

MATH2361: Probability and Statistics 2024-25

Instructor: **Dr. Harsha K V**, Department of Mathematics

GITAM Hyderabad

Short Notes and Formula list for Unit 3

Let X be a continuous random variable with **probability density function (pdf)** denoted by $f(x)$ (or more precisely $f_X(x)$). Note that $f(x)$ must satisfies two conditions:

- 1) $f(x) \geq 0$ for all $x \in \mathbb{R}$, where \mathbb{R} denote the set of real numbers.
- 2) $\int_{-\infty}^{\infty} f(x)dx = 1$.

For a random variable with pdf $f(x)$, probability of any event is computed by

$$P(a < X < b) = P(a \leq X \leq b) = \int_a^b f(x)dx, \quad \text{where } a < b. \quad (1)$$

Similarly, we have for any $c \in \mathbb{R}$, probability is computed as

$$P(X \leq c) = P(X < c) = \int_{-\infty}^c f(x)dx \quad (2)$$

$$P(X \geq c) = P(X > c) = \int_c^{\infty} f(x)dx. \quad (3)$$

For X with pdf $f(x)$, the **cumulative distribution function (CDF)** denoted by $F(x)$ (or $F_X(x)$) is defined as:

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t)dt, \quad \text{for all } x \in \mathbb{R}. \quad (4)$$

Once we know the formula for CDF $F(x)$, for any numbers a, b , we can compute

$$P(a \leq X \leq b) = F(b) - F(a), \quad a < b. \quad (5)$$

For a continuous random variable X with pdf $f(x)$, **Expectation** $E(X)$, **Variance** $\text{Var}(X)$ are defined as

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx \quad (6)$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx \quad (7)$$

$$\text{Var}(X) = E((X - \mu)^2) = E(X^2) - (E(X))^2, \quad \mu = E(X). \quad (8)$$

List of Distributions: In this course, we will learn about the following distributions:

- 1) Binomial distribution
- 2) Poisson distribution
- 3) Geometric distribution
- 4) Hypergeometric distribution
- 5) Exponential distribution
- 6) Normal distribution

Binomial distribution: Consider a random experiment with exactly two possible outcomes, “**success**” and “**failure**”. Such an experiment is called a **Bernoulli trial**. We denote the probability of success by p , then probability of failure is $1 - p$, where $0 \leq p \leq 1$.

- We repeat this Bernoulli trial n number of times.
- We assume that the trials are independent
- The probability of success is the same across all the trials.

Now define a random variable X as **the number of successes in n trials**. Here X can take values $i = 0, 1, \dots, n$. Hence this is a discrete random variable. The probability mass function $p(x)$ of X is called the **binomial distribution with parameters n and p** given by

$$p(i) = P(X = i) = \binom{n}{i} p^i (1 - p)^{n-i}, \quad \text{where} \quad \binom{n}{i} = \frac{n!}{i!(n-i)!}.$$

The binomial distribution is denoted by $B(n, p)$, and we write $X \sim B(n, p)$ which reads as “ X follows binomial distribution $B(n, p)$ ”. For a binomial random variable X , we have

$$E(X) = np$$

$$\text{Var}(X) = np(1 - p).$$

Poisson distribution: The random variable X is said to have Poisson distribution with a real parameter $\lambda > 0$ if X has the probability mass function $p(x)$ given by

$$p(i) = P(X = i) = \frac{e^{-\lambda} \lambda^i}{i!}, \quad \text{where } i = 0, 1, 2, \dots$$

Here X can take countably infinite values, hence X is a discrete random variable. The Poisson distribution is denoted by $\text{Pois}(\lambda)$, and we write $X \sim \text{Pois}(\lambda)$. Here the parameter λ is called the rate parameter. Poisson random variable can be used to model given number of events occurring in a fixed interval of time (or space) if these events occur with a known constant mean rate and independently of the time since the last event. For a Poisson random variable X , we have

$$E(X) = \lambda$$

$$\text{Var}(X) = \lambda.$$

Geometric distribution: Consider a Bernoulli trial with probability of success taken as p . Then define a random variable X **as the number of Bernoulli trials to get first success**. Here X takes countably infinite values from $1, 2, \dots$, and hence it is a discrete random variable. X follows the geometric distribution with parameter p , and we denote it as $X \sim G(p)$. Then probability mass function is given by

$$p(i) = P(X = i) = (1 - p)^{i-1} p, \quad i = 1, 2, \dots$$

The expectation and variance are given by

$$E(X) = \frac{1}{p}$$

$$\text{Var}(X) = \frac{1 - p}{p^2}.$$

Hypergeometric distribution: Consider a finite population with size N which contains K items with some features, which are described as “successes”. Suppose we draw n items from this population *without replacement*. Then define a random variable X **as the number of successes in n draws from the finite population (without replacement)**. Then X follows the hypergeometric distribution with parameters N, K, n , and the probability mass function is given by

$$p(i) = P(X = i) = \frac{\binom{K}{i} \binom{N-K}{n-i}}{\binom{N}{n}}$$

where, $i = a$ to b with

$$a = \max \{0, n - (N - K)\}$$

$$b = \min \{n, K\}.$$

Mean and variance are given by

$$\begin{aligned} E(X) &= \frac{nK}{N} \\ \text{Var}(X) &= \frac{nK}{N} \frac{(N - K)(N - n)}{N(N - 1)} \end{aligned}$$

Exponential distribution: A random variable X is said to follow an exponential distribution if the pdf of X is given by

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0.$$

It is a continuous random variable as X takes uncountable values $[0, \infty)$. Exponential random variable can be used to model the time until an event occurs if the events occur independently and continuously over time, and the rate of occurrence λ is constant over time. Examples: Time until a customer arrives at a service point, time until a light bulb burns out, time between phone calls at a call center, time between radioactive particle emissions.

Expectation, variance, and CDF are given by

$$\begin{aligned} E(X) &= \frac{1}{\lambda} \\ \text{Var}(X) &= \frac{1}{\lambda^2} \\ F(x) &= P(X \leq x) = 1 - e^{-\lambda x}, \quad x \geq 0. \end{aligned}$$

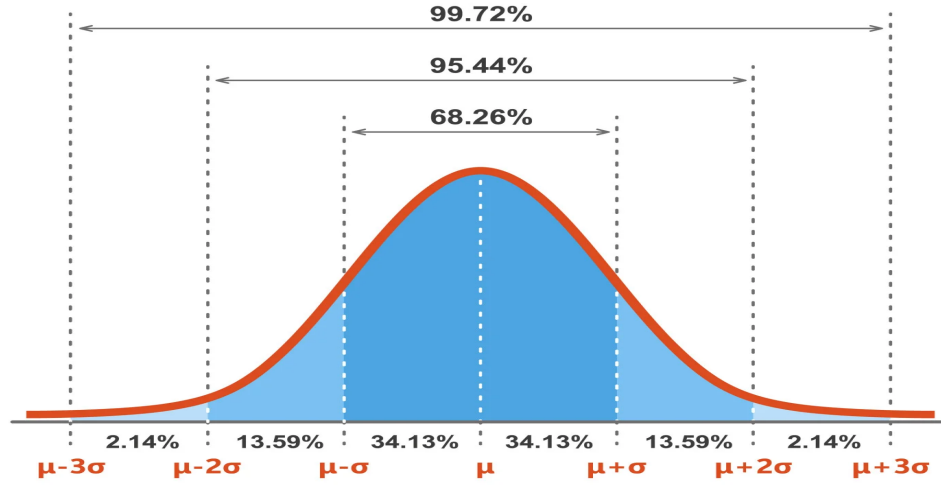
Normal distribution: The random variable X is said to have Normal distribution with real parameters $\mu \in \mathbb{R}, \sigma^2 > 0$ if X has the probability density function $f(x)$ given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x - \mu)^2}{2\sigma^2}\right) \quad \text{where } x \in \mathbb{R}.$$

X is a continuous random variable as $x \in \mathbb{R}$. The Normal distribution is denoted by $N(\mu, \sigma^2)$, and we write $X \sim N(\mu, \sigma^2)$. For a normal random variable X , we have

$$\begin{aligned} E(X) &= \mu \\ \text{Var}(X) &= \sigma^2. \end{aligned}$$

Here, σ is the standard deviation. For normal distribution $N(\mu, \sigma^2)$, approximately 68%, 95%, and 99.7% of the values lie within one, two, and three standard deviations (σ) of the mean (μ), respectively, as depicted in the figure.



Normal pdf is symmetric around μ , and mean, mode and median of the normal distribution are the same which is μ .

For $\mu = 0$ and $\sigma = 1$, the normal distribution $N(0, 1)$ is called the **standard normal distribution**. The standard normal random variable is usually denoted by Z , and we write $Z \sim N(0, 1)$. The pdf of the standard normal variable Z is given by

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \quad \text{where } z \in \mathbb{R}.$$

Let random variable $X \sim N(\mu, \sigma^2)$. Then one can always convert X to the standard normal random variable Z by the change of variable $Z = \frac{X-\mu}{\sigma}$. Then it is easy to see from the properties of expectation that

$$\begin{aligned} E(Z) &= E\left(\frac{X}{\sigma} - \frac{\mu}{\sigma}\right) = \frac{1}{\sigma}E(X) - \frac{\mu}{\sigma} = 0 \\ \text{Var}(Z) &= \text{Var}\left(\frac{X}{\sigma} - \frac{\mu}{\sigma}\right) = \frac{1}{\sigma^2}\text{Var}(X) = \frac{\sigma^2}{\sigma^2} = 1. \end{aligned}$$

Hence, we have

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

For any $a > 0$, the probability $P(0 < Z < a)$ can be computed from the standard normal table, see the Z-table. Example, for $a = 1.32 = 1.3 + 0.02$, from the table first look for the row starting

with 1.3, then in that row, look for the column with 0.02, the intersection value is 0.40658. Hence $P(0 < Z < 1.32) = 0.40658$. Due to the symmetry of normal distribution, we have

$$P(Z < 0) = P(Z > 0) = 0.5 \quad (9)$$

$$P(0 < Z < a) = P(-a < Z < 0), \quad a > 0 \quad (10)$$

$$P(Z < a) = 0.5 + P(0 < Z < a), \quad a > 0. \quad (11)$$

Example 9:: Let X be a random variable which follows normal distribution with mean $\mu = 30$ and standard deviation $\sigma = 4$. Find the following probabilities: i) $P(X < 40)$ and ii) $P(X > 21)$?
Answer: Given that $X \sim N(30, 16)$. First we convert X to $Z = \frac{X-\mu}{\sigma} = \frac{X-30}{4}$.


1) To calculate $P(X < 40)$, we have

$$\begin{aligned} P(X < 40) &= P\left(\frac{X-30}{4} < \frac{40-30}{4}\right) \\ &= P(Z < 2.5) \\ &= 0.5 + P(0 < Z < 2.5) \\ &= 0.5 + 0.49379 \quad (\text{value from the Z-table}) \\ &= 0.99379. \end{aligned}$$

2) To calculate $P(X > 21)$, we have

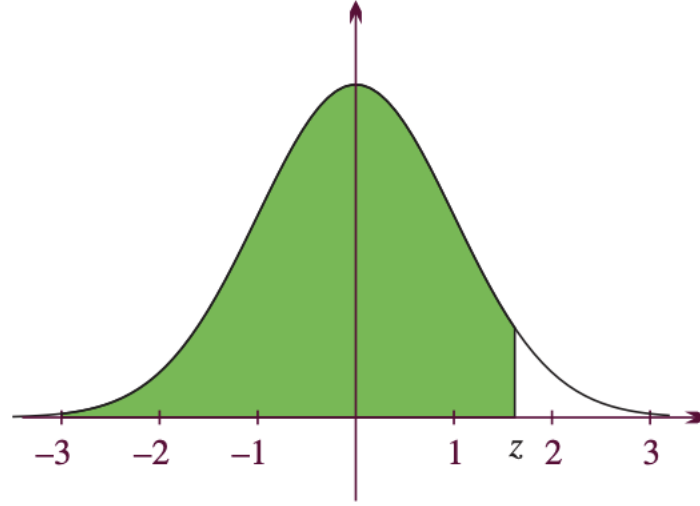
$$\begin{aligned} P(X > 21) &= P\left(\frac{X-30}{4} > \frac{21-30}{4}\right) \\ &= P(Z > -2.25) \\ &= P(-2.25 < Z < 0) + 0.5 \\ &= P(0 < Z < 2.25) + 0.5 \quad (\text{from Eq (10)}) \\ &= 0.48778 + 0.5 \quad (\text{value from the Z-table}) \\ &= 0.98778. \end{aligned}$$

A. Z-table for Standard Normal Distribution for $P(0 \leq Z \leq z)$

Z-table										
					<p>Entries in the table give the percentage of the area under the curve from zero to the z-score. For example, if $z = 1.32$ the area under the curve between 0 and 1.32 is 0.40658. If $z = -1.32$, the area under the curve between -1.32 and 0 is still 0.40658.</p>					
$\pm z$	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0	0.00399	0.00798	0.01197	0.01595	0.01994	0.02392	0.0279	0.03188	0.03586
0.1	0.03983	0.0438	0.04776	0.05172	0.05567	0.05962	0.06356	0.06749	0.07142	0.07535
0.2	0.07926	0.08317	0.08706	0.09095	0.09483	0.09871	0.10257	0.10642	0.11026	0.11409
0.3	0.11791	0.12172	0.12552	0.1293	0.13307	0.13683	0.14058	0.14431	0.14803	0.15173
0.4	0.15542	0.1591	0.16276	0.1664	0.17003	0.17364	0.17724	0.18082	0.18439	0.18793
0.5	0.19146	0.19497	0.19847	0.20194	0.2054	0.20884	0.21226	0.21566	0.21904	0.2224
0.6	0.22575	0.22907	0.23237	0.23565	0.23891	0.24215	0.24537	0.24857	0.25175	0.2549
0.7	0.25804	0.26115	0.26424	0.2673	0.27035	0.27337	0.27637	0.27935	0.2823	0.28524
0.8	0.28814	0.29103	0.29389	0.29673	0.29955	0.30234	0.30511	0.30785	0.31057	0.31327
0.9	0.31594	0.31859	0.32121	0.32381	0.32639	0.32894	0.33147	0.33398	0.33646	0.33891
1	0.34134	0.34375	0.34614	0.34849	0.35083	0.35314	0.35543	0.35769	0.35993	0.36214
1.1	0.36433	0.3665	0.36864	0.37076	0.37286	0.37493	0.37698	0.379	0.381	0.38298
1.2	0.38493	0.38686	0.38877	0.39065	0.39251	0.39435	0.39617	0.39796	0.39973	0.40147
1.3	0.4032	0.4049	0.40658	0.40824	0.40988	0.41149	0.41308	0.41466	0.41621	0.41774
1.4	0.41924	0.42073	0.4222	0.42364	0.42507	0.42647	0.42785	0.42922	0.43056	0.43189
1.5	0.43319	0.43448	0.43574	0.43699	0.43822	0.43943	0.44062	0.44179	0.44295	0.44408
1.6	0.4452	0.4463	0.44738	0.44845	0.4495	0.45053	0.45154	0.45254	0.45352	0.45449
1.7	0.45543	0.45637	0.45728	0.45818	0.45907	0.45994	0.4608	0.46164	0.46246	0.46327
1.8	0.46407	0.46485	0.46562	0.46638	0.46712	0.46784	0.46856	0.46926	0.46995	0.47062
1.9	0.47128	0.47193	0.47257	0.4732	0.47381	0.47441	0.475	0.47558	0.47615	0.4767
2	0.47725	0.47778	0.47831	0.47882	0.47932	0.47982	0.4803	0.48077	0.48124	0.48169
2.1	0.48214	0.48257	0.483	0.48341	0.48382	0.48422	0.48461	0.485	0.48537	0.48574
2.2	0.4861	0.48645	0.48679	0.48713	0.48745	0.48778	0.48809	0.4884	0.4887	0.48899
2.3	0.48928	0.48956	0.48983	0.4901	0.49036	0.49061	0.49086	0.49111	0.49134	0.49158
2.4	0.4918	0.49202	0.49224	0.49245	0.49266	0.49286	0.49305	0.49324	0.49343	0.49361
2.5	0.49379	0.49396	0.49413	0.4943	0.49446	0.49461	0.49477	0.49492	0.49506	0.4952
2.6	0.49534	0.49547	0.4956	0.49573	0.49585	0.49598	0.49609	0.49621	0.49632	0.49643
2.7	0.49653	0.49664	0.49674	0.49683	0.49693	0.49702	0.49711	0.4972	0.49728	0.49736
2.8	0.49744	0.49752	0.4976	0.49767	0.49774	0.49781	0.49788	0.49795	0.49801	0.49807
2.9	0.49813	0.49819	0.49825	0.49831	0.49836	0.49841	0.49846	0.49851	0.49856	0.49861
3	0.49865	0.49869	0.49874	0.49878	0.49882	0.49886	0.49889	0.49893	0.49896	0.499

B. CDF Z-table for Standard Normal Distribution

Not that the CDF table can also be used to compute probabilities for the standard normal distribution $Z \sim N(0, 1)$. In the given figure, $F(z) = P(Z \leq z)$ is given by the shaded area.



This is same as $F(z) = P(Z \leq z) = 0.5 + P(0 \leq Z \leq z)$. The CDF table for standard normal is given page 9 for negative values, and page 10 for positive values:

Example, To calculate $P(Z \leq 2.5)$, we have two methods:

1) **Using Z-table for Standard Normal Distribution for $P(0 \leq Z \leq z)$:**

$$\begin{aligned} P(Z \leq 2.5) &= 0.5 + P(0 \leq Z \leq 2.5) \\ &= 0.5 + 0.49379 \quad (\text{value from the Z-table}) \\ &= 0.99379. \end{aligned}$$

2) **Using CDF Z-table for Standard Normal Distribution:**

$$P(Z \leq 2.5) = 0.99379 \quad (\text{direct from the CDF-table}).$$

STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z score.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.9	.00005	.00005	.00004	.00004	.00004	.00004	.00004	.00004	.00003	.00003
-3.8	.00007	.00007	.00007	.00006	.00006	.00006	.00006	.00005	.00005	.00005
-3.7	.00011	.00010	.00010	.00010	.00009	.00009	.00008	.00008	.00008	.00008
-3.6	.00016	.00015	.00015	.00014	.00014	.00013	.00013	.00012	.00012	.00011
-3.5	.00023	.00022	.00022	.00021	.00020	.00019	.00019	.00018	.00017	.00017
-3.4	.00034	.00032	.00031	.00030	.00029	.00028	.00027	.00026	.00025	.00024
-3.3	.00048	.00047	.00045	.00043	.00042	.00040	.00039	.00038	.00036	.00035
-3.2	.00069	.00066	.00064	.00062	.00060	.00058	.00056	.00054	.00052	.00050
-3.1	.00097	.00094	.00090	.00087	.00084	.00082	.00079	.00076	.00074	.00071
-3.0	.00135	.00131	.00126	.00122	.00118	.00114	.00111	.00107	.00104	.00100
-2.9	.00187	.00181	.00175	.00169	.00164	.00159	.00154	.00149	.00144	.00139
-2.8	.00256	.00248	.00240	.00233	.00226	.00219	.00212	.00205	.00199	.00193
-2.7	.00347	.00336	.00326	.00317	.00307	.00298	.00289	.00280	.00272	.00264
-2.6	.00466	.00453	.00440	.00427	.00415	.00402	.00391	.00379	.00368	.00357
-2.5	.00621	.00604	.00587	.00570	.00554	.00539	.00523	.00508	.00494	.00480
-2.4	.00820	.00798	.00776	.00755	.00734	.00714	.00695	.00676	.00657	.00639
-2.3	.01072	.01044	.01017	.00990	.00964	.00939	.00914	.00889	.00866	.00842
-2.2	.01390	.01355	.01321	.01287	.01255	.01222	.01191	.01160	.01130	.01101
-2.1	.01786	.01743	.01700	.01659	.01618	.01578	.01539	.01500	.01463	.01426
-2.0	.02275	.02222	.02169	.02118	.02068	.02018	.01970	.01923	.01876	.01831
-1.9	.02872	.02807	.02743	.02680	.02619	.02559	.02500	.02442	.02385	.02330
-1.8	.03593	.03515	.03438	.03362	.03288	.03216	.03144	.03074	.03005	.02938
-1.7	.04457	.04363	.04272	.04182	.04093	.04006	.03920	.03836	.03754	.03673
-1.6	.05480	.05370	.05262	.05155	.05050	.04947	.04846	.04746	.04648	.04551
-1.5	.06681	.06552	.06426	.06301	.06178	.06057	.05938	.05821	.05705	.05592
-1.4	.08076	.07927	.07780	.07636	.07493	.07353	.07215	.07078	.06944	.06811
-1.3	.09680	.09510	.09342	.09176	.09012	.08851	.08691	.08534	.08379	.08226
-1.2	.11507	.11314	.11123	.10935	.10749	.10565	.10383	.10204	.10027	.09853
-1.1	.13567	.13350	.13136	.12924	.12714	.12507	.12302	.12100	.11900	.11702
-1.0	.15866	.15625	.15386	.15151	.14917	.14686	.14457	.14231	.14007	.13786
-0.9	.18406	.18141	.17879	.17619	.17361	.17106	.16853	.16602	.16354	.16109
-0.8	.21186	.20897	.20611	.20327	.20045	.19766	.19489	.19215	.18943	.18673
-0.7	.24196	.23885	.23576	.23270	.22965	.22663	.22363	.22065	.21770	.21476
-0.6	.27425	.27093	.26763	.26435	.26109	.25785	.25463	.25143	.24825	.24510
-0.5	.30854	.30503	.30153	.29806	.29460	.29116	.28774	.28434	.28096	.27760
-0.4	.34458	.34090	.33724	.33360	.32997	.32636	.32276	.31918	.31561	.31207
-0.3	.38209	.37828	.37448	.37070	.36693	.36317	.35942	.35569	.35197	.34827
-0.2	.42074	.41683	.41294	.40905	.40517	.40129	.39743	.39358	.38974	.38591
-0.1	.46017	.45620	.45224	.44828	.44433	.44038	.43644	.43251	.42858	.42465
-0.0	.50000	.49601	.49202	.48803	.48405	.48006	.47608	.47210	.46812	.46414

List of Discrete Distributions

	Probability mass function $p(x)$	Properties
1	<p>Binomial Distribution ($X \sim B(n, p)$)</p> $p(i) = P(X = i) = \binom{n}{i} p^i (1-p)^{n-i}$ <p>where, $\binom{n}{i} = \frac{n!}{i!(n-i)!}$, and $i = 0, 1, \dots, n$.</p>	<p>Parameters $\rightarrow n, p$</p> $E(X) = np$ $\text{Var}(X) = np(1-p).$
2	<p>Poisson distribution ($\text{Pois}(\lambda)$)</p> $p(i) = P(X = i) = \frac{e^{-\lambda} \lambda^i}{i!} \quad i = 0, 1, \dots$	<p>Parameter $\rightarrow \lambda$</p> $E(X) = \lambda$ $\text{Var}(X) = \lambda$
3	<p>Geometric distribution ($G(p)$)</p> $p(i) = P(X = i) = (1-p)^{i-1} p, \quad i = 1, 2, \dots$	<p>Parameter $\rightarrow p$</p> $E(X) = \frac{1}{p}$ $\text{Var}(X) = \frac{1-p}{p^2}$
4	<p>Hypergeometric Distribution</p> $p(i) = P(X = i) = \frac{\binom{K}{i} \binom{N-K}{n-i}}{\binom{N}{n}}$ <p>where, $i = a$ to b with</p> $a = \max \{0, n - (N - K)\}$ $b = \min \{n, K\}$	<p>Parameters $\rightarrow N, K, n$</p> $E(X) = \frac{nK}{N}$ $\text{Var}(X) = \frac{nK}{N} \frac{(N-K)(N-n)}{N(N-1)}$

List of Continuous Distributions

	pdf $f(x)$ with domain	Properties
1	<p>Exponential Distribution ($\text{Exp}(\lambda)$)</p> $f(x) = \lambda e^{-\lambda x}, \quad x \geq 0$ <p>Parameter: $\lambda > 0$ (rate)</p>	<p>Notation $X \sim \text{Exp}(\lambda)$</p> $E(X) = \frac{1}{\lambda}$ $\text{Var}(X) = \frac{1}{\lambda^2}$ $F(x) = 1 - e^{-\lambda x}, \quad x \geq 0.$
2	<p>Standard Normal Distribution ($N(0, 1)$)</p> $f(z) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-z^2}{2}\right), \quad z \in \mathbb{R}.$ <p>($N(0, 1)$ is when $\mu = 0$, $\sigma = 1$ in $N(\mu, \sigma^2)$)</p>	<p>Notation $Z \sim N(0, 1)$</p> $E(Z) = 0$ $\text{Var}(Z) = 1$ $F(z) = \Phi(z) \quad (\text{CDF notation})$ <p>(Computed from Z-table)</p>
3	<p>Normal Distribution ($N(\mu, \sigma^2)$)</p> $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x - \mu)^2}{2\sigma^2}\right), \quad x \in \mathbb{R}.$ <p>$\mu \in \mathbb{R}$ (location), $\sigma^2 > 0$ (variance)</p>	<p>Notation $X \sim N(\mu, \sigma^2)$</p> $E(X) = \mu$ $\text{Var}(X) = \sigma^2$ $F(x) = \Phi\left(\frac{x - \mu}{\sigma}\right).$