1179: Probability
Lecture 5 — Bayes' Rule, Independence, and Combinatorics

Ping-Chun Hsieh (謝秉均)

September 29, 2021

Announcement

- Online office hour for today, 1pm-1:30pm:
  - https://nycu.webex.com/nycu/j.php?MTID=ma2106f2503f60807a6dedb2d5d777756 (same as the Webex link for the lectures)

About Problem 3 of HW1

Problem 3 (Continuity of Probability Functions) (12+12=24 points)

(a) Let $A_1, A_2, A_3, \cdots$ be a countably infinite sequence of events. Prove that if $\sum_{n=1}^{\infty} P(A_n) < \infty$, then $P(\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n) = 0$. This property is known as the *Borel-Cantelli Lemma*. (Hint: Consider the continuity of probability function for $\bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n$ and then apply the union bound)

Quick Review

- What is "reduction of sample space" (or conditional universe)?
- Multiplication rule?
- Total probability theorem?

1. $P(A|B) \geq 0$
2. $P(\Omega|B) = 1$
3. $A_1, A_2, A_3 \cdots$ mutually exclusive events

$$P\left(\bigcup_{i=1}^{\infty} A_i \,\middle|\, B\right) = \sum_{i=1}^{\infty} P(A_i|B)$$

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1)\cdot P(A_2|A_1)\cdot P(A_3|A_1 \cap A_2)\cdots P(A_n|A_1 \cap \cdots A_{n-1})$$

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + \cdots + P(A_n \cap B)$$

$$= P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \cdots + P(A_n)\cdot P(B|A_n)$$

This Lecture

1. Conditioning and Independence

2. Review: Combinatorial Methods

- Reading material: Chapter 2 and 3.1~3.5

Example: Gambler's Ruin

$P_n = \frac{1}{2} P_{n-1} + \frac{1}{2} P_{n+1}$, for any $n = 1, \cdots, (a+b)-1$,

$P_3 = \frac{1}{2} P_2 + \frac{1}{2} P_4$,

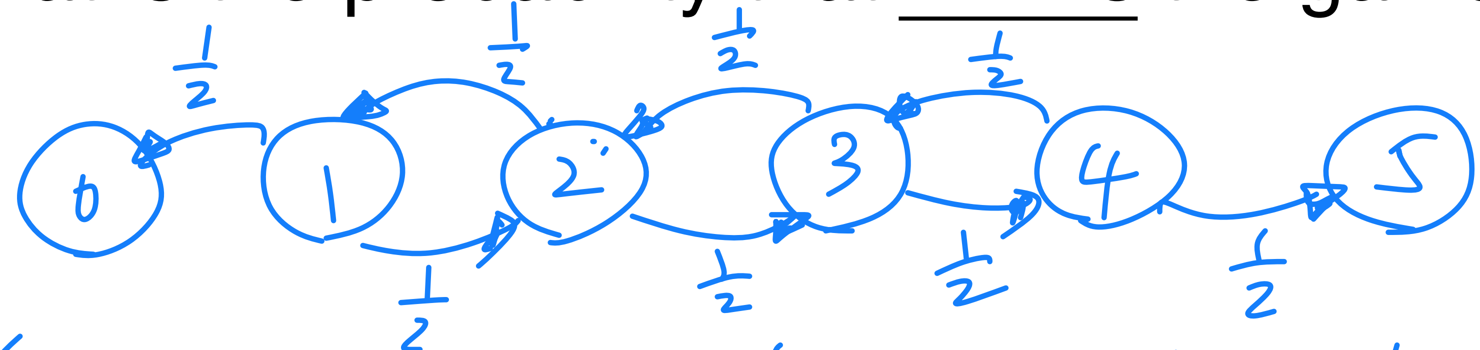
$P_4 = \frac{1}{2} P_3 + \frac{1}{2} P_5$.

$P_1 = \frac{1}{2} P_0 + \frac{1}{2} P_2$

a $\quad$ b

- Example: Two gamblers A and B keep tossing a fair coin
  - If "head" occurs, A pays \$1 to B; otherwise, B pays \$1 to A
  - Initially, A has 2 dollars, and B has 3 dollars
  - The game ends when either A or B has zero dollar
  - What is the probability that A wins the game?

A



$\frac{a}{a+b}$

$$\begin{cases} P_4 - P_3 = P_3 - P_2 \\ P_5 - P_4 = P_4 - P_3 \\ P_2 - P_1 = P_1 - P_0 \end{cases}$$

$P_2 = P(\text{ultimately A wins the game} \mid \text{A starts with 2 dollars})$

$P_n = P(\quad " \quad\quad " \quad \mid \text{A starts with } n \text{ dollars})$

$P_0 = 0, P_5 = 1$

$P_1 = \frac{1}{5}, P_2 = \frac{2}{5}, P_3 = \frac{3}{5}, P_4 = \frac{4}{5}$

$P_2 = P(\text{1st toss is H}) \cdot P(\text{A ultimately wins} \mid \text{A starts with 1 dollar}) + P(\text{1st toss is T}) \cdot P(\text{A wins} \mid \text{A starts with 3 dollars})$

$= \frac{1}{2} P_1 + \frac{1}{2} P_3 \Rightarrow P_3 - P_2 = P_2 - P_1$

A_1, A_2 partition

$$\left(\begin{array}{l} A_1 = \{ \text{1st toss is } H \} \\ A_2 = \{ \text{1st toss is } T \} \end{array} \right)$$

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B)$$

$$= P(A_1) \cdot P(B \mid A_1) + P(A_2) \cdot P(B \mid A_2)$$

2

Bayes' Rule

Tool #3: Bayes' Rule

"prior" distribution" $P(A_i)$

Theorem: Let $A_1, A_2, \cdots, A_n$ be mutually exclusive events that form a partition of Ω, and $P(A_i) > 0$, for all i. Then, for any event B, we have "likelihood"

prior

$$P(A_i \mid B) = \frac{P(A_i)P(B \mid A_i)}{P(B)} = \frac{P(A_i)P(B \mid A_i)}{P(A_1)P(B \mid A_1) + \cdots + P(A_n)P(B \mid A_n)}$$

"posterior" "evidence" Total probability theorem

A_1, A_2, A_3

- Why is Bayes' rule useful? → Inference

"Bayesian inference"

$P(A_1 \mid B)$ ✓

$P(A_2 \mid B)$ ✓

$P(A_3 \mid B)$ ✓

Bayesian Inference: Crush and Dates

$$P(A_i \mid B) = \frac{P(A_i)P(B \mid A_i)}{P(B)} = \frac{P(A_i)P(B \mid A_i)}{P(A_1)P(B \mid A_1) + \cdots + P(A_n)P(B \mid A_n)}$$

- Example: Bill has a crush on Amy, and Bill wants to ask Amy out to see whether Amy likes him or not.
  - $A_1 = \{\text{Amy likes Bill}\}$, $A_2 = \{\text{Amy does not like Bill}\}$
  - $B = \{\text{Amy looks happy during the date}\}$
  - Suppose $P(A_1) = P(A_2) = 0.5$, $P(B \mid A_1) = 0.9$, and $P(B \mid A_2) = 0.3$
  - What are $P(A_1 \mid B)$ and $P(A_1 \mid B^c)$?

A_1, A_2 form a partition

$$P(A_1 \mid B) = \frac{P(A_1) \cdot P(B \mid A_1)}{P(A_1) \cdot P(B \mid A_1) + P(A_2) \cdot P(B \mid A_2)} = \frac{0.5 \times 0.9}{0.5 \times 0.9 + 0.5 \times 0.3} = 0.75$$

$$P(A_1 \mid B^c) = \frac{P(A_1) \cdot P(B^c \mid A_1)}{P(A_1) P(B^c \mid A_1) + P(A_2) \cdot P(B^c \mid A_2)} = \frac{0.5 \times 0.1}{0.5 \times 0.1 + 0.5 \times 0.7} = 0.125$$

Example: Answer an Exam Question

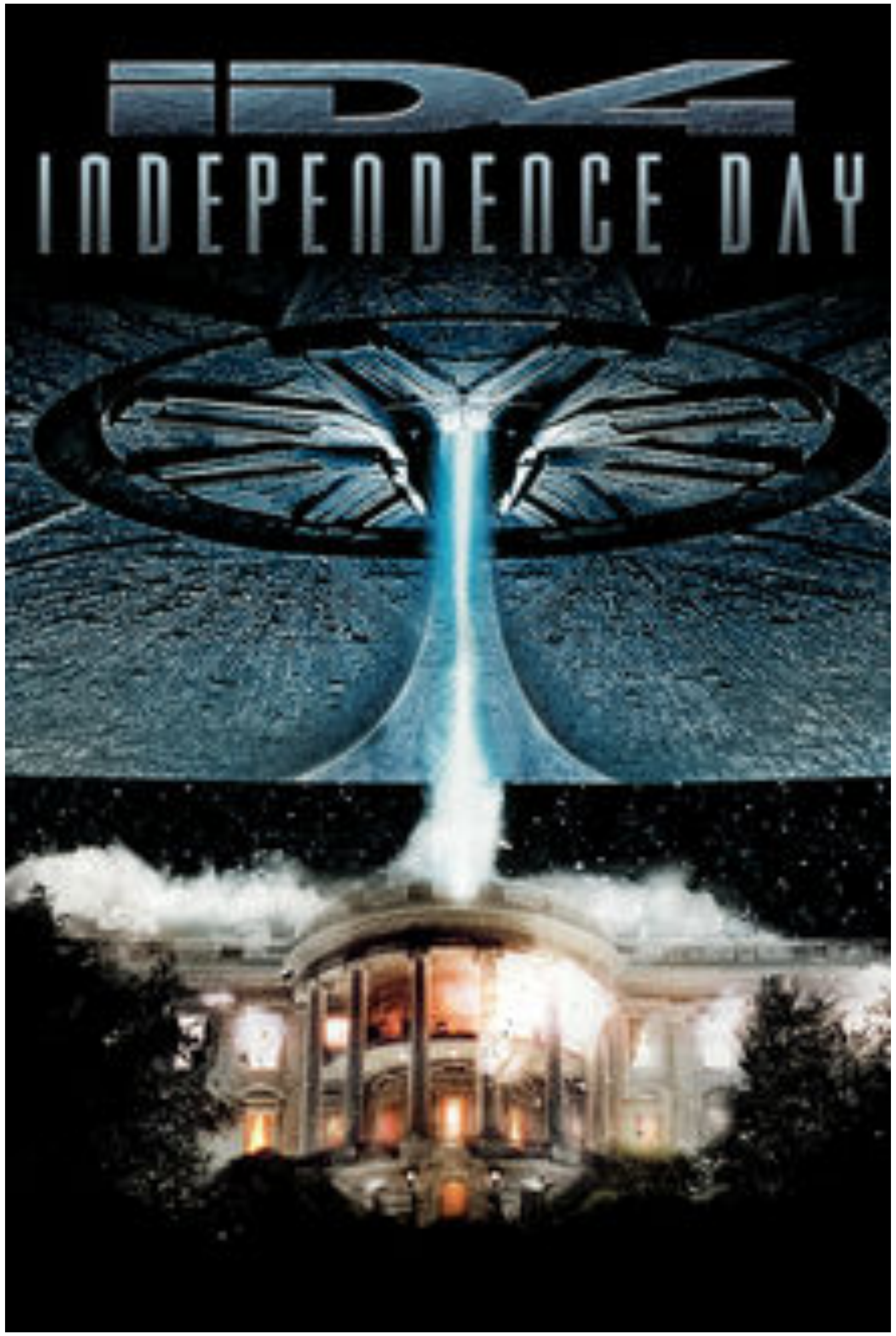
multile-choice question

- Example: Bill answers a question with 4 choices (A, B, C, D)
  - Bill either knows the correct answer or makes a random guess
  - $P(\text{Bill knows the correct answer}) = 2/3$ E
  - $P(\text{Bill does not make a random guess} \mid \text{answer is correct}) = ?$ F

$$P(E|F) = \frac{\underbrace{P(E)}_{2/3}\,\underbrace{P(F|E)}_{=1}}{\underbrace{P(E)}_{=2/3}\,\underbrace{P(F|E)}_{=1} + \underbrace{P(E^c)}_{=1/3}\,\underbrace{P(F|E^c)}_{=1/4}} = \frac{\frac{2}{3}}{\frac{2}{3}+\frac{1}{12}} = \frac{8}{9}$$

Independence

Independence?



Independence of 2 Events

$P(A \cap B^c) = P(A) \cdot P(B^c)$

pf: $P(A) = P(A \cap B) + P(A \cap B^c)$

$= P(A) \cdot P(B) + P(A \cap B^c)$

Definition: Two events A and B are said to be **independent** if $P(A \cap B) = P(A)P(B)$ ...... operational

Moreover, if $P(B) > 0$, then independence is equivalent to the condition

$$P(A \mid B) = P(A)$$

.... intuition

- Example:
- If A and B are independent, then are A and B^c also independent? Yes

$P(A \cap B^c) = P(A) - P(A) \cdot P(B)$

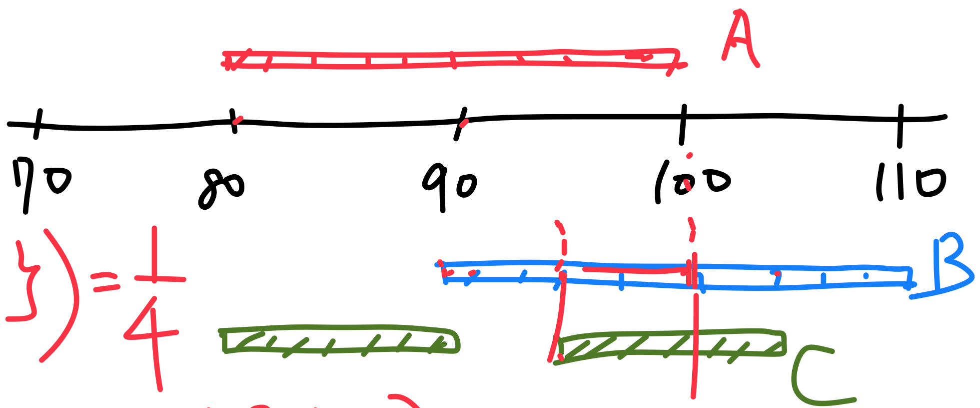
$= P(A) \cdot (1 - P(B))$

$= P(A) \cdot P(B^c)$

Example

Yellowstone NP

- Example: Old Faithful Geyser
  - Erupts every 70-110 minutes (at random) (uniformly)
  - Let X = waiting time before next eruption
  - $A = \{80 \le X \le 100\}$ $P(A) = \frac{1}{2}$
  - $B = \{90 \le X \le 110\}$ $P(B) = \frac{1}{2}$ $P(C) = \frac{1}{2}$
  - $C = \{80 \le X \le 90$ or $95 \le X \le 105\}$



Yes

1. A and B independent?

$P(A \cap B) = P(\{90 \le X \le 100\}) = \frac{1}{4} = P(A)P(B)$

Yes

2. B and C independent?

$P(B \cap C) = P(\{95 \le X \le 105\}) = \frac{1}{4} = P(B)P(C)$

No

3. A and C independent?

$P(A \cap C) = P(\{80 \le X \le 90 \text{ or } 95 \le X \le 100\}) = \frac{3}{8} \neq P(A)P(C)$

No

4. B and $A \cap C$ independent?

$P(B \cap (A \cap C)) = P(\{95 \le X \le 100\}) = \frac{1}{8} \neq P(B)P(A \cap C)$

Independence of Several Events

Definition: Events $A_1, A_2, \cdots, A_n$ are said to be **independent** if

$$P(\bigcap_{i \in S} A_i) = \prod_{i \in S} P(A_i), \quad \text{for every } S \subseteq \{1,2,\cdots,n\}$$

- When $n = 3$?

$$\begin{cases} P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2) \cdot P(A_3). & (1) \\ P(A_1 \cap A_2) = P(A_1) \cdot P(A_2) & (2) \\ P(A_2 \cap A_3) = P(A_2) \cdot P(A_3) & (3) \\ P(A_1 \cap A_3) = P(A_1) \cdot P(A_3) & (4) \end{cases}$$

Pairwise Independence $\not\Rightarrow$ Independence

- Example: Toss a fair coin twice
  - $T_1 = \{\text{1st toss is a tail}\}$ $P(T_1) = \frac{1}{2}$
  - $T_2 = \{\text{2nd toss is a tail}\}$ $P(T_2) = \frac{1}{2}$
  - $D = \{\text{2 tosses have different results}\}$

$P(D) = \frac{1}{2}$

HH $\frac{1}{4}$	HT $\frac{1}{4}$
TH $\frac{1}{4}$	TT $\frac{1}{4}$

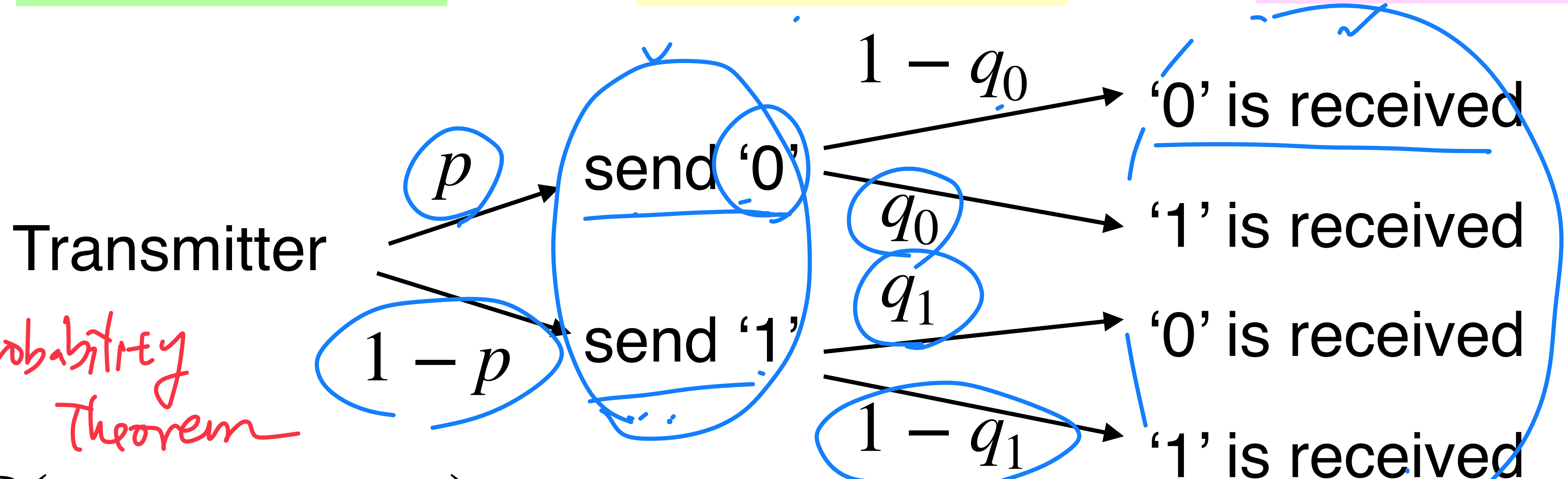
T_2 T_1 D

1. $P(T_1 \cap T_2) = \frac{1}{4} = P(T_1) \cdot P(T_2)$ $\Rightarrow$ T_1, T_2 are independent
2. $P(D \cap T_1) = \frac{1}{4} = P(D)P(T_1)$ $\Rightarrow$ D, T_1 are independent
3. $P(D \cap T_2) = \frac{1}{4} = P(D) \cdot P(T_2)$ $\Rightarrow$ D, T_2 are " "

(pair-wise independence)

4. $P(D \cap T_1 \cap T_2) = 0 \neq P(D) \cdot P(T_1) \cdot P(T_2)$
5. $P(D \mid T_1 \cap T_2) = \frac{P(D \cap T_1 \cap T_2)}{P(T_1 \cap T_2)} = 0$ $\Rightarrow$ D is not independent from $T_1 \cap T_2$

Example: Communication Over a Noisy Channel



Total probability Theorem

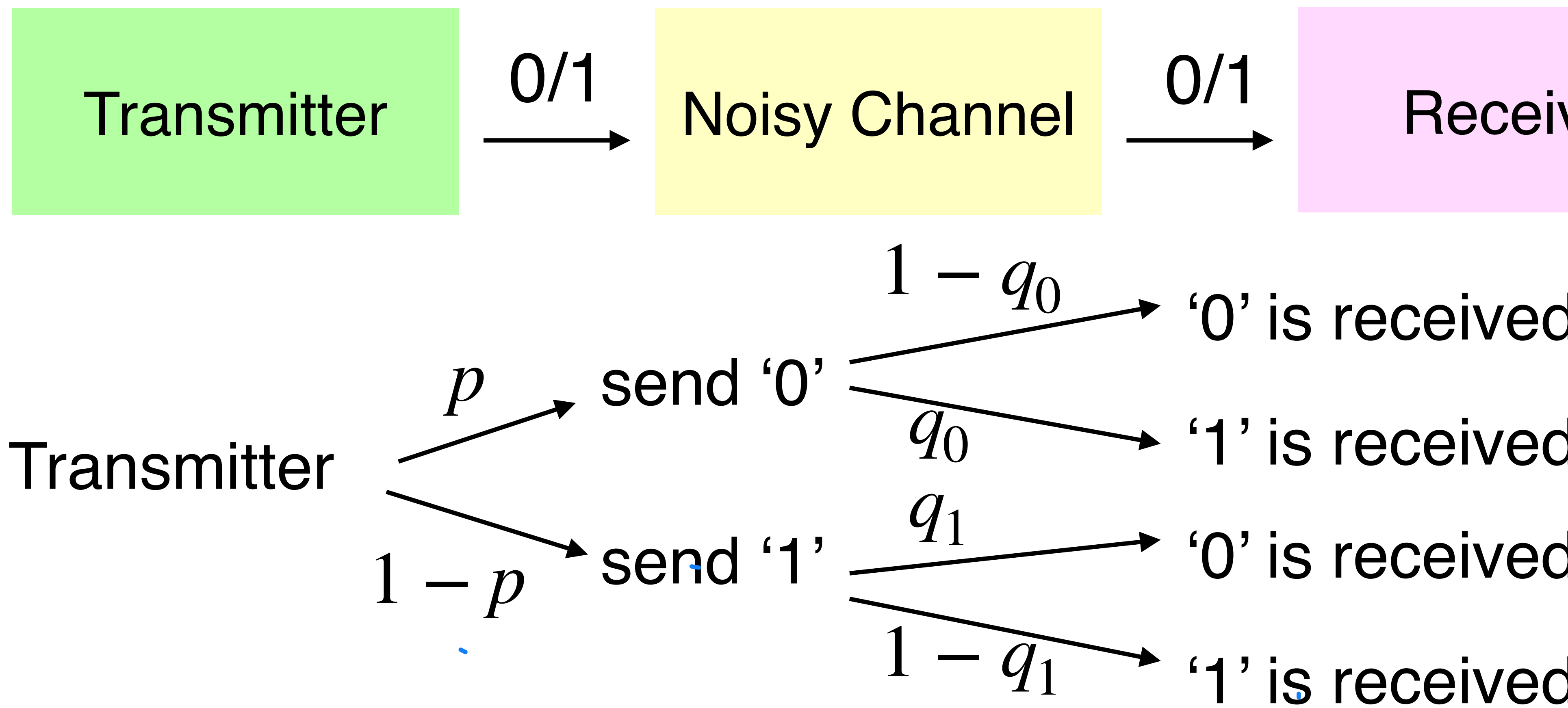
- P(0 is received) = ?

$$\underbrace{P(0 \text{ is sent})}_{p} \cdot \underbrace{P(0 \text{ is recv} \mid 0 \text{ is sent})}_{1-q_0}$$

$$+ \underbrace{P(1 \text{ is sent})}_{1-p} \cdot \underbrace{P(0 \text{ is recv} \mid 1 \text{ is sent})}_{q_1}$$

$$= p\cdot(1-q_0) + (1-p)\cdot q_1.$$

Example: Communication Over a Noisy Channel



- P(01 sent and 01 received) = ?

$A_1 = \{$1st sent bit is 0$\}$

$A_2 = \{$2nd sent bit is 1$\}$

$A_3 = \{$1st recv bit is 0$\}$

$A_4 = \{$2nd recv bit is 1$\}$

$= P(A_1 \cap A_2 \cap A_3 \cap A_4)$

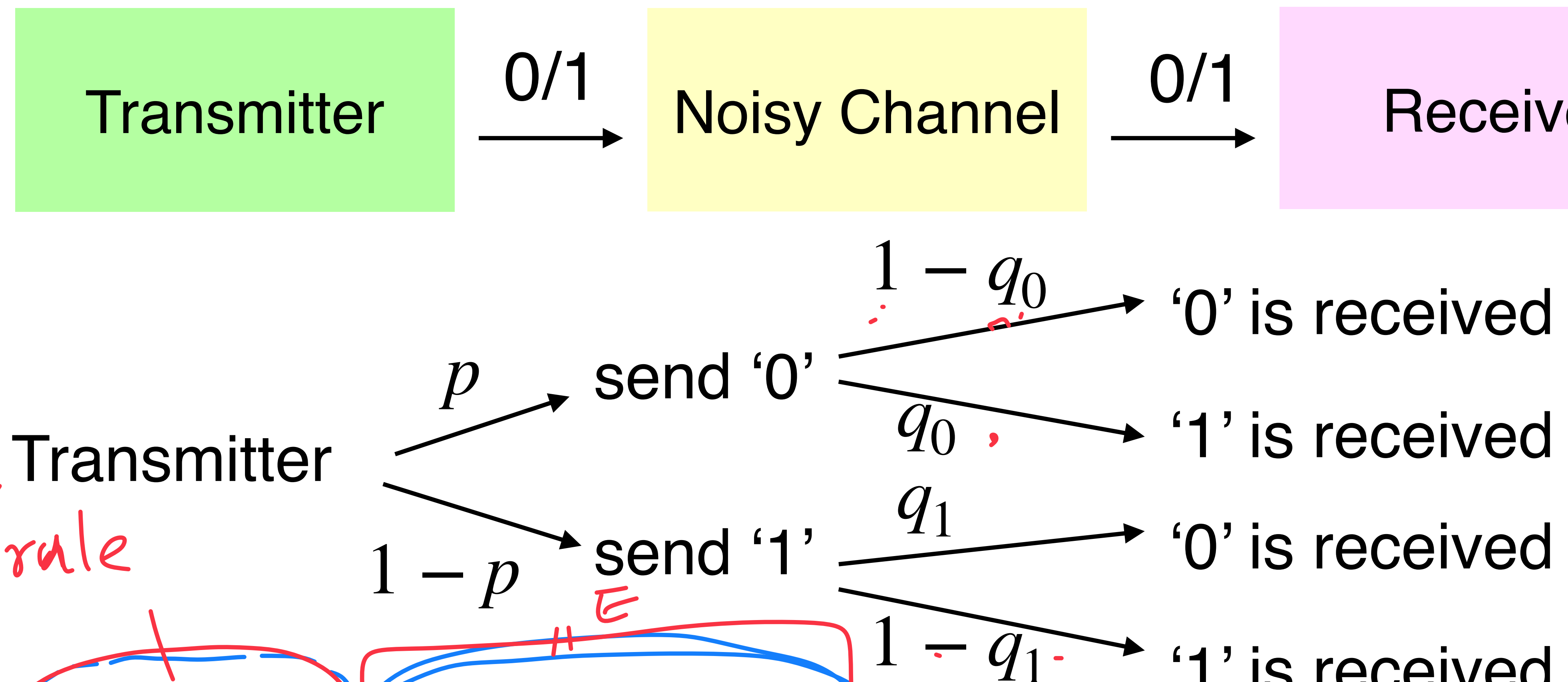
$= P((A_1 \cap A_3) \cap (A_2 \cap A_4))$

$= P(A_1 \cap A_3) \cdot P(A_2 \cap A_4)$

$= P(A_1) P(A_3 | A_1) \cdot P(A_2) P(A_4 | A_2)$

p $\quad 1-q_0 \quad 1-p \quad 1-q_1$

Example: Communication Over a Noisy Channel



Bayes' rule

- $P(01 \text{ is sent} \mid \underbrace{01 \text{ is received}}_{=E}) = ?$

$$= \frac{\overbrace{P(01 \text{ is sent})}^{p\cdot(1-p)} \cdot \overbrace{P(01 \text{ recv} \mid 01 \text{ sent})}^{(1-q_0)\cdot(1-q_1)}}{\underbrace{P(01 \text{ sent})}_{p(1-p)} \cdot \underbrace{P(E \mid 01 \text{ sent})}_{(1-q_0)(1-q_1)} + \underbrace{P(00 \text{ sent})}_{p\cdot p} \cdot \underbrace{P(E \mid 00 \text{ sent})}_{(1-q_0)\cdot q_0} + P(10 \text{ sent}) \cdot P(E \mid 10 \text{ sent}) + P(11 \text{ sent}) \cdot P(E \mid 11 \text{ sent})}$$

Review: Combinatorial Methods

Why Counting?

- "Principle of indifference": All outcomes are equally likely

Discrete uniform probability law: Let Ω be the sample space of an experiment. If Ω has N elements that are equally likely to occur, then for any event A of Ω, we have

$$P(A) = \frac{\text{Number of elements in A}}{N}$$

"counting"

Basic Counting Principle

- Example: Buy a sandwich at Subway

1. Size: 6-inch or 12-inch?
2. Meat: Chicken, meatball, beef, or tuna?
3. Vegetable: Lettuce or tomato?
4. Cheese: Mozzarella, Parmesan, or Cheddar?

Question: How many different types of sandwich?

Replacement

- Example: Suppose we want to draw 3 cards from 52 poker cards. How many possible ways?

1. With replacement:

2. Without replacement:

Permutation

- Example: Count # of passwords that consist of 8 distinct English letters (case sensitive)

Definition: Given n distinct objects, and let k be some positive integer with $k \leq n$. Then, an ordered arrangement of k objects is called a k**-element permutation from** n **objects**. The number of k-element permutation from n objects is denoted by P_k^n, and

$$P_k^n = n \cdot (n-1) \cdots (n-k+1) = \frac{n!}{(n-k)!}$$

Combination

- Example: Count # of possible collections that consist of 8 distinct letters (case sensitive)

Definition: Given n distinct objects, and let k be some positive integer with $k \leq n$. Then, an unordered arrangement of k objects is called a **k-element combination from n objects**. The number of k-element combination from n objects is denoted by C_k^n, and

$$C_k^n = \frac{P_k^n}{k!} = \frac{n!}{(n-k)!k!}$$

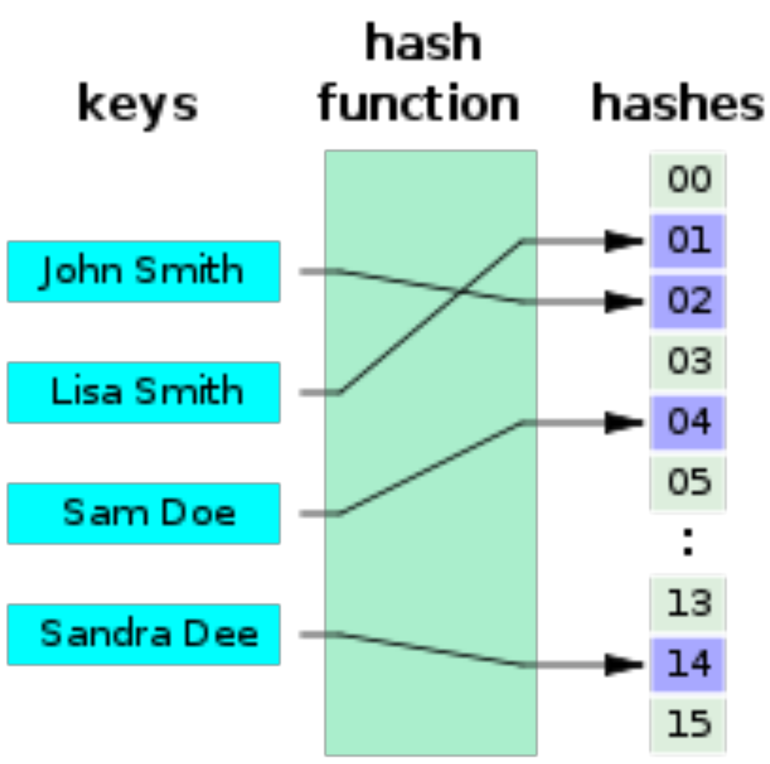
Example: Birthday Problems

- What is the probability that at least 2 students of a class of size N have the same birthday?

- What if $N = 23$? How about $N = 60$?

Example: Hash Collision

- Suppose there are K possible hash values
- What is the probability of <u>at least 1 hash collision</u> of a random group of N English words (keys)?



- What if $N \ll K$?

Example: Sum of Integers

- Let $X_1, X_2, \cdots, X_{10}$ be integers and $X_1 + X_2 + \cdots + X_{10} = 6$

1. If $X_1, X_2, \cdots, X_{10}$ are all binary (0 or 1), how many different combinations do we have?

2. If $X_1, X_2, \cdots, X_{10}$ are all nonnegative integers, how many different combinations do we have?

Binomial Expansion

- Example: $(x+y)^3$= ?

Theorem: For any $n \geq 0$, we have

$$(x+y)^n = \sum_{i=0}^{n} C_i^n x^{(n-i)} y^i$$

- Example: $C_0^n + C_1^n + \cdots + C_n^n$= ?

Multinomial Expansion

- Example: $(x + y + z)^3$= ?

Theorem: In the expansion of $(x_1 + x_2 + \cdots + x_k)^n$, the coefficient of the term $x_1^{n_1} x_2^{n_2} \cdots x_k^{n_k}$ with $n_1 + n_2 + \cdots + n_k = n$ is

$$\frac{(n_1 + n_2 + \cdots + n_k)!}{n_1! n_2! \cdots n_k!}$$

- How to interpret this?

1-Minute Summary

1. Conditioning and Independence

- Bayes' rule
- Gambler's ruin / Communication over a noisy channel
- Independence of two or several events
- Pair-wise independence $\nRightarrow$ independence

2. Review: Combinatorial Methods

- Permutation / Combination / Binomial expansion