CS70 @ UC Berkeley, Spring 2026

Lecture 20
Random Variables I:
Distribution and Expectation

April 7, 2026

Get the Largest Number

Consider a deck of N cards each with a number written on one side, facing down. **Assume:**

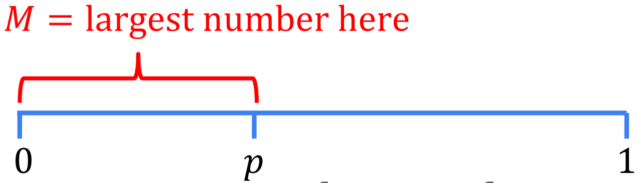
- N is large and all numbers are distinct.
- The deck is well shuffled.

Goal: Get the largest number.

Rules:

1. Reveal one card at a time starting from top.
2. STOP at the current card or reveal the next card.
3. If you pass on a card that has been revealed, you can't choose to it later.

Strategy: Reveal a certain proportion, say p, of the cards and record the largest number, denoted M, you have seen. Then, STOP when you see a number larger than M.



Question: What should p be to maximize your chance of winning?

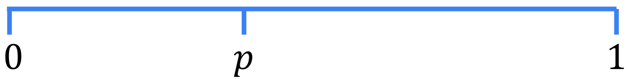
Get the Largest Number (Solution)

Success = get the largest number. $\mathbb{P}(\text{success})$?

Denote the sorted numbers as $X_1 > X_2 > \cdots > X_N$

Case 1: $M = X_1 \Rightarrow$ fail

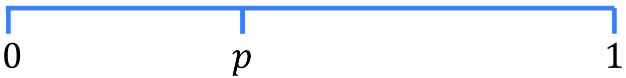
$\mathbb{P}(M = X_1) = p$



$\mathbb{P}(\text{success}|M = X_1) = 0$

Case 2: $M = X_2 \Rightarrow X_1$ is in $(p, 1]$

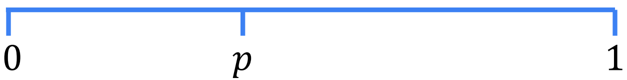
$\mathbb{P}(M = X_2) \approx p(1-p)$



$\mathbb{P}(\text{success}|M = X_2) = 1$

Case 3: $M = X_3 \Rightarrow X_1, X_2$ are in $(p, 1]$

$\mathbb{P}(M = X_3) \approx p(1-p)^2$



$\mathbb{P}(\text{success}|M = X_3) = \mathbb{P}(X_1 \text{ appears before } X_2 \text{ in } (p, 1]) = \frac{1}{2}$

General Case: $\mathbb{P}(M = X_{k+1}) \approx p(1-p)^k$ Good approximation for $k \ll N$

$\mathbb{P}(\text{success}|M = X_{k+1}) = \frac{1}{k}$

$$\mathbb{P}(\text{success}) = \sum_{i=1}^{(1-p)N} \mathbb{P}(\text{success}|M = X_i)\mathbb{P}(M = X_i)$$

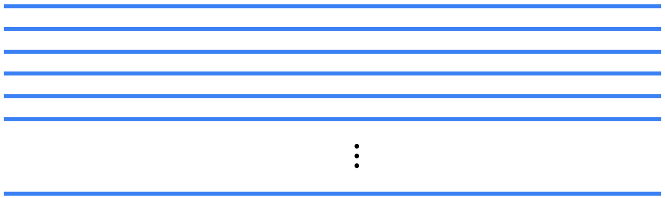
decays exponentially fast as k increases

$$\approx \sum_{k=1}^{(1-p)N} \frac{1}{k} p(1-p)^k \approx -p \ln p$$

this is maximized when $p = \frac{1}{e} \approx 0.37$

Average Number of Loops

n identical strings

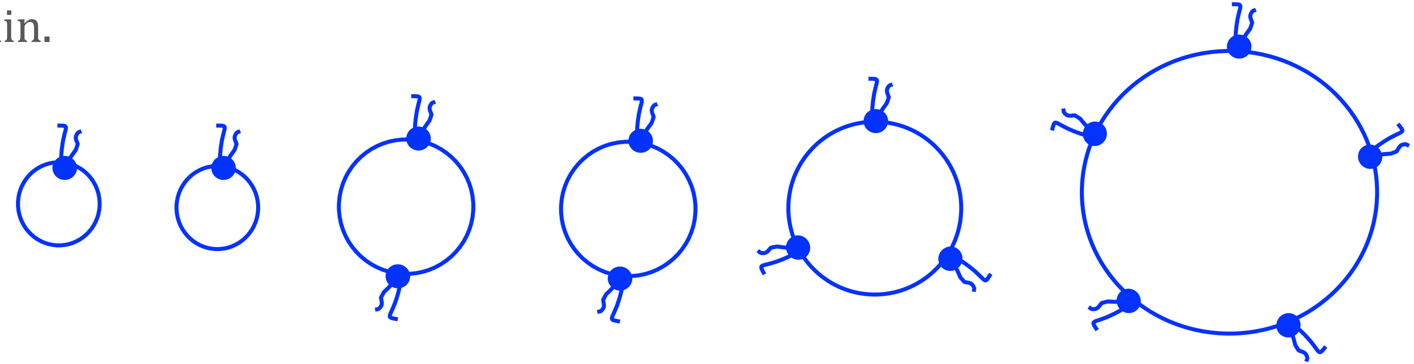


Fold them in halves and put the folded ends in a bag

$2n$ ends

Suppose you tie two free ends uniformly at random and continue until no more free ends remain.

Example outcome for $n = 14$:



6 loops formed

Question: What is the **expected** number of loops formed?

Solution will be provided in the next lecture.

Random Variables (RV)

- Why do we care about random variables?

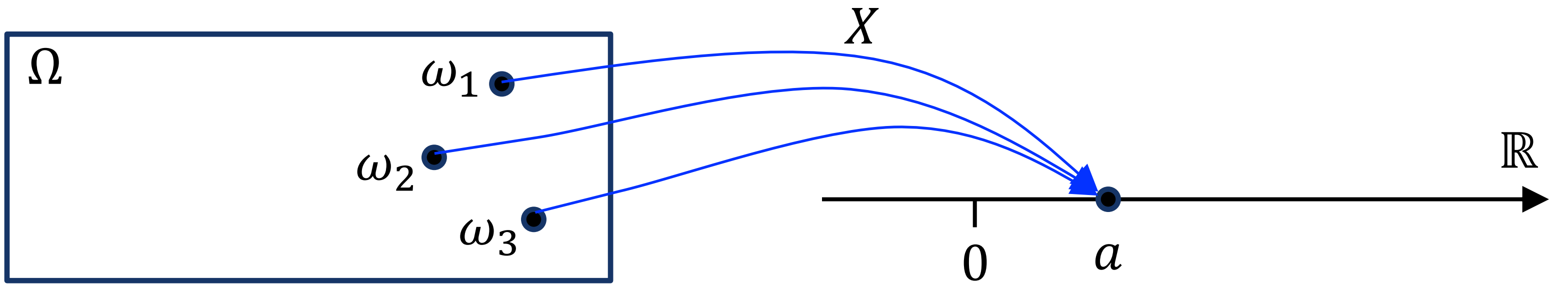
- They allow us
  - to quantify experimental outcomes,
  - extract a numerical summary of each outcome

- Once we assign a number $X(\omega)$ to each outcome $\omega \in \Omega$, we can
  - compute averages: $\mathbb{E}[X]$ (this lecture)
  - measure variability: $\mathrm{Var}(X)$ (next lecture)
  - estimate probabilities: $\mathbb{P}[X \geq a]$ (next week)

- Big Picture:
  - Experiments produce outcomes.
  - Random variables turn those outcomes into numbers.
  - Probability theory tells us how those numbers behave.

Random Variables (RV)

- A random variable is a function $X: \Omega \to \mathbb{R}^n$ (we will mostly deal with the $n = 1$ case) satisfying some technical conditions (which you will not need to worry about in CS70).



Example: Toss a fair coin twice.

$$\Omega = \{(H, H), (H, T), (T, H), (T, T)\}$$

For each $\omega \in \Omega$, let $X(\omega)$ be the number of **Heads** in ω.

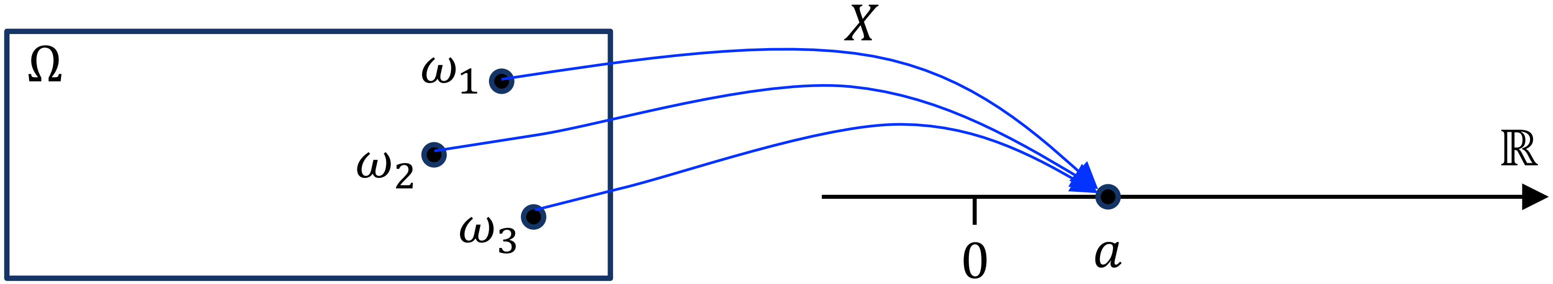
$X\big((H, H)\big) = 2$

$X\big((H, T)\big) = X\big((T, H)\big) = 1$

$X\big((T, T)\big) = 0$

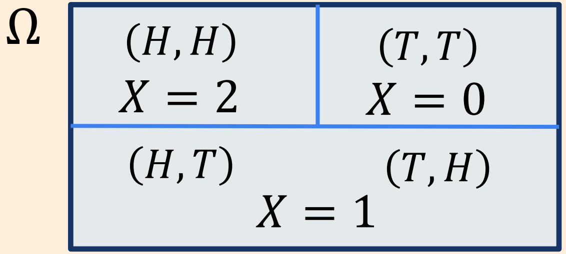
Events Defined by Random Variables

Probability measure $\mathbb{P}$ is defined on the set of subsets of Ω



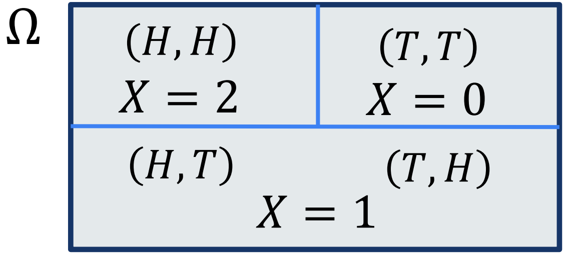
- Let $\mathcal{A} = \{a \in \mathbb{R} \mid X(\omega) = a \text{ for some } \omega \in \Omega\}$ denote the range of X.
- For a given $a \in \mathcal{A}$, the event "$X = a$" is defined as the pre-image
$$X^{-1}(a) = \{\omega \in \Omega \mid X(\omega) = a\} \subseteq \Omega.$$
- The probability of the event "$X = a$" is defined by measuring the "its size" using $\mathbb{P}$.

Example: Toss a fair coin twice.
$X(\omega)$ be the number of **Heads** in $\omega \in \Omega$.
"$X = 2$" $= X^{-1}(2) = \{(H,H)\}$
"$X = 1$" $= X^{-1}(1) = \{(H,T),(T,H)\}$
"$X = 0$" $= X^{-1}(0) = \{(T,T)\}$
$\mathbb{P}(X = 2) = \mathbb{P}(X = 0) = ¼$, and $\mathbb{P}(X = 1) = ½$.

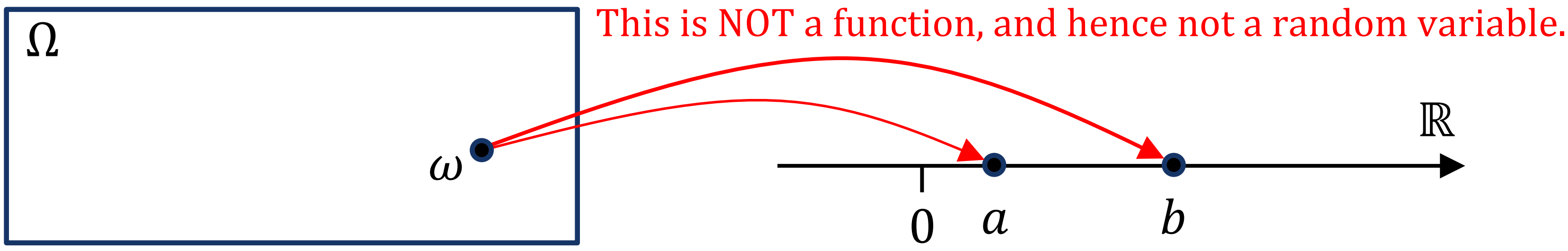


Partitions of Ω Defined by Random Variables

- For now, assume that the range $\mathcal{A}$ is discrete (finite or countable). We will later consider the continuous case (Lectures 24 and 25) .
- As illustrated in the example, the collection of events $\{X = a, \text{for } a \in \mathcal{A}\}$ partitions the sample space Ω



- Clearly $\cup_{a \in \mathcal{A}} (X = a) = \Omega$
- Why is $(X = a) \cap (X = b) = \emptyset$ for all $a, b \in \mathcal{A}$, where $a \neq b$?

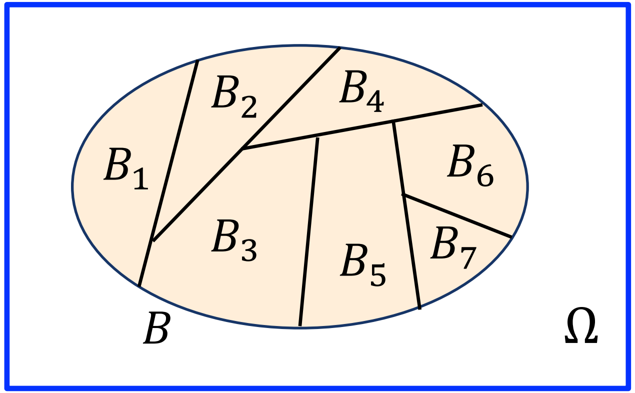


Recall the Rules of Probability

Definition (Partition): An event $B \subseteq \Omega$ is said to be partitioned into n events $B_1, \dots, B_n$ if

1. $B = B_1 \cup \cdots \cup B_n$,
2. $B_i \cap B_j = \emptyset$, for all $i \neq j$ (that is, $B_1, \dots, B_n$ are mutually exclusive).

More generally, infinitely many mutually exclusive B_ks may be involved ($B = \bigcup_{k=1}^{\infty} B_k$).



1. **(Non-negativity)** $\mathbb{P}(A) \geq 0$, for all $A \subseteq \Omega$.
2. **(Countable Additivity)** If $B_1, B_2, B_3, \dots$ is a partition of B, then

$$\mathbb{P}(B) = \sum_{k=1}^{\infty} \mathbb{P}(B_k)$$

3. **(Normalization)** $\mathbb{P}(\Omega) = 1$.

Probability Distribution

Remarks: Since $\{X = a, \text{for } a \in \mathcal{A}\}$ defines a partition of Ω,

1. $\sum_{a\in\mathcal{A}} \mathbb{P}(X = a) = \mathbb{P}(\Omega) = 1$, and

2. for any event $E \in \Omega$, the Law of Total Probability implies

$$\mathbb{P}(E) = \sum_{a\in\mathcal{A}} \mathbb{P}(E|X = a)\mathbb{P}(X = a).$$

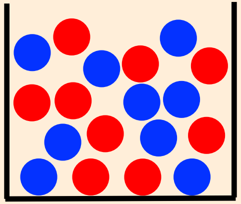
Definition (Probability Distribution): The probability distribution of a random variable X is the set $\{(a, \mathbb{P}(X = a) \mid a \in \mathcal{A}\}$.

B blue balls, $R = N - B$ red balls.
Want to sample a blue ball.
Success probability $p = B/N$

"distributed as"

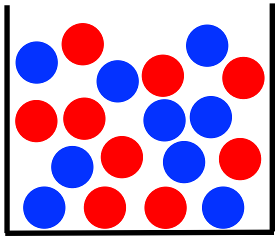
Example (Bernoulli): $I \sim \text{Bernoulli}(p)$, for $0 < p < 1$.

$$\mathbb{P}(I = a) = \begin{cases} p, & \text{if } a = 1, \\ 1 - p, & \text{if } a = 0. \end{cases}$$



Binomial Distribution

- $X \sim \text{Binomial}(n, p)$, where $n \in \mathbb{Z}_+$ and $0 < p < 1$.
- n **independent** trials with success probability p
- $X =$ number of successes



Sample n times **with** replacement.

Trials	1	2	3	4	5	6	7	…	n
$\omega \in \Omega$	F	$\boldsymbol{S}$	$\boldsymbol{S}$	F	F	$\boldsymbol{S}$	F		F
Prob	$1-p$	p	p	$1-p$	$1-p$	p	$1-p$		$1-p$

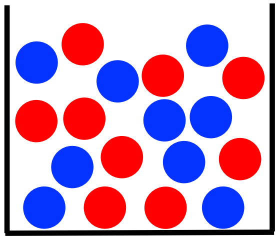
- For all $\omega \in \Omega$ with k successes and $n-k$ failures, $\mathbb{P}(\{\omega\}) = p^k(1-p)^{n-k}$.
- So, for $k = 0, \ldots, n$, we get $\mathbb{P}(X = k) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) = k\})$

$$= |\{\omega \in \Omega \mid X(\omega) = k\}| \times p^k(1-p)^{n-k}$$

$$= \binom{n}{k} p^k (1-p)^{n-k}$$

Indicator Random Variables

- $X \sim \text{Binomial}(n, p)$, where $n \in \mathbb{Z}_+$ and $0 < p < 1$.
- n independent trials with success probability p
- $X =$ number of successes
- For $i = 1, \dots, n$, let



Sample n times with replacement.

$$I_i(\omega) = \begin{cases} 1, & \text{if the } i\text{th trial in } \omega \text{ is success,} \\ 0, & \text{otherwise.} \end{cases}$$

Trials	1	2	3	4	5	6	7	…	n
$\omega \in \Omega$	F	$\boldsymbol{S}$	$\boldsymbol{S}$	F	F	$\boldsymbol{S}$	F		F
Prob	$1-p$	p	p	$1-p$	$1-p$	p	$1-p$		$1-p$

	$\boldsymbol{I_1}$	$\boldsymbol{I_2}$	$\boldsymbol{I_3}$	$\boldsymbol{I_4}$	$\boldsymbol{I_5}$	$\boldsymbol{I_6}$	$\boldsymbol{I_7}$	…	$\boldsymbol{I_n}$
$\omega \in \Omega$	0	1	1	0	0	1			0

- Then, $X = I_1 + \cdots + I_n$, where $I_1, \dots, I_n$ are mutually independent Bernoulli(p) random variables.

Joint and Marginal Distributions

- **The Joint Probability** of a pair of discrete random variables X and Y on the same probability space with ranges $\mathcal{A}$ and $\mathcal{B}$, respectively:
$$\mathbb{P}\big((X = a) \cap (Y = b)\big).$$
- For convenience, this is often written as $\mathbb{P}(X = a, Y = b)$
- More generally, the joint probability for n random variables $X_1, \ldots, X_n$ is written as $\mathbb{P}(X_1 = a_1, X_2 = a_2, \ldots, X_n = a_n)$.

- Given a joint distribution of X and Y, the **Marginal Probabilities** of X and Y are given by:

$$\mathbb{P}(X = a) = \sum_{b \in \mathcal{B}} \mathbb{P}(X = a, Y = b) \text{ and } \mathbb{P}(Y = b) = \sum_{a \in \mathcal{A}} \mathbb{P}(X = a, Y = b)$$

Mutual Independence

Recall from Lecture 18:

Definition (Mutual Independence of n events). Equivalently, events $A_1, \dots, A_n$ are mutually independent if for all $B_i \in \{A_i, A_i^c\}$, $i = 1, \dots, n$,

$$\mathbb{P}(B_1 \cap \cdots \cap B_n) = \prod_{i=1}^{n} \mathbb{P}(B_i).$$

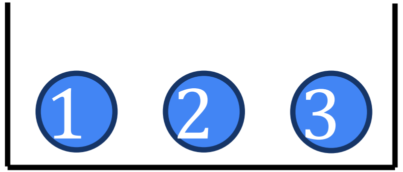
Generalization to random variables:

Definition (Mutual Independence of n random variable). Let $X_1, \dots, X_n$ be random variables defined on the same probability space. Then, they are said to be mutually independent if

$$\mathbb{P}((X_1 = a_1) \cap \cdots \cap (X_n = a_n)) = \prod_{i=1}^{n} \mathbb{P}(X_i = a_i),$$

for all $a_1 \in \mathcal{A}_1, \dots, a_n \in \mathcal{A}_n$, where $\mathcal{A}_i$ denotes the range of X_i.

Sampling with replacement vs. without replacement



Sample 2 balls at random

$X =$ label of the first ball

$Y =$ label of the second ball

Case 1: Sample with replacement

	$X = 1$	$X = 2$	$X = 3$	Row sums
$Y = 1$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{3}$
$Y = 2$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{3}$
$Y = 3$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{3}$
Column sums	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	

X, Y are independent and identically distributed (i.i.d.)

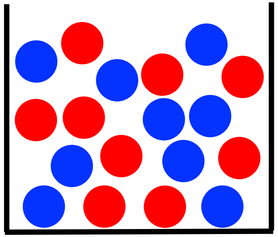
Case 2: Sample without replacement

	$X = 1$	$X = 2$	$X = 3$	Row sums
$Y = 1$	0	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3}$
$Y = 2$	$\frac{1}{6}$	0	$\frac{1}{6}$	$\frac{1}{3}$
$Y = 3$	$\frac{1}{6}$	$\frac{1}{6}$	0	$\frac{1}{3}$
Column sums	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	

X, Y are **not** independent, but identically distributed

Hypergeometric Distribution

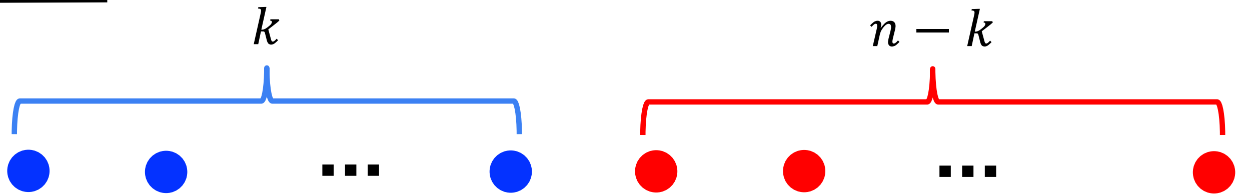
B blue balls, $R = N - B$ red balls.



Sample n times **without** replacement.

Theorem (Product Rule, Lecture 18). For any events $A_1, \dots, A_n$ on the same probability space,
$\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2)\cdots\mathbb{P}(A_n|A_1 \cap \cdots \cap A_{n-1})$.

$\omega =$



$$\mathbb{P}[\omega] = \frac{B}{N} \times \frac{B-1}{N-1} \times \cdots \times \frac{B-k+1}{N-k+1} \times \frac{R}{N-k} \times \frac{R-1}{N-k-1} \times \cdots \times \frac{R-(n-k)+1}{N-n+1} = \frac{\binom{B}{k}\binom{N-B}{n-k}}{\binom{N}{n}} \frac{1}{\binom{n}{k}}$$

- Another outcome ω' with k blue balls and $n-k$ red balls has exactly the sample probability as ω
- There are $\binom{n}{k}$ distinct sequences of k blue balls and $n-k$ red balls.
- Let Y = number of blue balls in the sample.

$$\mathbb{P}[Y = k] = \binom{n}{k}\mathbb{P}[\omega] = \frac{\binom{B}{k}\binom{N-B}{n-k}}{\binom{N}{n}}$$

$$\boxed{Y \sim \text{Hypergeometric}(N, B, n)}$$

Exchangeable Indicator Random Variables

- $Y \sim \text{Hypergeometric}(N, B, n)$
- Sample n times **without** replacement.
- Y = number of blue balls in the sample
- For $i = 1, \ldots, n$, let

$$I_i(\omega) = \begin{cases} 1, & \text{if the } i\text{th ball in } \omega \text{ is blue,} \\ 0, & \text{otherwise.} \end{cases}$$

- Then, $Y = I_1 + \cdots + I_n$, where $I_1, \ldots, I_n$ are NOT independent but **exchangeable**.

$$\mathbb{P}[I_1 = a_1, \ldots, I_n = a_n] = \frac{\binom{B}{\sum_{i=1}^n a_i}\binom{N-B}{n-\sum_{i=1}^n a_i}}{\binom{N}{n}} \cdot \frac{1}{\binom{n}{\sum_{i=1}^n a_i}}$$

The right-hand side depends on $(a_1, \ldots, a_n)$ only through their sum $\sum_{i=1}^n a_i$

For all permutations π of $\{1, \ldots, n\}$,

$$\mathbb{P}[I_{\pi(1)} = a_1, \ldots, I_{\pi(n)} = a_n] = \mathbb{P}[I_1 = a_1, \ldots, I_n = a_n]$$

Exchangeable Random Variables

Definition (Exchageability): A collection of random variables $X_1, \dots, X_n$ with common range $\mathcal{A}$ is said to be exchangeable if $\left(X_{\pi(1)}, \dots, X_{\pi(n)}\right)$ has the same joint distribution as $(X_1, \dots, X_n)$ for every permutation π of $\{1, \dots, n\}$.

Claim: Exchangeable random variables $X_1, \dots, X_n$ have identical marginal distributions:

$$\mathbb{P}[X_1 = a] = \mathbb{P}[X_2 = a] = \cdots = \mathbb{P}[X_n = a], \text{ for all } a \in \mathcal{A}$$

Proof. For any $i = 2, \dots, n$ and any $a \in \mathcal{A}$,

$$\begin{aligned}
\mathbb{P}[X_i = a] &= \sum_{b_1, \dots, b_{i-1}, b_{i+1}, \dots, b_n \in \mathcal{A}} \mathbb{P}[\boxed{X_1 = b_1}, \dots, X_{i-1} = b_{i-1}, \boxed{X_i = a}, X_{i+1} = b_{i+1}, \dots, X_n = b_n] \\
&= \sum_{b_1, \dots, b_{i-1}, b_{i+1}, \dots, b_n \in \mathcal{A}} \mathbb{P}[\boxed{X_1 = a}, \dots, X_{i-1} = b_{i-1}, \boxed{X_i = b_1}, X_{i+1} = b_{i+1}, \dots, X_n = b_n] \\
&= \mathbb{P}[X_1 = a]
\end{aligned}$$

Summarizing Probability Distributions

Value Frequency

Arithmetic average of 1, 1, 1, 3, 3, 4

$$= \frac{1+1+1+3+3+4}{6} = \frac{1\times 3 + 3\times 2 + 4\times 1}{6} = 1\times\frac{3}{6} + 3\times\frac{2}{6} + 4\times\frac{1}{6}$$

Definition (Expectation). The expectation of a random variable X with range $\mathcal{A}$ is defined as

$$\mathbb{E}[X] = \sum_{\omega\in\Omega} X(\omega)\mathbb{P}(\omega) = \sum_{a\in\mathcal{A}} a\,\mathbb{P}(X=a),$$

provided that $\sum_{a\in\mathcal{A}} |a|\,\mathbb{P}(X=a) < \infty$ (absolutely convergent).

Aside: If $\sum_{i=0}^{\infty}|a_i| = \infty$, then the sum $\sum_{i=0}^{\infty} a_i$ is not well defined.
Riemann Rearrangement Theorem: If $\sum_{i=0}^{\infty} a_i$ converges but $\sum_{i=0}^{\infty}|a_i| = \infty$, then for any given $r \in [-\infty, \infty]$, there exist a permutation π of $\mathbb{N}$ such that $\sum_{i=0}^{\infty} a_{\pi(i)} = r$.

Linearity of Expectation

Theorem (Linearity of Expectation): For any two random variables X and Y on the same probability space and fixed **constants** α and β,

$$\mathbb{E}[\alpha X + \beta Y] = \alpha \mathbb{E}[X] + \beta \mathbb{E}[Y].$$

(Note: this holds irrespective of whether X and Y are independent.)

Proof: Let $\mathcal{A}$ and $\mathcal{B}$ denote the ranges of X and Y, respectively

$$\mathbb{E}[\alpha X + \beta Y] = \sum_{a\in\mathcal{A}} \sum_{b\in\mathcal{B}} (\alpha a + \beta b) \mathbb{P}\big((X = a) \cap (Y = b)\big)$$

$$= \alpha \sum_{a\in\mathcal{A}} a \sum_{b\in\mathcal{B}} \mathbb{P}\big((X = a) \cap (Y = b)\big) + \beta \sum_{b\in\mathcal{B}} b \sum_{a\in\mathcal{A}} \mathbb{P}\big((X = a) \cap (Y = b)\big)$$

Since $\{Y = b, b \in \mathcal{B}\}$ partition Ω, $\{(X = a) \cap (Y = b), b \in \mathcal{B}\}$ partition $X = a$.

Similarly, since $\{X = a, a \in \mathcal{A}\}$ partition Ω, $\{(X = a) \cap (Y = b), a \in \mathcal{A}\}$ partition $Y = b$. So, additivity implies

$$= \alpha \sum_{a\in\mathcal{A}} a \mathbb{P}\big((X = a)\big) + \beta \sum_{b\in\mathcal{B}} b \mathbb{P}\big((Y = b)\big) = \alpha \mathbb{E}[X] + \beta \mathbb{E}[Y]$$

□

Examples

Example (Bernoulli): $I \sim \text{Bernoulli}(p)$, for $0 < p < 1$.

$$\mathbb{P}(I = a) = \begin{cases} p, & \text{if } a = 1, \\ 1 - p, & \text{if } a = 0. \end{cases}$$

$$\mathbb{E}[I] = 1 \cdot \mathbb{P}(I = 1) + 0 \cdot \mathbb{P}(I = 0) = p$$

Example (Binomial): $X \sim \text{Binomial}(n, p)$, where $n \in \mathbb{Z}_+$ and $0 < p < 1$.
$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$, for $k = 0, \dots, n$.

$$\mathbb{E}[X] = \sum_{k=0}^{n} k\, \mathbb{P}(X = k) = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k}$$

Cumbersome calculation

$$\mathbb{E}[X] = \mathbb{E}[I_1 + \cdots + I_n] = \mathbb{E}[I_1] + \cdots + \mathbb{E}[I_n] = np$$

Much easier

Linearity of expectation

Examples

Example (Hypergeometric): $Y \sim \text{Hypergeometric}(N, B, n)$

$$\mathbb{P}[Y = k] = \frac{\binom{B}{k}\binom{N-B}{n-k}}{\binom{N}{n}}.$$

$$\mathbb{E}[\mathrm{Y}] = \sum_{k=0}^{n} k\, \mathbb{P}(\mathrm{Y} = k) = \sum_{k=0}^{n} k \frac{\binom{B}{k}\binom{N-B}{n-k}}{\binom{N}{n}}$$

Cumbersome calculation

$$\mathbb{E}[\mathrm{Y}] = \mathbb{E}[I_1 + \cdots + I_n] = \mathbb{E}[I_1] + \cdots + \mathbb{E}[I_n] = n\mathbb{E}[I_1] = n\frac{B}{N}$$

Much easier

Linearity of expectation

Exchangeability

Examples

Example (Fixed points, Lecture 15):
$\Omega =$ set of all permutations of $\{1, \ldots, n\}$.
Suppose $\mathbb{P}[\omega] = \frac{1}{|\Omega|} = \frac{1}{n!}$ for all $\omega \in \Omega$
Let $X(\omega) =$ denote the number of fixed points of $\omega \in \Omega$.
What is $\mathbb{E}[X]$?

$$\begin{array}{rccccccc} i: & 1 & 2 & \boxed{3} & 4 & \boxed{5} & \ldots & n \\ \text{Permutation } \pi_i: & 7 & 1 & \boxed{3} & 8 & \boxed{5} & \ldots & 2 \end{array}$$

Fixed points

Let $I_i(\omega) = \begin{cases} 1, & \text{if } i \text{ is a fixed point of } \omega \in \Omega, \\ 0, & \text{otherwise.} \end{cases}$

By linearity of expectation

Then, $X = I_1 + \cdots + I_n$ and $\mathbb{E}[X] = \mathbb{E}[I_1 + \cdots + I_n] = \mathbb{E}[I_1] + \cdots + \mathbb{E}[I_n]$

These indicators are NOT independent

$$\mathbb{P}[I_i = 1] = \frac{(n-1)!}{n!} = \frac{1}{n} \text{ and } \mathbb{E}[I_i] = \frac{1}{n} \text{ for all } i = 1, \ldots, n.$$

Therefore, $\mathbb{E}[X] = 1$.