Machine Learning and Data Analytics
ME 5013- Fall 2019

Lecture 02

Review:

- Linear Algebra
- Probability Distributions

UTSA®
The University of Texas at San Antonio™

Adel Alaeddini, PhD
Associate Professor of Mechanical Engineering
Advanced Data Engineering Lab
adel.alaeddini@utsa.edu

Disclosure: some of the slides are adopted from Dr. Andrew Ng lecture notes on "Machine Learning" at Stanford university

Matrix: Rectangular array of numbers:

$$A = \begin{bmatrix} 1402 & 191 \\ 1371 & 821 \\ 949 & 1437 \\ 147 & 1448 \end{bmatrix}$$

Dimension of matrix: number of rows x number of columns

Matrix Elements (entries of matrix)

$$A = \begin{bmatrix} 1402 & 191 \\ 1371 & 821 \\ 949 & 1437 \\ 147 & 1448 \end{bmatrix}$$

$A_{ij} =$ "i, j entry" in the i^{th} row, j^{th} column.

Vector: An n x 1 matrix.

$$y = \begin{bmatrix} 460 \\ 232 \\ 315 \\ 178 \end{bmatrix}$$

$y_i = i^{th}$ element

1-indexed vs 0-indexed:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} \qquad y = \begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \end{bmatrix}$$

Matrix Addition

$$\begin{bmatrix} 1 & 0 \\ 2 & 5 \\ 3 & 1 \end{bmatrix} + \begin{bmatrix} 4 & 0.5 \\ 2 & 5 \\ 0 & 1 \end{bmatrix} =$$

$$\begin{bmatrix} 1 & 0 \\ 2 & 5 \\ 3 & 1 \end{bmatrix} + \begin{bmatrix} 4 & 0.5 \\ 2 & 5 \end{bmatrix} =$$

Scalar Multiplication

$$3 \times \begin{bmatrix} 1 & 0 \\ 2 & 5 \\ 3 & 1 \end{bmatrix} =$$

$$\begin{bmatrix} 4 & 0 \\ 6 & 3 \end{bmatrix} /4 =$$

Combination of Operands

$$3 \times \begin{bmatrix} 1 \\ 4 \\ 2 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 5 \end{bmatrix} - \begin{bmatrix} 3 \\ 0 \\ 2 \end{bmatrix} / 3$$

Details:

$$
\begin{array}{ccccc}
A & \times & x & = & y \\
\left[\begin{array}{cccc} & & & \\ & & & \\ & & & \end{array}\right] & \times & \left[\begin{array}{c} \\ \\ \\ \\ \end{array}\right] & = & \left[\begin{array}{c} \\ \\ \\ \end{array}\right]
\end{array}
$$

m x n matrix (m rows, n columns)	n x 1 matrix (n-dimensional vector)	m-dimensional vector

To get y_i, multiply A's i^{th} row with elements of vector x, and add them up.

Example

$$\begin{bmatrix} 1 & 3 \\ 4 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 5 \end{bmatrix} =$$

$$\begin{bmatrix} 1 & 2 & 1 & 5 \\ 0 & 3 & 0 & 4 \\ -1 & -2 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 2 \\ 1 \end{bmatrix} =$$

Details:

$$A \times B = C$$

$$\begin{bmatrix} & & \end{bmatrix} \times \begin{bmatrix} & & \end{bmatrix} = \begin{bmatrix} & & \end{bmatrix}$$

m x n matrix (m rows, n columns)

n x o matrix (n rows, o columns)

m x o matrix

The i^{th} column of the matrix C is obtained by multiplying A with the i^{th} column of B. (for i = 1,2,…,o)

Example

$$\begin{bmatrix} 1 & 3 & 2 \\ 4 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 0 & 1 \\ 5 & 2 \end{bmatrix} =$$

$$\begin{bmatrix} 1 & 3 & 2 \\ 4 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 1 \\ 0 \\ 5 \end{bmatrix} =$$

$$\begin{bmatrix} 1 & 3 & 2 \\ 4 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix} =$$

Matrix Multiplication Properties

Let A and B be matrices. Then in general,

$A \times B \neq B \times A$. (not commutative.)

E.g. $$\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 2 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 0 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 2 & 2 \end{bmatrix}$$

$A \times B \times C$.

Let $D = B \times C$. Compute $A \times D$.

Let $E = A \times B$. Compute $E \times C$.

Identity Matrix

Denoted I (or $I_{n \times n}$).

Examples of identity matrices:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$
2 x 2

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$
3 x 3

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$
4 x 4

For any matrix A,

$$A \cdot I = I \cdot A = A$$

Matrix Operation Properties

Given A, B, C $n \times m$ matrices and $\alpha, \beta \in \mathbb{R}$

Addition

i. $A + B = B + A$ Conmotativity

ii. $A + (B + C) = (A + B) + C$ Associativity

iii. $A + \mathbf{0} = A$ → Identity element.

iv. $A + (-A) = \mathbf{0}$

Scalar Multiplication

v. $(\alpha\beta)A = \alpha(\beta A)$

vi. $(\alpha + \beta)A = \alpha A + \beta A$

vii. $\alpha(A + B) = \alpha A + \alpha B$

distributive

viii. $1A = A$ Scalar multp. identity

Vector Space.

Matrix Transpose

Example: $A = \begin{bmatrix} 1 & 2 & 0 \\ 3 & 5 & 9 \end{bmatrix}$ $A^T = \begin{bmatrix} 1 & 3 \\ 2 & 5 \\ 0 & 9 \end{bmatrix}$

Let A be an m x n matrix, and let $B = A^T$.
Then B is an n x m matrix, and

$$B_{ij} = A_{ji}.$$

Determinant

It is a scalar function defined over square matrices

$$\det A = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1j} & \dots & a_{1n} \\ \vdots & & & & & \\ a_{i1} & a_{i2} & \dots & a_{ij} & \dots & a_{in} \\ \vdots & & & & & \\ a_{n1} & a_{n2} & \dots & a_{nj} & \dots & a_{nn} \end{vmatrix} = \sum_{k=1}^{n} a_{ik} C_{ik} = \sum_{k=1}^{n} a_{kj} C_{kj}$$

i

Cofactor $C_{ij} = (-1)^{i+j} M_{ij}$

↳ the minor determinant

Example

$$A = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 3 & 2 & 1 \\ 4 & 3 & 6 \\ 2 & 3 & 1 \end{bmatrix} \end{array}$$

$$C_{12} = (-1)^{1+2} M_{12}$$

$$= (-1)^3 \begin{vmatrix} 4 & 6 \\ 2 & 1 \end{vmatrix} = -(4-12)$$

$$= 8$$

Example

$$\sum a_{1k} C_{1k}$$

$$\begin{vmatrix} 4 & 7 & -2 \\ 0 & 3 & 2 \\ 1 & 3 & 6 \end{vmatrix} = a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13}$$

$$4\,C_{11} + 7\,C_{12} - 2\,C_{13}$$

$$4(-1)^{1+1}\begin{vmatrix} 3 & 2 \\ 3 & 6 \end{vmatrix} + 7(-1)^{1+2}\begin{vmatrix} 0 & 2 \\ 1 & 6 \end{vmatrix} - 2(-1)^{1+3}\begin{vmatrix} 0 & 3 \\ 1 & 3 \end{vmatrix}$$

$$4(18-6) - 7(-2) - 2(-3)$$

$$48 + 14 + 6 = 68$$

Not all numbers have an inverse.

Matrix inverse:

If A is an m x m matrix, and if it has an inverse,

$$AA^{-1} = A^{-1}A = I.$$

Matrices that don't have an inverse are "singular" or "degenerate"

Rank of a Matrix

Rank

The rank of a matrix is the number if linearly independent rows or "columns"

Method

Reduce the matrix to row echelon form and count the numbers of row different from zero

Example 1: $\begin{bmatrix} 1 & 1 & -1 \\ 3 & 1 & 1 \\ 1 & -1 & 4 \end{bmatrix} = B \xrightarrow[\text{Elimination}]{} \begin{bmatrix} 1 & 1 & -1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix}$ $\text{rank}(B) = 3$

dimension = 3

Matrix Inverse

Adjoint Matrix:

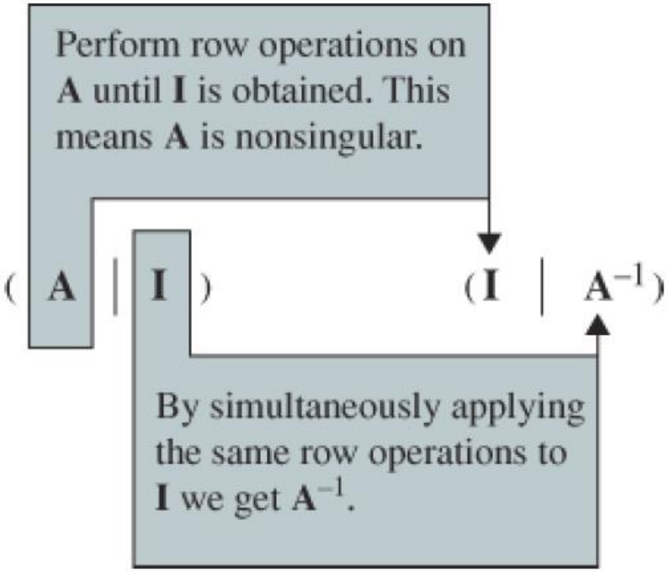
Let A be an n × n matrix. The matrix that is the transpose of the matrix of cofactors corresponding to the entries of A:

$$\begin{pmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ C_{21} & C_{22} & \cdots & C_{2n} \\ \vdots & & & \vdots \\ C_{n1} & C_{n2} & \cdots & C_{nn} \end{pmatrix}^T = \begin{pmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \vdots & & & \vdots \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{pmatrix}$$

Inverse Matrix

$$\mathbf{A}^{-1} = \left(\frac{1}{\det \mathbf{A}}\right) \operatorname{adj} \mathbf{A}.$$

Matrix Inverse



$$\mathbf{A} = \begin{pmatrix} 2 & 0 & 1 \\ -2 & 3 & 4 \\ -5 & 5 & 6 \end{pmatrix}$$

$$\left(\begin{array}{ccc|ccc} 2 & 0 & 1 & 1 & 0 & 0 \\ -2 & 3 & 4 & 0 & 1 & 0 \\ -5 & 5 & 6 & 0 & 0 & 1 \end{array}\right) \longrightarrow \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -2 & 5 & -3 \\ 0 & 1 & 0 & -8 & 17 & -10 \\ 0 & 0 & 1 & 5 & -10 & 6 \end{array}\right)$$

Probability Distributions

Probability

- **Probability:** is a measure of the expectation that an event will occur or a statement is true.
- **Random experiment/event:** an experiment whose outcome cannot be predicted with certainty, before the experiment is run
- **Random variable** or **stochastic variable** is a variable whose value is subject to variations due to chance
  - A coin is tossed ten times. The random variable X is the number of tails that are noted.
  - A light bulb is burned until it burns out. The random variable Y is its lifetime in hours.
- **Probability Distribution** assigns a probability to each of the possible outcomes of a random experiment

Discrete Random variable

- **Discrete random variable**
  - Takes on one of a finite (or at least countable) number of different values.
  - X = 1 if heads, 0 if tails
  - Y = 1 if male, 0 if female (phone survey)
  - Z = # of spots on face of thrown die

Continuous Random variable

- **Continuous random variable (r.v.)**
  - Takes on one in an infinite range of different values
  - W = % GDP grows (shrinks?) this year
  - V = hours until light bulb fails
- For a **discrete** r.v., we have Prob(X=x), i.e., the probability that
- r.v. X takes on a given value x.

- What is the probability that a **continuous** r.v. takes on a specific value? E.g. Prob(X_light_bulb_fails = 3.14159265 hrs) = ?? 0

- However, **ranges of values** can have non-zero probability.
- E.g. Prob(3 hrs <= X_light_bulb_fails <= 4 hrs) = 0.1

  - *Ranges* of values have a probability

Probability Distribution

- The probability distribution is a complete probabilistic description of a random variable.
- All other statistical concepts (expectation, variance, etc) are derived from it.
- Once we know the probability distribution of a random variable, we know everything we can learn about it from statistics.

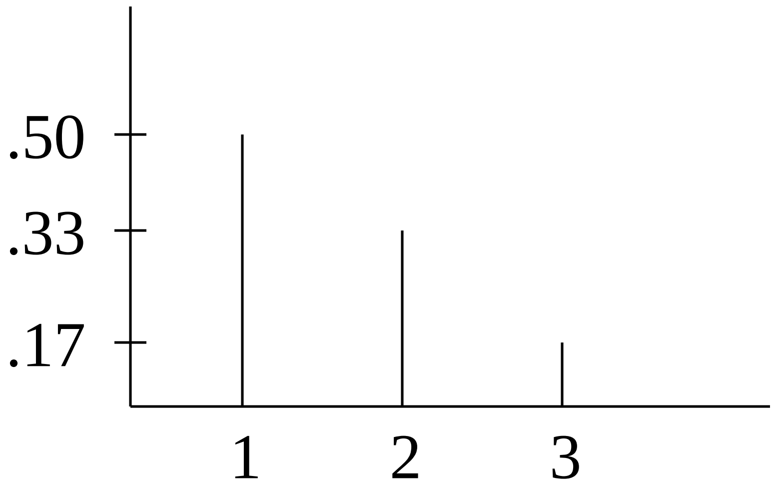
- Probability function
  - One form the probability distribution of a discrete random variable may be expressed in.
  - Expresses the probability that X takes the value x as a function of x (as we saw before):

$$P_X(x) = P(X = x)$$

- The probability function
  - May be tabular:

$$X = \begin{cases} 1 & w.p. & 1/2 \\ 2 & w.p. & 1/3 \\ 3 & w.p. & 1/6 \end{cases}$$

- The probability function
  - May be graphical:



- The probability function
  - May be formulaic:

$$P(X = x) = \frac{4 - x}{6} \quad \text{for x} = 1,2,3$$

- The probability function, properties

$$P_X(x) \geq 0 \quad \text{for each } x$$

$$\sum_x P_X(x) = 1$$

Cumulative Probability Distribution

- Cumulative probability distribution
  - The cdf is a function which describes the probability that a random variable does not exceed a value.

$$F_X(x) = P(X \leq x)$$

Does this make sense for a continuous r.v.?

Yes!

Cumulative Probability Distribution

- The cdf, properties

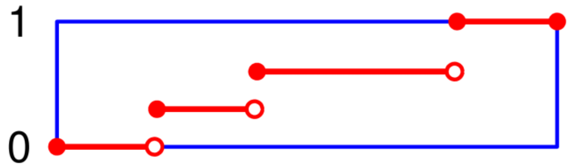
$0 \leq F_X(x) \leq 1 \quad \text{for each } x$

$F_X(x) \quad \text{is non−decreasing}$

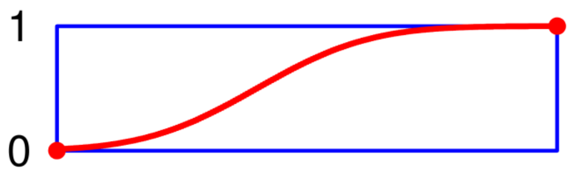
$F_X(x) \quad \text{is continuous from the right}$

$$\lim_{x \to -\infty} F(x) = 0, \qquad \lim_{x \to +\infty} F(x) = 1.$$

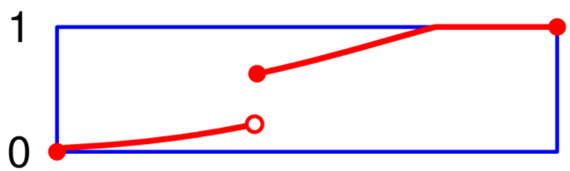
Example CDFs



Of a discrete probability distribution



Of a continuous probability distribution



Of a distribution which has both a continuous part and a discrete part.

1) --- Bernoulli distribution
2) --- Binomial
3) --- Geometric
4) --- Poisson
5)--- Negative binomial
6)---Hyper-geometric
7)---Uniform

Bernoulli distribution

- The Bernoulli distribution is the “coin flip” distribution.

- X is Bernoulli if its probability function is:

$$X = \begin{cases} 1 & w.p. \quad p \\ 0 & w.p. \quad 1-p \end{cases}$$

X=1 is usually interpreted as a “success.” E.g.:

X=1 for heads in coin toss
X=1 for male in survey
X=1 for defective in a test of product
X=1 for “made the sale” tracking performance

1) --- Normal distribution
2) --- Exponential
3) --- Chi-Squared
4) --- Weibull
5) ---Student –t
6)---Log-normal
7)---Beta
8)---Gamma
9)---Uniform

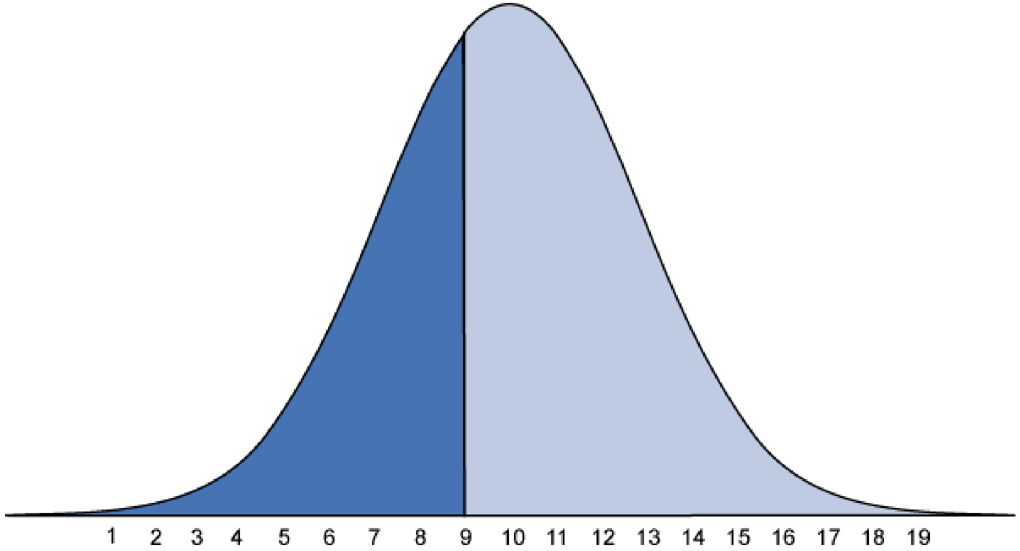
Normal distribution

Characteristics:

- Bell-shaped with a single peak
- Symmetrical so two halves are mirror images
  - There are numerous normal distributions that have the same mean, but different standard deviations.

The normal distribution is very important :

- It can be used as an approximation for many other distributions.
- Many random variables in the real world follow a normal distribution.

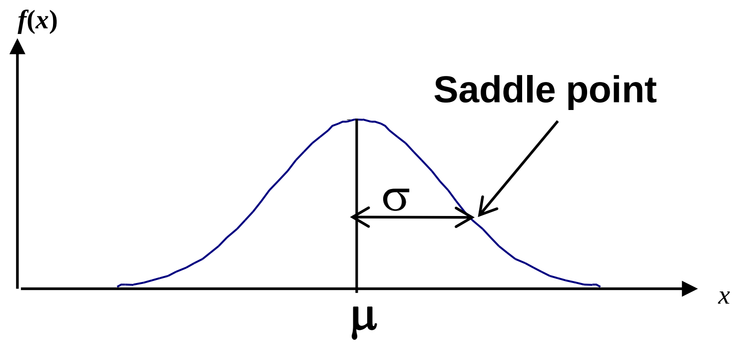


Normal Distribution

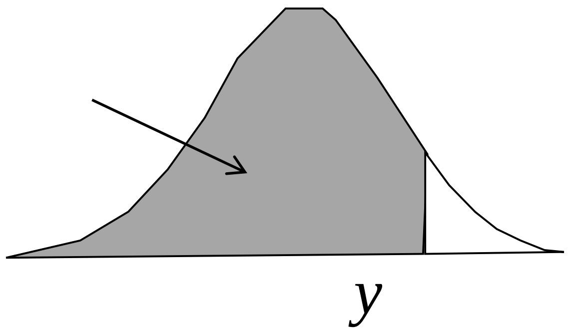
- This *pdf* is the most popular distribution for continuous random variables
- First described de Moivre in 1733, elaborated in 1812 by Laplace

$$X \sim N(\mu, \sigma^2); \qquad -\infty < x < +\infty$$

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2 \right\}$$



$$F(x) = \Pr\{X \le x\} = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2 \right\} dx$$



Map any X into Z

$X \sim N(\mu, \sigma^2); \quad -\infty < x < +\infty$

$$Z = \frac{X - \mu}{\sigma}$$

$Z \sim N(0, 1^2); \quad -\infty < z < +\infty$

$$f(z; \mu = 0, \sigma^2 = 1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right)$$

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-0.00	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
-0.10	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.20	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.30	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.40	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.50	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.60	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.70	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.80	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.90	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-1.00	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-1.10	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.20	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.30	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.40	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.50	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.60	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.70	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.80	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.90	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-2.00	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-2.10	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.20	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.30	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.40	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.50	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.60	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.70	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.80	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.90	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-3.00	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-3.10	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.20	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.30	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.40	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.50	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
-3.60	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
-3.70	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
-3.80	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001

Example

$X \sim N(40, 5^2)$

$$p(x \leq 37.9) = \Phi\left(\frac{37.9 - 40}{5}\right) = \Phi(-0.42) = 0.3772$$