Parametric continuous distributions

Li Li

Department of Mathematics and Statistics

Continuous uniform distribution

A continuous uniform random variable X has probability density function (PDF)

$$f(x) = \frac{1}{b-a}, \boxed{a \leq x \leq b,}$$

Sample space

and 0 elsewhere. Here a and b are two real numbers. We refer the distribution as uniform (a, b).

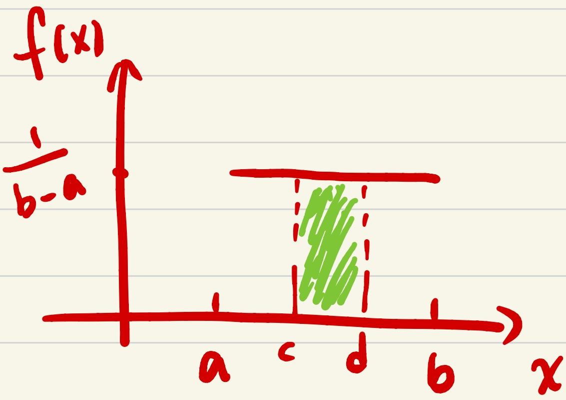
$a = 0$ $b = 1$

- For example, a continuous uniform distribution over $[0, 1]$ (often referred as uniform $(0, 1)$) has density $f(x) = 1$ for $0 < x < 1$ and 0 elsewhere.

  $f(x) = 0$ $x \notin (0,1)$

- Interesting fact: given any continuous random variable Y and its cumulative distribution function $F(y)$, a CDF transformation of Y, i.e $F(Y)$ has a uniform $(0, 1)$ distribution.
- For example, if Y has a density $f(y) = \lambda e^{-\lambda y}$ for $y \geq 0$, then its CDF is $F(y) = 1 - e^{-\lambda y}$ for $y \geq 0$. We have

$$1 - e^{\lambda Y} \sim \text{Uniform}(0, 1).$$



$$P(c < X < d) = \int_c^d \frac{1}{b-a}\,dx = \frac{d-c}{b-a}$$

$$P(a < X < x^*) = \frac{x^* - a}{b-a}$$

$X \sim \text{Uniform}(0, 1)$

$F(\cdot)$: CDF function

$Y: F^{-1}(X)$ $\quad$ Y has distribution $F(\cdot)$

Where is continuous uniform distribution used?

- Simulations are used to model complicated processes, estimate distributions of estimators (using methods such as bootstrap), and have dramatically increased the use of an entire field of statistics.
- In simulations, we often generate random numbers from a desired distribution.
- Uniform $(0, 1)$ is the where random number generation start.
- To generate a random variable that has CDF $F(y) = 1 - e^{-\lambda y}$ for $y \geq 0$, we can use the following steps
  (a) generate a random number u from Uniform $(0, 1)$.
  (b) transform the random number u by the inverse of the CDF for the density we desire, i.e. $-\log(1 - u)/\lambda$.

CDF

CDF of Uniform (a, b)

$$F(x) = \begin{cases} 0 & x \le a \\ \dfrac{x-a}{b-a} & a < x \le b \\ 1 & x > b \end{cases}$$

$$F(b) = \frac{b-a}{b-a} = 1$$

Mean and variance of uniform (a, b)

Mean and variance of uniform (a, b):

$$\mu = \frac{a+b}{2}$$

$$\sigma^2 = \frac{(b-a)^2}{12}$$

$$b^3 - a^3 = (b-a)(b^2 + ba + a^2)$$

$$\mu = \int_a^b x \cdot \underbrace{\frac{1}{b-a}}_{f(x)} \, dx$$

$$\left[\mu = \int_{-\infty}^{+\infty} x \cdot f(x) \, dx \right]$$

$$= \left. \frac{x^2}{2(b-a)} \right|_a^b$$

$$= \frac{b^2 - a^2}{2(b-a)} = \frac{b+a}{2}$$

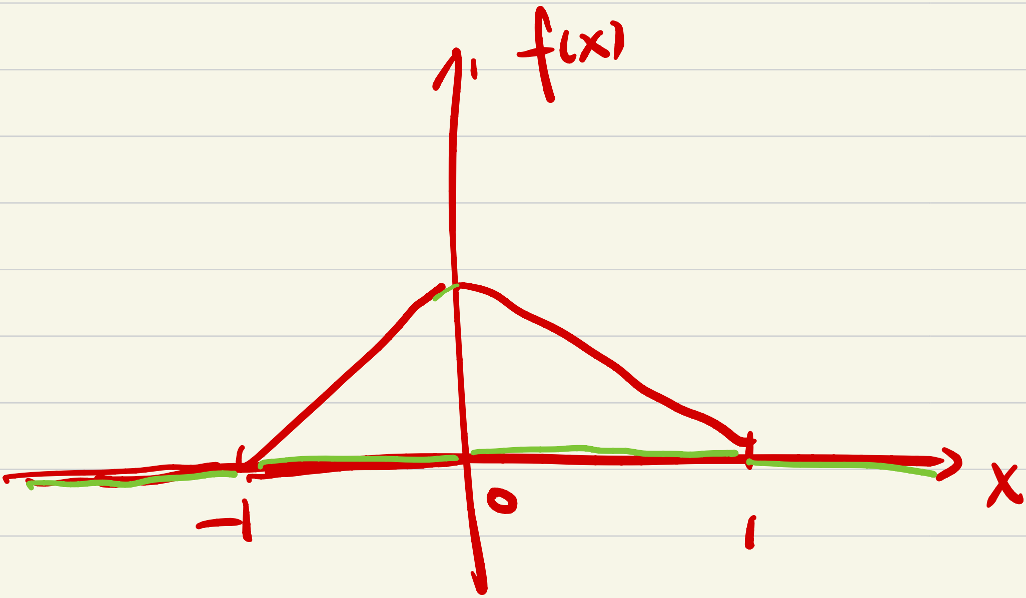
$$\sigma^2 = E(X^2) - \mu^2$$

$$= \int_a^b \frac{x^2}{b-a} \, dx - \frac{(b+a)^2}{4}$$

$$= \left. \frac{x^3}{3(b-a)} \right|_a^b - \frac{(b+a)^2}{4}$$

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

$$f(x) = \begin{cases} f_1(x) & -1 < x < 0 \\ f_2(x) & 0 \le x < 1 \end{cases}$$



$$E(X) = \int_{-\infty}^{+\infty} x\, f(x)\, dx = \int_{-1}^{0} x \cdot f_1(x)\, dx$$

$$+ \int_{0}^{1} x \cdot f_2(x)\, dx$$

Normal distribution

1) Charateristics

2) $P(a < X < b)$

3) Percentiles

4) Normal approximations to Binomial distribution and Poisson distribution

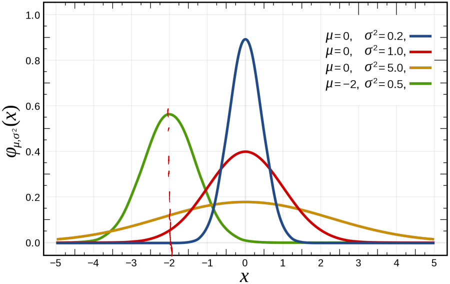
Normal distributions

Given parameters μ and σ, PDF of the Normal distribution is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < +\infty.$$

$E(X) = \mu$

$Var(X) = \sigma^2$



$\mu=0, \ \sigma=1$

$\mu=1, \ \sigma=1$

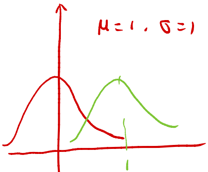


Figure: Normal distribution density functions. Density function is denoted as $\phi_{\mu,\sigma^2}(x)$.

μ– location parameter or mean of the distribution and σ–scale parameter or standard deviation of the distribution.

- A Normal distribution with $\mu = 0$ and $\sigma = 1$ is referred as "standard Normal distribution".
- Normal distributions are also called "Gaussian distributions".
- Normal distribution is sometimes informally called the "bell curve".
- A random variable with a Gaussian distribution is said to be normally distributed, denoted by $X \sim N(\mu, \sigma^2)$. For example, $X \sim N(3, 2^2)$.

Important facts about Normal distributions

- All Normal curves have the same overall shape: symmetric, single-peaked, bell-shaped.
- Any specific Normal curve is completely described by giving its mean μ and its standard deviation σ.
- The mean is located at the center of the symmetric curve and is the same as the median. Changing μ without changing σ moves the Normal curve along the horizontal axis without changing its spread.
- The standard deviation σ controls the spread of a Normal curve. Curves with larger standard deviations are more spread out or wider.

Important facts about Normal distributions

- The average of many independent processes (such as measurement errors) often have distributions that are nearly normal.
- If $X_1, X_2, \ldots, X_n$ are independent Bernoulli random variables with the same success rate p, $\bar{X}$ (average of $X_1, \ldots, X_n$) follows a Normal distribution $N(p, \frac{p(1-p)}{n})$ approximately.

p.m.f

x	0	1
$f(x)$	$1-p$	p

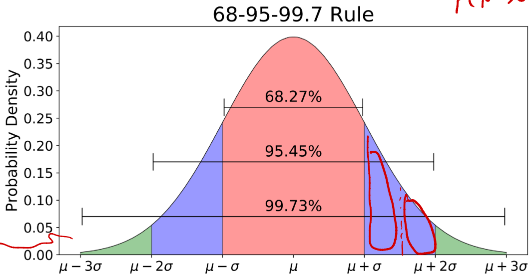
- If $X_1, X_2, \ldots, X_n$ are independent Poisson random variables with the same rate of event λ and time interval T, $\bar{X}$ follows a Normal distribution $N(\lambda T, \frac{\lambda T}{n})$ approximately.

For a Normal distribution with mean μ and standard deviation σ:

- Approximately 68% of the observations fall within σ of the mean μ.
- Approximately 95% of the observations fall within 2σ of the mean μ.
- Approximately 99.7% of the observations fall within 3σ of the mean μ.

The empirical rule

Figure: The 68–95–99.7 rule.



$$P(X > \mu + 3\sigma) = \frac{0.0027}{2} \approx 0.00135$$

$$P(X > \mu + \sigma) = (1 - 0.6827)/2 = 0.5 - \frac{0.6827}{2}$$

$$P(X > \mu + 1.5\sigma)$$

$$P(\mu - 3\sigma < X < \mu + 3\sigma) \approx 99.73\%$$

$$P(X > \mu + 3\sigma \text{ or } X < \mu - 3\sigma) = 1 - 99.73\% = 1 - 0.9973 \approx 0.0027$$

$$P(\mu - \sigma < X < \mu + \sigma) \approx 68.27\%$$

$$P(\mu - 2\sigma < X < \mu + 2\sigma) \approx 95.45\%$$

How do we find $P(a < X < b)$

For a Normal distribution with mean μ and variance σ^2 and $X \sim N(\mu, \sigma^2)$, probability

$$P(a < X < b) = \int_a^b \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

can be calculated by $pnorm((b-\mu)/\sigma) - pnorm((a-\mu)/\sigma)$ in R.

pnorm(x) : returns F(x) where F(·) is the CDF of Standard Normal distribution.

$$P(a < X < b) = F\left(\frac{b-\mu}{\sigma}\right) - F\left(\frac{a-\mu}{\sigma}\right)$$

How do we find the *p*-th percentile?

p-th percentile of a continuous distribution with density $f(x)$:

Solve $F(c) = \int_{-\infty}^{c} f(x)\,dx = \frac{p}{100}$ $\quad 0 \le p \le 100$

The *p*-th percentile *c* where

$$P(X < c) = \int_{\infty}^{c} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = p/100$$

can be calculated by $qnorm(p/100) * \sigma + \mu$.

CDF

The cumulative distribution function (CDF) of a Normal distribution function is denoted as

$\Phi_{1,2^2}(x)$

$$\Phi_{\mu,\sigma^2}(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(u-\mu)^2}{2\sigma^2}} du.$$

$\Phi_{2,1^2}(x)$

1) Monotone increasing within the sample space of the distribution

2) $0 \leq CDF(x) \leq 1$

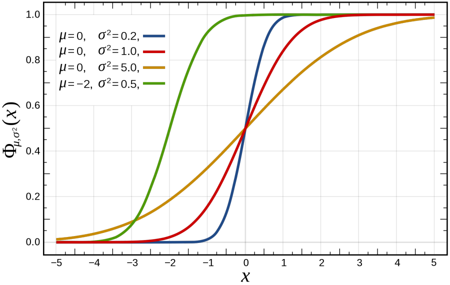


Figure: Normal distribution CDF functions. CDF functions are denoted as $\Phi_{\mu,\sigma^2}(x)$.

Standard Normal distribution and z-score

$P(a < X < b)$

Standard Normal distribution:

- If a variable X has any Normal distribution $N(\mu, \sigma^2)$, then the standardized variable

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

has the standard Normal distribution $N(0, 1)$. For example, if X follows a Normal distribution with mean 3 and variance 4, i.e. $X \sim N(3, 2^2)$, then $Z = \frac{X-\mu}{\sigma} \sim N(0, 1)$.

- For a real number a, the standardized value of a is

$$z = \frac{a - \mu}{\sigma}$$

is called the z-score of a.

$$F_Z(z) = P(Z \le z)$$

$$= P\left(\frac{X-\mu}{\sigma} \le z\right)$$

$$= P(X \le \sigma z + \mu)$$

$$= \int_{-\infty}^{\sigma z+\mu} \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx$$

show $\longrightarrow$ Let $y = \frac{x-\mu}{\sigma}$

$$= \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{y^2}{2}}\, dy$$

$$= \Phi_{0,1}(z)$$

Use of z-score

- When the z-score of an observation has an absolute value greater than 3, this observation can be viewed roughly as an outlier or unusual.
- z-score can be used to compare two observations from two populations that have different Normal distributions.

Example

Consider for two high school senior students,

$\sigma = 116$

- student A scored 670 on the Mathematics part of the SAT.4 The distribution of SAT Math scores in 2010 was Normal with mean 516 and standard deviation 116. $\mu = 516$
- student B took the ACT and scored 46 on the Mathematics portion. ACT Math scores for 2010 were Normally distributed with mean 21.0 and standard deviation of 5.3.

$\mu = 21, \quad \sigma = 5.3$

(a) Find the z- scores for both students.

(b) Assuming that both tests measure the same kind of ability, who had a higher score? Are any of these two test scores outlying?

a) $z_A = \dfrac{670 - 516}{116} = 1.3276$

$z_B = \dfrac{46 - 21}{5.3} = 4.7170$

Normal table

Goal: $X \sim N(\mu, \sigma^2)$; $P(a < X < b)$

from the Normal table.

- Only used in test situation these days.
- It is a one to one mapping of z to $\Phi_{0,1}(z)$ (Standard Normal CDF) for z goes from -3.99 to 3.99.
- Given z, use the table we can find $\Phi_{0,1}(z)$. Given p such that $\Phi_{0,1}(z) = p$, we can also find z. This gives us the $p * 100$-th percentile of the Standard Normal distribution.

Normal table

$\Phi_{0,1}(-3.90)$

-3.91 -3.92 $\Phi_{0,1}(-3.62)$ $\Phi_{0,1}(-3.615)$ $\Phi_{0,1}(z) = 0.00013$

$z = -3.65$

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

Figure: One to one mapping of z to $\Phi_{0,1}(z)$ for z from -3.99 to 0.

Normal 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
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
2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99492	.99506	.99520
2.6	.99534	.99547	.99560	.99573	.99585	.99598	.99609	.99621	.99632	.99643
2.7	.99653	.99664	.99674	.99683	.99693	.99702	.99711	.99720	.99728	.99736
2.8	.99744	.99752	.99760	.99767	.99774	.99781	.99788	.99795	.99801	.99807
2.9	.99813	.99819	.99825	.99831	.99836	.99841	.99846	.99851	.99856	.99861
3.0	.99865	.99869	.99874	.99878	.99882	.99886	.99889	.99893	.99896	.99900
3.1	.99903	.99906	.99910	.99913	.99916	.99918	.99921	.99924	.99926	.99929
3.2	.99931	.99934	.99936	.99938	.99940	.99942	.99944	.99946	.99948	.99950
3.3	.99952	.99953	.99955	.99957	.99958	.99960	.99961	.99962	.99964	.99965
3.4	.99966	.99968	.99969	.99970	.99971	.99972	.99973	.99974	.99975	.99976
3.5	.99977	.99978	.99978	.99979	.99980	.99981	.99981	.99982	.99983	.99983
3.6	.99984	.99985	.99985	.99986	.99986	.99987	.99987	.99988	.99988	.99989
3.7	.99989	.99990	.99990	.99990	.99991	.99991	.99992	.99992	.99992	.99992
3.8	.99993	.99993	.99993	.99994	.99994	.99994	.99994	.99995	.99995	.99995
3.9	.99995	.99995	.99996	.99996	.99996	.99996	.99996	.99996	.99997	.99997

$\Phi_{0,1}(0.5) = 0.69146$

$\Phi_{0,1}(-0.5) = 1 - 0.69146 = 0.30854$

0.929

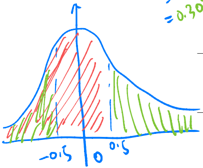


Figure: One to one mapping of z to $\Phi_{0,1}(z)$ for z from 0 to 3.99.

What is the probability for a standard Normal variable Z take values less than 1.47?

- Locate 1.4 in the left-hand column of the Normal table
- Then locate the remaining digit seven as .07 in the top row.
- The entry opposite 1.4 and under .07 is 0.9292. This is the cumulative proportion we seek.

What is the 92.9-th percentile of a Standard Normal distribution?—It is 1.47.

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

For a Normal random variable follows a distribution that has a mean μ and variance σ^2, how to find the probability for the random variable to fall within an interval $[a, b]$?

- Note that $P(a < X < b) = P(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma})$ and $\frac{X-\mu}{\sigma}$ follows a standard Normal distribution.
- Hence $P(a < X < b) = \Phi_{0,1}(\frac{b-\mu}{\sigma}) - \Phi_{0,1}(\frac{a-\mu}{\sigma})$.
- Calculate $\frac{b-\mu}{\sigma}$ and $\frac{a-\mu}{\sigma}$. Then use the Normal table to find the corresponding probabilities $\Phi_{0,1}(\frac{b-\mu}{\sigma})$ and $\Phi_{0,1}(\frac{a-\mu}{\sigma})$ respectively.

Suppose $X \sim N(\overset{\mu}{2}, 2^2)$, show that $Y \sim N(3, 1^2)$

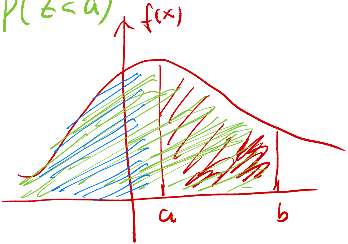
$P(1 < X < 3) = \Phi_{0,1}(0.5) - \Phi_{0,1}(-0.5) = 0.383.$

$$P(1 < X < 3) = P\left(\frac{1-2}{2} < \frac{X-2}{2} < \frac{3-2}{2}\right)$$

$$= P(-0.5 < Z < 0.5) \quad Z \sim N(0,1)$$

$$= \Phi_{0,1}(0.5) - \Phi_{0,1}(-0.5) =$$

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



Finding percentiles using the Normal table

For a Normal random variable follows a distribution that has a mean μ and standard deviation σ, how to find the the p-th percentile of the distribution?

(a) Note that our goal is to find c such that $P(X < c) = p/100$. Since $P(X < c) = \Phi_{0,1}(\frac{c-\mu}{\sigma})$, we are solving c from $\Phi_{0,1}(\frac{c-\mu}{\sigma}) = p/100$.

(b) Use the table to find the p-th percentile of the standard Normal distribution. Denoted it by z_p.

(c) Then $c = z_p * \sigma + \mu$.

Suppose $X \sim N(2, 2^2)$, show that the 92.9−th percentile of the distribution is $1.4\cancel{7} * 2 + 2 = 4.94$.

8

$\Phi_{0,1}(1.47) = 92.9$

Goal: find p-th percentile of distribution $N(\mu, \sigma^2)$

i.e find c such that

$$P(X \le c) = p/100 \quad \text{where } X \sim N(\mu, \sigma^2)$$

$$= \Phi_{0,1}\left(\frac{c-\mu}{\sigma}\right) = \frac{p}{100}$$

$$\frac{c-\mu}{\sigma} = z_p$$

$$c = z_p \cdot \sigma + \mu$$

Normal probability plot

How to tell whether observations from a population follows a Normal distribution? (Chapter 7)

- Normal probability plot or QQ plot.
- Shapiro-Wilk test.

Normal approximation to Binomial distributions

Normal approximation to the Binomial distribution: If X s a binomial random variable with parameters n and p,

$$Z = \frac{X - np}{\sqrt{np(1-p)}}$$

follows a standard Normal distribution approximately. The approximation is close if $np > 5$ and $n(1-p) > 5$.
`https://newonlinecourses.science.psu.edu/stat414/node/179/`

Normal approximation to Binomial distributions

- To approximate a binomial probability with a normal distribution, a continuity correction is applied as follows:

$$P(X \leq x) = P(X \leq x + 0.5) \approx P\left(Z \leq \frac{x + 0.5 - np}{\sqrt{np(1-p)}}\right)$$

and

$$P(X \geq x) = P(X \geq x - 0.5) \approx P\left(Z \geq \frac{x - 0.5 - np}{\sqrt{np(1-p)}}\right)$$

- For example, if $n = 20$ and $p = 0.3$, then
$P(X \leq 7) \approx \Phi_{0,1}(\frac{7+0.5-6}{\sqrt{4.2}}) = 0.758$,
$P(X \geq 7) \approx 1 - \Phi_{0,1}(\frac{7-0.5-6}{\sqrt{4.2}}) = 1 - 0.596 = 0.404$.
- Note that $P(X < x) = P(X \leq x - 1)$ and
$P(X > x) = P(X \geq x + 1)$.

Example

Assume that in a digital communication channel, the number of bits received in error can be modeled by a Binomial random variable, and assume that the probability that a bit is received in error is 1×10^{-5}. If 16 million bits are transmitted, what is the probability that 150 or fewer errors occur?

Normal approximation to Poisson distribution

Normal approximation to the Poisson distribution: If X is a Poisson random variable with $E(X) = \lambda T$ and $V(X) = \lambda T$,

$$P(X \leq x) = P(X \leq x + 0.5) = P\left(Z \leq \frac{x + 0.5 - \lambda T}{\sqrt{\lambda T}}\right)$$

and

$$P(X \geq x) = P(X \geq x - 0.5) = P\left(Z \geq \frac{x - 0.5 - \lambda T}{\sqrt{\lambda T}}\right)$$

The approximation is generally good for $\lambda T > 5$.

Chi-square distributions

If Z follows a standard Normal distribution then:

$$V = Z^2 \sim \chi_1^2,$$

where χ_1^2 is called a chi-square distribution with 1 degree of freedom which has density

$$f(v) = \frac{1}{\Gamma(1/2)2^{0.5}} v^{0.5-1} e^{-v/2}, v \geq 0.$$

Chi-square distributions

More generally, if $Z_1, Z_2, \ldots, Z_k$ are independent (one does not affect the distribution of another), standard Normal random variables, then

$$V = \sum_{i=1}^{k} Z_i^2 \sim \chi_k^2$$

which denotes a chi-squared distribution with k degrees of freedom. For example, $V = Z_1^2 + Z_2^2 \sim \chi_2^2$.The density of the distribution is

$$f(v) = \frac{1}{\Gamma(k/2)2^{k/2}} v^{k/2-1} e^{-v/2}, v \geq 0.$$

Chi-squared distributions are used primarily in hypothesis testing.

Distributions of random sample mean and random sample variance

Suppose $X_1, \ldots, X_n$ are independent Normal random variables, which all follow distribution $N(\mu, \sigma^2)$, which means they are identical.

- Denote $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ as the random sample mean.
- Denote $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ as the random sample variance.

Then the results are true regardless what values of μ and σ^2 are:

- $\bar{X} \sim N(\mu, \sigma^2/n)$
- $(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$

Example

Suppose $X_1, \ldots, X_{20}$ are independent Normal random variables, which all follow distribution $N(2, 3^2)$, which means they are identical.

- Denote $\bar{X} = \frac{1}{20}\sum_{i=1}^{20} X_i$ as the random sample mean.
- Denote $S^2 = \frac{1}{19}\sum_{i=1}^{20}(X_i - \bar{X})^2$ as the random sample variance.

- $\bar{X} \sim N(2, 9/20)$
- $19S^2/9 \sim \chi^2_{19}$

Simple random sample

If $X_1, \ldots, X_n$ is called a simple random sample if

- $X_1, X_2, \ldots, X_n$ are independent random variables.
- $X_1, X_2, \ldots, X_n$ follow the same distribution, i.e. they are identical.

Central limit theorem (CLT)

If $X_1, \ldots, X_n$ is a random sample of size n taken from a population or a distribution (not necessarily Normal distribution) with mean μ and variance σ^2 and if $\bar{X}$ is the sample mean, then

$$\bar{X} \sim N(\mu, \sigma^2/n)$$

for large n. For example,

- If $X_1, X_2, \ldots, X_{10}$ are independent random variables following an uniform distribution $(0, 1)$, then $\bar{X}$ follows a Normal distribution $N(0.5, 1/12/10)$, i.e. $N(0.5, 0.0083)$.
- If $X_1, X_2, \ldots, X_n$ are independent Bernoulli random variables with the same success rate 0.4, $\bar{X}$ (average of $X_1, \ldots, X_n$) follows a Normal distribution $N(0.4, \frac{0.24}{n})$ approximately.

Animation of CLT

https://www.youtube.com/watch?v=Pujol1yC1_A