# Chapter 2 Part 2: Conditional probability and independence

Jess Kunke

MATH/STAT 394: Probability I (Summer 2022 A-term)

# Outline

# Outline

The full house problem

Recap of Monday's lecture

Conditional independence

Independent random variables, independent trials

Some common discrete distributions

Odds and ends

## The full house problem

Suppose you are dealt five cards from a standard 52-card deck. How many ways can you get a full house?

- Let's see two ways to solve this problem
- Start with an easier, related problem: choosing captain and co-captain from 12-member basketball team

# Outline

## Recap of Monday's lecture

- **Conditional probability:** $P(A \mid B) = \frac{P(AB)}{P(B)}$
- **Multiplication rule:**

$$P(A_1, ..., A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2) \cdots P(A_n|A_1 \cdots A_n).$$

- **Law of total probability:** If $B_1, ..., B_n$ is a partition of $\Omega$ with $P(B_i) > 0 \; \forall \, i = 1, ..., n$, then for any event $A$ we have

$$P(A) = \sum_{i=1}^{n} P(AB_i) = \sum_{i=1}^{n} P(A|B_i)P(B_i).$$

- **Bayes' formula:** If $B_1, ..., B_n$ partition the sample space $\Omega$ and $P(B_i) > 0$ for all $i$, then for any event $A$ with $P(A) > 0$, and any $k = 1, ..., n$,

$$P(B_k \mid A) = \frac{P(AB_k)}{P(A)} = \frac{P(A \mid B_k)P(B_k)}{\sum_{i=1}^{n} P(A \mid B_i)P(B_i)}.$$

- Two events $A$ and $B$ are **independent** if

$$P(AB) = P(A)P(B).$$

# Independence

If two events A and B are independent, is their intersection empty (no overlap between A and B)?

i It has to be. (If so, show why)

ii It cannot be. (If so, show why)

iii It can be. (If so, demonstrate two events A and B for which it is and another choice of A and B for which it is not.)

# Recap of Monday's lecture, continued

- **Mutual** vs. **pairwise independence**
- Independence of complements
- **Random variable** $X : \Omega \to \mathbb{R}$
- The **probability distribution** of $X$ is the collection of probabilities $P\{X \in B\}$ for sets $B \subseteq \mathbb{R}$.
- $X$ is a **discrete random variable** if $\exists$ a finite or countably infinite set $\{k_1, k_2, ...\}$ of real numbers such that

$$\sum_i P(X = k_i) = 1.$$

- The **probability mass function** or pmf of a discrete random variable $X$ is the function $p$ (or $p_X$) defined by

$$p(k) = P(X = k)$$

  for all possible values $k$ of $X$.

# Outline

## Conditional independence

We said before that conditioning on an event gives us a new probability measure: Given $B$ s.t. $P(B) > 0$, then

$$P(\cdot \mid B) : A \to P(A \mid B) \text{ is a probability measure.}$$

We can now define events as conditionally independent if they are independent under this new measure.

### Definition

Let $B \subseteq \Omega$ s.t. $P(B) > 0$. The events $A_1, A_2$ are **conditionally independent given $B$** if

$$P(A_1 \cap A_2 \mid B) = P(A_1 \mid B)P(A_2 \mid B).$$

More generally, events $A_1, ..., A_n$ are conditionally independent given $B$ if for any $k \in \{2, ..., n\}$ and $1 \leq i_1 \leq i_2 \leq ... \leq i_k \leq n$,

$$P(A_{i_1} \cdots A_{i_k} \mid B) = P(A_{i_1} \mid B)P(A_{i_2} \mid B) \cdots P(A_{i_k} \mid B)$$

# Conditional Independence

## Example

Suppose 90% of coins in the circulation are fair and 10% are biased with $P(\mathrm{T}) = \frac{3}{5}$. I have a random coin and flip it twice. Denote $A_1 = \{\text{1st flip is T}\}$ and $A_2 = \{\text{2nd flip is T}\}$. Are $A_1, A_2$ independent?

- Denote $F = \{\text{the coin is fair}\}$, $B = \{\text{the coin is biased}\}$
- By the law of total probability, for $i = 1$ or $2$,

$$P(A_i) = P(A_i \mid F)P(F) + P(A_i \mid B)P(B)$$

- For **a given coin** the events $A_i$ have the same probability:

$$P(A_1 \mid F) = P(A_2 \mid F) = \frac{1}{2}, \quad P(A_1 \mid B) = P(A_2 \mid B) = \frac{3}{5}$$

- Therefore

$$P(A_i) = P(A_i \mid F)P(F) + P(A_i \mid B)P(B) = \frac{1}{2} \cdot \frac{9}{10} + \frac{3}{5} \cdot \frac{1}{10} = \frac{51}{100}.$$

# Conditional Independence

## Example

Suppose 90% of coins in the circulation are fair and 10% are biased with $P(\mathrm{T}) = \frac{3}{5}$. I have a random coin and flip it twice. Denote $A_1 = \{\text{1st flip is T}\}$ and $A_2 = \{\text{2nd flip is T}\}$. Are $A_1, A_2$ independent?

- Now assume that for **a given coin**, the two events are conditionally independent (natural assumption), i.e.,

$$P(A_1 \cap A_2 \mid F) = P(A_1 \mid F)P(A_2 \mid F), \quad P(A_1 \cap A_2 \mid B) = P(A_1 \mid B)P(A_2 \mid B)$$

- Then by the law of total probability,

$$\begin{aligned} P(A_1 \cap A_2) &= P(A_1 \cap A_2 \mid F)P(F) + P(A_1 \cap A_2 \mid B)P(P) \\ &= P(A_1 \mid F)P(A_2 \mid F)P(F) + P(A_1 \mid B)P(A_2 \mid B)P(B) \\ &= \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{9}{10} + \frac{3}{5} \cdot \frac{3}{5} \cdot \frac{1}{10} = \frac{261}{1000} \end{aligned}$$

- Then $P(A_1 \cap A_2) = \frac{261}{1000} \neq \left(\frac{51}{100}\right)^2 = P(A_1)P(A_2)$, the two events **are not independent**.

## Example

Suppose 90% of coins in the circulation are fair and 10% are biased with $P(\mathrm{T}) = \frac{3}{5}$. I have a random coin and flip it twice. Denote $A_1 = \{\text{1st flip is T}\}$ and $A_2 = \{\text{2nd flip is T}\}$. Are $A_1, A_2$ independent?

Discussion

- Intuitively, why are the two events not independent?
- Seeing the result of one flip gives us some information about the coin, which influences the probability of getting a tail a second time
- e.g. if $A_1 = \{\text{first 100 flips are T}\}$ and $A_2 = \{\text{101th flip is T}\}$, then clearly if $A_1$ is true, the coin has more chances to be biased and so $P(A_2)$ is influenced by this information

## Another interpretation

Conditional independence tells us that, given some information $B$, another event $A_2$ is no longer relevant

### Lemma

*If $A_1$ and $A_2$ are conditionally independent given $B$, then*

$$P(A_2 \mid A_1, B) = P(A_2 \mid B).$$

### Proof.

□

# Conditional Independence

## Example

Every day I walk a random number of kilometers. The distance I walk one day is independent of the distance I walked another day. Let $X_n$ the distance that I walked after $n$ days. Are the events $\{X_1 = 10\}$ and $\{X_3 = 20\}$ conditionally independent given $\{X_2 = 15\}$?

- This is just an intuitive example; we won't dive into this kind of problem during the course
- Yes, if we know $X_2$, then $X_1$ is not relevant:

$$P(X_3 = 20 \mid X_2 = 15, X_1 = 10) = P(X_3 = 15 \mid X_2 = 15).$$

- This is an example of a Markov chain, a sequence of events such that the future is independent of the past given the present
- This is a very common model that can be used e.g. to predict the weather

# Outline

# Independence of random variables

## Definition

Random variables $X_1, ..., X_n$ defined on the same probability space $(\Omega, \mathcal{F}, P)$ —i.e. with the same sample space, event space, and probability measure — are **independent** if

$$P(X_1 \in B_1, ..., X_n \in B_n) = P(X_1 \in B_1)...P(X_n \in B_n)$$

for any (Borel) subsets $B_1, ..., B_n \subseteq \mathbb{R}$.

- This means that the distribution of the r.v. can be factorized in the distributions of each r.v.
- We will generally not use this more abstract, general definition

Specifically, discrete random variables $X_1, ..., X_n$ on the same probability space are independent if and only if

$$P(X_1 = x_1, ..., X_n = x_n) = P(X_1 = x_1)...P(X_n = x_n)$$

for any possible choices $x_1, ..., x_n$ of the values of the random variables.

# Independent trials

Several of the examples we have seen can be viewed as independent repeated trials of an experiment:

- Flipping the same coin $n$ times
- Rolling the same die $n$ times
- Drawing a ball from the same urn $n$ times with replacement

Each trial (coin flip, die roll) has the same possible outcomes whose probabilities are not affected by previous trials (flips/rolls)

Recall that we can represent the sample space in these examples as a Cartesian product of a set that represents the simpler experiment:

- Coin toss: $\Omega = \{\text{H,T}\}^n$
- Die roll: $\Omega = ?$
- Drawing a ball: $\Omega = ?$

# Identically distributed variables

This experimental setup is related to the concept of **independent and identically distributed (i.i.d.)** random variables.

## Definition

Random variables $X_1, ..., X_n$ are **identically distributed** if each $X_i$ has the same probability distribution.

- iid random variables have very nice properties and in some applications they either come up naturally or can be a reasonable assumption
  - Independent repeated trials
- In some other applications, i.i.d. is not a reasonable assumption

## Example

Consider sampling $k$ balls one at a time from an urn with $n$ balls labeled $1, \ldots, n$. Denote

$$X_i = \text{label of the } i\text{th ball drawn.}$$

- If sampling with replacement, then $X_1, \ldots, X_n$ are iid
- If sampling without replacement, are $X_1, \ldots, X_n$ independent? Identically distributed? Both? Neither?

## iid: continuing the example

Suppose you draw $k$ balls one at a time without replacement from an urn with $n$ balls labeled $1, \ldots, n$. Denote

$$X_i = \text{label of the } i\text{th ball drawn.}$$

- The distribution of balls you can get changes with each draw; that means the *conditional probability* $P(X_i | X_1, X_2, ..., X_{i-1})$ is different for different $i$. Show this with an example.

# iid: continuing the example

Suppose you draw $k$ balls one at a time without replacement from an urn with $n$ balls labeled $1, \ldots, n$. Denote

$$X_i = \text{label of the } i\text{th ball drawn.}$$

- The distribution of balls you can get changes with each draw; that means the *conditional probability* $P(X_i|X_1, X_2, ..., X_{i-1})$ is different for different $i$. Show this with an example.
- What about $P(X_i)$? $P(X_i)$ is actually the same for all the variables

# iid: continuing the example

Suppose you draw $k$ balls one at a time without replacement from an urn with $n$ balls labeled $1, \ldots, n$. Denote

$$X_i = \text{label of the } i\text{th ball drawn.}$$

- The distribution of balls you can get changes with each draw; that means the *conditional probability* $P(X_i|X_1, X_2, ..., X_{i-1})$ is different for different $i$. Show this with an example.
- What about $P(X_i)$? $P(X_i)$ is actually the same for all the variables
  - This is called **marginal probability**
- The $X_i$ are not independent

## iid: continuing the example

Suppose you draw $k$ balls one at a time without replacement from an urn with $n$ balls labeled $1, \ldots, n$. Denote

$$X_i = \text{label of the } i\text{th ball drawn.}$$

- The distribution of balls you can get changes with each draw; that means the *conditional probability* $P(X_i | X_1, X_2, ..., X_{i-1})$ is different for different $i$. Show this with an example.
- What about $P(X_i)$? $P(X_i)$ is actually the same for all the variables
  - This is called **marginal probability**
- The $X_i$ are not independent
- Therefore $X_1, \ldots, X_n$ are identically distributed but not independent.

Takeaway: be careful about whether your intuition is based on conditional or marginal probability

# Outline

# Bernoulli distribution

## Definition

A random variable $X$ has a Bernoulli distribution with success probability $p$ if $X \in \{0, 1\}$ (binary) and satisfies $P(X = 1) = p$. We denote this by $X \sim \text{Ber}(p)$, read "$X$ is distributed Bernoulli with success probability $p$."

- What must $P(X = 0)$ be equal to? Why?
- We call $p$ a **parameter** of the distribution, a fixed number used as part of computing the probability of any given outcome.

A sequence of independent $\text{Ber}(p)$ trials: Let $X_i$ denote the outcome of trial $i$. Then, for example,

$$P(X_1 = 0, X_2 = 1, X_3 = 1, X_4 = 0, X_5 = 1) = p^3(1 - p)^2.$$

- What is $P(X_1 = 1, X_2 = 1, X_3 = 1, X_4 = 0, X_5 = 1)$?
- Suppose heads are "successes" with probability $p$. What is the probability of 3 heads in a row followed by 2 tails in a row?
- What is the probability of the sequence $(H, T, H, T, H)$?
- What is the probability that you get either $(H, H, H, T, T)$ or $(H, T, H, T, H)$?

# Binomial distribution

The binomial distribution counts the number of successes in a sequence of $n$ independent Ber($n$) trials

## Definition

Let $n$ be a positive integer and $0 \leq p \leq 1$. A random variable $Y$ **is binomially distributed** or **follows a binomial distribution** with parameters $n$ and $p$ if the pmf of $Y$ is given by

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{for } k = 0, 1, ..., n.$$

We write that $Y \sim \text{Bin}(n, p)$.

## Example

What is the probability that five rolls of a fair die yield two or three sixes?

- Repeated trial = die roll; success = roll a six
- Let $S$ be the number of sixes that appear in 5 rolls
- Then $S$ is binomial. What are $n$ and $p$?
- Finish solving this problem.

## Binomial distribution: Solution to previous example

# Geometric distribution

In a series of independent repeated trials, the geometric distribution gives the probability that the $k$th trial is the first success. Let's compute its pmf:

## Definition

A random variable $G$ follows the **geometric distribution** with success probability $0 \leq p \leq 1$ if the pmf of $G$ is given by

$$P(G = g) = (1 - p)^{g-1} p \quad \forall\, g \in \mathbb{N}.$$

(Recall that $\mathbb{N}$ is the set of positive integers.) We write $G \sim \text{Geom}(p)$.

# Geometric distribution

1. What is the probability that the first tails happens on the 5th flip of a fair coin?

2. What if the coin is biased and $P(\text{H}) = 0.3$? (Tip: what is a "success" here?)
    - Should this value be larger or smaller than the answer to #1?

3. What is the probability that it takes more than 7 rolls of a fair die to roll a six?

# Hypergeometric distribution

- Draw $n$ things without replacement from a finite population of $N$ things that contains $K$ things with some feature of interest
- What is the probability that $k$ of the $n$ things you drew have that feature?
- Example: urn contains 5 balls of each of 4 colors (green, yellow, blue, red), and you draw 3 balls. $P(2 \text{ of the 3 balls are green}) = ?$
- Binomial distribution is the same but for sampling *with* replacement
  - Therefore as $N$ and $K$ become large (and if $p$ is not too small or too large), the binomial becomes a good approximation (think about previous DNA vs Scrabble problem)

## Definition

A random variable $H$ follows a **hypergeometric distribution** with parameters $N$, $K$, and $n$ if the pmf of $H$ is

$$P(H = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} \quad \text{for } k = 0, 1, ..., n.$$

# Outline

# The birthday problem

You meet someone at a party and find out they have the same birthday as you (month and day)! What is the probability of that happening?
No, really, what is the probability of that happening?

Note that there are two different questions we could be asking:

- What is the probability that among $k$ (or $k-1$) people I meet, someone has my birthday?
- What is the probability that among $k$ people, two people have the same birthday?

These two questions have very different answers; see the book for details if you are curious.

# Degenerate random variables

### Definition

A random variable is **degenerate** if $\exists\ b \in \mathbb{R}$ such that $P(X = b) = 1$.

### Example

Consider drawing a number uniformly at random from the interval $[0, 10]$.

- Define a random variable $X$ such that $X$ is degenerate.