Day 1 - A few notes on the probability theory

Flip a coin and compute the likelihood

Kevin Cazelles
University of Guelph

August 14, 2017

UNIVERSITÉ DE SHERBROOKE

PART 1

Probability theory from scratch

Chevalier de Méré (1607-1684), Pascal (1623-1662) and Fermat (1607-1665):

A few notes on History

Chevalier de Méré (1607-1684), Pascal (1623-1662) and Fermat (1607-1665):

- de Méré: wit / courtier (Louis XIV) brought Pascal's attention to 2 problems (1650-1655):

A few notes on History

Chevalier de Méré (1607-1684), Pascal (1623-1662) and Fermat (1607-1665):

- de Méré: wit / courtier (Louis XIV) brought Pascal's attention to 2 problems (1650-1655):

1. dice problem
2. division problem (problèmes des parties)

A few notes on History

Chevalier de Méré (1607-1684), Pascal (1623-1662) and Fermat (1607-1665):

- de Méré: wit / courtier (Louis XIV) brought Pascal's attention to 2 problems (1650-1655):

1. dice problem
2. division problem (problèmes des parties)

- Pascal solved the hardest problem (correspondence with Fermat)

A few notes on History

- **dice problem**:

A few notes on History

- **dice problem**:

  "When one throws with two dice, how many throws must one be allowed in order to have a better than even chance of getting two sixes at least once?"

A few notes on History

- **dice problem**:

  "When one throws with two dice, how many throws must one be allowed in order to have a better than even chance of getting two sixes at least once?"

- **division problem**:

A few notes on History

- **dice problem**:

  "When one throws with two dice, how many throws must one be allowed in order to have a better than even chance of getting two sixes at least once?"

- **division problem**:

  "An elderly nobleman, staying at his country house, was extremely fond of watching ball games, and so he called in two young farmhands, saying, 'Here are four ducats for which you may play; the one who first takes eight games is the winner.' So they began to play, but when one had five games and the other three games, they lost the ball and were unable to finish. The question is how the prize should be divided."

Ore, O. Pascal and the Invention of Probability Theory (1960).

A few notes on History

- Reverend Thomas Bayes (1701 – 1761) *An Essay Towards Solving a Problem in the Doctrine of Chances* (published posthumously)

- Reverend Thomas Bayes (1701 – 1761) *An Essay Towards Solving a Problem in the Doctrine of Chances* (published posthumously)
- Pierre-Simon Laplace (1749–1827) *Théorie Analytique des Probabilités in 1812*

A few notes on History

- Reverend Thomas Bayes (1701 – 1761) *An Essay Towards Solving a Problem in the Doctrine of Chances* (published posthumously)
- Pierre-Simon Laplace (1749–1827) *Théorie Analytique des Probabilités in 1812*

- Huygens, de Moivre, Galton, Gauss, von Mises, Kolmogorov, Neyman, Wiener, Wald, Pearson, Shannon, Fisher

References

1. Hacking I., The Emergence of Probability (2006).

References

1. Hacking I., The Emergence of Probability (2006).
2. Hendricks V. F., Pedersen S. A., Jørgensen K. F. , Probability Theory Philosophy, Recent History and Relations to Science (2001).

1. Hacking I., The Emergence of Probability (2006).
2. Hendricks V. F., Pedersen S. A., Jørgensen K. F. , Probability Theory Philosophy, Recent History and Relations to Science (2001).
3. Brémaud P., An Introduction to Probability Modelling (1988).

Examples

- Allesina and Tang, **Stability criteria for complex ecosystems**, *Nature* (2012),

- Allesina and Tang, **Stability criteria for complex ecosystems**, *Nature* (2012), based on Tao, Vu & Krishnapur, **Random matrices: Universality of ESDs and the circular law**, *The Annals of Probability* (2010).

Examples

- Allesina and Tang, **Stability criteria for complex ecosystems**, *Nature* (2012), based on Tao, Vu & Krishnapur, **Random matrices: Universality of ESDs and the circular law**, *The Annals of Probability* (2010).

- Population genetics: a population = probability distribution of traits (random variables)

Examples

- Allesina and Tang, **Stability criteria for complex ecosystems**, *Nature* (2012), based on Tao, Vu & Krishnapur, **Random matrices: Universality of ESDs and the circular law**, *The Annals of Probability* (2010).

- Population genetics: a population = probability distribution of traits (random variables)

- Expending the classical TIB: short talk this pm!

Let's flip a coin

Let's flip a coin and define $(\Omega, \mathcal{F}, \mathrm{P})$

Let's flip a coin and define $(\Omega, \mathcal{F}, \mathrm{P})$

1. Ω: Sample space, *i.e.* set of all possible outcomes

Let's flip a coin and define $(\Omega, \mathcal{F}, \mathrm{P})$

1. Ω: Sample space, *i.e.* set of all possible outcomes

2. $\mathcal{F}$: set of events *i.e.* combinations of outcomes (σ-field)

Let's flip a coin and define $(\Omega, \mathcal{F}, \mathrm{P})$

1. Ω: Sample space, *i.e.* set of all possible outcomes

2. $\mathcal{F}$: set of events *i.e.* combinations of outcomes (σ-field)

3. P: maps events occurrence into [0,1]

Probability space

Let's flip a coin and define $(\Omega, \mathcal{F}, \mathrm{P})$

1. Ω: Sample space, *i.e.* set of all possible outcomes

   - "Head" "Tail"

2. $\mathcal{F}$: set of events *i.e.* combinations of outcomes (σ-field)

3. P: maps events occurrence into [0,1]

Let's flip a coin and define $(\Omega, \mathcal{F}, \mathrm{P})$

1. Ω: Sample space, *i.e.* set of all possible outcomes
   - "Head" "Tail"
2. $\mathcal{F}$: set of events *i.e.* combinations of outcomes (σ-field)
   - $\emptyset$, "Head", "Tail", "Head or Tail" (Ω)
3. P: maps events occurrence into [0,1]

Let's flip a coin and define $(\Omega, \mathcal{F}, P)$

1. Ω: Sample space, *i.e.* set of all possible outcomes

- "Head" "Tail"

2. $\mathcal{F}$: set of events *i.e.* combinations of outcomes (σ-field)

- $\emptyset$, "Head", "Tail", "Head or Tail" (Ω)

3. P: maps events occurrence into [0,1]

- $P(\emptyset) = 0$; P("Head") = p; P("Tail") = 1 – p; $P(\Omega) = 1$

Occurrence of species 1 on an island

Occurrence of species 1 on an island

1. Ω: Sample space set of all possible outcomes

   - "Present" or "1", "Absent" or "0"

Occurrence of species 1 on an island

1. Ω: Sample space set of all possible outcomes

- "Present" or "1", "Absent" or "0"

2. $\mathcal{F}$: set of events *i.e.* 0 or more outcomes

- $\emptyset$, "0", "1", "1 or 0" (Ω)

Occurrence of species 1 on an island

1. Ω: Sample space set of all possible outcomes
   - "Present" or "1", "Absent" or "0"
2. $\mathcal{F}$: set of events *i.e.* 0 or more outcomes
   - $\emptyset$, "0", "1", "1 or 0" (Ω)
3. P: assign a probability / map events occurrence into [0,1]
   - $P(\emptyset) = 0$; $P(''1'') = p$; $P(''0'') = 1 - p$; $P(\Omega) = 1$

Occurrence of species 1 and species 2 on an island

Occurrence of species 1 and species 2 on an island

1. Ω: {"00", "01", "10", "11"}

Occurrence of species 1 and species 2 on an island

1. Ω: $\{$"00", "01", "10", "11"$\}$

2. $\mathcal{F}$: $\emptyset$, Ω, "01", "at least one species"

Occurrence of species 1 and species 2 on an island

1. Ω: $\{$"00", "01", "10", "11"$\}$
2. $\mathcal{F}$: $\emptyset$, Ω, "01", "at least one species"
3. P: p_{00}, p_{01}, ...

Let “A” and “B” denotes two distinct events:

Combining events

Let "A" and "B" denotes two distinct events:

1. $A \cup B$: "A or B"

Let “A” and “B” denotes two distinct events:

1. $A \cup B$: “A or B”
2. $A \cap B$: “A and B”

Let "A" and "B" denotes two distinct events:

1. $A \cup B$: "A or B"
2. $A \cap B$: "A and B"
3. $\bar{A}$: "complement of A"

Let "A" and "B" denotes two distinct events:

1. $A \cup B$: "A or B"
2. $A \cap B$: "A and B"
3. $\overline{A}$: "complement of A"

NB: $A \cup \overline{A} = \Omega$ and $A \cap \overline{A} = \emptyset$

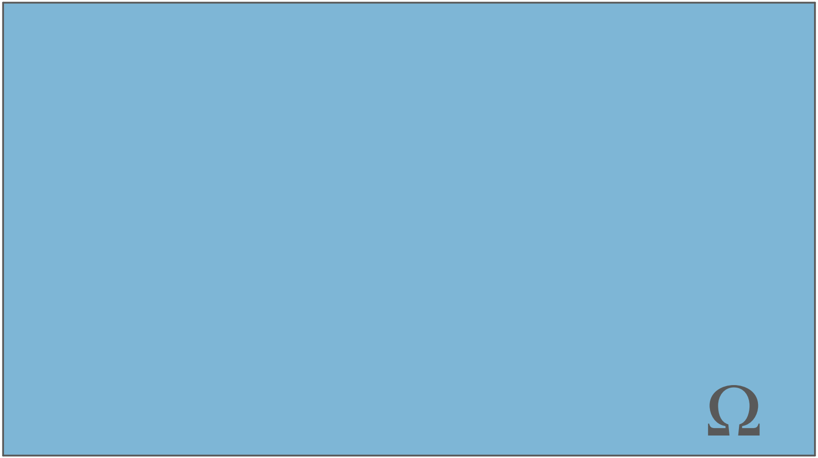
Let "A" and "B" denotes two distinct events:

1. $A \cup B$: "A or B"
2. $A \cap B$: "A and B"
3. $\overline{A}$: "complement of A"

NB: $A \cup \overline{A} = \Omega$ and $A \cap \overline{A} = \emptyset$

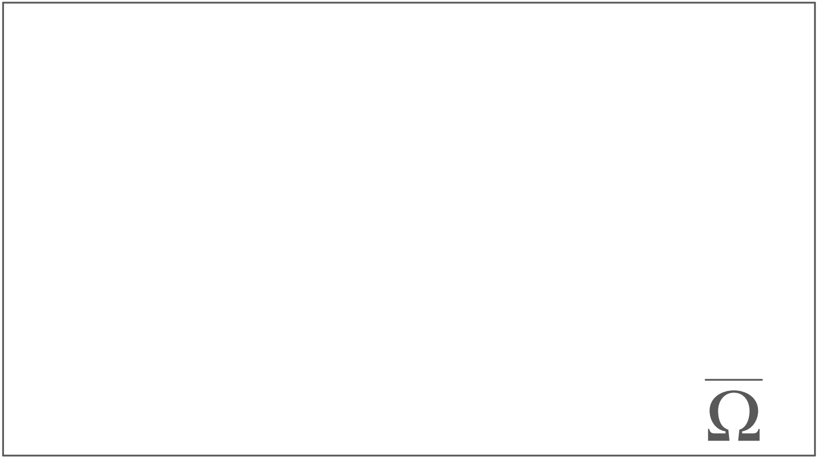
NB: $P(\overline{A}) = 1 = P(A)$

Combining events

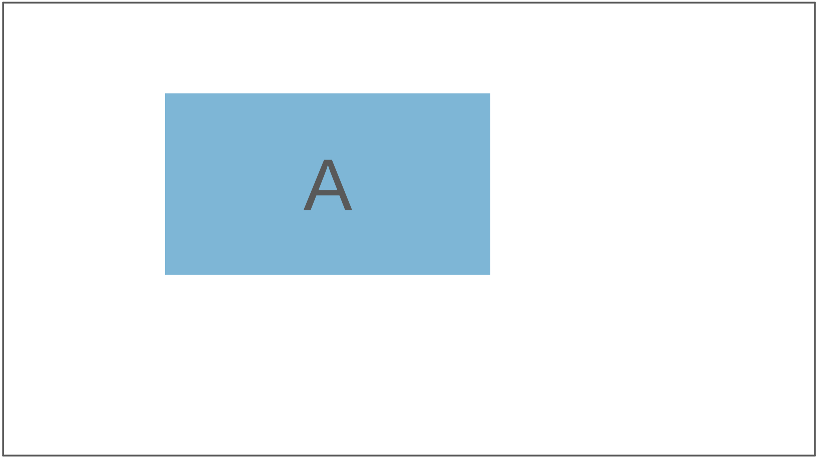


$P(\Omega) = 1$

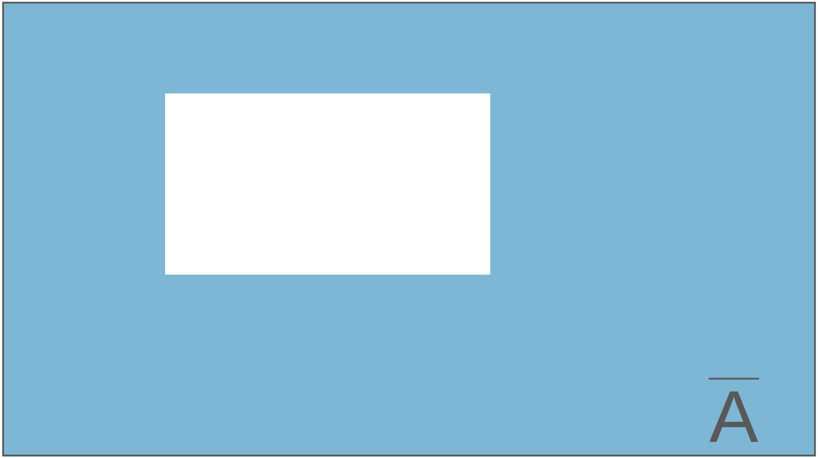
Combining events



$$P(\overline{\Omega}) = P(\emptyset) = 1 - P(\Omega) = 0$$

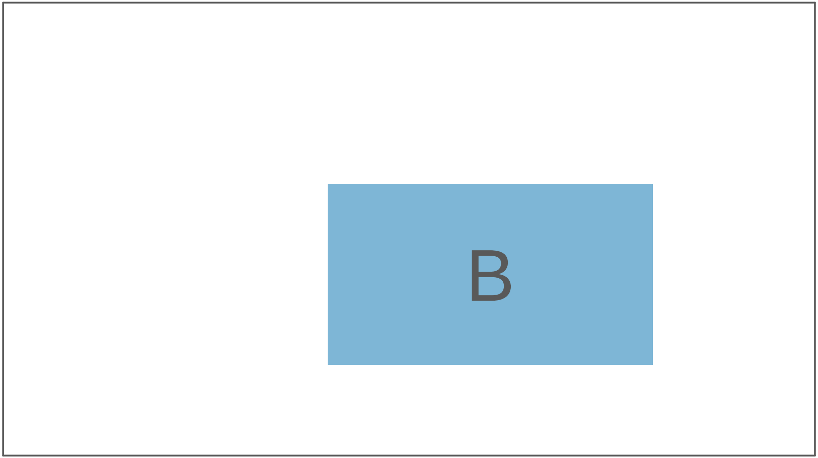


P(A)

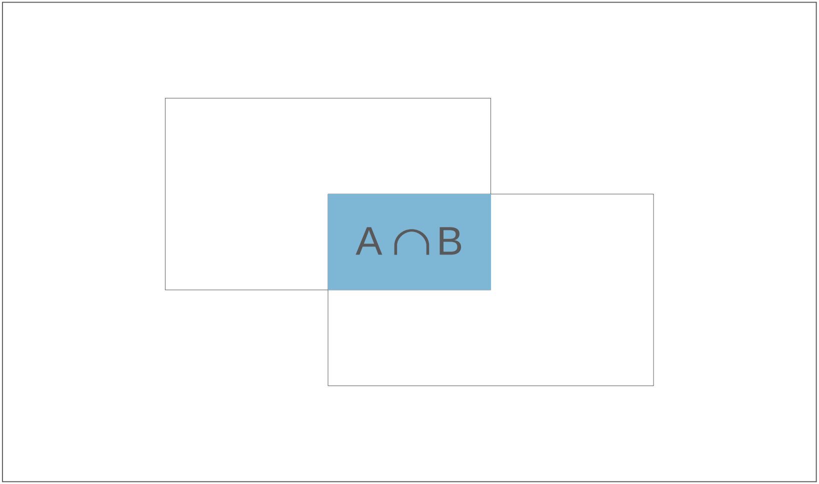


$$P(\overline{A}) = 1 - P(A)$$

Combining events

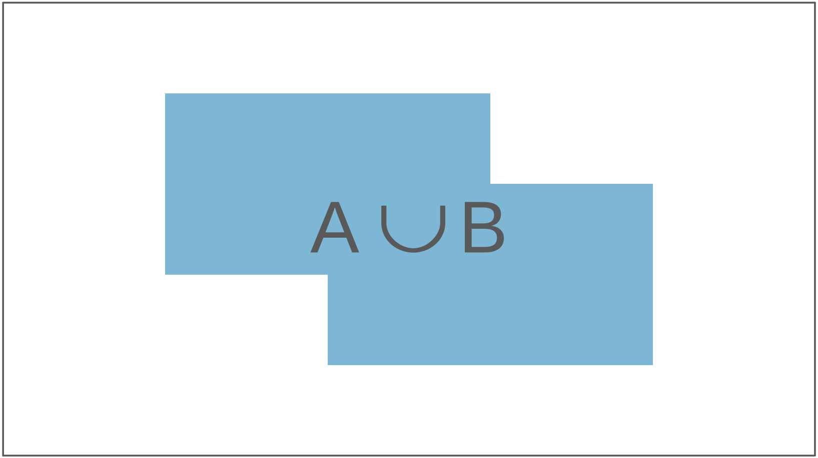


P(B)

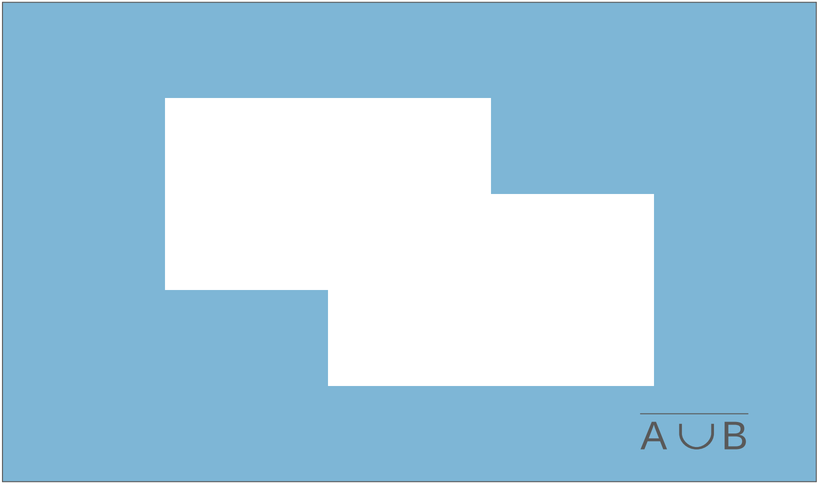


$P(A \cap B) = ?$

Combining events

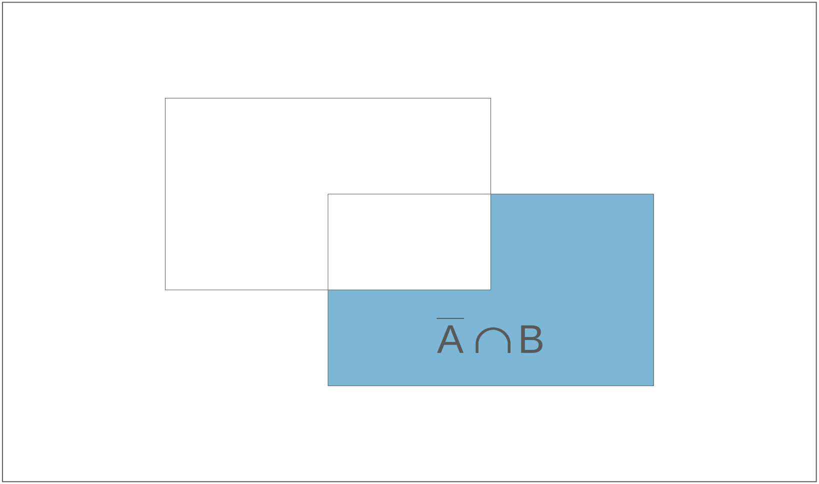


$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



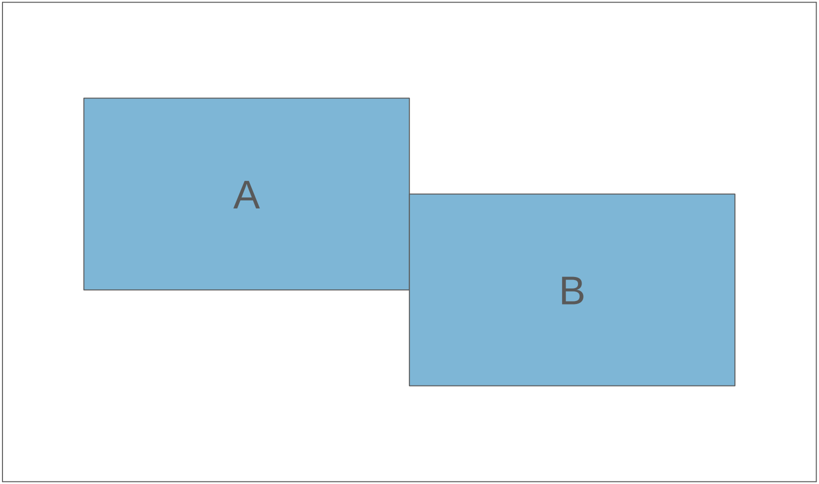
$$P(\overline{A \cup B}) = 1 - P(A \cup B)$$

Combining events



$$P(\overline{A} \cap B) = P(B) - P(A \cap B)$$

Combining events - disjoint events



$A \cap B = \emptyset$; $P(A \cap B) = 0$

Combining events - partition

Consider B

Combining events - partition

Consider B and A_i where $i \in \{1,...,n\}$:

Consider B and A_i where $i \in \{1,\dots,n\}$:

1. $\forall \{i, j\} \setminus i \neq j$, $P(A_i \cap A_j) = 0$ **pairwise disjoint**

Consider B and A_i where $i \in \{1,...,n\}$:

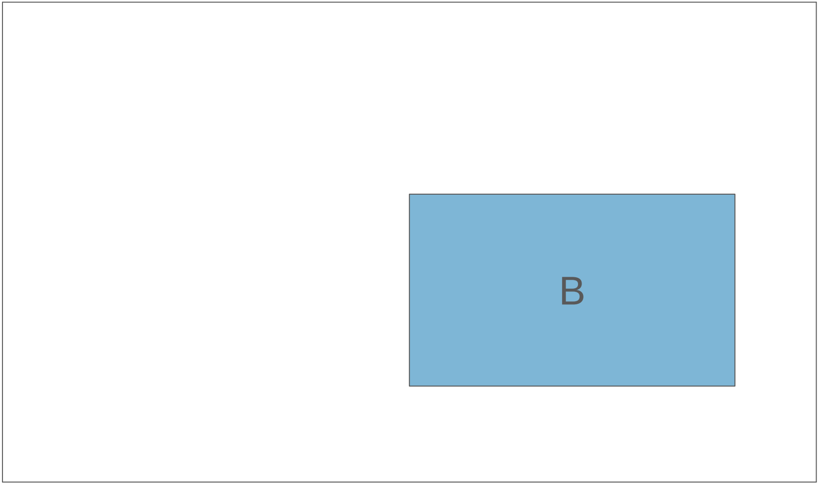
1. $\forall \{i, j\} \setminus i \neq j$, $P(A_i \cap A_j) = 0$ **pairwise disjoint**
2. $\bigcap_i^n A_i = B \Rightarrow \sum_i^n P(A_i) = P(B)$

Consider B and A_i where $i \in \{1, \ldots, n\}$:

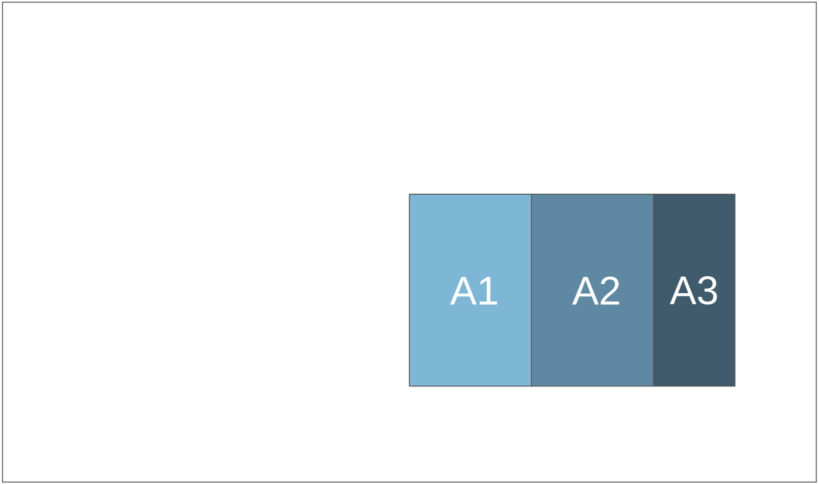
1. $\forall \{i, j\} \setminus i \neq j, P(A_i \cap A_j) = 0$ **pairwise disjoint**
2. $\bigcap_i^n A_i = B \Rightarrow \sum_i^n P(A_i) = P(B)$

then, the set A_i is a **partition** of B.

Combining events - partition

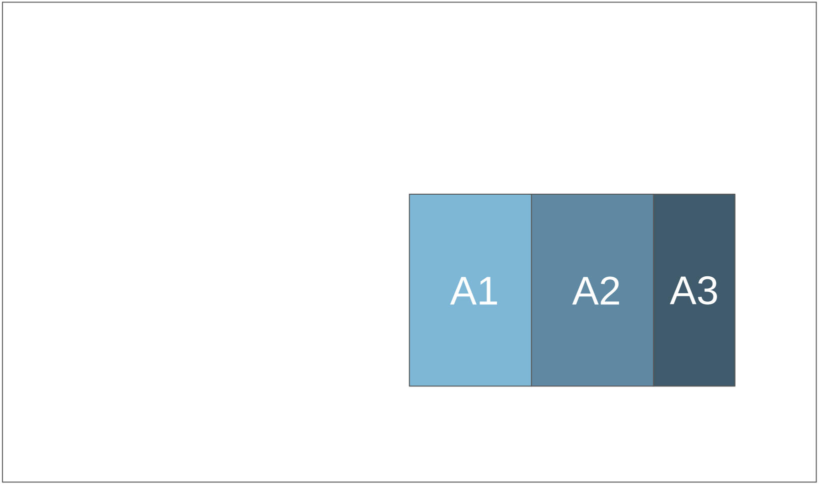


Combining events - partition



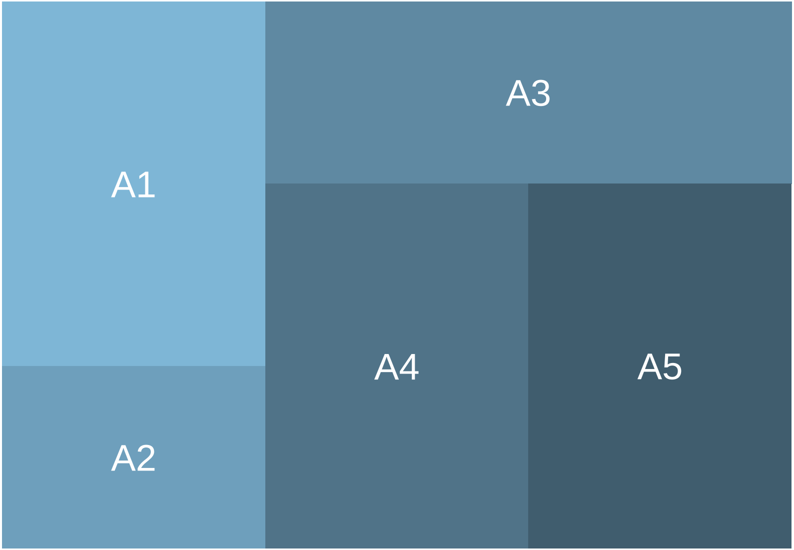
$$P(B) = P(A_1 \cup A_2 \cup A_3)$$

Combining events - partition



$$P(B) = P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3)$$

Combining events - partition



Combining events - partition

A_i where $i \in \{1, 2, 3, 4, 5\}$ is a partition of Ω

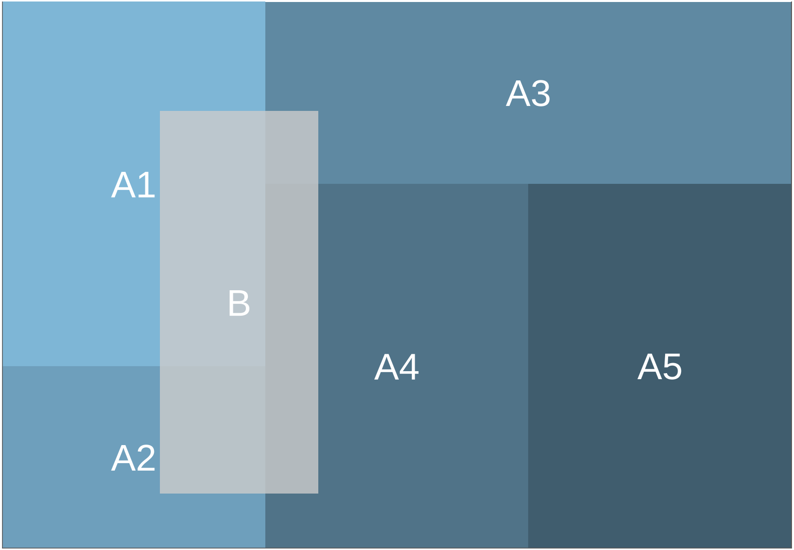
$$\sum_{i}^{5} P(A_i) = 1$$

Combining events - formula (law) of total probability

A_i a partition of Ω and B an event:

$$P(B) = \sum_{i}^{n} P(B \cap A_i)$$

Combining events - formula (law) of total probability



Combining events - formula (law) of total probability

A_i where $i \in \{1, 2, 3, 4, 5\}$ is a partition of Ω

$$P(B) = \sum_{i}^{5} P(B \cap A_i)$$

Occurrence of species 1 and species 2 on an island

- Events: $\{$"00", "01", "10", "11"$\}$

Occurrence of species 1 and species 2 on an island

- Events: {“00”, “01”, “10”, “11”}
- $P(''00'' \cap ''01'') = 0$

Occurrence of species 1 and species 2 on an island

- Events: $\{``00", ``01", ``10", ``11"\}$
- $P("00" \cap "01") = 0$
- $"00" \cup "01" \cup "10" \cup "11" = \Omega$

Occurrence of species 1 and species 2 on an island

- Events: $\{$"00", "01", "10", "11"$\}$
- $P(''00'' \cap ''01'') = 0$
- $''00'' \cup ''01'' \cup ''10'' \cup ''11'' = \Omega$
- $\{$"00", "01", "10", "11"$\}$ is a partition of Ω

Occurrence of species 1 and species 2 on an island

- Events: {“00”, “01”, “10”, “11”}
- P(”00” ∩ ”01”) = 0
- ”00” ∪ ”01” ∪ ”10” ∪ ”11” = Ω
- {“00”, “01”, “10”, “11”} is a partition of Ω
- $p_{00} + p_{01} + p_{10} + p_{11} = 1$

Occurrence of species 1 and species 2 on an island

- "00", "01", "10", "11" are **singleton sets** (a.k.a unit sets)

Occurrence of species 1 and species 2 on an island

- "00", "01", "10", "11" are **singleton sets** (a.k.a unit sets)
- $\{$"00", "01", "10", "11"$\}$ a partition of Ω made of singleton sets

Occurrence of species 1 and species 2 on an island

- "00", "01", "10", "11" are **singleton sets** (a.k.a unit sets)
- {"00", "01", "10", "11"} a partition of Ω made of singleton sets
- p_{00}, p_{01}, p_{10}, p_{11}

Occurrence of species 1 and species 2 on an island

- "00", "01", "10", "11" are **singleton sets** (a.k.a unit sets)
- {"00", "01", "10", "11"} a partition of Ω made of singleton sets
- p_{00}, p_{01}, p_{10}, p_{11}

Describes a **probability distribution**

Let’s practice 1 (15 min)

PRACTICE 1

- $P(A \cup B \cup C)$

Let's practice 1 (15 min)

PRACTICE 1

- $P(A \cup B \cup C)$
- the duck hunter 1 bullet

Let's practice 1 (15 min)

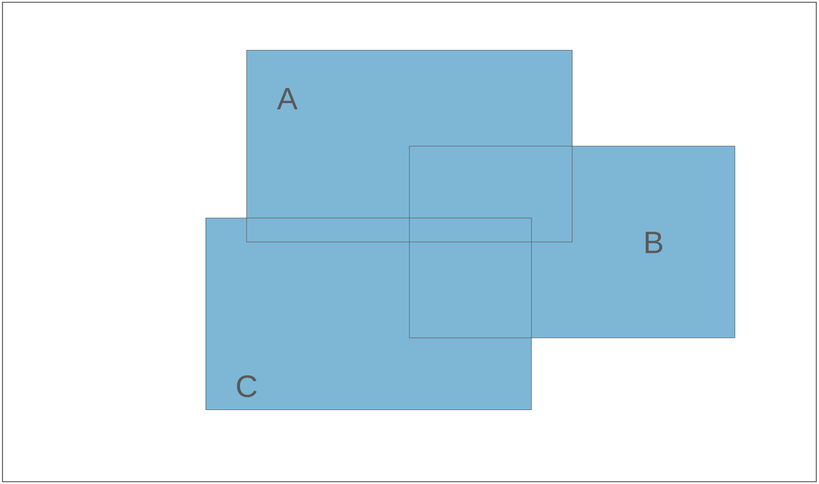
PRACTICE 1

- $P(A \cup B \cup C)$
- the duck hunter 1 bullet
- the duck hunter 2 bullets / 2 ducks - 1 duck

Let's practice 1 (15 min)

PRACTICE 1

- $P(A \cup B \cup C)$
- the duck hunter 1 bullet
- the duck hunter 2 bullets / 2 ducks - 1 duck
- bonus: how to simulate a dice with a coin?



$P(A \cup B \cup C) = ?$

Let's practice 1 - Elmer, the duck hunter

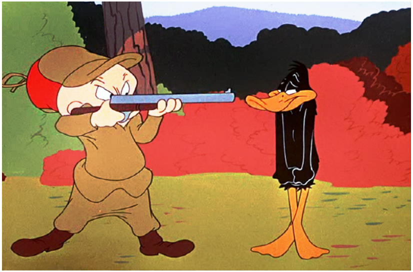
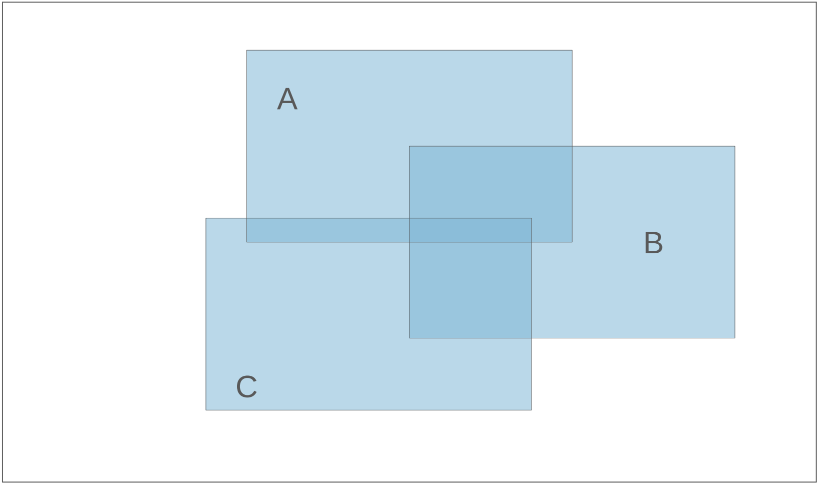


Figure 1: Daffy & Elmer

- Elmer, one bullet, one duck
- Elmer, two bullets, two ducks
- Elmer, two bullets, one duck

Solution 1 - P($A \bigcup B \bigcup C$)



Solution 1 - $P(A \bigcup B \bigcup C)$

- See the Inclusion–exclusion principle article on wikipedia (formule du crible de Poincaré).

Solution 1 - the duck paradigm

- Elmer, one bullet, one duck
- “sucess” (“1”)/ “failure” (“0”)
- P(”sucess”) = p ; P(”failure”) = 1 – p

- Flipping a coin / occurrence of 1 species on an island / shooting a duck

- Flipping a coin / occurrence of 1 species on an island / shooting a duck $\rightarrow$ we apply a similar probabilistic approach.

Random variables

- Flipping a coin / occurrence of 1 species on an island / shooting a duck $\rightarrow$ we apply a similar probabilistic approach.

- *Success*: "Head", "Presence", "one dead duck" $\rightarrow$ **1**

- Flipping a coin / occurrence of 1 species on an island / shooting a duck $\rightarrow$ we apply a similar probabilistic approach.
- *Success*: "Head", "Presence", "one dead duck" $\rightarrow \mathbf{1}$
- *Failure*: "Tails", "Absence", "no dinner tonight" $\rightarrow \mathbf{0}$

Random variables

- Flipping a coin / occurrence of 1 species on an island / shooting a duck $\rightarrow$ we apply a similar probabilistic approach.
- *Success*: “Head”, “Presence”, “one dead duck” $\rightarrow$ **1**
- *Failure*: “Tails”, “Absence”, “no dinner tonight” $\rightarrow$ **0**
- Now let X denote a variable such as:
  - X = 1 success: $P(X) = p$;
  - X = 0 failure: $P(\bar{X}) = 1 - p$

Random variables

- Flipping a coin / occurrence of 1 species on an island / shooting a duck $\rightarrow$ we apply a similar probabilistic approach.
- *Success*: "Head", "Presence", "one dead duck" $\rightarrow$ **1**
- *Failure*: "Tails", "Absence", "no dinner tonight" $\rightarrow$ **0**
- Now let X denote a variable such as:
  - X = 1 success: $P(X) = p$;
  - X = 0 failure: $P(\bar{X}) = 1 - p$
- Define a random variable + assign a probability distribution.

Random variables and probability distribution

- A **random variable** $X : \Omega \rightarrow M$ where M is a measurable space (natural numbers, real numbers, ...)

Random variables and probability distribution

- A **random variable** $X : \Omega \rightarrow M$ where M is a measurable space (natural numbers, real numbers, ...)

- A **probability distribution** f is a function that assigns probability under certain constraints:

Random variables and probability distribution

- A **random variable** $X : \Omega \rightarrow M$ where M is a measurable space (natural numbers, real numbers, ...)

- A **probability distribution** f is a function that assigns probability under certain constraints:

1. f specifies P for the partition of Ω made of singletons

Random variables and probability distribution

- A **random variable** $X : \Omega \rightarrow M$ where M is a measurable space (natural numbers, real numbers, ...)

- A **probability distribution** f is a function that assigns probability under certain constraints:

1. f specifies P for the partition of Ω made of singletons
2. f defines p_0, p_1, p_2, ..., p_n such as $\sum p_i = 1$

Random variables and probability distribution

- A **random variable** $X : \Omega \rightarrow M$ where M is a measurable space (natural numbers, real numbers, ...)

- A **probability distribution** f is a function that assigns probability under certain constraints:

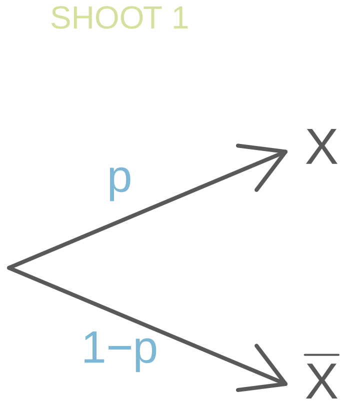
1. f specifies P for the partition of Ω made of singletons
2. f defines $p_0, p_1, p_2, \ldots, p_n$ such as $\sum p_i = 1$
   - Coin P("Head") = P(1) = p; P("Tail") = P(0) = 1 – p

Random variables and probability distribution

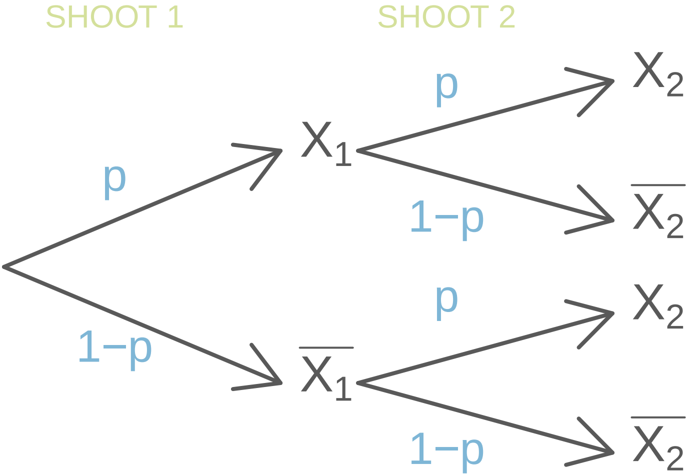
- A **random variable** $X : \Omega \rightarrow M$ where M is a measurable space (natural numbers, real numbers, ...)

- A **probability distribution** f is a function that assigns probability under certain constraints:

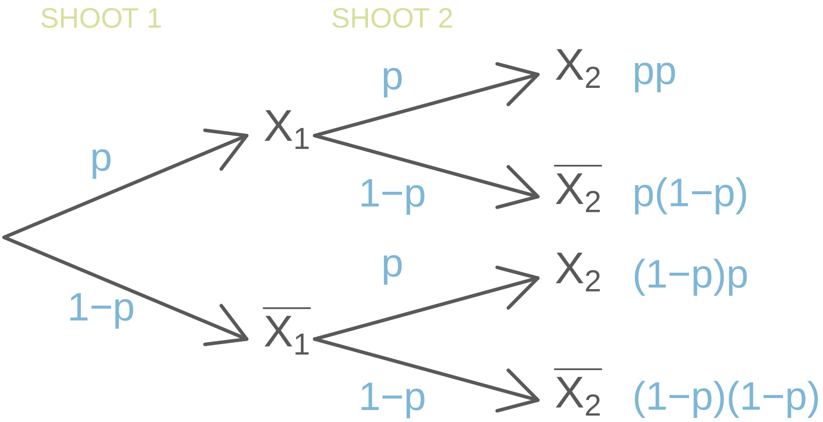
1. f specifies P for the partition of Ω made of singletons
2. f defines $p_0, p_1, p_2, \dots, p_n$ such as $\sum p_i = 1$
   - Coin P("Head") = P(1) = p; P("Tail") = P(0) = 1 – p
   - Dice P(1) = P(2) = ... = P(6) = 1/6


SHOOT 1
p
1–p
X
$\overline{X}$

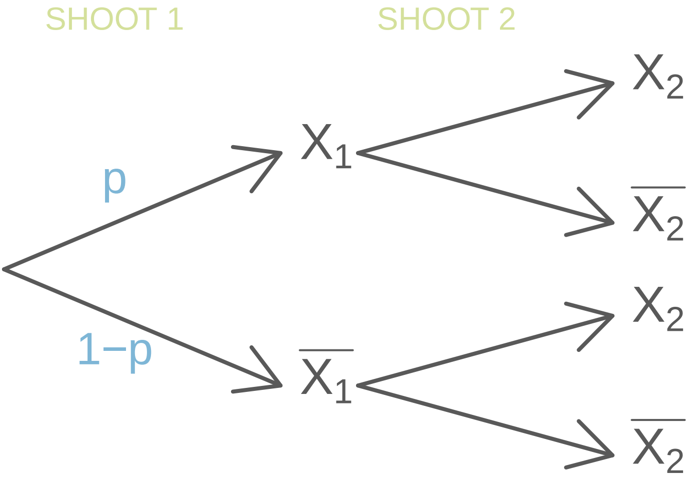
Independence - Intuition - 2 ducks



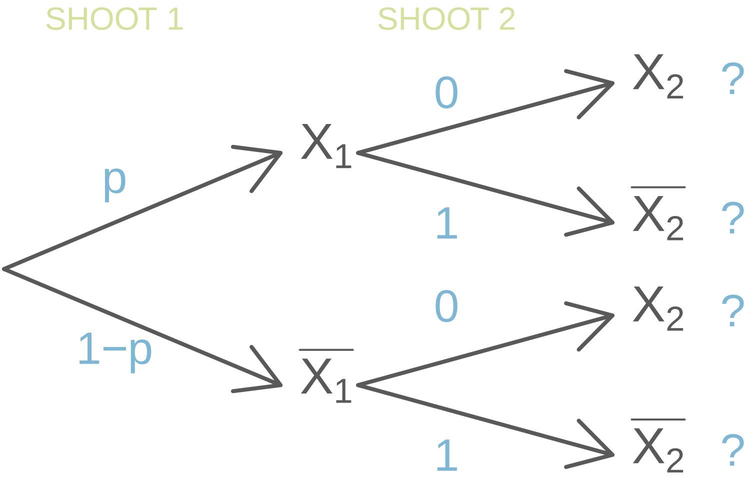
Independence - Intuition - 2 ducks



Independence - Intuition - 1 duck



Independence - Intuition - 1 duck



Independence - Definition

Two events are independent iif:

- $P(A \cap B) = P(A)P(B)$

Two events are independent iif:

- $P(A \cap B) = P(A)P(B)$

Remarks:

1. this is an assumption often implicit (notably in statistics)

Independence - Definition

Two events are independent iif:

- $P(A \cap B) = P(A)P(B)$

Remarks:

1. this is an assumption often implicit (notably in statistics)

2. events that may not seem independent (intuitively) may be independent according to the definition

Independence - Definition

Two events are independent iif:

- $P(A \cap B) = P(A)P(B)$

Remarks:

1. this is an assumption often implicit (notably in statistics)
2. events that may not seem independent (intuitively) may be independent according to the definition
3. A and B independent then $P(A \cup B) = P(A) + P(B) - P(A)P(B)$

Let's practice 2 (15 min)

Elmer shoots 3 independent ducks with a success rate of $p = 0.4$

Let's practice 2 (15 min)

Elmer shoots 3 independent ducks with a success rate of p = 0.4

1. Find the probability he misses the first 2 ducks and kills the last one.

Let's practice 2 (15 min)

Elmer shoots 3 independent ducks with a success rate of $p = 0.4$

1. Find the probability he misses the first 2 ducks and kills the last one.

2. Find the probability he kills 2 ducks.

Let's practice 2 (15 min)

Elmer shoots 3 independent ducks with a success rate of $p = 0.4$

1. Find the probability he misses the first 2 ducks and kills the last one.

2. Find the probability he kills 2 ducks.

He now shoots n independent ducks with a success rate of p

3. Find the probability he misses the $n - 1$ first ducks and kills the last one

Let's practice 2 (15 min)

Elmer shoots 3 independent ducks with a success rate of $p = 0.4$

1. Find the probability he misses the first 2 ducks and kills the last one.

2. Find the probability he kills 2 ducks.

He now shoots n independent ducks with a success rate of p

3. Find the probability he misses the $n - 1$ first ducks and kills the last one

4. Find the probability he kills k ducks.

Let's practice 2 (15 min)

Elmer shoots 3 independent ducks with a success rate of $p = 0.4$

1. Find the probability he misses the first 2 ducks and kills the last one.

2. Find the probability he kills 2 ducks.

He now shoots n independent ducks with a success rate of p

3. Find the probability he misses the $n - 1$ first ducks and kills the last one

4. Find the probability he kills k ducks.

- bonus: solve the **dice problem**

Solution 2

Y = "number of duck Elmer killed",

Solution 2

$Y =$ "number of duck Elmer killed", $Y \in 0, 1, 2, 3$

Solution 2

Y = "number of duck Elmer killed", $Y \in 0, 1, 2, 3$

Z = "number of failure before first success"

Solution 2

$Y =$ "number of duck Elmer killed", $Y \in 0, 1, 2, 3$

$Z =$ "number of failure before first success" $Z = 0, 1, ..., n$

Finite and countably infinite support sets

1. Finite set: $X = \{1, 2, \ldots, n\}$

Finite and countably infinite support sets

1. Finite set: $X = \{1, 2, \ldots, n\}$

- Rolling n dices
- presence of n species on an island
- killing k/n ducks

Finite and countably infinite support sets

1. Finite set: $X = \{1, 2, \dots, n\}$

- Rolling n dices
- presence of n species on an island
- killing k/n ducks

2. Countably infinite set $X = \{1, 2, 3, \dots, +\infty\}$

- number of species on a given island
- number of failure before the first success
- missing n ducks before killing one

Finite and countably infinite support sets

1. Finite set: $X = \{1, 2, \ldots, n\}$

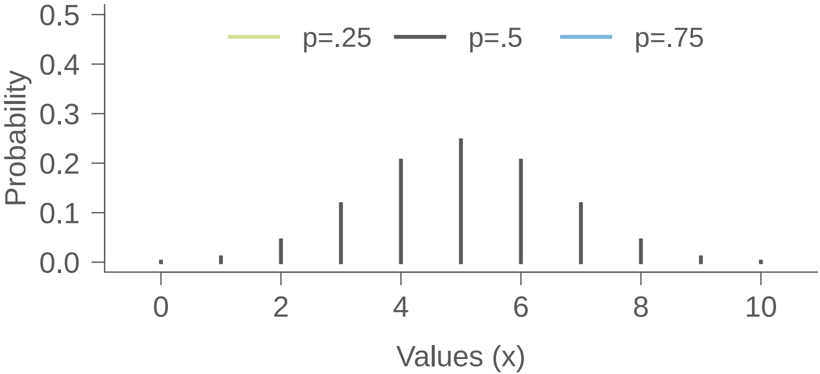
- Rolling n dices
- presence of n species on an island
- killing k/n ducks

2. Countably infinite set $X = \{1, 2, 3, \ldots, +\infty\}$

- number of species on a given island
- number of failure before the first success
- missing n ducks before killing one

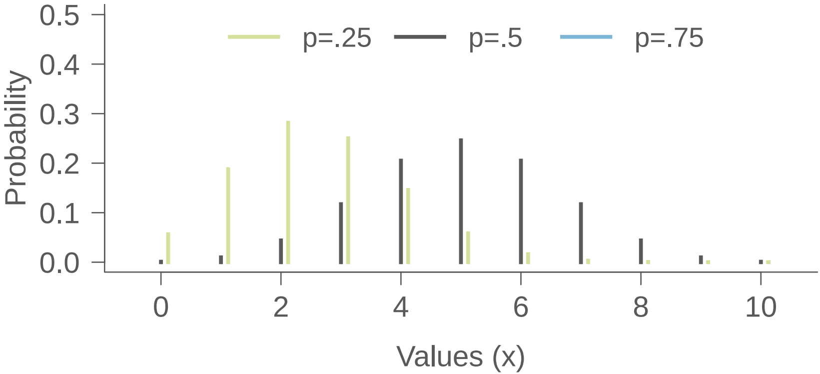
$$\sum_{i}^{+\infty} P(X_i) = 1$$

Binomial distribution *dbinom*



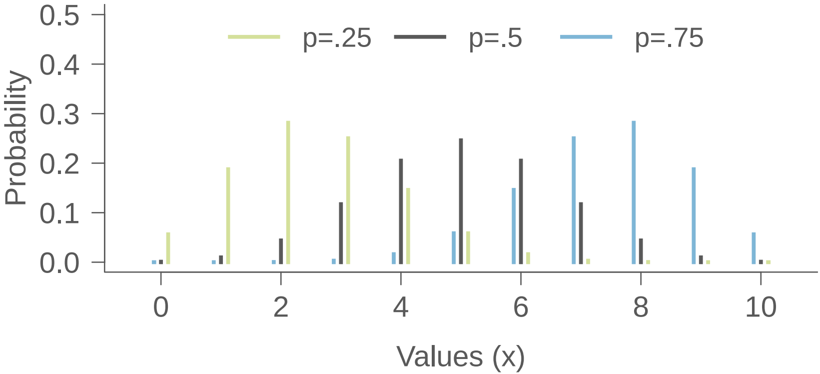
$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Binomial distribution *dbinom*



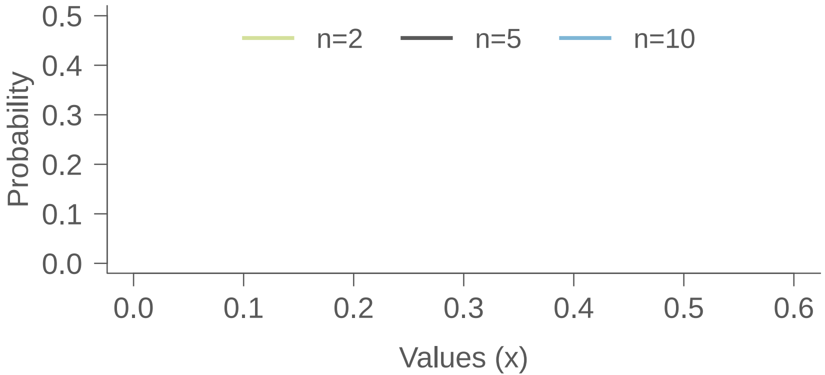
$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Binomial distribution *dbinom*



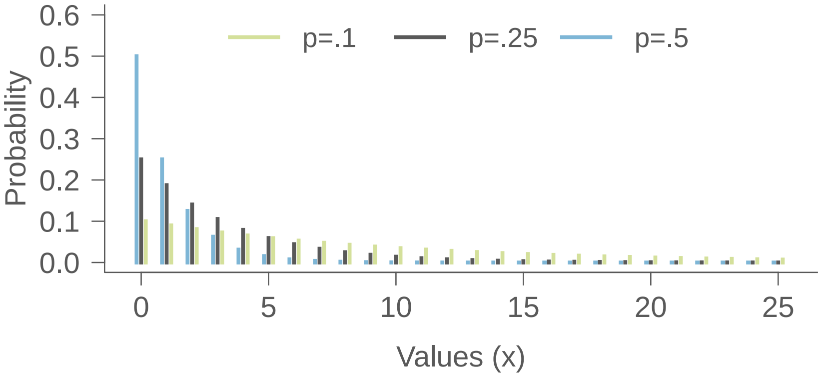
$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Uniform distribution



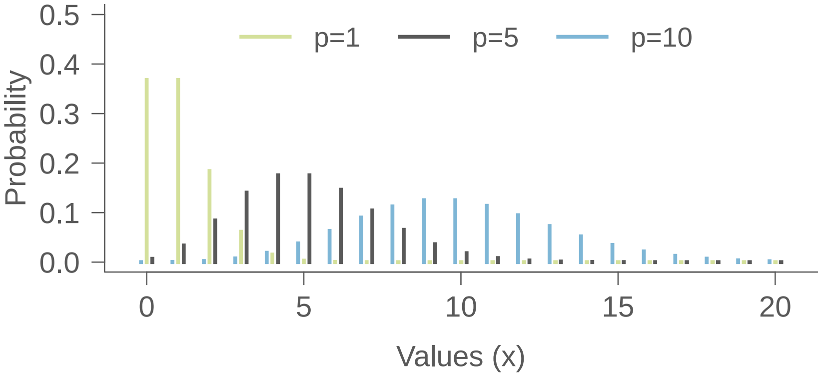
$$P(X = k) = \frac{1}{n}$$

Negative binomial distribution *dnbinom*



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

Poisson distribution *dpois*



$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

PAUSE PAUSE PAUSE PAUSE PAUSE PAUSE

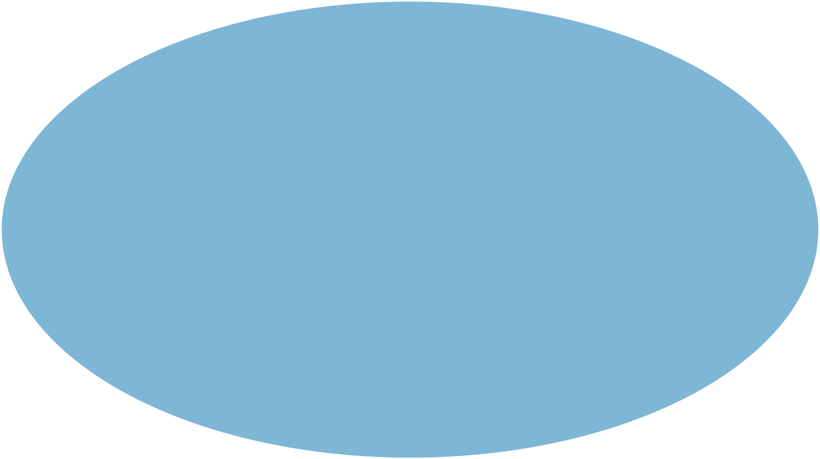
PART 2

Infinite sets

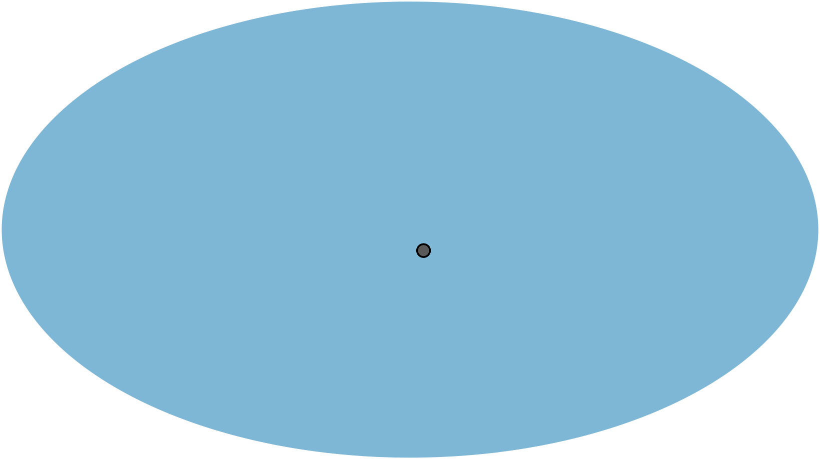
Moments

The Bayes theorem

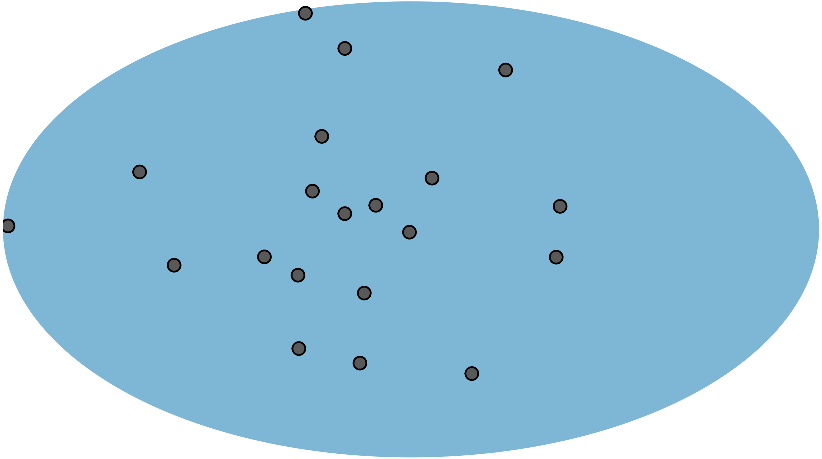
Infinite set - where is the duck?



Infinite set - where is the duck?



Infinite set - where is the duck?



Infinite set - where is the duck?

Let X be the random values x-coordinate

Infinite set - where is the duck?

Let X be the random values x-coordinate

- values: $x \in [0,10]$

Infinite set - where is the duck?

Let X be the random values x-coordinate

- values: $x \in [0,10]$
- $P(X = x) =?$

Infinite set - where is the duck?

Let X be the random values x-coordinate

- values: $x \in [0,10]$
- $P(X = x) =?$
- $P(X \in [a, b])$ makes sense!

Infinite set - where is the duck?

Let X be the random values x-coordinate

- values: $x \in [0,10]$
- $P(X = x) =?$
- $P(X \in [a, b])$ makes sense!
- $P(X \in [0, 10]) = 1$

Infinite set - where is the duck?

Let X be the random values x-coordinate

- values: $x \in [0{,}10]$
- $P(X = x) = ?$
- $P(X \in [a, b])$ makes sense!
- $P(X \in [0, 10]) = 1$
- $P(X \in [0, 5]) = 0.5$

Infinite set - where is the duck?

Let X be the random values x-coordinate

- values: $x \in [0,10]$
- $P(X = x) =?$
- $P(X \in [a, b])$ makes sense!
- $P(X \in [0, 10]) = 1$
- $P(X \in [0, 5]) = 0.5$

Infinite set - where is the duck?

- $P(X = x) = \frac{1}{\infty}$

Infinite set - where is the duck?

- $P(X = x) = \frac{1}{\infty} = 0$

Infinite set - where is the duck?

- $P(X = x) = \frac{1}{\infty} = 0$ but...

Infinite set - where is the duck?

- $P(X = x) = \frac{1}{\infty} = 0$ but...

We need something else!

Infinite set - probability density function (p.d.f)

f is a **p.d.f** iif:

1. defined on [a,b] (a may be $-\infty$ / b may be $+\infty$)

Infinite set - probability density function (p.d.f)

f is a **p.d.f** iif:

1. defined on [a,b] (a may be $-\infty$ / b may be $+\infty$)
2. positive

Infinite set - probability density function (p.d.f)

f is a **p.d.f** iif:

1. defined on [a,b] (a may be $-\infty$ / b may be $+\infty$)
2. positive
3. regular

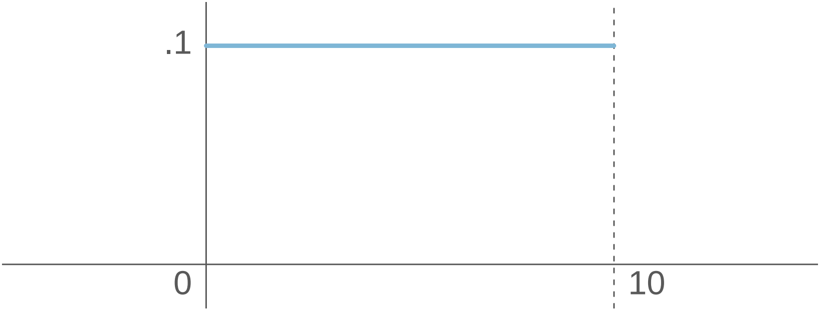
Infinite set - probability density function (p.d.f)

f is a **p.d.f** iif:

1. defined on [a,b] (a may be $-\infty$ / b may be $+\infty$)
2. positive
3. regular
4. and:

$$\int_a^b f(x)dx = 1$$

Infinite set - where's the duck?

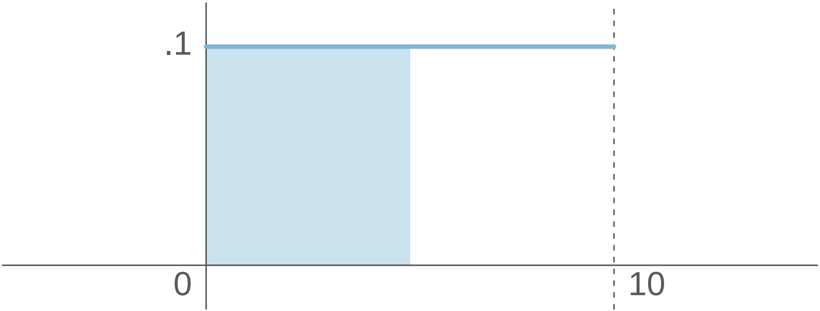


$$\forall\ x \in [0, 10] \quad f(x) = .1 \qquad (\mathcal{U}_{[0,10]})$$

Infinite set - where's the duck?

$$\int_0^{10} f(x)dx = 1$$

Infinite set - where's the duck?



$$\int_0^5 f(x)dx = .5$$

Probability distribution - act 2

Probability distribution function:

Probability distribution function:

- **probability mass function, p.m.f.**: random variables with a discrete support set (or countable infinite)

Probability distribution - act 2

Probability distribution function:

- **probability mass function, p.m.f.**: random variables with a discrete support set (or countable infinite)
- **probability density function, p.d.f.**: random variables with a infinite support set

Probability distribution - act 2

- f(x)   [x] (pmf or pdf)
- $\int$ f(x)dx $\int$[x]dx

Probability distribution - act 2

- $f(x) \quad [x]$ (pmf or pdf)
- $\int f(x)dx \quad \int [x]dx$

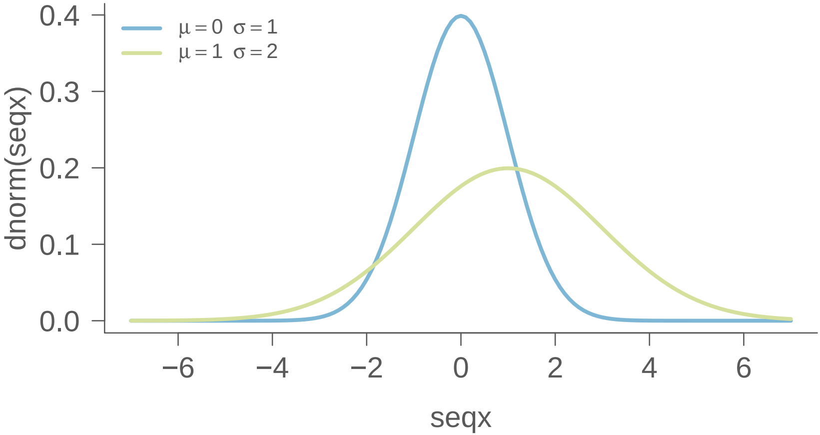
Conditional probability:

- $f(x|y) \quad [x|y]$
- $f(x) = f(x|y)f(y)$
- $f(x) = f(x|y)P(y)$
- $f(x_1)f(x_2)$

Cumulative distribution function (c.d.f.)

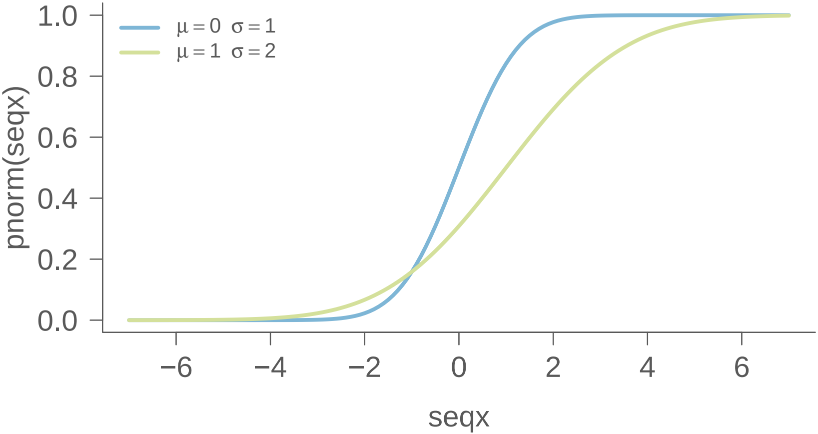
$$F(y) = P(X \leq y) = \int_{-\infty}^{y} f(x)dx$$

Normal distribution - p.d.f. *dnorm*

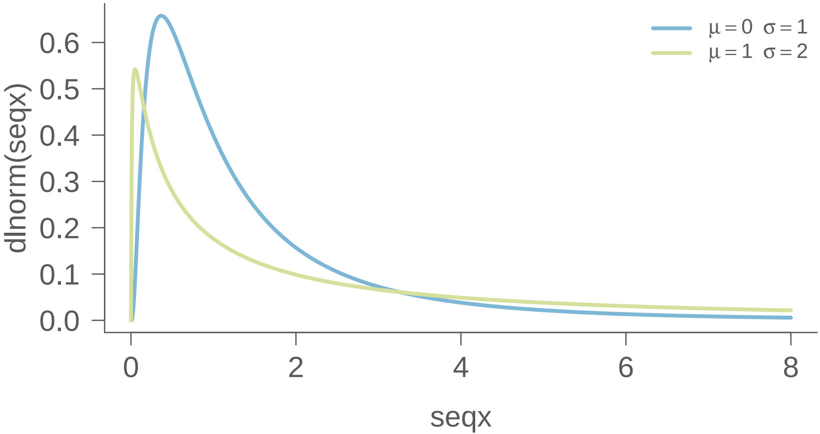


$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Lognormal distribution - c.d.f. *pnorm*

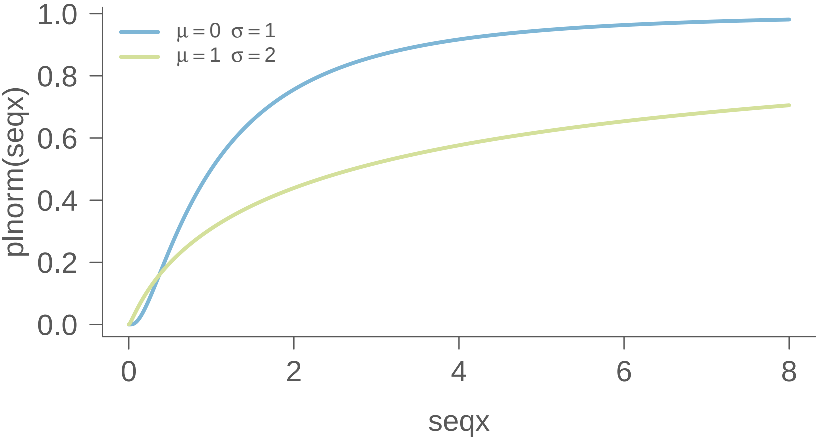


Lognormal distribution - p.d.f. *dlnorm*

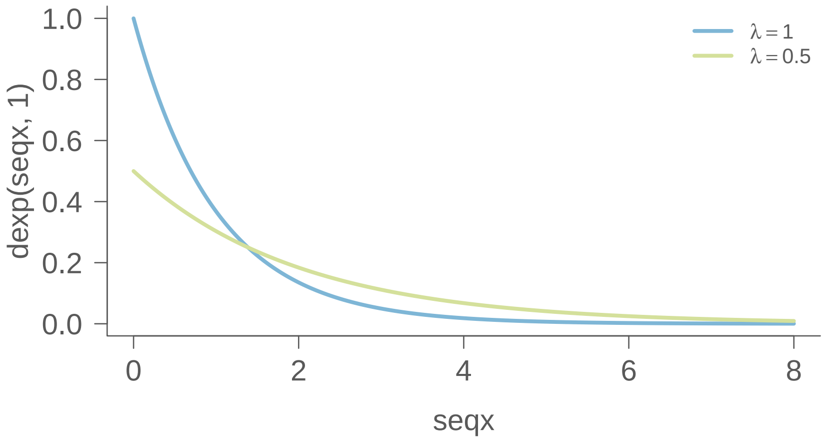


$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{\ln(x)-\mu}{\sigma}\right)^2}$$

Lognormal distribution - c.d.f. *plnorm*

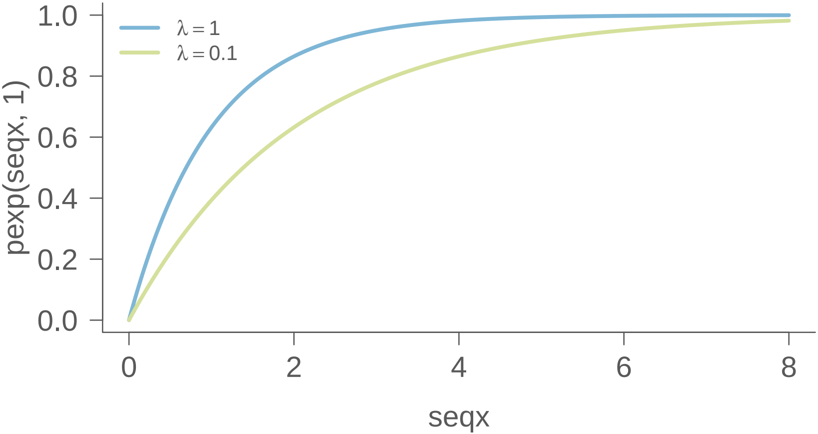


Exponential distribution - p.d.f. *dexp*



$$f(x) = \lambda e^{-\lambda x}$$

Exponential distribution - c.d.f. *pexp*



Practice 3 (10 min)

- You've set up a meeting with 2 colleagues:

Practice 3 (10 min)

- You've set up a meeting with 2 colleagues:

1. you = regular behavior
2. colleague 1 = often late but you never really know when he'll show up
3. colleague 2 = regular behavior if the starting time were 30 min earlier...

Practice 3 (10 min)

- You've set up a meeting with 2 colleagues:

1. you = regular behavior
2. colleague 1 = often late but you never really know when he'll show up
3. colleague 2 = regular behavior if the starting time were 30 min earlier...

- How to model tpon arrival?

Practice 3 (10 min)

- You've set up a meeting with 2 colleagues:

1. you = regular behavior
2. colleague 1 = often late but you never really know when he'll show up
3. colleague 2 = regular behavior if the starting time were 30 min earlier...

- How to model tpon arrival?

- Find the probability that you get started on time?

Practice 3 (10 min)

- You've set up a meeting with 2 colleagues:

1. you = regular behavior
2. colleague 1 = often late but you never really know when he'll show up
3. colleague 2 = regular behavior if the starting time were 30 min earlier...

- How to model tpon arrival?

- Find the probability that you get started on time?

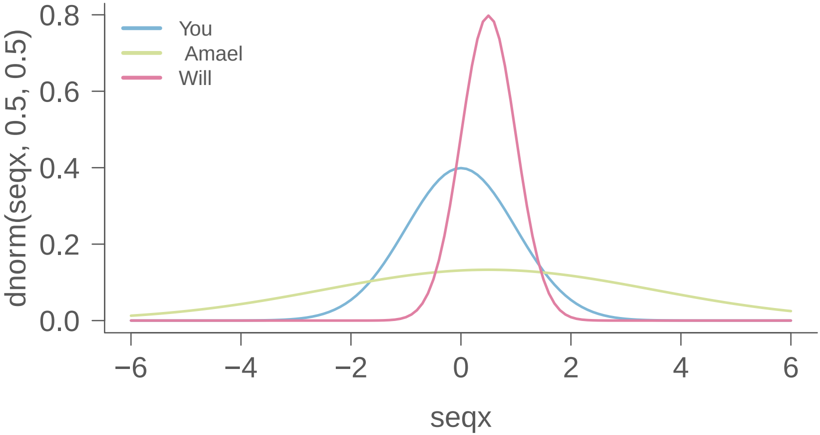
- Find the probability that the meeting is delayed by *at least* half an hour?

Solution 3

1- You: $\mathcal{N}(0, 1)$

2- Amael: $\mathcal{N}(.5, 3)$

3- Will: $\mathcal{N}(.5, .5)$



Solution 3

1. Starting on time
2. Meeting delayed by at least half an hour

Dealing with joint distributions

Dealing with joint distributions

1. $P(X \cap Y)$ or $P(X, Y)$

Dealing with joint distributions

1. $P(X \cap Y)$ or $P(X, Y)$ if independent $P(X)P(Y)$

Dealing with joint distributions

1. $P(X \cap Y)$ or $P(X, Y)$ if independent $P(X)P(Y)$
2. $P(X|Y)$ or $P(X|Y)$

Dealing with joint distributions

1. $P(X \cap Y)$ or $P(X, Y)$ if independent $P(X)P(Y)$
2. $P(X|Y)$ or $P(X|Y)$
3. $f(x, y)$

Dealing with joint distributions

1. $P(X \cap Y)$ or $P(X, Y)$ if independent $P(X)P(Y)$
2. $P(X|Y)$ or $P(X|Y)$
3. $f(x, y)$ if independent $f(y)f(x)$

Dealing with joint distributions

1. $P(X \cap Y)$ or $P(X, Y)$ if independent $P(X)P(Y)$
2. $P(X|Y)$ or $P(X|Y)$
3. $f(x, y)$ if independent $f(y)f(x)$
4. $f(x|y)$, $f(y|x)$

Expectation (*a.k.a* expected value, mean):

$$E(X) = \int xf(x)dx$$

Expectation and moments

Expectation (*a.k.a* expected value, mean):

$$E(X) = \int xf(x)dx$$

Variance:

$$V(X) = \int (x - E(x))^2 f(x)dx$$

Expectation and moments

Expectation (*a.k.a* expected value, mean):

$$E(X) = \int xf(x)dx$$

Variance:

$$V(X) = \int (x - E(x))^2 f(x)dx$$

n-th moment:

$$E(X^n) = \int x^n f(x)dx$$

Expectation and moments

Expectation (*a.k.a* expected value, mean):

$$E(X) = \int xf(x)dx$$

Variance:

$$V(X) = \int (x - E(x))^2 f(x)dx$$

n-th moment:

$$E(X^n) = \int x^n f(x)dx$$

Moment-generating function (MGF) alternative speciation of the distribution.

Quantile α:

$$x_\alpha \quad P(X \leq x_\alpha) = \alpha$$

Quantiles

Quantile α:

$$x_\alpha \quad P(X \leq x_\alpha) = \alpha$$

Examples:

- median ($\alpha = .5$)
- 1st and 3rd quartile ($\alpha = .25$ $\alpha = .75$)
- 5 / 95 percentile ($\alpha = .05$ $\alpha = .95$)

Quantile α:

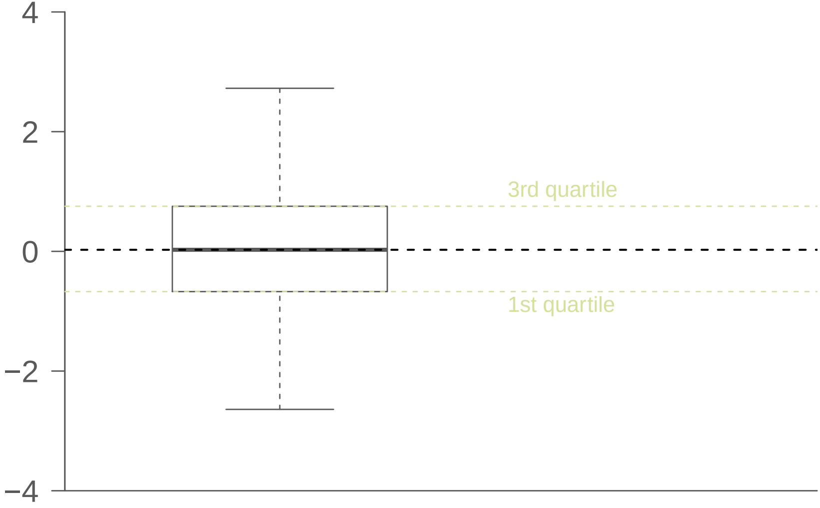
$$x_\alpha \quad P(X \leq x_\alpha) = \alpha$$

Examples:

- median ($\alpha = .5$)
- 1st and 3rd quartile ($\alpha = .25$ $\alpha = .75$)
- 5 / 95 percentile ($\alpha = .05$ $\alpha = .95$)

R: `qbinom`, `qpois`, `qnorm`, ...

Quantiles



More about expectation

$$E(g(X)) = \int g(x)f(x)dx$$

More about expectation

$$E(g(X)) = \int g(x)f(x)dx$$

$$Z = X^2 \quad E(Z) = \int x^2 f(x)dx$$

$$E(g(X)) = \int g(x)f(x)dx$$

$$Z = X^2 \quad E(Z) = \int x^2 f(x)dx$$

$$Z = \cos(X) \quad E(Z) = \int \cos(x)f(x)dx$$

Expectation / variance

- Binomial: $X : \mathcal{B}(n, p)$

- Binomial: $X : \mathcal{B}(n, p)$

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

$$E(X) = \sum_k^n kP(X = k) = \sum_k^n k \binom{n}{k} p^k (1 - p)^{n-k}$$

$$E(X) = np$$

- Binomial: $X : \mathcal{B}(n, p)$

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

$$E(X) = \sum_k^n kP(X = k) = \sum_k^n k \binom{n}{k} p^k (1 - p)^{n-k}$$

$$E(X) = np$$

$$V(X) = npq$$

Example Expectation / variance

- Poisson $\mathcal{P}(\lambda)$: $E(X) = \lambda$; $V(X) = \lambda$
- Binomial negative: $\mathcal{NB}(r, p)$: $E(X) = \frac{(1-p)r}{p}$; $V(X) = \frac{(1-p)r}{(p)^2}$
- Binomial negative: $\mathcal{NB}(1, p)$: $E(X) = \frac{(1-p)}{p}$; $V(X) = \frac{(1-p)}{(p)^2}$
- Normal $\mathcal{N}(\mu, \sigma)$: $E(X) = \mu$; $V(X) = \sigma^2$
- Exponential $\mathcal{E}(\lambda)$: $E(X) = \lambda$; $V(X) = \lambda^2$

Example Expectation / variance

Notation	$\mathcal{N}(\mu, \sigma^2)$
Parameters	$\mu \in \mathbb{R}$ — mean (location) $\sigma^2 > 0$ — variance (squared scale)
Support	$x \in \mathbb{R}$
PDF	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
CDF	$\frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$
Quantile	$\mu + \sigma\sqrt{2}\operatorname{erf}^{-1}(2F-1)$
Mean	μ
Median	μ
Mode	μ
Variance	σ^2
Skewness	0
Ex. kurtosis	0

Figure 2: Normal distribution's properties on Wikipedia

Let's practice 4 (15 min)

Elmer and the frightening question!

Let's practice 4 (15 min)

Elmer and the frightening question!

- Elmer's success rate is p

Let's practice 4 (15 min)

Elmer and the frightening question!

- Elmer's success rate is p
- a bullet is 3$

Let's practice 4 (15 min)

Elmer and the frightening question!

- Elmer's success rate is p
- a bullet is 3$
- a duck of the same quality is 60$

Let's practice 4 (15 min)

Elmer and the frightening question!

- Elmer's success rate is p
- a bullet is 3$
- a duck of the same quality is 60$
- "Should Elmer better stop hunting?"

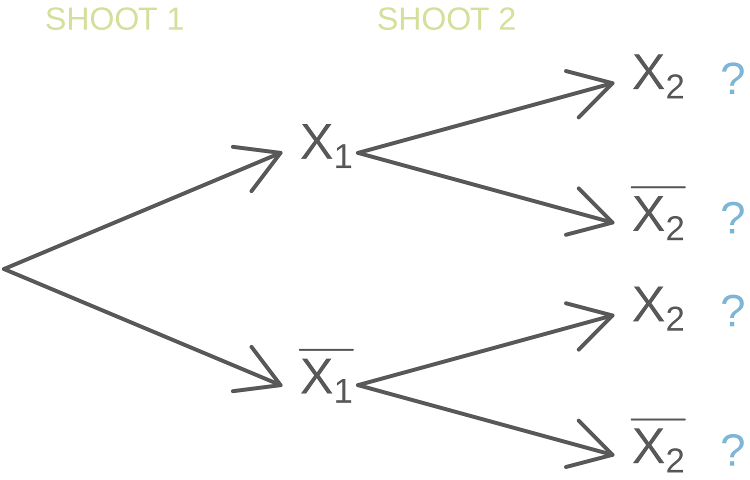
Let's practice 4 (15 min)

Elmer and the frightening question!

- Elmer's success rate is p
- a bullet is 3$
- a duck of the same quality is 60$
- "Should Elmer better stop hunting?"
- Find p_{sh} the success rate below which Elmer should better stay at home?

Solution 4

Independence act 2



Independence act 2

Let's A and B be two events, the conditional probability $P(A|B)$ is defined as:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Independence act 2

Let's A and B be two events, the conditional probability $P(A|B)$ is defined as:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(X_1 = 1|X_2 = 1) = \frac{P(X_1 = 1 \cap X_2 = 1)}{P(X_2 = 1)}$$

Independence act 2

Let's A and B be two events, the conditional probability $P(A|B)$ is defined as:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(X_1 = 1|X_2 = 1) = \frac{P(X_1 = 1 \cap X_2 = 1)}{P(X_2 = 1)}$$

consequently:

$$P(A \cap B) = P(A|B)P(B)$$

Independence act 2

Independence:

$$P(A|B) = P(A)$$

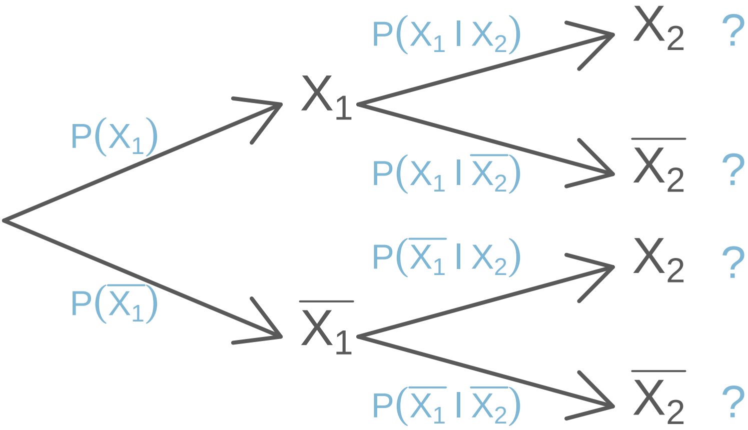
Independence act 2

Independence:

$$P(A|B) = P(A)$$

$$P(A \cap B) = P(A|B)P(B)$$

Independence act 2



Bayes theorem

Bayes theorem

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

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

$$P(A|B)P(B) = P(B|A)P(A)$$

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

$$P(A|B)P(B) = P(B|A)P(A)$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Bayes theorem

"Given the number of times in which an unknown event has happened and failed: Required the chance that the probability of its (specific event) happening in a single trial lies somewhere between any two degrees of probability that can be named."

"Given the number of times in which an unknown event has happened and failed: Required the chance that the probability of its (specific event) happening in a single trial lies somewhere between any two degrees of probability that can be named."

Proposition 3:

"The probability that two subsequent events will both happen is a ratio compounded of the probability of the 1st, and the probability of the 2nd on supposition the 1st happens."

"Given the number of times in which an unknown event has happened and failed: Required the chance that the probability of its (specific event) happening in a single trial lies somewhere between any two degrees of probability that can be named."

Proposition 3:

"The probability that two subsequent events will both happen is a ratio compounded of the probability of the 1st, and the probability of the 2nd on supposition the 1st happens."

An Essay Towards Solving a Problem in the Doctrine of Chances

Proposition 5:

> *"If there be two subsequent events, the probability of the 2nd b/N and the probability both together P/N, and it being first discovered that the 2nd event has happened, from hence I guess that the 1st event has also happened, the probability I am in the right is P/b"*

Proposition 5:

> *"If there be two subsequent events, the probability of the 2nd b/N and the probability both together P/N, and it being first discovered that the 2nd event has happened, from hence I guess that the 1st event has also happened, the probability I am in the right is P/b"*

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

- information

Proposition 5:

> *"If there be two subsequent events, the probability of the 2nd b/N and the probability both together P/N, and it being first discovered that the 2nd event has happened, from hence I guess that the 1st event has also happened, the probability I am in the right is P/b"*

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

- information
- inferences

Proposition 5:

> *"If there be two subsequent events, the probability of the 2nd b/N and the probability both together P/N, and it being first discovered that the 2nd event has happened, from hence I guess that the 1st event has also happened, the probability I am in the right is P/b"*

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

- information
- inferences
- cause/consequence

Bayes theorem

C_i $(i \in 1, ..., n)$ is a partition of Ω, let's use the law of total probability

Bayes theorem

C_i $(i \in 1, ..., n)$ is a partition of Ω, let's use the law of total probability

$$P(A|B) = \frac{P(B|A)P(A)}{\sum P(B \cap C_i)}$$

C_i $(i \in 1, ..., n)$ is a partition of Ω, let's use the law of total probability

$$P(A|B) = \frac{P(B|A)P(A)}{\sum P(B \cap C_i)}$$

$$P(A|B) = \frac{P(B|A)P(A)}{\sum P(B|C_i)P(C_i)}$$

Bayes theorem

C_i $(i \in 1, ..., n)$ is a partition of Ω, let's use the law of total probability

$$P(A|B) = \frac{P(B|A)P(A)}{\sum P(B \cap C_i)}$$

$$P(A|B) = \frac{P(B|A)P(A)}{\sum P(B|C_i)P(C_i)}$$

$$f(A|B) = \frac{f(B|A)f(A)}{\int f(b|c)f(c)dc}$$

Practice 5 - Are you infected? (20 min)

- Prevalence is π (0.01)

Practice 5 - Are you infected? (20 min)

- Prevalence is π (0.01)
- test to determine whether or not you are infected
  - error type I is $\alpha = .025$
  - error type II $\beta = .05$

Practice 5 - Are you infected? (20 min)

- Prevalence is π (0.01)
- test to determine whether or not you are infected
    - error type I is $\alpha = .025$
    - error type II $\beta = .05$

- You take the test, it is positive, are you infected?

Practice 5 - Are you infected? (20 min)

- Prevalence is π (0.01)
- test to determine whether or not you are infected
  - error type I is $\alpha = .025$
  - error type II $\beta = .05$

- You take the test, it is positive, are you infected?
- You take the test, it is negative, are you infected?

Practice 5 - Are you infected? (20 min)

- Prevalence is π (0.01)
- test to determine whether or not you are infected
  - error type I is $\alpha = .025$
  - error type II $\beta = .05$

- You take the test, it is positive, are you infected?
- You take the test, it is negative, are you infected?
- bonus: build a function to answer the questions above for any parameters' value.

Solution 5

Let's use 2 random variables:

- $X = 1$ ("sick"); $X = 0$ ("sane")
- $T = 1$ ("test positive"); $T = 0$ ("test negative")

LUNCH LUNCH LUNCH LUNCH LUNCH

PART 3

Let's practice more

Practice 6 - Elmer is back (25 min)

- Elmer's precision decreases as distance increases

Practice 6 - Elmer is back (25 min)

- Elmer's precision decreases as distance increases
- Ducks escape when Elmer gets too close

Practice 6 - Elmer is back (25 min)

- Elmer's precision decreases as distance increases
- Ducks escape when Elmer gets too close

1. Model $P(X = 1 | D = d)$

Practice 6 - Elmer is back (25 min)

- Elmer's precision decreases as distance increases
- Ducks escape when Elmer gets too close

1. Model $P(X = 1 | D = d)$
2. Find the effective rate of success p

Practice 6 - Elmer is back (25 min)

- Elmer's precision decreases as distance increases
- Ducks escape when Elmer gets too close

1. Model $P(X = 1 | D = d)$
2. Find the effective rate of success p (how to model $P(D = d)$)

Practice 6 - Elmer is back (25 min)

- Elmer's precision decreases as distance increases
- Ducks escape when Elmer gets too close

1. Model $P(X = 1 | D = d)$
2. Find the effective rate of success p (how to model $P(D = d)$)
3. Elmer brings 10 bullets, what's the probability he'll have a nice diner?

Practice 6 - Elmer is back (25 min)

- Elmer's precision decreases as distance increases
- Ducks escape when Elmer gets too close

1. Model $P(X = 1|D = d)$
2. Find the effective rate of success p (how to model $P(D = d)$)
3. Elmer brings 10 bullets, what's the probability he'll have a nice diner?

- bonus: solve the **division problem**

Solution 6

Practice 7 - Elmer... the truth (25 min)

`val1.csv` (or `val1.Rds`) are the results of 1000 shoots Elmer took.

1. Create a function to compute the probability obtaining such results for any `p`.

Practice 7 - Elmer... the truth (25 min)

`val1.csv` (or `val1.Rds`) are the results of 1000 shoots Elmer took.

1. Create a function to compute the probability obtaining such results for any `p`.
2. Deduce the probability of having these results.

Practice 7 - Elmer... the truth (25 min)

`val1.csv` (or `val1.Rds`) are the results of 1000 shoots Elmer took.

1. Create a function to compute the probability obtaining such results for any p.
2. Deduce the probability of having these results.
3. Create a function that computes P(p|X) for any value of p.

Practice 7 - Elmer... the truth (25 min)

`val1.csv` (or `val1.Rds`) are the results of 1000 shoots Elmer took.

1. Create a function to compute the probability obtaining such results for any `p`.
2. Deduce the probability of having these results.
3. Create a function that computes P(p|X) for any value of `p`.
4. We have a new set of data `val2.csv` or `val2.Rds`, what should you do?

Practice 7 - Elmer... the truth (25 min)

`val1.csv` (or `val1.Rds`) are the results of 1000 shoots Elmer took.

1. Create a function to compute the probability obtaining such results for any p.
2. Deduce the probability of having these results.
3. Create a function that computes $P(p|X)$ for any value of p.
4. We have a new set of data `val2.csv` or `val2.Rds`, what should you do?
5. bonus: 1-3 including the distance (see `val3.csv` or `val3.Rds`)
6. bonus 2: Answer Bayes' original question

Solution 7

Let's step back

What do we do when we do statistics? (simple case)

Let's step back

What do we do when we do statistics? (simple case)

- **observations**: $x_1, x_2, \dots, x_n$

Let's step back

What do we do when we do statistics? (simple case)

- **observations**: $x_1, x_2, \ldots, x_n$
- Hypothesis: outcomes of **random variables independent and identically distributed (i.i.d.)** X_i

Let's step back

What do we do when we do statistics? (simple case)

- **observations**: $x_1, x_2, \dots, x_n$

- Hypothesis: outcomes of **random variables independent and identically distributed (i.i.d.)** X_i

- The **distribution** is given by θ (*i.e.* $\mathcal{N}(\theta)$ where $\theta = (\mu, \sigma)$)

Let's step back

What do we do when we do statistics? (simple case)

- **observations**: $x_1, x_2, \dots, x_n$
- Hypothesis: outcomes of **random variables independent and identically distributed (i.i.d.)** X_i
- The **distribution** is given by θ (*i.e.* $\mathcal{N}(\theta)$ where $\theta = (\mu, \sigma)$)
- We try to find out θ 's value(s) given x_i : **inference**

Let's take a step back

- To do so, we build **estimators**

Let's take a step back

- To do so, we build **estimators**
- Normal: $\theta = (\mu, \sigma)$
  - $\hat{\mu} = \frac{1}{n} \sum_i^n x_i$

Let's take a step back

- To do so, we build **estimators**

- Normal: $\theta = (\mu, \sigma)$
  - $\hat{\mu} = \frac{1}{n}\sum_i^n x_i$
  - $\hat{\sigma} = \frac{1}{n}\sum_i^n (x_i - \mu)^2$

Let's take a step back

- To do so, we build **estimators**

- Normal: $\theta = (\mu, \sigma)$
  - $\hat{\mu} = \frac{1}{n}\sum_i^n x_i$
  - $\hat{\sigma} = \frac{1}{n}\sum_i^n (x_i - \mu)^2$
  - $\hat{\sigma} = \frac{1}{n-1}\sum_i^n (x_i - \hat{\mu})^2$

Let's take a step back

- To do so, we build **estimators**
- Normal: $\theta = (\mu, \sigma)$
  - $\hat{\mu} = \frac{1}{n}\sum_i^n x_i$
  - $\hat{\sigma} = \frac{1}{n}\sum_i^n (x_i - \mu)^2$
  - $\hat{\sigma} = \frac{1}{n-1}\sum_i^n (x_i - \hat{\mu})^2$
- Then we assess the goodness of our estimation : IC / tests

Let's take a step back

- To do so, we build **estimators**
- Normal: $\theta = (\mu, \sigma)$
  - $\hat{\mu} = \frac{1}{n}\sum_i^n x_i$
  - $\hat{\sigma} = \frac{1}{n}\sum_i^n (x_i - \mu)^2$
  - $\hat{\sigma} = \frac{1}{n-1}\sum_i^n (x_i - \hat{\mu})^2$
- Then we assess the goodness of our estimation : IC / tests
- Bayesian framework offers few other possibilities.

Why normal, why?

Central limit theorem:

$$X_i, i \in 1, 2, ..., n$$

i.i.d. $\mathcal{L}(\theta)$,

$$\frac{X_i - \mu}{\sigma} \rightarrow N(\mu, \sigma)$$

Why normal, why?

Central limit theorem:

$$X_i, i \in 1, 2, ..., n$$

i.i.d. $\mathcal{L}(\theta)$,

$$\frac{X_i - \mu}{\sigma} \rightarrow N(\mu, \sigma)$$

Poincaré:

> *"Tout le mode croit à la loi normale : les physiciens parcequ'ils pensent que les mathématiciens l'ont démontrée et les mathématiciens parcequ'ils croient qu'elle a été vérifiée par les physiciens."*

To be continued