# Probability Theory and Exact Inference in Bayesian Networks

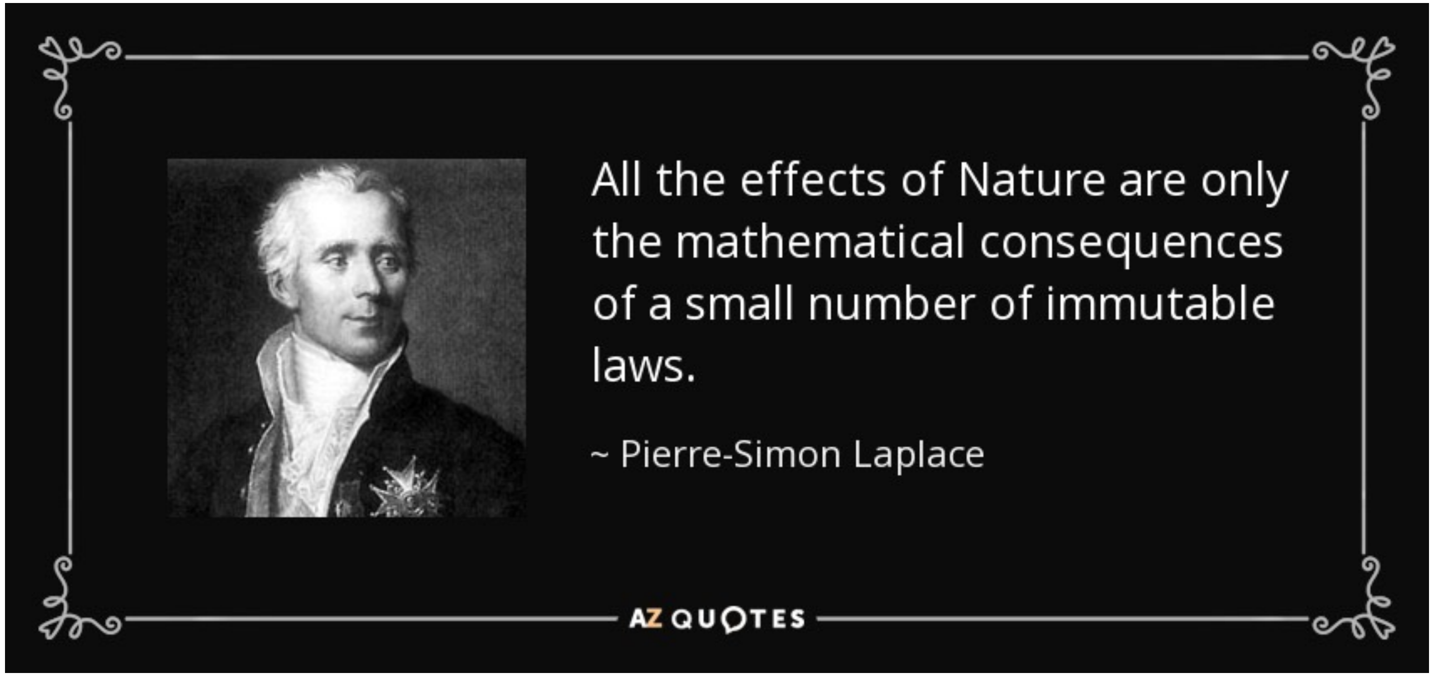
By: Aryeh Freud

# Finite Probability

If $S$ is a finite nonempty sample space of equally likely outcomes, and $E$ is an event, that is, a subset of S, then the probability of $E$ is $p(E) = \frac{|E|}{|S|}$

According to Pierre-Simon Laplace's definition, the probability of an event happening is between 0 and 1.

# Pierre-Simon Laplace



# Assigning Probabilities

- Suppose that *S* is a set with *n* elements. The *uniform distribution* assigns the probability 1/*n* to each element of *S*.

- The probability of an event as the sum of the probabilities of the event is

$$p(E) = \sum_{s \in E} p(s)$$

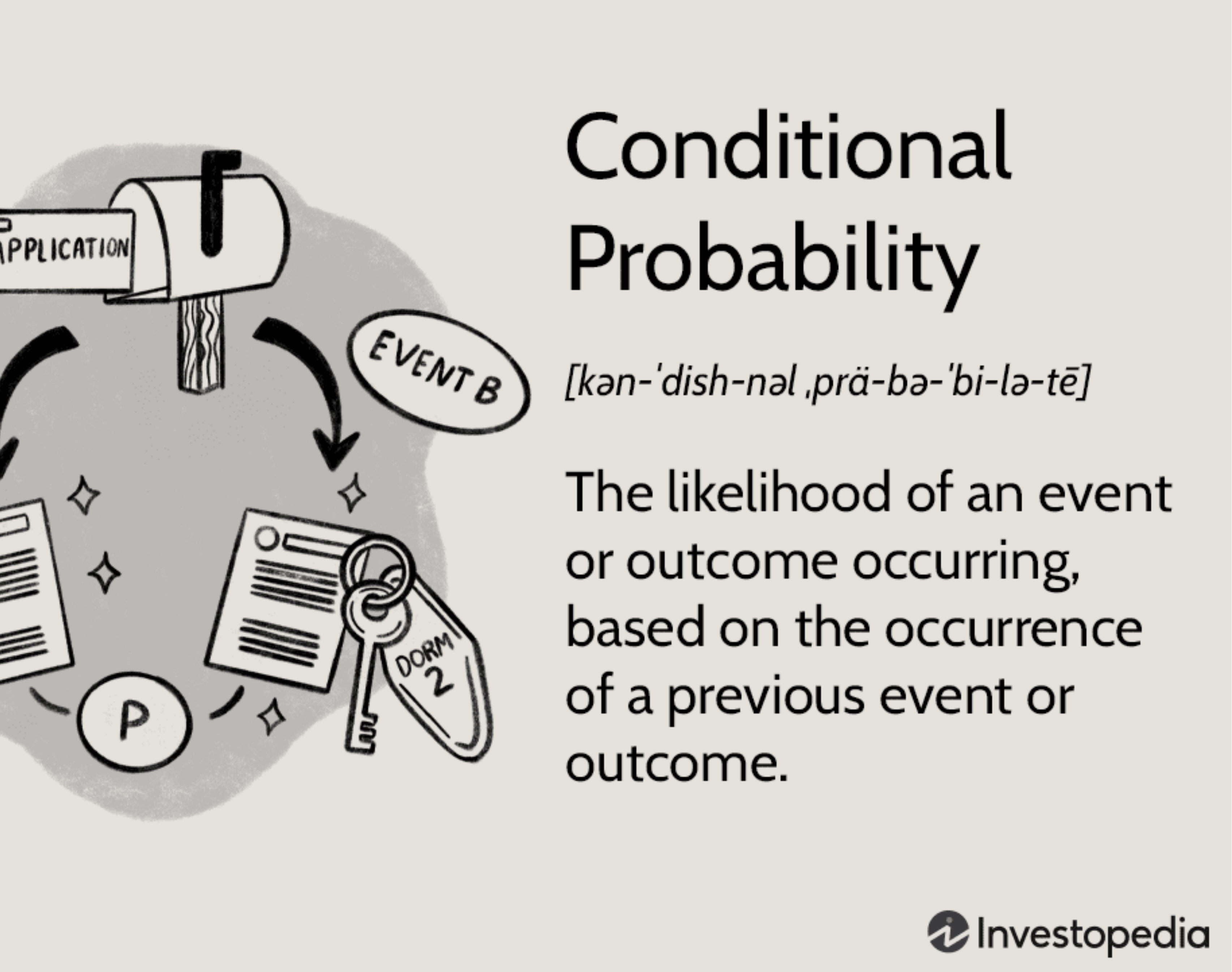
- Because |E| = m and |S| = n, it follows that $p(E) = \frac{m}{n} = \frac{|E|}{|S|}$ which is Laplace’s definition of the probability of event E.

# Probabilities of Complements and Unions of Events

- Let *E* be an event in a sample space *S*. The probability of the event is $\overline{E} = S - E$.

- The complementary event of *E*, is given by $p(\overline{E}) = 1 - p(E)$, where $\overline{E}$ is the complementary event of the event E. This holds true when looking at the equation $\sum_{s \in S} p(s) = 1 = p(E) + p(\overline{E})$ because the sum of probabilities of the n possible outcomes is 1 and each outcome is either in E or in $\overline{E}$ but not in both.

- The generalization of this is in the following formula for the probability of the union of pairwise disjoint events: $p(\bigcup_i E_i) = \sum_i p(E_i)$.




- Let $E$ and $F$ be events with $p(F) > 0$. The conditional probability of E given F, denoted by $p(E \mid F)$ is defined as:
  $$p(E \mid F) = \frac{p(E \cap F)}{p(F)}$$

# Independence

- Let E and F be events. E and F are independent if and only if $p(E \cap F) = p(E)p(F)$.

- We know that $p(E \mid F) = \frac{p(E \cap F)}{p(F)}$, so asking if $p(E \mid F) = p(E)$ is the same as $p(E \cap F) = p(E)p(F)$, which is what led us to our definition of independence.

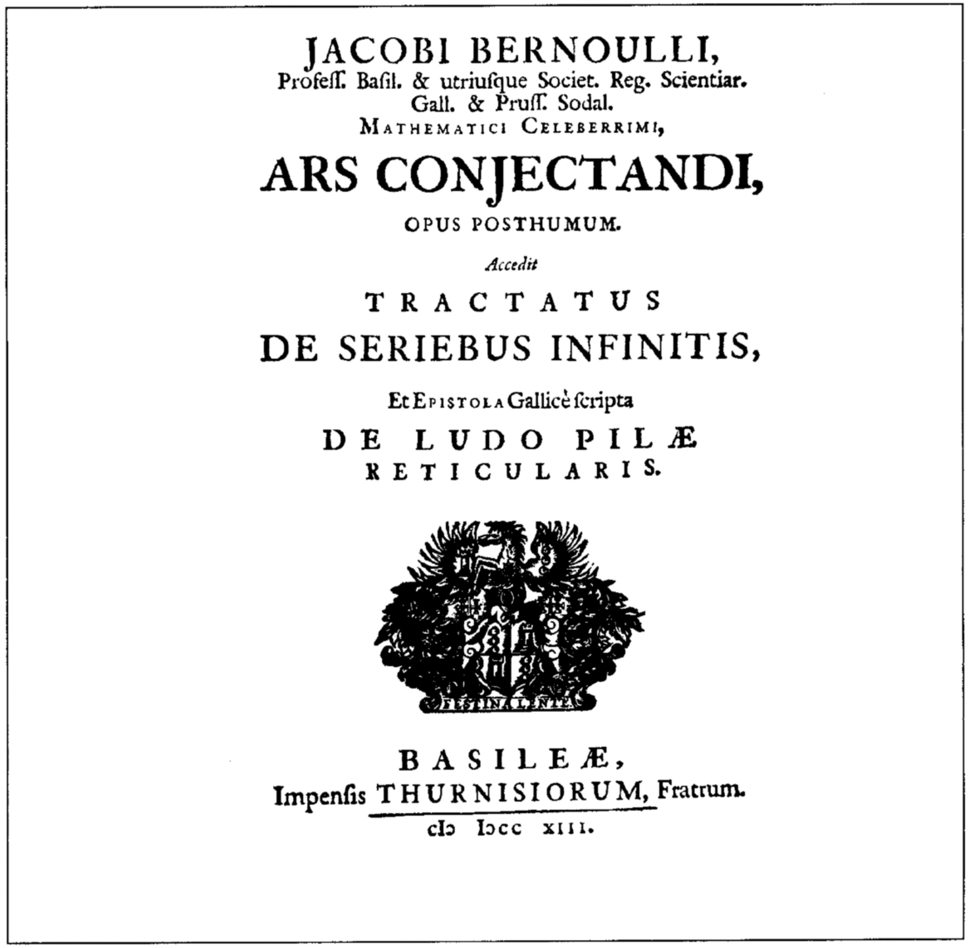
# Bernoulli Trials

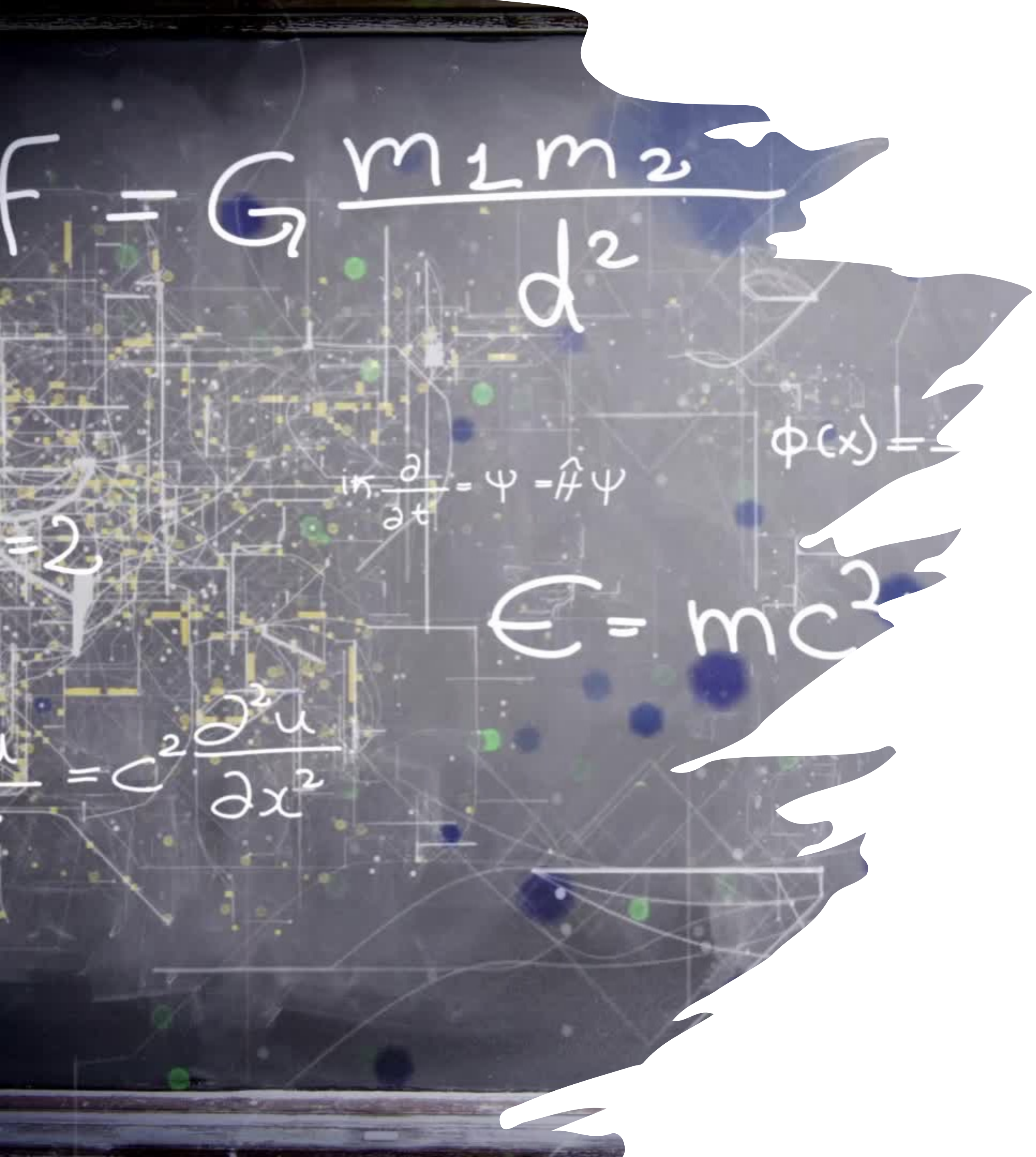
- The probability of exactly *k* successes in *n* independent Bernoulli trials, with probability of success *p* and probability of failure *q* = 1 – *p* is the following equation: $C(n,k)p^k q^{n-k}$.

# Background on Bernoulli

The law of large numbers:

If $\epsilon > 0$, as $n$ becomes arbitrarily large the probability approaches 1 that the fraction of times an event $E$ occurs during $n$ trials is within $\epsilon$ of $p(E)$.



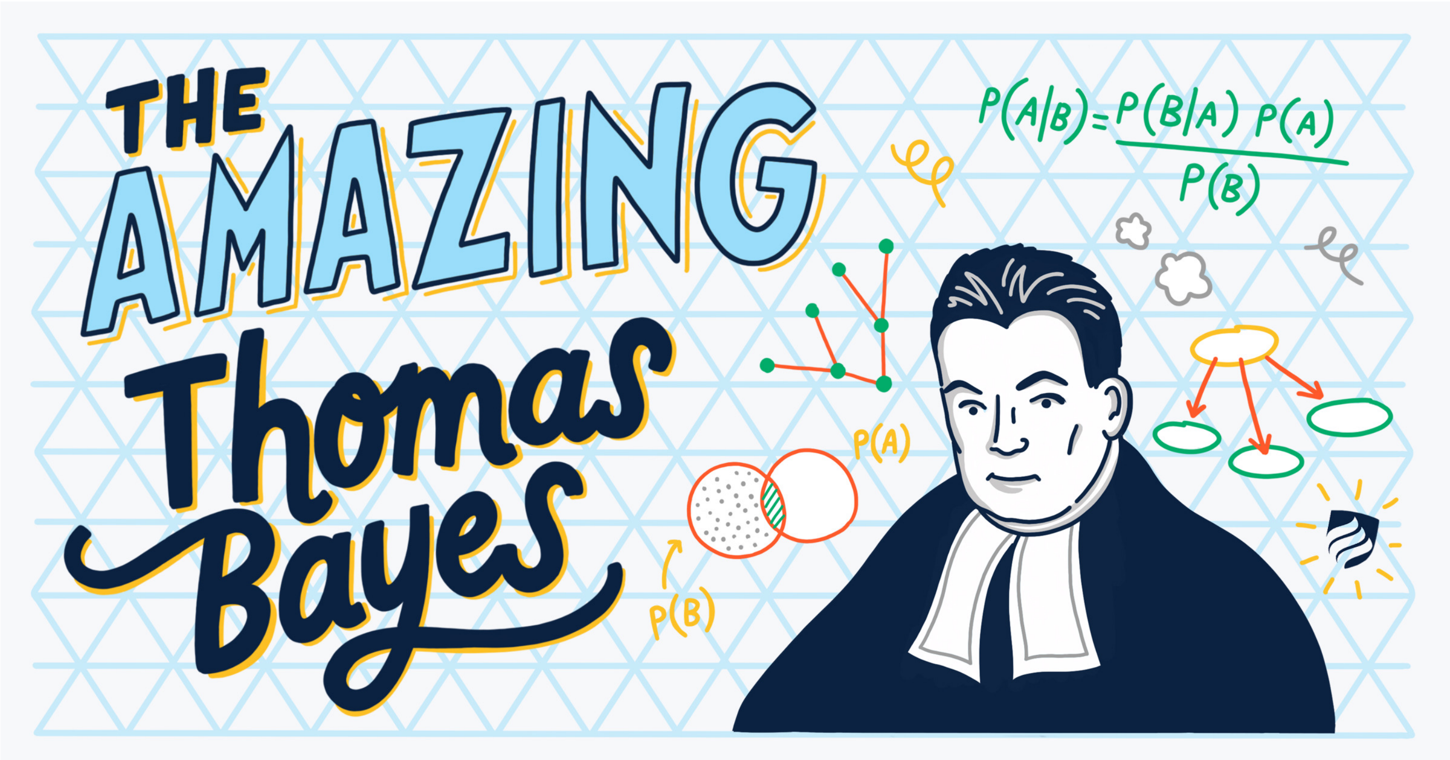


# Random Variables

- The *distribution* of a random variable *X* on a sample space *S* is the set of pairs ($r$, $p(X = r)$) for all $r \in X(S)$, where $p(X = r)$ is the probability that *X* takes the value *r*.

# Bayes' Theorem

- Suppose that $E$ and $F$ are events from a sample space $S$ such that $p(E) \neq 0$ and $p(F) \neq 0$. Then we have,
  $p(F \mid E) = \dfrac{p(E \mid F)p(F)}{p(E \mid F)p(F) + p(E \mid \overline{F})p(\overline{F})}$



# The Enumeration Algorithm - Ask Function

**function** ENUMERATION-ASK($X$, **e**, $bn$) **returns** a distribution over $X$
  **inputs**: $X$, the query variable
    **e**, observed values for some set of variables **E**
    $bn$, a Bayes net
  **Q** ← a distribution over $X$, where **Q**($x_i$) is P($X$=$x_i$)
  **for each** value $x_i$ that $X$ can have **do**
    **Q**($x_i$) ← ENUMERATE-ALL($bn$.VARS, $\mathbf{e}_{x_i}$), where $\mathbf{e}_{x_i}$ is the evidence **e** plus the assignment $X$=$x_i$
  **return** NORMALIZE(**Q**)

# The Enumeration Algorithm – All Function

**function** ENUMERATE-ALL(*vars*, **e**) **returns** a probability (a real number in [0,1])
  **inputs**: *vars*, a list of all the variables
    **e**, observed values for some set of variables **E**
  **if** EMPTY(*vars*) **then return** 1.0
  $Y \leftarrow$ FIRST(*vars*)
  **if** $Y$ is assigned a value (call it $y$) in **e then**
    **return** P($Y$=$y$ | values assigned to $Y$'s parents in **e**) × ENUMERATE-ALL(REST(*vars*), **e**)
  **else**
    **return** $\sum_{y_i}$ [P($Y$=$y_i$ | values assigned to $Y$'s parents in **e**) × ENUMERATE-ALL(REST(*vars*), $\mathbf{e}_{y_i}$)],
      where $\mathbf{e}_{y_i}$ is the evidence **e** plus the assignment $Y$=$y_i$

# The Variable-Elimination Algorithm

```
function ELIMINATION-ASK(X, e, bn) returns a distribution over X
  inputs: X, the query variable
          e, observed values for some set of variables E
          bn, a Bayes net
  factors ← [for each variable v in bn.VARS, the CPT for v given e]
  for each var in bn.vars if var is not in e and var is not X do
      relevant-factors ← [all factors that contain var]
      factors.remove(relevant-factors)
      factors.append(SUM-OUT(var, POINTWISE-PRODUCT(relevant-factors)))
  return NORMALIZE(POINTWISE-PRODUCT(factors))
```

# Sources


Discrete Mathematics and Its Applications, 8th Edition By Kenneth H. Rosen

Exact Inference in Bayes Nets – Pseudocode from MIT's Artificial Intelligence course (6.034)

Bayesian Network – www.wikipedia.com