Conditional Probability

Conditional Probability

Given two events A, B in a σ-field $\mathscr{F}$ on a sample space S we can calculate the probability that

- "event A occurs",
- "event A does not occur",
- "events A and B occur" and
- "event A or event B occurs".

The axioms do not, however, provide us with a way to calculate the probability that

- "event B occurs if event A has occurred."

In other words, given information about whether an event A has occurred, we would like to (re-)calculate the probability of B occurring.

Conditional Probability

Let us denote by

$$P[B \mid A]$$

the ***conditional probability*** that "B occurs given that A has occurred".

2.1. Example. recall from the previous Example **??** that in rolling two dice we considered the events

$$\begin{aligned} A_1 &= \{\boxed{(1,1)}, (1,2), (2,1)\}, \\ A_2 &= \{\boxed{(1,1)}, (2,2), (3,3), (4,4), (5,5), (6,6)\}. \end{aligned}$$

What is then $P[A_1 \mid A_2]$? A_2发生，所以A_2可以考虑为new sample space

If we somehow have the information that the die rolls were equal, we can then conclude that A_1 is only possible if among the six results in A_2 the single result $(1,1)$ has occurred. We should have

$$P[A_1 \mid A_2] = \frac{1}{6}.$$

Generalizing the Counting Approach

What we have done is calculate

$$P[A_1 \mid A_2] = \frac{|A_1 \cap A_2|}{|A_2|}$$

where $|A|$ denotes the number of elements of A. We could re-write this as

$$\begin{aligned} P[A_1 \mid A_2] &= \frac{|A_1 \cap A_2|/|S|}{|A_2|/|S|} \\ &= \frac{P[A_1 \cap A_2]}{P[A_2]}. \end{aligned}$$

This last expression is independent of the "counting" approach and uses only known probabilistic quantities, so we now define

$$P[B \mid A] := \frac{P[A \cap B]}{P[A]}$$

whenever $P[A] \neq 0$.

The Two Children Problem

Girl Scout in Uniform. File:Girls scout leader in uniform in Trebir, Trebc District.jpg. (2019, September 4).

Suppose that in any birth the probability of the baby being male or female is equal and that the sex is not influenced by that of any siblings.

At a gathering of parents of Girl Scouts, you meet a mother with her daughter. She says "Actually, I also have a second child."

(i) What is the probability that the other child is a boy?

The sample space has three elements: $\{(g, g), (b, g), (g, b)\}$, where the first element of the tuple is the gender of the first child, the second element that of the second child. Then we obtain

$$P[\text{other child is a boy}] = \frac{2}{3}.$$

The Two Children Problem

(ii) She then tells you, "My daughter here is my older child." What is the probability that the other child is a boy?

Now

$$\begin{aligned} & P[\text{other child is a boy} \mid \text{the older child is a girl}] \\ &= \frac{P[\text{other child is a boy and the older child is a girl}]}{P[\text{the older child is a girl}]} \\ &= \frac{1/3}{2/3} = \frac{1}{2}. \end{aligned}$$

The same result would be true if she had said, "My daughter here is my younger child." Even though the duaghter must either be younger or older than the second child, knowing which of the two options applies yields a probability of $1/2$ for each sex of the other child. Not knowing the relative age makes it $2/3$ probable that the other child is a boy.

Independence of Events

If one event does not influence another, then we say that the two events are independent. Formally, we say that two events A and B are ***independent*** if

$$P[A \cap B] = P[A]P[B]. \tag{2.1}$$

Equation (2.1) is equivalent to

$$\boxed{\begin{aligned} P[A \mid B] &= P[A] && \text{if } P[B] \neq 0, \\ P[B \mid A] &= P[B] && \text{if } P[A] \neq 0, \end{aligned}}$$

which correspond to the intuitive idea that the probability of A is not affected by B occurring and vice-versa.

The Birthday Problem

2.2. Example. The birthdays (day and month) of a group of people are generally assumed to be independent. Disregarding leap years, any person is assumed to have a $1/365$ chance of being born on a given day. (Do you think that this is a reasonable assumption?) How many people should a group have so that there is a better than even chance of two people in the group having the same birthday?

We consider the complementary problem and start with a single person in the group. If we add a second person, there is a $364/365$ chance of them ***not*** sharing a birthday. Adding a third person, for no two people to share a birthday, this person must have his birthday on one of the other 363 days of the year, so there is now a

$$\frac{364}{365}\frac{363}{365}$$

chance of no two people in the group sharing a birthday.

The Birthday Problem

Continuing this argument, in a group of $n \geq 2$ people there is a

$$\prod_{k=2}^{n} \frac{366-k}{365} = \frac{1}{365^{n-1}} \frac{364!}{(365-n)!}$$

chance of no two people having the same birthday. It turns out that for $n = 23$ this number is less than 0.5, so the probability of two people having the same birthday is > 0.5.

This statement has been verified empirically; in a soccer match there are 2×11 players $+ 1$ referee on the pitch. On any given playing day in the Premier Division of the English league, about half the games should feature two participants with the same birthday.

Literature: *Coincidences: The truth is out there*, Teaching Statistics, Vol. 1, No. 1, 1998

Independence vs. Law of Large Numbers

On the one hand, successive flips of a coin are independent - the result of one coin flip should not influence the result of the following coin clips.

On the other hand, experience tells us that if we toss a fair coin many times, it should not come heads up all the time. On average, we expect about one-half of the results to be heads.

Jacob Bernoulli (1654-1705). Painting by Niklaus Bernoulli in 1687. File:Jakob Bernoulli.jpg. (2016, December 29). Wikimedia Commons, the free media repository.

This principle was formulated by Jacob Bernoulli in the early 18th century as the ***Law of Large Numbers***:

$$\text{Probability} \quad \longleftrightarrow \quad \text{Proportion of outcomes}$$

Heuristic Version of the Law of Large Numbers

2.3. Heuristic Law of Large Numbers. Let A be a random outcome of an experiment that can be repeated without this outcome influencing subsequent repetitions. Then the probability $P[A]$ of this event occurring may be approximated by

$$P[A] \approx \frac{\text{number of times } A \text{ occurs}}{\text{number of times experiment is performed}}$$

We will give a more precise statement of the law later.

Total Probability

Recall that two events A and B are mutually exclusive if $A \cap B = \emptyset$.

Consider a set of n, pairwise mutually exclusive events $A_1, \dots, A_n$ in a sample space S with the additional properties that $P[A_k] \neq 0$ for all $k = 1, \dots, n$ and $A_1 \cup \dots \cup A_n = S$. Let $B \subset S$ be any event. Then

$$\begin{aligned} P[B] &= P[B \cap S] = P\big[B \cap (A_1 \cup \dots \cup A_n)\big] \\ &= P\big[(B \cap A_1) \cup \dots \cup (B \cap A_n)\big] \\ &= P[B \cap A_1] + \dots + P[B \cap A_n] \\ &= P[B \mid A_1] \cdot P[A_1] + \dots + P[B \mid A_n] \cdot P[A_n] \end{aligned}$$

$P[B|A] = \dfrac{P[B \cap A]}{P[A]}$

与向量向基向量分解相似。(向量用基底表示)

The expression

$$P[B] = \sum_{k=1}^{n} P[B \mid A_k] \cdot P[A_k]. \tag{2.2}$$

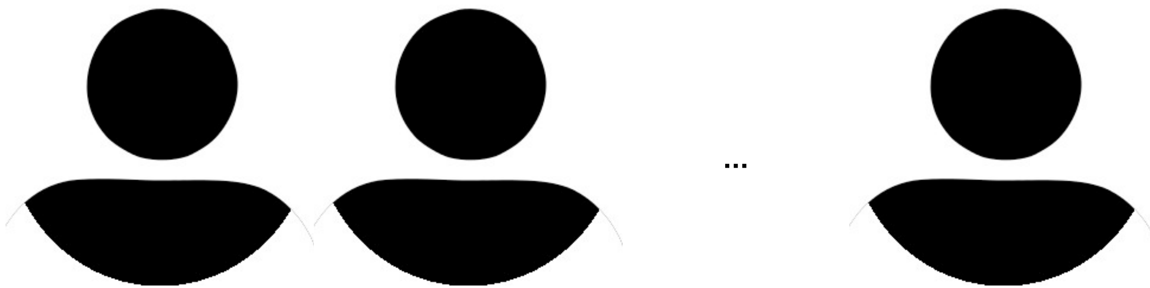
$\vec{x} = \sum_{i=1}^{n} \langle \vec{e}_i, \vec{x} \rangle \vec{e}_i$

is called the ***total probability*** formula for $P[B]$.

The Marriage Problem

As an application of the formula for total probability, consider the following ***marriage problem***:

Suppose you are trying to find the "perfect partner." There are n partners available, and they can be ranked from 1 to n with regard to "suitability", where the "most suitable" partner has rank 1 and the "least suitable" partner has rank n. We write $x_k \in \{1, \dots, n\}$ for the rank of the kth partner.

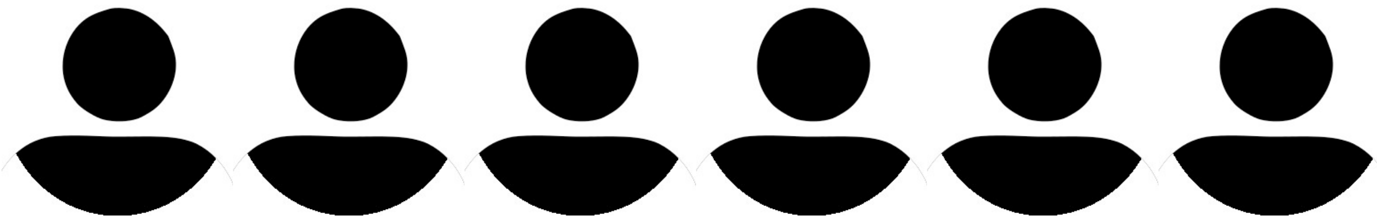


$x_1 = ?$ $x_2 = ?$ ... $x_n = ?$

The Marriage Problem

You can ***evaluate*** each partner, but only detect their ***relative rank*** y_k: ("best so far", "second best so far", etc.).

2.4. Example.



$x_1 = 3$	$x_2 = 6$	$x_3 = 2$	$x_4 = 4$	$x_5 = 1$	$x_6 = 5$
$y_1 = 1$	$y_2 = 2$	$y_3 = 1$	$y_4 = 3$	$y_5 = 1$	$y_6 = 5$

After evaluation: Accept or discard forever.

Goal: Find ***most suitable partner*** with $x_k = 1$.

Strategy and Outcomes

Optimal strategy: For some $r \geq 1$, evaluate and automatically

reject $r - 1$ ***potential partners***.

Then select the first candidate superior to all the previous ones, if possible.
To summarize:

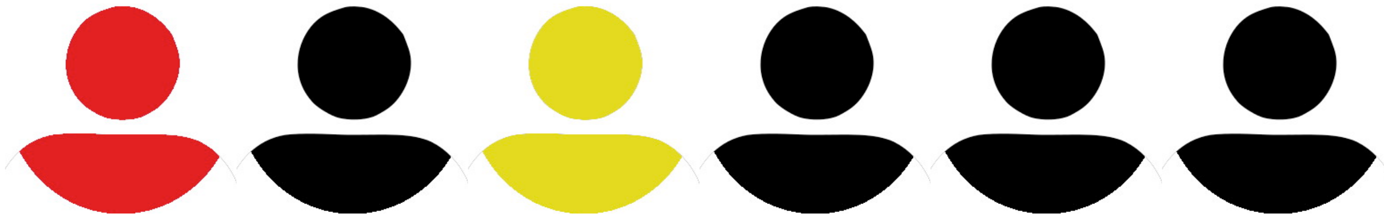
1. Choose $r \geq 1$.
2. Select k with $y_k = 1$ and $k \geq r$, if possible. Discard all others.
3. Otherwise, do not choose anyone.

Possible outcomes:

- The most suitable partner ($x_k = 1$) is selected;
- A less suitable partner ($x_k > 1$) is selected;
- No partner is selected.

Examples

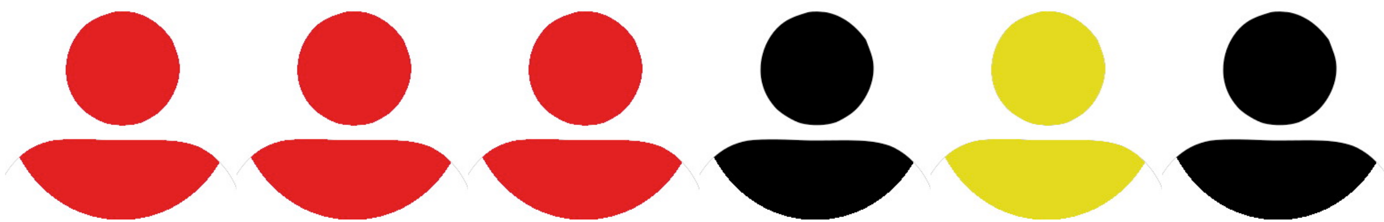
$r = 2$:



$x_1 = 3$ $x_2 = 6$ $x_3 = 2$ $x_4 = 4$ $x_5 = 1$ $x_6 = 5$

$y_1 = 1$ $y_2 = 2$ $y_3 = 1$ $y_4 = 3$ $y_5 = 1$ $y_6 = 5$

$r = 4$:



$x_1 = 3$ $x_2 = 6$ $x_3 = 2$ $x_4 = 4$ $x_5 = 1$ $x_6 = 5$

$y_1 = 1$ $y_2 = 2$ $y_3 = 1$ $y_4 = 3$ $y_5 = 1$ $y_6 = 5$

Finding the Optimal Strategy

The sample space can be taken to be

$$S = \{(k, j) \colon k \text{ is selected and } x_j = 1, \ k = 0, \dots, n, \ j = 1, \dots, n\}$$

where $k = 0$ indicates that no person was selected. We say that we "win" if we end up by selecting the most suitable partner, i.e., we select k with $x_k = 1$. We denote this event by

$$W_r = \big\{(k, k) \colon k = r, \dots, n\big\}, \qquad\qquad r \geq 1.$$

Given $r \geq 1$, the probability of winning is denoted

$$p_r = P[W_r].$$

Problem: ***How to choose r so that p_r is maximal?***

The Total Probability Formula

We denote the event that the mth person is the most suitable partner by

$$B_m = \{(k, m) \colon k = 0, \dots, n\}, \qquad m = 1, \dots, n.$$

Note that the B_m are mutually exclusive and that their union is S. Then by the formula for total probability,

$$p_r = P[W_r] = \sum_{m=1}^{n} P[W_r \mid B_m] P[B_m].$$

Evaluating the partners in random order,

$$P[B_m] = \frac{1}{n}, \qquad m = 1, \dots, n.$$

Probability of Selecting No Partner

No partner will be chosen if and only if the very best candidate was discarded. The probability of this happening is

$$P[\text{selecting no partner}] = \frac{r-1}{n},$$

Of course, in that case we can't win:

$$P[W_r \mid B_m] = 0 \qquad \text{for } m < r$$

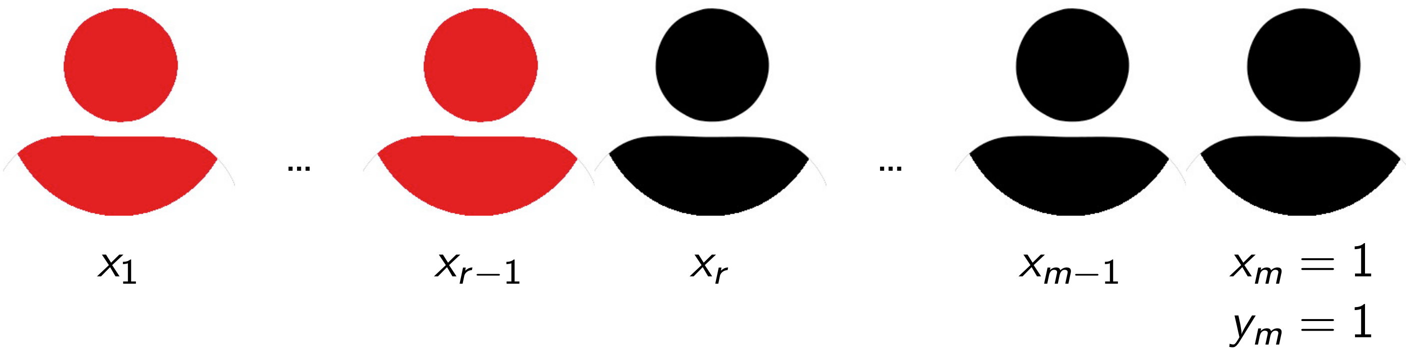
since the most suitable partner will have been discarded.

Therefore, the expression reduces to

$$p_r = \frac{1}{n} \sum_{m=r}^{n} P[W_r \mid B_m].$$

The Marriage Problem

Suppose that $x_m = 1$ for $m \geq r$:



We will win if there is no relative rank $y_k = 1$ for $r \leq k < m$.

Hence, the minimum of the finite sequence $(x_1, \ldots, x_{r-1}, x_r, \ldots, x_{m-1})$ must occur for one of the subscripts $1 \leq k \leq r - 1$.

This will happen with probability

$$P[W_r \mid B_m] = \frac{r-1}{m-1}.$$

The Marriage Problem

We hence have

$$p_r = \frac{r-1}{n} \sum_{m=r}^{n} \frac{1}{m-1}$$

To find the maximum of this expression, we set $x = r/n$ and use the approximation

$$\sum_{k=1}^{n} \frac{1}{k} \approx \ln(n) - \gamma$$

where γ is the Euler-Mascheroni constant. Then

$$\begin{aligned}
p_r &= \left(x - \frac{1}{n}\right) \left(\sum_{m=2}^{n} \frac{1}{m-1} - \sum_{m=2}^{r-1} \frac{1}{m-1}\right) \\
&\approx \underbrace{\left(x - \frac{1}{n}\right)}_{\approx x} \underbrace{(\ln(n-1) - \ln(r-2))}_{= \ln \frac{n-1}{r-2} \approx -\ln x} \\
&\approx -x \ln x.
\end{aligned}$$

The Marriage Problem

The maximum is now easily found using calculus, yielding $x_{\mathrm{max}} = 1/e$. We hence take $r = \lceil n/e \rceil$, which is about 37% of the partners. Note that p_r has the value $1/e$ at x_{max}.

The optimal strategy can be summarized as follows: evaluate and reject 37% of the partners, then choose the first partner that is more suitable than any of the preceding partners. This strategy will yield the most suitable partner 37% of the time, lead to no choice (rejection of all partners)

$$\frac{r-1}{n} \approx \frac{r}{n} = x_{\mathrm{max}} = 37\%$$

of the time and lead to an inferior choice 26% of the time.

Bayes's Theorem

From the formula for total probability we immediately obtain one of the most important theorems of elementary probability:

2.5. Bayes's Theorem. Let $A_1, \ldots, A_n \subset S$ be a set of pairwise mutually exclusive events whose union is S and who each have non-zero probability of occurring. Let $B \subset S$ be any event such that $P[B] \neq 0$. Then for any A_k, $k = 1, \ldots, n$,

$$P[A_k \mid B] = \frac{P[B \cap A_k]}{P[B]} = \frac{P[B \mid A_k] \cdot P[A_k]}{\sum_{j=1}^{n} P[B \mid A_j] \cdot P[A_j]}.$$

reverse

The theorem is due to the English mathematician ***Thomas Bayes (1701? - 1761)***. Unfortunately, no clearly authentic image of him survives.

Bayes's Theorem

2.6. Example. Suppose that a rare disease occurs at a rate of 0.1%, i.e., one out of a thousand people have that disease. Suppose a test for the disease is developed that is 99% accurate, i.e., if someone has the disease, the test determines this with 99% accuracy and if someone does not have the disease, the test is negative 99% of the time.

Suppose a patient is tested positive for the disease. What is the probability that she actually has the disease?

We know that

$$\begin{aligned} P[\text{has disease}] &= 0.001, \\ P[\text{test positive} \mid \text{has disease}] &= 0.99, \\ P[\text{test negative} \mid \text{does not have disease}] &= 0.99. \end{aligned}$$

不同

What we need is $P[\text{has disease} \mid \text{test positive}]$.

Bayes's Theorem

Let us write D for the event "has disease", $\neg D$ for "does not have disease", n for "test negative" and p for "test positive".

By Bayes's Theorem,

$$P[\text{has disease} \mid \text{test positive}] = P[D \mid p] = \frac{P[D \text{ and } p]}{P[p]}$$
$$= \frac{P[p \mid D] \cdot P[D]}{P[p \mid D] \cdot P[D] + P[p \mid \neg D] \cdot P[\neg D]}$$
$$= \frac{0.99 \cdot 0.001}{0.99 \cdot 0.001 + 0.01 \cdot 0.999}$$
$$= 0.0902 \approx 9\%.$$

$D \cap \neg D = \phi$

$D \cup \neg D = S$

Hence, for rare diseases, doctors will always perform a second test on receiving a positive first test. If possible, the second test uses a different principle, so as to be independent of the first test.

The Monty Hall Paradox

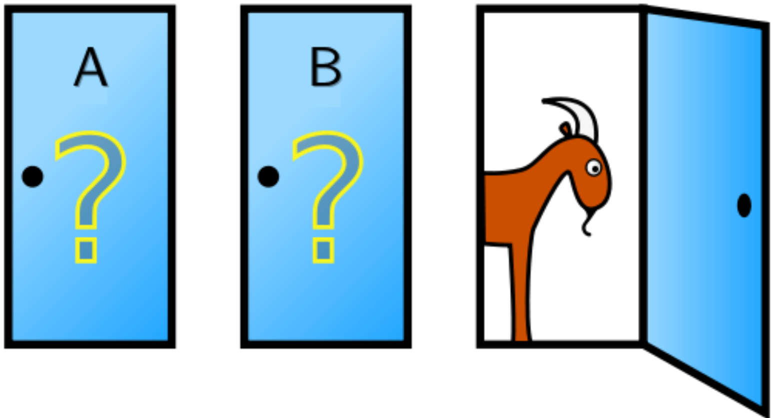
You are participating in a game show to win 10,000,000 RMB. The game master [Monty Hall] presents you with three closed doors. Behind one of the doors is the prize, behind the other two doors there is simply a goat. If you open the correct door, you will receive the money, if you open one of the other two doors you will get a goat.

Before opening any of the three doors, you can announce which door you intend to open. Obviously, at least one of the other two doors does not hide the money. The game master opens this (empty) door. You are then given the option of either

- sticking with your choice or
- switching to the other closed door.

What do you do and does it make a difference?

The Monty Hall Paradox



Question. You have chosen door A. Then door C is shown to harbor a goat. To win the money, should you switch to door B?

A) Yes

B) No

C) It doesn't matter.

The Monty Hall Paradox

To many people it seems counter-intuitive, but the best course of action is to ***change your choice to the other door***. There will be a $2/3$ probability that the prize is behind the remaining door that you have not chosen!

Why is that? By opening the door, the game master has not given you any information about the door you have chosen (he can always open one of the remaining doors, no matter which door you choose). The probability of this being the correct door was $1/3$ before he opens the other door, and it remains that way after he opens the door.

However, his opening a door ***does*** give you information on the other two doors, namely, it tells you which of the other two doors does definitely ***not*** hide the prize. The original $2/3$ probability that one of these doors hides the money is now concentrated on just the one door. Therefore, it is advantageous for you to change your choice.

The Monty Hall Paradox

We can use Bayes's formula to evaluate the probabilities explicitly. Suppose the doors are denoted A, B and C and denote by the same letter X the event "prize is behind door X" where $X = A, B, C$. Suppose that door A is initially selected and that the host opens door C; we denote the event "host opens door C" by C^*.

Then, by Bayes's formula,

$$P[A \mid C^*] = \frac{P[C^* \mid A] \cdot P[A]}{P[C^* \mid A] \cdot P[A] + P[C^* \mid C] \cdot P[C] + P[C^* \mid B] \cdot P[B]}$$
$$= \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2} \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3}} = \frac{1}{3}.$$

Of course, since $P[C \mid C^*] = 0$, this also implies $P[B \mid C^*] = 2/3$.