The Corona lecture notes

PROBABILITY

Mihael Perman

Note:

The following are the Corona lecture notes. They cover most of the material we covered by Zoom except Chapter 2.2 on continuous distributions.

1. Outcomes, events, probabilities

1.1. Outcomes and events

Example: Italian gamblers in the 17th century liked the game where they placed a bet on the outcome of rolling three dice. Popular bets were 9 and 10. The gamblers had a „theory" that the two popular bets are equivalent in the sense that the probability of winning is the same for both bets. They wrote down two lists:

Sum 9	Sum 10
1 2 6	1 3 6
1 3 5	1 4 5
1 4 4	2 2 6
2 2 5	2 3 5
2 3 4	2 4 4
3 3 3	3 3 4

Based on these two lists the two games were deemed equivalent. However, gambling experience suggested they were not.

The problem was solved by Galileo Galilei (1564 – 1642). He wrote down all possible outcomes.

111	112	113	114	115	116
121	122	123	124	125	126
⋮	⋮	⋮	⋮	⋮	⋮
661	662	663	664	665	666

There are $6^3 = 216$ possible triplets. Galileo found that 25 sum to 9 and 27 to 10. Assuming all triplets have the same chance of appearing the problem is solved.

The moral of the story is that we have to write down all possibilities when dealing with an experiment involving chance.

In mathematical language we will talk about the set of all possible outcomes and denote it by Ω.

Examples:

(i) In Galileo's example we have

$$\Omega = \{(i,j,k) : 1 \le i,j,k \le 6\}$$
$$= \{1,2,3,\dots,6\}^3$$

(ii) If we toss a coin n times we get a sequence of n heads and tails. In this case

$$\Omega = \{H,T\}^n .$$

(iii) Suppose we arrange n objects in random order. This means that we choose a random permutation or

$$\Omega = S_n = \text{set of all permutations.}$$

(iv) We can think of tossing a coin infinitely many times. In this case

$$\Omega = \{H, T\}^{\mathbb{N}}$$

which is the set of all countably infinite sequences of symbols H and T.

The next concept is the event. If we roll three dice an event is, say, that the sum is 9. An event can either happen or not. But what is an event mathematically? All the triplets that give a sum of 9 are a subset of $\Omega = \{1, 2, 3, 4, 5, 6\}^3$.

It is plausible to understand events as subsets of Ω.

We will denote events by $A, B, C, \dots$

For mathematical reasons denote the family of all events by $\mathcal{F}$.

We will require the following.

(i) $\Omega \in \mathcal{F}$.

(ii) if $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$.

(iii) if $A_1, A_2, \dots \in \mathcal{F}$, then

$$\bigcup_i A_i \in \mathcal{F}.$$

The union is either finite or countable.

Remark: In mathematics a family of subsets with the above properties is called a σ-algebra.

Remark: In cases of infinite sets Ω not all subsets are necessarily events. For finite Ω we will usually assume that all subsets are events.

In Galileo's example we assumed that all outcomes in Ω are equally likely. The probability of

$A = \{\text{sum is } 9\}$ is then $25/216$.

If $B = \{\text{sum is } 10\}$ then $P(B) = 27/216$.

We have $A \cap B = \emptyset$ and

$$P(A \cup B) = \frac{25 + 27}{216} = P(A) + P(B).$$

For mathematical reasons it turns out to be better to assign probabilities to events rather than outcomes. The example shows that for disjoint A and B we should have $P(A \cup B) = P(A) + P(B)$.

Any assignment of probabilities should have this property. The mathematical definition is more general.

<u>Definition</u>: Probability is an assignment to every event $A \in \mathcal{F}$ of a real number in such a way that

(i) $0 \le P(A) \le 1$, $P(\Omega) = 1$.

(ii) if $A_1, A_2, \ldots$ are disjoint we have

$$P\left(\bigcup_i A_i\right) = \sum_i P(A_i)$$

Remark: The sum in (ii) can be finite or infinite.

Remark: The property (ii) is called σ-additivity.

Let us look at some simple consequences of the above definition.

(i) we have $A \cup A^c = \Omega$ and $A \cap A^c = \emptyset$. By additivity

$$1 = P(\Omega) = P(A \cup A^c) = P(A) + P(A^c)$$

so $\boxed{P(A^c) = 1 - P(A)}$

(ii) Let A, B be events. We can write

$$A \cup B = \underbrace{(A \cap B^c) \cup (A \cap B) \cup (A^c \cap B)}_{\text{disjoint}}.$$

So

$$P(A \cup B) = P(A \cap B^c) + P(A \cap B) + P(A^c \cap B)$$

But $\underbrace{(A \cap B^c) \cup (A \cap B)}_{\text{disjoint}} = A$ so

$$P(A \cap B^c) + P(A \cap B) = P(A) \Rightarrow$$

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

and similarly

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

Using this in the above expression gives

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

If we have three events we get

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

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

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

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

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

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

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

$+ P(A \cap B \cap C)$

From this we generalize to

<u>Theorem 1.9</u> (inclusion-exclusion formula)

Let $A_1, A_2, \ldots, A_n$ be events. We have

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{1 \le i < j \le n} P(A_i \cap A_j)$$

$$+ \sum_{1 \le i < j < k \le n} P(A_i \cap A_j \cap A_k)$$

$$- \ldots$$

$$+ (-1)^{n-1} P(A_1 \cap A_2 \cap \ldots \cap A_n)$$

Proof: We know that the formula is valid for $n = 2$. Suppose it is valid for n. We write

$$\bigcup_{i=1}^{n+1} A_i = \bigcup_{i=1}^{n} A_i \cup A_{n+1} \quad \text{so}$$

$$P\Big(\bigcup_{i=1}^{n+1} A_i\Big) = P\Big(\bigcup_{i=1}^{n} A_i\Big) + P(A_{n+1})$$

$$- \underbrace{P\Big(\bigcup_{i=1}^{n} A_i \cap A_{n+1}\Big)}_{= P\left(\bigcup_{i=1}^{n} (A_i \cap A_{n+1})\right)}.$$

By the induction assumption the formula is valid for unions of n sets. This means

$$P\Big(\bigcup_{i=1}^{n} (A_i \cap A_{n+1})\Big) = \sum_{i=1}^{n} P(A_i \cap A_{n+1})$$

$$- \sum_{1 \le i < j \le n} P(A_i \cap A_j \cap A_{n+1})$$

$$+ \ldots + (-1)^{n-1} P(A_1 \cap \ldots \cap A_{n+1})$$

Using this gives the inclusion exclusion formula for $n+1$ sets, and the induction step is completed.

<u>Example</u>: n couples go dancing. When they are about to leave the power goes out and each woman grabs a man at random. What is the probability that no woman will grab her man?

In probability language we are talking about choosing a random permutation of n numbers. All permutations have the same probability $\frac{1}{n!}$.

<u>Figure</u>:

Women	1	2	3	$\ldots i \ldots$	n
Men	3	5	4	$\sigma(i)$	1

Define $A_i = \{ \text{woman } i \text{ grabs her man} \}$.

and $A = \{ \text{no woman grabs her man} \}$

We have

$$A^c = \bigcup_{i=1}^{n} A_i$$

To use the exclusion-exclusion formula we need the following probabilities:

\# of permutations sending $i \to i$.

$$P(A_i) = \frac{(n-1)!}{n!} = \frac{1}{n}$$

\# of all probabilities

$$P(A_i \cap A_j) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}$$

$$\vdots$$

$$P(A_{i_1} \cap \dots \cap A_{i_r}) = \frac{(n-r)!}{n!}$$

The inclusion - exclusion formula gives

$$P(A^c) = \binom{n}{1} \cdot \frac{1}{n} - \binom{n}{2} \cdot \frac{(n-2)!}{n!}$$

$$+ \binom{n}{3} \frac{(n-3)!}{n!} -$$

$$\vdots$$

$$\vdots$$

$$+ (-1)^n \quad \frac{0!}{n!}$$

$$= 1 - \frac{1}{2!} + \frac{1}{3!} - \dots + (-1)^{n-1} \cdot \frac{1}{n!}$$

The symbol $\binom{n}{r}$ counts the number of different intersections.

Finally

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

$$= \frac{1}{2!} - \frac{1}{3!} + \dots + (-1)^n \cdot \frac{1}{n!}$$

From Analysis we know

$$e^{x} = \sum_{k=0}^{\infty} \frac{x^{k}}{k!} = 1 + \frac{x}{1!} + \frac{x^{2}}{2!} + \dots\dots$$

Take $x = -1$ to get

$$e^{-1} = \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \dots\dots$$

The probability $P(A)$ is a partial sum of the above series which converges fast. We can approximate

$$P(A) \approx e^{-1} = 0.3697.$$

1.2. Conditional probabilities and independence

Example: Let us return to Galileo's example. We have $\Omega = \{1,2,3,4,5,6\}^3$ and all triplets are equally likely. Suppose you know that the first component is 1 but not the other two components. What is your opinion about the probability that the sum is 9? There are 36 triplets of the form $(1,j,k)$. Of these the triplets

$(1,2,6)$
$(1,3,5)$
$(1,4,4)$
$(1,5,3)$
$(1,6,2)$

give a sum of 9.

Is it reasonable to assum that given the information that the first component is 1 all the 36 triplets are equally likely ?

Yes. So the updated probability is $5/36$ which is different from $25/216$. We rewrite

$$5/36 = \frac{5/216}{36/216}$$

and denote $A = \{\text{sum is } 9\}$ and $B = \{\text{first component is } 1\}$.

We have

$$5/36 = \frac{P(A \cap B)}{P(B)}$$

Definition: The conditional probability of A given B is

$$\boxed{P(A \mid B) = \frac{P(A \cap B)}{P(B)}.}$$

Remark: If we have additional information about an outcome this usually means that the outcome is in a restricted subset of Ω. In the above example this restricted subset is B.

Rewriting the definition we get

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

If $A_1, A_2, \ldots, A_n$ are events we can write

$$P(A_1 \cap \cdots \cap A_{n-1} \cap A_n) =$$

$$P(A_n \mid A_1 \cap \cdots \cap A_{n-1}) \, P(A_1 \cap \cdots \cap A_{n-1})$$

Iterating the rule gives

$$\boxed{\begin{aligned} &P(A_1 \cap \dots \cap A_n) \\ &= P(A_n \mid A_1 \cap \dots \cap A_{n-1}) \cdot \\ &\qquad P(A_{n-1} \mid A_1 \cap \dots \cap A_{n-2}) \\ &\qquad\qquad \ddots \\ &\qquad\qquad P(A_2 \mid A_1) \cdot P(A_1) \end{aligned}}$$

Definition: A collection $\{H_1, H_2, \dots, H_n\}$ is a partition of Ω if $H_i \cap H_j = \emptyset$ for all $i \neq j$ and $\bigcup_{i=1}^{n} H_i = \Omega$.

Theorem 1.2 (law of total probabilities)

Let $\{H_1, H_2, \dots, H_n\}$ be a partition and A an event. We have

$$P(A) = \sum_{i=1}^{n} P(A \mid H_i) \cdot P(H_i)$$

Proof: We write

$$\begin{aligned} A &= A \cap \Omega \\ &= A \cap \bigcup_{i=1}^{n} H_i \\ &= \underbrace{\bigcup_{i=1}^{n} (A \cap H_i)}_{\text{disjoint events.}} \end{aligned}$$

It follows that

$$\begin{aligned} P(A) &= \sum_{i=1}^{n} P(A \cap H_i) \\ &= \sum_{i=1}^{n} \frac{P(A \cap H_i)}{P(H_i)} \cdot P(H_i) \\ &= \sum_{i=1}^{n} P(A \mid H_i) \cdot P(H_i) \end{aligned}$$

Remark: This formula is useful to compute probabilities. We can often guess conditional probabilities.

Example: In an Internet game of chance you have 12 tickets.

$\boxed{1}$ $\boxed{1}$ $\boxed{1}$ $\boxed{1}$ $\boxed{2}$ $\boxed{2}$ $\boxed{3}$ $\boxed{D}$ $\boxed{S}$ $\boxed{3}$ $\boxed{S}$ $\boxed{S}$

The tickets are randomly permuted and turned around so that the player sees

The player then turns around the tickets until the first $\boxed{S}$ = STOP. One example is

$\boxed{1}$ $\boxed{2}$ $\boxed{D}$ $\boxed{1}$ $\boxed{S}$

The payoff is the sum of digits multiplied by 2 if $\boxed{D}$ = DOUBLE is among the visible tickets.

In the above example the payoff is 8.

What is the probability that the player will see the ticket $\boxed{D}$?

Define $H_i = \{$ first $\boxed{S}$ is in position $i \}$ for $i = 1, 2, \dots, 9$. The collection $\{H_1, H_2, \dots, H_9\}$ is a partition.

Let $A = \{$ we see $\boxed{D} \}$.

First we compute

$$P(H_i) = \frac{8}{12} \cdot \frac{7}{11} \cdot \; \cdot \frac{8-i+2}{12-i+2} \cdot \frac{4}{12-i+1}$$

What about $P(A \mid H_i)$? Idea ?

If $\boxed{S}$ appears in position i then the first $i-1$ tickets are randomly chosen from

$\boxed{\alpha}$ $\boxed{1}$ $\boxed{1}$ $\boxed{1}$ $\boxed{2}$ $\boxed{2}$ $\boxed{3}$ $\boxed{D}$. So we choose $i-1$ tickets out of 8 and ask for the (conditional) probability that $\boxed{D}$ is among the tickets chosen. We have

$$P(A \mid H_i) = \frac{\binom{7}{i-2}}{\binom{8}{i-1}}$$

$\leftarrow$ # of samples of size $i-1$ containing $\boxed{D}$.

$\uparrow$

# of possible samples

Cancelling we get

$$P(A \mid H_i) = \frac{i-1}{8}$$

<u>Check</u>: For $i = 9$ we should get 1.

The rest is adding fractions.

$$P(A) = \sum_{i=1}^{9} \frac{8!\,(12-i)!}{12!\,(8-i+1)!} \cdot 4 \cdot \frac{(i-1)}{8}$$

$$= \frac{8!}{12!} \cdot \frac{1}{2} \sum_{i=1}^{9} \frac{(12-i)!\,(i-1)}{(8-i+1)!}$$

$$= 1/5$$

Example: Prisoner's Paradox

Three prisoners are in jail in a dark country. They are all sentenced to death but the ruler will choose one of them at random and pardon him. Here is a conversation between the guard in jail and prisoner A:

A: Guard, you already know who will be pardoned. If you tell me who of the other two will not be pardoned you do not give me any information.

G: If I tell you there will be only two of you left. Your probability of survival is then 1/2. I do give you some information.

Who is right? To talk about conditional probabilities we need a space of all possible outcomes. Here is a suggestion:

$$\Omega = \left\{ \begin{array}{ll} \boxed{AB|B} & 1/3 \\ \boxed{AC|C} & 1/3 \\ \boxed{BC|B} & \multirow{2}{*}{$\left.\begin{array}{l} \\ \\ \end{array}\right\} 1/3 \begin{array}{l} \to x/3 \\ \to (1-x)/3 \end{array}$} \\ \boxed{BC|C} & \end{array} \right.$$

↑

Last letter is what the guard says

First two letters are the wretched prisoners who will be hanged

There is no indication how the last probability of $1/3$ is distributed between the last two outcomes. Let us say $\frac{x}{3}$ and $\frac{1-x}{3}$ for $x \in [0,1]$.

We compute

$$P(\text{A survives} \mid \text{Guard says B})$$

$$= \frac{P(\{\text{A survives}\} \cap \{\text{Guard says B}\})}{P(\text{Guard says B})}$$

$$= \frac{x/3}{1/3 + x/3}$$

$$= \frac{x}{1+x}$$

This function has values from 0 to $1/2$ on $[0,1]$.

Two cases:

(i) if $x = 1/2$ the guard chooses at random when he has the choice. In this case the conditional probability is

$$\frac{1/2}{1+1/2} = 1/3$$ as before.

(ii) If $x = 0$ then the conditional probability is 0!

Why?

What if $P(A|B) = P(A)$? Then B does not „tell us anything" about the probability of A. The word we choose is independence. The above equality can be written as

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

by definition. If we have events A, B and C and they are „independent" then $A \cap B$ ought to be independent of C. This leads to the following definition.

<u>Definition</u> :

(i) The events A and B are independent if

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

(ii) Events $A_1, A_2, \ldots, A_n$ are independent if for all collections of indeces $1 \le i_1 < i_2 < \ldots < i_m \le n$ we have

$$P(A_{i_1} \cap \cdots \cap A_{i_m})$$

$$= P(A_{i_1}) P(A_{i_2}) \cdots P(A_{i_m}).$$

<u>Remark</u>: Typically independence is associated with physically different objects, like several dice, different coins.

Example (Paradox of Chevalier de Méré, Antoine Gombaud, 1607 – 1684).

Chevalier de Méré considered the following two games of chance.

(i) You roll a die 4 times. You win if you see at least one ace (ace = ⚀)

(ii) You roll two dice 24 times. You win if you see at least one double ace, i.e. ⚀⚀

Which of the two games has a higher probability of winning?

Let us look at the first game.

Define $A_i = \{$ the i-th roll is not ⚀ $\}$ and $A = \{$ we win $\}$.

We have

$$A^c = A_1 \cap A_2 \cap A_3 \cap A_4$$

It is reasonable to assume that subsequent rolls are independent which means that A_1, A_2, A_3, A_4 are independent. It follows

$$P(A^c) = P(A_1 \cap A_2 \cap A_3 \cap A_4)$$
$$= P(A_1) \cdot P(A_2) \cdot P(A_3) \cdot P(A_4)$$
$$= 5/6 \cdot 5/6 \cdot 5/6 \cdot 5/6$$
$$= (5/6)^4$$

Finally we get

$$P(A) = 1 - P(A^c) = 1 - (5/6)^4 \doteq 0.5177$$

For the second game we similarly define

$$A_i = \{\text{not a double ace on roll } i\},$$
$$i = 1, 2, \ldots, 24.$$

$A = \{ \text{we win} \}$

We have

$$A^c = A_1 \cap A_2 \cap \dots \cap A_{24}$$

We assume independence and get

$$P(A^c) = P(A_1) P(A_2) \cdots P(A_{24})$$

$$= 35/36 \cdot 35/36 \cdot \dots \cdot 35/36$$

$$= \left(35/36\right)^{24},$$

and finally

$$P(A) = 1 - P(A^c) = 1 - \left(35/36\right)^{24} \doteq 0.4914.$$

Comment: The difference is small but nevertheless important.

Example (Gambler's ruin).

Two gamblers A and B start out with m and n sequins (gold coins) respectively. In each round of the game they toss a coin. If it is heads A gets a coin from B; if it is tails B gets a coin from A. They play until one of them is left with no coins. What is the probability that A will get all the coins? We assume that tosses are independent and the probability of heads is $p \in (0,1)$.

Let $A = \{$ gambler A wins $\}$ and denote $p_{m,n} = P(A \text{ wins})$.

Let $H = \{$ first toss is heads $\}$.

Then

$$P(A) = \underbrace{P(A \mid H)}_{= p_{m+1,n-1}} \cdot P(H) + \underbrace{P(A \mid H^c)}_{= p_{m-1,n+1}} P(H^c)$$

We get the recursion

$$p_{m,n} = p \cdot p_{m+1,n-1} + (1-p)\, p_{m-1,n+1}.$$

Obviously $p_{0,m+n} = 0$ and $p_{m+n,0} = 1$.

Denote

$$\pi_m = p_{m,n}.$$

We can rewrite the recursion as

$$\pi_m = p \cdot \pi_{m+1} + (1-p)\, \pi_{m-1}.$$

with $\pi_{m+n} = 1$ and $\pi_0 = 0$.

Define $q := 1-p$. We rewrite

$$\underbrace{(p+q)}_{=1} \pi_m = p\,\pi_{m+1} + q\,\pi_{m-1} \quad , \text{ or}$$

$$p\,(\pi_{m+1} - \pi_m) = q\,(\pi_m - \pi_{m-1}) \quad , \text{ or}$$

$$\pi_{m+1} - \pi_m = q/p\,(\pi_m - \pi_{m-1})$$

Write

$$\pi_2 - \pi_1 = q/p\,(\pi_1 - \pi_0)$$

$$\pi_3 - \pi_2 = q/p\,(\pi_2 - \pi_1)$$

$$= (q/p)^2\,(\pi_1 - \pi_0)$$

$$\vdots$$

$$\pi_{m+n} - \pi_{m+n-1} = (q/p)^{m+n-1}\,(\pi_1 - \pi_0)$$

From this we get by adding

$$\pi_1 \left(1 + q/p + \cdots + (q/p)^{m+n-1}\right)$$

$$= \pi_{m+n}$$

$$= 1$$

It follows that

$$\pi_1 = \frac{1}{1 + q/p + \cdots + q/p^{m+n-1}}.$$

As a consequence we have

$$\pi_m = p_{m,n} = \frac{1 + (q/p) + \cdots + (q/p)^{m-1}}{1 + q/p + \cdots + (q/p)^{m+n-1}}.$$

How do we know that the game will end?

Let B_k be defined as

$B_k = \{$ tosses $(m+n)k+1, (m+n)k+2, \ldots$

$\ldots (m+n)(k+1) - 1$

produce heads $\}$

By independence

$$P(B_k) = p^{m+n}.$$

But B_k depend on disjoint blocks of events so they are independent. But

$$\{\text{game ends}\} \supseteq \bigcup_{k=1}^{\infty} B_k$$

We compute

$$P\left(\bigcup_{k=1}^{\infty} B_k\right) = \lim_{k \to \infty} P\left(\bigcup_{k=1}^{r} B_k\right)$$

$$= \lim_{r \to \infty} \left(1 - P\left(\bigcap_{k=1}^{n} B_k^c\right)\right)$$

$$= 1 - \lim_{r \to \infty} P(B_k^c)^r$$

$$= 1 - \lim_{r \to \infty} (1 - p^{m+n})^r$$

$$= \underline{1}.$$

<u>Lemma 1.3</u>: Let $A_1, A_2, \ldots$ be events.

(i) If $A_1 \subseteq A_2 \subseteq \ldots$ then

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = \lim_{n\to\infty} P(A_n)$$

(ii) If $A_1 \supseteq A_2 \supseteq \ldots$ then

$$P\left(\bigcap_{k=1}^{\infty} A_k\right) = \lim_{n\to\infty} P(A_n)$$

<u>Proof</u>: We only prove (i). The second assertion follows by de Morgan rules. Write

$$\bigcup_{k=1}^{\infty} A_k = A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_1 \cup A_2) \cup \ldots$$

The events in the union on the right are disjoint. It follows that

$$P\left(\bigcup_{k=1}^{\infty} A_k\right) = P(A_1) + P(A_2 \setminus A_1) + \ldots$$

The infinite series on the right converges and its partial sum is

$$P(A_1) + P(A_2 \setminus A_1) + \cdots + P(A_n \setminus A_1 \cup \cdots \cup A_{n-1})$$

$$= P\left(\bigcup_{k=1}^{n} A_k\right) = P(A_n).$$

The assertion follows.