Statistical Literacy — MINT

Lecture 4: Probability and discrete random variables

Rémi Viné

The Graduate Institute | Geneva

October 16th, 2023

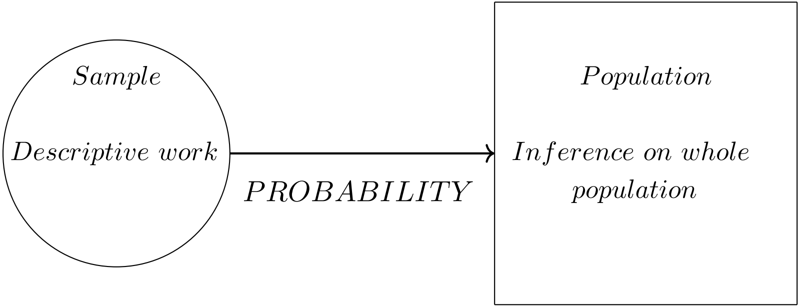
Outline

Housekeeping

- Problem set 4 is now available
- The first assignment on the survey is due next week (by Monday 23rd, 2.00pm)
  - You are expected to describe four variables, each of different type
  - 200 words per variable
  - Include one visualization per variable
  - Be to the point, contextualize
- Send the assignment to your TA

What is a probability about?

- Assign a measurable *likelihood* of occurrence to *events*
  - What is *likely* is not **certain**: probability is the science of the uncertain
  - Probability is about giving a measure to something we do not measure
    - It gives: $\frac{number\ of\ occurrences\ of\ event\ of\ the\ interest}{total\ number\ of\ possible\ outcomes}$
- This is a conceptual tool to bridge practical approaches



Different conceptualizations around probability

1. *Subjective probability*
    - Make guesses
    - *E.g., "Throwing a fair die, I suppose that the probability of obtaining an odd number (the event of interest) is 4/10"*

Different conceptualizations around probability

1. *Subjective probability*
2. *Empirical probability*
    - Conduct an experiment and assume it approaches the *real* probability
    - *E.g., "Throwing this die 100 times, I obtained 45 odd numbers and therefore conclude that the probability of obtaining an odd number in a die throw is 45%"*

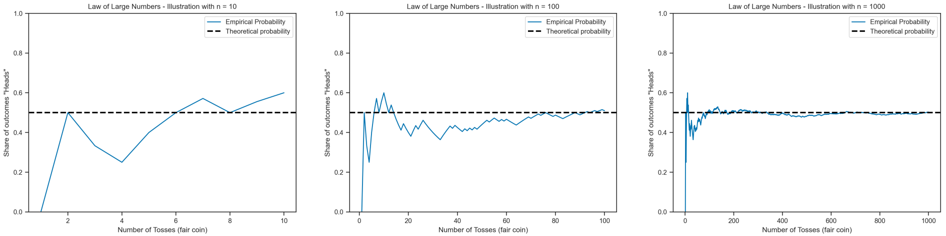
Different conceptualizations around probability

1. *Subjective probability*
2. *Empirical probability*
3. *Theoretical (deductive) probability*
    - Use some (mathematical) properties to construct a measurement of the uncertain event's occurrence
    - *E.g., "Since each outcome (die result per throw) is* **equally likely**, *and there are 6 outcomes, each is granted a probability of 1/6. Then, there are as many odd numbers as even numbers. Hence,* $P(\text{"}odd\ numbers\ in\ die\ throw\text{"}) = 3 \times \frac{1}{6} = 0.5$ *"*

From empirical probability to theoretical probability

The Law of Large Numbers

- By repeating an experiment many times, the observed probability converges (in probability) to the *true* probability
  - Formally (no need to remember that by heart): $\bar{X}_n \xrightarrow{p} \mu$ as n increases *[this can give a hint that probability is the bridge between descriptive and inferential statistics]*
- Imagine a game where a (fair coin) is tossed many times:



Different conceptualization around probability

1. *Subjective probability*
2. *Empirical probability*
3. *Theoretical (deductive) probability*

- Note that the probability notation can be fractions, percentages, or decimals
- Note that it is written $P(Event) = \cdot$

The foundations of probability

Kolmogorov axioms (1933)

1. Non-negativity

$$P(A) \geq 0, \text{for all } A \; in \; the \; set \; of \; all \; sets \; of \; outcomes$$

2. Normalization

$$P(Universe) = 1$$

The sum of probabilities of all the possible outcomes is equal to 1

3. Additivity of disjoint (mutually exclusive) events

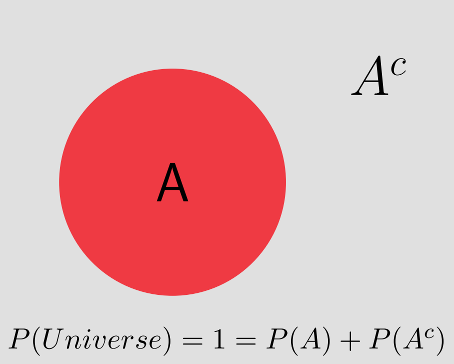
$$P(A \; or \; B) = P(A \cup B) = P(A) + P(B)$$

$$\text{for all } A, B \; in \; the \; set \; of \; all \; sets \; of \; outcomes$$

$$\text{such that } \; P(A \; and \; B) = P(A \cap B) = \emptyset$$

Probability and complement

- $P(A)$ is the probability of event A
- $P(A^c)$ is the probability of the **complement** of event A
  - *E.g.*, $P(\text{"}Dark\ hair\text{"})$ and the complement $P(\text{"}Not\ having\ dark\ hair\text{"})$

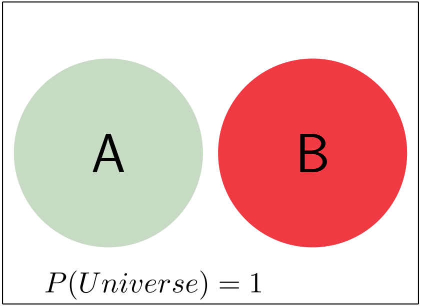


Venn diagram - complement

Add events in probability

Mutually exclusive case

- $P(A)$ and $P(B)$ do not overlap
  - $P(A \cup B) = P(A) + P(B)$
- *E.g.*, throwing a die
  - $P(A) = P(\text{“}Obtain\ a\ 6\text{”}) = 1/6$
  - $P(B) = P(\text{“}Obtain\ a\ number\ strictly\ lower\ than\ 4\text{”}) = 1/2$
  - $P(A \cup B) = 1/6 + 1/2 = 4/6$



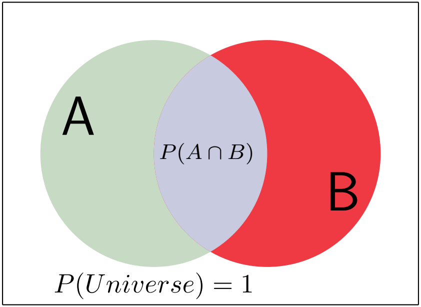
Add events in probability

Non-exclusive events

- $P(A)$ and $P(B)$ overlap (*e.g.*, *"Landlock"* & *"Asian country"*)

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

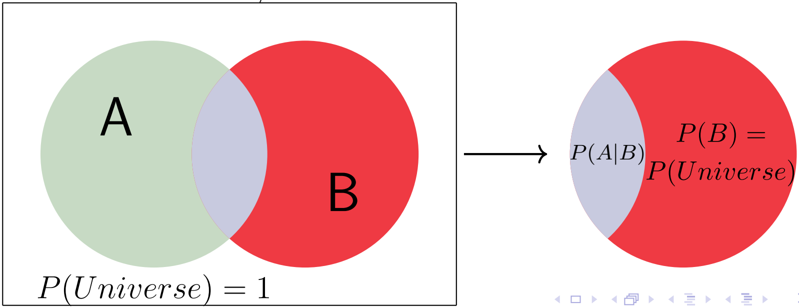
- *E.g.*, throwing a die
  - $P(A) = P(\text{"}Obtain\ an\ even\ number\text{"})$
  - $P(B) = P(\text{"}Obtain\ a\ number\ strictly\ lower\ than\ 4\text{"})$
  - $P(A \cup B) = 1/2 + 1/2 - 1/6 = 5/6$



Probability with conditions

A first step in Bayesian statistics

- What if you assume that event B is given
  - It writes: $P(A|B) = \frac{P(A \cap B)}{P(B)}$
- *E.g.*, we want "$Obtain\ an\ even\ number$" and we assume that "$Obtain\ a\ number\ strictly\ lower\ than\ 4$" is observed
  - Hence, we have either $1\ or\ 2\ or\ 3$ as a die throw outcome
  - $P(A|B) = \frac{1/6}{1/2} = 1/3$



Independence in probability

- Two events are **independent** if the occurrence of one does not affect the probability of the occurrence of the other
- It writes: $P(A \textbf{ given } B) = P(A|B) = P(A)$ if A, B are two independent events
- It also writes $P(A \cap B) = P(A) \times P(B)$ *[Do you see why from the formula above?]*
- In the last example:
  - $P(A) = P(\text{“}Obtain\ an\ even\ number\text{”}) = 1/2$
  - $P(A|B) = P(\text{“}Obtain\ an\ even\ number\ \mathbf{given}\ that\ the\ number\ obtained < 4\text{”}) = 1/3$
  - $P(A) \neq P(A|B)$ so the events are not independent

    $\rightarrow$ Here, one can help *predict* the other

A convenient tool: the contingency table

- A contingency table relates 2 variables (nominal, ordinal, grouped in interval,*etc.*)
- Here, whether the country is landlocked or not, and the form of government

Landlock / Form of government	Autocracy (A)	Democracy (D)	Theocracy (T)
Landlock (L)	17	24	11
Sea access (S)	6	23	19

A convenient tool: the contingency table

Landlock / Form of government	Autocracy (A)	Democracy (D)	Theocracy (T)
Landlock (L)	17	24	11
Sea access (S)	6	23	19

- How many countries were part of the study?
- $P(\text{“}S\text{”})$
- $P(\text{“}T\text{”})$
- $P(\text{“}D\text{”} \cap \text{“}L\text{”})$
- $P(\text{“}A\text{”} \cup \text{“}L\text{”})$
- $P(\text{“}T\text{”} \mid \text{“}S\text{”})$
- Are “T” and “L” independent?

A convenient tool: the contingency table

Landlock / Form of government	Autocracy (A)	Democracy (D)	Theocracy (T)	*Total*
Landlock (L)	17	24	11	*52*
Sea access (S)	6	23	19	*48*
Total	*23*	*47*	*30*	*100*

- How many countries were included in the study? $\rightarrow 100$
- $P(\text{“}S\text{”}) = \frac{48}{100} = 48\%$
- $P(\text{“}T\text{”}) = \frac{30}{100} = 30\%$
- $P(\text{“}D\text{”} \cap \text{“}L\text{”}) = \frac{24}{100} = 24\%$
- $P(\text{“}A\text{”} \cup \text{“}L\text{”}) = \frac{23}{100} + \frac{52}{100} - \frac{17}{100} = \frac{58}{100} = 58\%$
- $P(\text{“}T\text{”} | \text{“}S\text{”}) = \frac{19}{48} \approx 39.58\%$
- Are “T” and “L” independent? No because $P(\text{“}T\text{”}) = 30\% \neq 21.15\% \approx \frac{11}{52} = P(\text{“}T\text{”} | \text{“}L\text{”})$

Random variables (usually denoted X)

- A random variable is a collection of numerical values
  - This is common with all usual functions
- A random variable is derived from random experiments
  - This is the specificity of the random variable as compared to other functions

Discrete	Rolling a die
	Number of casualties in a bombing
	Passengers in a flight
	Number of votes in favor of a resolution
Continuous	Noise in a flight
	Waiting time for the bus
	Radioactivity levels next to a uranium mine
	Exchange rate CHF-€

Random variables: the discrete case

- A discrete variable is **countable**
- Each outcome has a defined probability associated
  - Formally: $0 \leq P(X = x) \leq 1$
    - It gives the probability that the random variable X has outcome of value x
    - Remember first Kolmogorov's axiom? Non-negative probability
  - Another property: $\sum_i P(X = x_i) = 1$
    - Remember third Kolmogorov's axiom? $P(universe) = 1$
- Many possible calculations: $P(X = x_i)$, $P(X = x_i, x_j)$, $P(X < x_i)$, $P(X > x_i)$, *etc.*

Discrete random variables: a simple example

Throw one die

- There are 6 outcomes
- The outcomes are equally likely

x	1	2	3	4	5	6
$P(X = x)$	1/6	1/6	1/6	1/6	1/6	1/6

Probability distribution of a random variable X

- Note that this probability distribution is identical to the **relative frequency** of each outcome
- $P(X < 4)$ is easy to compute:

$$P(X < 4) = P(X = 1) + P(X = 2) + P(X = 3) = 3/6 = 0.5$$

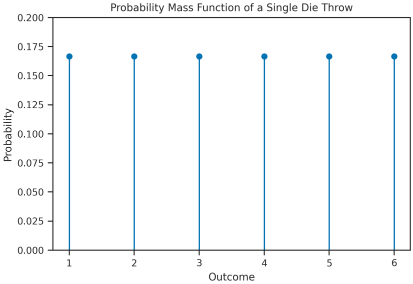
Random variables: a simple example and a visualization

Throwing one die is a **Uniform distribution**: $X \sim U(1;6)$

- There are 6 outcomes
- The outcomes are equally likely

x	1	2	3	4	5	6
$P(X=x)$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$

Probability distribution of random variable X



Random variables: another example

Throw two dice

- Now, let us focus on the **sum** of two dice throws
  - $s = die_1 + die_2$
- There are 11 possible outcomes
- The outcomes are **not** equally likely

s	2	3	4	5	6	7	8	9	10	11	12
$P(S = s)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

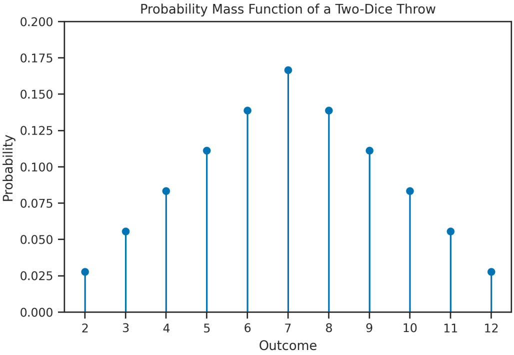
Probability distribution of random variable S

Random variables: another example

Throw two dice

s	2	3	4	5	6	7	8	9	10	11	12
$P(S=s)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Probability distribution of random variable S



Expected value (discrete case)

- The **expected value** of X, $E(X)$ is the sum of products of each possible outcome with its corresponding probability.

Expected value

$$E(X) = \sum_{i=1} x_i P(X = x_i) = x_1 P(X = x_1) + ... + x_k P(X = x_k)$$

- Remember the weighted average's formula?
  - $\bar{x} = \sum_i x_i w_i$
  - "Weights" are equivalent to relative frequencies

Expected value and average, what is the difference?

- The expected value is about an **uncertain** measure
  - It is based on probability
- The average **describes** data used

Variance (discrete case)

- **Variance** is a measure of the variability around the expected value. It is denoted as $Var(X)$ computed as follows:

Variance

$$Var(X) = \sum_i (x_i - E(X))^2 P(X = x_i) =$$

$$(x_1 - E(X))^2 \times P(X = x_1) + ... + (x_k - E(X))^2 \times P(X = x_k)$$

- Another useful formula: $Var(X) = E(X^2) - E(X)^2$
  - Do you see why (*difficult*)?
- Standard deviation: $\sigma = \sqrt{Var(X)}$
  - Remember the units of measurement

Compute $E(\cdot)$ and $Var(\cdot)$ (and the standard deviation) of the sum of two dice random variable

s	2	3	4	5	6	7	8	9	10	11	12
$P(S=s)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Probability distribution of random variable S

- $E(S)$
- $Var(S)$
- σ_S

Compute $E(\cdot)$ and $Var(\cdot)$ (and the standard deviation) of the sum of two dice random variable

s	2	3	4	5	6	7	8	9	10	11	12
$P(S=s)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Probability distribution of random variable S

- $E(S) = 2 * \frac{1}{36} + 3 * \frac{2}{36} + ... = 7$
- $Var(S) = 2^2 * \frac{1}{36} + 3^2 * \frac{2}{36} + ... + 12^2 * \frac{2}{36} - 7^2 \approx 5.83$
- $\sigma_S = \sqrt{Var(x)} \approx 2.42$

  $\rightarrow$ Do not forget the units of measurement

Next session

- Next session is on continuous random variable, and in particular on the Normal distribution