x}$ and standard deviation = σ



Normal Probabilistic Distribution

- P(x) decreases symmetrically as x increases from x_0 or x decreases from x0
- The PDF of the normal distribution is such that
- 68% of area under the PDF is within ($\bar{x} + \sigma$) and ($\bar{x} - \sigma$)

Normal PDF

- 95% of area under the PDF within x= ($\bar{x} + 2\sigma$) and ($\bar{x} - 2\sigma$) and
- 99.7% is within x = ($\bar{x} + 3\sigma$) and ($\bar{x} - 3\sigma$).

Summary

We learnt meanings of:

- Population and Sample
- Probability
- Probability Distribution Function (PDF)
- Standard deviation σ
- Normal PDF features

End of Lesson 3 on Sample, Population, Variance, and Probabilistic Distribution