Continuous random variables

Outline

- Discrete vs continuous random variables
- Probability mass function vs Probability density function
- Properties of the pdf
- Cumulative distribution function
- Properties of the cdf
- Expectation, variance and properties
- The normal distribution

Recap

Till now, we discussed

Discrete random variables: can take a finite, or at most countably infinite, number of values,

For example:

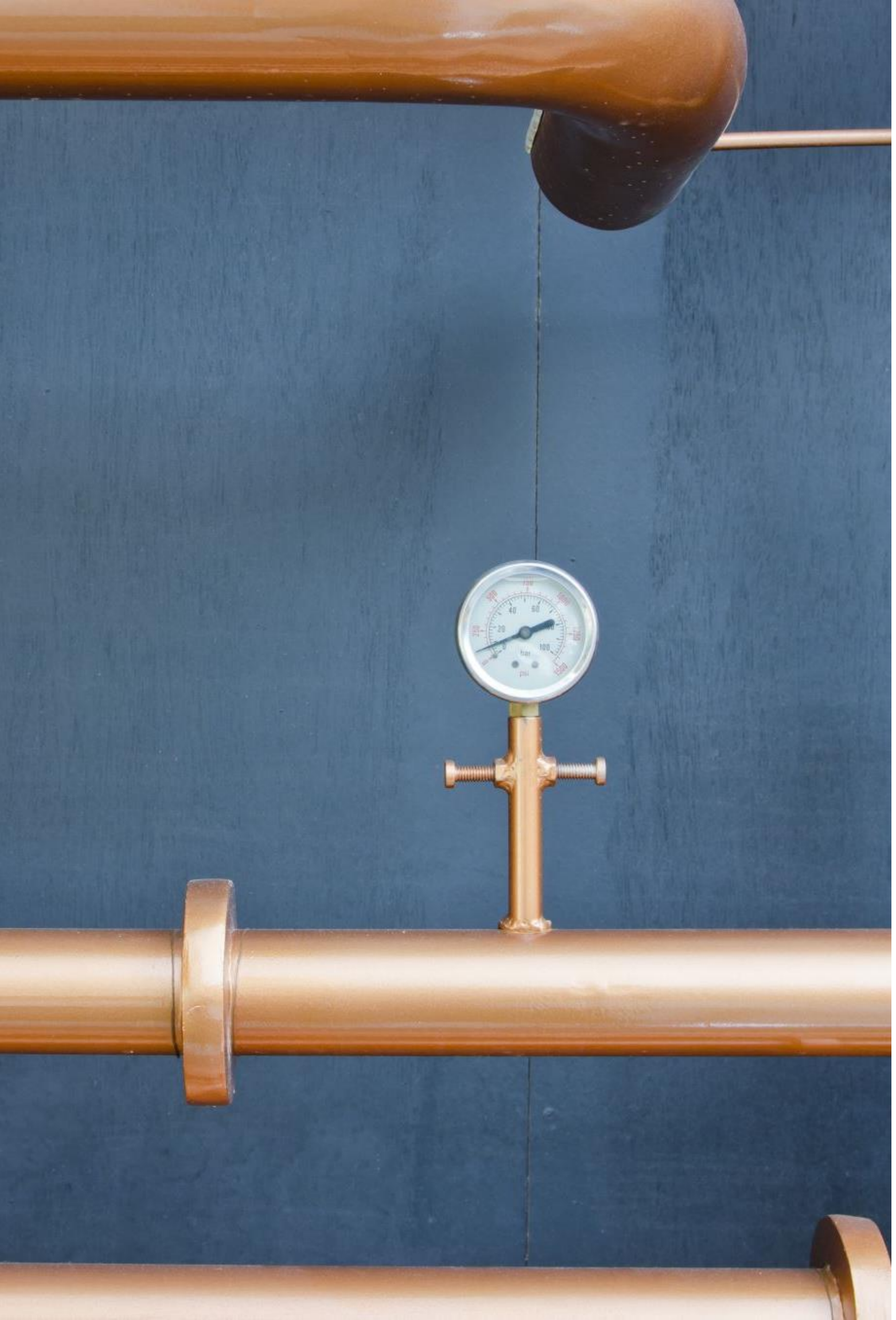
- Binomial random variable
- Bernoulli random variable
- Geometric random variable

Continuous random variable

- A continuous random variable is a random variable that Can take on an uncountably infinite range of values.

- Due to the above definition, the probability that a continuous random variable will take on an exact value is 0.

- For any specific value $X = x$, $P(X = x) = 0$



Continuous random variable

<u>Examples:</u>

- The volume of water passing through a pipe over a given time period.
- The height of a randomly selected individual.

$$\int x^n dx = \frac{x^{n+1}}{n+1}$$

$$\int_{1/2}^{1} 4x^3 dx = \left[\frac{4x^4}{4}\right]_{1/2}^{1}$$

Continuous random variable

Example:

Suppose the **probability density function** of a continuous random variable, X, is given by $4x^3$, where $x \in [0, 1]$.

The probability that X takes on a value between 1/2 and 1 needs to be determined.

$$\left(x^4\right)_{1/2}^{1}$$

$$1 - \frac{1}{16} = \frac{15}{16}$$

Solution

Solution:

This can be done by integrating $4x^3$ between 1/2 and 1. Thus, the required probability is

$$P = \int_{1/2}^{1} 4x^3$$

P = 15/16.

Probability density function (pdf)

- The probability density function (pdf) and the cumulative distribution function (CDF) are used to describe the probabilities associated with a continuous random variable.

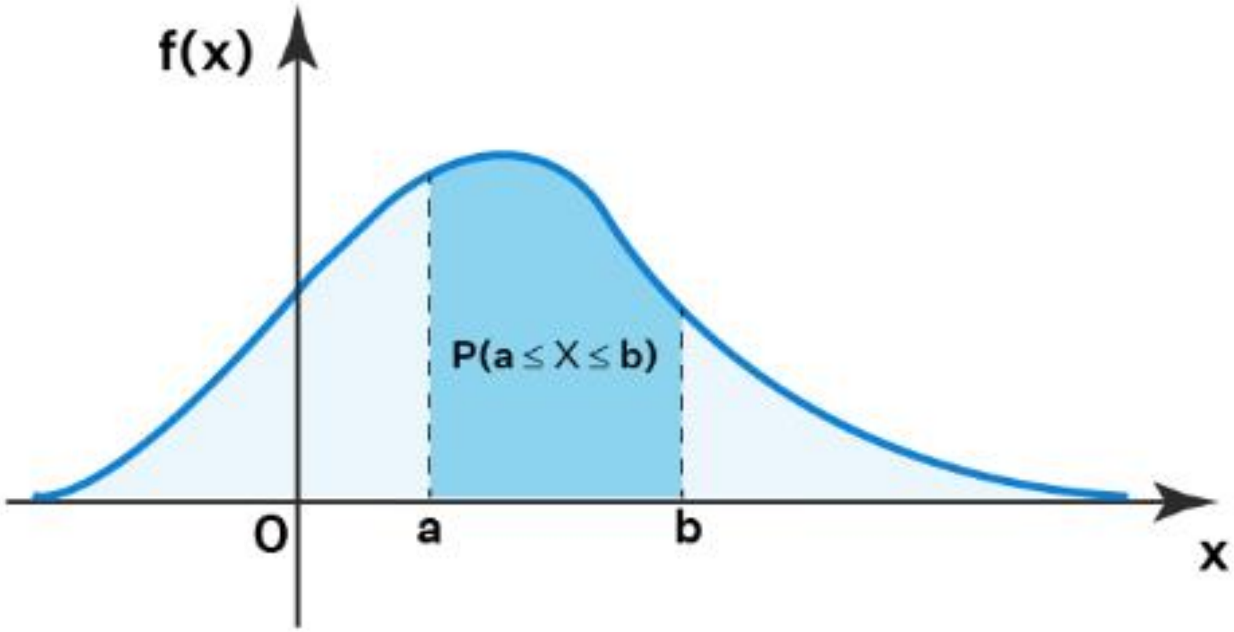
- For a continuous random variable, we cannot construct a PMF (discussed earlier for discrete) – each specific value has zero probability.

- Instead, we use a continuous, non-negative function $f_X\ (x)$ called the probability density function, or PDF, of X

Probability density function (pdf)

The probability of X lying between two values x_1 and x_2 is simply the area under the PDF, i.e

$$P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} f_X(x)dx$$



Example

The pdf of a continuous random variable, X, is given as follows:

$$f(x) = \begin{cases} x & 0 \leq x \leq 1 \\ x+3 & 1 < x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

Find the probability that the value of the continuous random variable will lie between 0 and 0.5, i,e., Find $P(0 \leq X \leq 0.5)$.

Ans: 0.125

Solution

-

$$F(x) = \int_0^{0.5} f(x)\,dx$$

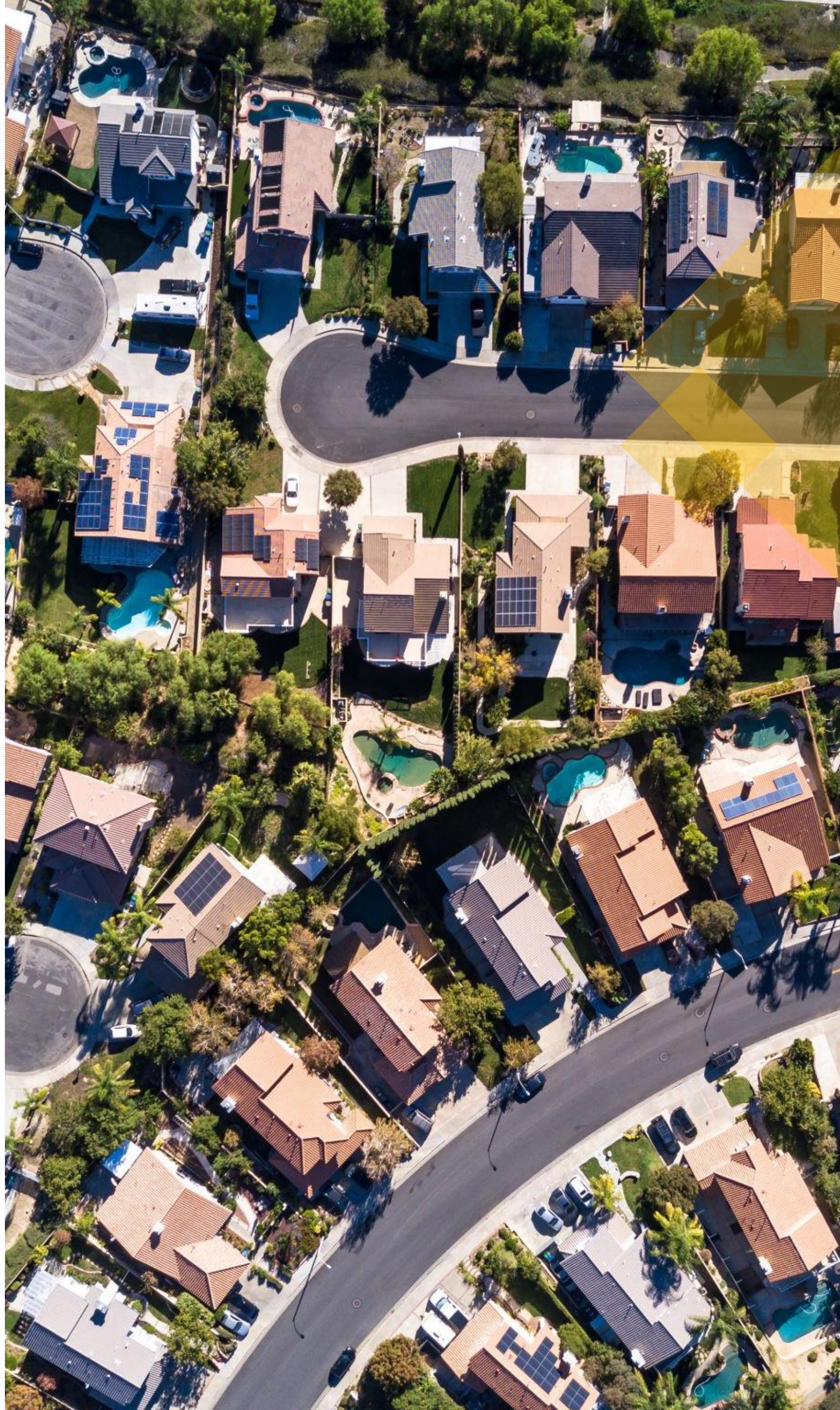
$$= \int_0^{0.5} x\,dx = \left.\frac{x^2}{2}\right|_0^{0.5}$$

$$= \frac{1}{2}\left(0.5 \times 5 - 0\right)$$

$$= 1/8 = 0.125$$

Properties of the pdf

- For any single value a, $P(X = a) = \int_a^a f_X(x)dx = 0$
- f(x) $\geq$ 0. This implies that the probability density function of a continuous random variable cannot be negative.
- $\int_{-\infty}^{-\infty} f_X(x)dx$ =1, this means that the total area under the graph of the pdf must be equal to 1.
- Note that fX (x) can be greater than 1 – even infinite! – for certain values of x, provided the integral over all x is 1.



Cumulative distribution function (cdf)

Now, we are interested in $P(X \leq x)$

Examples:

- What is the probability that the bus arrives before 1:30?
- What is the probability that a randomly selected person is under 5'7"

We can get this from our PDF:

$$F_X(x) = P(X \leq x') = \int_{-\infty}^{x'} f_X(x)dx$$

Note: If X is discrete, $f_X\ (x)$ is a piecewise-constant function of x

Cumulative distribution function (cdf)

- The CDF is monotonically non-decreasing:

if x ≤ y, then FX (x) ≤ FX (y)

- $F_X(x) \rightarrow 0$ as x → −∞

- $F_X(x) \rightarrow 1$ as x → ∞

Expectation of a continuous random variable

- Similar to the discrete case…

but we are integrating rather than summing

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x f_X(x)dx$$

- Just as in the discrete case, we can think of E[X] as the "center of gravity" of the PDF

Expectation of a continuous random variable

Expectation of a function g(X) of a continuous random variable is

$$\mathrm{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx$$

Note, g(X) can be a continuous random variable, e.g. $g(X) = X^2$, or a discrete random variable, e.g

$$g(X) = \begin{cases} 1 & \text{if } X \geq 0 \\ 0 & \text{if } X < 0 \end{cases}$$

Variance of a continuous random variable

$$var[X] = E[X^2] - E[X]^2$$

$$var[X] = \int_{-\infty}^{\infty} (x - E[x])^2 f_X(x) dx$$

Note:

$$E[aX + b] = aE[X] + b$$

$$var(aX + b) = a^2 \, var(X)$$

Geometric dist: $E(x) = x p_i$

mean $\Rightarrow 1/p$

Variance $\Rightarrow q/p^2$

S.D $\Rightarrow \sqrt{q/p^2}$

$P(x) = q^{x-1} \cdot p$

Binomial Dist

mean $= pn$

Var $= npq$

S.D $= \sqrt{npq}$

${}^nC_x \, p^x \, q^{n-x}$

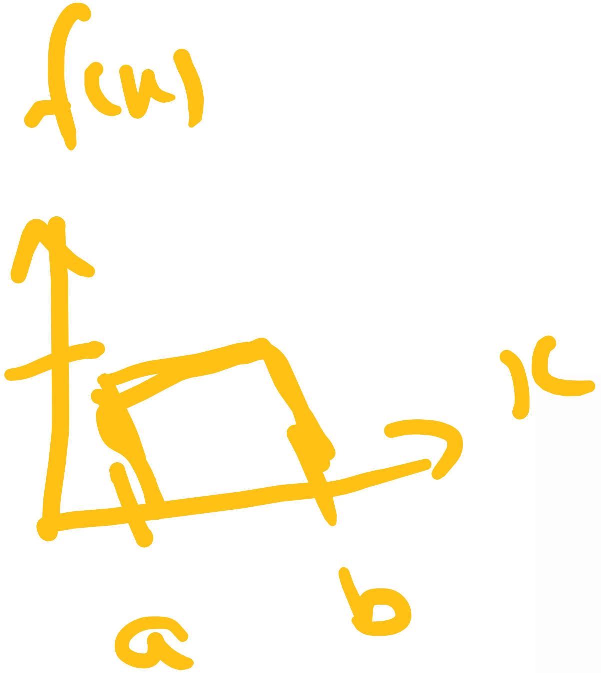
Poisson

mean $= \frac{r}{p}$

Var $= \frac{rq}{p}$

S.D $= \sqrt{\frac{rq}{p}}$

$\frac{\lambda^x e^{-\lambda}}{x!}$



Uniform Distribution

$$\Downarrow$$

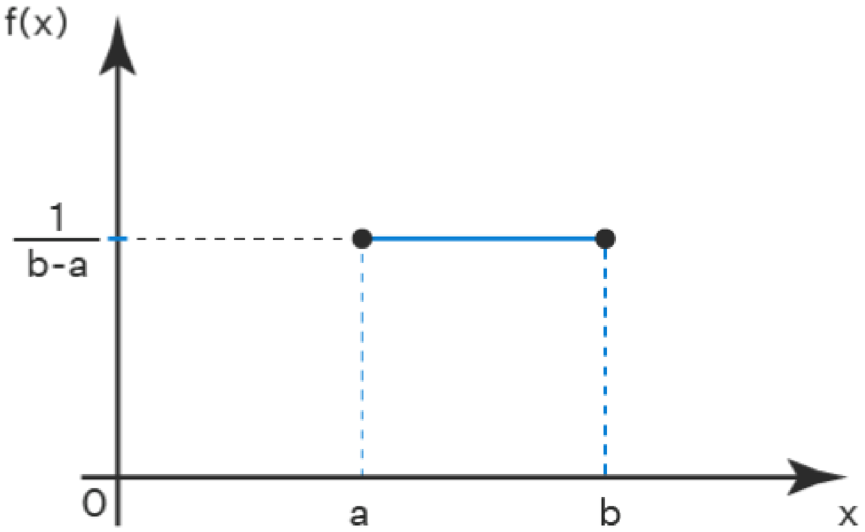
Constant Probability within Domain

Uniform Random Variable

- A continuous random variable that is used to describe a uniform distribution is known as a uniform random variable.

- Such a distribution describes events that are equally likely to occur.

- The pdf of a uniform random variable is as follows:



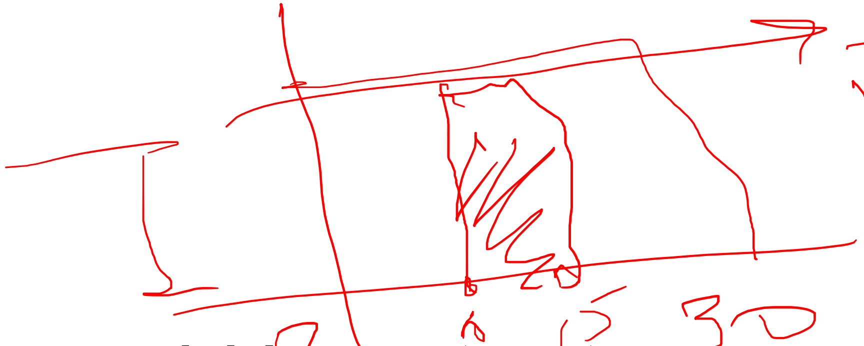
$$f(x) = \begin{cases} \frac{1}{b-a} & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

Example

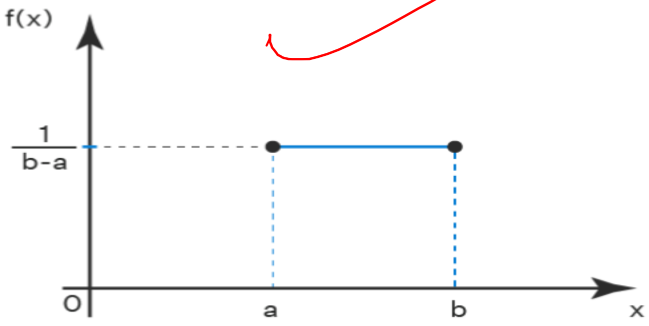
Uniform Distribution:

The average marks gained by a student for the first quiz is uniformly distributed and ranges from 0 to 30. Find the probability of a student that he will gain between 10 and 15 marks in the quiz.

0 to 30



$\frac{1}{b-a} = \frac{1}{30}$



$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

Probability = 5 × (1 / 30) = 1 / 6

area $= (15-10) \times \frac{1}{30} = 1/6$

Example

- Determine P(X ≤ 10) for the previous-given question.

$P(X \leq 10)$

- 10 × (1/ 30) = 1 / 3

$= (10 - 0) \times \frac{1}{30 - 0}$

$= 1/3$

Mean of a uniform random variable

$$E(X) = \int_{-\infty}^{\infty} x f(x)\,dx$$

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

$$\int_{-\infty}^{a} x f(x)\,dx + \int_{a}^{b} x f(x)\,dx + \int_{b}^{\infty} x f(x)\,dx$$

Mean of a uniform random variable

Let X be a uniform random variable over [a, b]. What is its expected value?

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\,dx$$

$$f(x) = \begin{cases} \frac{1}{b-a} & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

$$E[X] = \int_{-\infty}^{a} x \times 0dx + \int_{a}^{b} \frac{x}{b-a} dx + \int_{b}^{\infty} x \times 0dx$$

$$\mathrm{E}[X] = \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2}$$

Variance of a uniform random variable

$$var[X] = E[X^2] - (E[X])^2$$

$$E(x^2) = \frac{a^2+b^2+ab}{3}$$

$$(E(x))^2 = \left(\frac{a+b}{2}\right)^2$$

$$var(x) = \frac{a^2+b^2+ab}{3} - \left(\frac{a+b}{2}\right)^2$$

$$\frac{a^2+b^2+ab}{3} - \frac{a^2+b^2+2ab}{4} = \frac{(b-a)^2}{12}$$

Variance of a uniform random variable

$$
\begin{aligned}
E[X^2] &= \int_{-\infty}^{\infty} x^2 f_X(x)dx \\
&= \int_a^b \frac{x^2}{b-a}dx \\
&= \left[\frac{x^3}{3(b-a)}\right]_a^b \\
&= \frac{b^3-a^3}{3(b-a)} = \frac{a^2+ab+b^2}{3}
\end{aligned}
$$

So, the variance is

$$
\mathrm{var}(X) = E[X^2] - E[X]^2 = \frac{a^2+ab+b^2}{3} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}
$$

The normal distribution

A normal, or Gaussian, random variable is a continuous random variable with PDF

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$$

where,

μ = mean

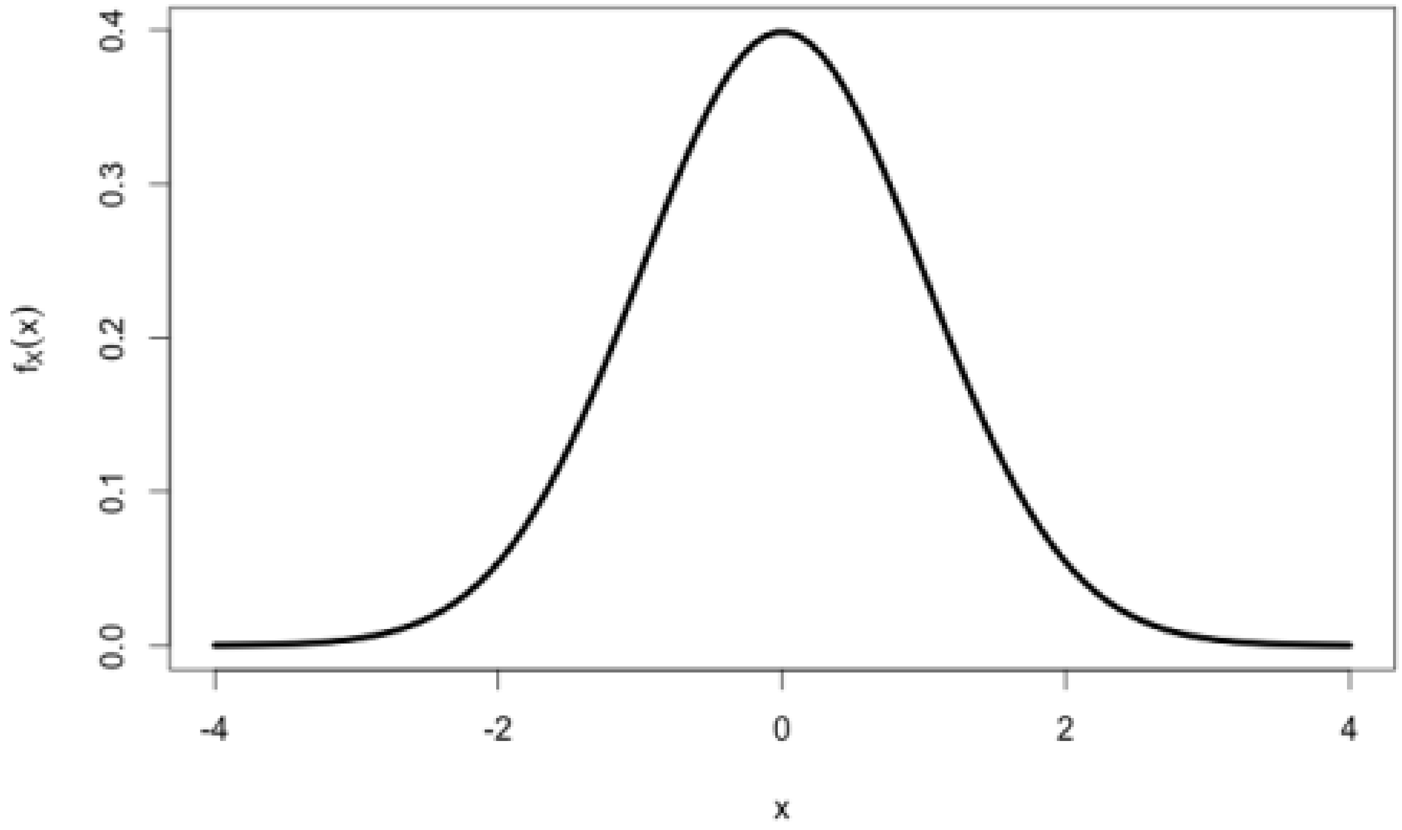
σ = standard deviation

σ^2 = variance

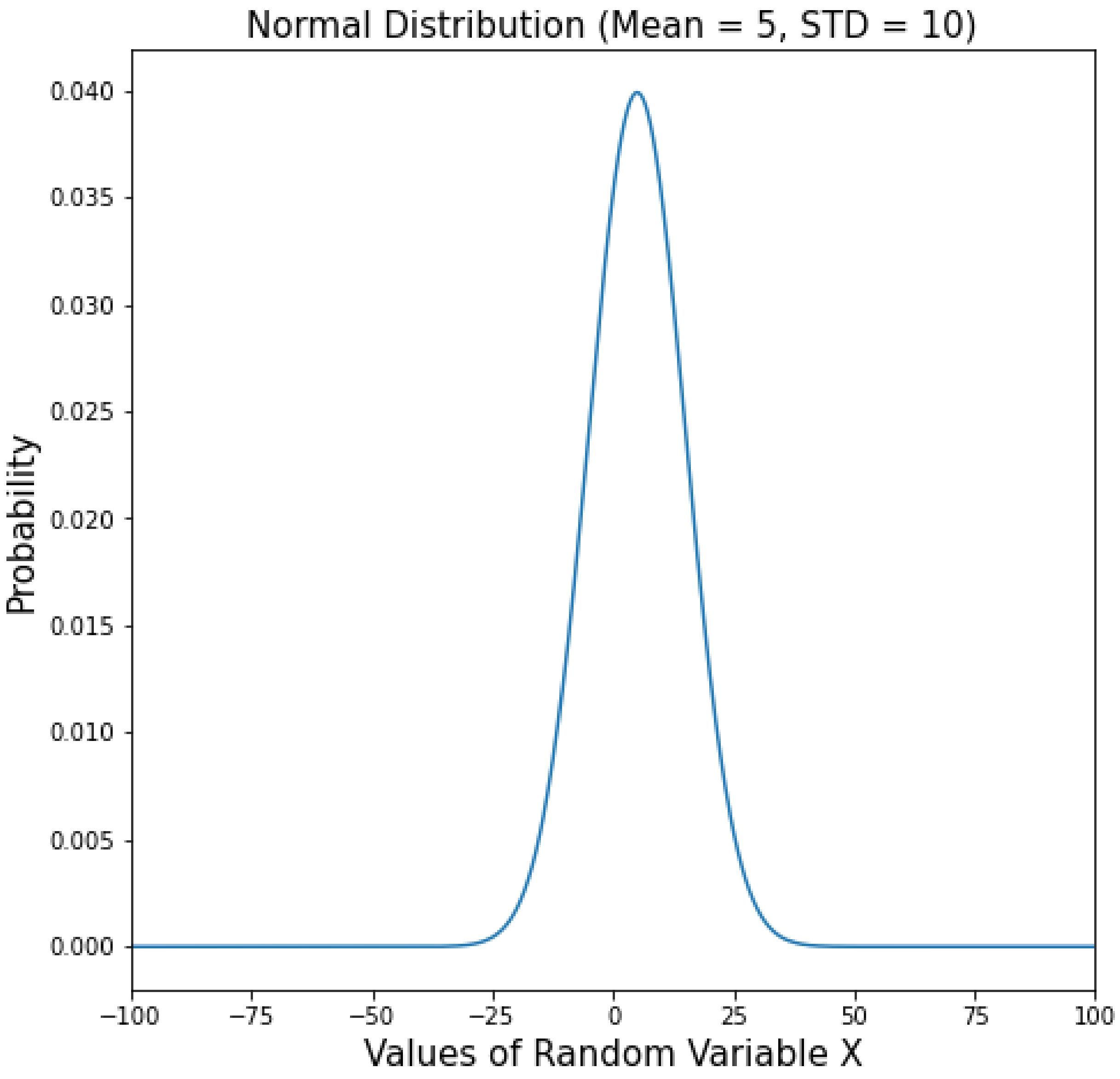
We write $X \sim N(\mu, \sigma^2)$

A normal distribution where μ = 0 and σ^2 = 1 is known as a standard normal distribution

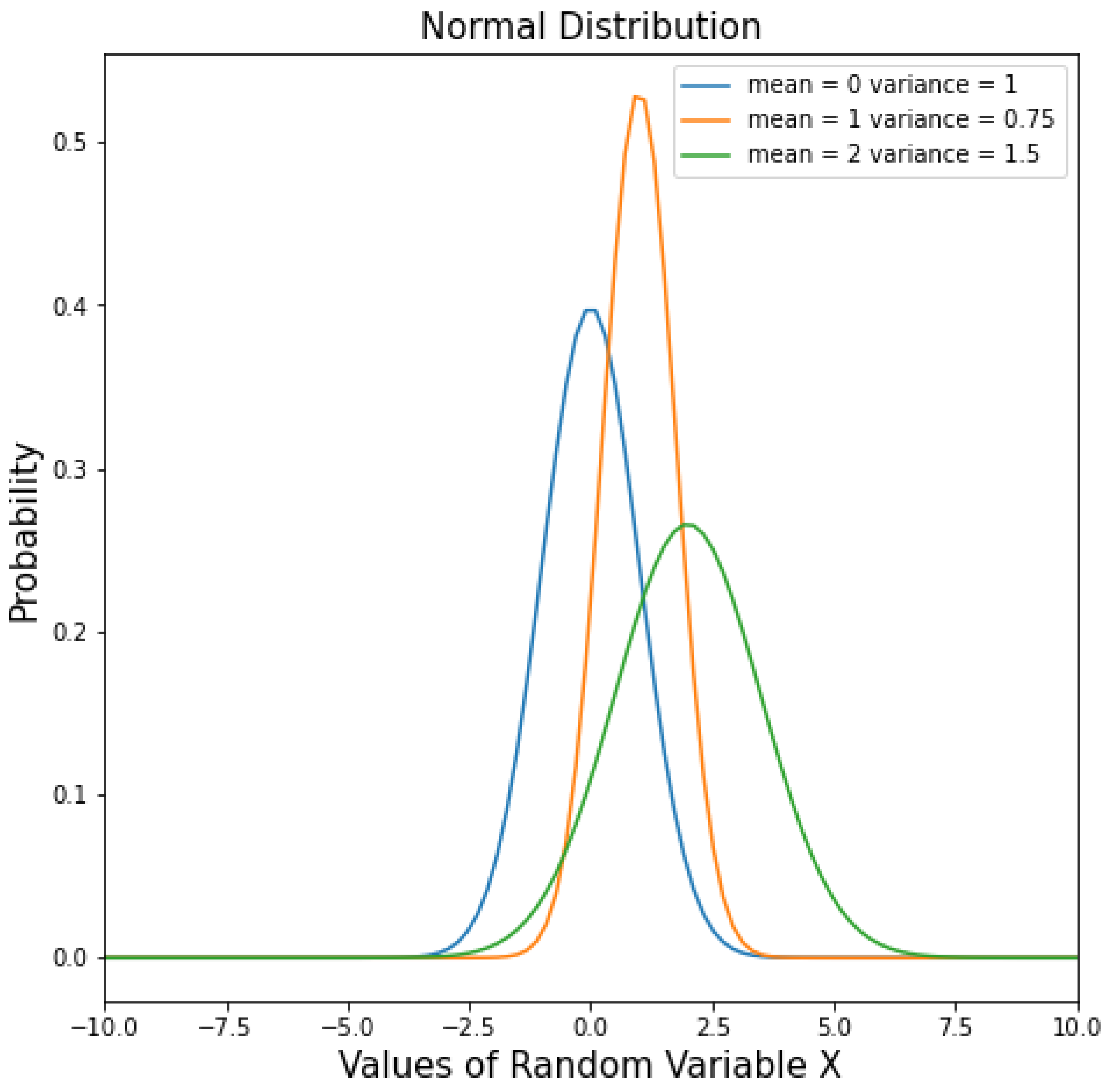
The normal distribution



The normal distribution



The normal distribution



The normal distribution

- The normal distribution is the classic "**bell-shaped curve**"

- Further, it has a number of nice properties that make it easy to work with.
  Like symmetry.

$P(X \geq 2) = P(X \leq -2)$ (from the curve)

Let $X \sim N(\mu, \sigma^2)$ and Let Y = aX + b

Then What are the mean and variance of Y ?

- $E[Y] = a\mu + b$
- $var[Y] = a^2\sigma^2$.
- Then Y is also a normal random variable with mean $a\mu + b$ and variance $var[Y] = a^2\sigma^2$.

Example

If 95% of students at school are between **1.1m and 1.7m** tall. Assuming this data is **normally distributed, then** calculate the mean?

- The mean is halfway between 1.1m and 1.7m:

- Mean = (1.1 + 1.7) / 2 = 1.4m

The standard normal

- $X \sim N(0,1)$
- The pdf formula is as follows:

$$f(x) = \frac{1}{\sqrt{2\Pi}} e^{-\frac{x^2}{2}}$$

The standard normal

It is often helpful to map our normal distribution with mean and
variance 2 onto a normal distribution with mean 0 and variance 1.

If $X \sim N(\mu, \sigma^2)$, then $Z = \dfrac{X - \mu}{\sigma} \sim N(0, 1)$.

(Note, we often use the letter Z for standard normal random variables)

Z Score

- A z-score (also called a standard score) gives you an idea of how far from the mean a data point is.

- But more technically it's a measure of how many standard deviations below or above the population mean a raw score is.

- In order to use a z-score, you need to know the mean μ and also the population standard deviation σ.

- The **basic z score formula**

$$\boldsymbol{z = (x - \mu) / \sigma}$$

Z Score

- For example, let's say you have a test score of 190. The test has a mean (μ) of 150 and a standard deviation (σ) of 25. Assuming a normal distribution, your z score would be:

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

$$= (190 - 150) / 25 = 1.6.$$

- In this example, your score is 1.6 standard deviations above the mean

Z Table

- The z-table is short for the “Standard Normal z-table”.
- The Standard Normal model is used in hypothesis testing, including tests on proportions and on the difference between two means.
- The area under the whole of a normal distribution curve is 1, or 100 percent.
- The z-table helps by telling us what percentage is under the curve at any particular point.

Mid Term

$Q_1 \Rightarrow \mu, \sigma^2$ Discrete Distribution (Binomial / Poisson)

5 μ, σ^2 Continuous Distribution (Normal / Exp. / Uni)

$Q_2 \Rightarrow$ Conditional Dist.

$Q_5 \Rightarrow$ Bayes theorem

$Q_3 \Rightarrow$ Normal Dist.

$Q_4 \Rightarrow$ Uniform dist.

Z Table

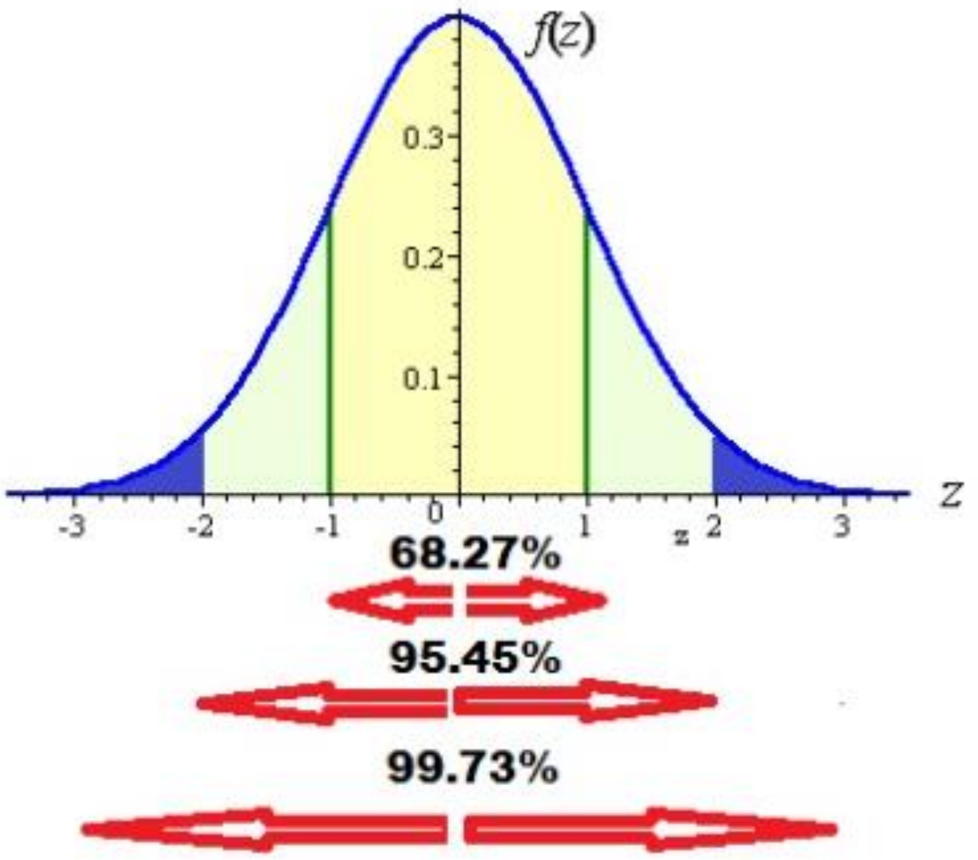


Image Source: https://www.statisticshowto.com/tables/z-table/

Z Table: [Link]

Example

$1 - P(Z < -0.24)$ or $P(-.24 < Z < 0) + 0.5$

Most colleges require applicants for admission to take JEE Council's examination. Scores on the JEE exam are roughly normally distributed with a mean of 527 and a standard deviation of 112. What is the probability of an individual scoring above 500 on the JEE?

Sol^n

$\mu = 527$

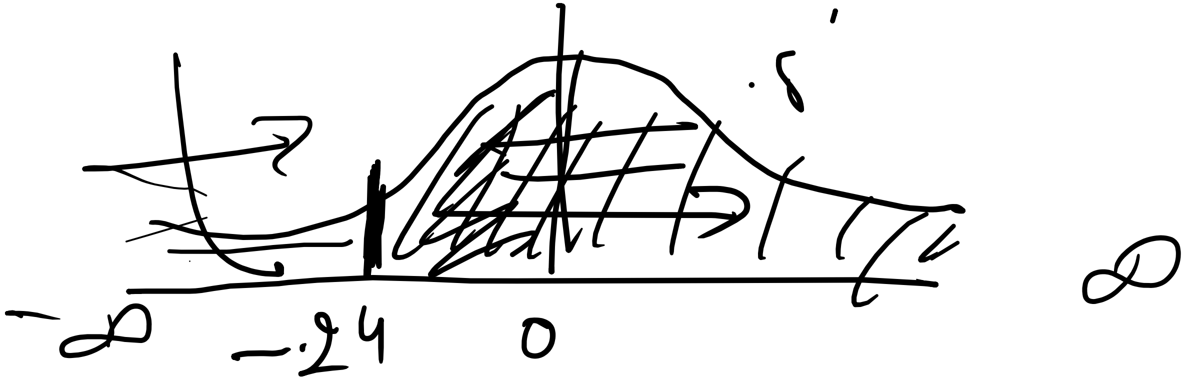
$\sigma = 112$

$$Z = \frac{X - \mu}{\sigma} = \frac{500 - 527}{112} = -0.24$$

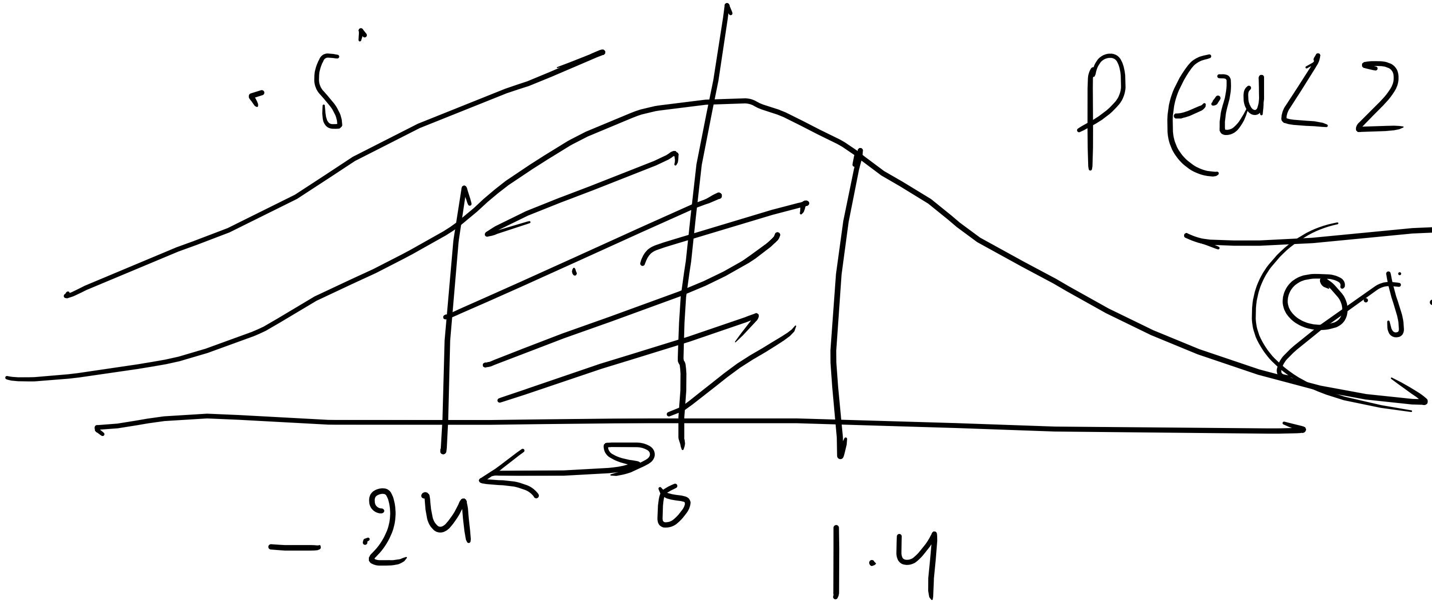
$P(Z > -0.24)$

$P(Z > -0.24)$

0.42



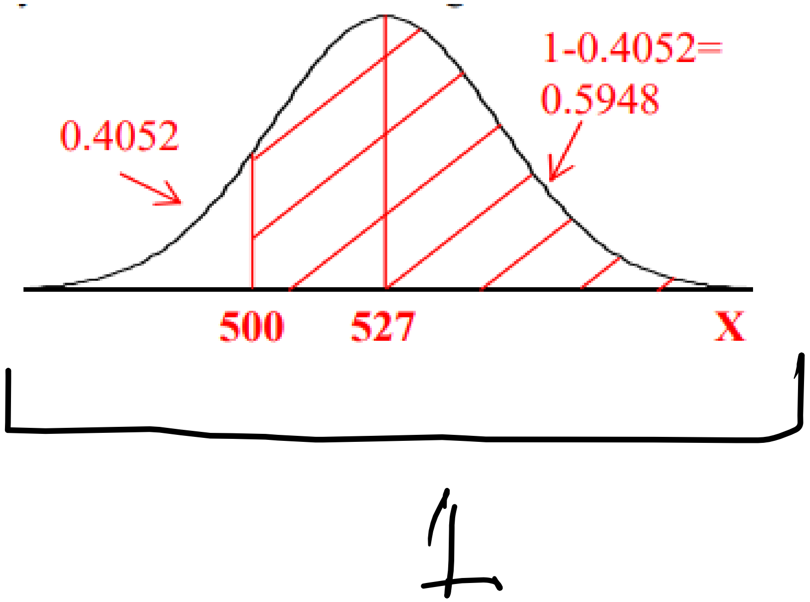
$$P(-0.24 < Z < 1.4)$$



$$P(-.24 < Z < 0) + P(0 < Z < 1.4)$$

$$0.5 - 0.4090 + 0.9.21 - 0.5$$

Solutions



Normal Distribution

$$Z = \frac{500 - 527}{112} = -0.24107$$

$\mu = 527$

$\sigma = 112$

$\Pr\{X > 500\} = \Pr\{Z > -0.24\} = 1 - 0.4052 = \boxed{\mathbf{0.5948}}$

Direct from z score Table: 0.5 + 0.0948

Example

You take the GATE exam and score 1100. The mean score for the GATE Exam is 1026 and the standard deviation is 209. How well did you score on the test compared to the average test taker? Note: It is normally distributed

$\mu = 1026$

$X = 1100$

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

$\sigma = 209$

Solutions

$$Z = \frac{1100 - 1026}{\sigma} \qquad Z = \frac{1100 - \mu}{\sigma} \qquad Z = \frac{1100 - 1026}{209}$$

Z = 0.354   This means that your score was .354 std devs above the mean.

A z-score of .354 is .1368

Example-

If the height of 300 students are normally distributed with mean 64.5 inches and standard deviation 3.3 inches.

How many students have height

1. Less than 5 feet

2. Between 5 feet and 5 feet 9 inches

Example-

The distribution of 500 workers in a factory is approximately Normal with mean and SD Rs 75 and Rs 15, respectively.

Find the No. of workers who receive weekly wages

1. More than 90

2. Less than 45

Exponential Distribution

The exponential distribution is often concerned with the amount of time until some specific event occurs.

Examples:

- Predict the time when an Earthquake might occur
- Life Span of Electronic Gadgets
- Time that an Interviewer spends with a candidate

For this purpose, the only requirement is that the average time that the interviewer takes to finish the interview of previous candidates is well known.

Q:- The length of Telephone Call is exponential variate with mean 3 min. Find Prob that call

(1) End less then 3 min

(2) takes betⁿ 3 to 5 min.

$\lambda = \frac{1}{\mu}$

$\mu = 3$ mins ;

$$\boxed{f(x) = \int_0^\infty \lambda e^{-\lambda x}\, dx}$$

(2) bet^n 3 to 5 min

$$\int_3^5 2e^{-2x}\,dx = \int_3^5 \frac{1}{3}\cdot e^{-1/3x}\,dx$$

$$= \frac{1}{3}\left[\frac{e^{-x/3}}{-1/3}\right] = -\left[e^{-5/3} - e^{-3/3}\right]$$

$$= e^{-1} - e^{-5/3}$$

$$\lambda = \frac{1}{\mu} = \frac{1}{3} \qquad e = 2.7\text{—}$$

$$f(x) = \int_0^{\infty} \lambda \cdot e^{-\lambda x}\,dx \Rightarrow \int_0^{3} \frac{1}{3} \cdot e^{-1/3 \cdot x}\,dx$$

$$= \frac{1}{3}\left[\frac{e^{-1/3 x}}{-1/3}\right]_0^3 = -\left[e^{-3/3} - e^{0}\right]$$

$$= -e^{-1} - 1$$

$$= 1 - 1/e$$

Exponential Distribution

- Notation: X ~ Exp(λ)
- An exponential random variable has pdf and cdf:
- A continuous random variable X is said to have an exponential distribution with parameter λ>0, shown as X~Exp(λ), if its PDF is given by

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

- **λ:** the rate parameter (calculated as λ = 1/μ)
- **e:** A constant roughly equal to 2.718

Exponential Distribution

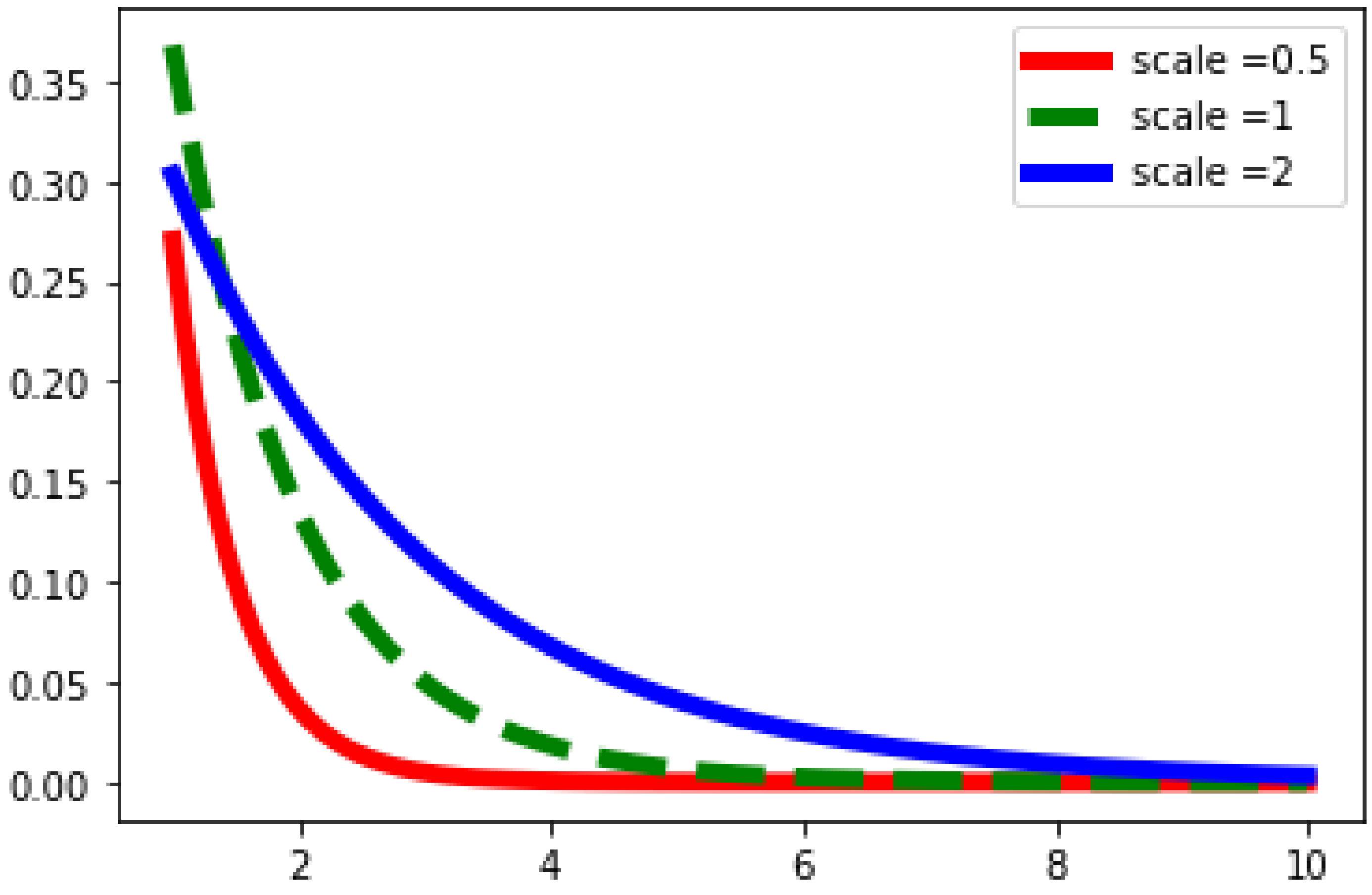
- A continuous random variable X is said to have an exponential distribution with parameter λ>0, shown as X∼Exp(λ), if its CDF is given by

$$CDF \qquad F_x(x) = 1 - e^{-\lambda x} \qquad \text{for x > 0}$$

Or

$$CDF \qquad F_x(x) = \left(1 - e^{-\lambda x}\right)u(x)$$

Exponential Distribution



Mean of Exponential Distribution

$$\mathrm{E}[X] = \int_0^{\infty} x f_X(x)\,dx$$

where

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & \text{otherwise} \end{cases}$$

- **Note:** Check hand written notes for full steps

$$E[X] = \frac{1}{\lambda}$$

Variance of Exponential Distribution

$$\mathrm{E}[X] = \int_0^\infty x f_X(x) dx$$

$$var(X) \ = \ E[X^2] \ - E[X]^2$$

Note: Check hand written notes for full steps

$$Var[X] = \frac{1}{\lambda^2}$$

Example

Suppose the mean number of minutes between eruptions for a certain geyser is 40 minutes. What is the probability that we'll have to wait less than 50 minutes for an eruption?

Solution

$\lambda = 1/\mu$
$\lambda = 1/40$
$\lambda = .025$

We can plug in $\lambda = .025$ and $x = 50$ to the formula for the CDF:

$P(X \leq x) = 1 - e^{-\lambda x}$
$P(X \leq 50) = 1 - e^{-.025(50)}$
$P(X \leq 50) = 0.7135$

The probability that we'll have to wait less than 50 minutes for the next eruption is **0.7135**.

Note: See Python code as well

Example

Let X = amount of time (in minutes) a postal clerk spends with his or her customer. The time is known to have an exponential distribution with the average amount of time equal to four minutes. Find the probability that a clerk spends four to five minutes with a randomly selected customer.

Solution

We need to find $P(4 < x < 5)$.

$\lambda = 1/\mu = ¼ = 0.25$

We can plug in $\lambda = .25$ and $x = 4$ to 5

$P(4 < x < 5) = P(x < 5) – P(x < 4)$

$P(X \leq x) = 1 – e^{-\lambda x}$
$P(X \leq 5) = 1 – e^{-.25(5)}$
$P(X \leq 5) = 0.7135$

Similarly, $P(X \leq 4) = 0.0814$
Hence, $P(4 < x < 5) = P(x < 5) – P(x < 4) = 0.7135 − 0.6321 = 0.0814$.