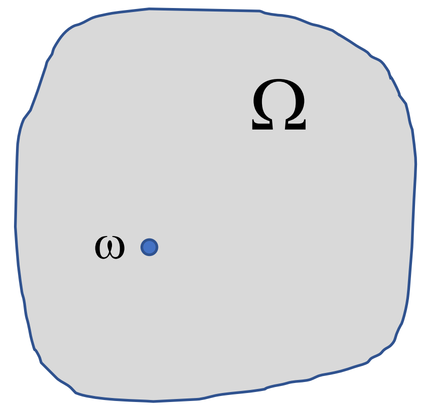
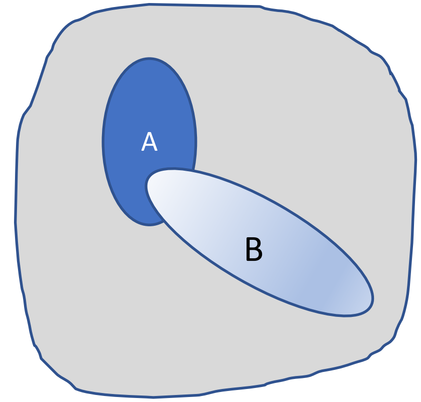
Basics of probability

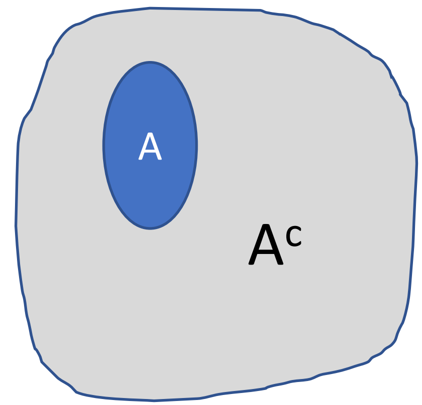
- **Sample space** Ω is the set of all possible **sample points** $\omega \in \Omega$
  - **Example 0**. Tossing a coin: $\Omega = \{\mathrm{H},\mathrm{T}\}$
  - **Example 1**. Casting a die: $\Omega = \{1,2,3,4,5,6\}$
  - **Example 2**. Number of customers in a queue: $\Omega = \{0,1,2,\ldots\}$
  - **Example 3**. Call holding time (e.g. in minutes): $\Omega = \{x \in \Re \mid x > 0\}$
- **Events** $A, B, C, \ldots \subset \Omega$ are subsets of the sample space Ω
  - **Example 1**. "Even numbers of a die": $A = \{2,4,6\}$
  - **Example 2**. "No customers in a queue": $A = \{0\}$
  - **Example 3**. "Call holding time greater than 3.0 (min)": $A = \{x \in \Re \mid x > 3.0\}$
- Denote by $\mathcal{F}$ the set of all events $A \in \mathcal{F}$
  - **Sure event**: The sample space $\Omega \in \mathcal{F}$ itself
  - **Impossible event**: The empty set $\emptyset \in \mathcal{F}$

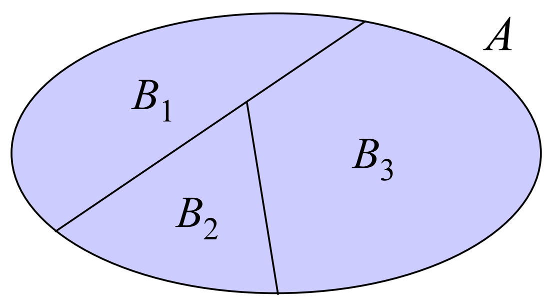


Combining events



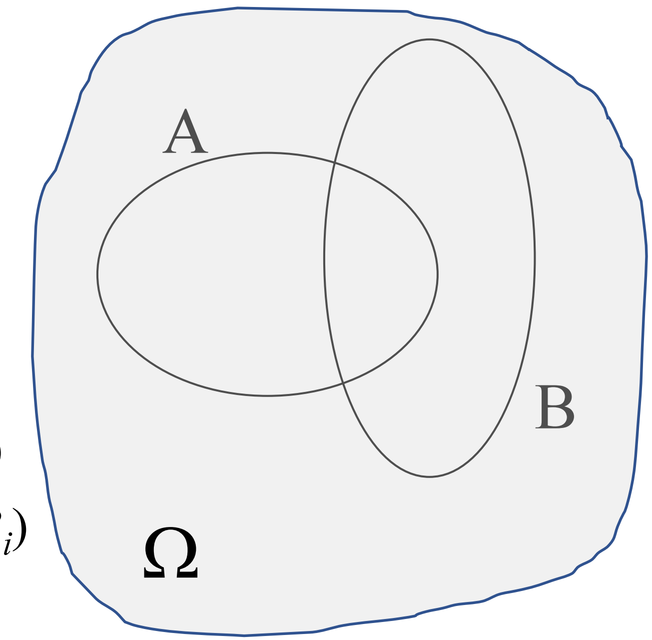
- **Union** "A or B": $A \cup B = \{\omega \in \Omega \mid \omega \in A \text{ or } \omega \in B\}$
- **Intersection** "A and B": $A \cap B = \{\omega \in \Omega \mid \omega \in A \text{ and } \omega \in B\}$
- **Complement** "not A": $A^c = \{\omega \in \Omega \mid \omega \notin A\}$
- Events A and B are **disjoint** if
  - $A \cap B = \emptyset$
- A set of events $\{B_1, B_2, \ldots\}$ is a **partition** of event A if
  - (i) $B_i \cap B_j = \emptyset$ for all $i \neq j$
  - (ii) $\cup_i B_i = A$





Probability rules

- **Probability** of event A is denoted by $P(A)$, $P(A) \in [0,1]$
  - Probability measure P is thus
    a real-valued set function defined on the set of events $\mathscr{F}$, P: $\mathscr{F} \to [0,1]$
- **Properties**:
  - (i) $\quad 0 \leq P(A) \leq 1$
  - (ii) $\quad P(\varnothing) = 0$
  - (iii) $\quad P(\Omega) = 1$
  - (iv) $\quad P(A^c) = 1 - P(A)$
  - (v) $\quad P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  - (vi) $\quad A \cap B = \varnothing \Rightarrow \quad P(A \cup B) = P(A) + P(B)$
  - (vii) $\{B_i\}$ is a partition of $A \Rightarrow \; P(A) = \Sigma_i \, P(B_i)$
  - $(\mathrm{v}iii)$ $\; A \subset B \Rightarrow P(A) \leq P(B)$



Examples

Tossing a coin: Ω={H,T}

- Fair coin: P(H)=P(T)=0.5

Casting a die: Ω={1,2,3,4,5,6}

- Fair die: P(1)=P(2)= P(3)=P(4)= P(5)=P(6)= 1/6

Number of individuals in a population: Ω={0,1,2,3,4,5,6,...}

- Countable infinity of sample points
- P(0)=probability of extinction, to be estimated
- $\sum_{i=0,\ldots\infty} P(i) = 1$

Conditional probability

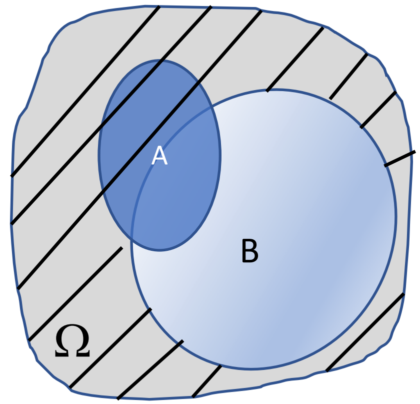
- Assume that $P(B) > 0$
- **Definition**: The **conditional probability** of event A **given** that event B occurred is defined as

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

- It follows that

$$P(A \cap B) = P(B)P(A \mid B) = P(A)P(B \mid A)$$

Example
- Draw an ace from a deck of cards
- Draw a black ace



Statistical independence

- **Definition**: Events A and B are **independent** if

$$P(A \cap B) = P(A)P(B)$$

- It follows that

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$$

- Correspondingly:

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A)P(B)}{P(A)} = P(B)$$

Borel’s law of large numbers

- If an experiment is repeated a large number of times, independently under identical conditions, then the proportion of times that any specified event occurs approximately equals the probability of the event's occurrence on any particular trial; the larger the number of repetitions, the better the approximation tends to be.
- More precisely, if E denotes the event in question, $p(E)$ its probability of occurrence, and $N_n(E)$ the number of times E occurs in the first n trials, then with probability one

  $$N_n(E) / n \to p(E) \text{ as } n \to \infty$$

- This theorem makes rigorous the intuitive notion of probability as the long-run relative frequency of an event's occurrence. It is a special case of any of several more general laws of large numbers in probability theory.

Random variables

Definition: Real-valued random variable X is a real-valued function that associates each sample point $\omega \in \Omega$ with a real number $X(\omega)$

Example

A coin is tossed three times

•Sample space:

$\Omega=\{(\omega 1,\omega 2,\omega 3)\,|\,\omega_i \in\{H,T\}, i=1,2,3\}$

• Let X be the random variable that tells the total number of tails in these three experiments:

Ω	HHH	HHT	HTH	THH	HTT	THT	TTH	TTT
$X(\omega)$	0	1	1	1	2	2	2	3

Random variables

In many cases each sample point $\omega \in \Omega$ is indeed a number. So the association is already there

Examples

- Number of individuals in a population: $\Omega=\{0,1,2,3,4,5,6,...\}$, all the nonnegative integers
- Body weight (g) of individuals in a population: Ω= all real positive numbers
- Fraction of populations becoming extinct in a landscape: Ω = all real numbers between 0 and 1

Discrete random variables

The random variable X can assume only a finite or countably infinite set A of values.

- finite, $A = \{x_1,\dots, x_n\}$, or
- countably infinite, $A = \{x_1, x_2,\dots\}$

The distribution P of X is determined by the point probabilities p_i, $p_i := P\{X=x_i\}$, $x_i \in A$

Example

- Probability that the number N of individuals in a population is 6: $p_6 = P\{N=6\}$
- Probability that the number b of alleles type B among all the genes of a population are 36: $p_{36} = P\{b=36\}$

Expected value

- Definition: The expectation (mean value) of X is defined by

  $$E[X]= \sum_i p_i x_i$$

- Properties:
- (i) c is a constant $\Rightarrow E[cX]=cE[X]$
- (ii) $E[X+Y]=E[X]+E[Y]$
- (iii) X and Y independent , that is

  $$P\{X = x_i, Y = y_j\}= P\{X = x_i\}P\{Y = y_j\}$$

  $$\Rightarrow E[XY] = E[X]E[Y]$$

Variance

- Definition: The variance of X is the expected value of the square deviation from its mean $E[X]$, namely it is defined by

$$\sigma^2 = \mathrm{Var}[X]=E[(X-E[X])^2] = \sum_i p_i \, (x_i - E[X])^2$$

- Useful formula (prove!):

$$\mathrm{Var}[X]= E[X^2]-E[X]^2$$

- Properties:

– (i) c is a constant $\Rightarrow \mathrm{Var}[cX]=c^2\mathrm{Var}[X]$

– (ii) X and Y independent $\Rightarrow \mathrm{Var}[X+Y]= \mathrm{Var}[X] + \mathrm{Var}[Y]$

The standard deviation σ is the square root of the variance

Example

A coin is tossed three times

•Sample space:

$\Omega=\{(\omega1,\omega2,\omega3)|\omega_i \in \{H,T\},i=1,2,3\}$

• Let *X* be the random variable that tells the total number of tails in these three experiments:

Ω	HHH	HHT	HTH	THH	HTT	THT	TTH	TTT
X(ω)	0	1	1	1	2	2	2	3

$p_0 = 1/8$ $p_1 = 3/8$ $p_2 = 3/8$ $p_3 = 1/8$

$E[X]= \sum_i p_i x_i = 0\times1/8+1\times3/8+2\times3/8+3\times1/8= (0+3+6+3)/8 = 1.5$

$\sigma^2 = Var[X]=E[(X-E[X])^2] = \sum_i p_i (x_i - E[X])^2 = (0-1.5)^2\times1/8$ etc.

Bernoulli distribution

$X \sim$Bernoulli(p), p $\in$ (0,1)

– describes a simple random experiment with two possible outcomes: success or failure, e.g. coin tossing; X is the number of successes (e.g. allele A), so either $X = 0$ or $X = 1$

– success with probability p (and failure with probability $1 - p$)

$$P\{X=0\}=1-p,\ P\{X=1\}=p$$

- Mean value: $E[X]=(1-p)\cdot 0+p\cdot 1=p$
- Second moment: $E[X^2] = (1 - p)\cdot 0^2 + p\cdot 1^2 = p$
- Variance: $Var[X]= E[X^2]-E[X]^2 = p-p^2=p(1-p)$

$Var[X]= E[(X\text{-}E[X])^2]=(0\text{-}p)^2 \cdot(1\text{-}p)+(1\text{-}p)^2 \cdot p=p^2\text{-}p^3+p\text{-}2p^2+p^3$

Binomial distribution

$X \sim \text{Bin}(n,p)$, $n \in \{1,2,\ldots\}$, $p \in (0,1)$

number of successes in an independent series of simple random experiments (of Bernoulli type); $X = X_1 + \ldots + X_n$ (with $X_i \sim \text{Bernoulli}(p)$)

- n = total number of experiments
- p = probability of success in any single experiment
- Point probabilities

$$P\{X = i\} = \binom{n}{i} p^i (1-p)^{n-i}$$

$$\binom{n}{i} = \frac{n!}{i!(n-i)!}$$

$$n! = n \cdot (n-1) \cdots 2 \cdot 1$$

- Mean value: $E[X]=E[X_1]+\ldots+E[X_n]=np$
- Variance: $\text{Var}[X] = \text{Var}[X_1] + \ldots + \text{Var}[X_n] = np(1-p)$ (independence!)

Covariance, correlation, coefficient of variation

- **Definition**: The **covariance** between X and Y is defined by

$$\sigma_{XY}^2 := \mathrm{Cov}[X,Y] := E[(X - E[X])(Y - E[Y])]$$

- Useful formula (prove!):

$$\mathrm{Cov}[X,Y] = E[XY] - E[X]E[Y]$$

- Correlation

  $\rho_{XY} = \mathrm{Cov}[X,Y]/(\sigma_X \sigma_Y)$ $\quad$ $-1 \leq \rho_{XY} \leq 1$

Variables are uncorrelated if $\mathrm{Cov}[X,Y] = 0$ or $\rho_{XY} = 0$

Two independent variable are always uncorrelated, in particular

  $E[XY]=E[X]\,E[Y]$

- Coefficient of variation

  $CV = \sigma_X/E[X]$

Continuous random variables

The random variable *X* can assume any real value ($-\infty < X < \infty$) or any value in an interval of the real axis (e.g. $A \le X \le B$, $X \ge 0$).

- We define the **cumulative distribution function**

$$F_X(x) := P\{X \le x\}$$

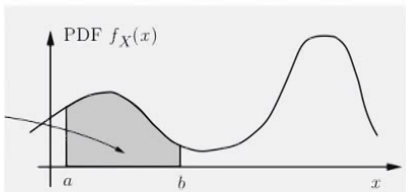
- and the **probability density function**

$$F_X(x) := P\{X \le x\} = \int_{-\infty}^{x} f_X(y)\, dy$$

$$P\left\{A \le X \le B\right\} = \int_{A}^{B} f_x(y) dy$$

$$\int_{-\infty}^{+\infty} f_x(y) dy = 1$$

P$(a \le X \le b)$



$$\mathbf{P}(a \le X \le b) = \int_a^b f_X(x)\, dx$$

Expected value and variance

$$E\left[X\right]=\mu_x=\int_{-\infty}^{+\infty} f_x(y)y\,dy$$

$$Var\left[X\right]=E\left[(X-\mu_x)^2\right]=\int_{-\infty}^{+\infty} f_x(y)(y-\mu_x)^2\,dy$$

Same properties as in the discrete case

$$SD=\sigma_x=\sqrt{Var\left[X\right]}$$ Standard deviation

$$CV={\sigma_x}/{\mu_x}$$ Coefficient of variation

Standard normal (Gaussian) distribution

$$X \sim \mathrm{N}(0,1)$$

- Probability density function (pdf):

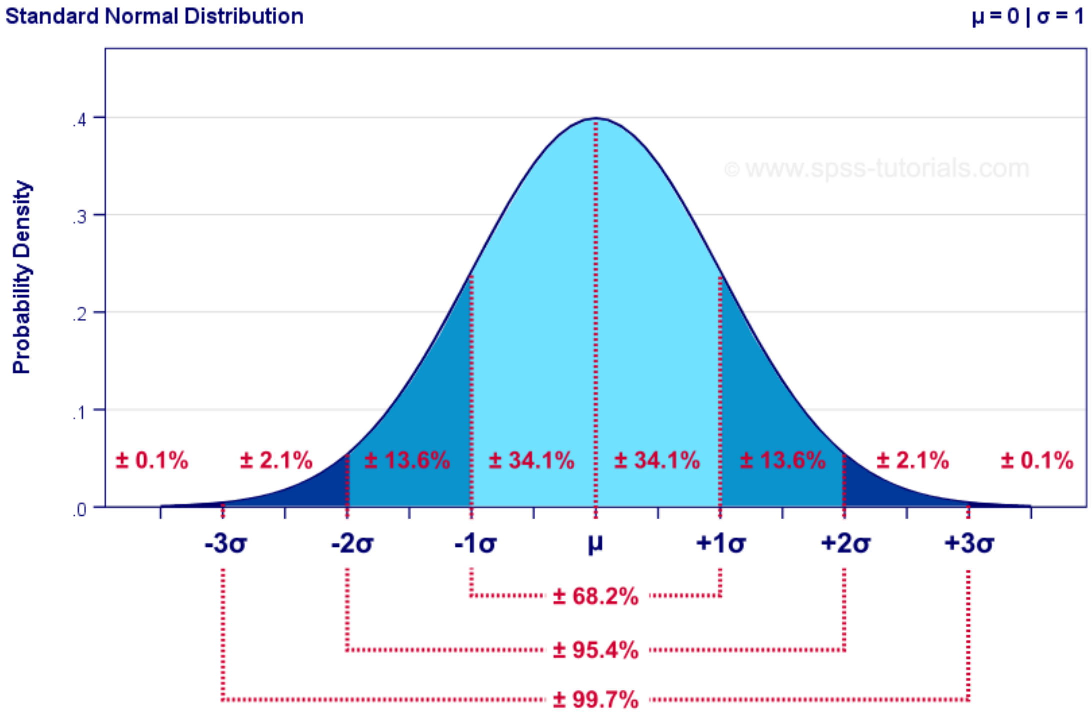
$$f_X(x) = \varphi(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

- Cumulative distribution function (cdf):

$$F_X(x) := P\{X \le x\} = \Phi(x) := \int_{-\infty}^{x} \varphi(y)\, dy$$

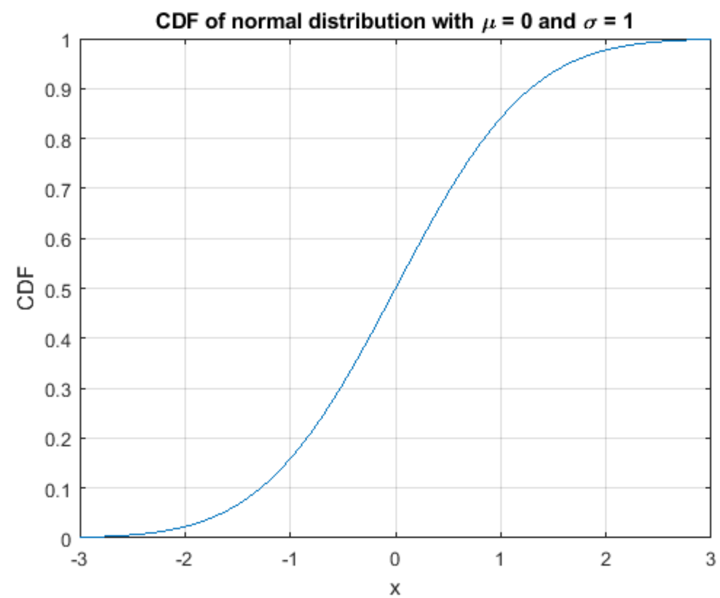
E[X] = 0

Var [X] = 1



pdf probability density

cdf cumulative probability



Normal (Gaussian) general distribution

$$X \sim \mathrm{N}(\mu, \sigma^2), \quad \mu \in \Re, \quad \sigma > 0$$

μ is the mean value, σ^2 the variance, σ the standard deviation

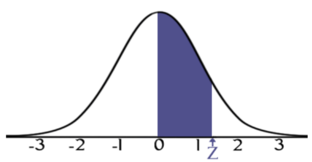
Probability density function (pdf):

$$\varphi(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

$$f_X(x) = F_X{}'(x) = \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)$$

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

Tables of (standard) normal distribution

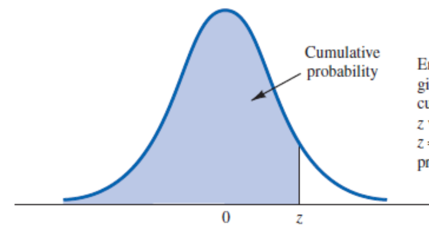


STANDARD NORMAL TABLE (Z)

Entries in the table give the area under the curve between the mean and z standard deviations above the mean. For example, for $z = 1.25$ the area under the curve between the mean (0) and z is 0.3944.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0190	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2969	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3513	0.3554	0.3577	0.3529	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998

TABLE 1 CUMULATIVE PROBABILITIES FOR THE STANDARD NORMAL DISTRIBUTION (*Continued*)



Entries in the table give the area under the curve to the left of the z value. For example, for $z = 1.25$, the cumulative probability is .8944.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.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

Meaning of the normal distribution

- Let $X_1, \ldots, X_n$ be independent and identically distributed (**IID**) with mean μ and variance σ^2
- Denote the average (sample mean) as follows:

$$\overline{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i$$

- Then (prove!)

$$E[\overline{X}_n] = \mu$$

In the limit (large *n*) the sample mean is distributed like a Normal

$$\overline{X}_n \approx \mathrm{N}(\mu, \tfrac{1}{n}\sigma^2)$$

Sample mean and sample variance

Theoretically the variance is

σ^2 = Var[X]=E[(X−E[X])^2]= E[(X−μ)^2]

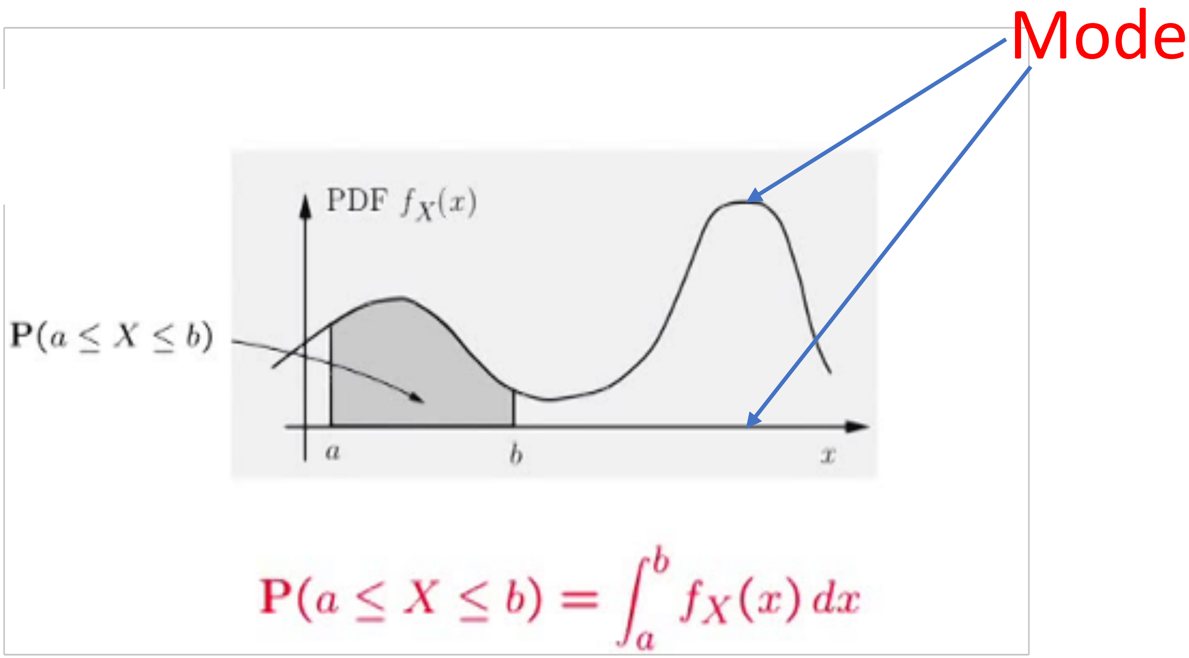
However, one does not know μ, but only its sample mean, an estimate of μ

$$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$$

The sample variance is calculated using the sample mean and subtracting one degree of freedom

$$s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}_n\right)^2$$

Median and mode



P(X≤Median) = 0.5 = value of x where cdf is 0.5

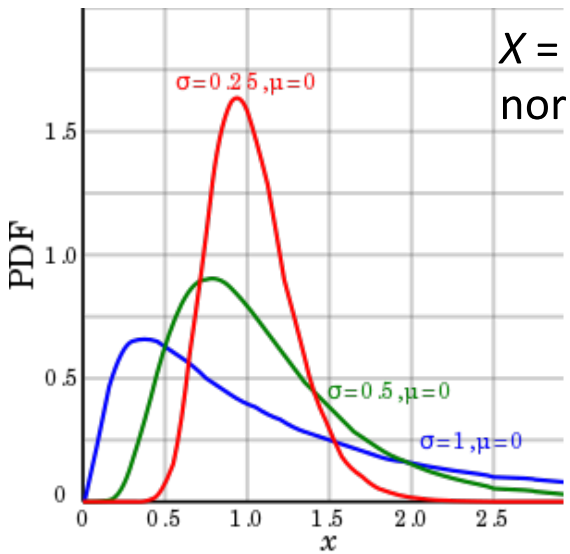
Mode = most probable value = value at which pdf is maximum i.e. $f_X(x_{mode})$ = max

For Normal distribution mean, median and mode coincide, but in general they do not

Lognormal distribution

A random variable whose logarithm is normally distributed

- In many cases a random variable can be seen as the product of *n* independent variables (e.g. total survival is the product of survival from egg to larva times survival from larva to stage 1 nymph times survival from stage 1 to stage 2 etc. Thus the logarithm of product is the sum of single logarithms. Apply the limit of the sum of independent variables.



$X = \exp(\mu + \sigma Z)$ where Z is the standard normal distribution

$$E\left[X\right] = \exp\left(\mu + \frac{\sigma^2}{2}\right)$$

$$Median\left[X\right] = \exp\left(\mu\right)$$

$$Mode\left[X\right] = \exp\left(\mu - \sigma^2\right)$$