CS 215
Data Analysis and Interpretation

Probability Axioms

Suyash P. Awate

Probability Axioms

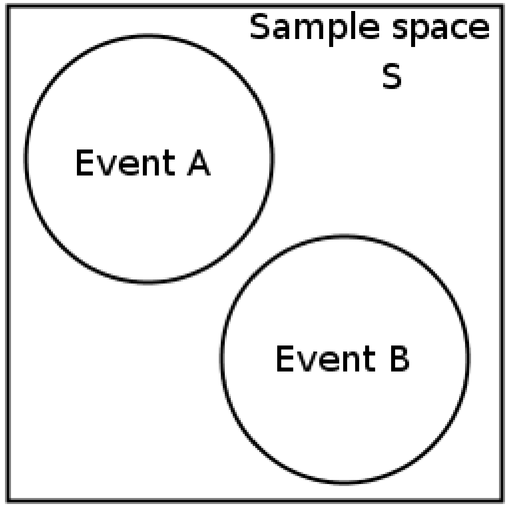
- **Set**: a collection of unique objects
  - Doesn’t care about the ordering of the objects
  - Interesting examples
    - Set of natural numbers, integers, rational numbers, real numbers
    - Set of fruits, people
- **Experiment**: an empirical procedure
- **Random Experiment**: an experiment whose outcome is uncertain
  - Examples
    - Tossing a coin
    - Rolling a die
    - Measuring the lifetime of a laptop battery
    - Measuring the temperature at a specified outdoor location after 50 hours

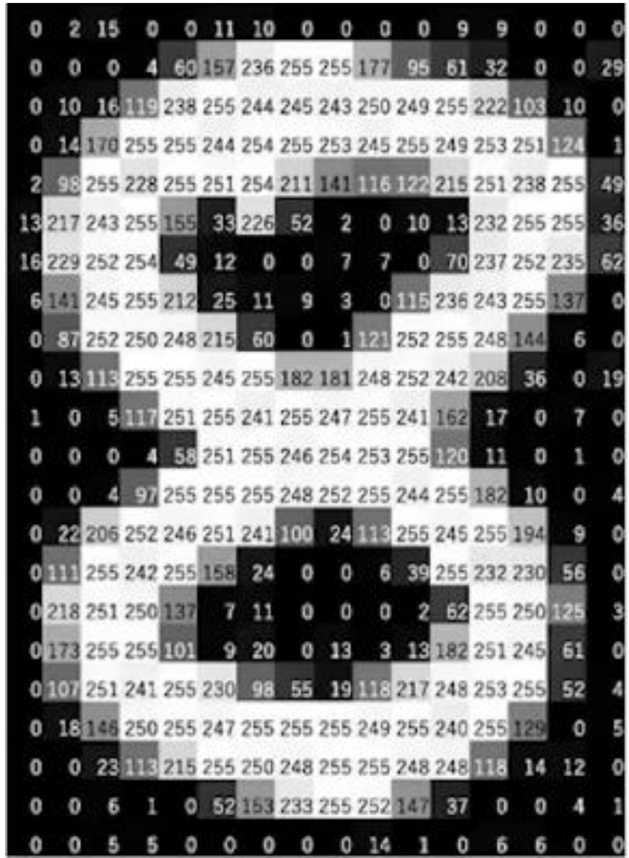
Probability Axioms

- **Sample Space Ω**: Set of all possible outcomes of a random experiment
  - Countably finite sample spaces
    - For coin toss: {H,T}
    - For die roll: {1,2,3,4,5,6}
  - Countably infinite sample spaces
    - For the experiment that counts the number of coin tosses required to get a head
  - Uncountably infinite sample spaces
    - For the experiments that will report the temperature, at some location in the universe, after 50 hours:
      Subset of real numbers in the interval from
      absolute zero (0 Kelvin) to Planck temperature (1.416785 x 10^{32} Kelvin)

Probability Axioms

- **Event**: A subset of the sample space
  - Examples
    - Event A: The die roll produces an even number
      Event B: The die roll produces 3
    - Event X: The temperature is between 298K and 300K (25˚C and 30˚C)
    - Event Y: Within an image,
      intensity at chosen pixel falls between 10 and 20





- **Operations on events**
  - Union, “or”, AUB, A happens or B happens
  - Intersection, “and”, A∩B, A happens and B happens
  - Complement, “negation”, $\overline{A}$, A doesn’t happen
  - Subtraction, A–B, A happens but B doesn’t happen

Probability Axioms

- **Event Space ß**: Space of all possible events (that we want to model)
  - (A subset of) power set of sample space
    - Includes empty set ∅
    - Includes sample space Ω
  - Given a sample space, the event space ß must satisfy several rules (basically, for logical consistency and to deal with sample spaces that are infinite):
    1. Empty set ∅ ∈ ß
    2. Closed under Countable Unions:
       If events $A_1, A_2, A_3, \cdots$ belong to ß, then $A_1 \cup A_2 \cup A_3 \cup \cdots \in$ ß
       - We can only have a union of a countably infinite number of sets here
    3. Closed under Complementation:
       If event A ∈ ß, then event $\overline{A}$ (i.e., Ω – A) ∈ ß
  - A set ß that satisfies these axioms/rules is called a σ–algebra
    - The term “algebra” is used because the theory defines an algebra on sets (union, intersection, difference, ...) analogous to the algebra on numbers

Probability Axioms

- Event Space ß
  - Example:
    - If sample space Ω = { a, b, c, d }, then:
      - One possible design of the event space is: ß = { Φ, {a,b}, {c,d}, {a,b,c,d} }
      - Another possible design of the event space is: ß = power set of Ω
    - (A more useful example) If sample space is the real line, then:
      - Event space ß is the set of subsets of the real line formed by:
        first taking all open intervals,
        then adding in all countable unions, countable intersections, and relative complements,
        then continuing this process until the relevant closure properties are achieved
      - This construction is known as the Borel σ–algebra

Probability Axioms

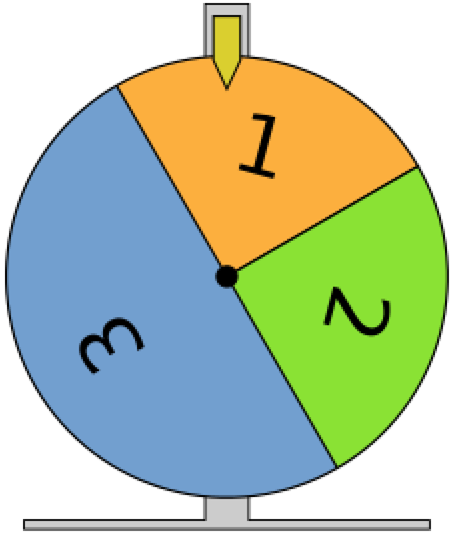
- **Probability Function** (or Probability Measure)
  - Gives a precise notion of the "size" or "volume" of the sets (i.e., events)
  - A probability function/measure on an event space ß is a function P:ß→[0,1] such that:
    - $P(\Phi) = 0$;
    - $P(\Omega) = 1$; this needn't hold for a general (non-probability) measure
    - For pairwise disjoint sets (events) $A_1, A_2, A_3, \ldots$, we have $P(A_1 \cup A_2 \cup A_3 \cup \ldots) = P(A_1) + P(A_2) + P(A_3) + \ldots$
      - We want the size of the union of disjoint sets to be the sum of their individual sizes, even for an countably infinite sequence of disjoint sets
  - In simpler words, the probability function/measure P(.) on a sample space Ω assigns every event A, a number within interval [0,1], such that:
    - $P(\Omega) = 1$, and, when A and B are disjoint events, $P(A \cup B) = P(A) + P(B)$
    - This "number" assigned to event A is the probability/chance that the event occurs

Probability Axioms

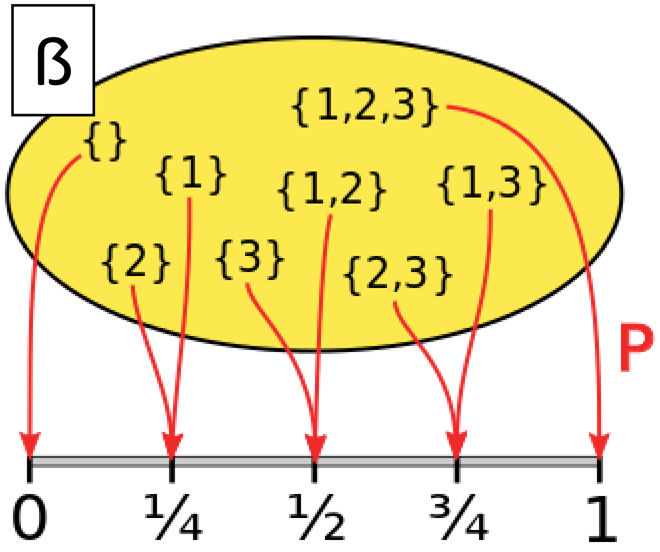
- **Properties of Probability Functions**/Measures
  - Complement of an event A
    - $P(\overline{A}) = 1 - P(A)$
  - Union of two overlapping events
    - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  - Difference of events
    - $P(A - B) = P(A) - P(A \cap B)$

Probability Axioms

- A **probability space** is a **triplet (Ω,ß,P)** where:
  - Ω is the **sample space** (= space of all possible outcomes)
  - ß is the **event space** (= space of all possible events; that we want to model)
  - P(.) is the **probability function**/measure on ß with P(Ω)=1



Ω={1,2,3}



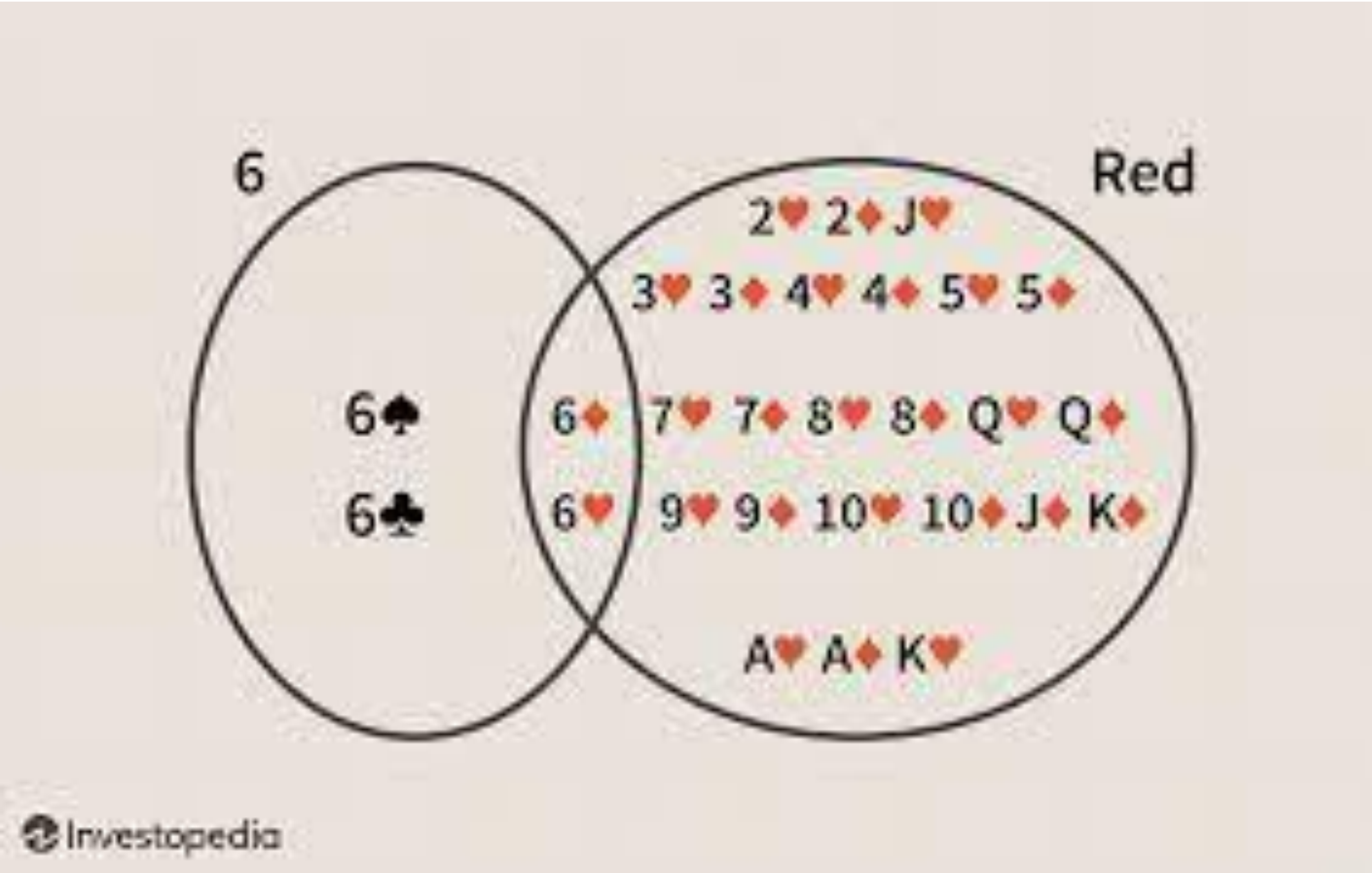
- Examples
  - Toss fair coin: What is Ω ? What is ß ? What is P ?
  - Roll fair die: P({1}) = 1/6. All outcomes equally likely.

Probability Axioms

- Probability space for multiple random experiments
  - **Two Different Kinds of Experiments**
    - If we do two experiments, sequentially, with sample spaces Ω_1 and Ω_2, then we equivalently have a new experiment with (induced) sample space $\Omega' = \Omega_1 \times \Omega_2 = \{ (x,y): x \in \Omega_1, y \in \Omega_2 \}$
    - Example: First, toss a coin. Then, roll a die.
      - What is Ω' ? What is ß' ? What is P' ?
  - **Repeated Experiments**
    - If we do two runs, sequentially, of an experiment with sample space Ω, then we equivalently have a new experiment with (induced) sample space $\Omega' = \Omega \times \Omega = \{ (x,y): x \in \Omega, y \in \Omega \}$
      - Note: element (x,y) is an ordered pair; not a set. So, order matters.
    - Example: Toss two coins.
      - What is Ω' ? What is ß' ? What is P' ?

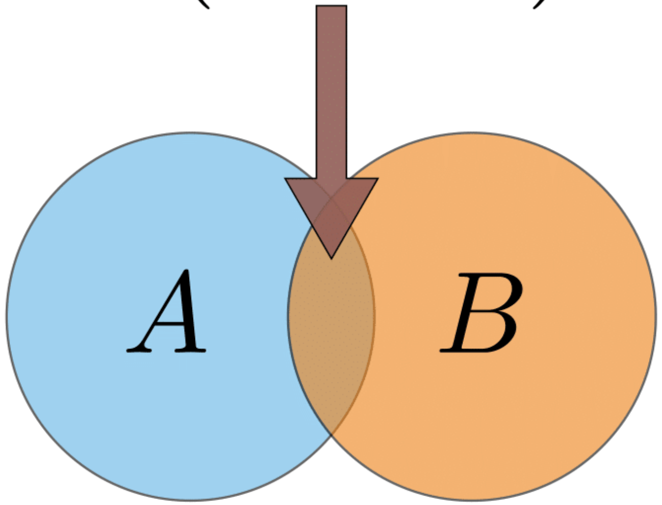
Joint Probability

- P(A and B) := P(A,B) := P(A∩B)
  - ":=" symbol indicates definition



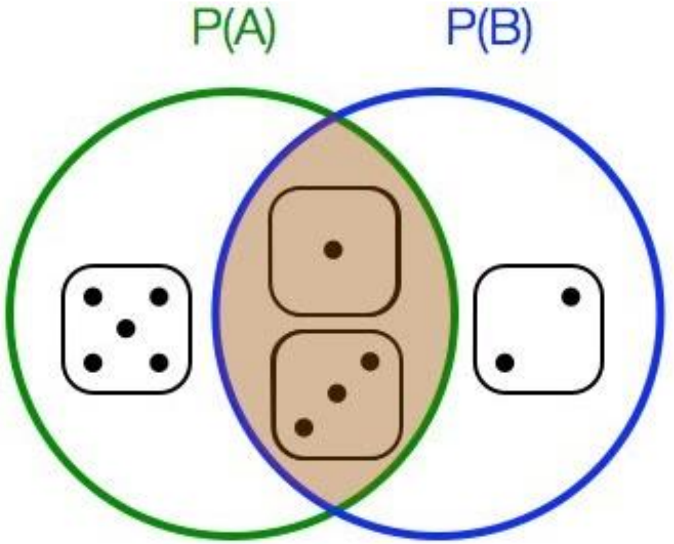
Joint Probability

$$P(A \cap B)$$



Conditional Probability

- Conditional probability of event A, given event B, is:
  $P(A|B) := P(A \cap B) / P(B)$
  - Example: Dice
    - A = {number odd}, B ={number<4}, P(B)=P(A)=1/2, P(B|A)=2/3



  - Example: Image-Generation Process
    - For a pixel,
      first choose an object label
      (e.g., sky, rock),
      then choose pixel color given object label
    - P(pixel takes specific color)
      versus
      P(pixel takes specific color | object label)

Conditional Probability

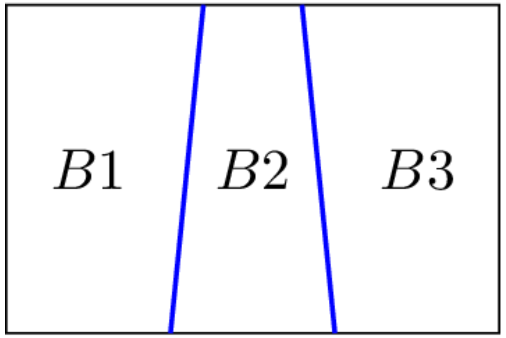
- Conditional probability of event A, given event B, is:
  $P(A|B) := P(A \cap B) / P(B)$
  - Example: Image-Acquisition Process
    - Observed color is a randomly perturbed version of the true color, due to measurement errors
      - Larger ISO setting → larger perturbation magnitude
    - P(pixel takes specific color | true color at that pixel) → perturbation model

Conditional Probability

- Conditional probability of event A, given event B, is:
  $P(A|B) := P(A \cap B) / P(B)$
  - Example: Image-Generation Process
    - P(pixel takes specific color) usually differs from
      P(pixel takes specific color | color at adjacent pixel)
    - Former relates to the image histogram
    - Latter relates to inter-pixel color relationship

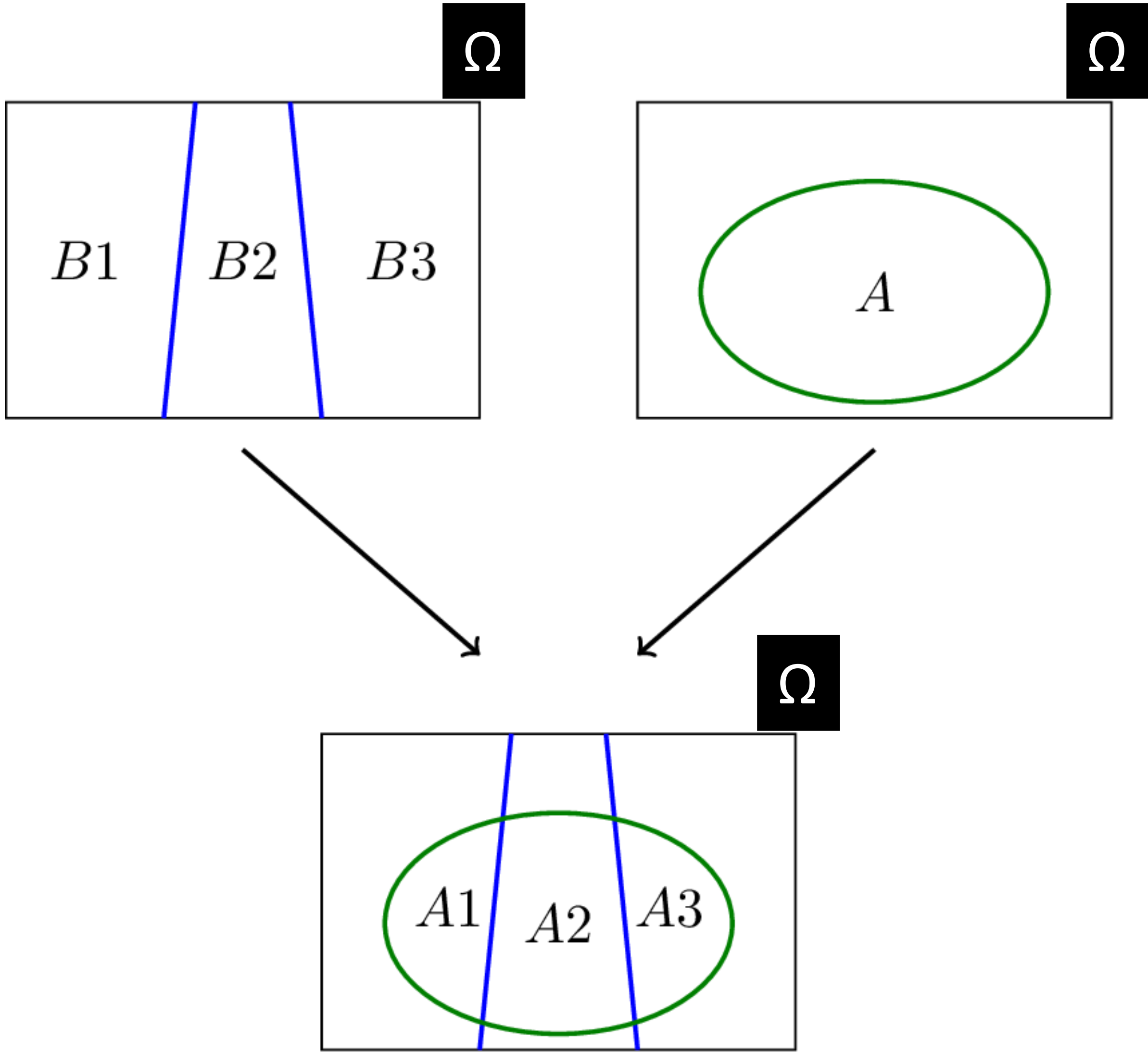
Partition

- Definition: A set of events $\{B_1, B_2, \ldots, B_n\}$ is a **partition** of the sample space if the events are:
  - Mutually exclusive: pairwise disjoint where $B_i \cap B_j = \Phi$, for all i,j; and
  - Exhaustive: $\cup_i B_i = \Omega$



Total Probability

- Given a partition $\{B_1, B_2, \ldots, B_n\}$, total probability $P(A) := \sum_i P(A \cap B_i) = \sum_i P(A|B_i)P(B_i)$



Independence

- Definition 1:
  Two events A and B are independent iff $P(A|B) := P(A)$
- Definition 2:
  Two events A and B are independent iff $P(A \cap B) := P(A)P(B)$
- Example: Roll two dice
  - Event A = get 1 on first die
  - Event B = get 1 on second die
- Example: Image
  - Event A: perturbation in observed intensity at pixel j
  - Event B: perturbation in observed intensity at pixel k

Conditional Independence

- Definitions
  - Events A and B are conditionally independent given event C iff $P(A,B|C) := P(A|C)P(B|C)$
  - Interpretation: Given event C, event B doesn't add any information to the probability of event A (or vice versa)
  - Events A and B are conditionally independent given event C iff $P(A|B,C) := P(A|C)$
- Example: consider the sequence of values $x_0 = rand(); x_{i+1} = x_i + 1$
  - Then, $P(x_5 | x_4, x_3, x_2, x_1, x_0) = P(x_5 | x_4)$. Given value of x_4, value of x_5 is conditionally independent of x_3, x_2, x_1, x_0, but x_5 isn't independent of them.
- Example
  - Observed colors at pixels j and k in a constant-color object are dependent, but given true object color, colors at j and k are conditionally independent

Random Number Generation

- A software function generating random numbers generates/**simulates** a sequence of numbers or symbols that cannot be reasonably predicted better than by a random chance
  - **Pseudo-random-number** generating functions() generate numbers that seem random, but are actually deterministic, and can be reproduced if the "internal state" of the generator is known
    - The internal state is typically set by a "seed" value
    - e.g., commands that "set seed to 0 and then generate 10 random numbers" will always return the same sequence
    - This is greatly useful in making code reproducible, and during debugging
  - To remove the determinism, in actual code deployment where this is needed, random-number generators can set their seed/state, just before generating the random sequence, using the value of an environmental variable (e.g., date-time in milliseconds, temperature) that is difficult to predict/model

Random Number Generation

- An example (pseudo)random-number generator: Mersenne Twister
  - Invented by Makoto Matsumoto and Takuji Nishimura (alphabetical order) of Hiroshima University
    - http://www.math.sci.hiroshima-u.ac.jp/m-mat/MT/emt.html
  - Why this name ?
    - http://www.math.sci.hiroshima-u.ac.jp/m-mat/MT/ename.html
    - Uses Mersenne primes
      - Primes of the form 2^p–1, where p itself is prime
    - Building upon earlier work: Twisted Generalized Feedback Shift Register Sequence
  - Code
    - http://www.math.sci.hiroshima

```
MT was firstly named "Primitive Twisted Generalized Feedback Shift
  Register Sequence" by a historical reason.
Makoto: Prof. Knuth said in his letter "the name is mouthful."
Takuji: ........
a few days later
Makoto: Hi, Takkun, How about "Mersenne Twister?" Since it uses Mersenne
  primes, and it shows that it has its ancestor Twisted GFSR.
Takuji: Well.
Makoto: It sounds like a jet coaster, so it sounds quite fast, easy to
  remember and easy to pronounce. Moreover, although it is a secret, it
  hides in its name the initials of the inventors.
Takuji: .......
Makoto: Come on, let's go with MT!
Takuji: ....well, affirmative.
Later, we got a letter from Prof. Knuth saying "it sounds a nice name." :-)
```

Random Number Generation

- Makoto Matsumoto
  - http://www.math.sci.hiroshima-u.ac.jp/m-mat/eindex.html

Profile

A professor at Department of Mathematics at Hiroshima University. He considers himself as a mathematician, but is not confident. There are so many Makoto Matsumoto's in Japan (quite a common name). The famous senior mathematician Makoto Matsumoto, who majors Finsler geometry, is another person, but in MathReview there is a confusion. Sorry.

Studies

Every Mathematics; in particular algebraic researches. Weak side is analysis. Arithmetic fundamental groups, mapping class groups, combinatorics, pseudorandom number generation, computer science, etc. are his research areas.

A dream

To be a mathematician. But it seems hard.

An R-rated profile. Warning: the below contains possibly harmful contents to serious researchers. X-profile

Random Number Generation

- “Random number generation is too important to be left to chance”, Studies in Applied Math. 1969, 3:70-111.
- Robert R. Coveyou
  - American research mathematician
  - Worked at Oak Ridge National Laboratory

Monte Carlo Methods

- Random number generation is important in statistical inference, where we can get good approximations to analytically-intractable problems using **simulated random experiments**
- Such methods were called Monte Carlo methods
- Invented while working on nuclear weapons projects, at Los Alamos National Laboratory, by:
  - Stanislaw Ulam: mathematician, physicist
  - John von Neumann: mathematician, physicist, computer scientist, engineer
- Term coined by their colleague Nicholas Metropolis
  - Physicist working on statistical mechanics
  - Refers to the Monte Carlo Casino in Monaco, where Ulam's uncle would borrow money from relatives to gamble

Monte Carlo Methods

- Monte Carlo casino in Monaco, France
  - Roulette tables

Population, Sample

- **Population** = a set of similar items or events which is of interest for some question or experiment
- **Sample** = a **set** of individuals or objects collected or selected from a statistical population by a defined procedure
- Elements of a sample are known as sample points, sampling units, or observations