Lecture 3: Probability Theory

COMP 411, Fall 2021

Victoria Manfredi

WESLEYAN UNIVERSITY

Acknowledgements: These slides are based primarily on material in the book, Introduction to Probability Models by Sheldon M. Ross, 11th edition, 2014, as well as on slides created by Vivek Srikumar (Utah) and Dan Roth (Penn), andcourses at Stanford (http://cs229.stanford.edu/) and UC Berkeley (http://ai.berkeley.edu/)

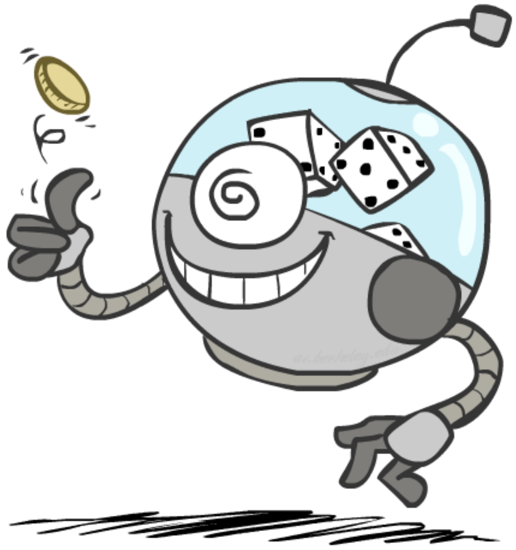
Today's Topics

Probability theory

- Overview

Definitions

- Sample space and events
- Probabilities defined on events
- Conditional probabilities
- Independent events
- Bayes' Formula



vumanfredi@wesleyan.edu

Probability Theory

OVERVIEW

Probability: who needs it?

Learning without probability is possible

- version spaces
- explanation-based
- but rare …

Learning almost always involves

- noise in data (training, testing)
- prediction about the future

Learning systems that don't use probability tend to be very, very brittle

Probabilities

Natural way to represent uncertainty

There exist intuitive notions about probabilities

- many notions are wrong are inconsistent
- many people don't get what probabilities mean

Therefore have FORMAL description that is consistent and useful

- overall framework is understood
- fine details of "meaning" still debated

Example

Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice and also participated in antinuclear demonstrations.

Rank the following by probability (1=most probable; 8 = least probable)

- Linda is a teacher in elementary school
- Linda works in a bookstore and takes yoga classes
- Linda is an active feminist
- Linda is a psychiatric social worker
- Linda is a member of the League of Women Voters
- Linda is a bank teller
- Linda is an insurance salesperson
- Linda is a bank teller and is an active feminist

Understanding probabilities

Probabilities have dual meanings

- relative frequencies (frequentist view)
- degree of belief (Bayesian view)

Neither is entirely satisfying

- no two events are truly the same
- statements should be grounded reality in some way

Probability as Relative Frequency

What is probability of event E?

Over long sequence of experiments, ratio of

- n_E: number of times E occurs in sequence
- n_{trial}: total number of trials

Estimate

$$P(E) \approx \frac{n_E}{n_{trial}}$$

As $n_{trial} \to \infty$, ratio approaches true probability

Examples

P(swimmer succeeds)

- Swimmer S
  - Tries 100 times to swim 50 feet in 15 secs
  - Succeeds 20 occasions
- Estimate: probability that
  - S can swim 50 feet in 15 sec
  - P(S can swim 50 feet in 15 sec) ≈ 20 /100 = 0.2

For probability to be meaningful, must clearly define

- Experiments, Sample space, Events

Q: What is the probability of a nuclear accident?

Interpretations of probability: a can of worms!

Frequentists

- $P(\alpha)$ = the frequency of α in the limit
- Many arguments against this interpretation: What is the frequency of the event “it will rain tomorrow”? “nuclear war tomorrow”?

Subjective interpretation

- $P(\alpha)$ = my degree of belief that α will happen
- Where “degree of belief’’ means: if I say $P(\alpha) = 0.8$, then I am willing to bet!!!

For this class …

- We (mostly) don’t care what camp you are in

Probability Theory

DEFINITIONS

Set theory notation

Subset relation

$$A \subset B \iff x \in A \implies x \in B$$

Equality

$$A = B \iff A \subset B \text{ and } B \subset A$$

Union

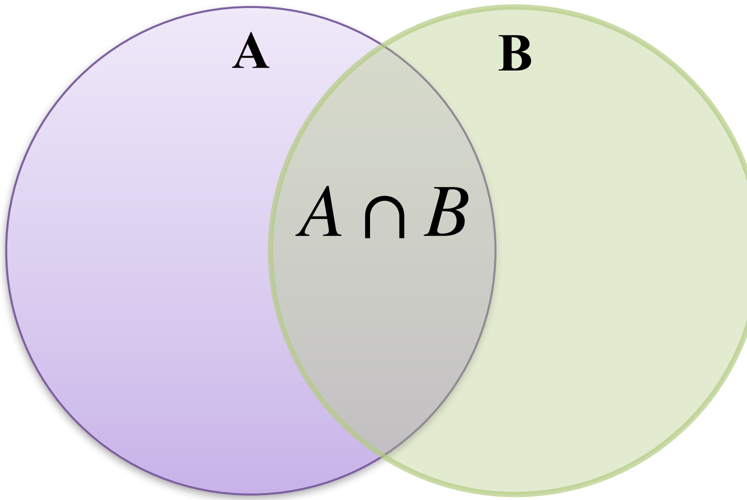
$$A \cup B = \{x : x \in A \text{ or } x \in B\}$$

Intersection

$$A \cap B = \{x : x \in A \text{ and } x \in B\}$$

Complement

$$A^c = \{x : x \notin A\}$$



Properties of set operations

Commutativity

$$A \cup B = B \cup A$$

$$A \cap B = B \cap A$$

Associativity

$$A \cup (B \cup C) = (A \cup B) \cup C$$

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

Distributivity

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

DeMorgans Laws

$$(A \cup B)^c = A^c \cap B^c$$

$$(A \cap B)^c = A^c \cup B^c$$

Sample space

The set of all possible outcomes of an experiment is called the sample space of the experiment and is denoted by S

Examples

Experiment 1: flipping a coin

- $S = \{H, T\}$

Experiment 2: a single roll of an ordinary die

- $S = \{1,2,3,4,5,6\}$

Experiment 3: flipping two coins

- $S = \{(H,H),(H,T),(T,H),(T,T)\}$

Event

Any subset E of the sample space S is known as an event. Event E occurs when the outcome of the experiment lies in E

Examples

Experiment 1: flipping a coin

- $S = \{H, T\}$
- $E = \{H\}$: event that a head appears on the flip of the coin
- $E = \{T\}$: event that a tail appears on the flip of the coin

Experiment 2: a single roll of an ordinary die

- $S = \{1,2,3,4,5,6\}$
- $E = \{1\}$: event that one appears on roll of the die
- $E = \{2,4,6\}$: event that an even number appears on roll of the die

Union of events

For any two events E and F of a sample space S we define the new event $E \cup F$ to consist of all outcomes that are either in E or F or in both E and F. That is, the event E ∪ F will occur if either E or F occurs.

Examples

Experiment 1: flipping a coin

- $S = \{H, T\}$
- $E = \{H\}, F = \{T\}$
- $E \cup F = \{H, T\}$

Experiment 2: a single roll of an ordinary die

- $S = \{1,2,3,4,5,6\}$
- $E = \{1,3,5\}, F = \{1,2,3\}$
- $E \cup F = \{1,2,3,5\}$

Intersection of events

For any two events E and F of a sample space S we define the new event $E \cap F$ to consist of all outcomes which are both in E and F. That is, the event $E \cap F$ will occur only if both E and F occur

Examples

Experiment 1: flipping a coin

- $S = \{H, T\}$
- $E = \{H\}, F = \{T\}$
- $E \cap F = \varnothing$: E and F are mutually exclusive

Experiment 2: a single roll of an ordinary die

- $S = \{1,2,3,4,5,6\}$
- $E = \{1,3,5\}, F = \{1,2,3\}$
- $E \cap F = \{1,3\}$

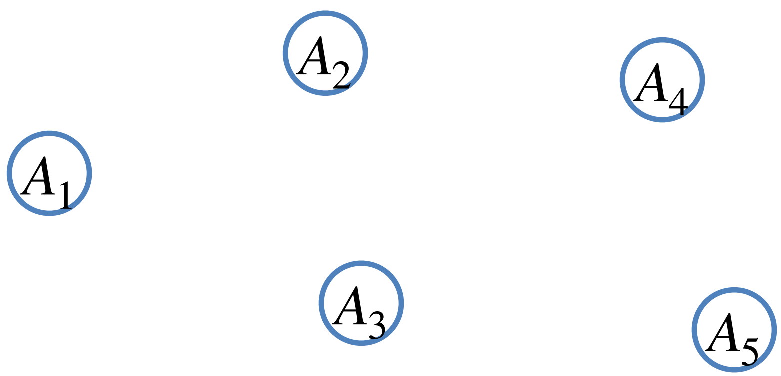
Unions and intersections of > 2 sets

Two events A and B are disjoint (or mutually exclusive if

$$A \cap B = \varnothing$$

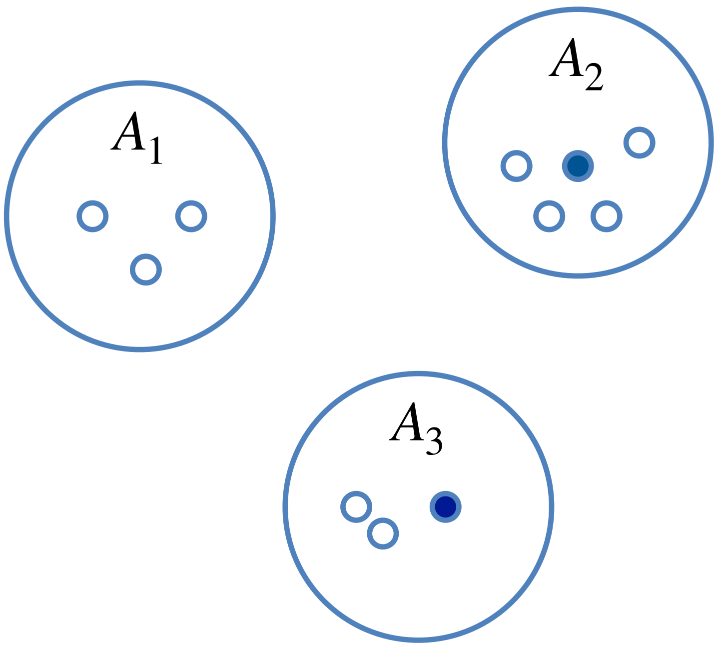
A sequence of events $A_1, A_2, A_3, \ldots$ are pairwise disjoint if

$A_i \cap A_j = \varnothing$ for all $i \neq j$



Partition

If $A_1, A_2, \ldots$ are pairwise disjoint and $\bigcup_{i=1}^{\infty} A_i = \Omega$, then the collection $A_1, A_2, \ldots$ forms a partition of the sample space Ω



Probability Theory

PROBABILITIES DEFINED ON EVENTS

Komolgorov's axioms

Probabilities defined on events. Consider an experiment whose sample space is S. For each event E of the sample space S, we assume that a number $P(E)$ is defined and satisfies the following conditions:

1. $0 \leq P(E) \leq 1$
2. $P(S) = 1$
3. For any sequence of events $E_1, E_2, \ldots$ that are mutually exclusive, that is events for which $E_n \cap E_m = \varnothing$ when $n \neq m$, then
$$P\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} P(E_n)$$

We refer to $P(E)$ as the probability of the event E. Intuitively, if our experiment is repeated over and over again then (with probability 1) the proportion of time that event E occurs will just be $P(E)$

Example: flipping a coin

Sample space: $S = \{H, T\}$

Events: $E = \{H\}, F = \{T\}$

If head is equally likely to appear as tail

$$P(\{H\}) = P(\{T\}) = \frac{1}{2}$$

if biased coin and head twice as likely to appear as tail

$$P(\{H\}) = \frac{2}{3}$$

$$P(\{T\}) = \frac{1}{3}$$

Example: tossing a die

Sample space: $S = \{1,2,3,4,5,6\}$

Events: $E = \{1,3,5\}$, $F = \{1,2,3\}$, …

Suppose all 6 numbers are equally likely to appear

$$P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = P(\{6\}) = \frac{1}{6}$$

From Condition 3, probability of getting an even number

$$P(\{2,4,6\}) = P(\{2\}) + P(\{4\}) = P(\{6\}) = \frac{1}{2}$$

Probability of mutually exclusive events

Since the events E and E^c are always mutually exclusive and since $E \cup E^c = S$, then we have by Conditions 2 and 3 that

$$\boxed{\begin{gathered} 1 = P(S) = P(E \cup E^c) = P(E) + P(E^c) \\ \text{or} \\ P(E^c) = 1 - P(E). \end{gathered}}$$

Probability of union of events

Consider $P(E) + P(F)$, which is the probability of all outcomes in E plus the probability of all points in F. Since any outcome that is in both E and F will be counted twice in $P(E) + P(F)$, and only once in $P(E \cup F)$, we must have

$$P(E) + P(F) = P(E \cup F) + P(EF)$$

or equivalently

$$P(E \cup F) = P(E) + P(F) - P(EF)$$

When E and F are mutually exclusive (that is, when $EF = \varnothing$), then

$$P(E \cup F) = P(E) + P(F) - P(\varnothing) = P(E) + P(F)$$

a result which also follows from condition 3. (Why is P(Ø) = 0?)

Example

Suppose that we toss two coins, and suppose that we assume that each of the four outcomes in the sample space $S = \{(H,H),(H,T),(T,H),(T,T)\}$ is equally likely and hence has probability $\frac{1}{4}$.

Let $E = \{(H,H),(H,T)\}$ and $F = \{(H,H),(T,H)\}$. That is, E is the event that the first coin falls heads, and F is the event that the second coin falls heads.

The probability that either the first or the second coin falls heads, is given by

$$P(E \cup F) = P(E) + P(F) - P(E \cap F) = \frac{1}{2} + \frac{1}{2} - P(\{H,H\}) = 1 - \frac{1}{4} = \frac{3}{4}$$

This probability could have been computed directly since

$$P(E \cup F) = P(\{(H,H),(H,T),(T,H)\}) = \frac{3}{4}$$

Probability of union of events

Probability that any one of the three events E or F or G occurs is given by

$P(E \cup F \cup G) = P((E \cup F) \cup G)$ which equals $P(E \cup F) + P(G) - P((E \cup F) \cap G)$

We an show the events $(E \cup F)G$ and $(E \cap G) \cup (F \cap G)$ are equivalent, and the preceding equals

$$
\begin{aligned}
&P(E \cup F \cup G) \\
&= P(E) + P(F) - P(E \cap F) + P(G) - P((E \cap G) \cup (F \cap G)) \\
&= P(E) + P(F) - P(E \cap F) + P(G) - P(E \cap G) - P(F \cap G) + P(E \cap G \cap F \cap G) \\
&= P(E) + P(F) + P(G) - P(E \cap F) - P(E \cap G) - P(F \cap G) + P(E \cap F \cap G)
\end{aligned}
$$

It can be shown by induction that, for any n events $E_1, E_2, E_3, \ldots, E_n$

$$P(E_1 \cup E_2 \cup \cdots \cup E_n) = \sum_i P(E_i) - \sum_{i<j} P(E_i \cap E_j) + \sum_{i<j<k} P(E_i \cap E_j \cap E_k)$$

Inclusion-Exclusion Identity

$$- \sum_{i<j<k<l} P(E_i \cap E_j \cap E_k \cap E_l) + \cdots + (-1)^{n+1} P(E_1 \cap E_2 \cdots E_n)$$

Aside

Probability of A and/intersect B

- $P(A \cap B),\ P(A \wedge B), P(A, B), P(AB)$

Probability of A or/union B

- $P(A \cup B), P(A \vee B)$

Probability Theory

CONDITIONAL PROBABILITY

Example

Suppose we toss two fair dice. There are 36 possible outcomes, each occurs with probability $\frac{1}{36}$. Define the events, $E = \{\text{Sum of the dice is } 6\}$, and $F = \{\text{First die is a } 4\}$. What is the conditional probability that E occurs given that F has occurred, $P(E \mid F)$?

Solution:

There are six possible outcomes of experiment:

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

Conditional probabilities:

$$P((4,1) \mid F) = \frac{1}{6},\ P((4,2) \mid F) = \frac{1}{6},\ P((4,3) \mid F) = \frac{1}{6},$$
$$P((4,4) \mid F) = \frac{1}{6},\ P((4,5) \mid F) = \frac{1}{6},\ P((4,6) \mid F) = \frac{1}{6}$$

The (conditional) probability of the other 30 points in the sample space is 0. Hence

$P(E \mid F) = \frac{1}{6}$

Conditional probabilities

A general formula for $P(E \mid F)$ that is valid for all events E and F

$$P(E \mid F) = \frac{P(E \cap F)}{P(F)}, \qquad \text{when } P(F) > 0$$

If F occurs, then in order for E to occur it is necessary for the actual occurrence to be a point in both E and in F, that is, it must be in $E \cap F$. Because we know that F, has occurred, it follows that F becomes our new sample space and hence the probability that the event $E \cap F$ occurs will equal the probability of $E \cap F$ relative to the probability of F

Intuitively: Conditional probabilities capture beliefs taking evidence into consideration. Conditional distributions are probability distributions over some events given fixed values of other events

Example

Suppose cards numbered 1 through 10 are placed in a hat, mixed up, and then one of the cards is drawn. If we are told that the number on the drawn card is at least 5, then what is the conditional probability that it is 10?

Solution: Let E denote the event that the number of the drawn card is 10, and let F be the event that it is at least 5. The desired probability is $P(E \mid F)$.

$$P(E \mid F) = \frac{P(E \cap F)}{P(F)}$$

However, $E \cap F = E$ since the number of the card will be both 10 and at least 5 if and only if it is number 10. Hence,

$$P(E \mid F) = \frac{\frac{1}{10}}{\frac{6}{10}} = \frac{1}{6}$$

Example

A family has two children. What is the conditional probability that both are boys given that at least one of them is a boy? Assume that the sample space S is given by $S = \{(b,b),(b,g),(g,b),(g,g)\}$, and all outcomes are equally likely. ((b,g) means, for instance, that the older child is a boy and the younger child a girl.)

Solution: Let B be the event that both children are boys, and A the event that at least one of them is a boy. Then the desired probability is given by

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

$$= \frac{\{(b,b)\}}{\{(b,b),(b,g),(g,b)\}}$$

$$= \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}$$

Example

Bev can either take a course in computers or in chemistry. If Bev takes the computer course, then she will receive an A grade with probability $\frac{1}{2}$; if she takes the chemistry course then she will receive an A grade with probability $\frac{1}{3}$. Bev decides to base her decision on the flip of a fair coin. What is the probability that Bev will get an A in chemistry?

Solution: Let C be the event that Bev takes chemistry and A be the event that she receives an A in whatever course she takes. Then the desired probability is $P(A \cap C)$

$$P(A \cap C) = P(C)P(A \mid C)$$
$$= \frac{1}{2}\frac{1}{3} = \frac{1}{6}$$

Example

Suppose an urn contains 7 black balls and 5 white balls. We draw 2 balls from the urn without replacement. Assuming that each ball in the urn is equally likely to be drawn, what is the probability that both drawn balls are black?

Solution: Let F and E denote, respectively, the events that the 1st and 2nd balls drawn are black.

Given that the first ball selected is black, there are 6 remaining black balls and 5 white balls, and so $P(E|F) = \frac{6}{11}$.

As $P(F)$ is clearly $\frac{7}{12}$, our desired probability is:

$$P(E \cap F) = P(F)P(E|F)$$
$$= \frac{7}{12}\frac{6}{11} = \frac{42}{132}$$

Example

Suppose that each of 3 people at a party throws their hats into the center of the room. The hats are first mixed up and then each person randomly selects a hat. What is the probability that none of the 3 people selects their own hat?

Solution: Let $E_i, i = 1,2,3,$ be the event that the ith person selects their own hat. Then:

$$P(E_i) = \frac{1}{3}, \quad i = 1,2,3$$

(People are equally likely to select any of the 3 hats)

$$P(E_i \cap E_j) = P(E_i)P(E_j \mid E_i) = \frac{1}{3}\frac{1}{2} = \frac{1}{6}, \qquad i \neq j$$

(Given that i selects their own hat, there remain two hats that j may select, and one of these two is j's hat)

$$P(E_1 \cap E_2 \cap E_3) = P(E_1 \cap E_2)P(E_3 \mid E_1 \cap E_2) = \frac{1}{6} \cdot 1 = \frac{1}{6}$$

(Given that the first 2 people get their own hats, the third person must also get their own hat, since no other hats left)

Example cont'd

Suppose that each of 3 people at a party throws their hats into the center of the room. The hats are first mixed up and then each person randomly selects a hat. What is the probability that none of the 3 people selects their own hat?

Solution: Using the inclusion-exclusion principle we first compute probability that at least one person selects their own hat:

$$P(E_1 \cup E_2 \cup E_3) = P(E_1) + P(E_2) + P(E_3) - P(E_1 \cap E_2) - P(E_1 \cap E_3) - P(E_2 \cap E_3) + P(E_1 \cap E_2 \cap E_3)$$
$$= 1 - \frac{1}{2} + \frac{1}{6} = \frac{2}{3}$$

Then the probability that none of the people select there own hat is $1 - \frac{2}{3} = \frac{1}{3}$

Probability Theory

INDEPENDENT EVENTS

Independent events

Two events E and F are said to be independent if

$$\boxed{P(E \cap F) = P(E)P(F)}$$

From $P(E \mid F) = \dfrac{P(E \cap F)}{P(F)}$, this implies that E and F are independent if $P(E \mid F) = P(E)$, which also implies that $P(F \mid E) = P(F)$.

That is, E and F are independent if knowledge that F has occurred does not affect the probability that E occurs. That is, the occurrence of E is independent of whether or not F occurs.

Two events E and F that are not independent are said to be dependent.

Example

Suppose we toss two fair dice.

E_1 : event that the sum of the dice is 6

F: event that the first die equals 4.

$$P(E_1 \cap F) = P(\{4,2\}) = \frac{1}{36}$$

$$P(E_1)P(F) = \frac{5}{36}\frac{1}{6} = \frac{5}{216}$$

Hence E_1 and F are not independent

Why? Our chance of getting a total of 6 depends on the outcome of the first die and hence E_1 and F cannot be independent.

Example

Suppose we toss two fair dice.

E_2 : event that the sum of the dice is 7

F: event that the first die equals 4.

$$P(E_2 \cap F) = P(\{4,3\}) = \frac{1}{36}$$

Hence E_2 and F are independent.

$$P(E_2)P(F) = \frac{1}{6}\frac{1}{6} = \frac{1}{36}$$

Why is the event that the sum of the dice equals seven independent of the outcome on the first die?

Independence with more than two events

The events $E_1, E_2, \ldots, E_n$ are said to be independent if for every subset $E_{1'}, E_{2'}, \ldots, E_{r'},\ r \leq n$ of these events

$$P(E_{1'} \cap E_{2'} \cap \cdots \cap E_{r'}) = P(E_{1'})P(E_{2'})\cdots P(E_{r'})$$

Intuitively, the events $E_1, E_2, \ldots, E_n$ are independent if knowledge of the occurrence of any of these events has no effect on the probability of any other event.

Example

Pairwise Independent Events That Are Not Independent: Let a ball be drawn from an urn containing four balls, numbered $1,2,3,4$. Define the events $E = \{1,2\}, F = \{1,3\}, G = \{1,4\}$. Recall that an event occurs when the outcome of the experiment lies in E.

If all four outcomes are assumed equally likely, then

$$P(E \cap F) = P(E)P(F) = \frac{1}{2}\frac{1}{2} = \frac{1}{4}$$
$$P(E \cap G) = P(E)P(G) = \frac{1}{2}\frac{1}{2} = \frac{1}{4}$$
$$P(F \cap G) = P(F)P(G) = \frac{1}{2}\frac{1}{2} = \frac{1}{4}$$

However, $\frac{1}{4} = P(E \cap F \cap G) \neq P(E)P(F)P(G)$

Hence, even though the events E, F, G are pairwise independent, they are not jointly independent

Independent trials

Suppose that a sequence of experiments, each of which results in either a “success” or a “failure,” is to be performed.

Let $E_i,\ i \geq 1$, denote the event that the ith experiment results in a success. If, for all $i_1, i_2, \ldots, i_n$,

$$P(E_{i_1}E_{i_2}\ldots E_{i_n}) = \prod_{j=1}^{n} E_{i_j}$$

we say that the sequence of experiments consists of independent trials