Discrete Probability

Mahmuda Naznin

CSE 103

Random Variables

- A *"random variable"* V is any variable whose value is unknown, or whose value depends on the precise situation.
- The proposition $V=v_i$ may have an uncertain truth value, and may be assigned a *probability*.

Example

- A fair coin is flipped 3 times. Let S be the sample space of 8 possible outcomes, and let X be a random variable that assignees to an outcome the number of heads in this outcome.

- **Random variable X is a function** $X:S \rightarrow X(S)$,
  where $X(S)=\{0, 1, 2, 3\}$ is the range of X, which is the number of heads H, and
  $S=\{$ (TTT), (TTH), (THH), (HTT), (HHT), (HHH), (THT), (HTH) $\}$

- X(TTT) = 0 (no head)
  X(TTH) = X(HTT) = X(THT) = 1 (1 head)
  X(HHT) = X(THH) = X(HTH) = 2 (2 heads)
  X(HHH) = 3 (3 heads)

- The **probability distribution (pdf) of random variable** X is given by
  Pr(X=3) = 1/8, Pr(X=2) = 3/8, Pr(X=1) = 3/8, Pr(X=0) = 1/8.

Experiments and Sample Spaces

- A (stochastic) *experiment* is any process by which a given random variable V gets assigned some *particular* value, and where this value is not necessarily known in advance.
- The *sample space* S of the experiment is just the domain of the random variable, $S = \mathrm{dom}[V]$.
- The *outcome* of the experiment is the specific value v_i of the random variable that is selected.

Probability

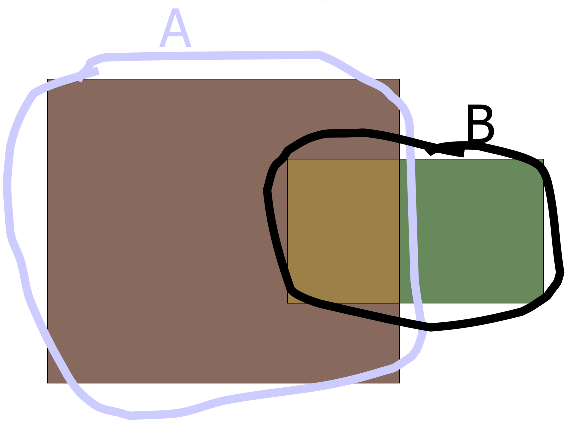
- The *probability* $p = \Pr[E] \in [0,1]$ of an event E is a real number representing our degree of certainty that E will occur.
  - If $\Pr[E] = 1$, then E is absolutely certain to occur,
    - thus $V \in E$ has the truth value **True**.
  - If $\Pr[E] = 0$, then E is absolutely certain *not* to occur,
    - thus $V \in E$ has the truth value **False**.
  - If $\Pr[E] = ½$, then we are *maximally uncertain* about whether E will occur with 50% probability.

The Axioms of Probability

- 0 <= P(A) <= 1
- Pr(True) = 1
- Pr(False) = 0
- Pr(A or B) = Pr(A) + Pr(B) - Pr(A and B)

Interpreting the axioms

- 0 <= Pr(A) <= 1
- Pr(True) = 1
- Pr(False) = 0
- Pr(A or B) = Pr(A) + Pr(B) - Pr(A and B)



More can be computed from the Axioms

- 0 <= Pr(A) <= 1, Pr(True) = 1, Pr(False) = 0
- Pr(A or B) = Pr(A) + Pr(B) - Pr(A and B)

From these we can find:

Pr(not A) = Pr(~A) = 1-Pr(A)

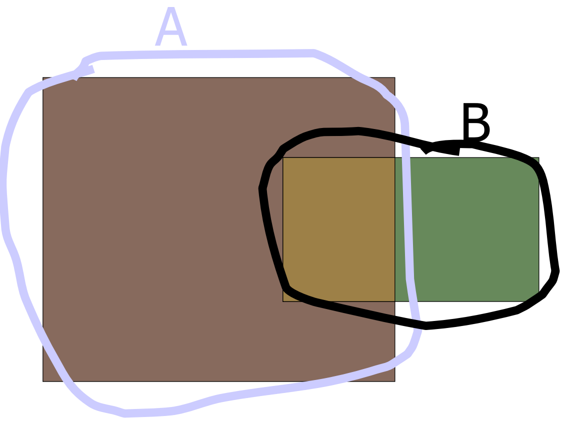
Total Probability

- $0 <= Pr(A) <= 1$, $Pr(True) = 1$, $Pr(False) = 0$
- $Pr(A \text{ or } B) = Pr(A) + Pr(B) - Pr(A \text{ and } B)$

From these we can find

$$Pr(A) = Pr(A \cap B) + Pr(A \cap \sim B)$$

It is called *total probability* rule



Probability of an Event E

The probability of an event E is the sum of the probabilities of the outcomes in E. That is

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

Note that, if there are *n* outcomes in the event E, that is, if $E = \{a_1, a_2, \ldots, a_n\}$ then

$$p(E) = \sum_{i=1}^{n} p(a_i)$$

Example

- What is the probability that, if we flip a coin three times, we get an odd number of tails?

(TTT), (TTH), (THH), (HTT), (HHT), (HHH), (THT), (HTH)

Each outcome has probability 1/8,

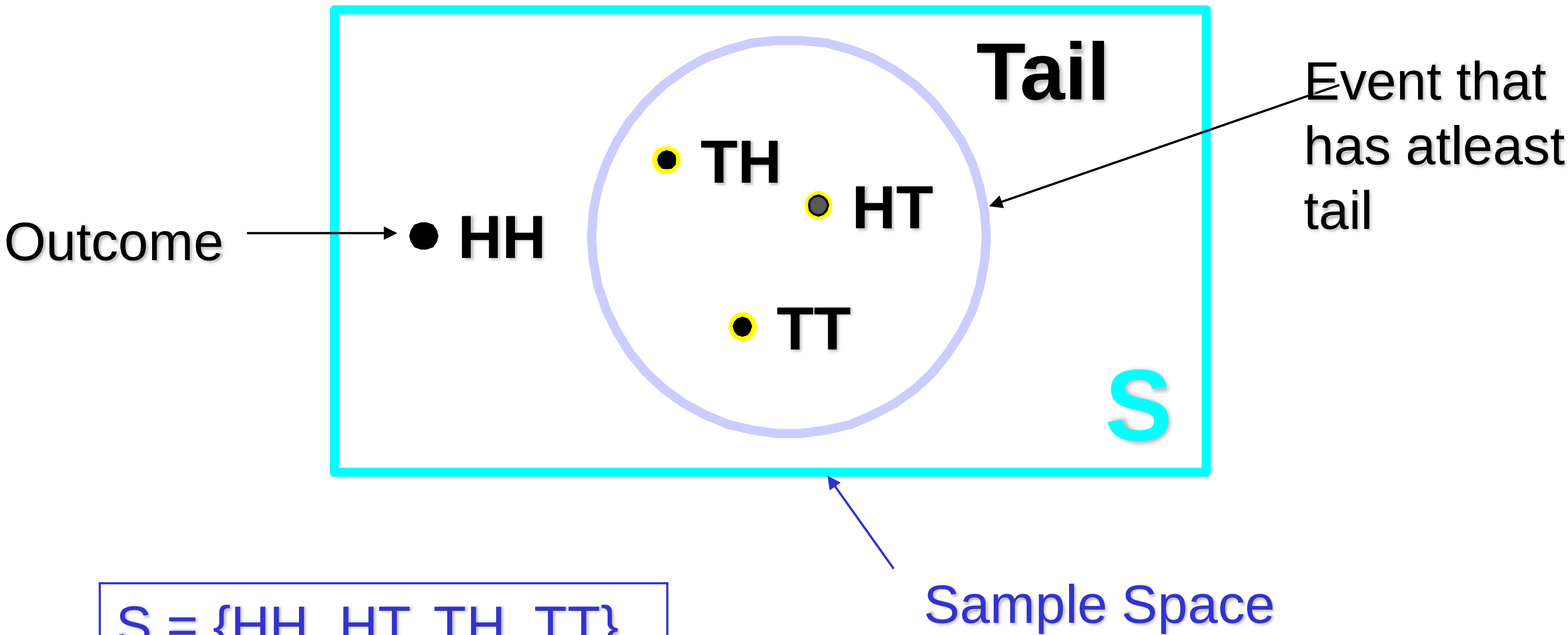
p(odd number of tails) = 1/8+1/8+1/8+1/8 = ½

Visualizing Sample Space

- 1. Listing
  - S = {Head, Tail}
- 2. Venn Diagram
- 3. Contingency Table
- 4. Decision Tree Diagram

Venn Diagram

Experiment: Toss 2 Coins. Note Faces.



S = {HH, HT, TH, TT}

Contingency Table

Experiment: Toss 2 Coins. Note Faces.

1st Coin	2nd Coin Head	2nd Coin Tail	Total
Head	HH	HT	HH, HT
Tail	TH	TT	TH, TT
Total	HH, TH	HT, TT	S

Simple Event (Head on 1st Coin)

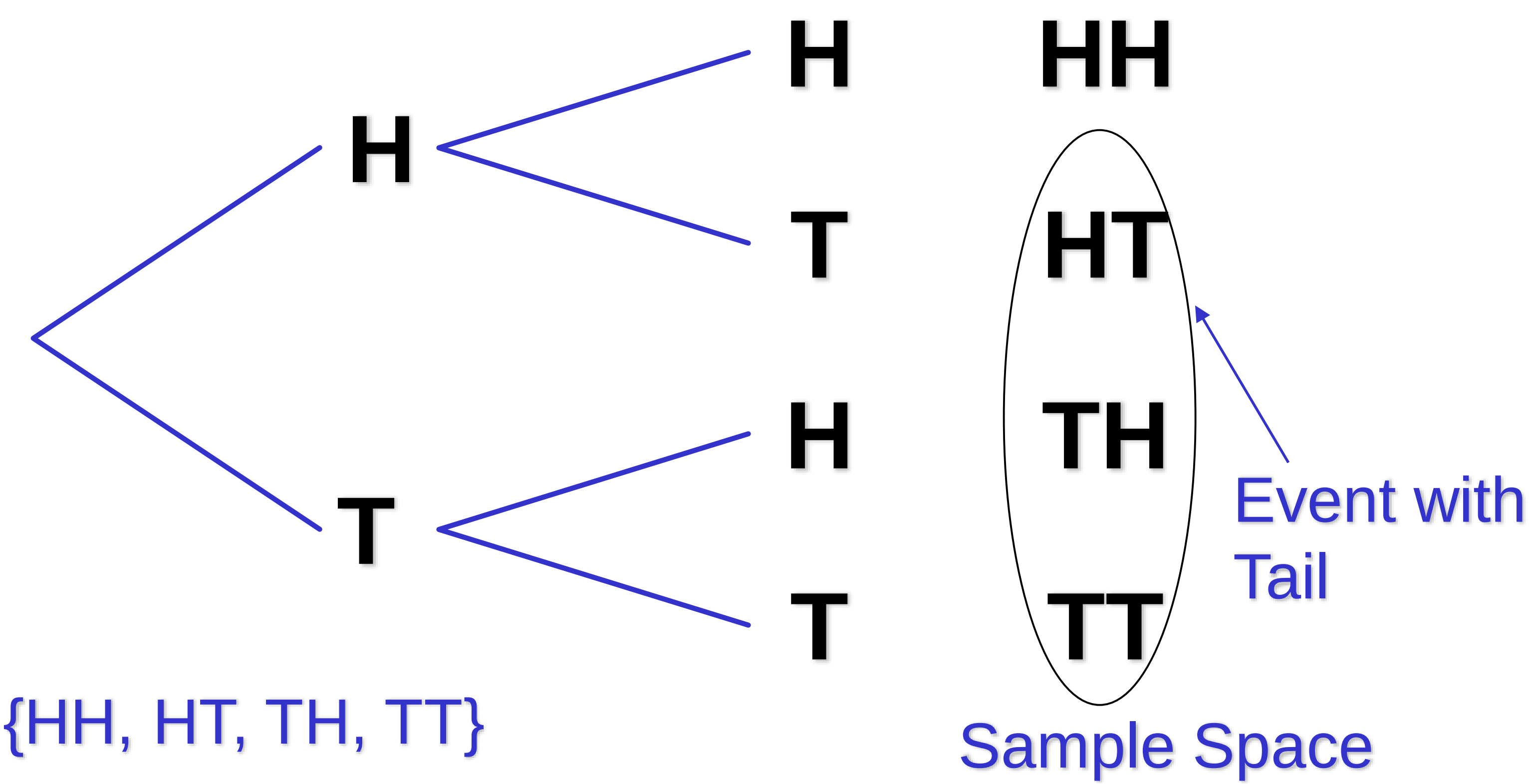
Outcome

Sample Space

S = {HH, HT, TH, TT}

Tree Diagram

Experiment: Toss 2 Coins. Note Faces.



S = {HH, HT, TH, TT}

Discrete Random Variable

- Possible values (outcomes) are discrete
  - E.g., natural number (0, 1, 2, 3 etc.)
- Obtained by Counting
- Usually Finite Number of Values
  - But could be infinite (must be “countable”)

Discrete Probability Distribution

1. List of All possible [x, $\Pr(x)$] pairs
   - x = Value of Random Variable (Outcome)
   - $p(x)$ = Probability Associated with Value
2. Mutually Exclusive (No Overlap)
3. Collectively Exhaustive (Nothing Left Out)
4. $0 \leq \Pr(x) \leq 1$
5. $\sum \Pr(x) = 1$

Visualizing Discrete Probability Distributions

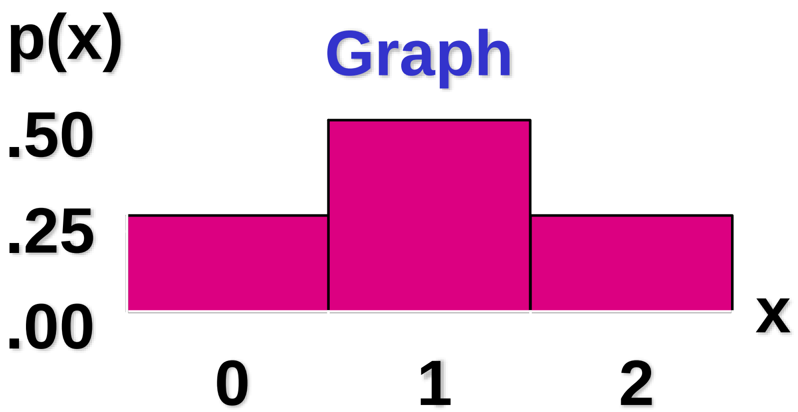
Listing

{ (0, .25), (1, .50), (2, .25) }

Table

# Tails	f(x) Count	p(x)
0	1	.25
1	2	.50
2	1	.25

Graph



Equation

$$p(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

Mutually Exclusive Events

- Two events E_1, E_2 are called *mutually exclusive* if they are disjoint: $E_1 \cap E_2 = \varnothing$
  - Note that two mutually exclusive events *cannot both occur* in the same instance of a given experiment.
- For mutually exclusive events,

$$\Pr[E_1 \cup E_2] = \Pr[E_1] + \Pr[E_2].$$

Exhaustive Sets of Events

- A set $\boldsymbol{E} = \{E_1, E_2, \ldots\}$ of events in the sample space S is called *exhaustive* iff $\bigcup E_i = S$.
- An exhaustive set $\boldsymbol{E}$ of events that are all mutually exclusive with each other has the property that

$$\sum \Pr[E_i] = 1.$$

Bernoulli Trials

- Each performance of an experiment with only two possible outcomes is called a **Bernoulli trial**.
- In general, a possible outcome of a Bernoulli trial is called a **success** or a **failure**.
- If p is the probability of a success and q is the probability of a failure, then $p+q=1$.

Probability of *k* successes in *n* independent Bernoulli trials.

The probability of *k* successes in *n* independent Bernoulli trials, with probability of success p and probability of failure q = 1-p is $C(n,k)\, p^k q^{n-k}$

Independent Events

- Two events E, F are called *independent* if

  $$\Pr[E \cap F] = \Pr[E] \cdot \Pr[F].$$

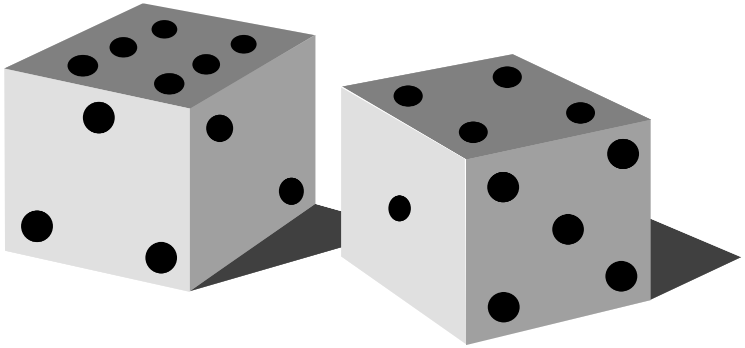
- Relates to the product rule for the number of ways of doing two independent tasks.
- **Example:** Flip a coin, and roll a dice.

  Pr[(coin shows heads) $\cap$ (dice shows 1)] =

  Pr[coin is heads] $\times$ Pr[dice is 1] = $\frac{1}{2} \times 1/6 = 1/12$.

Example

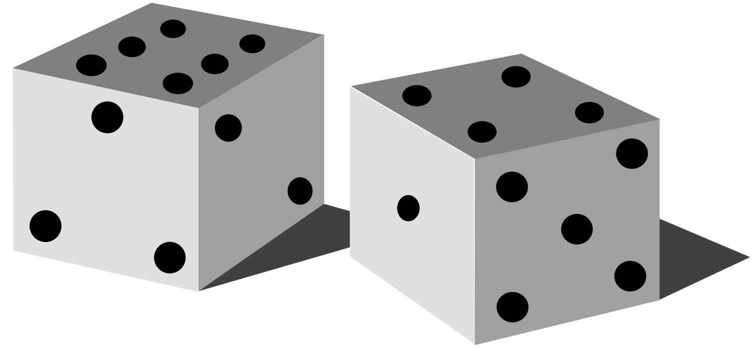
A red die and a blue die are rolled. The sample space:



	1	2	3	4	5	6
1	x	x	x	x	x	x
2	x	x	x	x	x	x
3	x	x	x	x	x	x
4	x	x	x	x	x	x
5	x	x	x	x	x	x
6	x	x	x	x	x	x

Are the events sum is 7 and the blue die is 3 independent?

The events sum is 7 and
the blue die is 3 are independent:



$|S| = 36$

$|\text{sum is } 7| = 6$

$|\text{blue die is } 3| = 6$

$|\text{ in intersection }| = 1$

	1	2	3	4	5	6
1	x	x	x	x	x	x
2	x	x	x	x	x	x
3	x	x	x	x	x	x
4	x	x	x	x	x	x
5	x	x	x	x	x	x
6	x	x	x	x	x	x

p(sum is 7 and blue die is 3) =1/36

p(sum is 7) p(blue die is 3) =6/36*6/36=1/36

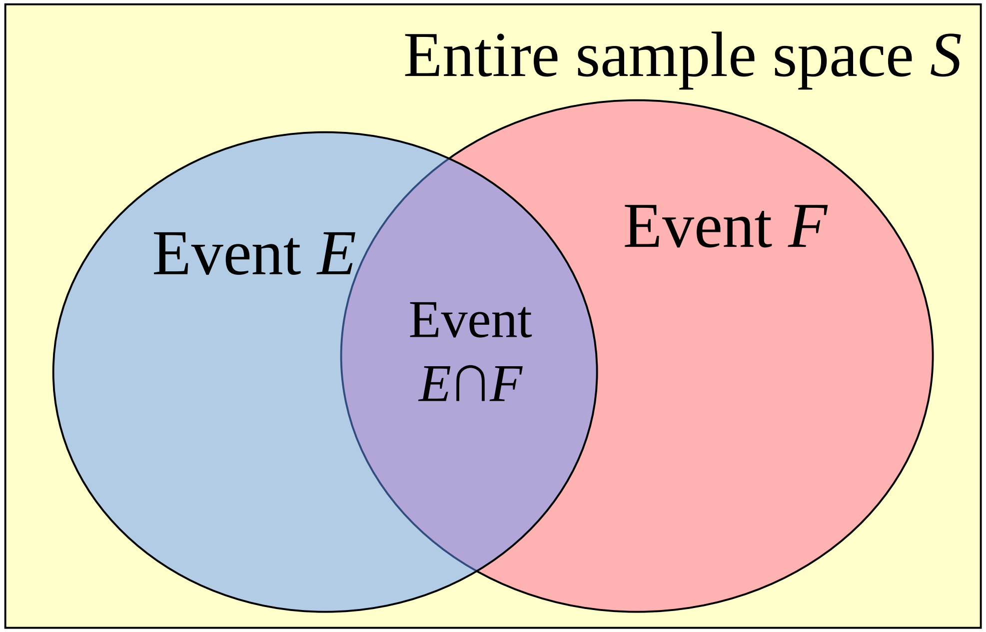
Thus, p((sum is 7) and (blue die is 3)) = p(sum is 7) p(blue die is 3)

Conditional Probability

- Let E,F be any events such that $\Pr[F]>0$.
- Then, the *conditional probability of* E *given* F, written $\Pr[E|F]$, is defined as
$$\Pr[E|F] = \Pr[E\cap F]/\Pr[F].$$
- This is what our probability that E would turn out to occur should be, if we are given *only* the information that F occurs.
- If E and F are independent then $\Pr[E|F] = \Pr[E]$.
  - $\because \Pr[E|F] = \Pr[E\cap F]/\Pr[F] = \Pr[E]\times\Pr[F] / \Pr[F] = \Pr[E]$

Visualizing Conditional Probability

- If we are given that event *F* occurs, then
  - Our attention gets restricted to the subspace *F*.
- Our *posterior* probability for *E* (after seeing *F*) corresponds to the *fraction* of *F* where *E* occurs also.
- Thus, $Pr'(E) = Pr(E \cap F)/Pr(F)$.



Example

- What is the probability that, if we flip a coin three times, that we get an odd number of tails (=event *E*), if we know that the event *F*, the first flip comes up tails occurs?

(TTT), (TTH), (THH), ~~(HTT)~~,
~~(HHT)~~, ~~(HHH)~~, (THT), ~~(HTH)~~

Each outcome has probability 1/4, (First Head will be discarded)

$\Pr(E\,|F) = 1/4+1/4 = ½$, where E=odd number of tails

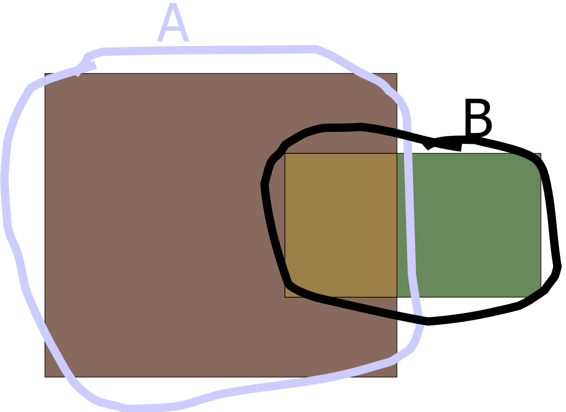
or $\Pr(E|F) = \Pr(E \cap F)/\Pr(F) = (1/4\,)/\ (1/2) = 2/4 = ½$

Conditional Probability Rules

It is called *multiplicative* rule-

$$p(A\,|B) = [p(A \cap B)]/p(B)$$
$$= [p(B|A)\,p(A)]/\,p(B)$$

$$p(A\,|B)p(B) = p(A \cap B)$$
$$= p(B \cap A)$$
$$= p(B|A)\,p(A)$$



Conditional Probability Rules

It is called *multiplicative* rule-

$p(A\,|B) = [p(A \cap B)]/p(B)$

$\qquad = [p(B|A)\; p(A)]/\; p(B)$

$p(A\,|B)p(B) = p(A \cap B) = p(B \cap A) = p(B|A)\; p(A)$

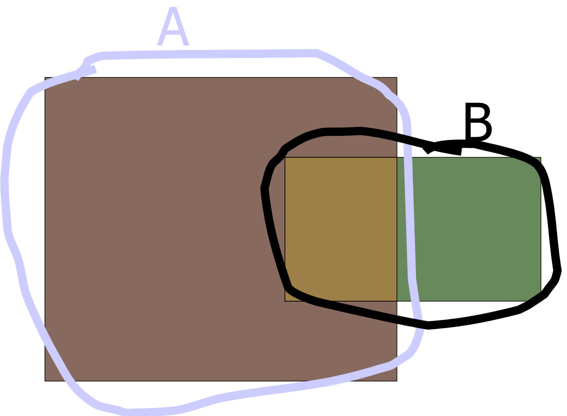
If A, *B* are *independent,*

$p(A \cap B) = p(A).p(B)$,

$\mathbf{p(\boldsymbol{B}) = p(\boldsymbol{B}|\boldsymbol{A})p(\boldsymbol{A}) + p(\boldsymbol{B}|\sim \boldsymbol{A})\; p(\sim \boldsymbol{\acute{A}})}$

$\mathbf{[\boldsymbol{P(B) = P(B \cap A) + P(B \cap \sim A)}]}$

It is called *total probability* rule



Bayes' Rule

- One way to compute the probability that a hypothesis *H* is correct, given a data *D*:

$$\Pr[H \mid D] = \frac{\Pr[D \mid H] \cdot \Pr[H]}{\Pr[D]}$$

- This follows directly from the definition of conditional probability

Extended Bayes' Theorem

- Allows one to compute the probability that any hypothesis H_i is correct, given data D:

$$\Pr[H_i \mid D] = \frac{\Pr[D \mid H_i] \cdot \Pr[H_i]}{\sum_j \Pr[D \mid H_j] \cdot \Pr[H_j]}$$

Set of H_j is exhaustive

Conditional Probability

What is the probability of a random bit string of length four contains at least two consecutive 0s, given that its first bit is a 0

E: "bit string contains at least two consecutive 0s"

condition: "first bit of the string is a 0"

- We know the formula $\Pr(E \mid F) = \Pr(E \cap F)/\Pr(F)$.
- $E \cap F$ = {0000, 0001, 0010, 0011, 0100}
- F={0000, 0001, 0010, 0011, 0100, 0101, 0110, 0111, 1000, 1001, 1010, 1011, 1100, 1101, 1110, 1111}
- $\Pr(E \cap F) = 5/16$
- $\Pr(F) = 8/16 = 1/2$
- $\Pr(E \mid F) = (5/16)/(1/2) = 10/16 = 5/8 = 0.625$

Bayes' Rule

- One way to compute the probability that a hypothesis *H* is correct, given a data *D*:

$$\Pr[H \mid D] = \frac{\Pr[D \mid H] \cdot \Pr[H]}{\Pr[D]}$$

- This follows directly from the definition of conditional probability

Extended Bayes' Theorem

- Allows one to compute the probability that any hypothesis H_i is correct, given data *D*:

$$\Pr[H_i \mid D] = \frac{\Pr[D \mid H_i] \cdot \Pr[H_i]}{\sum_j \Pr[D \mid H_j] \cdot \Pr[H_j]}$$

Set of H_j is exhaustive

Rules

Summary of Probability Rules and Laws

1) *The Rule of Complement:* $Pr(\overline{A}) = 1 - Pr(A)$
2) *The Additive Rule:* $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$.
   When A, B are disjoint, $Pr(A \cup B) = Pr(A) + Pr(B)$.
3) *Conditional Probability:* $Pr(A|B) = \dfrac{Pr(A \cap B)}{Pr(B)}$, $Pr(B) \neq 0$
4) *Multiplicative Rule:* $Pr(A)Pr(B|A) = Pr(A \cap B) = Pr(B)Pr(A|B)$.
   When A, B are independent, $Pr(A \cap B) = Pr(A)Pr(B)$.

5) *The Law of Total Probability:* $Pr(B) = Pr(A)Pr(B|A) + Pr(\overline{A})Pr(B|\overline{A})$

6) *The Law of Total Probability (Extended Version):* If $A_1, A_2, \ldots, A_n \subseteq \mathcal{S}$, where $n \geq 3$, $A_i \cap A_j = \emptyset$ for all $1 \leq i < j \leq n$, and $\mathcal{S} = \cup_{i=1}^{n} A_i$, then for any event B,

$$Pr(B) = Pr(A_1)Pr(B|A_1) + \cdots + Pr(A_n)Pr(B|A_n) = \sum_{i=1}^{n} Pr(A_i)Pr(B|A_i).$$

7) *Bayes' Theorem:* $Pr(A|B) = \dfrac{Pr(A \cap B)}{Pr(B)} = \dfrac{Pr(A)Pr(B|A)}{Pr(A)Pr(B|A) + Pr(\overline{A})Pr(B|\overline{A})}$

8) *Bayes' Theorem (Extended Version):* If $A_1, A_2, \ldots, A_n \subseteq \mathcal{S}$, where $n \geq 3$, $A_i \cap A_j = \emptyset$ for all $1 \leq i < j \leq n$, and $\mathcal{S} = \cup_{i=1}^{n} A_i$, then for any event B, and each $1 \leq k \leq n$,

$$Pr(A_k|B) = \frac{Pr(A_k \cap B)}{Pr(B)} = \frac{Pr(A_k)Pr(B|A_k)}{Pr(A_1)Pr(B|A_1) + \cdots + Pr(A_n)Pr(B|A_n)}$$

$$= \frac{Pr(A_k)Pr(B|A_k)}{\sum_{i=1}^{n} Pr(A_i)Pr(B|A_i)}.$$

Example-Conditional Probability

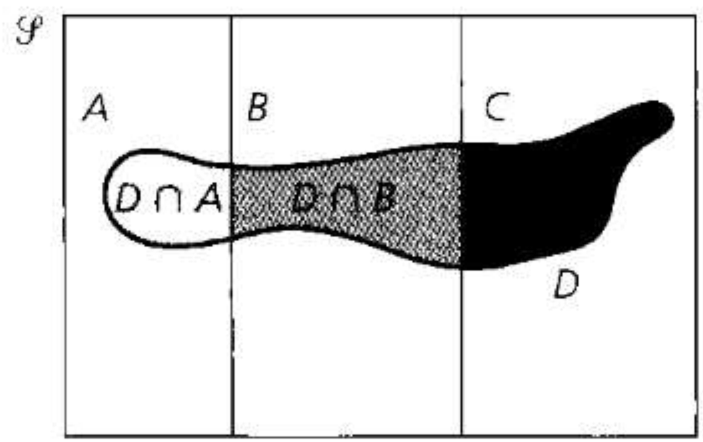
Emilio is a system integrator for personal computers. As such he finds himself using keyboards from three companies. Company 1 supplies 60% of the keyboards, company 2 supplies 30% of the keyboards, and the remaining 10% comes from company 3. From past experience Emilio knows that 2% of company 1's keyboards are defective, while the percentages of defective keyboards for companies 2, 3 are 3% and 5%, respectively. If one of Emilio's computers is selected, at random, and then tested, what is the probability it has a defective keyboard?

Let A denote the event

A: The keyboard comes from company 1.

Events B, C are defined similarly for companies 2, 3, respectively. Event D, meanwhile, is

D: The keyboard is defective.



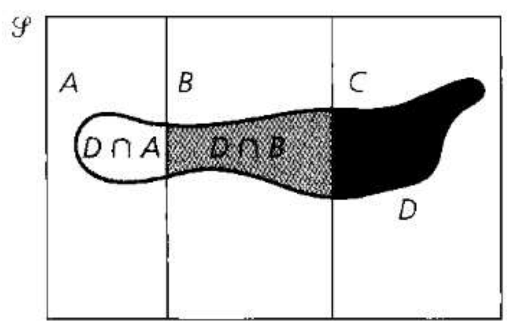
$D = D \cap \mathcal{S} = D \cap (A \cup B \cup C) = (D \cap A) \cup (D \cap B) \cup (D \cap C)$. But here $A \cap B = A \cap C = B \cap C = \emptyset$. So now, for example, the Laws of Set Theory show us that $(D \cap A) \cap (D \cap B) = D \cap (A \cap B) = D \cap \emptyset = \emptyset$. Likewise, $(D \cap A) \cap (D \cap C) = (D \cap B) \cap (D \cap C) = \emptyset$, and $(D \cap A) \cap (D \cap B) \cap (D \cap C) = \emptyset$. [redacted], we have

$$\begin{aligned} Pr(D) &= Pr(D \cap A) + Pr(D \cap B) + Pr(D \cap C) \\ &= Pr(A)Pr(D|A) + Pr(B)P(D|B) + Pr(C)Pr(D|C). \end{aligned}$$

From the information given at the start of this example we know that

$Pr(A) = 0.6$	$Pr(B) = 0.3$	$Pr(C) = 0.1$
$Pr(D\|A) = 0.02$	$Pr(D\|B) = 0.03$	$Pr(D\|C) = 0.05.$

So $Pr(D) = (0.6)(0.02) + (0.3)(0.03) + (0.1)(0.05) = 0.026$, and this tells us that 2.6% of the personal computers integrated by Emilio will have defective keyboards.



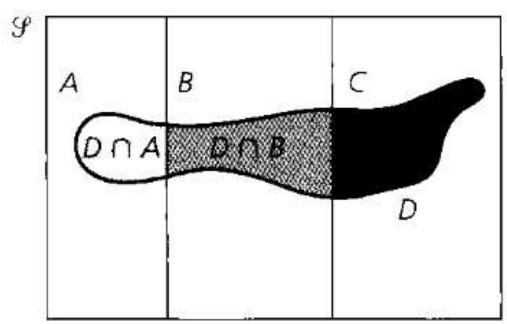
Example-Bay's Theorem

Referring back to the information in the preceding example, now we ask the question "If one of Emilio's personal computers is found to have a defective keyboard, what is the probability that keyboard came from company 3?"

Using the notation in [redacted] we see that here the given condition is D and that we want to find $Pr(C|D)$.

$$Pr(C|D) = \frac{Pr(C \cap D)}{Pr(D)} = \frac{Pr(C)Pr(D|C)}{Pr(A)Pr(D|A) + Pr(B)Pr(D|B) + Pr(C)Pr(D|C)}$$

$$= \frac{(0.1)(0.05)}{(0.6)(0.02) + (0.3)(0.03) + (0.1)(0.05)} = \frac{0.005}{0.026} = \frac{5}{26} \doteq 0.192308.$$



Acknowledgement

- Susana Epp (Reading Materials)
- Ralph P. Grimaldi, B. V. Ramana
- Kenneth Rosen (Reading Materials)