Probability Theory Review

Last Edited: February 14, 2022

ő A Note

This review covers some of the basics of calculus based probability theory.

For those of you interested in measure theory based probability theory see Rick Durrett's Probability: Theory and Examples.

Probability Spaces

A *probability space* is a triple ($\mathcal{S}$, $\mathcal{F}$, P) where:

- $\mathcal{S}$ is known as the *sample space*, the collection of all possible simple outcomes
  - Example: $\mathcal{S}$ = {H, T}, when flipping a coin
- $\mathcal{F}$ is known as the *event space*, the collection of all possible subsets of the sample space
  - Example $\mathcal{F}$ = {∅, {H}, {T}, {H, T}}, when flipping a coin
- P is a probability function that assigns each possible subset of the sample space a value in [0, 1].
  - Example P(H) = P(T) = 0.5, P(Ø) = 0 and P({H, T}) = 1 when flipping a balanced coin.

Events

An *event* is a collection of simple outcomes.

Example: When flipping a coin:

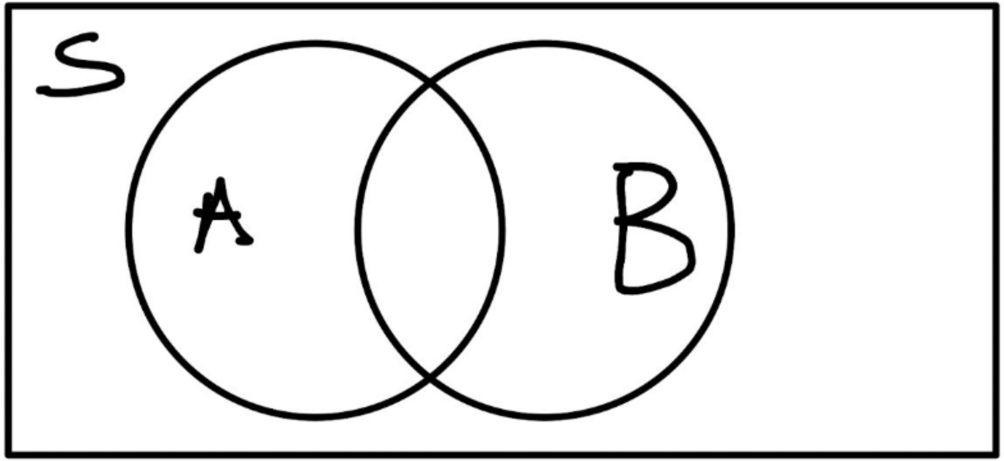
- The coin landing heads, {H}, is an event
- The coin landing heads or tails, {H, T} is an event

Example: When Rolling a 6-Sided Die:

- Rolling a 2, {2}, is an event
- Rolling an odd number, {1, 3, 5}, is an event

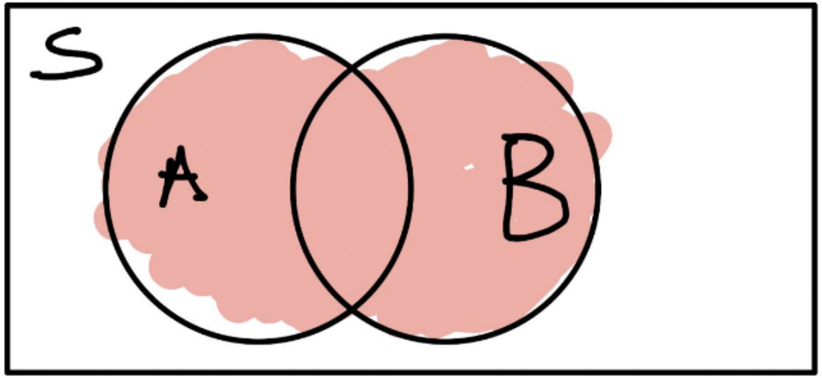
ő Probability Spaces Continued

For slides 4-8 suppose we have the probability space, ($\mathcal{S}$, $\mathcal{F}$, P), with events A and B pictured to the right.

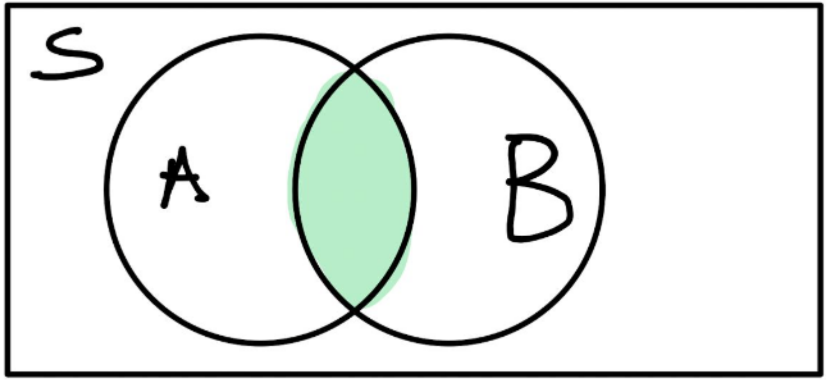


Unions and Intersection

The *union* of events A and B, denoted $A \cup B$, is the collection of outcomes that occur in A **or** B.



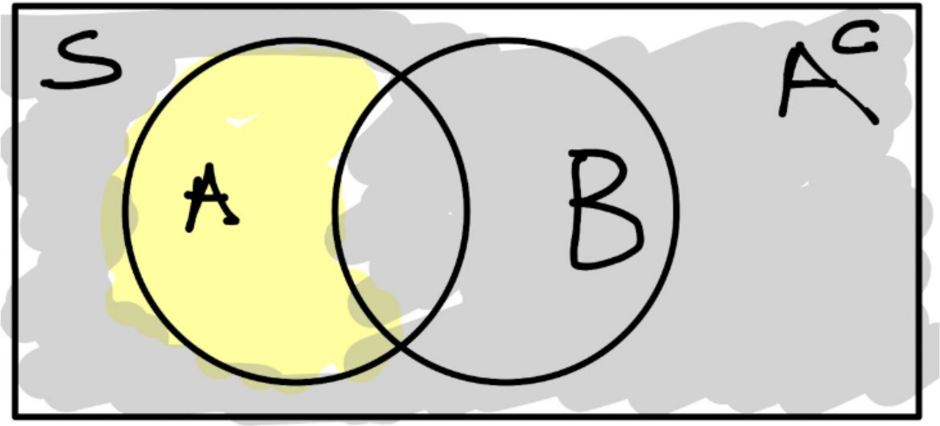
The *intersection* of events A and B, denoted $A \cap B$, is the collection of outcomes that occur in both A **and** B. We sometimes denote this as AB.



Complements

For an event, A, in a probability space, ($\mathcal{S}$, $\mathcal{F}$, P) we define the complement of A as all outcomes in $\mathcal{S}$ that are not in A.

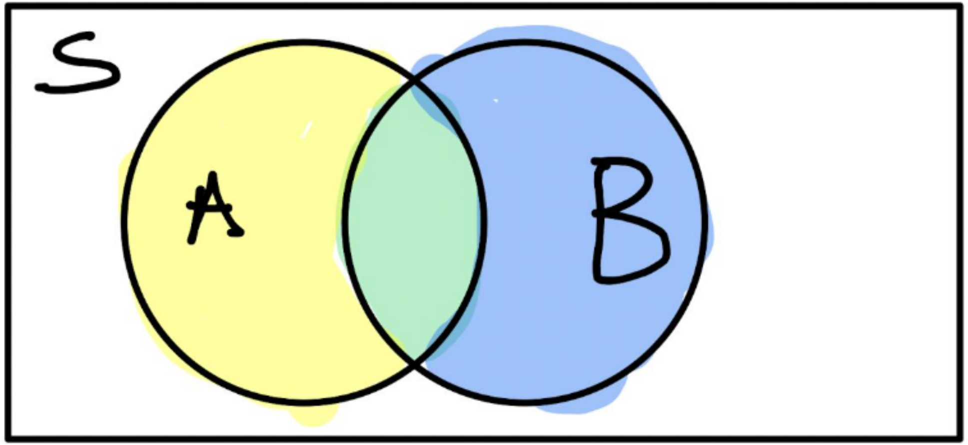
We denote the complement of A as A^c.



Basic Probability Properties

P has the following properties:

1. $P(\varnothing) = 0$, where $\varnothing$ denotes the empty event,
2. $P(\mathcal{S}) = 1$,
3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.



It can be also be shown that:

- $P(A^c) = 1 - P(A)$

ő Conditional Probability

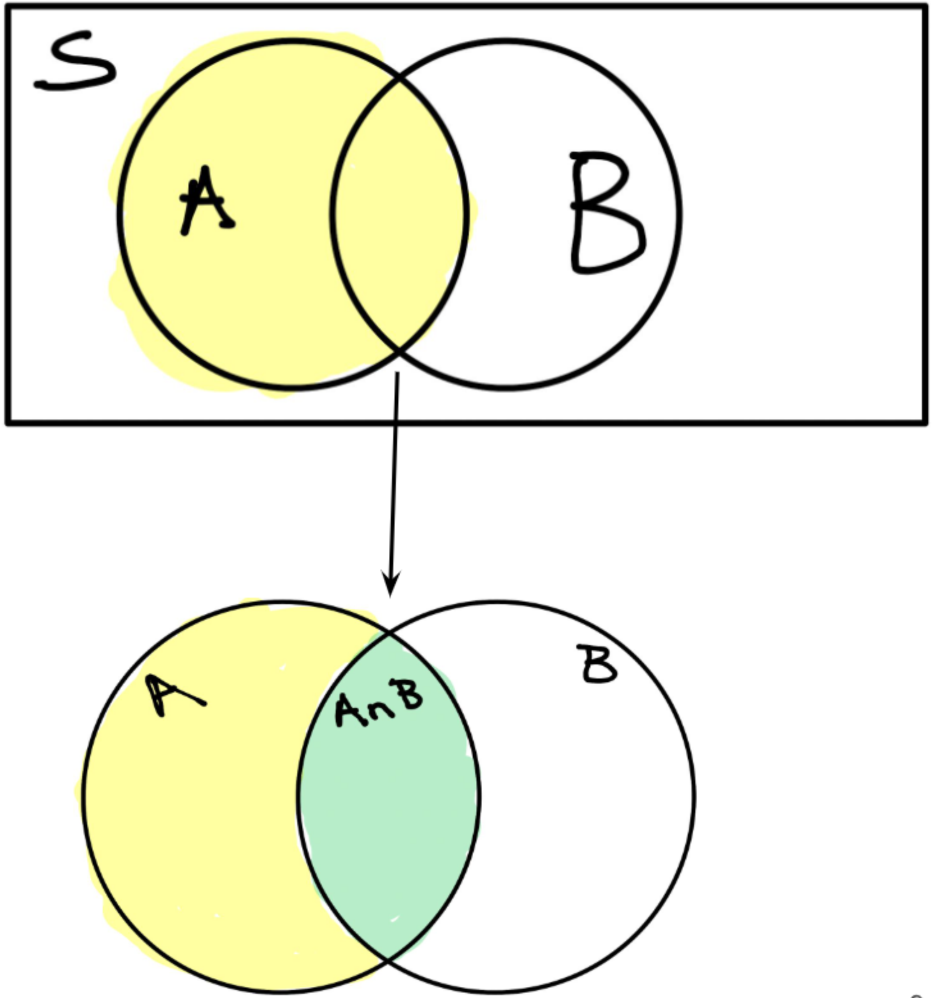
The *conditional probability of event B given that event A has occurred* is defined as:

P(B | A) = P(A∩B) / P(A),

assuming P(A) ≠ 0.

We can see from this that:

P(A∩B) = P(A) P(B | A)



ő Independent Events

Two events, A and B, are *independent* if:

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

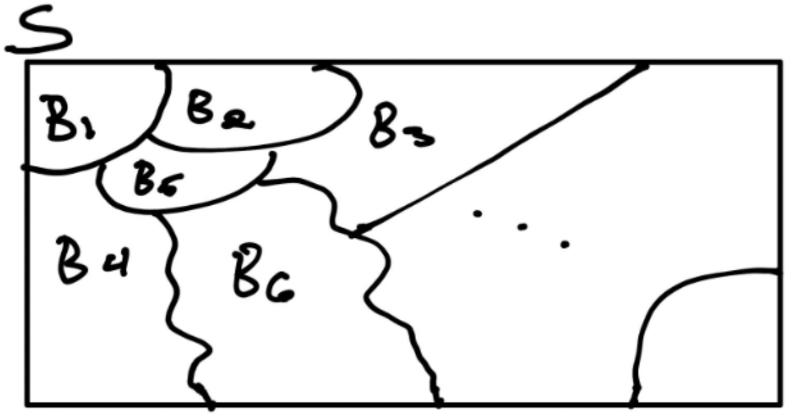
or equivalently

$$P(B \mid A) = P(B), P(A|B) = P(A).$$

Law of Total Probability

Suppose there are events $B_1, B_2, \ldots, B_n$ so that:

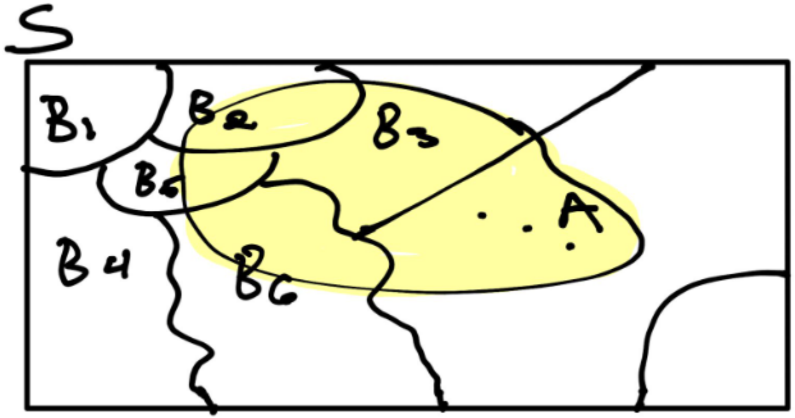
- $B_i \cap B_j = \varnothing$ for all pairs i, j and
- $\mathcal{S} = B_1 \cup B_2 \cup \ldots \cup B_n$.



The Law of Total Probability says for an event A,

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + \ldots + P(A \cap B_n)$$

$$= P(A \mid B_1)\, P(B_1) + \ldots + P(A \mid B_n)\, P(B_n)$$



Bayes' Rule

Suppose A and B are events and $P(B) \neq 0$, then *Bayes' rule* says:

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

Note that this is often combined with the law of total probability to break up the denominator into a sum.

Random Variables

Let ($\mathcal{S}$, $\mathcal{F}$, P) be a probability space.

A *random variable* is a function from $\mathcal{S}$ to the real numbers.

Example: Suppose the probability space corresponds to flipping a balanced coin two times. X = the number of heads is a random variable with:

- X(T,T) = 0,
- X(H,T) = X(T,H) = 1 and
- X(H,H) = 2.

ő Random Variables Continued

Discrete Random Variables

A random variable that takes on values from an at most countably infinite set.

Example: The sum of rolling two 6-sided dice.

Example: Selecting an integer at random.

Non-Discrete Random Variables

A random variable that takes on values from an uncountably infinite set.

Example: The height of a randomly selected person.

Example: A random value taken from the interval [-71, 199].

ő Probability Distributions

Let Y be a random variable.

The *probability distribution* of Y, is defined to be p(y) = P(Y=y) for all real numbers y.

Example: If Y is the number of heads from the result of flipping a balanced coin twice then:

- P(Y=0) = 0.25,
- P(Y=1) = 0.50,
- P(Y=2) = 0.25 and
- P(Y=y) = 0 for all other real numbers y.

Note: For any non-discrete random variable, Y, P(Y=y) = 0 for all real numbers, y.

Cumulative Distribution Functions (CDF)

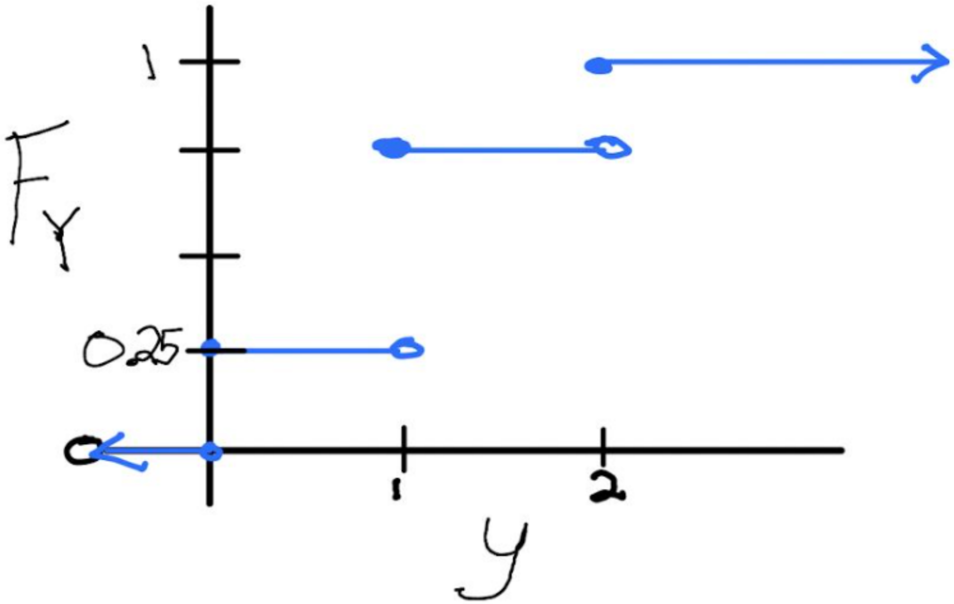
Let Y be a random variable.

The *cumulative distribution function (CDF)* of Y, is defined to be $F_Y(y) = P(Y \leq y)$ for all real numbers y.

CDFs have the following properties:

- $F_Y(y) \to 0$ as $y \to -\infty$,
- $F_Y(y) \to 1$ as $y \to \infty$ and
- $F_Y(y) \leq F_Y(x)$ for $y \leq x$.

Example: The graph of the CDF for Y = the number of heads from the result of flipping a balanced coin twice is given on the right.



ő Probability Density Functions (PDFs)

Let Y be a random variable.

The *probability density function* of Y, if it exists, is defined to be the derivative of $F_Y(y)$ and is denoted as $f_Y(y)$

Some properties of $f_Y(y)$ include:

- $f_Y(y) \geq 0$
- $\int_{-\infty}^{\infty} f_Y(y)\, dy = 1$
- $F_Y(y) = \int_{-\infty}^{y} f_Y(x)\, dx$

ő Independent Random Variables

Let X and Y be random variables. We say that X and Y are *independent random variables* if:

$$F_{X,Y}(x, y) = P(X \le x \text{ and } Y \le y) = P(X \le x)P(Y \le y) = F_X(x)F_Y(y).$$

ő Expectation of a Random Variable

For a random variable Y, the *expectation* of Y, denoted E(Y), can be thought of as the "average" value of Y. More specifically:

<u>Discrete Random Variables</u>

$$E(Y) = \sum_{y \in \text{range}(Y)} y p(y)$$

<u>Non-Discrete Random Variables</u>

$$E(Y) = \int_{-\infty}^{\infty} y f(y) dy$$

Assuming Y has a probability density function f(y).

Note that for a continuous random variable without a pdf we can still define the expectation, see Durrett.

ő Expectation of a Random Variable: Example Calculation

If Y = number of heads from flipping two balanced coins then

E(Y) = 0*0.25 + 1*0.5 + 2*0.25 = 1.

ő Properties of Expectation

Suppose X and Y are random variables with a finite expected value, then:

- $E(aX + bY) = aE(X) + bE(Y)$, for constants a and b,
- If $X \geq Y$ almost surely, then $E(X) \geq E(Y)$ and
- If X and Y are independent, then $E(XY) = E(X)E(Y)$

ő Variance of a Random Variable

For a random variable Y, the *variance* of Y, denoted Var(Y), can be thought of as how far Y tends to be from its expected value.

Formally it is defined as:

$$\mathrm{Var}(Y) = E[\,(Y - E(Y))^2\,] = E(Y^2) - E(Y)^2.$$

Note, there is no guarantee that the variance of Y is finite.

Variance of a Random Variable: Example Calculation

If Y = number of heads from flipping two balanced coins then

$Var(Y) = E(Y^2) - E(Y)^2$,

$E(Y^2) = 0*0.25 + 1*0.5 + 4*0.25 = 1.5$ and $E(Y)^2 = 1$ so

$Var(Y) = 1.5 - 1 = 0.5$

ő Covariance of Two Random Variables

Let X and Y be two random variables.

The *covariance* between X and Y is defined to be:

$$Cov(X,Y) = E[(X - E(X))\,(Y - E(Y))] = E(XY) - E(X)E(Y).$$

ő Properties of Variance/Covariance

Suppose X and Y are random variables with a finite variance, then:

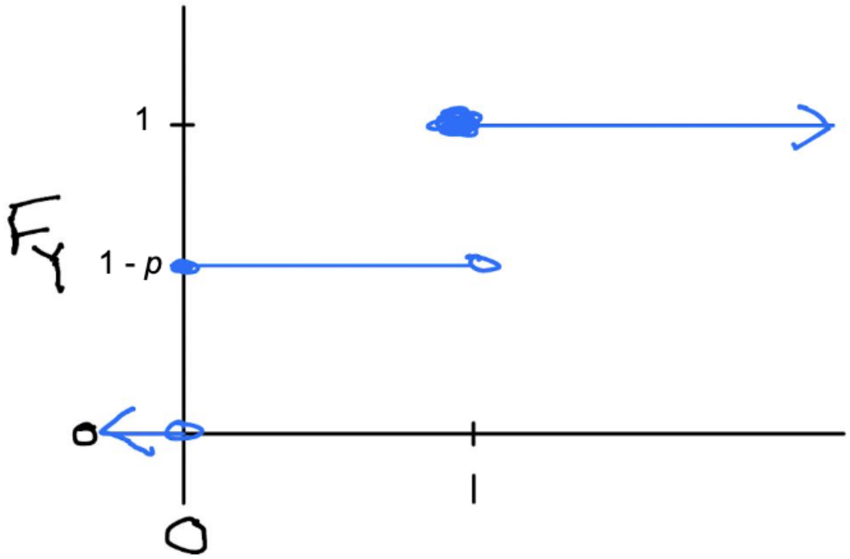
- $Var(X) \geq 0$,
- $Var(aX) = a^2 Var(X)$ for any constant a,
- $Var(X) = Cov(X, X)$
- $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X, Y)$,
- $Cov(X, Y) = Cov(Y, X)$ and
- If X and Y are independent, $Cov(X, Y) = 0$.

Common Probability Distributions

Bernoulli

If Y is a Bernoulli random variable then it takes the value 1 with probability p and the value 0 with probability $1 - p$.

- $E(Y) = p$ and
- $Var(Y) = p(1 - p)$.



Commonly used to model coin tosses.

CDF seen to the right.

ő Common Probability Distributions

<u>Binomial</u>

The number of successes from *n* independent identically distributed Bernoulli trials with *p* as the probability of success. If X is a binomial random variable then

- $P(X = x) = p(x) = \binom{n}{x} p^x (1-p)^{(n-x)}$
- $E(X) = np$
- $Var(X) = np(1-p)$

Commonly used to model the number of heads after tossing *n* coins.

Another example is the number of people that support a candidate in a poll of *n* people.

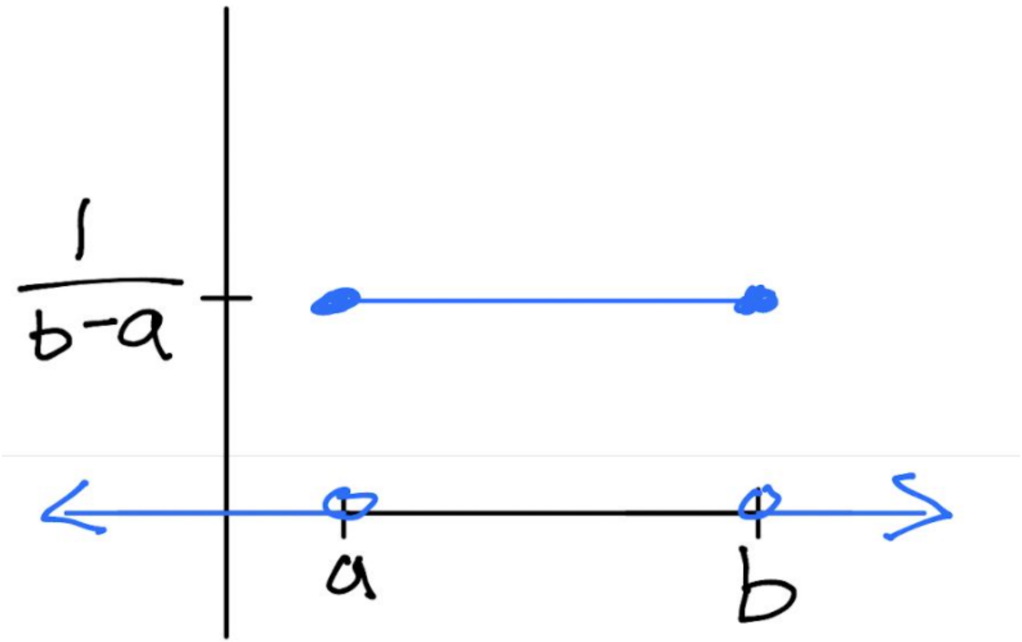
ő Common Probability Distributions

<u>Uniform</u>

If X is a uniform random variable over the interval [a, b] then:

- $f_X(x) = 1/(b-a)$, for $a \le x \le b$,
- $E(X) = (a + b)/2$ and
- $Var(X) = (b - a)^2/12$

PDF seen on the right.



ő Common Probability Distributions

Normal or Gaussian

If X is a normal random variable with mean μ and variance σ^2, then:

- We say that $X \sim N(\mu, \sigma^2)$,
- $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\left(\frac{1}{2\sigma^2}\right)(x-\mu)^2\right],$
- $E(X) = \mu$ and
- $Var(X) = \sigma^2$.

PDF seen on the right, called the "bell curve"

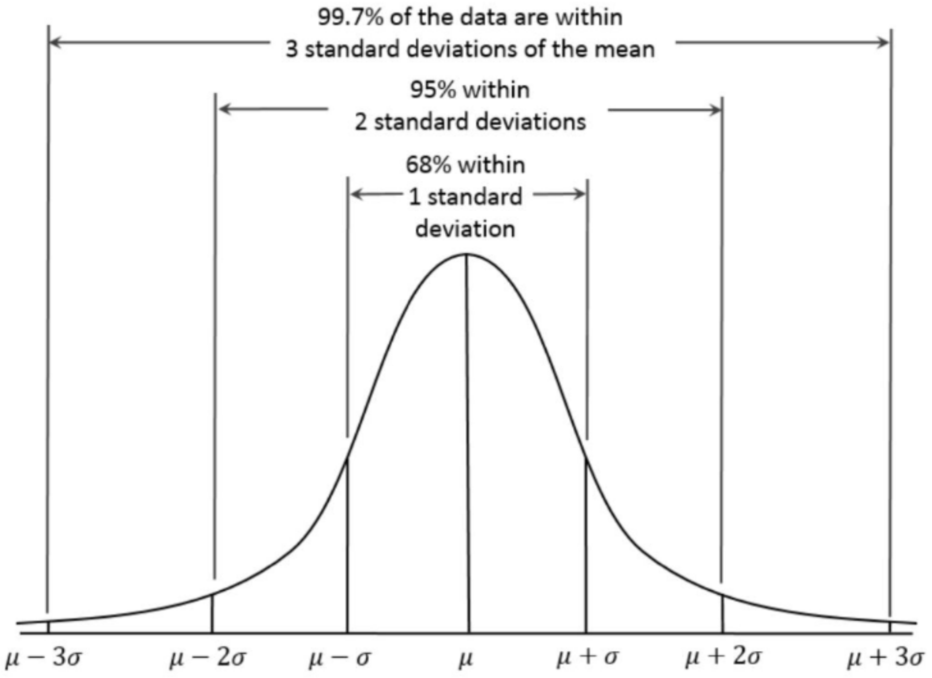


Photo Credit: wikipedia.com

Note: standard deviation is the square root of the variance.

ő Common Probability Distributions

A Standard Normal

Z is a *standard normal* random variable if: $Z \sim N(0, 1)$.

Any normal random variable, $X \sim N(\mu, \sigma^2)$, can be transformed into a standard normal variable using the transformation: $Z = (X - \mu)/\sigma$.

ő The Law of Large Numbers

Let $X_1, X_2, \ldots, X_n$ denote a sequence of independent identically distributed random variables with mean μ, then the law of large numbers says that:

$(X_1 + X_2 + \ldots + X_n)/n \to \mu$ as $n \to \infty$.

ő The Central Limit Theorem

Let $X_1, X_2, \ldots, X_n$ denote a sequence of independent identically distributed random variables with mean μ and finite variance σ^2.

Let $\overline{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$, then the central limit theorem states that:

as $n \to \infty$ the distribution of $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ approaches that of a standard normal.

This theorem was a massive achievement and is the foundation for much of modern statistics.