

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
 - Examples of interest
 - Set of natural numbers, integers, real numbers
 - Set of fruits, people
- **Experiment:** an empirical procedure
 - Empirical: based on what is experienced or seen, rather than on theory
- **Statistical/Random Experiment:**
an experiment whose outcome is uncertain, e.g.,
 - Tossing a coin, Rolling a die
 - Measuring the lifetime of a laptop battery
 - Measuring the temperature at a specified outdoor location after 50 hours
 - Note: NO uncertainty about the procedure, only the outcome

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 a specific location in the universe, after 50 hours:
Subset of real numbers in the interval from absolute zero (0 Kelvin) to Planck temperature (1.416785×10^{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:

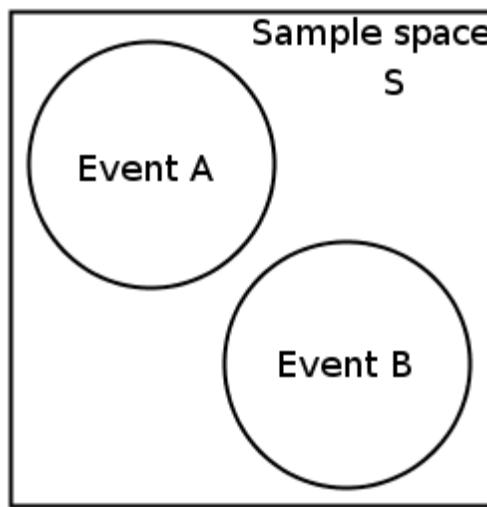
The temperature is between 298 Kelvin and 300 Kelvin

- Event Y: Within an image,

intensity at chosen pixel falls between 10 and 20

- **Operations on events**

- Union, “or”, $A \cup B$, A happens or B happens (or both)
- Intersection, “and”, $A \cap B$, A happens and B happens
- Complement, “negation”, \bar{A} , A doesn’t happen
- Subtraction, $A - B$, A happens but B doesn’t happen



0	2	15	0	0	11	10	0	0	0	0	9	9	0	0	0	0
0	0	0	4	60	157	236	255	255	177	95	61	32	0	0	29	
0	10	16	119	238	255	244	245	245	243	250	249	255	222	103	10	0
0	14	170	255	255	244	254	255	253	245	255	249	253	251	124	1	1
2	98	255	228	255	251	254	211	141	116	122	215	251	238	255	49	
13	217	243	255	155	33	226	52	2	0	10	13	232	255	255	36	
16	229	252	254	49	12	0	0	7	7	0	70	237	252	235	62	
6	141	245	255	212	25	11	9	3	0	115	236	243	255	137	0	
0	87	252	250	248	215	60	0	1	121	252	255	248	144	6	0	
0	13	113	255	255	245	255	182	181	248	252	242	208	36	0	19	
1	0	5	117	251	255	241	255	247	255	241	162	17	0	7	0	
0	0	0	4	58	251	255	246	254	253	255	120	11	0	1	0	
0	0	4	97	255	255	255	248	252	255	244	255	182	10	0	4	
0	22	206	252	246	251	241	100	24	113	255	245	255	194	9	0	
0	111	255	242	255	158	24	0	0	6	39	255	232	230	56	0	
0	218	251	250	137	7	11	0	0	0	2	62	255	250	125	3	
0	173	255	255	101	9	20	0	13	3	13	182	251	245	61	0	
0	107	251	241	255	230	98	55	19	118	217	248	253	255	52	4	
0	18	146	250	255	247	255	255	249	255	240	255	129	0	5		
0	0	0	23	113	215	255	250	248	255	255	248	248	118	14	12	0
0	0	6	1	0	52	153	233	255	252	147	37	0	0	4	1	
0	0	5	5	0	0	0	0	0	14	1	0	6	6	0	0	

Probability Axioms

- **Event Space \mathcal{B} :** Space of all possible events (that we want to model)
 - (A subset of) power set of sample space
 - Includes empty set \emptyset
 - Includes sample space Ω
 - Given a sample space, the event space \mathcal{B} must satisfy several rules (basically, for logical consistency and to deal with sample spaces that are infinite):
 1. Empty set $\emptyset \in \mathcal{B}$
 2. Closed under Countable Unions:
If events A_1, A_2, A_3, \dots belong to \mathcal{B} , then $A_1 \cup A_2 \cup A_3 \cup \dots \in \mathcal{B}$
 - We can only have a union of a finite or countably infinite number of sets here
 3. Closed under Complementation:
If event $A \in \mathcal{B}$, then event \bar{A} (i.e., $\Omega - A$) $\in \mathcal{B}$
 - A set \mathcal{B} 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 \mathcal{B}

- Example:

- If sample space $\Omega = \{ a, b, c, d \}$, then:
 - One possible design of the event space is: $\mathcal{B} = \{ \emptyset, \{a,b\}, \{c,d\}, \{a,b,c,d\} \}$
 - Another possible design of the event space is: $\mathcal{B} = \text{power set of } \Omega$
 - (A more useful example) If sample space is the real line, then:
 - Event space \mathcal{B} 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)
- Gives the “probability”/chance that the event occurs
- A probability function/measure on an event space \mathcal{B} is a function $P:\mathcal{B} \rightarrow [0,1]$ such that:
 - $P(\emptyset) = 0$;
 - $P(\Omega) = 1$; this needn’t hold for a general (non-probability) measure
 - For pairwise disjoint sets (events) A_1, A_2, A_3, \dots , we have
$$P(A_1 \cup A_2 \cup A_3 \cup \dots) = P(A_1) + P(A_2) + P(A_3) + \dots$$
 - We want the size of the union of disjoint sets to be the sum of their individual sizes, for any finite or countably infinite sequence of disjoint sets
- In simpler words, the probability function/measure $P(\cdot)$ on an event space \mathcal{B} 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 (designed to be) the probability that A occurs

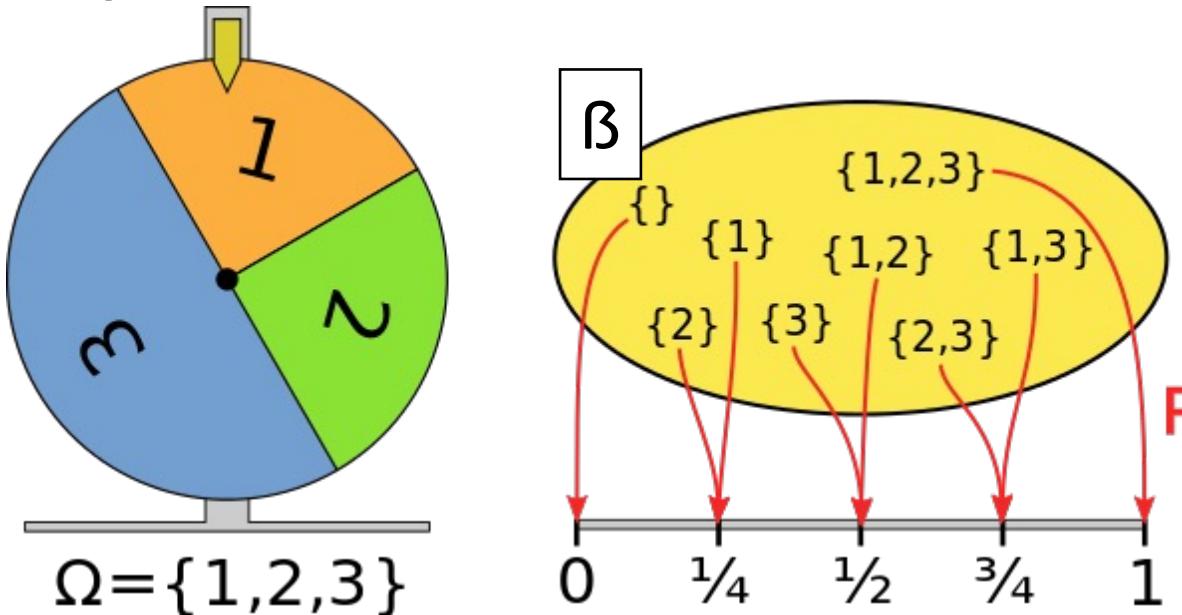
Probability Axioms

• Properties of Probability Functions/Measures

- Complement of an event A
 - $P(\bar{A}) = 1 - P(A)$
- 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, \mathcal{B}, P) where:
 - Ω is the **sample space** (= space of all possible outcomes)
 - \mathcal{B} is the **event space** (= space of all possible events; that we want to model)
 - $P(\cdot)$ is the **probability function/measure** on \mathcal{B} with $P(\Omega)=1$



- Examples

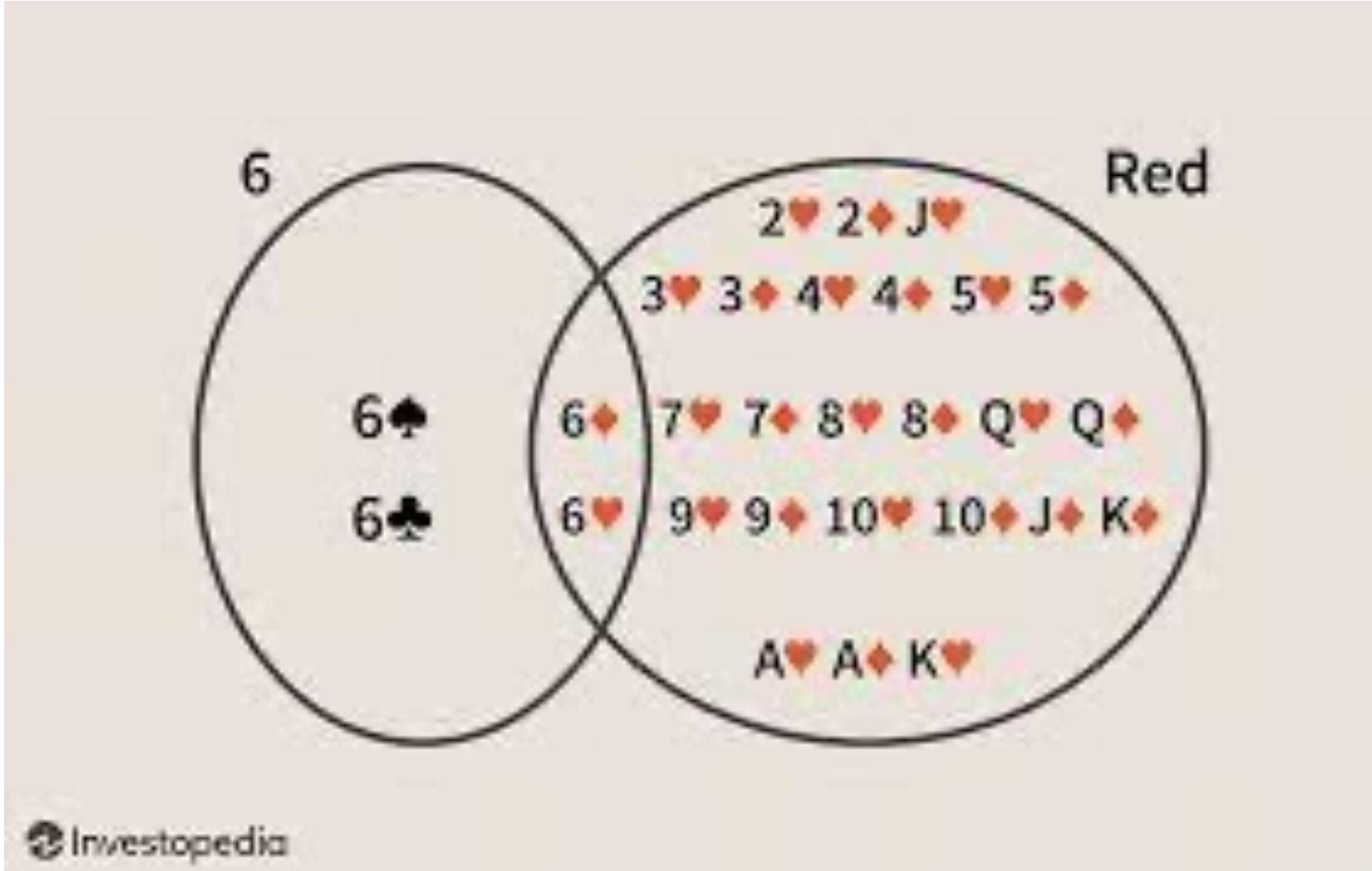
- Toss fair coin: What is Ω ? What is \mathcal{B} ? What is P ?
- Roll fair die: $P(\{1\}) = 1/6$. All outcomes equally likely.

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 $\Omega' = \Omega_1 \times \Omega_2 = \{ (s_1, s_2) : s_1 \in \Omega_1, s_2 \in \Omega_2 \}$
 - Example: First, toss a coin. Then, roll a die.
 - What is Ω' ? What is \mathcal{B}' ? 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 $\Omega' = \Omega \times \Omega = \{ (s_1, s_2) : s_1 \in \Omega, s_2 \in \Omega \}$
 - Note: element (x, y) is an ordered pair; not a set. So, order matters.
 - Example: Toss two coins.
 - What is Ω' ? What is \mathcal{B}' ? What is P' ?

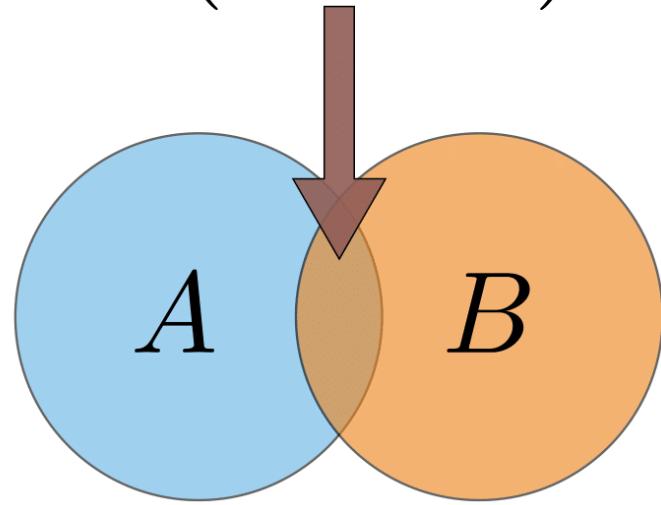
Joint Probability

- Definition: $P(A \text{ and } B) := P(A, B) := P(A \cap B)$
 - “:=” symbol indicates definition



Joint Probability

$$P(A \cap B)$$



Conditional Probability

- Definition:

Conditional probability of event A, given event B, is:

$$P(A|B) := P(A \cap B) / P(B)$$

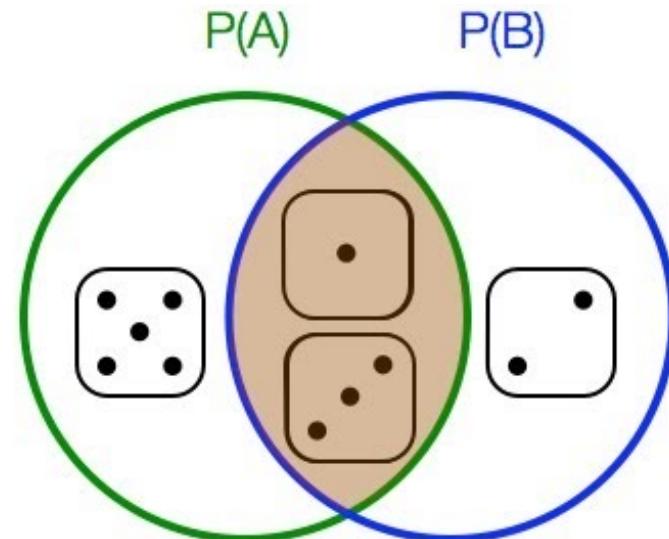
- Think of conditioning as restricting the sample space

- Example: Dice

- $A = \{\text{number odd}\}$, $B = \{\text{number} < 4\}$,
 $P(B)=P(A)=1/2$, $P(B|A)=2/3$

- Example: Image-Generation Process

- For a pixel,
first choose an object label
(e.g., sky, rock),
then choose pixel color given object label
 - $P(\text{pixel takes specific color})$
versus
 $P(\text{pixel takes specific color} | \text{object label})$



Conditional Probability

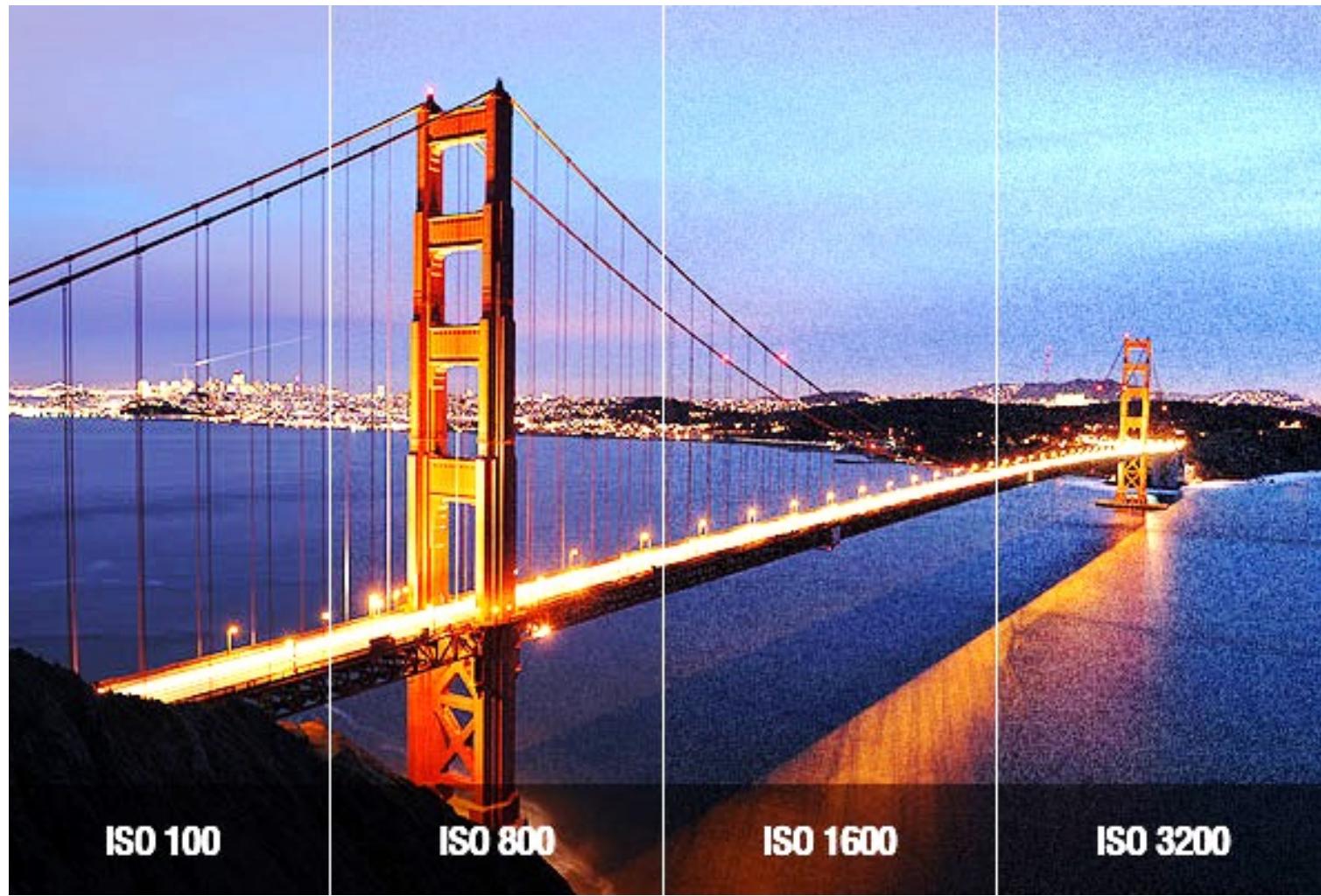
- Conditional probability of event A, given event B, is:

$$P(A|B) := P(A \cap B) / P(B)$$

- Example: Image-Acquisition Process

- Observed color is a perturbed version of the true color, due to measurement errors
 - Larger ISO setting → larger perturbation magnitude
- $P(\text{pixel takes specific color} | \text{true color at that pixel})$ → perturbation model

[larger ISO → film is more sensitive to light]



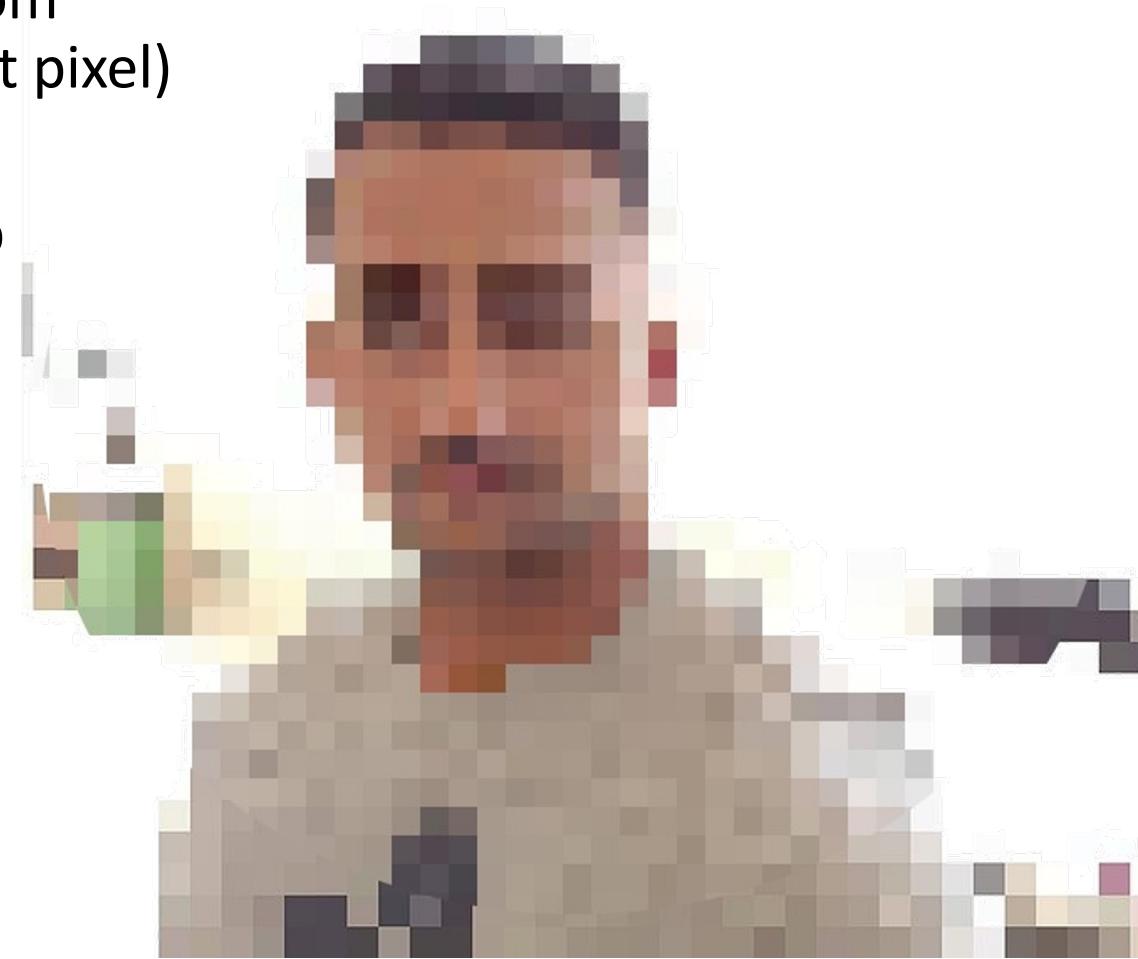
Conditional Probability

- Conditional probability of event A, given event B, is:

$$P(A|B) := P(A \cap B) / P(B)$$

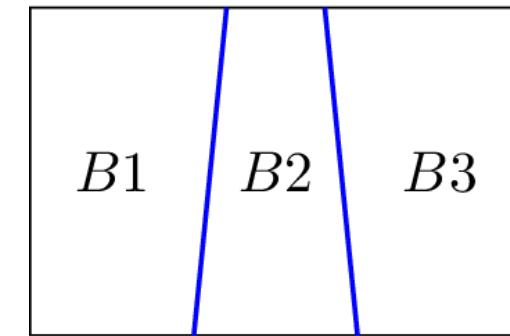
- Example: Image-Generation Process

- $P(\text{pixel takes specific color})$ usually differs from
 $P(\text{pixel takes specific color} | \text{color at adjacent pixel})$
- Former relates to the image histogram
- Latter relates to inter-pixel color relationship



Partition

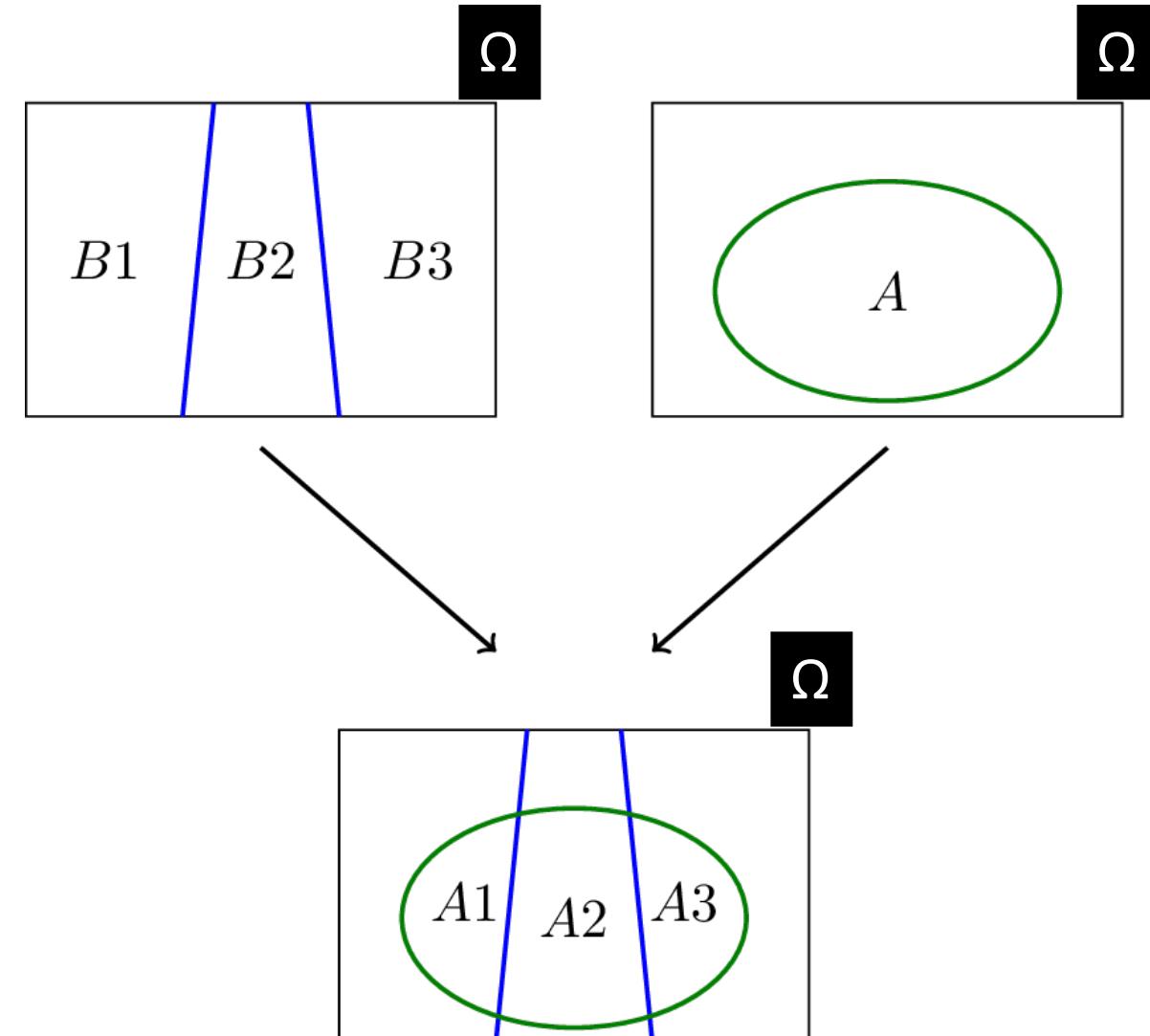
- Definition: A set of events $\{B_1, B_2, \dots, B_n\}$ is a **partition** of the sample space if the events are:
 - Mutually exclusive: pairwise disjoint where $B_i \cap B_j = \emptyset$, for all i, j ; and
 - Exhaustive: $\bigcup_i B_i = \Omega$



Total Probability

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

- Partition over sample-space (set) Ω induces a partition over event (set) A



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 $P(A,B) = P(A)P(B)$
- 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 $x_0 = \text{rand}(); x_{i+1} = x_i + 1$
 - 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.

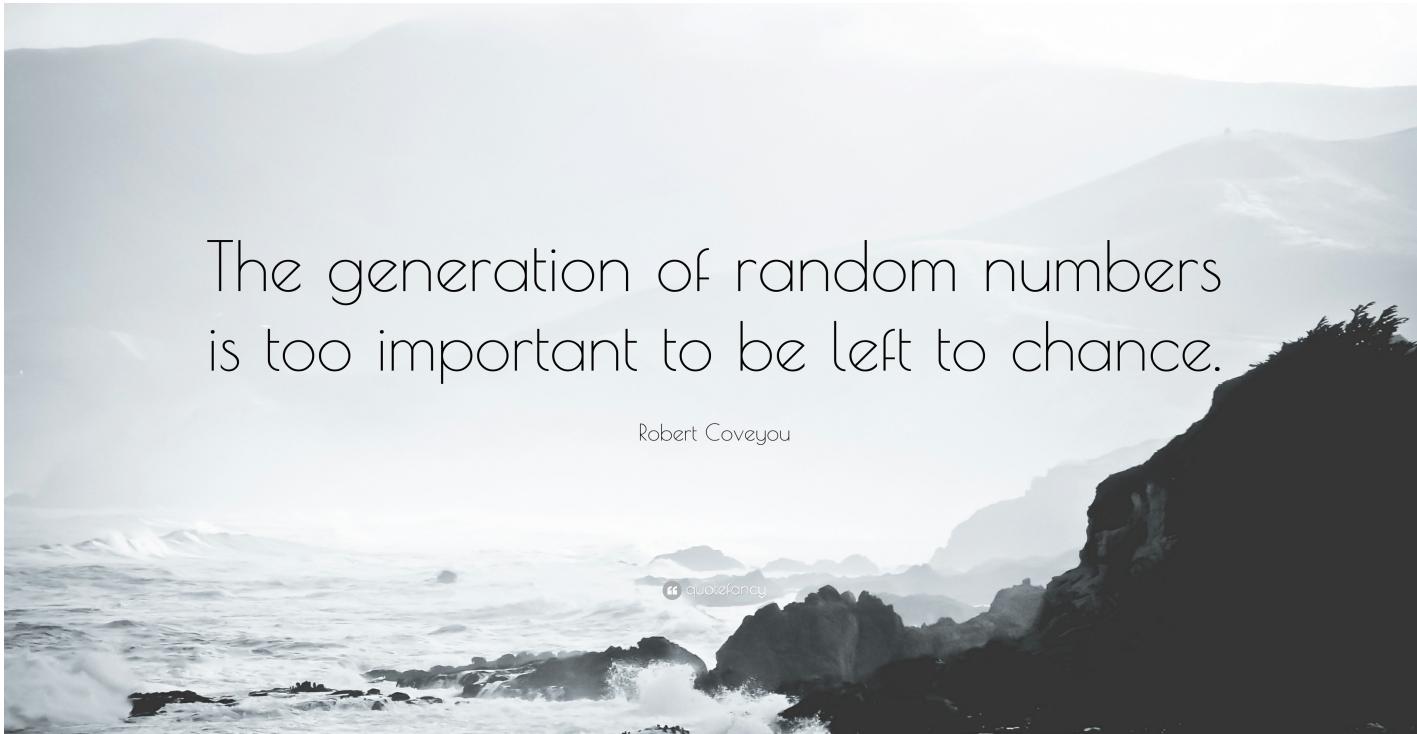
Studies

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.

Random Number Generation



The generation of random numbers
is too important to be left to chance.

Robert R. Coveyou

- “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

