SWAYAM NPTEL COURSE ON

MINE AUTOMATION AND DATA ANALYTICS

By

Prof. Radhakanta Koner

Department of Mining Engineering

Indian Institute of Technology (Indian School of Mines) Dhanbad

Module 8: Inferential Statistics

Lecture 20 B: Continuous Random Variable Part II

CONCEPTS COVERED

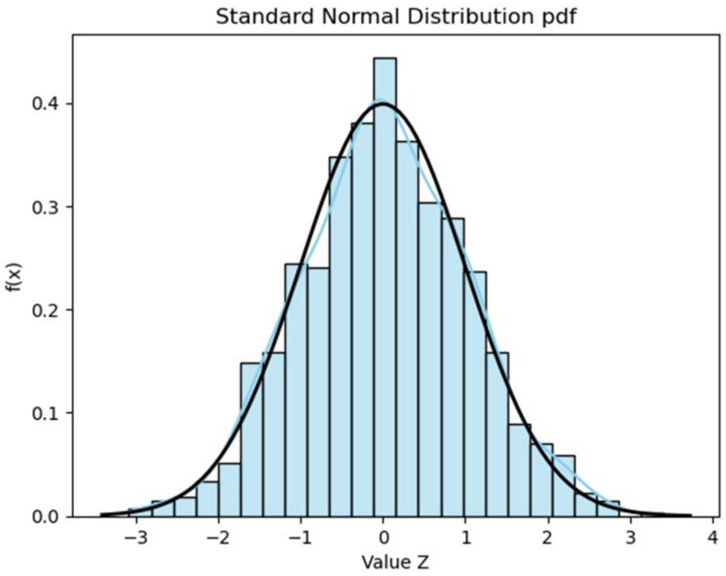
1. Standard Normal Distribution
2. T distribution
3. Chi-Squared Distribution

Standard Normal Distribution
Z ~ N(0,1)

A normal continuous random variable with an expected value (mean) of 0 and a standard deviation of 1 is referred to as a standard normal continuous random variable.

The density curve associated with it is termed the standard normal curve.

This type of random variable is symbolized by Z.



Standard Normal Distribution
$Z \sim N(0,1)$

Z-Score Formula

The statistical formula for a value's z-score is calculated using the following formula:

$z = (x - \mu) / \sigma$

Where:

z = Z-score

x = the value being evaluated

μ = the mean

σ = the standard deviation

Standard Normal Distribution
$Z \sim N(0,1)$

How to Calculate Z-Score

Calculating a z-score requires that you first determine the mean and standard deviation of your data. Once you have these figures, you can calculate your z-score. So, assume you have the following variables:

$x = 57$

$\mu = 52$

$\sigma = 4$

You would use the variables in the formula:

$z = (57 - 52) / 4$

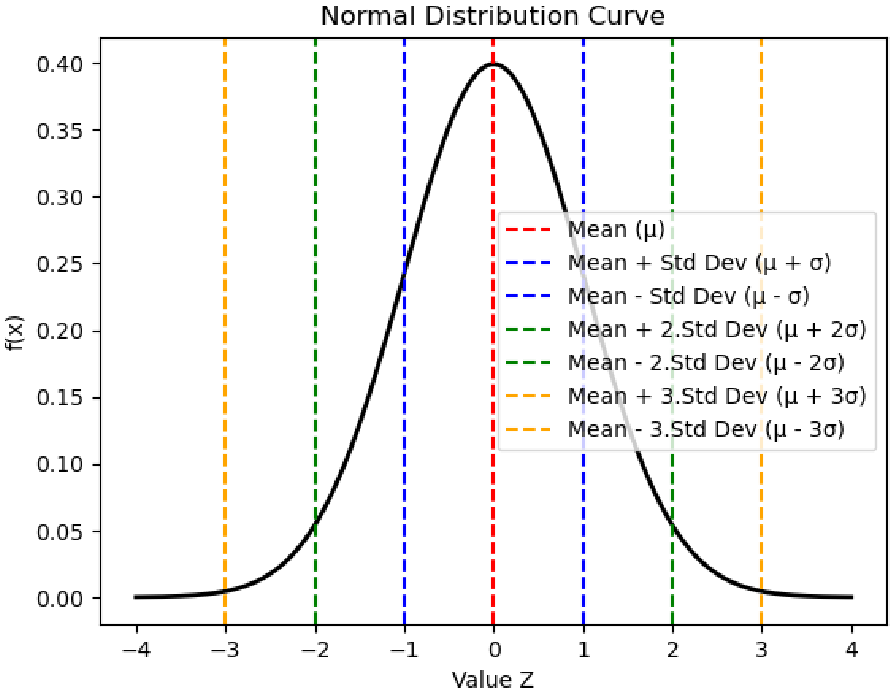
$z = 1.25$

So, your selected value has a z-score that indicates it is 1.25 standard deviations from the mean.

Approximate Rule to Standard Normal Distribution
$Z \sim N(0,1)$

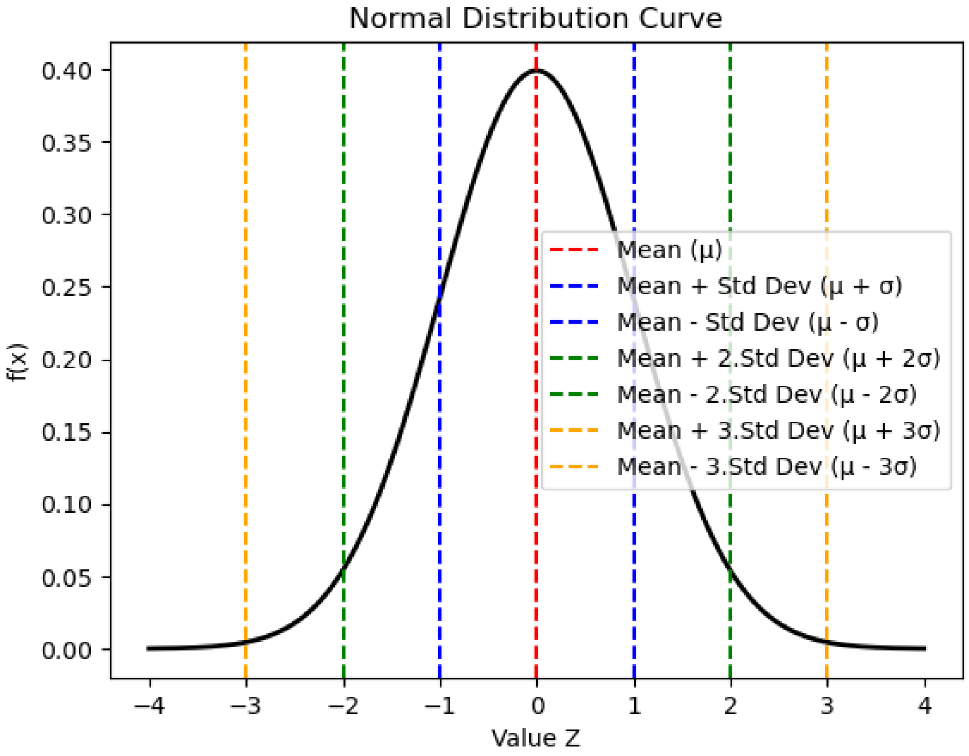
For a continuous random variable following a normal distribution with mean μ and standard deviation σ:

- **Between μ + σ and μ - σ, approximately 68% of the distribution is covered.**
- **Between μ + 2σ and μ - 2σ, approximately 95% of the distribution is encompassed.**
- **Between μ + 3σ and μ - 3σ, approximately 99.7% of the distribution is included.**



Approximate Rule to Standard Normal Distribution
Z ~ N(0,1)

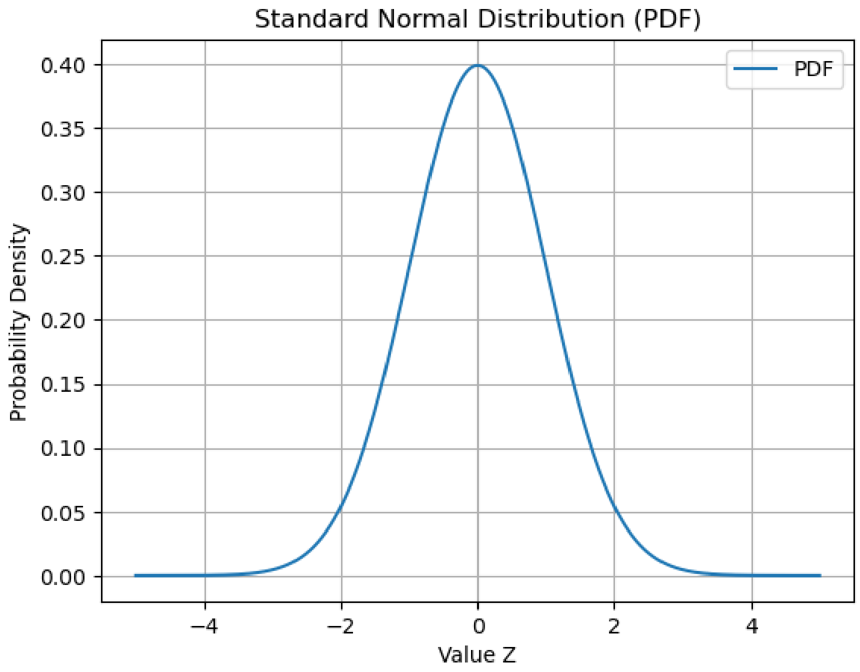
1) P(-1 < Z < 1) = ? 0.68
2) P (-2 < Z < 2) = ? 0.95
3) P(Z > -1) = ? 0.84
4) P(Z < -3) = ? 0
5) P(Z < 2) = 0.975



Properties of Z

1) 1^{st} Property

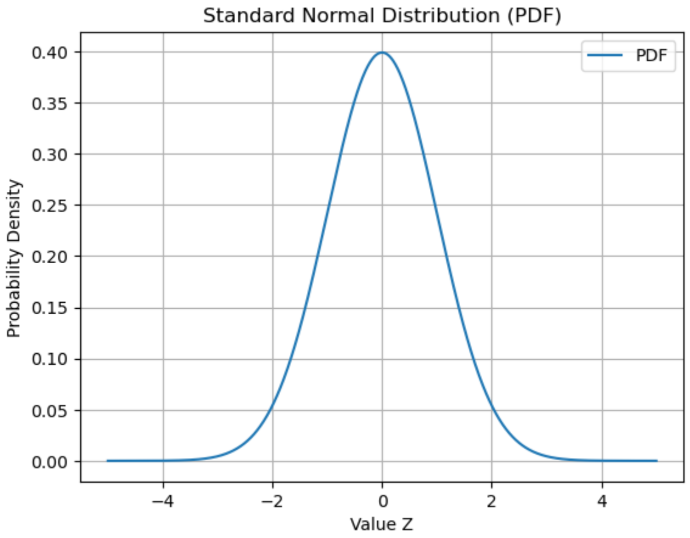
$$P\,(Z < -x) = 1 - P(Z < x)$$



Properties of Z

1) 2^{nd} Property

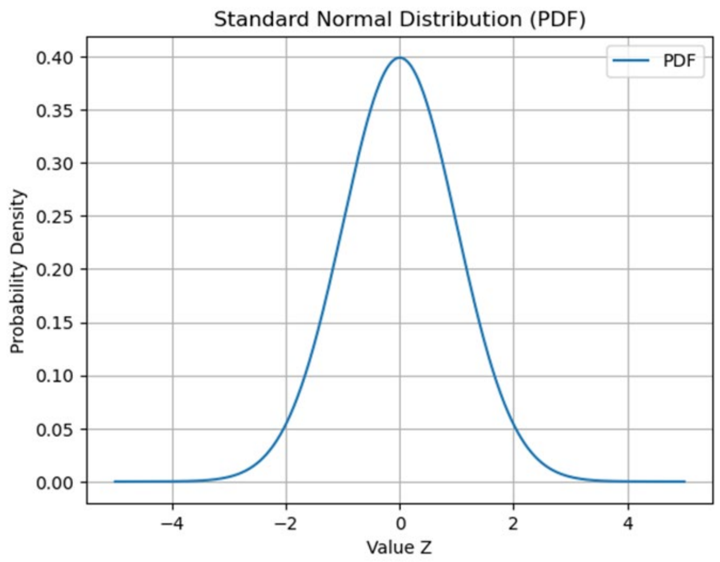
$$P\,(a < Z < b) = P(Z < b) - P(Z < a)$$



Properties of Z

1) 3rd Property

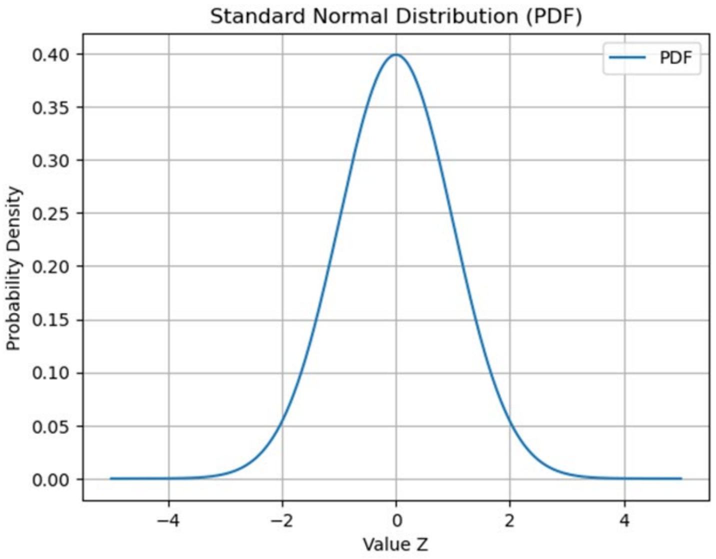
$$(|Z| > a) = 2(1 - P(Z < a))$$



Properties of Z

1) 4th Property

$$(\,|Z| < a\,) = 2\,P(\,Z < a\,) - 1$$



Standard Normal Table

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142	.57535
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91309	.91466	.91621	.91774
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
2.0	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361

Standard Normal Table

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142	.57535
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91309	.91466	.91621	.91774
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
2.0	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361

1. $P(Z < 1.09) = 0.86214$

2. $P(Z > 2.1) = 0.01786$

3. $P(0 < Z < 1)$
$= P(Z < 1) - P(Z < 0)$
$= 0.84134 - 0.5$
$= 0.34134$

Standard Normal Table

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142	.57535
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91309	.91466	.91621	.91774
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
2.0	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361

1. $P(-0.9 < Z < 2.3) = P(Z < 2.3) - P(Z < -0.9)$
   $= P(Z < 2.3) - P(Z > 0.9)$
   $= 0.98928 - (1 - P(Z < 0.9))$
   $= 0.98928 - (0.18406)$
   $= 0.80522$

2. $P(Z > -0.96) = P(Z < 0.96)$

3. $P(Z < -0.53) = P(Z > 0.53)$

Standard Normal Table

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142	.57535
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91309	.91466	.91621	.91774
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
2.0	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361

1. $P(|Z| < 0.2) = 2.\ P(Z < 0.2) - 1$

What Is a T-Distribution?

- The t-distribution, also known as the Student's t-distribution. is a type of probability distribution that is similar to the normal distribution with its bell shape but has heavier tails.
- It is used for estimating population parameters for small sample sizes or unknown variances.
- T-distributions have a greater chance for extreme values than normal distributions, and as a result have fatter tails.
- The t-distribution is the basis for computing t-tests in statistics

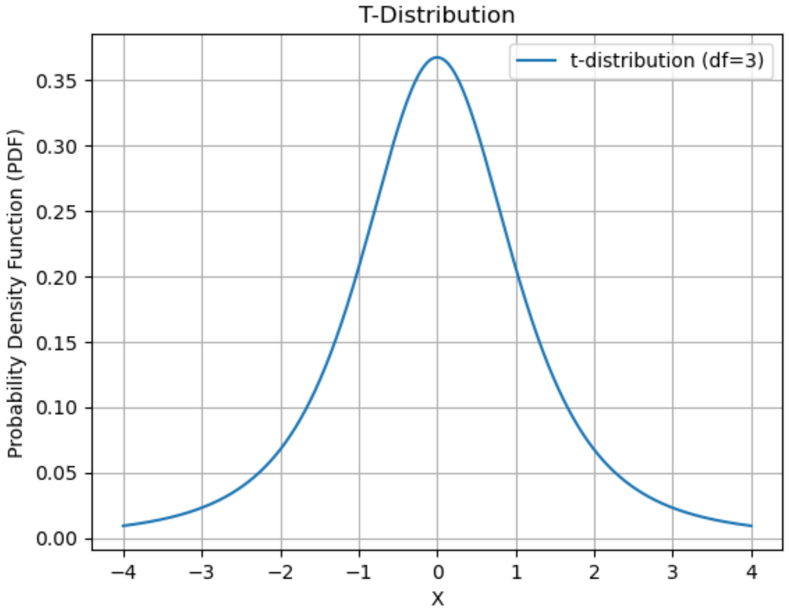
KEY TAKEAWAYS

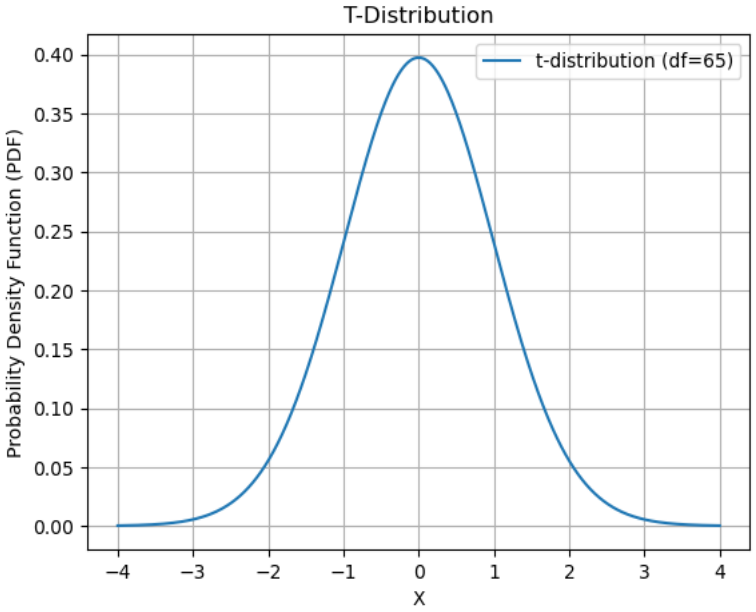
- The t-distribution is a continuous probability distribution of the z-score when the estimated standard deviation is used in the denominator rather than the true standard deviation.
- The t-distribution, like the normal distribution, is bell-shaped and symmetric, but it has heavier tails, which means that it tends to produce values that fall far from its mean.
- T-tests are used in statistics to estimate significance.

What Is a T-Distribution?

What Does a T-Distribution Tell You?

Tail heaviness is determined by a parameter of the t-distribution called degrees of freedom, with smaller values giving heavier tails, and with higher values making the t-distribution resemble a standard normal distribution with a mean of 0 and a standard deviation of 1.





What Is a T-Distribution?

When a sample of n observations is taken from a normally distributed population having mean (M) and standard deviation (D), the sample mean (m) and the sample standard deviation (d) will differ from M and D because of the randomness of the sample.

A z-score can be calculated with the population standard deviation as :

$$Z = (x - M)/D$$

The value Z has the normal distribution with mean 0 and standard deviation 1.

But when using the estimated standard deviation, a t-score is calculated as :

$$T = (m - M)/\{d/sqrt(n)\}$$

and the difference between d and D makes the distribution a t-distribution with (n - 1) degrees of freedom rather than the normal distribution with mean 0 and standard deviation 1.

T-Distribution vs. Normal Distribution

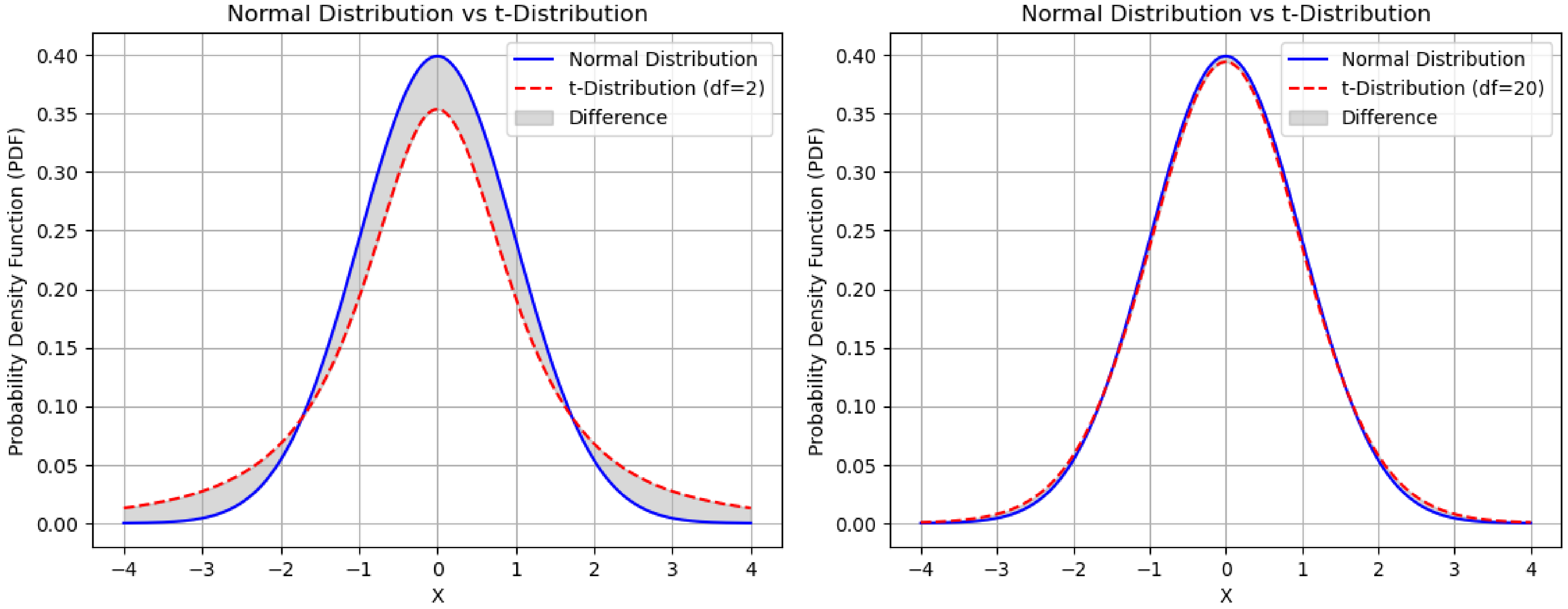
- Normal distributions are used when the population distribution is assumed to be normal.

- The t-distribution is similar to the normal distribution, just with fatter tails. Both assume a normally distributed population.

- T-distributions thus have higher kurtosis than normal distributions.

- The probability of getting values very far from the mean is larger with a t-distribution than a normal distribution.

T-Distribution vs. Normal Distribution



Important Note: Because the t-distribution has fatter tails than a normal distribution, it can be used as a model for financial returns that exhibit excess kurtosis, which will allow for a more realistic calculation of Value at Risk (VaR) in such cases.

Limitations of Using a T-Distribution

- The t-distribution can skew exactness relative to the normal distribution.
- Its shortcoming only arises when there's a need for perfect normality.
- The t-distribution should only be used when the population standard deviation is not known.
- If the population standard deviation is known and the sample size is large enough, the normal distribution should be used for better results.

When should the t-distribution be used?

The t-distribution should be used if the population sample size is small and the standard deviation is unknown. If not, then the normal distribution should be used.

The Bottom Line

The t-distribution is used in statistics to estimate the significance of population parameters for small sample sizes or unknown variations. Like the normal distribution, it is bell-shaped and symmetric. Unlike normal distributions, it has heavier tails, which result in a greater chance for extreme values.

Note: We will look into the main application of t distribution in the Hypothesis Testing lecture (i.e., T-test).

Chi-squared distribution

- The chi-squared distribution is a continuous probability distribution that arises in statistical inference, particularly in hypothesis testing and confidence interval construction.
- It is used in various statistical tests, such as the chi-squared test for independence and the chi-squared goodness-of-fit test.
- Definition: The chi-squared distribution is a continuous probability distribution of the sum of squared standard normal deviates.
- Symbol: χ^2
- It is widely used in statistical hypothesis testing.

Chi-squared distribution

The probability density function of the chi-squared distribution is given by:

$$f(x;k) = \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2}$$

where k is the degrees of freedom , and Γ is the gamma function.

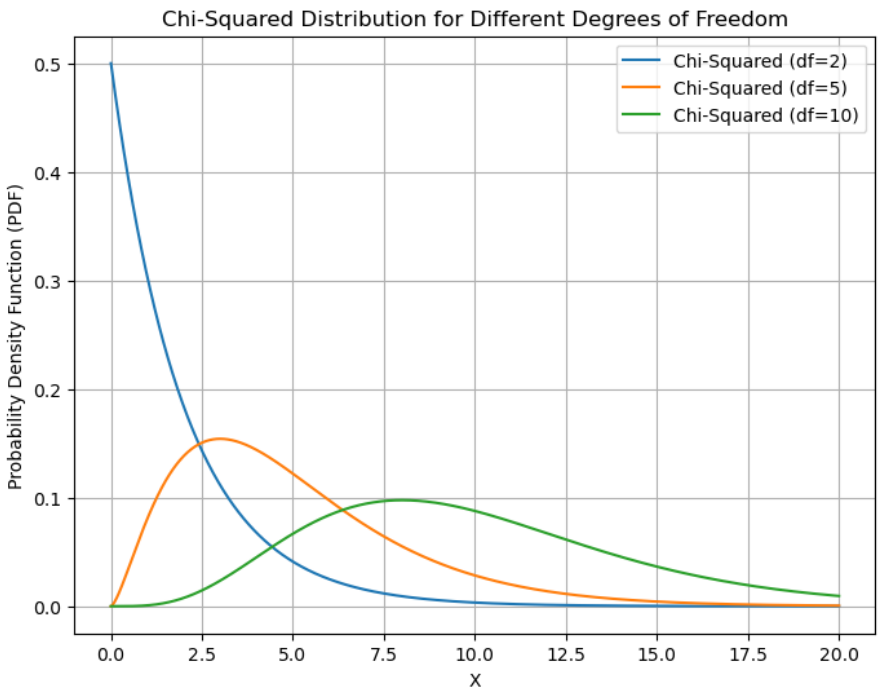
Chi-squared distribution

Degrees of freedom (k) in the chi-squared distribution determine its shape.

As k increases, the distribution becomes more symmetric and approaches normality.

Mean: E(X) = k

Variance: Var(X) = 2k



Chi-squared distribution

Chi-Squared Test for Independence

One of the main applications is the chi-squared test for independence.

Used to determine if there is a significant association between two categorical variables.

Chi-Squared Goodness-of-Fit Test

Another application is the chi-squared goodness-of-fit test.

Tests whether an observed frequency distribution matches an expected distribution

Relationship with Normal Distribution

The chi-squared distribution is a special case of the gamma distribution.

As k increases, the chi-squared distribution approaches a normal distribution.

Note: We will look into the main application examples of Chi-Squared distribution in the Hypothesis Testing lecture (in the Chi-Square test).

REFERENCES

- Introduction to Probability and Statistics for Engineers and Scientists, Sixth Edition, Sheldon M. Ross
- Statistical Methods Combined Edition (Volume I& II), N G Das

CONCLUSION

- Discussed the Standard Normal distribution and its Approximation Rule
- Discussed about the T distribution
- Discussed about the Chi-Squared distribution


THANK YOU