Notes on Probability

Allin Cottrell

The classical approach

The probability of any event A is the number of outcomes that correspond to A, n_A, divided by the total number of equiprobable outcomes, n, or the proportion of the total outcomes for which A occurs.

$$0 \leq P(A) = \frac{n_A}{n} \leq 1$$

Example: let A be the event of getting an even number when rolling a fair die. Three outcomes correspond to this event, namely 2, 4 and 6, out of a total of six possible outcomes, so $P(A) = \frac{3}{6} = \frac{1}{2}$.

Complementary probabilities

If the probability of some event A is $P(A)$ then the probability that event A does *not* occur, $P(\neg A)$, must be

$$P(\neg A) = 1 - P(A).$$

Example: if the chance of rain for tomorrow is 80 percent, the chance that it doesn't rain tomorrow must be 20 percent.

When trying to compute a given probability, it is sometimes *much* easier to compute the complementary probability first, then subtract from 1 to get the desired answer.

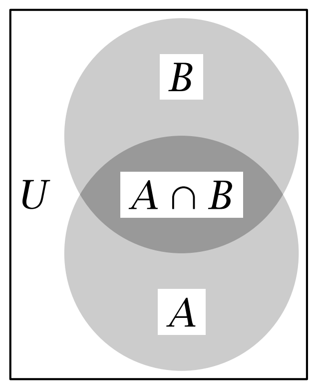
Addition Rule

A means of calculating the probability of $A \cup B$, the probability that either of two events occurs.

With equiprobable outcomes, $P(A) = \frac{n_A}{n}$ and $P(B) = \frac{n_B}{n}$.

First approximation: $P(A \cup B) = \frac{n_A + n_B}{n}$.

Problem: $n_A + n_B$ may overstate the number of outcomes corresponding to $A \cup B$: we must subtract the number of outcomes contained in the intersection, $A \cap B$, namely n_{AB}.



Thus the full version of the addition rule is:

$$
\begin{aligned}
P(A \cup B) &= \frac{n_A + n_B - n_{AB}}{n} \\
&= \frac{n_A}{n} + \frac{n_B}{n} - \frac{n_{AB}}{n} \\
&= P(A) + P(B) - P(A \cap B)
\end{aligned}
$$

Multiplication rule

Clearly

$$\frac{n_{AB}}{n} \equiv \frac{n_A}{n} \times \frac{n_{AB}}{n_A}$$

(the RHS is the LHS multiplied by $n_A/n_A = 1$).

$n_{AB}/n \equiv P(A \cap B)$ is the probability that A and B both occur.

$n_A/n \equiv P(A)$ represents the "marginal" (unconditional) probability of A.

n_{AB}/n_A represents the number of outcomes in $(A \cap B)$ over the number of outcomes in A, or "the probability of B given A".

The general form of the multiplication rule for joint probabilities is therefore:

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

Special case: A and B are *independent*. Then $P(B|A)$ equals the marginal probability $P(B)$ and the rule simplifies:

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

Exercises

- The probability of snow tomorrow is .20, and the probability of all members of ECN 215 being present in class is .8 (let us say). What is the probability of both these events occurring?

- A researcher is experimenting with several regression equations. Unknown to him, all of his formulations are in fact worthless, but nonetheless there is a 5 per cent chance that each regression will—by the luck of the draw—appear to come up with 'significant' results. Call such an event a 'success'. If the researcher tries 10 equations, what is the probability that he has exactly one success? What is the probability of at least one success?

Marginal probabilities

$$P(A) = \sum_{i=1}^{N} P(A|E_i) \times P(E_i)$$

where $E_1, \ldots, E_N$ represent N mutually exclusive and jointly exhaustive events.

Example:

	conditional on E_i:		
	snow ($P = \frac{2}{10}$)	$\neg$ snow ($P = \frac{8}{10}$)	
P(all here)	$\frac{6}{10}$	$\frac{9}{10}$	
product	$\frac{12}{100}$	$\frac{72}{100}$	$\Sigma = \frac{84}{100}$

Conditional probabilities

In general,

$$P(A|B) \neq P(B|A)$$

the probability of A given B is *not* the same as the probability of B given A.

Example: The police department of a certain city finds that 60 percent of cyclists involved in accidents at night are wearing light-colored clothing. *How can we express this in terms of conditional probability? Should we conclude that wearing light-colored clothing is dangerous?*

Discrete random variables

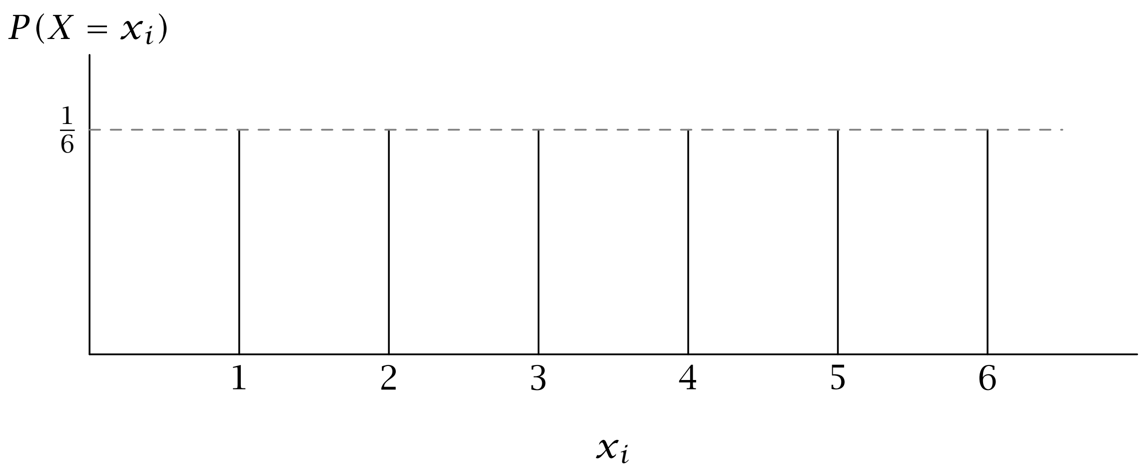
The *probability distribution* for a random variable X is a mapping from the possible values of X to the probability that X takes on each of those values.

x_i	$P(X = x_i)$	$x_i P(X = x_i)$
1	$\frac{1}{6}$	$\frac{1}{6}$
2	$\frac{1}{6}$	$\frac{2}{6}$
3	$\frac{1}{6}$	$\frac{3}{6}$
4	$\frac{1}{6}$	$\frac{4}{6}$
5	$\frac{1}{6}$	$\frac{5}{6}$
6	$\frac{1}{6}$	$\frac{6}{6}$
$\sum$	$\frac{6}{6} = 1$	$\frac{21}{6} = 3.5 = E(X)$

$$E(X) \equiv \mu_X = \sum_{i=1}^{N} x_i P(X = x_i)$$

The mean is the probability-weighted sum of the possible values of the random variable.

Uniform distribution (one die):



Variance

Probability-weighted sum of the squared deviations of the possible values of the random variable from its mean, or *expected value of the squared deviation from the mean*.

$$
\begin{aligned}
\mathrm{Var}(X) \equiv \sigma_X^2 &= \sum_{i=1}^{N} (x_i - \mu)^2 P(X = x_i) \\
&= E\,(X - \mu)^2 \\
&= E\left(X^2 - 2X\mu + \mu^2\right) \\
&= E(X^2) - 2E(X\mu) + \mu^2 \\
&= E(X^2) - 2\mu^2 + \mu^2 \\
&= E(X^2) - \mu^2 \\
&= E(X^2) - [E(X)]^2
\end{aligned}
$$

Note that in general $E(X^2) \neq [E(X)]^2$.

Example: variance for one die

x_i	$P(X = x_i)$	$x_i - \mu$	$(x_i - \mu)^2$	$(x_i - \mu)^2 P(X = x_i)$
1	$\frac{1}{6}$	-2.5	6.25	1.0417
2	$\frac{1}{6}$	-1.5	2.25	0.3750
3	$\frac{1}{6}$	-0.5	0.25	0.0833
4	$\frac{1}{6}$	$+0.5$	0.25	0.0833
5	$\frac{1}{6}$	$+1.5$	2.25	0.3750
6	$\frac{1}{6}$	$+2.5$	6.25	1.0417
$\sum$	1	0		$2.917 = \text{Var}(X)$

Two dice

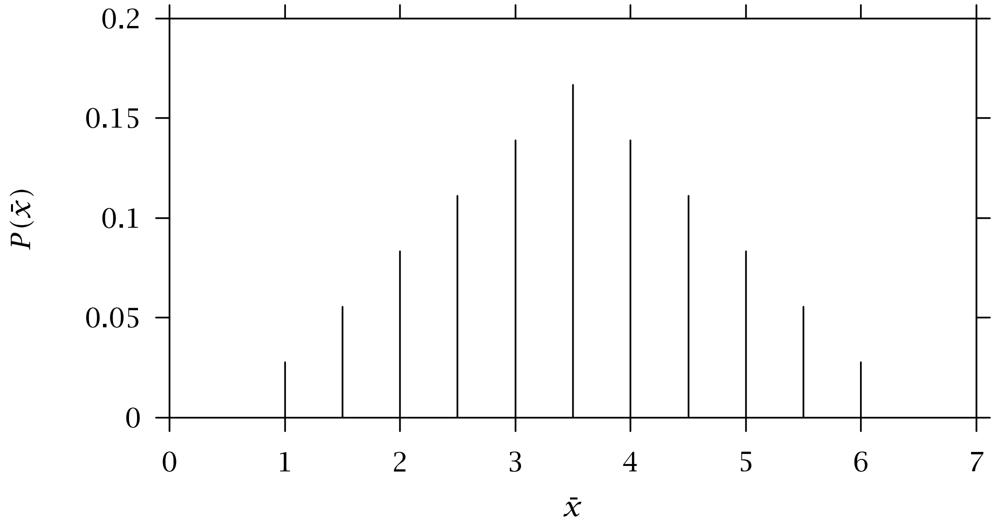
Sample space:

(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

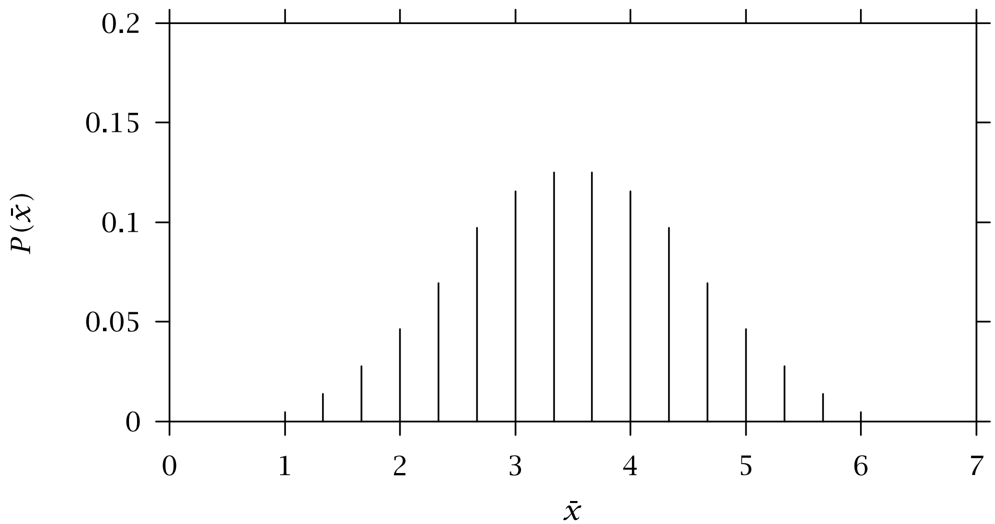
1.0	1.5	2.0	2.5	3.0	3.5
1.5	2.0	2.5	3.0	3.5	4.0
2.0	2.5	3.0	3.5	4.0	4.5
2.5	3.0	3.5	4.0	4.5	5.0
3.0	3.5	4.0	4.5	5.0	5.5
3.5	4.0	4.5	5.0	5.5	6.0

x_i	$P(x_i)$	$x_i P(x_i)$	$x_i - \mu$	$(x_i - \mu)^2$	$\times P(x_i)$
1.0	$\frac{1}{36}$	$\frac{1}{36}$	−2.5	6.25	0.17
1.5	$\frac{2}{36}$	$\frac{3}{36}$	−2.0	4.00	0.22
2.0	$\frac{3}{36}$	$\frac{6}{36}$	−1.5	2.25	0.19
2.5	$\frac{4}{36}$	$\frac{10}{36}$	−1.0	1.00	0.11
3.0	$\frac{5}{36}$	$\frac{15}{36}$	−0.5	0.25	0.03
3.5	$\frac{6}{36}$	$\frac{21}{36}$	0.0	0.00	0.00
4.0	$\frac{5}{36}$	$\frac{20}{36}$	0.5	0.25	0.03
4.5	$\frac{4}{36}$	$\frac{18}{36}$	1.0	1.00	0.11
5.0	$\frac{3}{36}$	$\frac{15}{36}$	1.5	2.25	0.19
5.5	$\frac{2}{36}$	$\frac{11}{36}$	2.0	4.00	0.22
6.0	$\frac{1}{36}$	$\frac{6}{36}$	2.5	6.25	0.17
$\sum$	$\frac{36}{36} = 1$	$\frac{126}{36} = 3.5$	0.0		1.46

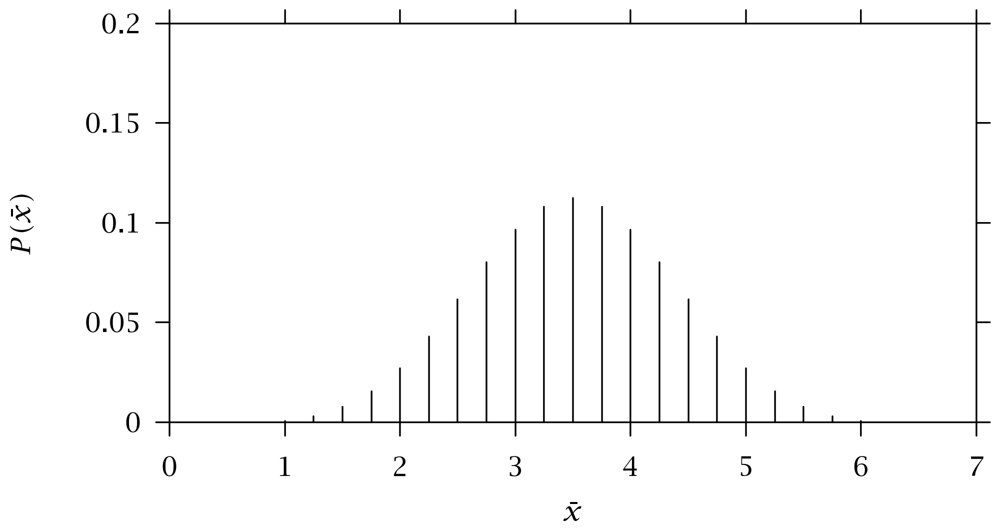
Two dice



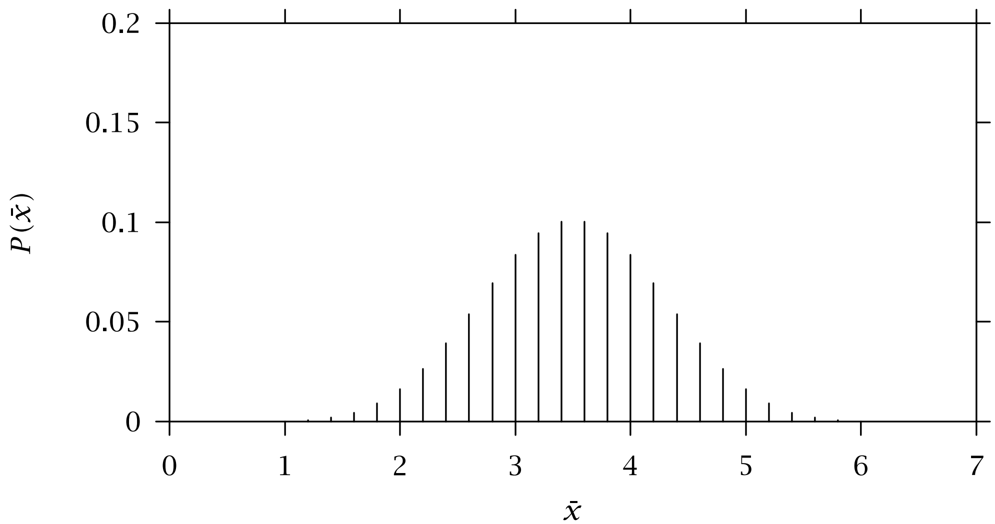
Three dice



Four dice



Five dice



Measures of Association

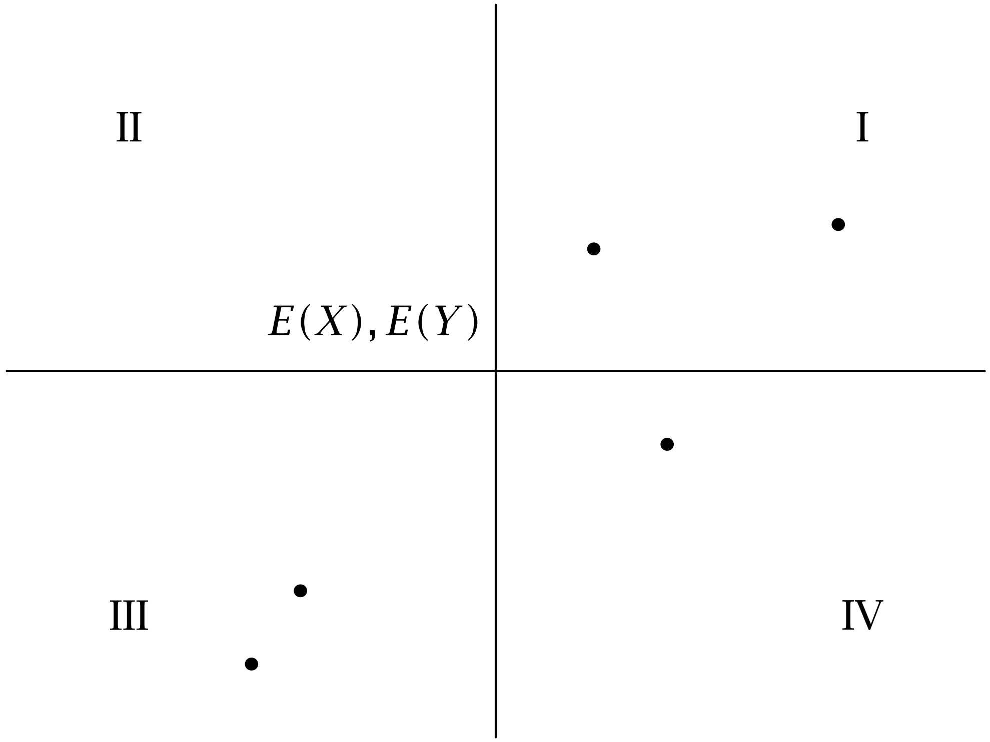
The *covariance* of X and Y is the expected value of the cross-product, deviation of X from its mean times deviation of Y from its mean.

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

or

$$\mathrm{Cov}(X,Y) = \frac{1}{N}\sum_{i=1}^{N}[x_i - E(X)][y_j - E(Y)]$$

It measures the linear association between X and Y.



Cross-products are positive in I and III, negative in II and IV.

The *correlation coefficient* for two variables X and Y is a scaled version of covariance: divide through by the product of the standard deviations of the two variables.

$$\rho_{XY} = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

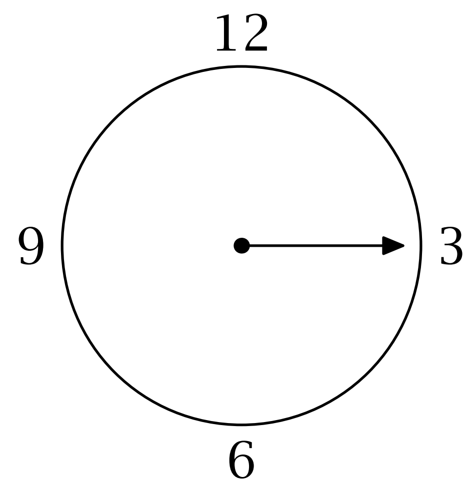
The *correlation coefficient* for two variables X and Y is a scaled version of covariance: divide through by the product of the standard deviations of the two variables.

$$\rho_{XY} = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

Note that $-1 \leq \rho \leq +1$.

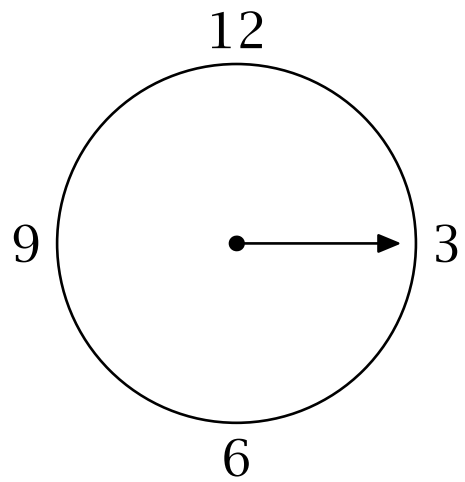
Continuous random variables

Let the random variable X = the number towards which the spinner points when it comes to rest.



Continuous random variables

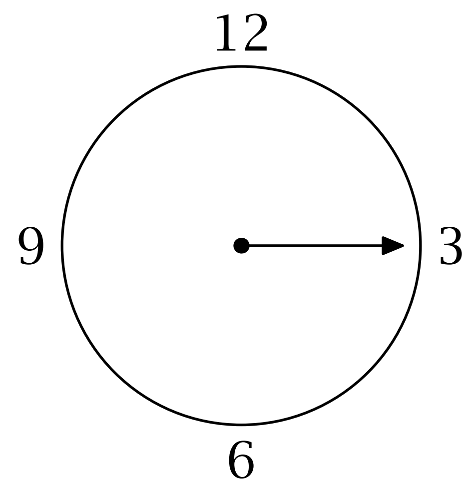
Let the random variable X = the number towards which the spinner points when it comes to rest.



To find probabilities, think in terms of *fractions* of the total measure.

Continuous random variables

Let the random variable X = the number towards which the spinner points when it comes to rest.

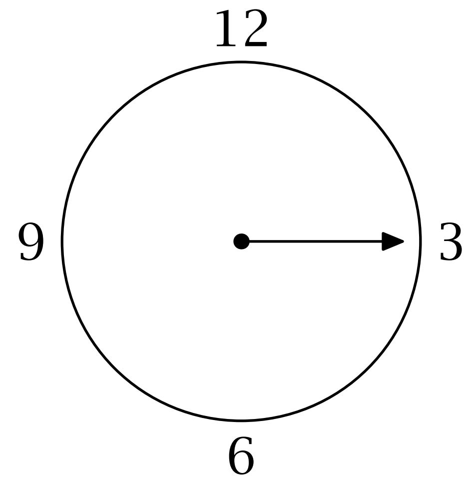


To find probabilities, think in terms of *fractions* of the total measure.

$$P(0 < X < 3) = 3/12 = 1/4$$

Continuous random variables

Let the random variable X = the number towards which the spinner points when it comes to rest.



To find probabilities, think in terms of *fractions* of the total measure.

$$P(0 < X < 3) = 3/12 = 1/4$$

$$P(7 < X < 9) = 2/12 = 1/6$$

Cumulative density function or cdf:

$$F(x) = P(X < x)$$

Cumulative density function or cdf:

$$F(x) = P(X < x)$$

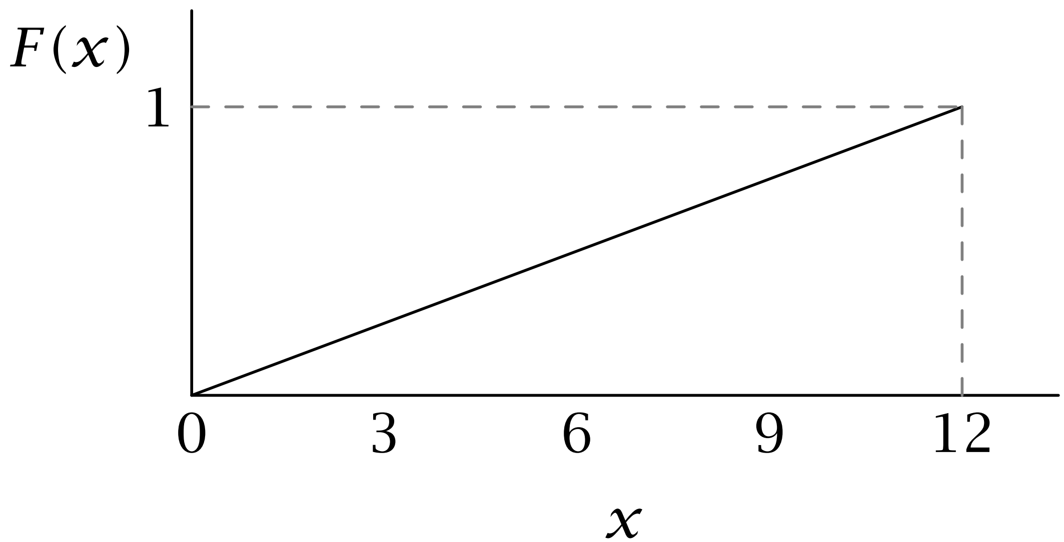
The probability that a random variable X has a value less than some specified value, x.

Cumulative density function or cdf:

$$F(x) = P(X < x)$$

The probability that a random variable X has a value less than some specified value, x.

cdf for the spinner example:



Probability density function or pdf:

$$f(x) = \frac{d}{dx} F(x)$$

Probability density function or pdf:

$$f(x) = \frac{d}{dx} F(x)$$

Derivative of the cdf with respect to x.

Probability density function or pdf:

$$f(x) = \frac{d}{dx} F(x)$$

Derivative of the cdf with respect to x.

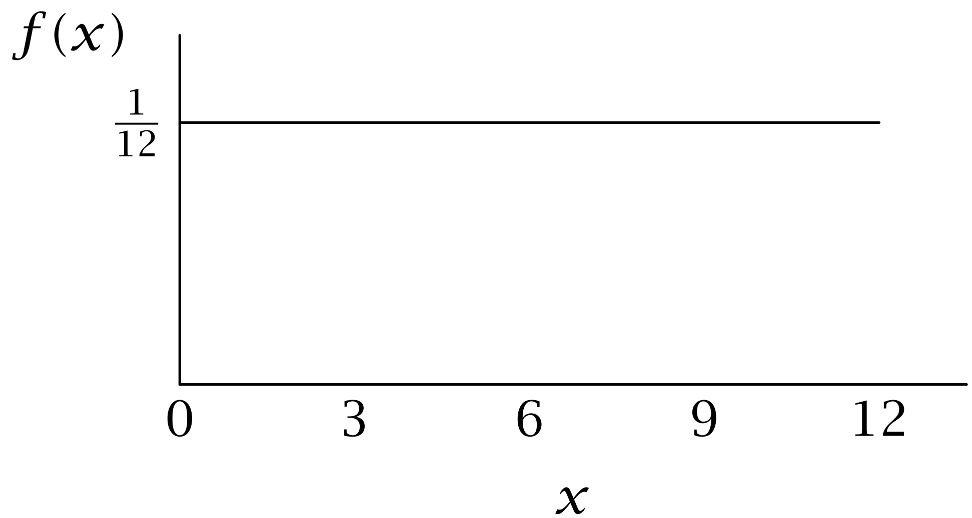
Determine the probability of X falling into any given range by taking the integral of the pdf over that interval.

Probability density function or pdf:

$$f(x) = \frac{d}{dx} F(x)$$

Derivative of the cdf with respect to x.

Determine the probability of X falling into any given range by taking the integral of the pdf over that interval.

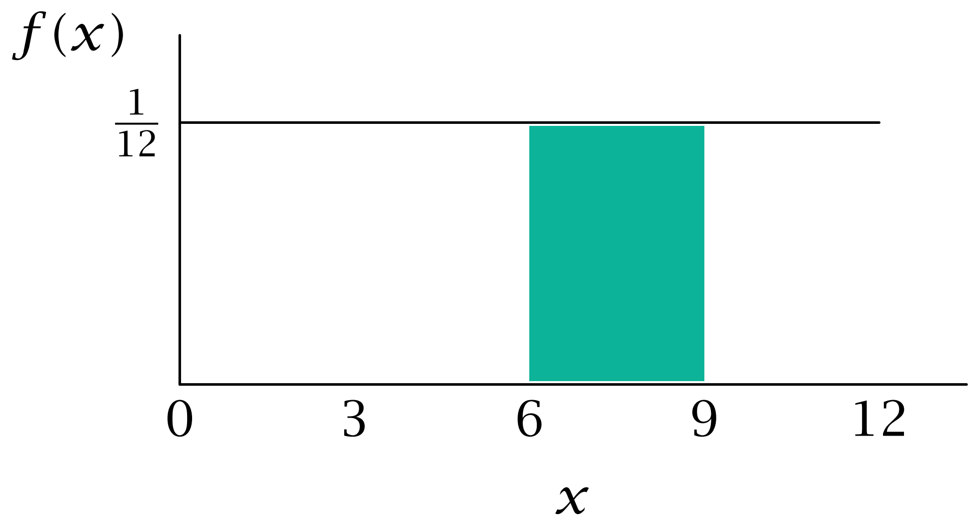


Probability density function or pdf:

$$f(x) = \frac{d}{dx} F(x)$$

Derivative of the cdf with respect to x.

Determine the probability of X falling into any given range by taking the integral of the pdf over that interval.



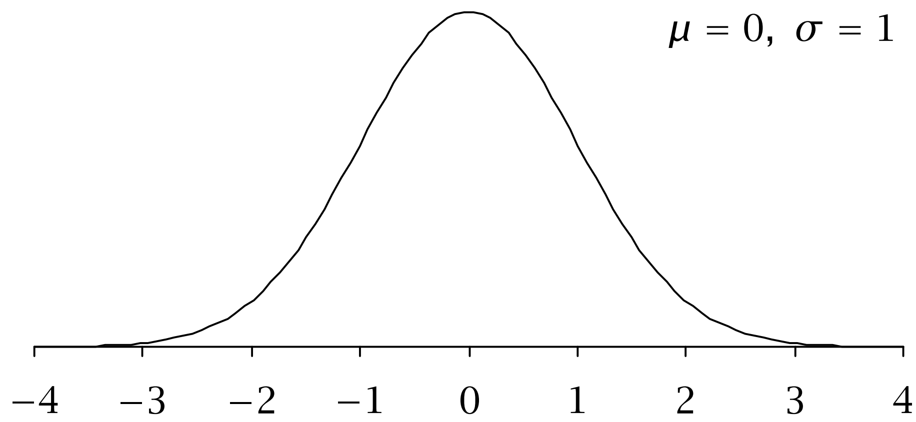
$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x)\, dx$$

Gaussian distribution

Central Limit Theorem: If a random variable X represents the summation of numerous independent random factors then, *regardless of the specific distribution of the individual factors*, X will tend to follow the normal or Gaussian distribution.

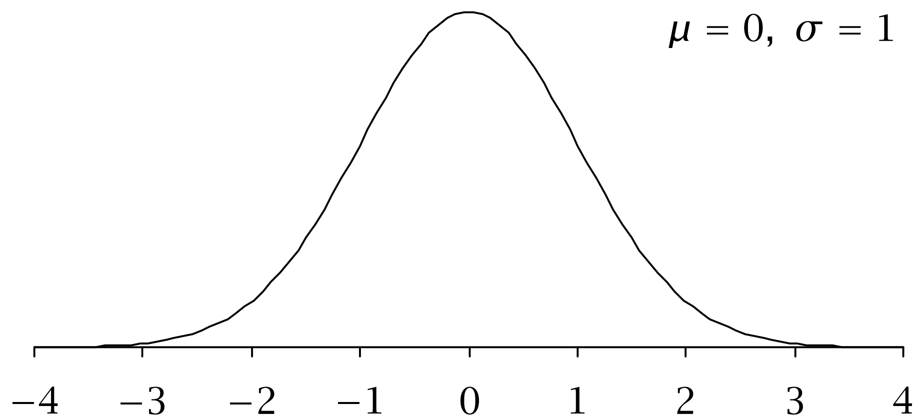
Gaussian distribution

Central Limit Theorem: If a random variable X represents the summation of numerous independent random factors then, *regardless of the specific distribution of the individual factors*, X will tend to follow the normal or Gaussian distribution.



Gaussian distribution

Central Limit Theorem: If a random variable X represents the summation of numerous independent random factors then, *regardless of the specific distribution of the individual factors*, X will tend to follow the normal or Gaussian distribution.



General formula for the normal pdf:

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

The *standard normal* distribution is obtained by setting $\mu = 0$ and $\sigma = 1$; its pdf is

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \qquad -\infty < x < \infty$$

The *standard normal* distribution is obtained by setting $\mu = 0$ and $\sigma = 1$; its pdf is

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \qquad -\infty < x < \infty$$

Commit to memory:

$P(\mu - 2\sigma < x < \mu + 2\sigma) \approx 0.95$

$P(\mu - 3\sigma < x < \mu + 3\sigma) \approx 0.997$

The *standard normal* distribution is obtained by setting $\mu = 0$ and $\sigma = 1$; its pdf is

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \qquad -\infty < x < \infty$$

Commit to memory:

$P(\mu - 2\sigma < x < \mu + 2\sigma) \approx 0.95$

$P(\mu - 3\sigma < x < \mu + 3\sigma) \approx 0.997$

A compact notation for saying that x is distributed normally with mean μ and variance σ^2 is $x \sim N(\mu, \sigma^2)$. □

Contents

- The classical approach ❖
- Addition Rule ❖
- Multiplication rule ❖
- Marginal probabilities ❖
- Conditional probabilities ❖
- Discrete random variables ❖
- Measures of Association ❖
- Continuous random variables ❖
- Gaussian distribution ❖