Probability and Statistics

Probability theory: introduction + discrete random variables

Joris Vankerschaver

Overview

1. **Introduction: What do we mean by probability?**
    1. Why study probability theory?
    2. Sample space, events, probability
2. Rules for probabilities
    1. Calculating with probabilities
    2. Conditional probabilities
3. Applications of Bayes' rule
4. Discrete random variables and probability distributions
    1. Bernoulli distribution
    2. Binomial distribution
    3. Poisson distribution
5. Solution of Monty Hall problem (optional)

Section 1

Introduction: What do we mean by probability?

What are random phenomena?

We study phenomena where

- indidual outcomes are uncertain, but ...
- taken as a whole, the outcome has a regular distribution.

Examples

- Tossing a coin, predicting if a radio-active nucleus will decay
- Measuring the height of a random person in the population
- Counting the number of mutated cells on a plate
- Predicting the behavior of the stock market

Why do we study probability theory?

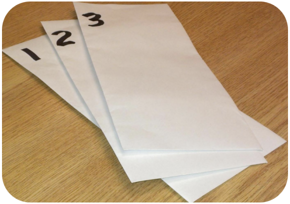
- It allows us to understand what happens in the population and relate it to what happens in a sample (see next chapter)
- Because it's fun/important/...

A famous problem...

A professor shows a student 3 envelopes. One contains the exam solutions, the others contain blank paper.

After the student chooses one of the envelopes, the professor opens one of the other envelopes. It contains blank paper.

The professor offers the student another choice. Should the student stick with their original choice, or should they pick the other envelope?



What do we mean by probability?

Probability = long-term relative frequency:

- Repeat the experiment n times
- Count the number of times n_E that E occurs

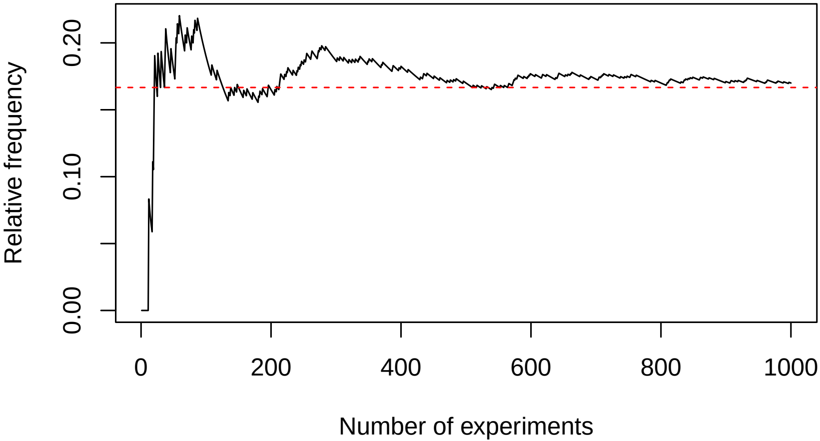
For n large, the relative frequency n_E/n is close to the probability $P(E)$ that E happens.

For example: tossing up a coin N times.

Number	N. Heads	$P(Heads)$
10	4	0.4
1,000	510	0.51
1,000,000	499,688	0.499688

Example: throwing a fair die

Intuitively, we expect the probability of a die coming up to 1 to be $p = 1/6$. Let's throw the die 1000 times and see what happens.



The sample space

The **sample space** S is the collection of all possible outcomes of an experiment.

Examples

- A single coin toss: $S = \{H, T\}$
- Three coin tosses in a row: $S = \{HHH, HHT, HTH, \ldots, TTT\}$
- Genotype: $S = \{CC, Cc, cc\}$
- Heights (in cm) of Korean people: $S = \mathbb{R}^+$

Note

The choice of sample space depends on the experiment!

Events

- An **event** E is a set of possible outcomes of an experiment
- An event is a subset of S: $E \subset S$.

Example

- Seeing heads in a coin toss: $E = \{H\}$
- Seeing exactly two heads in three coin tosses: $E = \{HHT, HTH, THH\}$
- Carrying at least one copy of a mutant allele: $E = \{Cc, cc\}$
- Height at least 190cm: $E = \{h \in \mathbb{R}^+ : h \geq 190\}$

Probabilities

The **probability** $P(E)$ of an event E describes how likely E is to happen.

Example

- Fair coin: $P(H) = P(T) = 1/2$
- Blood type of Korean people: $P(A) = 34\%$, $P(O) = 28\%$, $P(B) = 27\%$, $P(AB) = 11\%$.

Note

Probabilities must satisfy some rules!

Section 2

Rules for probabilities

Overview

1. Introduction: What do we mean by probability?
    1. Why study probability theory?
    2. Sample space, events, probability
2. **Rules for probabilities**
    1. Calculating with probabilities
    2. Conditional probabilities
3. Applications of Bayes' rule
4. Discrete random variables and probability distributions
    1. Bernoulli distribution
    2. Binomial distribution
    3. Poisson distribution
5. Solution of Monty Hall problem (optional)

Combining events

Consider two events:

$$A = \{\text{Married}, \text{Divorced}\} \quad \text{and} \quad B = \{\text{Married}, \text{Never married}\},$$

with $S = \{\text{Married}, \text{Divorced}, \text{Never married}\}$.

- **Union**: At least one of the events occurs ("OR' ').

$$A \cup B = A \text{ or } B = \{\text{Married}, \text{Divorced}, \text{Never married}\}$$

- **Intersection**: Both events occur ("AND' ').

$$A \cap B = A \text{ and } B = \{\text{Married}\}$$

- **Complement**: The event does not occur ("NOT' ').

$$A^c = \text{not } A = \{\text{Never married}\}$$

Disjoint events

Two events are **disjoint** if they have no outcomes in common:

$$A \cap B = \varnothing.$$

For example:

- $A = \{\text{Married}\}$
- $B = \{\text{Not married}\}$

Probability rules

Rule 1

The probability of an event E satisfies

$$0 \leq P(E) \leq 1$$

- Events with probability 0 occur (almost) never
- Events with probability 1 occur (almost) always

Rule 2

The sample space S satisfies

$$P(S) = 1$$

Some outcome must be realized on every trial.

Probability rules

Rule 3: Complement rule

For each event E,

$$P(E^c) = 1 - P(E)$$

For example:

$P(\text{blood type O}) = 0.28$, so $P(\text{not O}) = 1 - 0.28 = 0.72$.

Probability rules

Rule 4: Subset rule

For two events E and F such that $E \subseteq F$,

$$P(E) \leq P(F).$$

If an event contains fewer outcomes, it is less likely.

Probability rules

Rule 5: Addition rule (disjoint events)

For disjoint events E and F,

$$P(E \cup F) = P(E) + P(F)$$

Example (6-sided die)

What is the probability of obtaining 1 or an even number?

- $P(1) = \frac{1}{6}$
- $P(\text{even}) = \frac{1}{2}$
- Events $\{1\}$ and $\{\text{even}\}$ are disjoint.

Therefore $P(\{1\} \cup \text{even}) = P(1, 2, 4, 6) = \frac{1}{6} + \frac{1}{2} = \frac{2}{3}$.

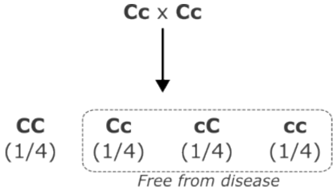
Example

Cystic fibrosis

Two parents are carriers for the cystic fibrosis gene but do not have the disease. What is the probability that their child will be free from the disease?

The events CC, Cc, and cC for the child

- are disjoint
- happen with probability $\frac{1}{4}$



Therefore,

$$P(\{CC, Cc, cC\}) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}$$

Probability rules

Rule 5: Addition rule (general events)

For arbitrary events E and F (not necessarily disjoint),

$$P(E \cup F) = P(E) + P(F) - P(E \cap F)$$

Example (6-sided die)

What is the probability of obtaining 2 or an even number?

- $P(2) = \frac{1}{6}$
- $P(\text{even}) = \frac{1}{2}$
- Events $\{2\}$ and $\{\text{even}\}$ are not disjoint: $\{2\} \cap \{\text{even}\} = \{2\}$

Therefore $P(\{2\} \cup \text{even}) = P(2, 4, 6) = \frac{1}{6} + \frac{1}{2} - \frac{1}{6} = \frac{1}{2}$.

Conditional probability

For two events A and B with $P(A) > 0$, the conditional probability of B given A is

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

What is the probability that B happens if we already know that A happened?

Example

What is the probability that a die comes up 6 if we know it came up on an even number?

$$P(\text{die} = 6|\text{even}) = \frac{P(\{\text{die} = 6\} \cap \{\text{even}\})}{P(\text{even})} = \frac{1/6}{1/2} = \frac{1}{3}.$$

Probability rules

Rule 6a: Multiplication rule

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

Example

There are 100 students in Ba3 and 90 students in Ba4. Campus president Han selects 2 students at random for a scholarship. What is the chance that both students are from Ba3?

Solution on the next slide.

Events:

- $A : \{\text{student 1 in Ba3}\}$
- $B : \{\text{student 2 in Ba3}\}$

Draw one student at random:

$$P(A) = \frac{100}{190}$$

When A happens, 99 students are left in Ba3. Hence

$$P(B|A) = \frac{99}{189}$$

Therefore,

$$P(A \cap B) = P(B|A)P(A) = \frac{100}{190}\frac{99}{189} = 0.276$$

Independence

Independence

Two events A and B that both have positive probabilities are independent if

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

Knowing that A occurred does not change the probability that B occurs.

Rule 6b: Multiplication rule for independent events

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

Example

Getting a 6 on two different rolls of a die.

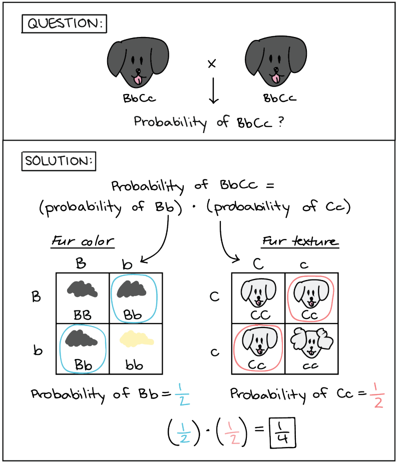
Independence: example

Genes that influence the fur type/color of dogs:

- Coat color: B black, b yellow
- Fur: C straight, c curly

Question

What is the probability that a cross of $BbCc$ and $BbCc$ will yield an offspring with genotype $BcCc$?



Example taken from KhanAcademy

Law of total probability

For 2 events

For two events A and B it holds that

$$P(B) = P(B|A)P(A) + P(B|A^C)P(A^C)$$

A partition of the sample space S is a collection of nonzero events $A_1, \ldots, A_k$ that are mutually disjoint and cover all of S.

In general

For any event B and partition $A_1, \ldots, A_k$

$$P(B) = \sum_{i=1}^{k} P(B|A_i)P(A_i)$$

Example

Problem statement

Professor Rao asks a random student to solve an ODE. The probability that a student answers the question successfully depends on the year they are in and is given by

$$P(\text{success}|\text{Ba3}) = 0.75, \quad \text{and} \quad P(\text{success}|\text{Ba4}) = 0.60.$$

There are 100 students in Ba3 and 90 students in Ba4. What is the probability that a random student can answer the question?

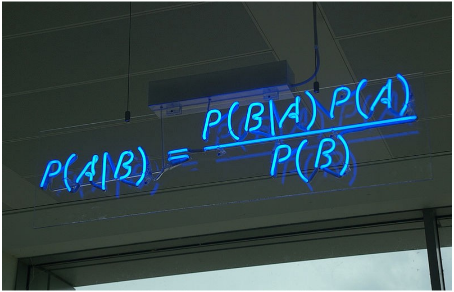
$$\begin{aligned} P(\text{success}) &= P(\text{success}|\text{Ba3})P(\text{Ba3}) + P(\text{success}|\text{Ba4})P(\text{Ba4}) \\ &= 0.75 \times \frac{100}{190} + 0.60 \times \frac{90}{190} \\ &= 0.679 \end{aligned}$$

Bayes' rule

For 2 events

For two events A and B whose probabilities are not 0 or 1,

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



"How does my belief about A change now that I know B?

Depends on

- Prior probability $P(A)$
- Likelihood $P(B|A)$

Example

Rainy days

Suppose there is a probability of 30% that any day is a rainy day. You observe someone enter your room carrying an umbrella. Does this make it more/less likely that it is raining outside?
You know that

$$P(\text{umbrella}|\text{rain}) = 0.8 \quad \text{and} \quad P(\text{umbrella}|\text{sunny}) = 0.25.$$

$$\begin{aligned} P(\text{rain}|\text{umbrella}) &= \frac{P(\text{umbrella}|\text{rain})P(\text{rain})}{P(\text{umbrella})} \\ &= \frac{0.8 \times 0.3}{0.8 \times 0.3 + 0.25 \times 0.7} \\ &= 0.578 \end{aligned}$$

Section 3

Applications of Bayes' rule

Overview

1. Introduction: What do we mean by probability?
   1. Why study probability theory?
   2. Sample space, events, probability
2. Rules for probabilities
   1. Calculating with probabilities
   2. Conditional probabilities
3. **Applications of Bayes' rule**
4. Discrete random variables and probability distributions
   1. Bernoulli distribution
   2. Binomial distribution
   3. Poisson distribution
5. Solution of Monty Hall problem (optional)

Diagnostic testing

Should you worry...

- ... when you take a positive COVID test?
- ... when you take two positive tests in a row?
- ... when you take a positive test for Ebola?

Bayes' rule can answer those questions, and much more.

Other uses of Bayes' rule

- Discovering planets
- Spam detection
- Learning from the environment (e.g. in self-driving cars)

Initial beliefs + New data = New and improved beliefs.

Sensitivity and specificity

Sensitivity

Probability that the test is positive, given that the patient is diseased:

$$\mathrm{Se} = P(\text{test positive} \mid \text{diseased}).$$

Specificity

Probability that the test is negative, given that the patient is not diseased:

$$\mathrm{Sp} = P(\text{test negative} \mid \text{not diseased}).$$

Neither the sensitivity nor the specificity depend on the prevalence of the disease.

- Useful for manufacturer: show that test is accurate enough.
- Not useful for consumer: wants to know chance of having disease.

Estimating the risk of disease

An individual taking the test is more interested in the probability that they have the disease.

- The positive predictive value:

$$\mathrm{PPV} = P(\text{dis.} \mid \text{test } +)$$

- The negative predictive value:

$$\mathrm{NPV} = P(\text{not dis.} \mid \text{test -})$$

Bayes' theorem allows us to calculate these probabilities.

Calculating the PPV/NPV

By Bayes' rule:

$$\mathrm{PPV} = P(\text{dis.} \mid \text{test } +) = \frac{P(\text{test } + \mid \text{dis.})P(\text{dis.})}{P(\text{test } +)}$$

By the rule of total probability:

$$P(\text{test } +) = P(\text{test } + \mid \text{dis.})P(\text{dis.}) + P(\text{test } + \mid \text{no dis.})P(\text{no dis.}).$$

Example: Covid diagnostics

(See whiteboard)

Section 4

Discrete random variables and probability distributions

Overview

1. Introduction: What do we mean by probability?
    1. Why study probability theory?
    2. Sample space, events, probability
2. Rules for probabilities
    1. Calculating with probabilities
    2. Conditional probabilities
3. Applications of Bayes' rule
4. **Discrete random variables and probability distributions**
    1. Bernoulli distribution
    2. Binomial distribution
    3. Poisson distribution
5. Solution of Monty Hall problem (optional)

Random variables

Definition

A random variable X is a variable that takes on values based on the outcome of a random operation.

- Nonnumerical outcomes are numerically encoded
- Usually denoted by capital letters $X, Y, \ldots$

Examples

- The height of a random Korean person
- The number of new cases of a disease in a given day
- The payoff in a bet on a coin: $X(H) = 0$, $X(T) = \$100$.

Discrete random variables

- A random variable X is discrete if X has a finite or countably infinite number of possible values
- The probability distribution of X lists these values and their probabilities

Examples

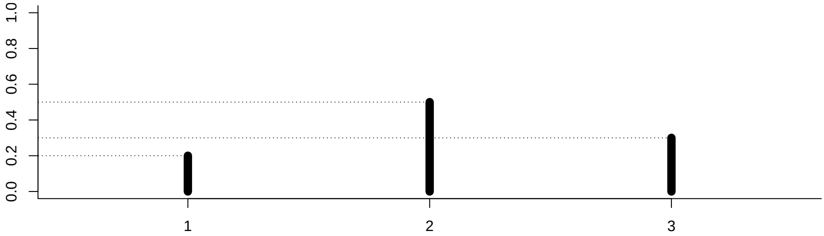
- $X = 0$ if person is male, $X = 1$ if female
- $Y = 0, 1, 2, \ldots$: number of cells on a plate
- $X = \$100$ if number on die is even, $X = \$0$ otherwise
- Z: gene expression level (count) for mutant/wild type allele

Discrete probability distribution

Describes for every value of X the probability that it occurs.

Properties: probabilities must ...

- Be non-negative: $p_i \geq 0$
- Sum to zero: $\sum_{i=1}^{N} p_i = 0$.



Mean of discrete random variables

If X is a discrete random variable with distribution

Outcome	x_1	x_2	$\ldots$	x_k
Probability	p_1	p_2	$\ldots$	p_k

Then

(population) mean μ of X

$$\mu = \sum_{j=1}^{k} p_j \, x_j = p_1 x_1 + p_2 x_2 + \cdots + p_k x_k$$

Mean of discrete random variables

- The mean of X is a long-run average of the outcomes of X
- μ is a weighted average of the outcomes, where each outcome is weighted by its probability
- The mean is a measure of the center of the distribution of X
- μ is also callled the expected value of X (denoted $E[X]$), but this does not mean that an outcome of X needs to be close to μ!

Variance of discrete random variables

(population) variance σ^2 of X

$$\sigma^2 = \sum_{j=1}^{k} p_j \left(x_j - \mu\right)^2$$

The standard deviation σ of X is the square root of the variance of X

- The variance of X is a long-run average of the squared deviations $(X - \mu)^2$
- σ^2 is a weighted average of the squared deviations, where each outcome is again weighted by its probability
- The variance is a measure of the spread of the distribution of X

Bernoulli distribution

Examples

- $X = 1$ if male, $X = 0$ if female
- $X = 1$ if mutated allele, $X = 0$ if wild type

Definition

X is Bernoulli if $X = 1$ ("success") with probability p, $X = 0$ with probability $1 - p$.

Mean:

$$\mu = 1 \cdot p + 0 \cdot (1 - p) = p$$

Variance:

$$\begin{aligned} \sigma^2 &= (1 - p)^2 \cdot p + (0 - p)^2 \cdot (1 - p) \\ &= p(1 - p) \end{aligned}$$

Binomial distribution

Definition

Consider n independent observations

- Each observation 2 possible outcomes: "success" or "failure"
- Same probability of success p for each observation

Binomial random variable X is number of "successes" among the n observations. Values: $0, 1, \ldots, n$

Notation

$X \sim B(n, p)$, **binomial distribution** with parameters n and p

Examples

Number of...

- Heads in 4 coin tosses
- Mutations in string of DNA of length N
- Votes obtained by candidate in runoff election

The distribution function

What do we know about the probability of having exactly k successes out of n trials?

- There are $\binom{n}{k}$ possibilities to organize the successes/failures
- Each success has probability p
- Each failure has probability $1 - p$.

Therefore

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

Mean and standard deviation

Binomial

Mean μ	np
Variance σ^2	$np(1-p)$

Prove using the binomial theorem.

Example

Counting lymphocytes

We can see 10 cells on a plate under the microscope. The probability of any one cell being a lymphocyte is $p = 0.20$. What is the probability of seeing

- Exactly 2 lymphocytes?
- At least 2 lymphocytes?

What number of lymphocytes do you expect on average?

Poisson distribution

Examples

Useful for counts of things per unit of space/time

- Number of seals in Alaska on aerial photographs
- Number of orders arriving at restaurant
- Number of goals in world championship

The distribution function

For $X \sim P(\lambda)$

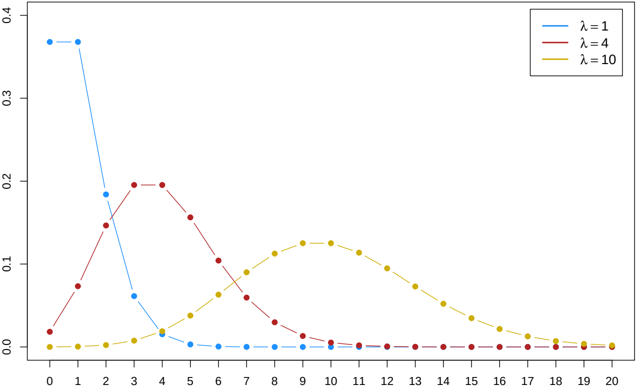
$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

Mean and standard deviation

Poisson

Mean μ $\quad\lambda$

Variance σ^2 $\quad\lambda$



Example

Counting deliveries

A delivery service gets on average 10 orders per hour. What is the probability that they will get more than 60 orders over a 5-hour timespan?

Section 5

Solution of the Monty Hall problem

Overview

1. Introduction: What do we mean by probability?
    1. Why study probability theory?
    2. Sample space, events, probability
2. Rules for probabilities
    1. Calculating with probabilities
    2. Conditional probabilities
3. Applications of Bayes' rule
4. Discrete random variables and probability distributions
    1. Bernoulli distribution
    2. Binomial distribution
    3. Poisson distribution
5. **Solution of Monty Hall problem (optional)**

Prior:

$$P(\text{sol} = 1) = P(\text{sol} = 2) = P(\text{sol} = 3) = \frac{1}{3}.$$

Let's say the student chooses envelope 1.

If it contains the solution, then the professor can open either envelope 2 or 3:

$$P(\text{opens} = 2 | \text{sol} = 1) = P(\text{opens} = 3 | \text{sol} = 1) = \frac{1}{2}.$$

If envelope 1 is not the winning one, then the solution must be in envelope 2 or 3. In either case, *the professor must open the other envelope*:

$$P(\text{opens} = 2 | \text{sol} = 3) = P(\text{opens} = 3 | \text{sol} = 2) = 1.$$

Using the law of total probability,

$$P(\text{opens} = 2) = \frac{1}{2} \times \frac{1}{3} + 0 \times \frac{1}{3} + 1 \times \frac{1}{3} = \frac{1}{2},$$

and similarly $P(\text{opens} = 3) = \frac{1}{2}$.

By Bayes' rule

$$P(\text{sol} = 1|\text{opens} = 2) = \frac{P(\text{opens} = 2|\text{sol} = 1)P(\text{sol} = 1)}{P(\text{opens} = 2)}$$
$$= \frac{1/2 \times 1/3}{1/2} = \frac{\mathbf{1}}{\mathbf{3}}.$$

while

$$P(\text{sol} = 3|\text{opens} = 2) = \frac{P(\text{opens} = 2|\text{sol} = 3)P(\text{sol} = 3)}{P(\text{opens} = 3)}$$
$$= \frac{1 \times 1/3}{1/2} = \frac{\mathbf{2}}{\mathbf{3}}.$$

Conclusion

By switching the student doubles their probability of choosing the correct envelope!