# Stat 230: Probability

## Lecture 3

Jeremy VanderDoes

University of Waterloo

Friday, May 6th

# Review

## Example (Review Question)

After graduating from a university not as good as Waterloo, Kronk applied for a job. At the interview, Yzma asked 3 yes-no questions–which Kronk randomly guessed–and recorded the correctness of each answer. Give:

(1) Sample Space

(2) Event Kronk answered at least 2 correctly

(3) Probability Kronk answered at least 2 answers correctly

Although Yzma said he has a 40% chance, as Kronk leaves, he sees 3 other candidates and thinks he has 25% chance of being offered a job.

(4) Give the type of probability each used

# Review

Last time we talked about

- Equally likely sample spaces
- Counting: Addition and Multiplication Rules

Schedule for today:

- Review Questions
- Permutations
- Combinations

Reading: Middle of chapter 3

# Review

- Text size
- Axioms of Probability
- Quiz

# Review

## Example

Suppose that three of the numbers $\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ are selected at random **without** replacement, and then put together in the order they are drawn to form a three digit number. What is the probability that:

(1) The number is larger than 500?

(2) The number is even?

(3) The number is larger than 700, and is even?

# Counting

Often in counting or combinatorial calculations, the following two general things come up:

(1) Select a few individuals from a group and order them somehow
   - How many ways are there to make a lineup in Baseball if you have 15 players? (ways to select pitcher, shortstop, ...)

(2) Select a few individuals without regard to the order they are selected
   - How many ways are there to choose 9 of 15 Baseball players to play? (but not giving them positions)

The general thing 1 can be thought of as adding a step to general thing 2.

# Counting

## Example

Suppose that the 5 members of the UofW hardcour parkour club must select a president and vice president. How many ways can they do this?

# Counting

## Definition

*Given $n$ distinct objects, a* **permutation** *of size $k$ is an ordered subset of $k$ of the individuals. The number of permutations of size $k$ taken from $n$ objects is denoted $n^{(k)}$ and*

$$n^{(k)} = n(n-1)\dots(n-k+1) = \frac{n!}{(n-k)!}.$$

The tricky part of this definition is the word "ordered". An ordering need not be numerical, for example assigning labels like "President" and "Vice-President" has the effect of ordering the individuals.

# Counting

## Example

Suppose instead that the 5 members of the UofW hardcour parkour club must instead simply select two members to be on the budget committee. How many ways can they do this?

# Counting

## Definition

*Given $n$ distinct objects, a* **combination** *of size $k$ is an unordered subset of $k$ of the individuals. The number of combinations of size $k$ taken from $n$ objects is denoted $\binom{n}{k}$ and*

$$\binom{n}{k} = \frac{n^{(k)}}{k!} = \frac{n!}{(n-k)!k!}.$$

Note: We assume that $0! = 1$ and hence $\binom{n}{0} = 1$.

# Counting

## Example

Suppose you have 20 distinct books, 7 of which are written by Mark Twain.

(1) How many ways can you arrange 12 books on a shelf if the order they are on the shelf matters?

(2) How many ways can you arrange 12 books on a shelf if exactly 3 of them must be Mark Twain books?

(3) A monkey picks books at random from the 20 books and puts them on the shelf until it contains 12 books. What is the probability that at least 3 of the books on the shelf are written by Mark Twain?

# Birthday Problem

## Example (The Birthday Problem)

Suppose a room contains $n$ people. What is the probability at least two people in the room share a birthday?

**Assumption**: Suppose that each of the $n$ people is equally likely to have any of the 365 days of the year as their birthday, so that all possible combinations of birthdays are equally likely.

# Birthday Problem

**Calculation**:

$$P(A_{100}) = .9999997, \quad P(A_{23}) = .5073$$

**Takeaway**: A room containing 23 or more randomly selected individuals most likely contains two that share a birthday.

**Schur's Convexity**: If $P(A_n^*)$ denotes the probability that at least two of $n$ share a birthday for randomly selected individuals according to any other distribution on Birthdays (e.g. other than the "equally likely" case): $P(A_n^*) \geq P(A_n)$.

# Birthday Problem

Approximation: $1 - x \approx e^{-x}$ for small $x$, and hence

$$P(A_n) = 1 - \prod_{i=0}^{n-1} \left(1 - \frac{i}{365}\right) \approx 1 - \prod_{i=0}^{n-1} e^{-i/365}$$
$$= 1 - e^{-\frac{1}{365}\sum_{i=1}^{n-1} i} \approx 1 - e^{\frac{-n^2}{2\times 365}}$$

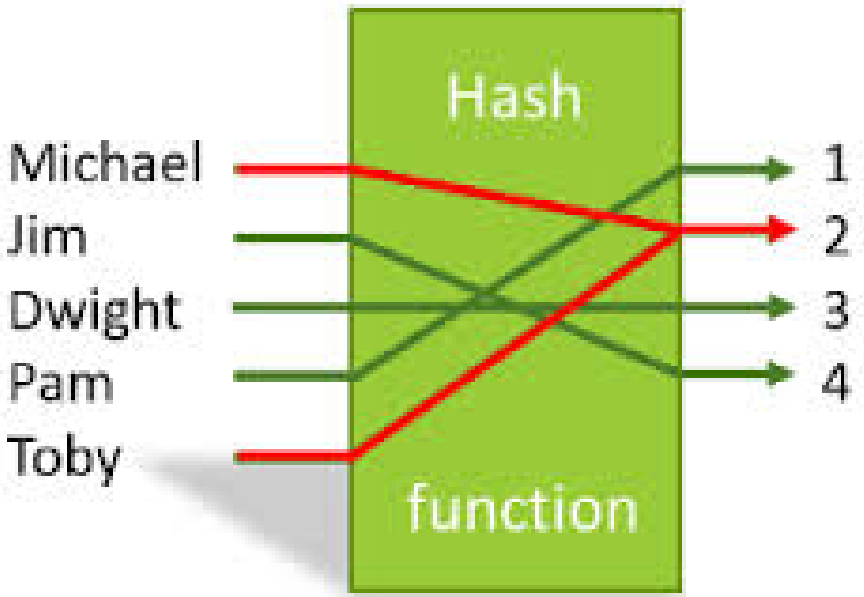
So if $n(p)$ = number of individuals required so that the probability of at least one match is $p$

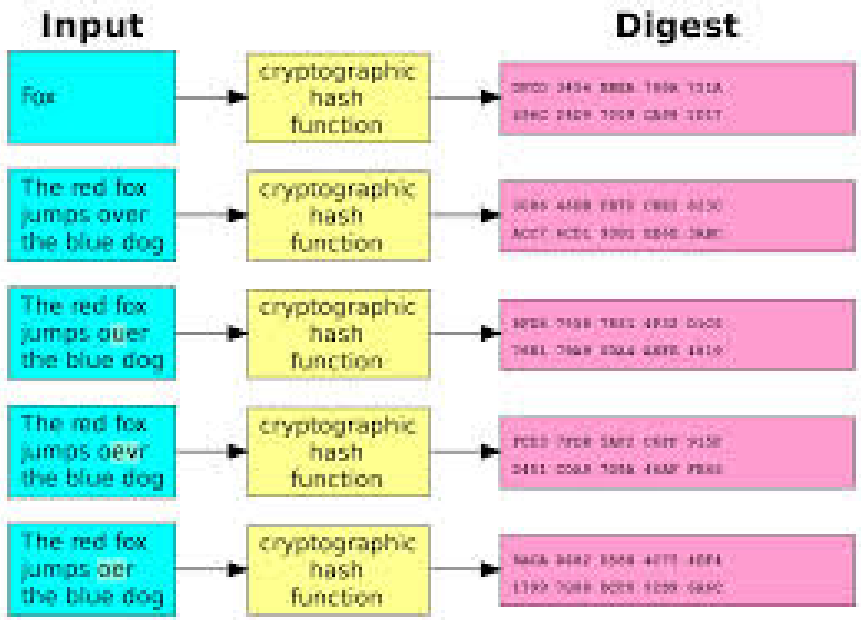
$$n(p) \approx \sqrt{2 \times 365 \log\left(\frac{1}{1-p}\right)}$$

Example: $n(.5) = 22.49$.

# Birthday Attack

A hash function is a function $f$ that maps data (for example a character string) to a set of hash values (for example an encrypted character string): Typically the data space is larger than the hash value space, and so it is possible to find two data points $x_1 \neq x_2$ so that $f(x_1) = f(x_2)$. Such an instance is called a collision.



# Birthday Attack

If the size of the hash space is $H$, then there is an approximately 50-50 chance that $\sqrt{2 \times H \log\left(\frac{1}{1-.5}\right)}$ data points will have a hash collision.

**Application**: If the hash space is comprised of 64-bit signatures, then $H = 2^{64} \approx 1.844674 \times 10^{19}$, and $\sqrt{2 \times H \log\left(\frac{1}{1-.5}\right)} = 5.05 \times 10^{9}$, which is a manageable search. This number is referred to as the "Birthday Bound"

# Birthday Attack

**Nefarious Application**: Consider generating messages with meanings $m_1$ and $m_2$, where $m_1$ is "expected" and $m_2$ is "fraudulent". By performing permutations of the wording/punctuation, one can come up with messages $m_1'$ and $m_2'$ with the same hash values...