Probability Overview (1A)

Young Won Lim
3/5/18

Please send corrections (or suggestions) to youngwlim@hotmail.com.

This document was produced by using LibreOffice and Octave.

Conditional Probability

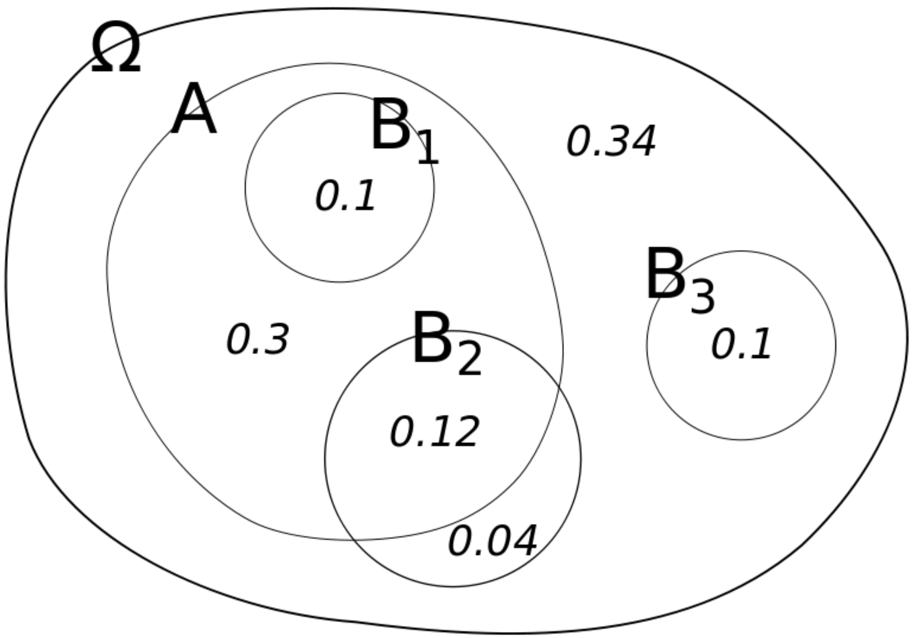
$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Given two events A and B with P(B) > 0,
the conditional probability of A given B
is defined as the quotient of the probability of the
joint of events A and B, and the probability of B:

https://en.wikipedia.org/wiki/Conditional_probability

Conditional Probability Examples (1)

The unconditional probability
P(A) = 0.52.
the conditional probability
P(A|B1) = 1,
P(A|B2) = 0.75, and
P(A|B3) = 0.



$$P(A \cap B1) = 0.1 \qquad P(B1) = 0.1$$
$$P(A \cap B2) = 0.12 \qquad P(B2) = 0.16$$
$$P(A \cap B3) = 0 \qquad P(B3) = 0.1$$

https://en.wikipedia.org/wiki/Conditional_probability

Conditional Probability Examples (2)

sample space

E : at least two consecutive zero's

F : starting with a zero

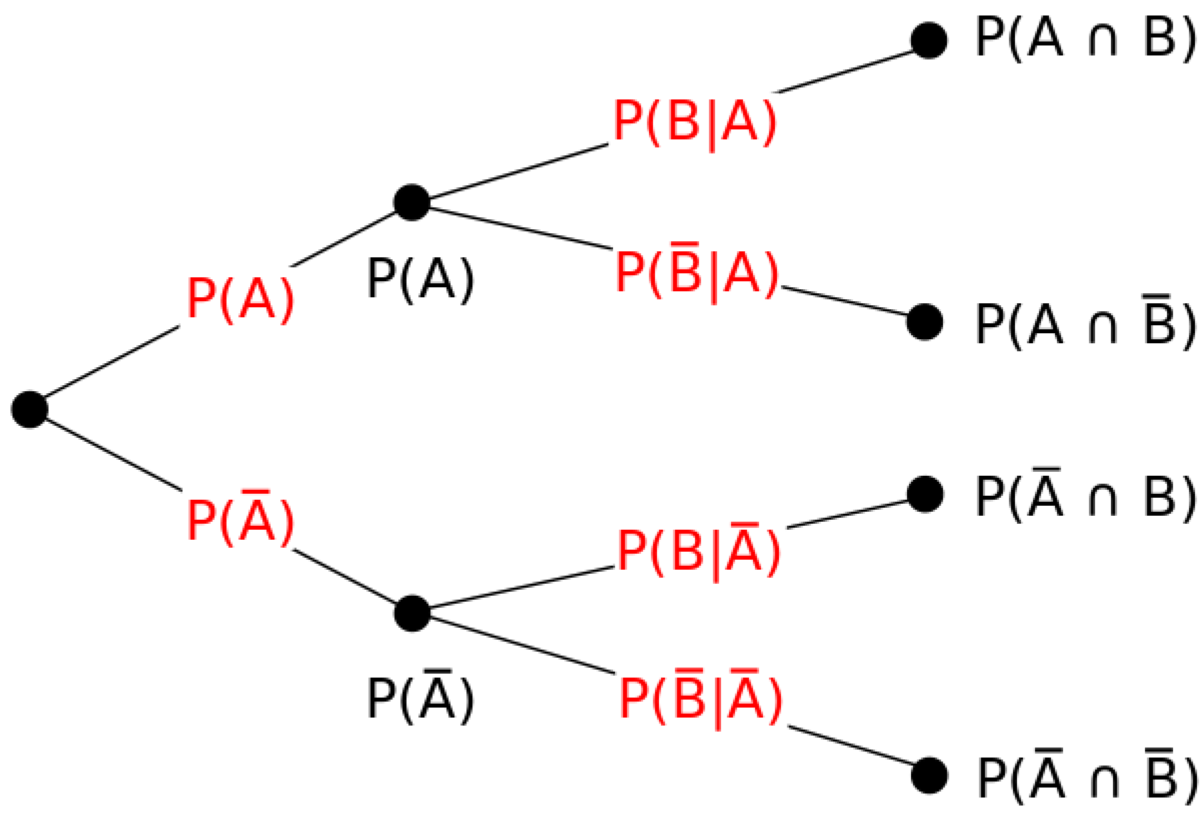
S	E	F	$E \cap F$
0, 0, 0, 0	0, 0, 0, 0	0, 0, 0, 0	0, 0, 0, 0
0, 0, 0, 1	0, 0, 0, 1	0, 0, 0, 1	0, 0, 0, 1
0, 0, 1, 0	0, 0, 1, 0	0, 0, 1, 0	0, 0, 1, 0
0, 0, 1, 1	0, 0, 1, 1	0, 0, 1, 1	0, 0, 1, 1
0, 1, 0, 0	0, 1, 0, 0	0, 1, 0, 0	0, 1, 0, 0
0, 1, 0, 1	~~0, 1, 0, 1~~	0, 1, 0, 1	~~0, 1, 0, 1~~
0, 1, 1, 0	~~0, 1, 1, 0~~	0, 1, 1, 0	~~0, 1, 1, 0~~
0, 1, 1, 1	~~0, 1, 1, 1~~	0, 1, 1, 1	~~0, 1, 1, 1~~
1, 0, 0, 0	1, 0, 0, 0		
1, 0, 0, 1	1, 0, 0, 1		
1, 0, 1, 0	~~1, 0, 1, 0~~		
1, 0, 1, 1	~~1, 0, 1, 1~~		
1, 1, 0, 0	1, 1, 0, 0		
1, 1, 0, 1	~~1, 1, 0, 1~~		
1, 1, 1, 0	~~1, 1, 1, 0~~		
1, 1, 1, 1	~~1, 1, 1, 1~~		

$$P(E) = \frac{8}{16} \qquad P(F) = \frac{8}{16} \qquad P(E \cap F) = \frac{5}{16}$$

$$P(E|F) = \frac{5}{8} \qquad \frac{P(E \cap F)}{P(F)} = \frac{5/16}{8/16}$$

Young Won Lim
3/5/18

Intersection Probability

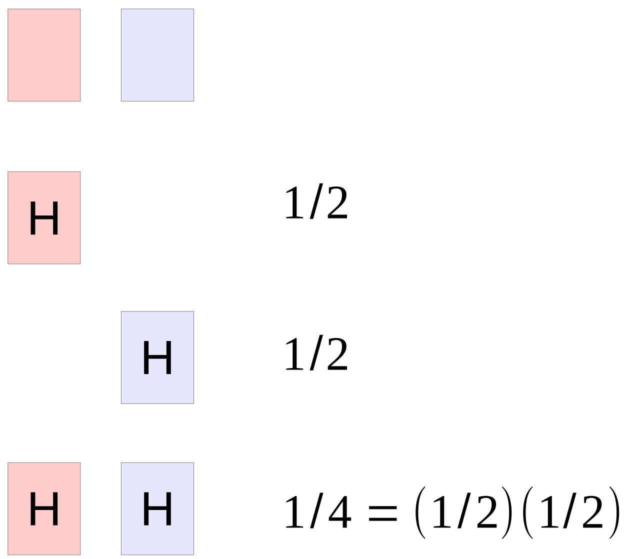


https://en.wikipedia.org/wiki/Conditional_probability

Independence

two events are (statistically) independent
if the occurrence of one does not affect the
probability of occurrence of the other.

Similarly, two random variables are
independent if the realization of one does
not affect the probability distribution of the
other.



$$P(A|B) = P(A)$$
$$P(B|A) = P(B)$$

https://en.wikipedia.org/wiki/Independence_(probability_theory)

Independence

$$P(A|B) = P(A)$$
$$P(B|A) = P(B)$$

⟺

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

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

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

https://en.wikipedia.org/wiki/Independence_(probability_theory)

Coin Tossing Experiment (1)

H	H	H	H	H	H	3 H's	⟷	0 T's			
H	H	T	H	H		2 H's	⟷	1 T's			T
H	T	H	H		H	2 H's	⟷	1 T's		T	
H	T	T	H			1 H's	⟷	2 T's		T	T
T	H	H		H	H	2 H's	⟷	1 T's	T		
T	H	T		H		1 H's	⟷	2 T's	T		T
T	T	H			H	1 H's	⟷	2 T's	T	T	
T	T	T				0 H's	⟷	3 T's	T	T	T

Counting heads only

Counting tails only

Coin Tossing Experiment (2)

H	H	H	H	H	H	3 H's	0 T's
H	H	T	H	H		2 H's	1 T's
H	T	H	H	H		2 H's	1 T's
H	T	T	H			1 H's	2 T's
T	H	H	H	H		2 H's	1 T's
T	H	T	H			1 H's	2 T's
T	T	H	H			1 H's	2 T's
T	T	T				0 H's	3 T's

Counting heads only → determines tails

Coin Tossing Experiment (3)

H	H	H
H	H	T
H	T	H
H	T	T
T	H	H
T	H	T
T	T	H
T	T	T

H	H	H	3 H's	0 T's	$\times 1$	$C(3,3)$
H	H		2 H's	1 T's	$\times 3$	$C(3,2)$
H			1 H's	2 T's	$\times 3$	$C(3,1)$
			0 H's	3 T's	$\times 1$	$C(3,0)$

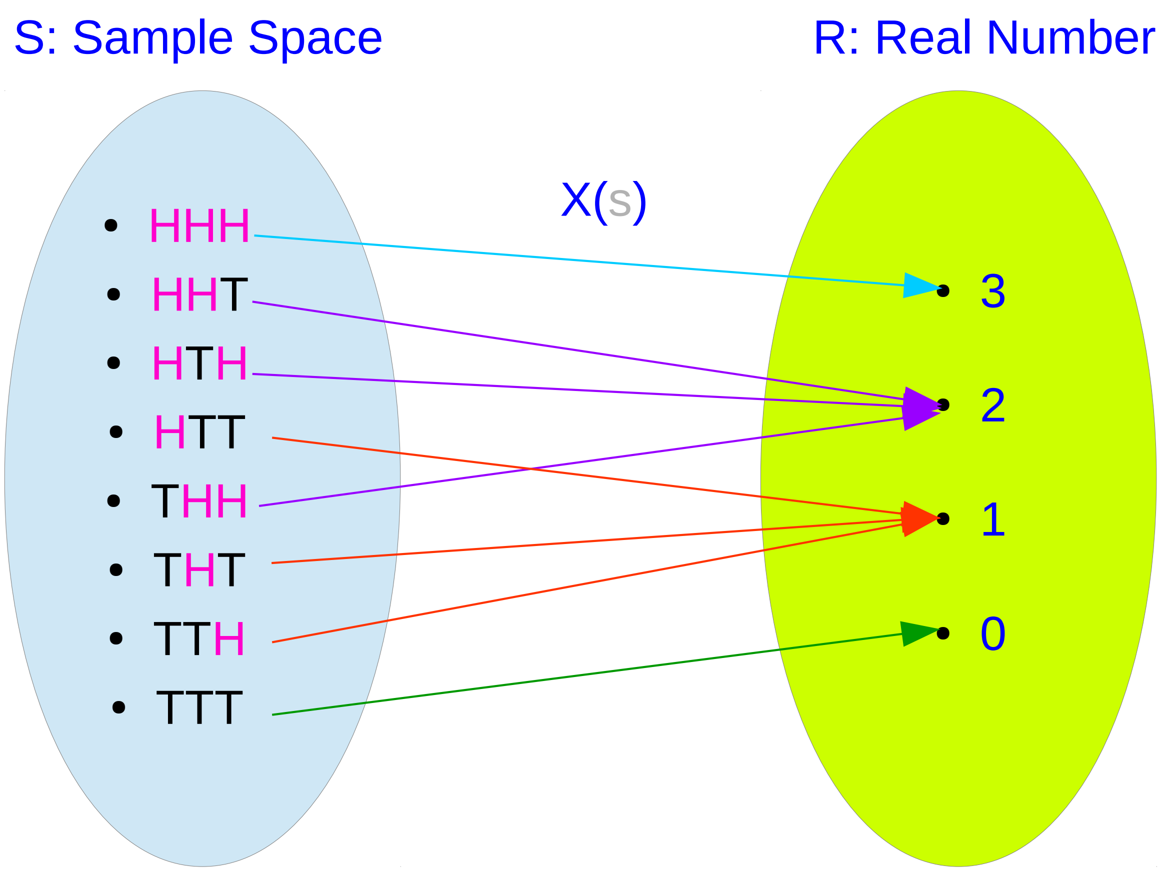
Consider a combination
Not a permutation

Coin Tossing Experiment (4)

H	H	H	p^3
H	H	T	$p^2 q$
H	T	H	$p^2 q$
H	T	T	$p q^2$
T	H	H	$p^2 q$
T	H	T	$p q^2$
T	T	H	$p q^2$
T	T	T	q^3

H (p)	H (p)	H (p)	3 H's	0 T's	$\times 1$	$p^3 \cdot C(3,3)$
H (p)	H (p)	(q)	2 H's	1 T's	$\times 3$	$p^2 q \cdot C(3,2)$
H (p)	(q)	(q)	1 H's	2 T's	$\times 3$	$p q^2 \cdot C(3,1)$
(q)	(q)	(q)	0 H's	3 T's	$\times 1$	$q^3 \cdot C(3,1)$

Random Variable is a function



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

Random Variable is related to events

A random variable does not return a probability.

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

X = 3	⇔	Event { HHH }
X = 2	⇔	Event { HHT, HTH, THH }
X = 1	⇔	Event { HTT, THT, TTH }
X = 0	⇔	Event { TTT }

looks like variables

Distribution

A random variable does not return a probability.

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

p(X = 3) ⟺ p(Event { HHH }) $P_3 = p^3 \cdot C(3,3)$

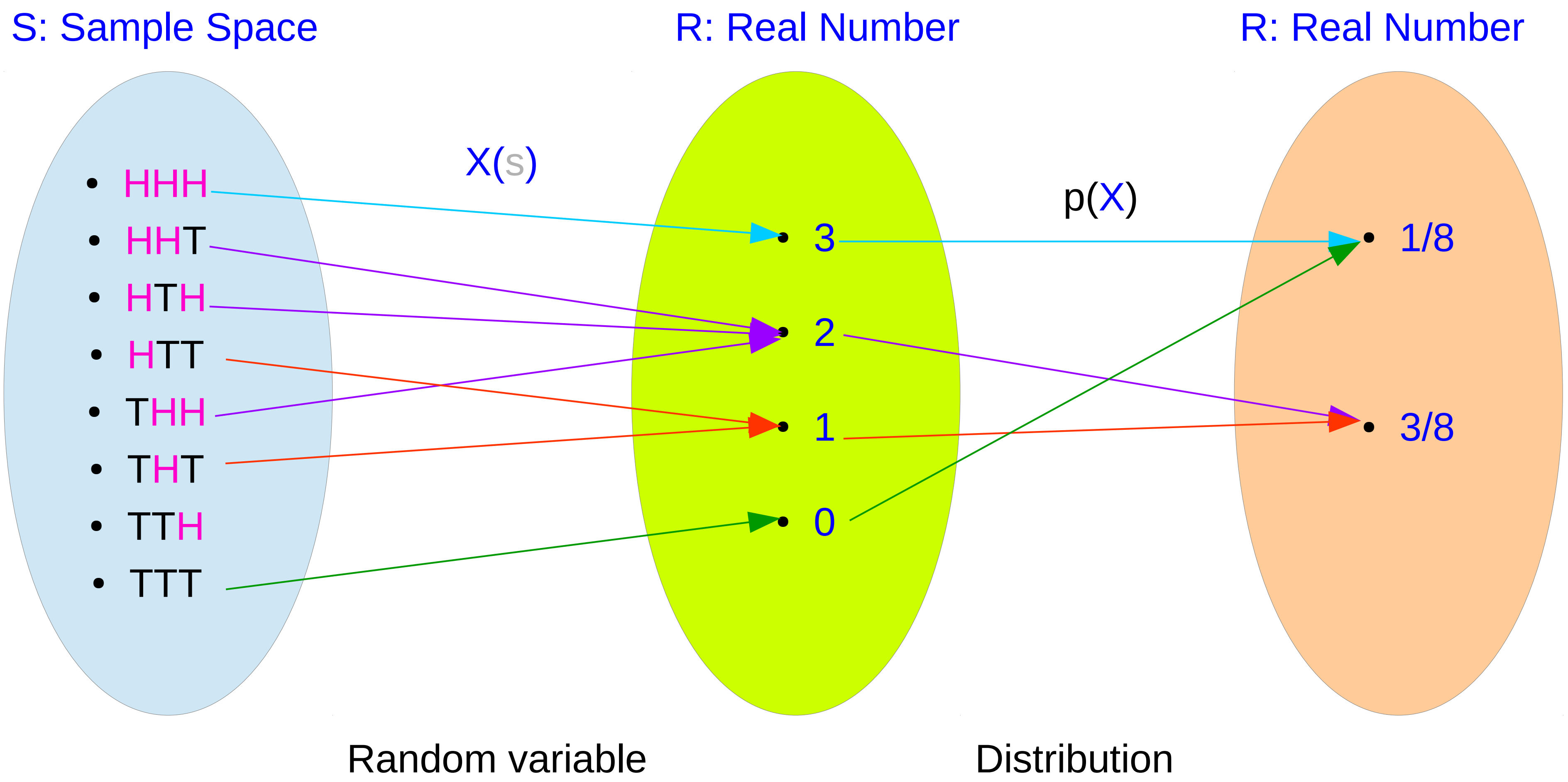
p(X = 2) ⟺ p(Event { HHT, HTH, THH }) $P_2 = p^2 q \cdot C(3,2)$

p(X = 1) ⟺ p(Event { HTT, THT, TTH }) $P_1 = p q^2 \cdot C(3,1)$

p(X = 0) ⟺ p(Event { TTT }) $P_0 = q^3 \cdot C(3,1)$

Distribution $\{ (0, P_0), (1, P_1), (2, P_2), (3, P_3) \}$

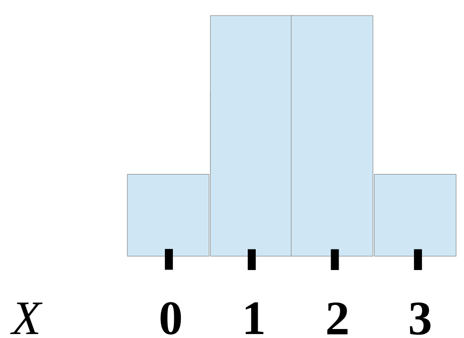
A random variable and its distribution



Different Random Variable Assignments

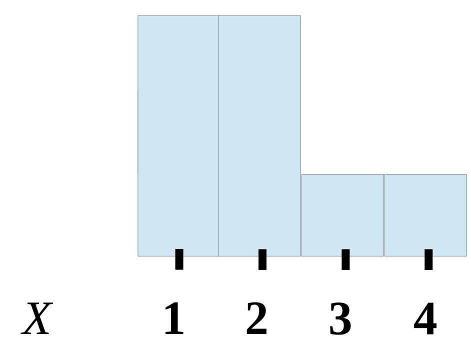
H	H	H	3 H's	$X = 3$
H	H		2 H's	$X = 2$
H			1 H's	$X = 1$
			0 H's	$X = 0$

$$p(X) \quad \frac{1}{8} \quad \frac{3}{8} \quad \frac{3}{8} \quad \frac{1}{8}$$



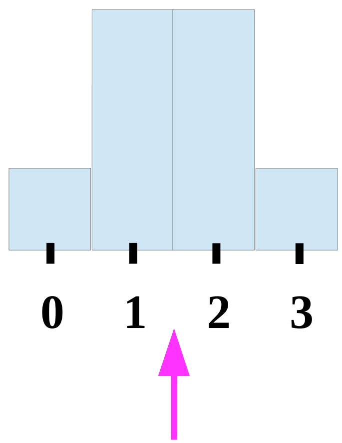
H	H	H	3 H's	$X = 4$
H	H		2 H's	$X = 1$
H			1 H's	$X = 2$
			0 H's	$X = 3$

$$p(X) \quad \frac{3}{8} \quad \frac{3}{8} \quad \frac{1}{8} \quad \frac{1}{8}$$



Different Expectation Values

$$p(X) \quad \frac{1}{8} \quad \frac{3}{8} \quad \frac{3}{8} \quad \frac{1}{8}$$

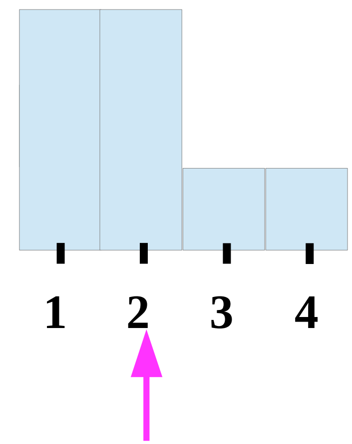


$$X \quad \mathbf{0} \quad \mathbf{1} \quad \mathbf{2} \quad \mathbf{3}$$

$$E(X) = \mathbf{0}\cdot\frac{1}{8} + \mathbf{1}\cdot\frac{3}{8} + \mathbf{2}\cdot\frac{3}{8} + \mathbf{3}\cdot\frac{1}{8}$$

$$= 1.5$$

$$p(X) \quad \frac{3}{8} \quad \frac{3}{8} \quad \frac{1}{8} \quad \frac{1}{8}$$



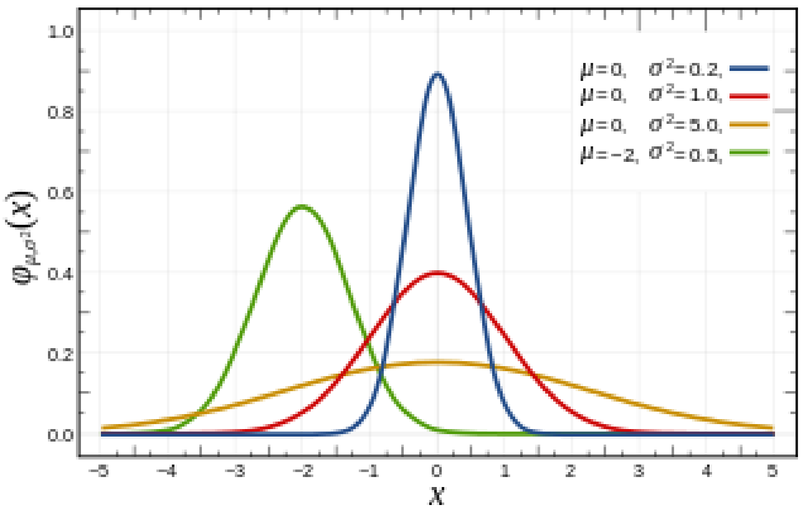
$$X \quad \mathbf{1} \quad \mathbf{2} \quad \mathbf{3} \quad \mathbf{4}$$

$$E(X) = \mathbf{1}\cdot\frac{3}{8} + \mathbf{2}\cdot\frac{3}{8} + \mathbf{3}\cdot\frac{1}{8} + \mathbf{4}\cdot\frac{1}{8}$$

$$= 2$$

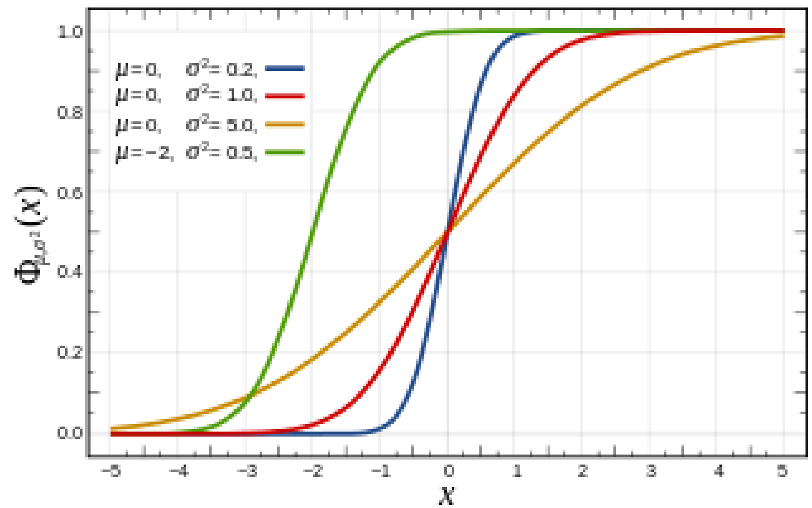
Normal Distribution

Probability mass function



Mean	μ
Median	μ
Mode	μ
Variance	σ^2

Cumulative distribution function

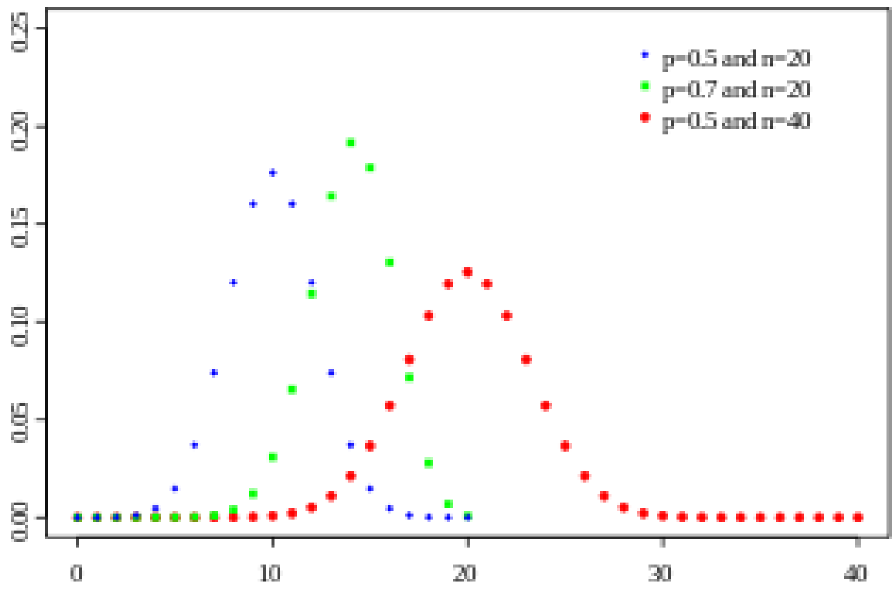


Notation	$\mathcal{N}(\mu, \sigma^2)$
Parameters	$\mu \in \mathbf{R}$ — mean (location) $\sigma^2 > 0$ — variance (squared scale)
Support	$x \in \mathbf{R}$
PDF	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
CDF	$\frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$

https://en.wikipedia.org/wiki/Normal_distribution

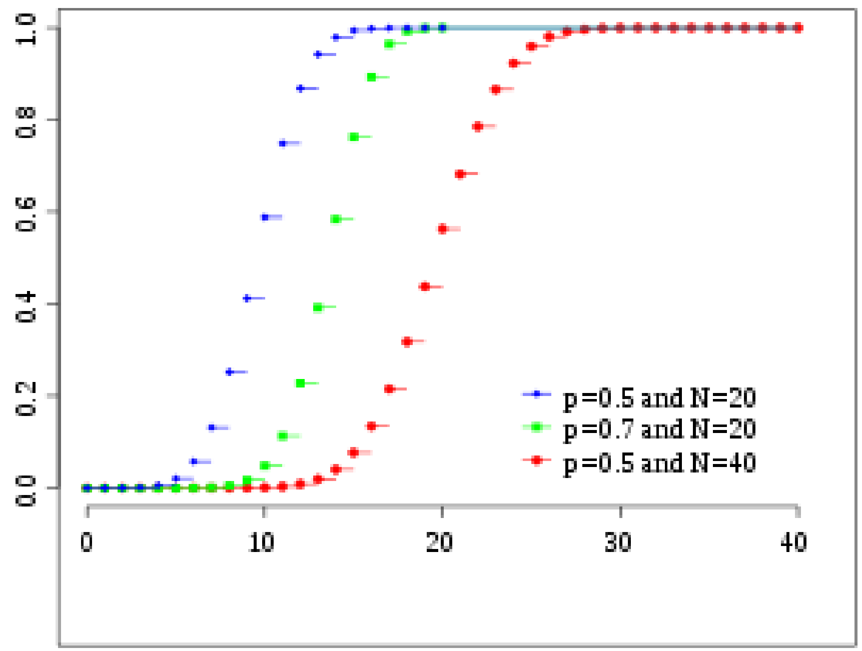
Binomial Distribution

Probability mass function



Notation	$B(n, p)$
Parameters	$n \in \mathbf{N}_0$ — number of trials $p \in [0,1]$ — success probability in each trial
Support	$k \in \{ 0, \ldots, n \}$ — number of successes
pmf	$\binom{n}{k} p^k (1-p)^{n-k}$
CDF	$I_{1-p}(n-k, 1+k)$

Cumulative distribution function



Mean	np
Median	$\lfloor np \rfloor$ or $\lceil np \rceil$
Mode	$\lfloor (n+1)p \rfloor$ or $\lceil (n+1)p \rceil - 1$
Variance	$np(1-p)$

https://en.wikipedia.org/wiki/Binomial_distribution

Binomial Distribution

the **binomial distribution** with parameters *n* and *p*

the discrete probability distribution of the number of successes
in a sequence of *n* independent experiments,
each asking a yes–no question,
and each with its own boolean-valued outcome:

a random variable containing single bit of information:
success / yes / true / one (with probability p)
failure / no / false / zero (with probability q = 1 − p).

A single success / failure experiment
a **Bernoulli trial** or **Bernoulli experiment**

a single trial, i.e., n = 1,
the binomial distribution is a **Bernoulli distribution**.

a sequence of outcomes
a Bernoulli process

https://en.wikipedia.org/wiki/Bernoulli_trial

Binomial Distribution

The binomial distribution is

the basis for the popular binomial test of statistical significance.

frequently used to *model* the number of successes
in a sample of size ***n*** drawn with replacement
from a population of size ***N***.

the sampling carried out without replacement
the draws are not independent
a **hypergeometric distribution**
not a **binomial distribution**

for ***N*** much larger than ***n***,
the **binomial distribution** remains
a good approximation
widely used.

https://en.wikipedia.org/wiki/Bernoulli_trial

Bernoulli Trial

a Bernoulli trial (or binomial trial) is
a random experiment
with exactly **two** possible outcomes,
"success" and "failure",
in which the probability of success is ***the same***
every time the experiment is conducted.

$$p = 1 - q$$
$$q = 1 - p$$
$$p + q = 1$$

$$P(k) = \binom{n}{k} p^k q^{n-k}$$

https://en.wikipedia.org/wiki/Bernoulli_trial

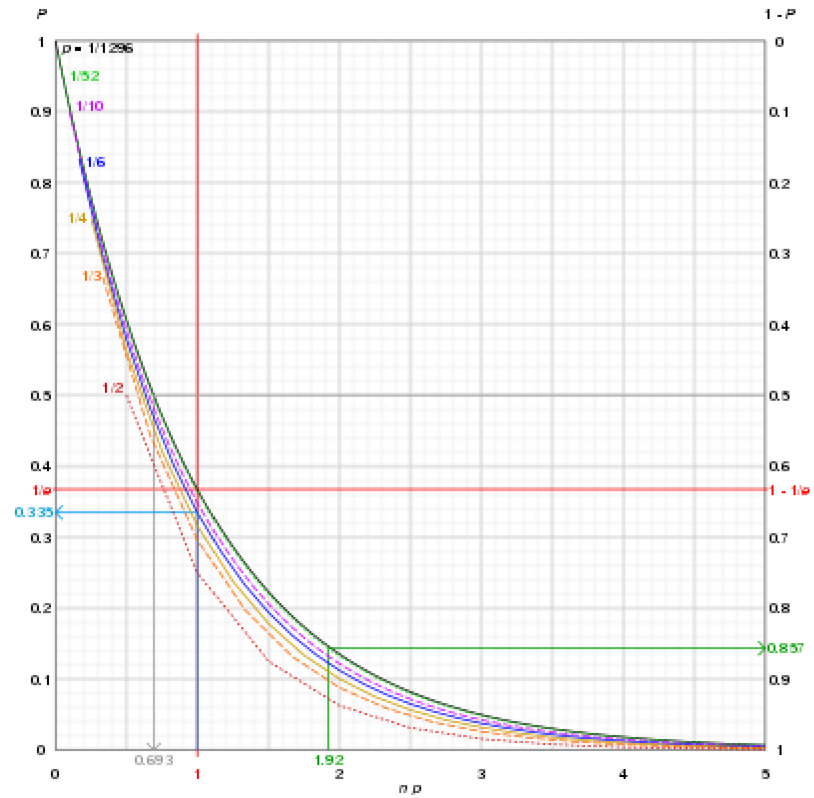
Dependent Events

Graphs of probability P of **not** observing independent events each of probability **p** after ***n*** Bernoulli trials vs **np** for various p.

Blue arrow: Throwing a 6-sided dice 6 times gives 33.5% chance that 6 (or any other given number) never turns up; it can be observed that as ***n*** increases, the probability of a 1/***n***-chance event never appearing after ***n*** tries rapidly converges to 0.

Grey arrow: To get 50-50 chance of throwing a Yahtzee (5 cubic dice all showing the same number) requires 0.69 × 1296 ~ 898 throws.

Green arrow: Drawing a card from a deck of playing cards without jokers 100 (1.92 × 52) times with replacement gives 85.7% chance of drawing the ace of spades at least once.



https://en.wikipedia.org/wiki/Bernoulli_trial

Tossing Coins Probability

$$q = 1 - p = 1 - \tfrac{1}{2} = \tfrac{1}{2}$$

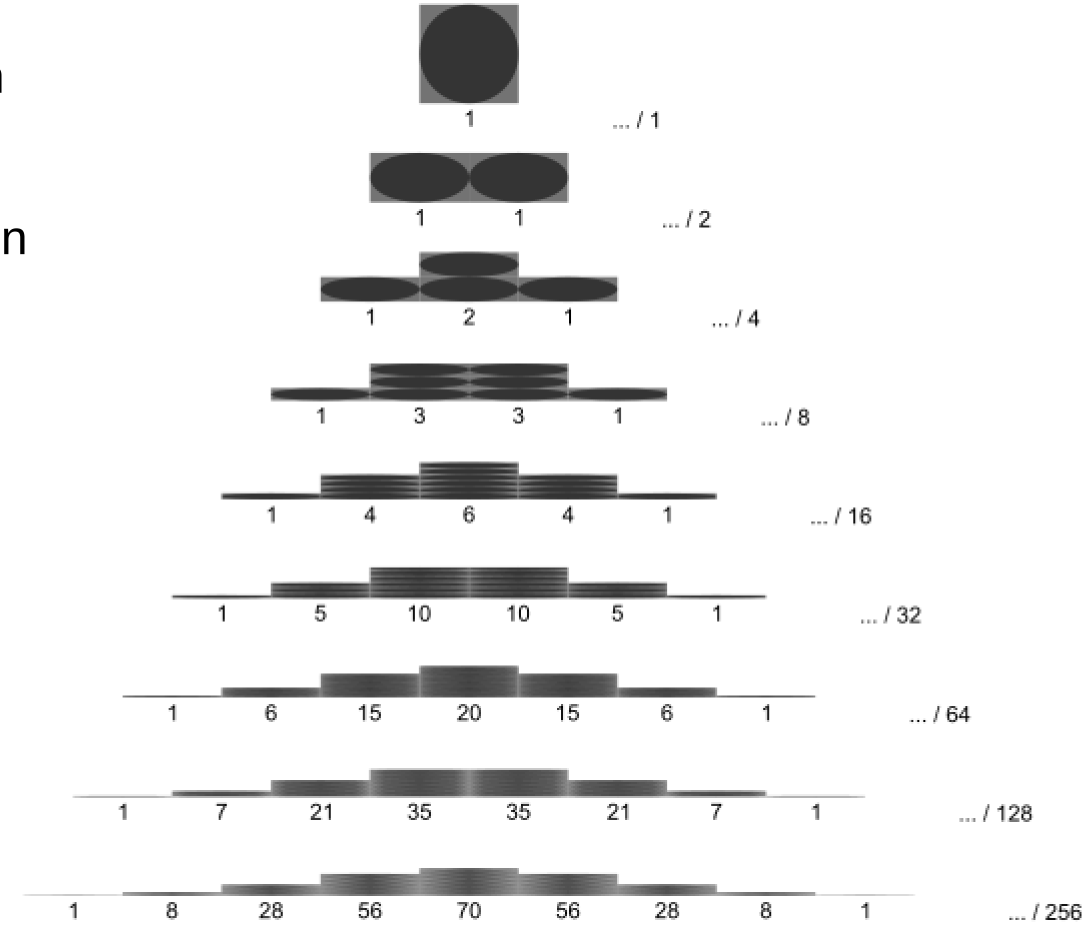
$$\begin{aligned} P(2) &= \binom{4}{2} p^2 q^2 \\ &= 6 \times (\tfrac{1}{2})^2 \times (\tfrac{1}{2})^2 \\ &= \frac{3}{8} \end{aligned}$$

https://en.wikipedia.org/wiki/Bernoulli_trial

Binomial Distribution Examples

Binomial distribution for p = 0.5 with n and k as in Pascal's triangle

The probability that a ball in a Galton box with 8 layers (n = 8) ends up in the central bin (k = 4) is 70 / 256



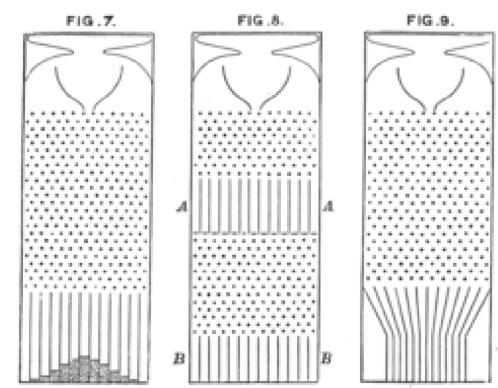
https://en.wikipedia.org/wiki/Binomial_distribution

Bean Machine

If a ball bounces to the right *k* times on its way down (and to the left on the remaining pins) it ends up in the *k*th bin counting from the left. Denoting the number of rows of pins in a bean machine by *n*, the number of paths to the *k*th bin on the bottom is given by the binomial coefficient $\binom{n}{k}$. If the probability of bouncing right on a pin is *p* (which equals 0.5 on an unbiased machine) the probability that the ball ends up in the *k*th bin equals $\binom{n}{k}p^k(1-p)^{n-k}$.

This is the probability mass function of a binomial distribution.

According to the central limit theorem (more specifically, the de Moivre–Laplace theorem), the binomial distribution approximates the normal distribution provided that *n*, the number of rows of pins in the machine, is large.



https://en.wikipedia.org/wiki/Bean_machine

Binomial Distribution – Mean

$$\begin{aligned}
\mu &= \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\
&= np \sum_{k=0}^{n} k \frac{(n-1)!}{(n-k)!k!} p^{k-1} (1-p)^{(n-1)-(k-1)} \\
&= np \sum_{k=1}^{n} \frac{(n-1)!}{((n-1)-(k-1))!(k-1)!} p^{k-1} (1-p)^{(n-1)-(k-1)} \\
&= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)} \\
&= np \sum_{\ell=0}^{n-1} \binom{n-1}{\ell} p^{\ell} (1-p)^{(n-1)-\ell} && \text{with } \ell := k-1 \\
&= np \sum_{\ell=0}^{m} \binom{m}{\ell} p^{\ell} (1-p)^{m-\ell} && \text{with } m := n-1 \\
&= np(p + (1-p))^m \\
&= np
\end{aligned}$$

https://en.wikipedia.org/wiki/Binomial_distribution

Binomial Distribution – Variance

$$X = X_1 + \cdots + X_n$$

$$E[X_i] = p$$

$$E[X] = E[X_1 + \cdots + X_n] = E[X_1] + \cdots + E[X_n] = \underbrace{p + \cdots + p}_{n \text{ times}} = np$$

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

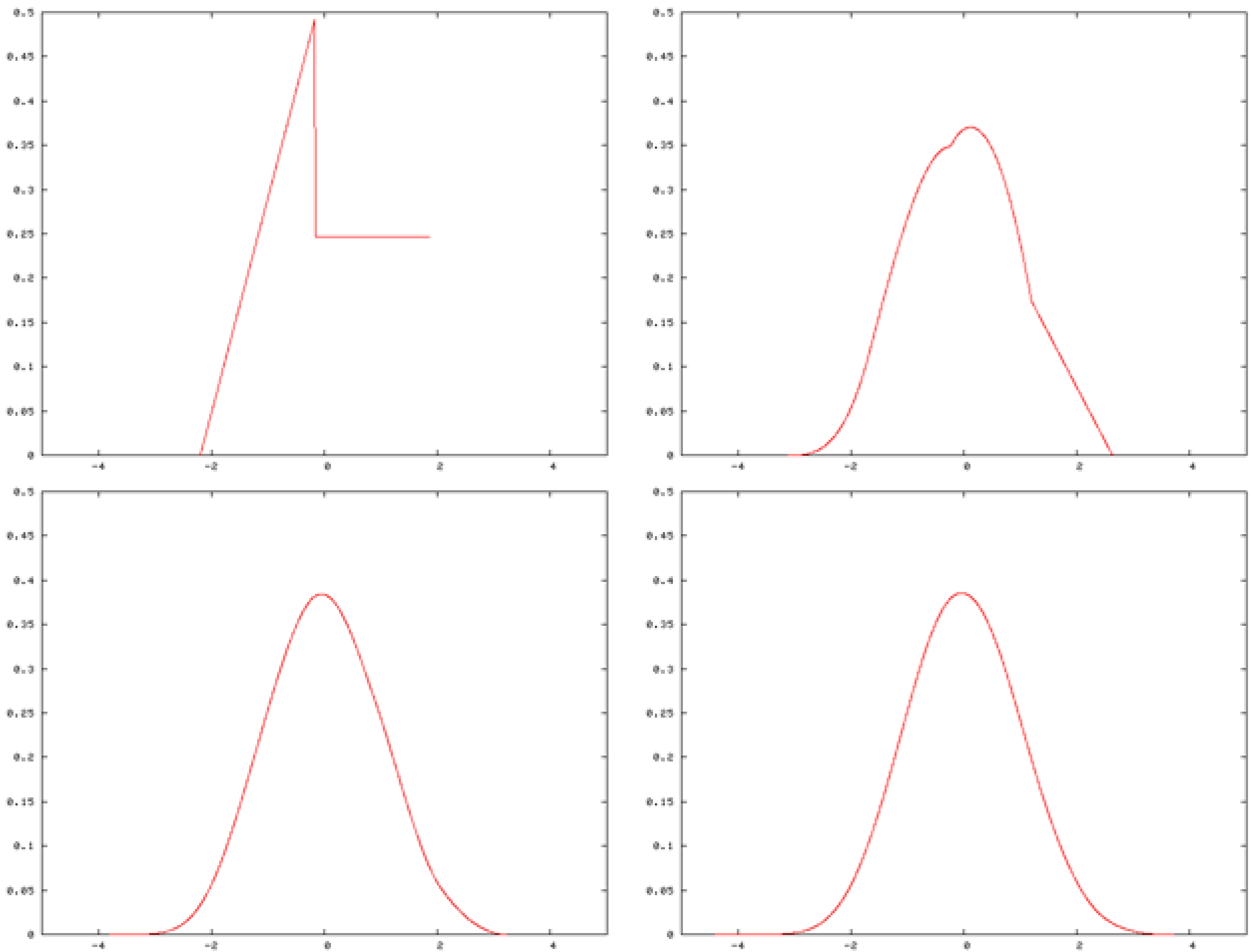
$$\operatorname{Var}(X_i) = p(1-p)$$

$$\operatorname{Var}(X) = \operatorname{Var}(X_1 + \cdots + X_n) = \operatorname{Var}(X_1) + \cdots + \operatorname{Var}(X_n) = n\operatorname{Var}(X_1) = np(1-p).$$

https://en.wikipedia.https://en.wikipedia.org/wiki/Binomial_distribution/wiki/Algorithm

Central Limit Theorem

A distribution being "smoothed out" by summation, showing original density of distribution and three subsequent summations;



https://en.wikipedia.org/wiki/Central_limit_theorem#/media/File:Central_limit_thm.png

Central Limit Theorem

Classical CLT [edit]

Let $\{X_1, \ldots, X_n\}$ be a random sample of size n — that is, a sequence of independent and identically distributed random variables drawn from distributions of expected values given by μ and finite variances given by σ^2. Suppose we are interested in the sample average

$$S_n := \frac{X_1 + \cdots + X_n}{n}$$

of these random variables. By the law of large numbers, the sample averages converge in probability and almost surely to the expected value μ as $n \to \infty$. The classical central limit theorem describes the size and the distributional form of the stochastic fluctuations around the deterministic number μ during this convergence. More precisely, it states that as n gets larger, the distribution of the difference between the sample average S_n and its limit μ, when multiplied by the factor $\sqrt{n}$ (that is $\sqrt{n}(S_n - \mu)$), approximates the normal distribution with mean 0 and variance σ^2. For large enough n, the distribution of S_n is close to the normal distribution with mean μ and variance $\frac{\sigma^2}{n}$. The usefulness of the theorem is that the distribution of $\sqrt{n}(S_n - \mu)$ approaches normality regardless of the shape of the distribution of the individual X_i. Formally, the theorem can be stated as follows:

https://en.wikipedia.org/wiki/Central_limit_theorem

References

[1] http://en.wikipedia.org/
[2] https://en.wikiversity.org/wiki/Discrete_Mathematics_in_plain_view