

## Section #1: Analytic Probability

---

### 1 Generative Processes: The Birthday Problem

**Preamble:** When solving a counting problem, it can often be useful to come up with a generative process, a series of steps that “generates” examples. A correct generative process to count the elements of set  $A$  will (1) *generate every element of  $A$*  and (2) *not generate any element of  $A$  more than once*. If our process has the added property that (3) *any given step always has the same number of possible outcomes*, then we can use the product rule of counting.

**Example:** Say we want to count the number of ways to roll two (distinct) dice where one die is even and one die is odd. Our process could be: (1) choose a number for the first die, (2) choose a number of opposite parity for the second die. Since the first step has 6 options and the second step has 3 options regardless of the outcome of the first step, the number of possibilities is  $6 * 3 = 18$ .

**Problem:** Assume that birthdays happen on any of the 365 days of the year with equal likelihood (we’ll ignore leap years).

- a. What is the probability that of the  $n$  people in class, at least two people share the same birthday?
- b. What is the probability that this class contains exactly one pair of people who share a birthday?

### 2 Permutations and Combinations: Flipping Coins

**Preamble:** One thing that students often find tricky when learning combinatorics is how to figure out when a problem involves permutations and when it involves combinations. Naturally, we will look at a problem that can be solved with both approaches. Pay attention to what parts of your solution represent distinct objects and what parts represent indistinct objects.

**Problem:** We flip a fair coin  $n$  times, hoping (for some reason) to get  $k$  heads.

- a. How many ways are there to get exactly  $k$  heads? Characterize your answer as a *permutation* of H’s and T’s.
- b. For what  $x$  and  $y$  is your answer to part a equal to  $\binom{x}{y}$ ? Why does this *combination* make sense as an answer?
- c. What is the probability that we get exactly  $k$  heads?

### 3 Bayes Rule: Song Identification

**Preamble:** In this class, seeing a problem written in English can often throw you off of its scent. In this problem, we will practice translating a problem from English to equations and then applying Bayes Rule, which you learned this week.

**Problem:** Shazam is an application which can predict what song is playing. Based on the frequency of requests it's been getting these days, Shazam has found that:

- 80% of songs are Hold Up by Beyonce
- 20% of songs are Can't Get Used to Losing You by Andy Williams

When a request is made, Shazam receives an audio sample that it uses to update its belief. From one particular audio sample,  $S$ , Shazam estimates that:

- $S$  would have a 50% chance of appearing if Hold Up were playing.
- $S$  would have a 90% chance of appearing if Can't Get Used to Losing You were playing.

What is the updated probability that the song is Hold Up given the audio sample heard? HINT: Define variables and write all of the information we have given to you in terms of those variables.

### 4 Probability Misunderstood: The Sally Clark Case

**Preamble:** Conditional probabilities are hard to interpret, especially if they are extremely close to zero or one. You should be careful about how you convey meaning to your audience, whether they are on a jury (as below) or are users of software that you have written.

**Problem:** Sally Clark was a British lawyer who was wrongly sentenced to life in prison in 1999 for the deaths of her two infant children. Her elder son Christopher died at age 11 weeks in December 1996 and her younger son Harry at 8 weeks in January 1998. At her trial, the defence argued that the deaths were due to sudden infant death syndrome (SIDS). Clark was convicted on the basis of testimony by pediatrician Sir Roy Meadow, who made the following argument:

- Hospital records show that the ratio of SIDS deaths to live births in affluent non-smoking families is about  $\frac{1}{8500}$ . (A live birth is a birth in which a child is born alive; not a still birth.)
- The chance of two SIDS deaths occurring in the same family is about  $\frac{1}{8500^2} \approx \frac{1}{73000000}$ .
- It is therefore extremely unlikely that Clark is innocent.

As a result of this prosecution, Clark spent more than 3 years in prison and was finally exonerated in 2003 after it was determined that Meadow's expert testimony was flawed. Two other women against whom Meadow provided expert testimony had their convictions overturned as well.

- a. Identify a flaw in Meadow's  $\frac{1}{73000000}$  figure.
- b. Even if we accept Meadow's  $\frac{1}{73000000}$  calculation as valid, what is wrong with a juror interpreting it as the probability of Clark's innocence?