Discrete Structures for Computer Science

William Garrison

bill@cs.pitt.edu

6311 Sennott Square

Lecture #24: Probability Theory

University of Pittsburgh

Based on materials developed by Dr. Adam Lee

Not all events are equally likely to occur…

Sporting events

Investments

Games of strategy

Nature

We can model these types of real-life situations by relaxing our model of probability

As before, let S be our sample space. Unlike before, we will allow S to be either finite or countable.

We will require that the following conditions hold:

1. $0 \leq p(s) \leq 1$ for each $s \in S$
2. $$\sum_{s \in S} p(s) = 1$$

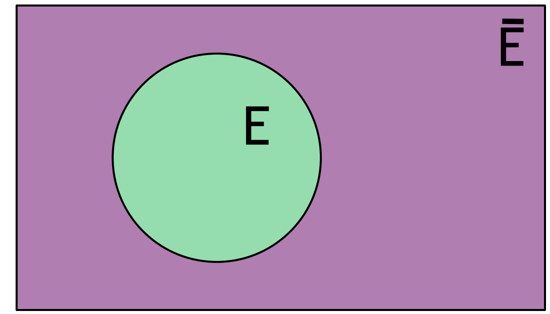
No event can have a negative likelihood of occurrence, or more than a 100% chance of occurence

In any given experiment, some event will occur

The function $p : S \rightarrow [0,1]$ is called a probability distribution

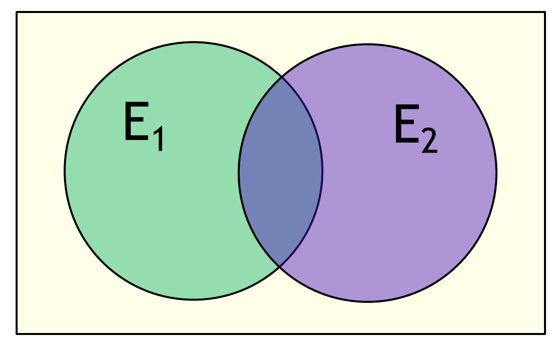
Recap our formulas for the probability of combinations of events

Property 1: $p(\overline{E}) = 1 - p(E)$

S = 

- Recall that $S = \overline{E} \cup E$ for any event E
- Further, $\sum_{s\in S} p(s) = 1$
- So, $p(S) = p(E) + p(\overline{E}) = 1$
- Thus, $p(\overline{E}) = 1 - p(E)$

Property 2: $p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2)$

S = 

- Recall that $p(E) = \sum_{s\in E} p(s)$
- Let x be some outcome in $E_1 \cup E_2$
- If x is in one of E_1 or E_2, then p(x) is counted once on the RHS of the equation
- If x is in both E_1 or E_2, then p(x) is counted 1 + 1 – 1 = 1 times on the RHS of the equation

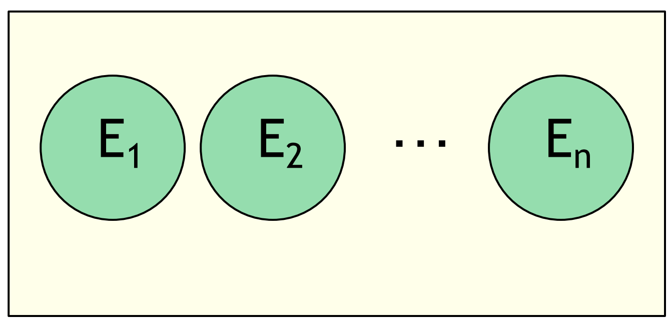
A formula for the probability of pairwise disjoint events

Theorem: If $E_1, E_2, \ldots, E_n$ is a sequence of pairwise disjoint events in a sample space S, then we have:

$$p\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} p(E_i)$$

Recall: $E_1, E_2, \ldots, E_n$ are pairwise disjoint iff $E_i \cap E_j = \emptyset$ for $1 \leq i,j \leq n$

S =



We can prove this theorem using mathematical induction!

How can we incorporate prior knowledge?

Sometimes we want to know the probability of some event given that another event has occurred.

Example: A fair coin is flipped three times. The first flip turns up tails. Given this information, what is the probability that an odd number of tails appear in the three flips?

Solution:

- Let F = “the first flip of three comes up tails”
- Let E = “tails comes up an odd number of times in three flips”
- Since F has happened, S is reduced to {THH, THT, TTH, TTT}
- We know:
- p(E) = |E|/|S|
- = |{THH, TTT}| / |{THH, THT, TTH, TTT}|
- = 2/4
- = 1/2

Conditional Probability

Definition: Let E and F be events with p(F) > 0. The conditional probability of E given F, denoted p(E | F), is defined as:

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

Intuition:

- Think of the event F as reducing the sample space that can be considered
- The numerator looks at the likelihood of the outcomes in E that overlap those in F
- The denominator accounts for the reduction in sample size indicated by our prior knowledge that F has occurred

Bit strings

Example: Suppose that a bit string of length 4 is generated at random so that each of the 16 possible 4-bit strings is equally likely to occur. What is the probability that it contains at least two consecutive 0s, given that the first bit in the string is a 0?

Solution:

- Let E = "A 4-bit string has at least two consecutive zeros"
- Let F = "The first bit of a 4-bit string is a zero"
- Want to calculate $p(E \mid F) = p(E \cap F)/p(F)$
- $E \cap F = \{0000, 0001, 0010, 0011, 0100\}$
- So, $p(E \cap F) = 5/16$
- Since each bit string is equally likely to occur, $p(F) = 8/16 = 1/2$
- So $p(E \mid F) = (5/16)/(1/2) = 10/16 = 5/8$

Kids

Example: What is the conditional probability that a family with two kids has two boys, given that they have at least one boy? Assume that each of the possibilities BB, BG, GB, GG is equally likely to occur.

Boy is older

Girl is older

Solution:

- Let E = "A family with 2 kids has 2 boys"
- E = {BB}
- Let F = "A family with 2 kids has at least 1 boy"
- F = {BB, BG, GB}
- E ∩ F = {BB}
- So p(E | F) = p(E ∩ F)/p(F)
- = (1/4) / (3/4)
- = 1/3

Does prior knowledge always help us?

Example: Suppose a fair coin is flipped twice. Does knowing that the coin comes up tails on the first flip help you predict whether the coin will be tails on the second flip?

Solution:

- S = {HH, HT, TH, TT}
- F = "Coin was tails on the first flip" = {TH, TT}
- E = "Coin is tails on the second flip" = {TT, HT}
- $p(E) = 2/4 = 1/2$
- $p(E \mid F) = p(E \cap F)/p(F)$
- $= (1/4) / (2/4)$
- $= 1/2$
- Knowing the first flip does not help you guess the second flip!

Independent Events

Definition: We say that events E and F are independent if and only if $p(E \cap F) = p(E)p(F)$.

Recall: In our last example…

- $S = \{HH, HT, TH, TT\}$
- $F = \{TH, TT\}$
- $E = \{HT, TT\}$
- $E \cap F = \{TT\}$

This checks out!

So:

- $p(E \cap F) = |E \cap F| / |S|$
- $= 1/4$
- $p(E)p(F) = 1/2 \times 1/2$
- $= 1/4$

Example: Bit Strings

Example: Suppose that E is the event that a randomly generated bit string of length four begins with a 1, and F is the event that this bit string contains an even number of 1s. Are E and F independent if all 4-bit strings are equally likely to occur?

Solution:

- By the product rule, $|S| = 2^4 = 16$
- E = {1111, 1110, 1101, 1011, 1100, 1010, 1001, 1000}
- F = {0000, 0011, 0101, 0110, 1001, 1010, 1100, 1111}
- So $p(E) = p(F) = 8/16 = 1/2$
- $p(E)p(F) = 1/4$
- $E \cap F$ = {1111, 1100, 1010, 1001}
- $p(E \cap F) = 4/16 = 1/4$
- Since $p(E \cap F) = p(E)p(F)$, E and F are independent events

Example: Distribution of kids

Example: Assume that each of the four ways that a family can have two children are equally likely. Are the events E that a family with two children has two boys, and F that a family with two children has at least one boy independent?

Solution:

- $E = \{BB\}$
- $F = \{BB, BG, GB\}$
- $p(E) = 1/4$
- $p(F) = 3/4$
- $p(E)p(F) = 3/16$
- $E \cap F = \{BB\}$
- $p(E \cap F) = 1/4$
- Since $1/4 \neq 3/16$, E and F are not independent

If probabilities are independent, we can use the product rule to determine the probabilities of combinations of events

Example: What is the probability of flipping heads 4 times in a row using a fair coin?

Answer: $p(H) = 1/2$, so $p(HHHH) = (1/2)^4 = 1/16$

Example: What is the probability of rolling the same number 3 times in a row using an unbiased 6-sided die?

Answer:

- First roll agrees with itself with probability 1
- 2nd roll agrees with first with probability 1/6
- 3rd roll agrees with first two with probability 1/6
- So probability of rolling the same number 3 times is $1 \times 1/6 \times 1/6 = 1/36$

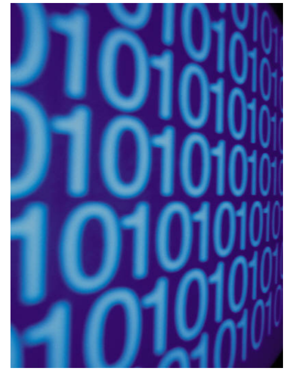
In-class exercises

Problem 1: What is the conditional probability that a randomly-generated 4-bit string contains two consecutive 1s, given that the first bit of the string is a 1?

Problem 2: What is the conditional probability that exactly 4 heads appear when a fair coin is flipped four times, given that the first flip came up heads?

Many experiments only have two outcomes

Coin flips: Heads or tails?



Bit strings: 0 or 1?

P(x)

Predicates: T or F?

These types of experiments are called Bernoulli trials

Two outcomes:

- Success — Probability p
- Failure — Probability $q = 1 - p$

Many problems can be solved by examining the probability of k successes in an experiment consisting of mutually-independent Bernoulli trials

Example: Coin flips

Example: A coin is biased so that the probability of heads is 2/3. What is the probability that exactly four heads come up when the coin is flipped seven times, assuming that each flip is independent?

Solution:

- $2^7 = 128$ possible outcomes for seven flips
- There are C(7,4) ways that heads can be flipped four times
- Since each flip is independent, the probability of each of these outcomes is $(2/3)^4(1/3)^3$
- So, the probability of exactly 4 heads occurring in 7 flips of this biased coin is $C(7,4)(2/3)^4(1/3)^3 = 560/2187$

Probability of each tails combined using product rule

7 Choose 4 outcomes to make heads

Probability of each heads combined using product rule

This general reasoning provides us with a nice formula…

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

Proof:

- The outcome of n Bernoulli trials is an n-tuple $(t_1, t_2, \ldots, t_n)$
- Each t_i is either S (for success) or F (for failure)
- C(n,k) ways to choose k t_is to label S
- Since each trial is independent, the probability of each outcome with k successes and n-k failures is p^kq^{n-k}
- So, the probability of exactly k successes is $C(n,k)p^kq^{n-k}$. ❑

Notation: We denote the probability of k successes in n independent Bernoulli trials with probability of success p as b(k; n, p).

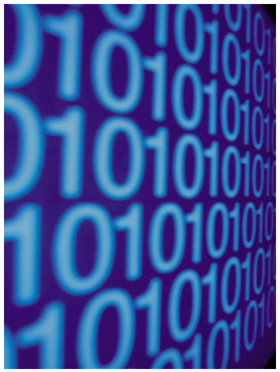
Bits (Again)

Example: Suppose that the probability that a 0 bit is generated is 0.9, that the probability that a 1 bit is generated is 0.1, and that bits are generated independently. What is the probability that exactly eight 0 bits are generated when ten random bits are generated?

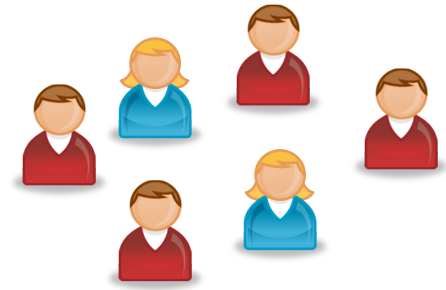
Solution:

- Number of trials
- Number of successes
- Probability of success
- Probability of failure
- Want to compute b(k; 10, 0.9)
- = $C(10, 8)0.9^8 0.1^2$
- = 0.1937102445

Many probability questions are concerned with some numerical value associated with an experiment

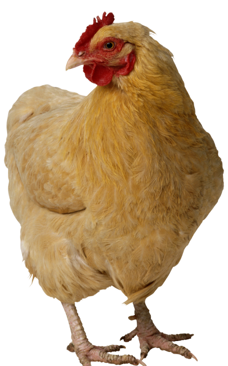


Number of 1 bits generated



Number of boys in a family

Beats per minute of a heart



Longevity of a chicken

Number of “heads” flips

What is a random variable?

Definition: A random variable is a function X from the sample space of an experiment to the set of real numbers **R**. That is, a random variable assigns a real number to each possible outcome.

Note: Despite the name, X is <u>not</u> a variable, and is <u>not</u> random. X is a function!

Example: Suppose that a coin is flipped three times. Let X(s) be the random variable that equals the numbers of heads that appear when s is the outcome. Then X(s) takes the following values:

- X(HHH) = 3
- X(HHT) = X(HTH) = X(THH) = 2
- X(TTH) = X(THT) = X(HTT) = 1
- X(TTT) = 0

Random variables and distributions

Definition: The distribution of a random variable X on a sample space S is the set of pairs (r, p(X=r)) for all r ∈ X(S), where p(X=r) is the probability that X takes the value r.

Note: A distribution is usually described by specifying p(X=r) for each r ∈ X(S)

Example: Assume that our coin flips from the previous slide were all equally likely to occur. We then get the following distribution for the random variable X:

- p(X=0) = 1/8
- p(X=1) = 3/8
- p(X=2) = 3/8
- p(X=3) = 1/8

Example: Rolling dice

Let X be the sum of the numbers that appear when a pair of fair dice is rolled. What are the values of this random variable for the 36 possible outcomes (i, j) where i and j are the numbers that appear on the first and second die, respectively?

Answer:

- X(1,1) = 2
- X(1,2) = X(2, 1) = 3
- X(1,3) = X(2,2) = X(3,1) = 4
- X(1,4) = X(2,3) = X(3,2) = X(4,1) = 5
- X(1,5) = X(2,4) = X(3,3) = X(4,2) = X(5,1) = 6
- X(1,6) = X(2,5) = X(3,4) = X(4,3) = X(5,2) = X(6,1) = 7
- X(2,6) = X(3,5) = X(4,4) = X(5,3) = X(6,2) = 8
- X(3,6) = X(4,5) = X(5,4) = X(6,3) = 9
- X(4,6) = X(5,5) = X(6,4) = 10
- X(5,6) = X(6,5) = 11
- X(6,6) = 12

Sometimes probabilistic reasoning can lead us to some interesting and unexpected conclusions…

Question: How many people need to be in the same room so that the probability of two people sharing the same birthday is greater than 1/2?

Assumptions:

1. There are 366 possible birthdays
2. All birthdays are equally likely to occur
3. Birthdays are independent

Solution tactic:

- Find the probability p_n that the n people in a room all have different birthdays
- Then compute $1-p_n$, which is the probability that at least two people share the same birthday

Let's figure this out…

Let's assess probabilities as people enter the room

- Person 1 clearly doesn't have the same birthday as anyone else in the room
- P_2 has a different birthday than P_1 with probability 365/366
- P_3 has a different birthday than P_1 and P_2 with probability 364/366
- …

In general, P_j has a different birthday than P_1, P_2, …, P_{j-1} with probability [366-(j-1)]/366 = (367-j)/366

Recall that p_n is the probability that n people in the room all have different birthdays. Using our above observations, this means:

$$p_n = \frac{365}{366}\frac{364}{366}\frac{363}{366}\cdots\frac{367-n}{366}$$

But we're interested in $1-p_n$ …

$$1 - p_n = 1 - \frac{365}{366}\frac{364}{366}\frac{363}{366}\cdots\frac{367-n}{366}$$

To check the minimum number of people need in the room to ensure that $p_n > 1/2$, we'll use trial and error:

- If $n = 22$, then $1 - p_n \approx 0.475$
- If $n = 23$, then $1 - p_n \approx 0.506$

So, you need only 23 people in a room to have a better than 50% chance that two people share the same birthday!

In-class exercises

Problem 3: What is the probability that exactly 2 heads occur when a fair coin is flipped 7 times?

Problem 4: Consider a game between Alice and Bob. Over time, Alice has been shown to win this game (against Bob) 75% of the time. If Alice and Bob play 6 games in a row, what is the probability that Alice wins every game?

Problem 5: Consider generating a uniformly-random 4-character bit string. Also consider R, a random variable that measures the longest run of 1 bits in the generated string. Determine the distribution of R.

Final Thoughts

- Today we covered
  - Conditional probability
  - Independence
  - Bernoulli trials
  - Random variables
  - Probabilistic analysis

- Next time:
  - Bayes' Theorem (Section 7.3)

The proof…

$P(n) \equiv p(\cup_{i=1}^{n} E_i) = \sum_{i=1}^{n} p(E_i)$

Base case: P(2): Let E_1, E_2 be disjoint events.
By definition, $p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2)$.
Since $E_1 \cap E_2 = \emptyset$, $p(E_1 \cup E_2) = p(E_1) + p(E_2)$

I.H.: Assume that P(k) holds for an arbitrary integer k

Inductive step: We will now show that $P(k) \rightarrow P(k+1)$

- Consider $E = E_1 \cup E_2 \cup \ldots \cup E_k \cup E_{k+1}$
- Let $J = E_1 \cup E_2 \cup \ldots \cup E_k$, so $E = J \cup E_{k+1}$
- $p(E) = p(J \cup E_{k+1})$ by definition of E
- $= p(J) \cup p(E_{k+1})$ by I.H.
- $= p(E_1 \cup E_2 \cup \ldots \cup E_k) \cup p(E_{k+1})$ by definition of J
- $= p(E_1) \cup p(E_2) \cup \ldots \cup p(E_k) \cup p(E_{k+1})$ by I.H.

Conclusion: Since we have proved the base case and the inductive case, the claim holds by mathematical induction ❑