1179: Probability
Lecture 6 — Combinatorics and Random Variables

Ping-Chun Hsieh (謝秉均)

October 1, 2021

Powerball Lottery and Fortune Cookies

- One day, in 2005, 110 lucky people in the US won the same Powerball lottery.
- Powerball: Pick 5 numbers from 1~69 + 1 number from 1~26
  - What is the probability of winning the lottery?


https://www.nytimes.com/2005/05/11/nyregion/who-needs-giacomo-bet-on-the-fortune-cookie.html

Quick Review

- Bayes' rule?

$$\underbrace{P(A_i|B)}_{\text{posterior}} = \frac{\underbrace{P(A_i)}_{\text{prior}}\underbrace{P(B|A_i)}_{\text{Likelihood}}}{\underbrace{P(B)}_{\text{evidence}}}$$

- Why do we care about "counting"?

Discrete Uniform Probability Law

Ω

$$P(E) = \frac{\#\text{ of elements in } E}{N}$$

This Lecture

1. Review: Combinatorial Methods

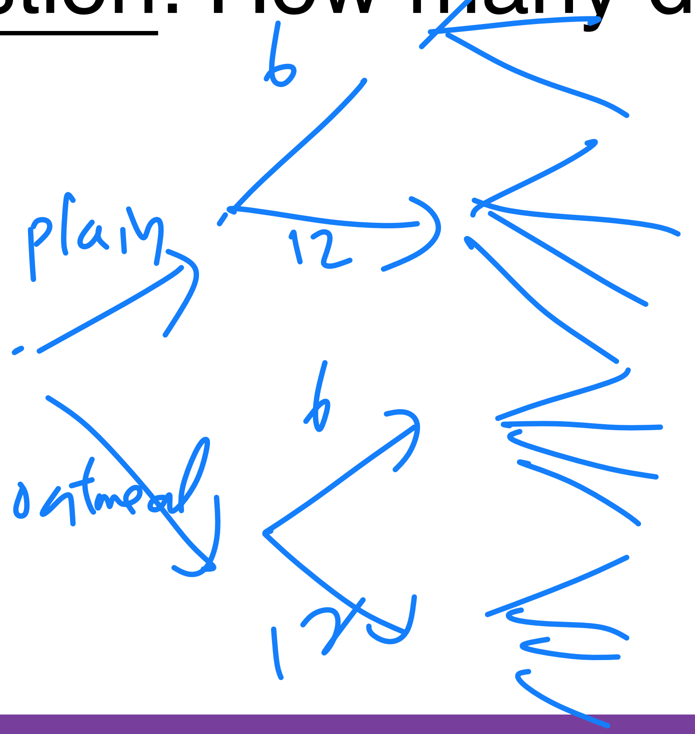
2. Random Variables and Cumulative Distribution Function (CDF)

- Reading material: Chapter 2 and 4.1~4.3

Basic Counting Principle

- Example: Buy a sandwich at Subway
  1. Bread: plain or oatmeal?
  2. Size: 6-inch or 12-inch?
  3. Meat: Chicken, meatball, beef, or tuna?
  4. Vegetable: Lettuce or tomato?
  5. Cheese: Mozzarella, Parmesan, or Cheddar?

Question: How many different types of sandwich?



$2 \times 2 \times 4 \times 2 \times 3$

Replacement

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

1. With replacement: (put back) $52 \times 52 \times 52$

2. Without replacement: (not put back) $52 \times 51 \times 50$

Permutation

- Example: Count # of passwords that consist of 8 distinct English letters (case sensitive) 52

Password: ABcDeFgh

$$\underset{52}{\_}\ \underset{51}{\_}\ \underset{50}{\_}\ \underset{49}{\_}\ \underset{48}{\_}\ \underset{47}{\_}\ \underset{46}{\_}\ \underset{45}{\_} = \frac{52!}{44!}$$

Definition: Given n distinct objects, and let k be some positive integer with $k \le n$. 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)

52, 51, 50, 49, 48, 47, 46, 45

8!

$$= \frac{P^{52}_{8}}{8!} = C^{52}_{8}$$

> **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^n_k, and
>
> $$C^n_k = \frac{P^n_k}{k!} = \frac{n!}{(n-k)!k!}$$

combination

Example: Birthday Problems

"leap month"

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

$N=2$: $|\Omega| = 365 \times 365$

Ω

$(1/1, 2/14)$

$E = \{$the two students have the same birthday$\}$

$|E| = 365$

$$P(E) = \frac{365}{365 \times 365}$$

$$P(E^c) = \frac{(365 \times 364)}{365 \times 365}$$

General N: $(N \leq 365)$

$E = \{$at least 2 students have the same birthday$\}$

$$P(E^c) = \frac{\overbrace{365 \times 364 \times \cdots \times (365-N+1)}^{P_N^{365}}}{\underbrace{365 \times 365 \times \cdots \times 365}_{365^N}}$$

$$= \frac{P_N^{365}}{365^N}$$

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

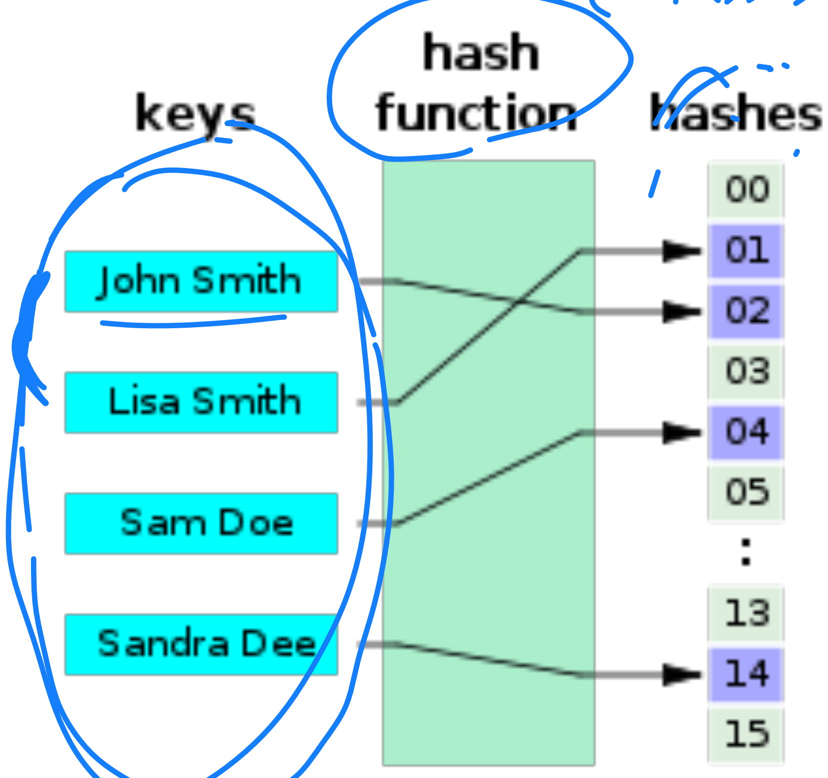
$P(E) = 0.5$

$P(E) = 0.99$

Example: Hash Collision (≡ Birthday Problem)

- Suppose there are K possible hash values (possible birthdays)
- What is the probability of at least 1 hash collision of a random group of N English words (keys)? (class size)

event E



$$P(E^c) = \frac{K(K-1)\cdots(K-N+1)}{\underbrace{K\cdot K\cdot K\cdots K}_{K^N}} \quad (\equiv P^K_N)$$

$$= 1\cdot\left(1-\frac{1}{K}\right)\left(1-\frac{2}{K}\right)\cdots\left(1-\frac{N-1}{K}\right)$$

$$\approx e^{-\frac{1}{K}} \quad \approx e^{-\frac{2}{K}} \quad \approx e^{-\frac{N-1}{K}}$$

$$\approx e^{-\frac{1}{K}(1+2+\cdots+N-1)}$$

- What if $N \ll K$?

Binomial Expansion

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

$$(C^3_3)x^3 + (C^3_2 \cdot C^1_1)x^2y^1 + (C^3_1 \cdot C^2_2)x^1y^2 + (C^3_0)y^3$$

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

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

- **Example**: $C^n_0 + C^n_1 + \cdots + C^n_n$= ?

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?

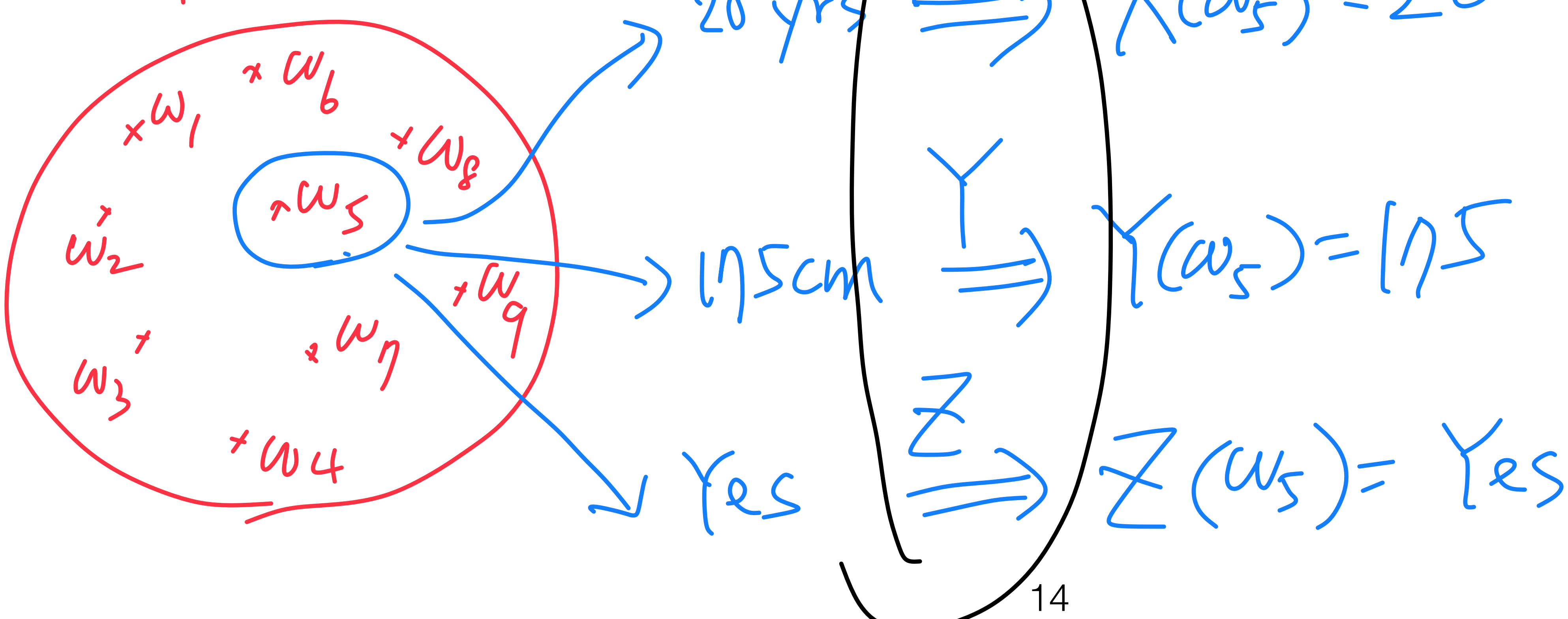
2. Random Variables

Why Do We Need Random Variables (r.v.)?

- Example: Our Probability class (127 students)
  - Suppose we are looking for the 3rd student leaving this Webex session
    1. How old is this student? 20
    2. How tall is this student? 175 cm
    3. Whether this student wears glasses or not? Yes

"random variables"

Sample space Ω

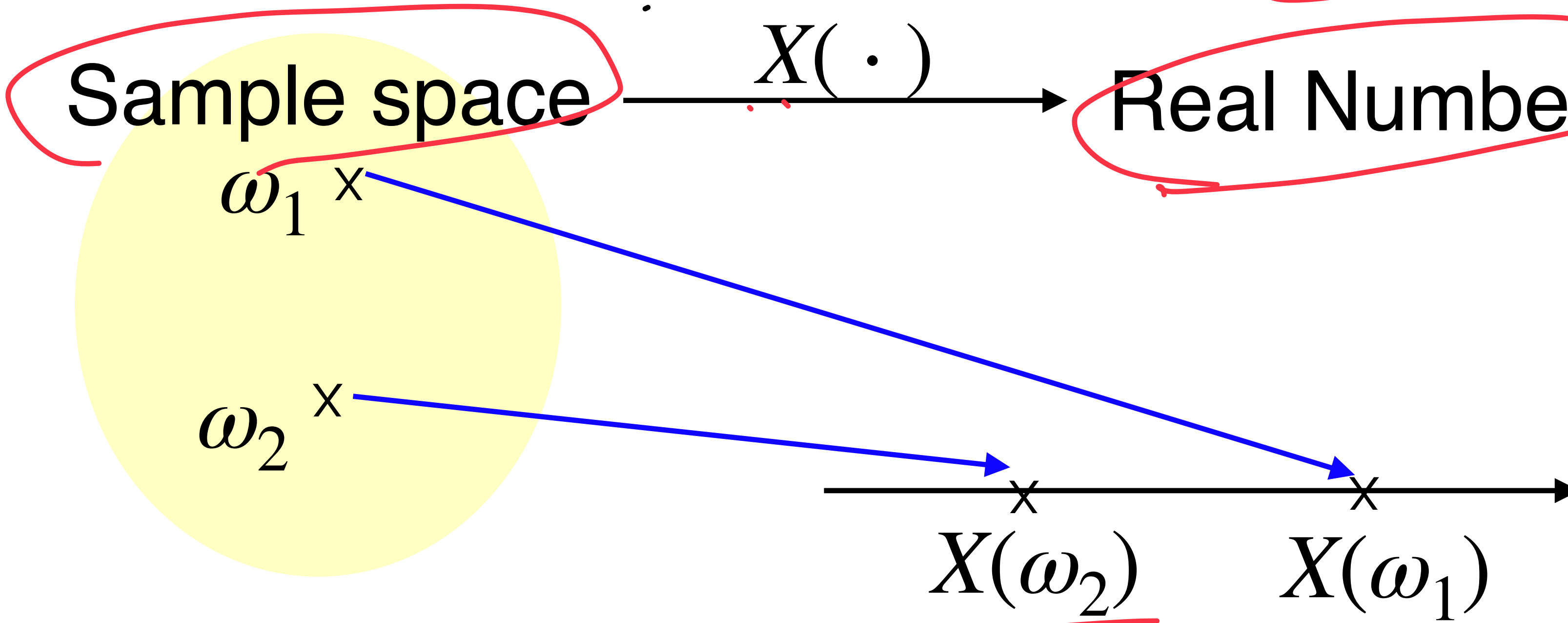


Why Do We Need Random Variables?

1. We are too lazy:
   - A random variable offers a shorthand for events
2. We are too curious about the world:
   - Multiple properties of interest from the same sample space and experiment
3. Random variables are powerful:
   - One type of random variable can capture the common features of different experiments

What is a Random Variable (Formally)?

- Random variable: a function that maps each outcome to a real number



- Example: Whether NCTU will merge with NYMU

- Example: # of people waiting in line at Shinemood

Function of a Random Variable

- Example: Buy a waffle at Shinemood
  - If it is sunny today, then you spend \$50 to order a Matcha-red-bean waffle
  - Otherwise, you spend \$70 to order a Fried-chicken waffle
  - Question: Is the price of your waffle a r.v.?

Discrete and Continuous Random Variables

- Example: # of people waiting in line at Shinemood

- Example: Amount of time needed for finishing HW1

Cumulative Distribution Function (CDF)

- Random variables are used to calculate the probabilities of events.

Cumulative Distribution Function (CDF): For any random variable X, the CDF of X is defined as:

$$F_X(t) = P(X \leq t), \text{ for all } t \in \mathbb{R}$$

- What's the range of $F_X(t)$?
- How to use the CDF?
- Example: $P(a < X \leq b) = ?$

CDF of a Discrete Random Variable

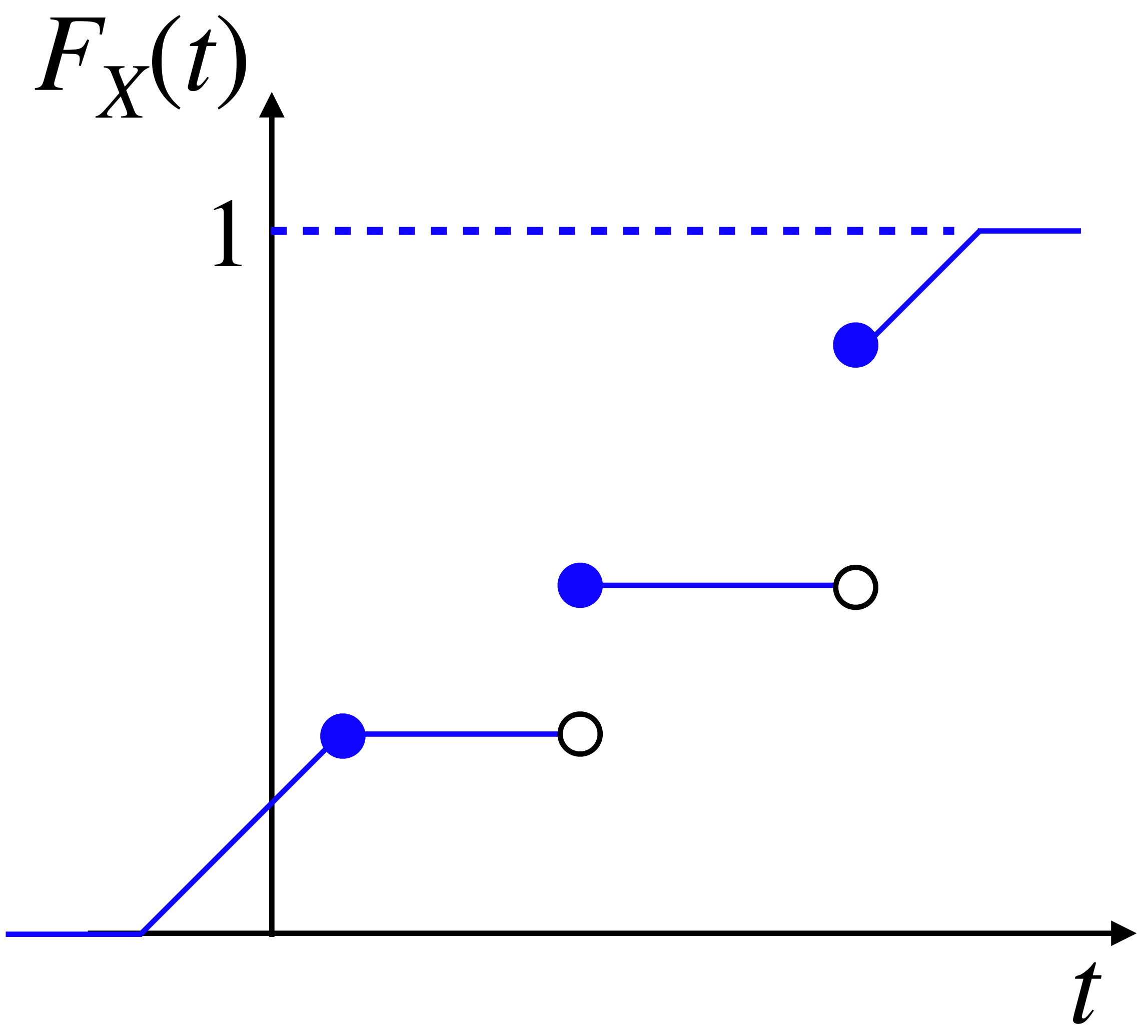
- Example: Roll a fair 4-sided die
  - $P(X = 1) = P(X = 2) = P(X = 3) = P(X = 4) = 1/4$
  - What is the CDF of X?

1. $P(X \leq 3)$ =
2. $P(X < 3)$ =
3. $P(1 < X \leq 3)$ =
4. $P(1 < X < 3)$ =
5. $P(X = 3)$ =

Use CDF to Find Probability of an Event (I)

$$F_X(t) = P(X \leq t), \text{ for all } t \in \mathbb{R}$$

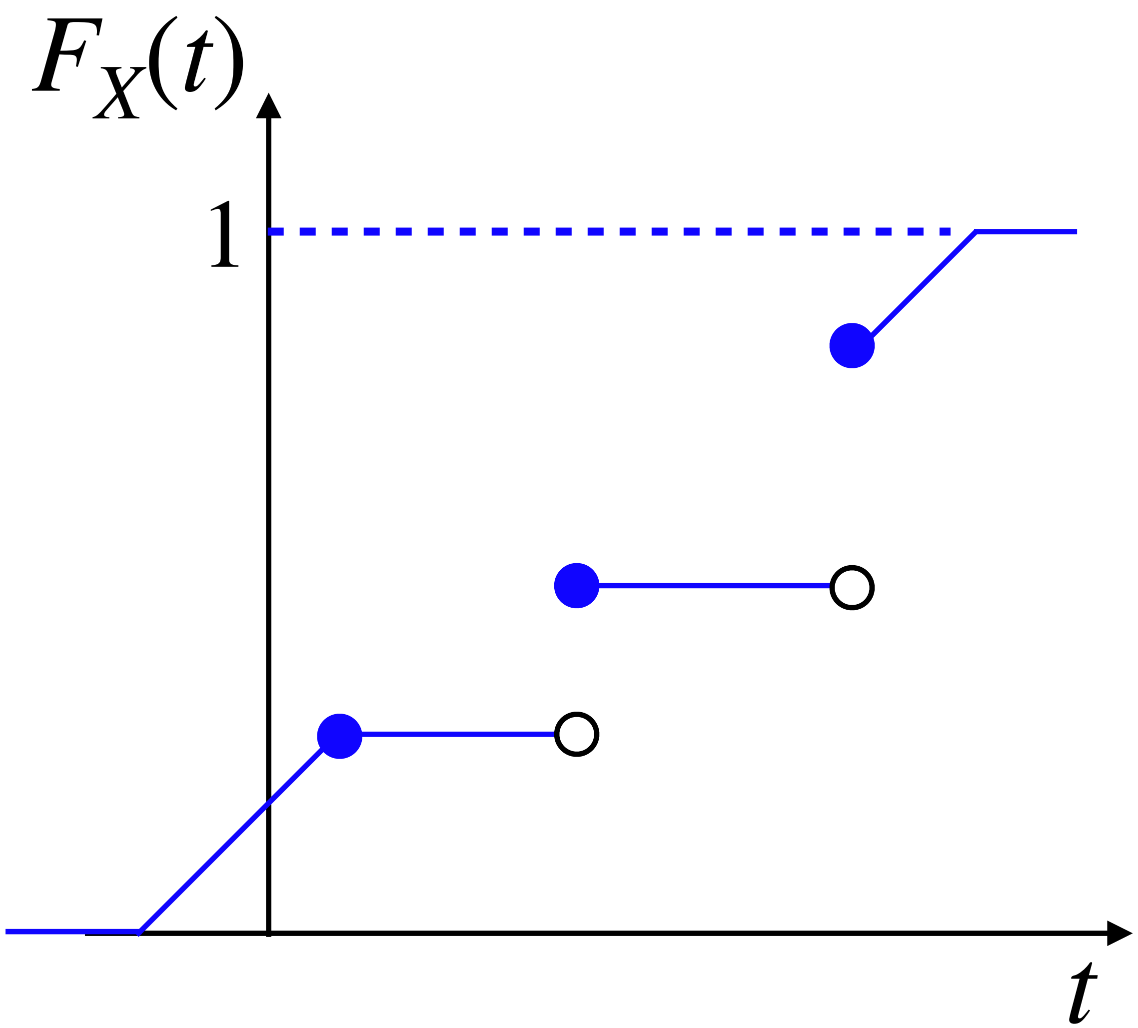
Event	Probability of the event
$X \leq a$	
$X > a$	
$X < a$	
$X \geq a$	
$X = a$	



Use CDF to Find Probability of an Event (II)

$$F_X(t) = P(X \leq t), \text{ for all } t \in \mathbb{R}$$

Event	Probability of the event
$a < X \leq b$	
$a < X < b$	
$a \leq X \leq b$	
$a \leq X < b$	



1-Minute Summary

1. Review: Combinatorial Methods

- Permutation / Combination / Binomial expansion

2. Random Variables and CDF

- Function from outcomes to real numbers
- Use CDF to find the probability of an event