Classical random variables
Conditional Independence

MATH/STAT 394: Probability I
Summer 2021 A Term

Introduction to Probability
D. Anderson, T.Seppäläinen, B. Valkó

§ 2.4, 2.5

Aaron Osgood-Zimmerman

Department of Statistics

Recap

Independent r.v.

- a **random variable** (r.v.) X is a function from Ω to $\mathbb{R}$

Recap

Independent r.v.

- a **random variable** (r.v.) X is a function from Ω to $\mathbb{R}$
- a r.v. is **discrete** if there exists a countable set $\mathcal{X}$ s.t. $\sum_{k \in \mathcal{X}} \mathbb{P}(X = k) = 1$

Recap

Independent r.v.

- a **random variable** (r.v.) X is a function from Ω to $\mathbb{R}$
- a r.v. is **discrete** if there exists a countable set $\mathcal{X}$ s.t. $\sum_{k\in\mathcal{X}} \mathbb{P}(X = k) = 1$
- the **probability mass function** (p.m.f.) of a discrete r.v. is

$$p_X : k \to \mathbb{P}(X = k) \quad \text{for } k \in \mathcal{X}$$

Recap

Independent r.v.

- a **random variable** (r.v.) X is a function from Ω to $\mathbb{R}$
- a r.v. is **discrete** if there exists a countable set $\mathcal{X}$ s.t. $\sum_{k\in\mathcal{X}} \mathbb{P}(X = k) = 1$
- the **probability mass function** (p.m.f.) of a discrete r.v. is

$$p_X : k \to \mathbb{P}(X = k) \quad \text{for } k \in \mathcal{X}$$

- r.v. $X_1, \dots X_n$ are **independent** if for any subsets $B_1, \dots, B_n \subseteq \mathbb{R}$

$$\mathbb{P}(X_1 \in B_1, \dots, X_n \in B_n) = \mathbb{P}(X_1 \in B_1) \dots \mathbb{P}(X_n \in B_n)$$

Recap

Independent r.v.

- a **random variable** (r.v.) X is a function from Ω to $\mathbb{R}$
- a r.v. is **discrete** if there exists a countable set $\mathcal{X}$ s.t. $\sum_{k\in\mathcal{X}} \mathbb{P}(X = k) = 1$
- the **probability mass function** (p.m.f.) of a discrete r.v. is

$$p_X : k \to \mathbb{P}(X = k) \quad \text{for } k \in \mathcal{X}$$

- r.v. $X_1, \dots X_n$ are **independent** if for any subsets $B_1, \dots, B_n \subseteq \mathbb{R}$

$$\mathbb{P}(X_1 \in B_1, \dots, X_n \in B_n) = \mathbb{P}(X_1 \in B_1) \dots \mathbb{P}(X_n \in B_n)$$

- **discrete** r.v. $X_1, \dots X_n$ are **independent** if for any possible $x_1, \dots, x_n$

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \mathbb{P}(X_1 = x_1) \dots \mathbb{P}(X_n = x_n)$$

Recap

Independent r.v.

- a **random variable** (r.v.) X is a function from Ω to $\mathbb{R}$
- a r.v. is **discrete** if there exists a countable set $\mathcal{X}$ s.t. $\sum_{k\in\mathcal{X}} \mathbb{P}(X=k) = 1$
- the **probability mass function** (p.m.f.) of a discrete r.v. is

$$p_X : k \to \mathbb{P}(X=k) \quad \text{for } k \in \mathcal{X}$$

- r.v. $X_1, \dots X_n$ are **independent** if for any subsets $B_1, \dots, B_n \subseteq \mathbb{R}$

$$\mathbb{P}(X_1 \in B_1, \dots, X_n \in B_n) = \mathbb{P}(X_1 \in B_1) \dots \mathbb{P}(X_n \in B_n)$$

- **discrete** r.v. $X_1, \dots X_n$ are **independent** if for any possible $x_1, \dots, x_n$

$$\mathbb{P}(X_1 = x_1, \dots, X_n = x_n) = \mathbb{P}(X_1 = x_1) \dots \mathbb{P}(X_n = x_n)$$

- r.v. $X_1, \dots X_n$ are **identically distributed** if

$$\mathbb{P}(X_i \in B) = \mathbb{P}(X_j \in B) \quad \text{for any } i, j \text{ and } B \subseteq \mathbb{R}$$

Recap

Independent r.v.

- a **random variable** (r.v.) X is a function from Ω to $\mathbb{R}$
- a r.v. is **discrete** if there exists a countable set $\mathcal{X}$ s.t. $\sum_{k \in \mathcal{X}} \mathbb{P}(X = k) = 1$
- the **probability mass function** (p.m.f.) of a discrete r.v. is

$$p_X : k \to \mathbb{P}(X = k) \quad \text{for } k \in \mathcal{X}$$

- r.v. $X_1, \ldots X_n$ are **independent** if for any subsets $B_1, \ldots, B_n \subseteq \mathbb{R}$

$$\mathbb{P}(X_1 \in B_1, \ldots, X_n \in B_n) = \mathbb{P}(X_1 \in B_1) \ldots \mathbb{P}(X_n \in B_n)$$

- **discrete** r.v. $X_1, \ldots X_n$ are **independent** if for any possible $x_1, \ldots, x_n$

$$\mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \mathbb{P}(X_1 = x_1) \ldots \mathbb{P}(X_n = x_n)$$

- r.v. $X_1, \ldots X_n$ are **identically distribute**d if

$$\mathbb{P}(X_i \in B) = \mathbb{P}(X_j \in B) \quad \text{for any } i, j \text{ and } B \subseteq \mathbb{R}$$

- "i.i.d." stands for **independent and identically distributed**

Outline

Bernoulli random variable

Reminder:

Definition

A r.v. X has a ***Bernoulli*** *dist. with param. $p \in [0,1]$ if it takes its values in $\{0,1\}$ and*

$$\mathbb{P}(X=1)=p \quad \mathbb{P}(X=0)=1-p$$

We denote it $X \sim \mathrm{Ber}(p)$

Binomial random variable

Many random variables *arise from repeated trials*.

Definition

*A r.v. X has a **Binomial** distribution with parameters $n \in \mathbb{N}$, $n > 0$, and $p \in [0, 1]$, if the possible values of X are $\{0, \dots, n\}$ and*

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

We denote it $X \sim \mathrm{Bin}(n, p)$.

Alternative definition

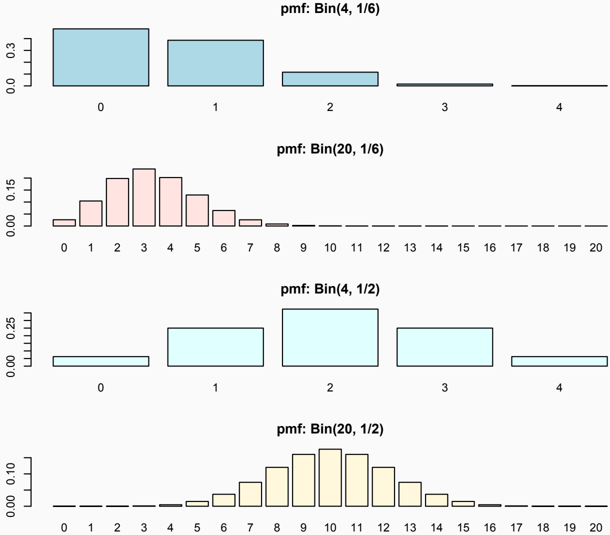
$X \sim \mathrm{Bin}(n, p)$ if and only if $X = Y_1 + \ldots + Y_n$ for $Y_i \overset{i.i.d.}{\sim} \mathrm{Ber}(p)$

where $Y_i \overset{i.i.d.}{\sim} \mathrm{Ber}(p)$ means that the Y_i are independent and identically distributed with a dist. $\mathrm{Ber}(p)$.

Notes:

- For $n = 1$, we retrieve the Bernoulli dist.

Binomial random variable



Geometric random variable

Example
Take a coin whose probability of $\mathtt{H}$ is p. Toss until the the first $\mathtt{H}$.

Recall that this is an experiment defined on

$$\Omega_\infty = \{\text{all infinite sequences of } \{\mathtt{H}, \mathtt{T}\}\}.$$

Let

$$Y = \text{the total number of tosses.}$$

What is the p.m.f. of Y?

Geometric random variable

Example
Take a coin whose probability of $\mathtt{H}$ is p. Toss until the the first $\mathtt{H}$.

Recall that this is an experiment defined on

$$\Omega_\infty = \{\text{all infinite sequences of } \{\mathtt{H}, \mathtt{T}\}\}.$$

Let

$$Y = \text{the total number of tosses.}$$

What is the p.m.f. of Y?

Solution

- Already seen:

$$\mathbb{P}(Y = k) = \mathbb{P}(X_1 = \cdots = X_{k-1} = 0, X_k = 1) = (1-p)^{k-1}p,$$

$$\boxed{p_Y(k) = (1-p)^{k-1}p, \quad k = 1, 2, 3, \ldots.}$$

- Exercise: show that $\sum_{k=1}^{+\infty} p_Y(k) = 1$.
  (Hint: this is the sum of a geometric sequence)

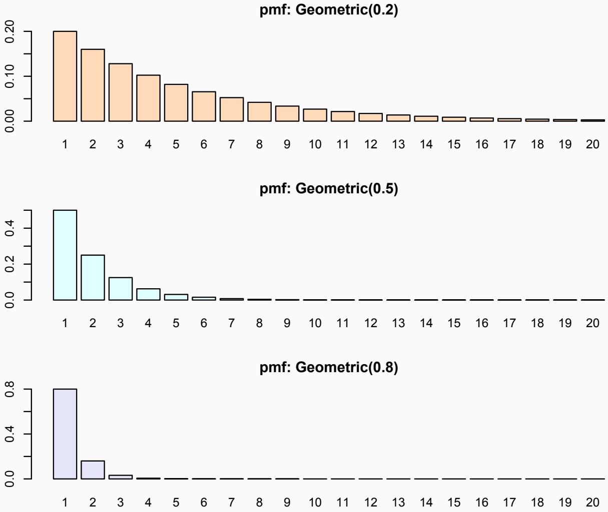
Geometric random variable

Definition

A r.v. X has a ***Geometric*** *dist. with param. $p \in [0,1]$ if it takes its values in $\{1, 2, \ldots\}$ and*

$$p_X(k) = \mathbb{P}(X = k) = (1-p)^{k-1} p$$

We denote it $X \sim \mathrm{Geom}(p)$.


pmf: Geometric(0.2)
0.00 0.10 0.20
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
pmf: Geometric(0.5)
0.0 0.2 0.4
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
pmf: Geometric(0.8)
0.0 0.4 0.8
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

credit: R. Guo

Hypergeometric distribution

Example
An urn contains N balls: N_A labelled as `A`, $N - N_A$ labelled as `B`.
Draw n ($n \leq N$) **without replacement**, and let

$$X = \text{the number of } \texttt{A} \text{ balls.}$$

What is the p.m.f. of X?

Hypergeometric distribution

Example

An urn contains N balls: N_A labelled as $\mathtt{A}$, $N - N_A$ labelled as $\mathtt{B}$.
Draw n ($n \leq N$) **without replacement**, and let

$$X = \text{the number of } \mathtt{A} \text{ balls.}$$

What is the p.m.f. of X?

Solution

- Denote k a possible value for X. Note that $0 \leq k \leq \min(N, N_A)$.
- Moreover one cannot take more than $N - N_A$ B balls from the urn, so $0 \leq n - k \leq N - N_A$, that is $n - (N - N_A) \leq k$. Overall the possible values of k are

$$\max(0, n - (N - N_A)) \leq k \leq \min(N_A, n).$$

- Now following the definition we have for a possible k,

$$\boxed{p_X(k) = P(X = k) = \frac{\binom{N_A}{k}\binom{N-N_A}{n-k}}{\binom{N}{n}}, \quad k = 0, 1, \ldots, n,}$$

Hypergeometric distribution

Definition

*A r.v. X has a **Hypergeometric** dist. with param. n, N_A, N if it takes values k such that*

$$\max(0, n-(N-N_A)) \leq k \leq \min(N_A, n).$$

and

$$p_X(k) = P(X=k) = \frac{\binom{N_A}{k}\binom{N-N_A}{n-k}}{\binom{N}{n}}, \quad k = 0, 1, \dots, n,$$

We denote it $X \sim \mathrm{Hypergeom}(N, N_A, n)$.

Hypergeometric distribution

Definition

A r.v. X has a ***Hypergeometric*** *dist. with param. n, N_A, N if it takes values k such that*

$$\max(0, n-(N-N_A)) \leq k \leq \min(N_A, n).$$

and

$$p_X(k) = P(X=k) = \frac{\binom{N_A}{k}\binom{N-N_A}{n-k}}{\binom{N}{n}}, \quad k = 0, 1, \ldots, n,$$

We denote it $X \sim \text{Hypergeom}(N, N_A, n)$.

Note:

The range of possible values of k is directly given in the p.m.f. since, for k an integer, $P_X(k) = 0$ if k does not satisfy $\max(0, n-(N-N_A)) \leq k \leq \min(N_A, n)$.

Sampling n times with/without replacement

- *Sampling without replacement*: $X \sim \text{Hypergeom}(N, N_A, n)$

$$\boxed{p_X(k) = \frac{\binom{N_A}{k}\binom{N-N_A}{n-k}}{\binom{N}{n}}, \quad k = 0, 1, \ldots, n.}$$

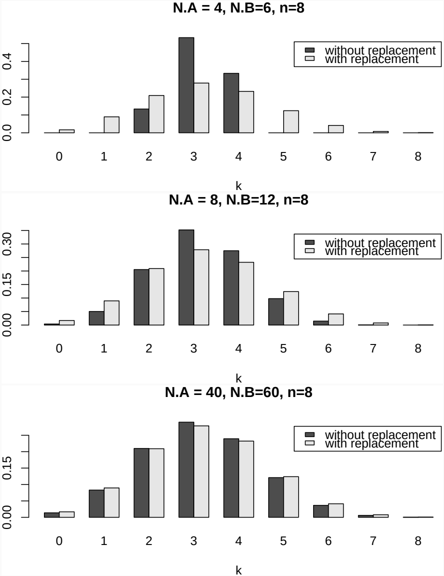
Sampling n times with/without replacement

- *Sampling without replacement*: $X \sim \text{Hypergeom}(N, N_A, n)$

$$\boxed{p_X(k) = \frac{\binom{N_A}{k}\binom{N-N_A}{n-k}}{\binom{N}{n}}, \quad k = 0, 1, \ldots, n.}$$

- *Sampling with replacement*: $Y \sim \text{Bin}(N_A/N, n)$

$$\boxed{p_Y(k) = \binom{n}{k}\left(\frac{N_A}{N}\right)^k\left(1 - \frac{N_A}{N}\right)^{n-k}, \quad k = 0, 1, \ldots, n.}$$


N.A = 4, N.B=6, n=8
without replacement
with replacement
0.0
0.2
0.4
0 1 2 3 4 5 6 7 8
k
N.A = 8, N.B=12, n=8
without replacement
with replacement
0.00
0.15
0.30
0 1 2 3 4 5 6 7 8
k
N.A = 40, N.B=60, n=8
without replacement
with replacement
0.00
0.15
0 1 2 3 4 5 6 7 8
k

credit: R. Guo

Poisson random variable

Example

Now we're going go try something different. Instead of observing some number of indexed discrete events, we're going to watch a process that generates events, with known average time between events, for *a fixed amount of time*, and count how many events occurred.

Poisson random variable

Example

Now we're going go try something different. Instead of observing some number of indexed discrete events, we're going to watch a process that generates events, with known average time between events, for *a fixed amount of time*, and count how many events occurred.

For example, you know the tree out your window well and you know in the fall 1 leaf falls, on average, every 7 minutes. Let's call the average time λ.

Poisson random variable

Example

Now we're going go try something different. Instead of observing some number of indexed discrete events, we're going to watch a process that generates events, with known average time between events, for *a fixed amount of time*, and count how many events occurred.

For example, you know the tree out your window well and you know in the fall 1 leaf falls, on average, every 7 minutes. Let's call the average time λ.

You watch the tree for t minutes and let

$$Y = \text{the number of leaves that fall.}$$

This is an experiment defined on

$$\Omega_\infty = \{0, 1, 2, \ldots\}.$$

What is the p.m.f. of Y?

Poisson random variable

Example

Now we're going go try something different. Instead of observing some number of indexed discrete events, we're going to watch a process that generates events, with known average time between events, for *a fixed amount of time*, and count how many events occurred.

For example, you know the tree out your window well and you know in the fall 1 leaf falls, on average, every 7 minutes. Let's call the average time λ.

You watch the tree for t minutes and let

$$Y = \text{the number of leaves that fall.}$$

This is an experiment defined on

$$\Omega_\infty = \{0, 1, 2, \dots\}.$$

What is the p.m.f. of Y?

Solution

$$\boxed{p_Y(k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}}$$

Geometric random variable

Definition

A r.v. X has a ***Poisson*** *dist. with param. $\lambda \geq 0$ if it takes its values in $\{0, 1, 2, \dots\}$ and*

$$p_X(k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$$

We denote it $X \sim Pois(\lambda)$.

Geometric random variable

Definition

A r.v. X has a ***Poisson*** *dist. with param. $\lambda \geq 0$ if it takes its values in $\{0, 1, 2, \ldots\}$ and*

$$p_X(k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$$

We denote it $X \sim Pois(\lambda)$.

- The Poisson process can be seen to describe a process where the single trial probability of success is very small, but in which the number of trials is so large that there is nonetheless a reasonable rate of events (eg, we could have viewed every millisecond while watching the tree to have been a trial)

Geometric random variable

Definition

A r.v. X has a ***Poisson*** *dist. with param. $\lambda \geq 0$ if it takes its values in $\{0, 1, 2, \dots\}$ and*

$$p_X(k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$$

We denote it $X \sim Pois(\lambda)$.

- The Poisson process can be seen to describe a process where the single trial probability of success is very small, but in which the number of trials is so large that there is nonetheless a reasonable rate of events (eg, we could have viewed every millisecond while watching the tree to have been a trial)
- the Poisson distribution occurs as the limiting form of the Binomial when $p \to 0$ and $N \to \infty$

Geometric random variable

Definition

A r.v. X has a ***Poisson*** *dist. with param. $\lambda \geq 0$ if it takes its values in $\{0, 1, 2, \ldots\}$ and*

$$p_X(k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$$

We denote it $X \sim Pois(\lambda)$.

- The Poisson process can be seen to describe a process where the single trial probability of success is very small, but in which the number of trials is so large that there is nonetheless a reasonable rate of events (eg, we could have viewed every millisecond while watching the tree to have been a trial)
- the Poisson distribution occurs as the limiting form of the Binomial when $p \to 0$ and $N \to \infty$
- Other examples include radioactive decay. ^{137}Cs has a half-life of 27 years, but $1\mu g$ has 10^{15} nuclei. Watching for decays follows a Poisson

Geometric random variable

Definition

A r.v. X has a ***Poisson*** *dist. with param. $\lambda \geq 0$ if it takes its values in $\{0, 1, 2, \ldots\}$ and*

$$p_X(k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$$

We denote it $X \sim Pois(\lambda)$.

- The Poisson process can be seen to describe a process where the single trial probability of success is very small, but in which the number of trials is so large that there is nonetheless a reasonable rate of events (eg, we could have viewed every millisecond while watching the tree to have been a trial)
- the Poisson distribution occurs as the limiting form of the Binomial when $p \to 0$ and $N \to \infty$
- Other examples include radioactive decay. ^{137}Cs has a half-life of 27 years, but $1\mu g$ has 10^{15} nuclei. Watching for decays follows a Poisson
- bombs during the Blitz

Identifying distributions

Exercise

For each of the following examples, select the distribution of the r.v. of interest

1. Draw 4 cards from a deck, $X =$ the number of hearts
2. Observe the weather in Seattle for 7 days. $Y =$ number of times it rains (unique rain events with a break since the last one).
3. Take the bus to school each day for 30 days. $X =$ number of times the bus is late.
4. Survey 100 people and ask which candidate they will vote for, among 4 candidates. $X =$ the number of votes for each candidate.
5. You're stuck in some bad traffic at a stop light. let $X =$ number of light cycles before you get through
6. You're outside a polling station and you ask people who they voted for until you find someone that voted for the socialist candidate in the local election

Identifying distributions

Identifying distributions

Solution

For each of the following examples, select the distribution of the r.v. of interest

1. Draw 4 cards from a deck, $X =$ the number of hearts
2. Observe the weather in Seattle for 7 days. $Y =$ number of times it rains (unique rain events with a break since the last one).
3. Take the bus to school each day for 30 days. $X =$ number of times the bus is late.
4. Survey 100 people and ask which candidate they will vote for, among 4 candidates. $X =$ the number of votes for each candidate.
5. You're stuck in some bad traffic at a stop light. let $X =$ number of light cycles before you get through
6. You're outside a polling station and you ask people who they voted for until you find someone that voted for the socialist candidate in the local election

1. Hypergeometric(52, 4, 4)
2. Pois(7 $*$ avg. number of rain events per day) or Pois(avg. number of rain events per week)
3. Binomial(30, p(bus is late))
4. Multinomial (haven't yet learned) note: it takes on vector values
5. ???
6. Geometric(p(socialist candidate gets a vote))

Outline

Classical random variables

Conditional Independence

Conditional Independence

Motivation

- The conditional probability with respect to an event $\mathbb{P}(\cdot \mid B) : A \to \mathbb{P}(A \mid B)$ is a prob. measure

Conditional Independence

Motivation

- The conditional probability with respect to an event $\mathbb{P}(\cdot \mid B) : A \to \mathbb{P}(A \mid B)$ is a prob. measure
- We can then define independence w.r.t. this prob. measure

Conditional Independence

Motivation

- The conditional probability with respect to an event $\mathbb{P}(\cdot \mid B) : A \to \mathbb{P}(A \mid B)$ is a prob. measure
- We can then define independence w.r.t. this prob. measure
- This can help to further simplify prob. computations.

Conditional Independence

Motivation

- The conditional probability with respect to an event $\mathbb{P}(\cdot \mid B) : A \to \mathbb{P}(A \mid B)$ is a prob. measure
- We can then define independence w.r.t. this prob. measure
- This can help to further simplify prob. computations.

Conditional Independence

Motivation

- The conditional probability with respect to an event $\mathbb{P}(\cdot \mid B) : A \to \mathbb{P}(A \mid B)$ is a prob. measure
- We can then define independence w.r.t. this prob. measure
- This can help to further simplify prob. computations.

Definition

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

$$\mathbb{P}(A_1 \cap A_2 \mid B) = \mathbb{P}(A_1 \mid B)\mathbb{P}(A_2 \mid B)$$

Conditional Independence

Exercise

Suppose 90% of coins in the circulation are fair and 10% are biased with $\mathbb{P}(\mathtt{T}) = \frac{3}{5}$. I have a random coin and flip it twice.

Denote $A_1 = \{\text{1st flip is tail}\}$ and $A_2 = \{\text{2nd flip is tail}\}$.

Are A_1, A_2 independent?

Conditional Independence

Exercise

Suppose 90% of coins in the circulation are fair and 10% are biased with $\mathbb{P}(\mathtt{T}) = \frac{3}{5}$. I have a random coin and flip it twice.

Denote $A_1 = \{\text{1st flip is tail}\}$ and $A_2 = \{\text{2nd flip is tail}\}$.

Are A_1, A_2 independent?

Solution

- Denote $F = \{\text{the coin is fair}\}$, $B = \{\text{the coin is biased}\}$

Conditional Independence

Exercise

Suppose 90% of coins in the circulation are fair and 10% are biased with $\mathbb{P}(\mathtt{T}) = \frac{3}{5}$. I have a random coin and flip it twice.

Denote $A_1 = \{\text{1st flip is tail}\}$ and $A_2 = \{\text{2nd flip is tail}\}$.

Are A_1, A_2 independent?

Solution

- Denote $F = \{\text{the coin is fair}\}$, $B = \{\text{the coin is biased}\}$
- For **a given coin** the events are identically distributed, i.e.,

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

Conditional Independence

Exercise

Suppose 90% of coins in the circulation are fair and 10% are biased with $\mathbb{P}(\mathtt{T}) = \frac{3}{5}$. I have a random coin and flip it twice.

Denote $A_1 = \{\text{1st flip is tail}\}$ and $A_2 = \{\text{2nd flip is tail}\}$.

Are A_1, A_2 independent?

Solution

- Denote $F = \{\text{the coin is fair}\}$, $B = \{\text{the coin is biased}\}$
- For **a given coin** the events are identically distributed, i.e.,

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

- Then by the law of total prob., for $i = 1$ or 2,

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

Conditional Independence

Conditional Independence

Solution (continued)

90% fair coins, 10% are biased with $\mathbb{P}(\mathtt{T}) = \frac{3}{5}$. I have a random coin that is flipped 2x

Denote $A_1 = \{\text{1st flip is tail}\}$ and $A_2 = \{\text{2nd flip is tail}\}$. Are A_1, A_2 independent?

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

$$\mathbb{P}(A_1 \cap A_2 \mid F) = \mathbb{P}(A_1 \mid F)\mathbb{P}(A_2 \mid F) \quad \mathbb{P}(A_1 \cap A_2 \mid B) = \mathbb{P}(A_1 \mid B)\mathbb{P}(A_2 \mid B)$$

Conditional Independence

Solution (continued)

90% fair coins, 10% are biased with $\mathbb{P}(\texttt{T}) = \frac{3}{5}$. I have a random coin that is flipped 2x

Denote $A_1 = \{\text{1st flip is tail}\}$ and $A_2 = \{\text{2nd flip is tail}\}$. Are A_1, A_2 independent?

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

$$\mathbb{P}(A_1 \cap A_2 \mid F) = \mathbb{P}(A_1 \mid F)\mathbb{P}(A_2 \mid F) \quad \mathbb{P}(A_1 \cap A_2 \mid B) = \mathbb{P}(A_1 \mid B)\mathbb{P}(A_2 \mid B)$$

- Then by the law of total prob.

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

Conditional Independence

Solution (continued)

90% fair coins, 10% are biased with $\mathbb{P}(\mathtt{T}) = \frac{3}{5}$. I have a random coin that is flipped 2x

Denote $A_1 = \{\text{1st flip is tail}\}$ and $A_2 = \{\text{2nd flip is tail}\}$. Are A_1, A_2 independent?

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

$$\mathbb{P}(A_1 \cap A_2 \mid F) = \mathbb{P}(A_1 \mid F)\mathbb{P}(A_2 \mid F) \quad \mathbb{P}(A_1 \cap A_2 \mid B) = \mathbb{P}(A_1 \mid B)\mathbb{P}(A_2 \mid B)$$

- Then by the law of total prob.

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

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

Conditional Independence

Solution (continued)

90% fair coins, 10% are biased with $\mathbb{P}(\mathtt{T}) = \frac{3}{5}$. I have a random coin that is flipped 2x

Denote $A_1 = \{\text{1st flip is tail}\}$ and $A_2 = \{\text{2nd flip is tail}\}$. Are A_1, A_2 independent?

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

$$\mathbb{P}(A_1 \cap A_2 \mid F) = \mathbb{P}(A_1 \mid F)\mathbb{P}(A_2 \mid F) \quad \mathbb{P}(A_1 \cap A_2 \mid B) = \mathbb{P}(A_1 \mid B)\mathbb{P}(A_2 \mid B)$$

- Then by the law of total prob.

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

- Then $\mathbb{P}(A_1 \cap A_2) = \frac{261}{1000} \neq \left(\frac{51}{100}\right)^2 = \mathbb{P}(A_1)\mathbb{P}(A_2)$, the two events **are not independent**
- Why? The first flip gives us some information about the coin, which influences the prob. of getting a tail a second time

Conditional Independence

Solution (continued)
90% fair coins, 10% are biased with $\mathbb{P}(\mathtt{T}) = \frac{3}{5}$. I have a random coin that is flipped 2x

Denote $A_1 = \{\text{1st flip is tail}\}$ and $A_2 = \{\text{2nd flip is tail}\}$. Are A_1, A_2 independent?

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

$$\mathbb{P}(A_1 \cap A_2 \mid F) = \mathbb{P}(A_1 \mid F)\mathbb{P}(A_2 \mid F) \quad \mathbb{P}(A_1 \cap A_2 \mid B) = \mathbb{P}(A_1 \mid B)\mathbb{P}(A_2 \mid B)$$

- Then by the law of total prob.

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

- Then $\mathbb{P}(A_1 \cap A_2) = \frac{261}{1000} \neq \left(\frac{51}{100}\right)^2 = \mathbb{P}(A_1)\mathbb{P}(A_2)$, the two events **are not independent**
- Why? The first flip gives us some information about the coin, which influences the prob. of getting a tail a second time
- Think for example that $A_1 = \{\text{first 100 flips are tail}\}$ and $A_2 = \{\text{101th flip is tail}\}$, clearly if A_1 is true, the coin has more chances to be biased and so the prob. of A_2 is influenced by this information.

Conditional Independence

- 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

$$\mathbb{P}(A_2 \mid A_1, B) = \mathbb{P}(A_2 \mid B)$$

Conditional Independence

- 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

$$\mathbb{P}(A_2 \mid A_1, B) = \mathbb{P}(A_2 \mid B)$$

Proof

$$\begin{aligned}\mathbb{P}(A_2 \mid A_1, B) := \mathbb{P}(A_2 \mid A_1 \cap B) &= \frac{\mathbb{P}(A_2 \cap A_1 \cap B)}{\mathbb{P}(A_1 \cap B)} \\ &= \frac{\mathbb{P}(A_2 \cap A_1 \mid B)}{\mathbb{P}(A_1 \mid B)} = \frac{\mathbb{P}(A_2 \mid B)\mathbb{P}(A_1 \mid B)}{\mathbb{P}(A_1 \mid B)} = \mathbb{P}(A_2 \mid B)\end{aligned}$$

Conditional Independence

Example

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

Conditional Independence

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

Solution This is just an intuitive example, we won't dive into this kind of problems during the course

- Yes, we naturally have that if we know X_2, X_1 is not relevant, namely

$$\mathbb{P}(X_3 = 20 \mid X_2 = 15, X_1 = 10) = \mathbb{P}(X_3 = 15 \mid X_2 = 15)$$

Note:

- 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 for example to predict the weather.

Practice next lecture

Practice

At a lottery, there are 10 out of 100 tickets that have prizes.

1. Consider picking 5 tickets with replacement, what is the prob. that you get exactly 2 prizes? (Namely you pick a ticket, look if you win or not and repeat that 5 times)
2. Consider picking 5 tickets without replacement, what is the prob. that you get exactly 2 prizes?

Practice

Roll a fair die twice, define

$$A = \{\text{first die is a 2 or a 3}\}, B = \{\text{4 appears at least once}\}$$

- Are A, B independent?
- Are A, B conditionally independent given that

$$C = \{\text{the sum of the dice is a 6}\}?$$