COMP170
Discrete Mathematical Tools for Computer Science

Lecture 14 *Version 5 Last updated, Nov 17 2005*

Discrete Math for Computer Science
K. Bogart, C. Stein and R.L. Drysdale
Section 5.1, pp. 213-221

Introduction to Probability

- Why Study Probability?

- Probability Spaces and Distributions
- Complementary Probabilities
- Probability and Hashing
- The Uniform Probability Distribution

Why Study Probability?

In Computer Science we often deal with random events. Some involve randomness imposed from the outside, e.g., networking, when requests from computers on the network enter the network at "random" time.

Some involve randomness that we introduce , e.g., hashing, which is a techinique often used to compactly store information in a computer for later quick retrieval

Studying the performance of computer systems in the presence of these types of randomness, requires understanding randomness, which is the study of probability.

Hashing

Imagine a company with one hundred employees. There's not enough room in the main office to give each one a mailbox. So, instead, they have one mailbox for each letter of the alphabet. When a letter arrived, it gets put into the box corresponding to the recipients last name. This is an example of a Hash Function.

Hashing is a very common programming tool that permits concise storage of data with quick lookups. The general idea is that we have a set of records that need to be stored. Each record is addressed using its key, e.g., name or ID number.

The records are stored in a table. each table location, called a bucket or slot, holds a list of records. We are also given a hash function $h(x)$. A record with key key is stored in the bucket with index $h(key)$.

Hashing

Keys are integers.

Our Hash Function:

$$h(x) = x \bmod m \qquad m = 8$$

Data (with Keys)

$4 \quad 7 \quad 10 \quad 13 \quad 15$

When searching for a record you might have to look at *every* record in the appropriate bucket, so

Good hash function spreads keys evenly among buckets.

Hash Table T

0	
1	
2	10
3	
4	4
5	13
6	
$m-1 = 7$	$7 \quad 15$

buckets/ slots

collision!

Given: Table with 100 buckets and 50 keys.

Is it possible that all 50 keys are assigned to *same* bucket? ← bad case

- Using **good** hash function, you'd never see this in a million years.
- Actually, you also wouldn't see that all the keys hash into *different* locations. ← best case

How can we calculate likelihood of such events?

$\longrightarrow$ Study of Probability

Introduction to Probability

- Why Study Probability?

- Probability Spaces and Distributions
- Complementary Probabilities
- Probability and Hashing
- The Uniform Probability Distribution

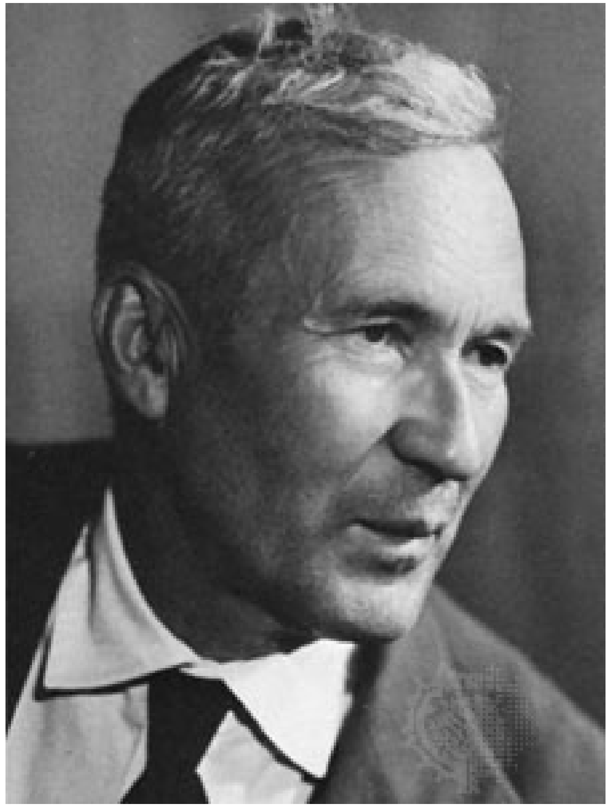
In Probability Theory we need to define
three related different concepts:

- The underlying Sample Space and
  Elements (Outcomes) in the sample space
- An event in the Sample Space
- The Weight of an element in the sample space
  Gives a Probability Distribution (Measure)

Andrei Nikolaevich Kolmogorov

Russian Mathematician

b. 1903. d. 1987



The birth of probabiity theory is often dated to 1654, when Pascal and Fermat, trying to solve a gambling problem, developed the fundamentals.

It wasn't until the work of Kolmogorav in 1933, though, that we had a "definition of probability that is precise enough for use in mathematics, yet comprehensive enough to be applicable to a wide range of phenomena".

Ref. A Short History of Probability by Tom Apostol
http://www.cc.gatech.edu/classes/cs6751_97_winter/Topics/stat-meas/probHist.html

Sample Space

$=$ set of possible outcomes of a process.

Example 1:

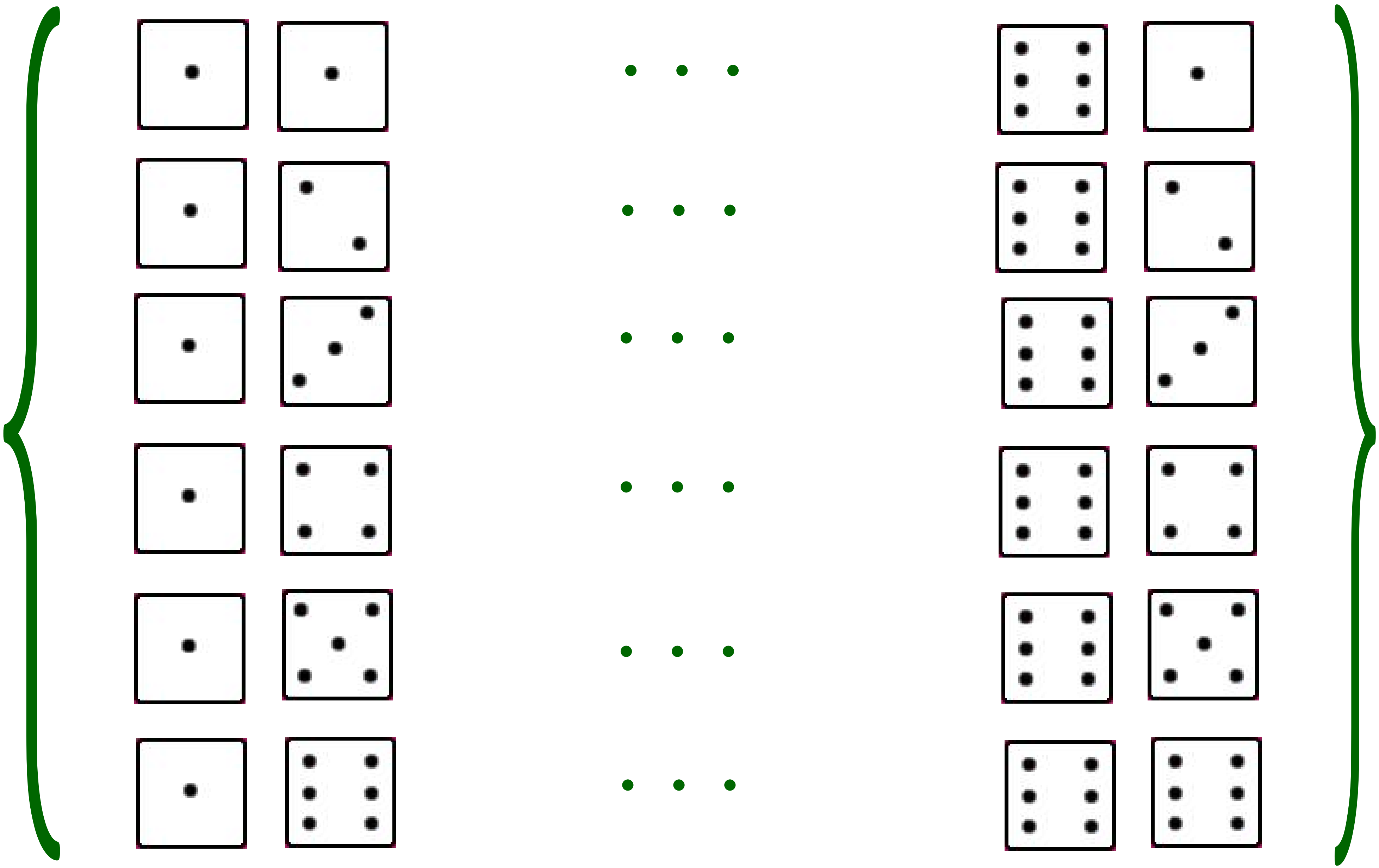
Professor starts each class
with a 3-question true-false quiz;
The sample space of all possible patterns
of correct answers is

$$S = \{\mathtt{TTT},\ \mathtt{TTF},\ \mathtt{TFT},\ \mathtt{FTT},\ \mathtt{TFF},\ \mathtt{FTF},\ \mathtt{FFT},\ \mathtt{FFF}\}.$$

Note: TTT corresponds to all answers being true, etc..

Example 2: Throwing two 6-sided dice.

The sample space contains 36 different outcomes:



Example 3:

Throw a two-sided (Heads, H, and Tails, T) coin, until you see a Head. This sample space contains an *infinite* number of outcomes.

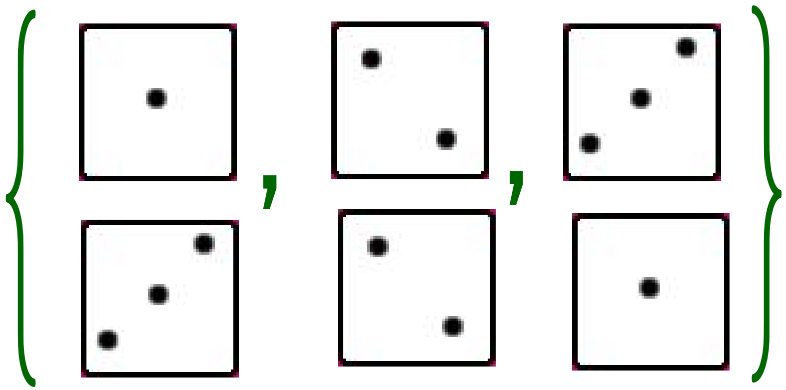
$$\{\mathtt{H},\ \mathtt{TH},\ \mathtt{TTH},\ \mathtt{TTTH},\ \mathtt{TTTTH},\ \mathtt{TTTTTH},\ \ldots\}$$

Event

$=$ some subset of outcomes in sample space.

Examples:

1) Event of the first two answers being true: {TTT, TTF}.

2) Event of the dice adding up to 4: 

3) Event of a Head occurring in the first 3 flips: {H, TH, TTH}.

Computing Probabilities:

- We must assign **probability weight** $P(x)$ to each element (outcome) x of the sample space

- This weight represents the relative likelihood of that outcome.

Rules for assigning weights

- weights must be nonnegative numbers
- sum of the weights of all the elements in a sample space must be 1

We now Define:

The **probability** $P(E)$ **of event** E is:
sum of the weights of the elements of E.

$$P(E) = \sum_{x:x \in E} P(x)$$

read: "The probability of event E *is the sum, over all* x *such that* x *is in* E*, of* $P(x)$*."*

A Probability function P on a sample space S must satisfy the following rules:

- $P(A) \geq 0$ for all $A \subseteq S$.
- $P(S) = 1$.
- $P(A \cup B) = P(A) + P(B)$
  for any two disjoint events A and B.

A function P satisfying these rules is called a **probability distribution** *or a* **probability measure**.

Example 1a:

The professor's three-question quiz with sample space

$$\{\texttt{TTT},\ \texttt{TTF},\ \texttt{TFT},\ \texttt{FTT},\ \texttt{TFF},\ \texttt{FTF},\ \texttt{FFT},\ \texttt{FFF}\}.$$

Suppose that each sequence of T and F is equally likely. We then want to assign each outcome the same *(uniform)* probability weight. Since the sum of the weights must add up to 1, we assign each of the 8 outcomes a weight of $\frac{1}{8}$.

The probability of event E is sum of weights of its elements:

The event "the first answer is true" is {TTT, TTF, TFT,TFF} so its probability is

$$\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{1}{2}$$

Example 1b:

The professor's three-question quiz with sample space

$$\{\texttt{TTT},\ \texttt{TTF},\ \texttt{TFT},\ \texttt{FTT},\ \texttt{TFF},\ \texttt{FTF},\ \texttt{FFT},\ \texttt{FFF}\}.$$

Suppose that each sequence of T and F is equally likely. We then want to assign each outcome the same *(uniform)* probabiity weight. Since the sum of the weights must add up to 1, we assign each of the 8 outcomes a weight of $\frac{1}{8}$.

The probability of event E is sum of weights of its elements:

Event
"there is exactly one true answer"
is {TFF, FTF, FFT}
so its probability is

$$\frac{1}{8}+\frac{1}{8}+\frac{1}{8}=\frac{3}{8}$$

Example 1c:

Now suppose that the professor has a bias towards true answers and a T is more likely than a F. Let s assume that a T has a probability of $2/3$ and a F has a probability of $1/3$. Then, we will see later, we will get the following different weights on the *same* sample space as before

$$\{\mathtt{TTT},\ \mathtt{TTF},\ \mathtt{TFT},\ \mathtt{FTT},\ \mathtt{TFF},\ \mathtt{FTF},\ \mathtt{FFT},\ \mathtt{FFF}\}.$$

$$\frac{8}{27} \quad \frac{4}{27} \quad \frac{4}{27} \quad \frac{4}{27} \quad \frac{2}{27} \quad \frac{2}{27} \quad \frac{2}{27} \quad \frac{1}{27}$$

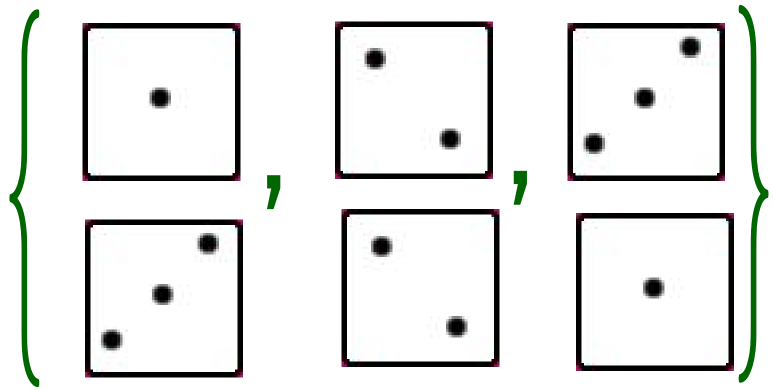
These add up to 1, so this is a legal probability distribution

The event "the first answer is true", {TTT, TTF, TFT,TFF}, now no longer has probability $1/2$. With these new probability weights, its probability is

$$\frac{8}{27}+\frac{4}{27}+\frac{4}{27}+\frac{2}{27}=\frac{2}{3}$$

Example 2:

There are 36 possible outcomes when rolling two dice. If the dice are *fair*, then each outcome should be equally likely and each outcome will have weight $1/36$.

The event of the dice adding up to 4 is: 

There are three outcomes in this event, so its probability is

$$\frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{1}{12}$$

Example 3:

When throwing a coin until the first H is seen,
the sample space is

$$\{\texttt{H},\ \texttt{TH},\ \texttt{TTH},\ \texttt{TTTH},\ \texttt{TTTTH},\ \texttt{TTTTTH},\ \ldots\}$$

$$\frac{1}{2} \quad \frac{1}{4} \quad \frac{1}{8} \quad \frac{1}{16} \quad \frac{1}{32} \quad \frac{1}{64}$$

Assuming that the coin is *fair*, a good probability distribution is

$$\text{P(Outcome of } (i-1) \text{ Ts followed by a H)} = \frac{1}{2^i}$$

Note that this is a legal probability distribution, since

$$\sum_{i=1}^{\infty} \frac{1}{2^i} = 1$$

Introduction to Probability

- Why Study Probability?

- Probability Spaces and Distributions

- Complementary Probabilities

- Probability and Hashing

- The Uniform Probability Distribution

Let's flip a fair coin 3 times. The sample space contains 8 possible outcomes. Since the coin is fair, each outcome should be equally likely. So, each one should have a weight of $\frac{1}{8}$.

$$S = \{\underbrace{\boxed{TTT, TTH, THT, THH, HTT, HTH, HHT,}}_{A}\underbrace{\boxed{HHH}}_{B}\}$$

Let A be event that there is at least one T.

$$P(A) = 7 \cdot \frac{1}{8}$$

Let B be event that there are no Ts.

$$P(B) = \frac{1}{8}$$

Note that $S = A \cup B$ (or $B = S - A$) and that $P(A) + P(B) = 1$.

Complementary Probabilities

- The **Complement** $S - E$
  of event E in sample space S
  is the set of all outcomes in S
  except those in E.

- Two events E and F are **complementary**
  (in S) if E is the complement of F
  in the sample space S.

Theorem 5.1:
If two events E and F are complementary, then
$P(E) = 1 - P(F)$.

Proof:

Sum of all probabilities of all elements of sample space is 1.

Because we can partition this sum into
sum of the probabilities of elements of E plus
sum of probabilities of elements of F,
we have

$$P(E) + P(F) = 1$$

$$\Rightarrow P(E) = 1 - P(F).$$

Introduction to Probability

- Why Study Probability?

- Probability Spaces and Distributions

- Complementary Probabilities

- Probability and Hashing

- The Uniform Probability Distribution

Probability and Hashing

What is a good sample space when we hash a list of n keys into a hash table with m locations?

> Sample space:
> Set of n-tuples $(\ell_1, \ell_2, \ldots, \ell_n)$ of numbers between 1 and m.

ℓ_1 is location that 1^{st} key hashes to
ℓ_2 is location that 2^{nd} key hashes to
ℓ_3 is location that 3^{rd} key hashes to
and so on ....

Example: $3, 12, 15, 8, 11, 5$

is outcome $(3, 4, 7, 0, 3, 5)$

0	8
1	
2	
3	$3 \quad 11$
4	12
5	5
6	
7	15

What is a good sample space when we hash a list of n keys into a hash table with m locations?

Sample space:
Set of n-tuples $(\ell_1, \ell_2, \ldots, \ell_n)$ of numbers between 1 and m.

The size of the sample space is m^n (why?).

Weight function: Assuming that hash function is "random" then every n-tuple is equally likely. So, every n-tuple should have (the same) weight $1/m^n$.

Hash a list of 3 keys into hash table with 20 locations.

What is the probability of event A in which all three keys hash to different locations, i.e., no collisions?

Sample space is 20^3 triples, (ℓ_1, ℓ_2, ℓ_3), in which each ℓ_i is in $[1, 2, \ldots 20]$.

Each triple has weight $1/20^3$.

In event A, ℓ_1 has 20 choices, ℓ_2 has 19 choices and ℓ_3 has 18 choices.

So, total number of triples in A is $20 \cdot 19 \cdot 18$

$$\Rightarrow \boxed{P(A) = \frac{20 \cdot 19 \cdot 18}{20^3} = .855}$$

Hash a list of 3 keys into hash table with 20 locations.

What is the probability of event B in which hashing the three keys causes a collision?

A collision can occur in two different ways. Either (i) two keys could hash to the same bucket and the third key hashes to a different bucket or, (ii) all keys could hash to the same bucket.

We could calculate the probabilities of all cases and add them together but it is easier to just note (why?) that event B is just the complement of event A on the previous slide so

$$P(B) = 1 - P(A) = 1 - .855 = .145$$

Hash a list of n keys into hash table with 20 locations.

What is the probability of event A in which all n keys hash to different locations, i.e., no collisions?

Sample space is 20^n triples, $(\ell_1, \ell_2, \ldots \ell_n)$, in which each ℓ_i is in $[1, 2, \ldots 20]$.

Each n-tuple has weight $1/20^n$.

Total number of n-tuples in A is $20 \cdot 19 \cdots (20 - n + 1) = 20^{\underline{n}}$

$$\boxed{\Rightarrow \ P(A) = \frac{20^{\underline{n}}}{20^n}}$$

Hash a list of n keys into hash table with 20 locations.

What is the probability of event B in which hashing the n keys causes at least one collision?

Since events A and B are complementary

$$P(B) = 1 - P(A) = 1 - \frac{20^{\underline{n}}}{20^n}$$

n		Probability of No Collisions
1		1
2		.95
3		.855
4		.72675
5		.5814
6		.43605
7		.305235
8		.19840275
9		.11904165
10		.065472908
11		.032736454
12		.014731404
13		.005892562
14		.002062397
15		.000618719
16		.00015468
17		.0000309359
18		.00000464039
19		.000000464039
20		.000000023202

Probabilities that all elements of set hash to different entries of hash table of size 20 is

$$p_n = \frac{20^{\underline{n}}}{20^n}.$$

Since

$$p_{n+1} = p_n \frac{20-n}{20} < p_n,$$

p_n decreases as n increases

Introduction to Probability

- Why Study Probability?

- Probability Spaces and Distributions
- Complementary Probabilities
- Probability and Hashing
- The Uniform Probability Distribution

The Uniform Probability Distribution

P is a **uniform probability measure** or a **uniform probability distribution** if it assigns *same* probability to all members of sample space.

Theorem 5.2:
Suppose P is uniform probability measure defined on sample space S. Then for any event E,

$$P(E) = \frac{|E|}{|S|},$$

which is size of E divided by size of S.

Note: We have implicitlly used this theorem many times already

Proof:

Let $S = \{x_1, x_2, \ldots, x_{|S|}\}$.

Because P is uniform probability measure, there must be some p such that $P(x_i) = p$ for each $x_i \in S$.

Combining this with 2^{nd} and 3^{rd} probability rules, we obtain

$$\begin{aligned} 1 &= P(S) \\ &= P(x_1 \cup x_2 \cup \cdots \cup x_{|S|}) \\ &= P(x_1) + P(x_2) + \cdots + P(x_{|S|}) \\ &= p|S|. \end{aligned}$$

$$\Rightarrow p = \frac{1}{|S|}$$

E is a subset of S with $|E|$ elements and, therefore,

$$P(E) = \sum_{x_i \in E} p(x_i) = |E|p.$$

Combining these equations gives

$$P(E) = |E|p = |E|(1/|S|) = |E|/|S|.$$

What is the probability of event E of seeing an odd number of heads in three tosses of a fair coin?

Sample Space:

$$\texttt{S = \{TTT, } \boxed{\texttt{TTH, THT, HTT,}} \texttt{ THH, HTH, HHT,} \boxed{\texttt{HHH}}\texttt{\}.}$$

8 sequences in sample space S.

3 sequences with one H **1 sequence with three H's**

Thus, 4 sequences in event E.

Use Theorem 5.2

$\Rightarrow$ probability is $4/8 = 1/2$ by Theorem 5.2.

Theorem 5.2 applies only to probabilities that come from the **equiprobable weighting** function.

In Example 1c, we already saw one example of a non-equiprobable weighting function. We will now see another.

Sample Space:	$\{0,$	$1,$	$2,$	$3\}$
with weights	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Theorem 5.2 doesn't apply to this distribution. For example, let E be the event that the outcome is not positive. Then $E = \{0\}$ but

$$P(E) = \frac{1}{8} \neq \frac{1}{4} = \frac{|E|}{|S|}$$

Where does this strange distribution come from?

Consider $\{0, 1, 2, 3\}$ as the sample space for flipping fair coin three times and counting number of H.

Determine appropriate weights $P(0), P(1), P(2), P(3)$.

{TTT, TTH, THT, HTT, THH, HTH, HHT, HHH}.

1 way to get no H.

3 ways to get 1 H. $P(1)$ should be 3 times $P(0)$

3 ways to get 2 H's. $P(2)$ should be 3 times $P(0)$

1 way to get 3 H's. $P(3)$ should be equal $P(0)$

We need

$$P(1) = 3P(0) \quad P(2) = 3P(0) \quad P(3) = P(0)$$

All weights add to 1:

$$\begin{aligned} 1 &= P(0) + P(1) + P(2) + P(3). \\ &= P(0) + 3P(0) + 3P(0) + P(0) = 8P(0) \end{aligned}$$

The **unique** solution:

$$\boxed{P(0) = \frac{1}{8}, \quad P(1) = \frac{3}{8}, \quad P(2) = \frac{3}{8}, \quad P(3) = \frac{1}{8}}$$

which is exactly the non-uniform distribution we just saw.