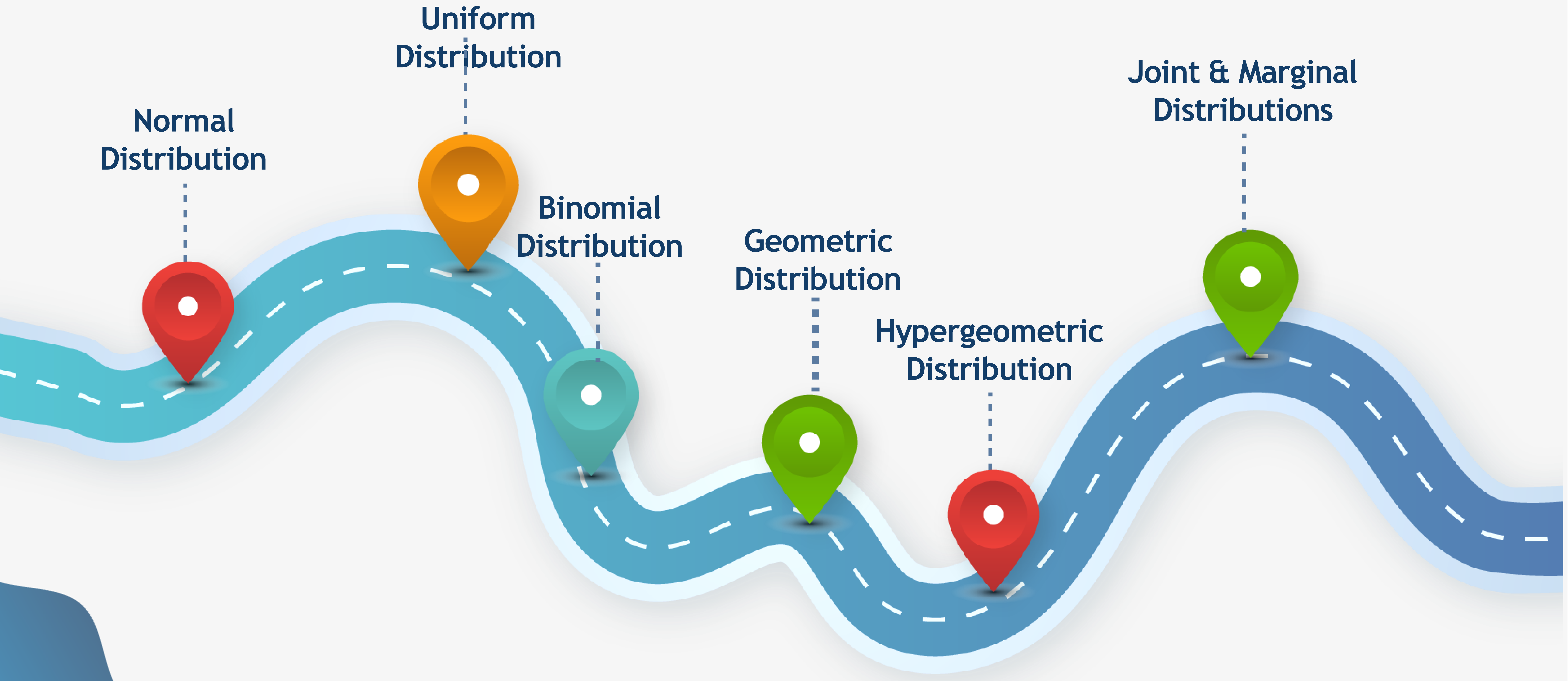
The Learning Hub

Data Science Fundamentals
Distributions

The Learning Hub



The Learning Hub

Normal Distribution

Normal Distribution Definition

- **Definition:**

$A\ random\ variable\ X\ is\ said\ to\ have\ a\ normal\ distribution\ with\ parameters\ \mu\ and\ \sigma^2$

$$if\ it\ has\ the\ density: f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < \infty$$

- We write as X ~N(μ,σ^2) and if X ~N(1,0) we call it a standard normal variable.
- The density of a standard normal variable is denoted as $\Phi(x)$ and equals the function:

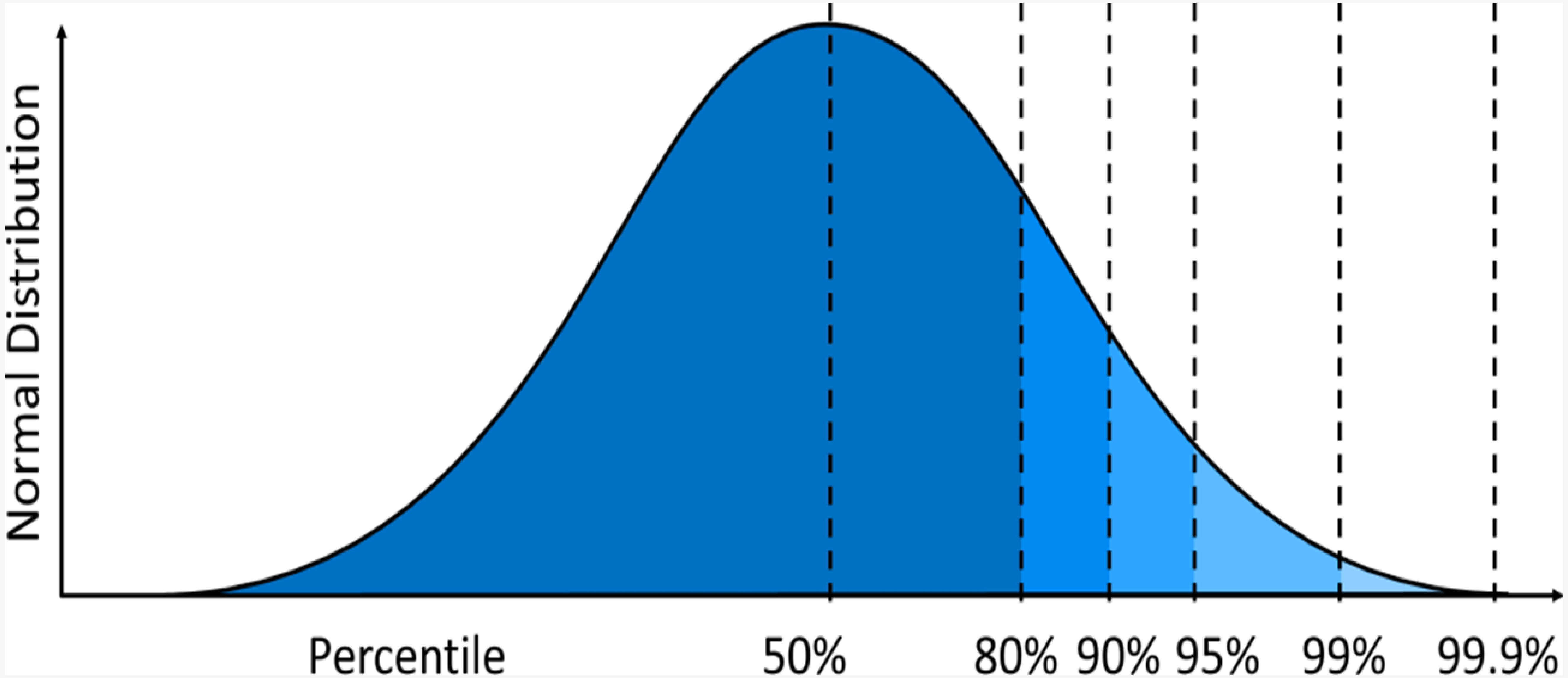
$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, -\infty < x < \infty$$

Normal Distribution Definition

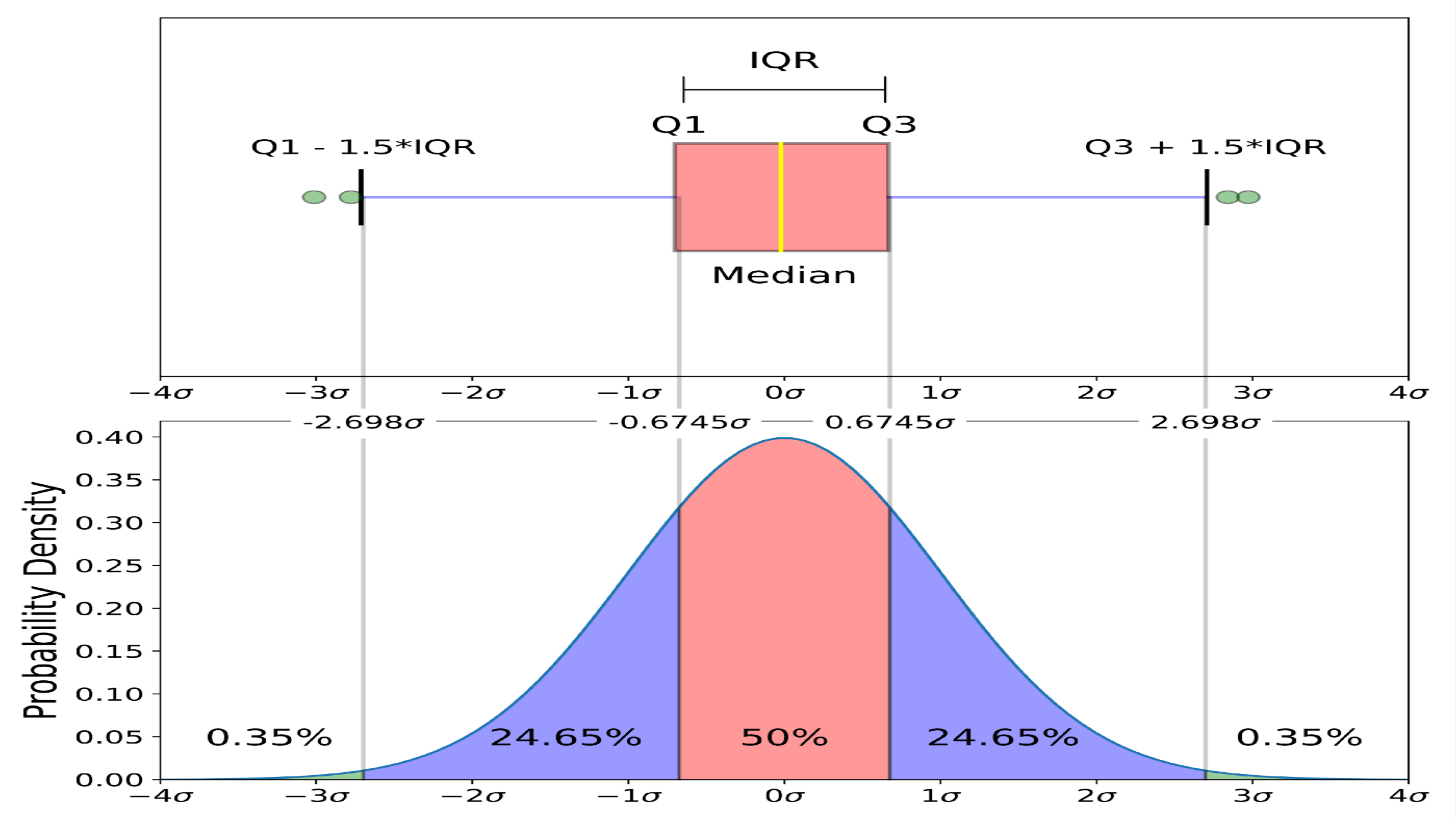
- The CDF is denoted as $\Phi(x)$, note that the standard normal density is symmetric and unimodal about zero. The general N(μ,σ^2) density is symmetric and unimodal about μ.
- The CDF cannot be written in terms of elementary functions but can be computed at a given value x.
- Because of the symmetry of the standard normal distribution about zero, we have:

$$\phi(-x) = 1 - \phi(x)\ \forall x.$$

Normal Distribution Definition



Normal Distribution Quantiles



Normal Distribution Properties

- $if\ X \sim N(\mu, \sigma^2), then\ Z = \frac{x-\mu}{\sigma} \sim N(0,1)\ and\ if\ Z \sim N(0,1)\ then\ X = \mu + \sigma Z \sim N(\mu, \sigma^2)$

- $\text{N}ormal\ random\ variable, then\ its\ standardized\ version\ is\ always\ a\ standard\ normal\ variable.$

- $if\ X \sim N(\mu, \sigma^2), then\ P(X = x) = \Phi\left(\frac{x-\mu}{\sigma}\right)\ \forall\ x\ , in\ particular, P(X \leq \mu) = P(Z \leq 0.5);$

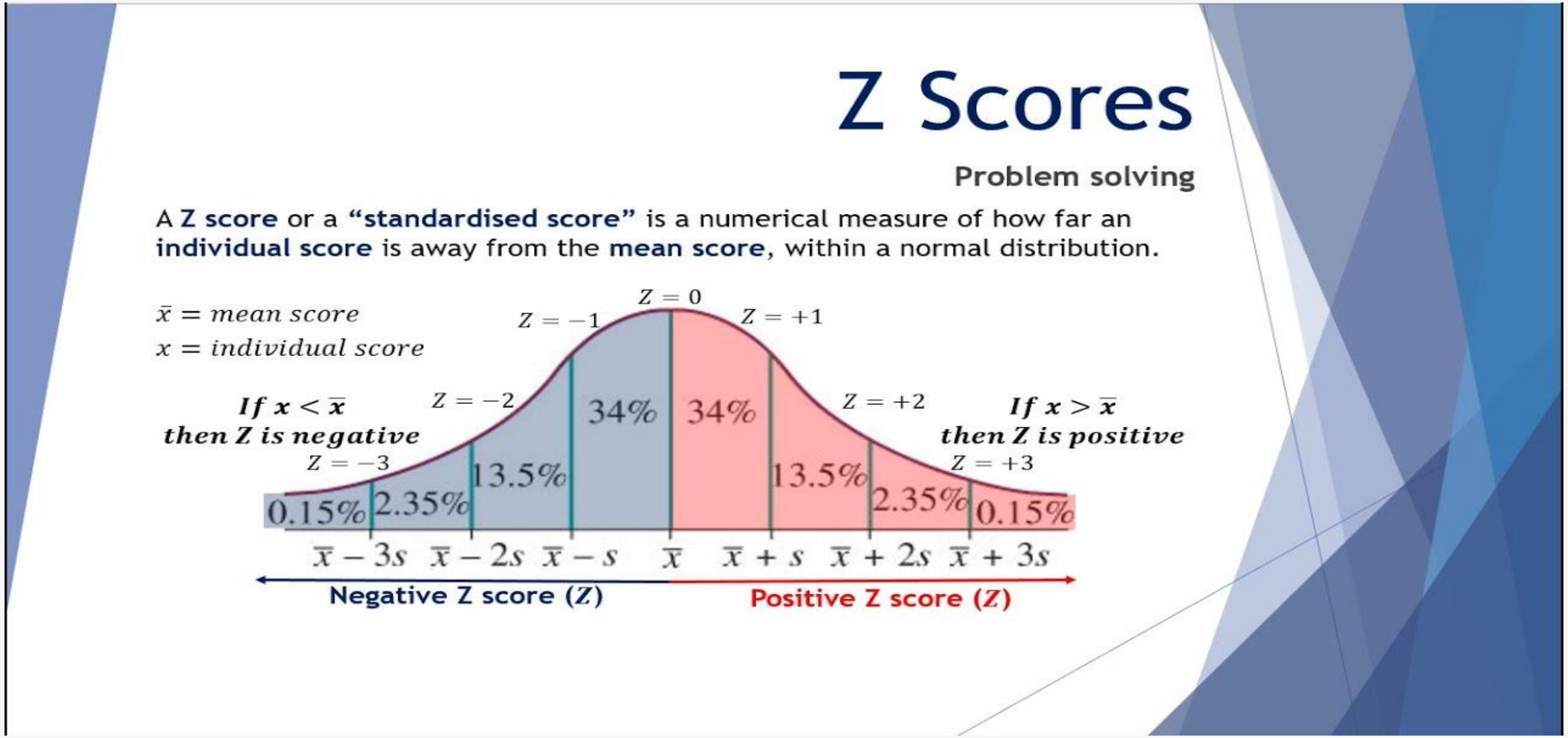
  $$the\ median\ of\ x = \mu$$

Normal Distribution Properties (Z-Values)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
–3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
–3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
–3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
–3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
–3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
–2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
–2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
–2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
–2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
–2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
–2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
–2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
–2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
–2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
–2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
–1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
–1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
–1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
–1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
–1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
–1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
–1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
–1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
–1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
–1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
–0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
–0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
–0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
–0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
–0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
–0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
–0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
–0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
–0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
–0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

Normal Distribution Properties (Z-Values)



Normal Distribution Examples

1. Given a random variable X having a normal distribution with $\mu = 50$ and $\sigma = 10$, find the probability that X assumes a value between 45 and 62.

- **Solution:**

  The z values corresponding to $x1 = 45$ and $x2 = 62$ are $z1 = 45 - 50 / 10 = -0.5$ and $z2 = 62 - 50 / 10 = 1.2$. Therefore, $P(45 < X < 62) = P(-0.5 < Z < 1.2)$.

  Using the Normal Distribution table

  $P(45 < X < 62) = P(-0.5 < Z < 1.2) = P(Z < 1.2) - P(Z < -0.5) = 0.8849 - 0.3085 = 0.5764$

Normal Distribution Examples

2. Given that X has a normal distribution, if the mean = 300 and the standard deviation = 50, find the probability that X assumes a value greater than 362.

- **Solution:**

$z = 362 - 300 / 50 = 1.24.$

Hence,

$P(X > 362) = P(Z > 1.24) = 1 - P(Z < 1.24) = 1 - 0.8925 = 0.1075$

Normal Distribution Examples

3. Given that X has a normal distribution with a mean = 40 and a standard deviation of 6, find the value of x such that it has.

A. 45% of the area to the left and

B. 14% of the area to the right.

Normal Distribution Examples

- **Solution:**

(a)An area of 0.45 to the left of the desired x value We require a z value that leaves an area of 0.45 to the left. From normal Table we find $P(Z < -0.13) = 0.45$, so the desired z value is -0.13. Hence, $x = (6)(-0.13) + 40 = 39.22$.

(b)An area equal to 0.14 to the right of the desired x value. This time we require a z value that leaves 0.14 of the area to the right and hence an area of 0.86 to the left. Again, from normal table we find $P(Z < 1.08) = 0.86$, so the desired z value is 1.08 and $x = (6)(1.08) + 40 = 46.48$

Normal Distribution Examples

4. An electronic components factory produces electrical resistors with a mean resistance of 40 ohms and a standard deviation of two ohms. Suppose that the resistance follows a normal distribution. find the percentage of resistors that will have a resistance exceeding 43 ohms?

- **Solution:**

$$z = 43 - 40/2 = 1.5$$

$$P(X > 43) = P(Z > 1.5) = 1 - P(Z < 1.5) = 1 - 0.9332 = 0.0668$$

Normal Distribution Examples

5. Assuming the average life of a certain type of storage battery is 3.0 years with a standard deviation of 0.5 year. suppose that battery life is normally distributed, find the probability that a given battery will last less than 2.3 years.

- **Solution:**

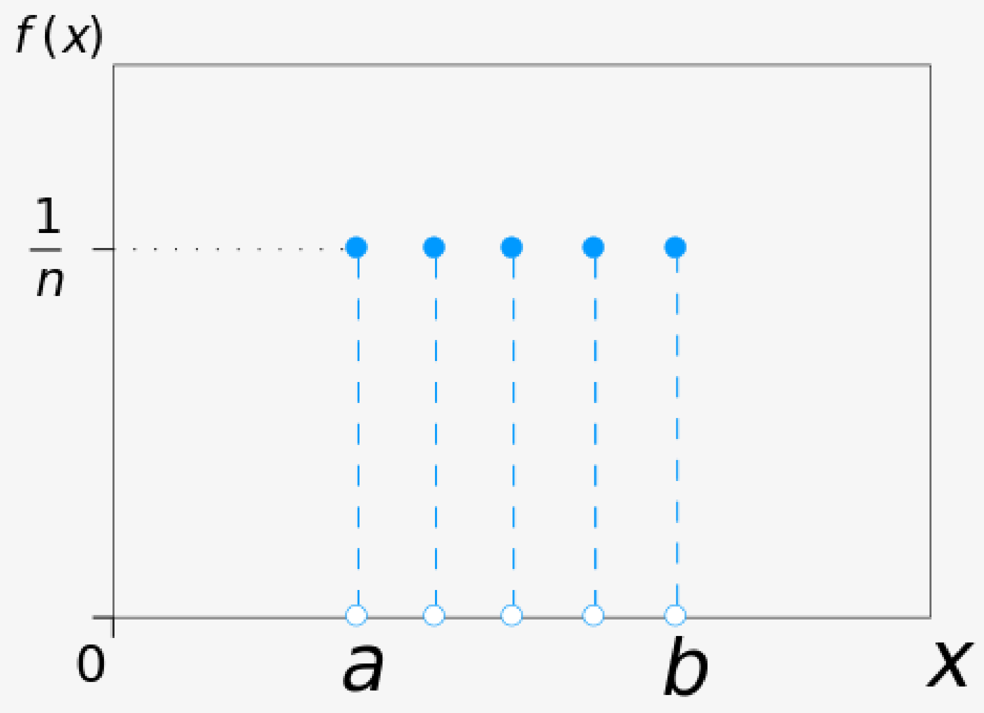
$z = (2.3 - 3)/0.5 = -1.4$

$P(X < 2.3) = P(Z < -1.4) = 0.0808$

The Learning Hub

Uniform Distribution

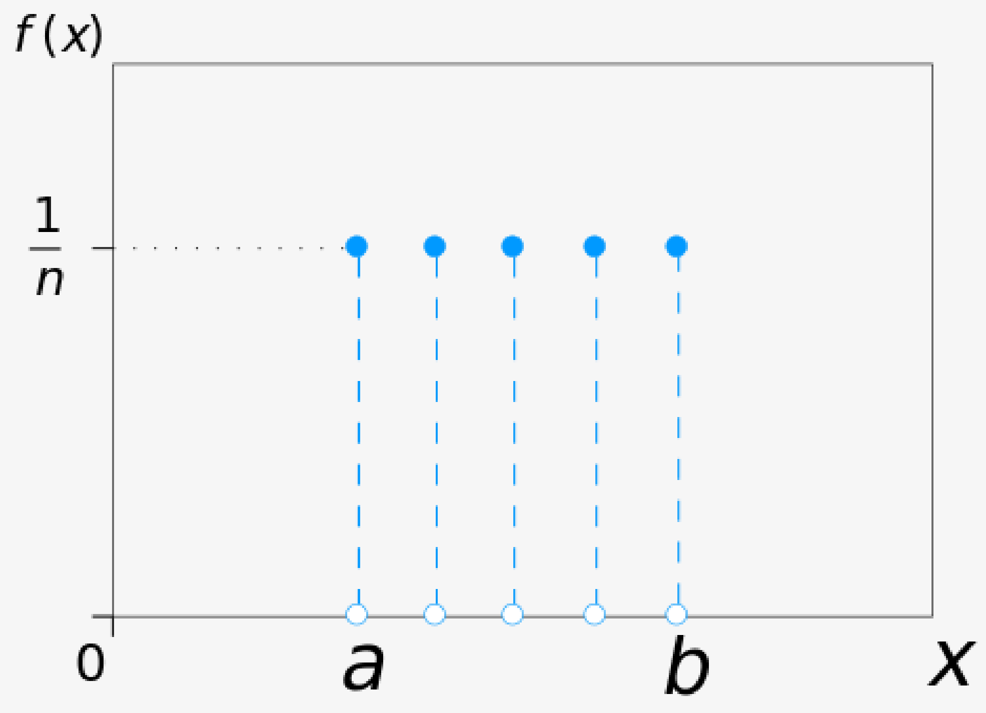
Uniform Distribution Definition



- Rectangular Distribution
- Represents a finite number of values are equally likely to happen
- It can also be written as $X \sim \text{Unif}\{1,2,..,n\}$
- **Definition :** The uniform distribution on a sample space Ω containing n elements is the function m defined by $m(\omega) = 1/n$ For every $\omega \in \Omega$

Uniform Distribution Properties

- $\mathbf{Let\ X \sim Unif\{1, 2, \dots, n\}. Then,}$
- $\boldsymbol{\mu = E(X) = \frac{N+1}{2}}$
- $\boldsymbol{\sigma^2 = var(X) = \frac{n^2-1}{12}}$
- $\boldsymbol{E(X-\mu)^3 = 0}$
- $\boldsymbol{E(X-\mu)^4 = \frac{(3n^2-7)(n^2-1)}{240}}$



The Learning Hub

Binomial Distribution

Binomial Distribution Definition

- Represents a sequence of independent coin-tossing experiments.
- Suppose a coin with probability p for getting heads is tossed n times where n >= 1. Let X be the number of times a toss

in those n tosses resulted in a head. PMF of X is then

$$\mathbf{P(X = x) = \binom{n}{x} p^{x}(1-p)^{n-x}} \qquad , x = 0,1, \ldots, n$$

Binomial Distribution Properties

- $\mu = E(X) = np$ $\quad\quad$ $\sigma^2 = var(X) = np(1-p)$
- The mgf of X equals $\psi(t) = (pe^t + 1 - p)^n$ at any t
- $E\left[(X-\mu)^3\right] = np\left(1 - 3p + 2p^2\right)$
- $E\left[(X-\mu)^4\right] = np(1-p)[1 + 3(n-2)p(1-p)]$

Binomial Distribution Examples

1. Suppose we have a biased coin that comes up heads with the probability of 2/3. This coin is flipped 5 times. What is the probability that it lands on heads exactly three times?

- **Solution:**

$$P=\binom{5}{3}\left(\frac{2}{3}\right)^{3}\left(\frac{1}{3}\right)^{2}=5*\left(\frac{8}{27}\right)*\left(\frac{1}{9}\right)=\frac{40}{243}$$

Binomial Distribution Examples

2. let's assume that according to the latest police reports, 80% of all petty crimes are unresolved, and in your town, at least 3 of such petty crimes are committed. The 3 crimes are all independent of each other. From the given data, what is the probability that one of the three crimes will be resolved?

- **Solution:**

N=3 , p(crime-solving)=0.2. The probability that one of the three crimes will be resolved=

$$\binom{3}{1}\left(\frac{2}{10}\right)^{1}\left(\frac{8}{10}\right)^{2}=0.0384$$

Negative Binomial Distribution Definition

- It is a generalization of the geometric distribution.
- In the same experiment, we toss the coin with the same probability p until we get r number of heads. When r = 1, the negative binomial distribution is equal to the geometric distribution.
- Let X be the number of the first toss at which the rth success is obtained. The PMF of X then

$$\boldsymbol{P(X=x)=\binom{x-1}{r-1}p^{r}(1-p)^{x-r}}, x=r, r+1, \ldots$$

Negative Binomial Distribution Properties

- $\mathbf{E(X) = \frac{r}{p}}$
- $\mathbf{var(X) = \frac{rq}{p^2}}$

26

Negative Binomial Distribution Example

- A multiple-choice questions exam with probability of attempting the question with the right answer is 70%. What is the probability that a student gives the second correct answer for the fourth attempted question?
- **Solution:**

P=0.7 ,1-p=0.3 ,r=2 , x=4

$$\boldsymbol{P(X=4)} = \binom{\mathbf{4-1}}{\mathbf{2-1}} \mathbf{0.7^2\,(0.3)^2} = 3 * 0.49 * 0.09 = 0.1323$$

The Learning Hub

Geometric Distribution

Geometric Distribution Properties

- Assume the coin tossing experiment with probability p where $0 < p < 1$ is the probability of getting a heads.
- We toss the coin repeatedly until the first time we get a heads. Let X be the number of the toss at which the first heads is obtained. Then the PMF of X is:

$$\boldsymbol{P(X = x) = p(1-p)^{x-1}}, \qquad x = 1,2,3,..$$

- X has a geometric distribution with parameter p written as $X \sim Geo(p)$

Hypergeometric Distribution

Hypergeometric Distribution Properties

- This distribution also specifies the number of successes in a prespecified number of Bernoulli trials, but in this case, the trials happen to be dependent.

- Assume a population of N objects. There are D objects of type 1 and N-D objects of type 2.

We choose n objects from the population without replacement and at random where 1 ≤n ≤N

- Let X be the number of individuals of type 1 among the n units chosen. Then the PMF of X is

$$P(X = x) = \frac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}}$$

Hypergeometric Distribution Example

- A deck of cards contains 10 cards: 6 red cards and 4 black cards. 6 cards are drawn randomly without replacement. What is the probability that exactly 3 red cards are drawn?

- **Solution:**

$$P(X = 3) = \frac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}} = \frac{\binom{6}{3}\binom{10-6}{6-3}}{\binom{10}{6}} = 20*4/210 = 8/21$$

The Learning Hub

Joint & Marginal Distribution

Joint Distribution Properties

- Deals with handling distributions of simultaneous random variables.
- The Join probability distribution of two random variables can be explained using the

$$f(x, y) = P(X = x, Y = y)$$

Joint Distribution Properties

- The function f(x,y) is a joint probability distribution or probability mass function of the discrete random variables X and Y if :
  - $f(x,y) \geq 0 for\ all\ (x,y)$
  - $\sum_x \sum_y f(x,y) = 1$
  - $P(X = x, Y = y) = f(x,y)$

$for\ any\ region\ A\ in\ the\ xy\ plane, P[(x,y) \in A] = \sum \sum_A f(x,y)$

Joint Distribution Properties

- The function f(x,y) is a joint density function of the continuous random variables X and Y if :

  - $f(x,y) \geq 0 for\ all\ (x,y)$
  - $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dx\ dy = 1$
  - $P[(x,y) \in A] =$
    $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\ dx\ dy\ for\ any\ region\ A\ in\ the\ xy\ plane$

Marginal Distribution Properties

- The marginal distribution of X alone and of Y alone are:
  - For the discrete case:

$$\mathbf{g(x)} = \sum_{\mathbf{y}} \mathbf{f(x, y)} \quad \& \quad \mathbf{h(y)} = \sum_{\mathbf{x}} \mathbf{f(x, y)}$$

  - For the continuous case:

$$\mathbf{g(x)} = \int_{-\infty}^{\infty} \mathbf{f(x, y)dy} \;\; \& \;\; \mathbf{h(y)} = \int_{-\infty}^{\infty} \mathbf{f(x, y)dx}$$

37

Marginal Distribution Properties

- Let X and Y be two random variables, discrete or continuous, with joint probability distribution f(x,y) and marginal distributions g(x) and h(y), retrospectively. The random variables X and Y are said to statistically independent if and only if:

$$f(x,y) = g(x)h(y)$$

for all (x,y) within their range.

38

Marginal Distribution Properties

- Let X and Y be two random variables, discrete or continuous, with joint probability Let X and Y be two random variables, discrete or continuous.
  - The conditional distribution of the random variables Y given that X=x is:

$$f(y|x) = \frac{f(x,y)}{g(x)}, provided\ g(x) > 0$$

  - The conditional distribution of the random variables X given that Y=y is:

$$f(x|y) = \frac{f(x,y)}{h(y)}, provided\ h(y) > 0$$

39

Marginal Distribution Properties

- Let X and Y be two random variables, discrete or continuous, with joint probability Let X and Y be two random variables, discrete or continuous.
  - The conditional distribution of the random variables Y given that X=x is:

$$f(y|x) = \frac{f(x,y)}{g(x)}, provided\ g(x) > 0$$

  - The conditional distribution of the random variables X given that Y=y is:

$$f(x|y) = \frac{f(x,y)}{h(y)}, provided\ h(y) > 0$$

Joint & Marginal Distribution Example

	Male	Female	Total
Football	120	75	195
Rugby	100	25	125
Other	50	130	180
	270	230	500

Probability Distribution

	Male	Female	Total
Football	0.24	0.15	0.39
Rugby	0.2	0.05	0.25
Other	0.1	0.26	0.36
	0.54	0.46	1

Joint Probability Distribution

	Male	Female	Total
Football	0.24	0.15	0.39
Rugby	0.2	0.05	0.25
Other	0.1	0.26	0.36
	0.54	0.46	1

Marginal Probability Distribution

	Male	Female	Total
Football	0.24	0.15	0.39
Rugby	0.2	0.05	0.25
Other	0.1	0.26	0.36
	0.54	0.46	1

References

1. [1] https://towardsdatascience.com/understanding-boxplots-5e2df7bcbd51
2. https://www.allaboutlean.com/storage-strategies-fixed-location/normal-distribution-and-percentiles/
3. How to Detect and Remove Outliers | Outlier Detection And Removal (analyticsvidhya.com)
4. Marginal, Joint and Conditional Probabilities explained By Data Scientist

THANK YOU