IE621: Probability and Stochastic Processes 1

K.S. Mallikarjuna Rao

Industrial Engineering & Operations Research
Indian Institute of Technology Bombay

IE 621 Probability and Stochastic Processes
August - November, 2023

What is Probability?

What is Probability?

- Classical: First introduction to Probability. Depends on equally likely principle.
- Roll a fair die, there are six outcomes and all of which are equally likely. Therefore each outcome has probability 1/6.
- Advantage is that it is conceptually simple.

What is Probability?

IEOR IIT BOMBAY

- Empirical (Frequentist): This perspective defines probability via a thought experiment and generalizes the first view.
- We are given a die. It is supposed to be not fair, but we do not know the weights. To deduce the probability, we roll the die again and again and consider the empirical averages. The probability is defined to be the limit of this average as the number of rolls grow.
- Disadvantage: Thought experiment could never be carried out in practice more than once.
- Disadvantage: It does not tell how large n has to be before we get a good approximation.

What is Probability

- Subjective: Subjective probability is an individual person’s measure of belief that an event will occur.
- The drawback of this approach is that one person’s view may be drastically different from other person’s view.
- Another drawback is that subjective probability disobeys coherence (consistency). The probabilities need not sum to 1.

What is Probability

- Axiomatic: This is a unifying perspective.
- The coherence conditions needed for subjective probability can be proved to hold for the classical and empirical definitions.
- The axiomatic perspective codifies these coherence conditions, so can be used with any of the above three perspectives.

Probability

Probability (Axiomatic Approach, A.N. Kolmogorov)

- Ω Set of possible outcomes; sample space.
- An event is a subset of Ω.
- $\mathcal{F}$ Set of all possible events satisfying certain conditions:
  - $\emptyset, \Omega \in \mathcal{F}$.
  - If $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$.
  - If $A_1, A_2, \cdots \in \mathcal{F}$, then $A_1 \cup A_2 \cup \cdots \in \mathcal{F}$
  - A collection of subsets satisfying above three conditions is called σ-field (σ-algebra).

Probability (Axiomatic Approach, A.N. Kolmogorov)

- Probability is a function $\mathbb{P} : \mathcal{F} \to \mathbb{R}$ satisfying
  - $\mathbb{P}(A) \geq 0$ for every $A \in \mathcal{F}$.
  - $\mathbb{P}(\Omega) = 1$.
  - If $A_1, A_2, \dots \in \mathcal{F}$ are pairwise disjoint, then

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n). \qquad \text{(Countable Additivity)}$$

- $(\Omega, \mathcal{F}, \mathbb{P})$ is called probability space.

Some Points

- Why σ-field?
- Why countable additivity?

Common Mistake

- If there are only two possible outcomes, and you don't know which is true, the probability of each of these outcomes is 1/2.

Tossing two coins and Equally Likely

Suppose a cup containing two similar coins is shaken, then turned upside down on a table. What is the chance that the two coins show heads? Consider the following solutions to this problem.

Tossing two coins and Equally Likely

- Either they both show heads, or they don't. These are the two possible outcomes. Assuming these are equally likely, the chance of both heads is 1/2.

Tossing two coins and Equally Likely

- Regard the number of heads showing on the coins as the outcome. There could be 0 heads, 1 head, or 2 heads. Now there are three possible outcomes. Assuming these are equally likely, the chance of both heads is 1/3.

Tossing two coins and Equally Likely

- Despite the fact that the coins are supposed to be similar, imagine that they are labeled in some way to distinguish them. Call one of the the first coin and the other the second. Now there are four outcomes which might be considered. Assume these four possible outcomes are equally likely. Then the event of both coins showing heads has a chance of 1/4.

Tossing two coins and Equally Likely

Question: Which of the solutions is correct?

All are correct as far as the formal theory is considered.

Another Problem

A fair coin is tossed twice. What's the probability to have at least one Tail?

The number of Tails possible is 0, 1, or 2. Therefore the probability is $2/3$.

Another Problem

A fair coin is tossed twice. What's the probability to have at least one Tail?

The number of Tails possible is 0, 1, or 2. Therefore the probability is $2/3$.

Is this correct?

Another Problem

A fair coin is tossed twice. What's the probability to have at least one Tail?

The number of Tails possible is 0, 1, or 2. Therefore the probability is $2/3$.

Is this correct?

The sample space $\Omega = \{HH, HT, TH, TT\}$. All are equally likely. Therefore the probability of desired event is $3/4$.

Consequences

- $\mathbb{P}(\emptyset) = 0$.
- For every $A \in \mathcal{F}$, $0 \leq \mathbb{P}(A) \leq 1$.
- Monotonicity: If $A \subseteq B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$.
  - ☞ $B = A \cup (B \setminus A)$ and hence

$$\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A) \geq \mathbb{P}(A).$$

- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.
  - ☞ $1 = \mathbb{P}(A \cup A^c) = \mathbb{P}(A) + \mathbb{P}(A^c)$.

Inclusion-Exclusion Principle

Theorem

For $A, B \in \mathcal{F}$,

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

Note that $A \cup B = A \cup (B \setminus A)$ and hence

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A).$$

Also $B = (B \setminus A) \cup (A \cap B)$ and hence

$$\mathbb{P}(B) = \mathbb{P}(B \setminus A) + \mathbb{P}(A \cap B)$$

Using the two equalities, we get the result.

Inclusion-Exclusion Principle

The above can be generalized to multiple sets.

Theorem

For any events $A_1, A_2, \cdots, A_n \in \mathcal{F}$, we have

$$\mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_i \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) + \sum_{i<j<k} \mathbb{P}(A_i \cap A_j \cap A_k) - \cdots + (-1)^{n+1}\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n).$$

We can use induction to prove this.

Inclusion-Exclusion Principle

Assume that the result is true for n and we will prove it for $n+1$.

$$\begin{aligned}
&\mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_n \cup A_{n+1}) \\
&= \mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_n) + \mathbb{P}(A_{n+1}) - \mathbb{P}((A_1 \cup A_2 \cup \cdots \cup A_n) \cap A_{n+1}) \\
&= \mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_n) + \mathbb{P}(A_{n+1}) - \mathbb{P}((A_1 \cap A_{n+1}) \cup (A_2 \cap A_{n+1}) \cup \cdots \cup (A_n \cap A_{n+1}))
\end{aligned}$$

Now expand the terms and rearrange to complete.

Inclusion-Exclusion Principle

$$
\begin{aligned}
&= \sum_{i=1}^{n} \mathbb{P}(A_i) - \sum_{1\le i_1<i_2\le n} \mathbb{P}(A_{i_1}\cap A_{i_2}) + \sum_{1\le i_1<i_2\le i_3<n} \mathbb{P}(A_{i_1}\cap A_{i_2}\cap A_{i_3}) \\
&\quad - \cdots + (-1)^{n+1}\mathbb{P}(A_1\cap A_2\cap\cdots\cap A_n) \\
&\quad + \mathbb{P}(A_{n+1}) \\
&\quad - \sum_{i=1}^{n} \mathbb{P}(A_i\cap A_{n+1}) + \sum_{1\le i_1<i_2\le n} \mathbb{P}(A_{i_1}\cap A_{i_2}\cap A_{n+1}) \\
&\quad - \sum_{1\le i_1<i_2<i_3\le n} \mathbb{P}(A_{i_1}\cap A_{i_2}\cap A_{i_3}\cap A_{n+1}) + \cdots - (-1)^{n+1}\mathbb{P}(A_1\cap A_2\cap\cdots\cap A_n\cap A_{n+1})
\end{aligned}
$$

Close inspection will complete the proof.

Boole's Inequality (Union Bound)

IEOR IIT BOMBAY

Theorem

Let $A_1, A_2, \cdots, A_n \in \mathcal{F}$. Then

$$\mathbb{P}(A) \leq \sum_{k=1}^{n} \mathbb{P}(A_k).$$

Define $B_1 = A_1$ and $B_{n+1} = A_{n+1} \setminus (A_1 \cup \cdots \cup A_n)$. Then $A = \bigcup_{k=1}^{n} B_k$ and

$$\mathbb{P}(A) = \sum_{k=1}^{n} \mathbb{P}(B_k) \leq \sum_{k=1}^{n} \mathbb{P}(A_k).$$

This proves Boole's Inequality.

A more general inequality (Boneferroni Inequality) will be proved in later

Probability of Derangements

A derangement is a permutation σ of $\{1, 2, \cdots, n\}$ such that $\sigma(i) \neq i$ for each $i \in \{1, 2, \cdots, n\}$. The probability that a random permutation is a derangement is $\sum_{k=0}^{n}(-1)^k \frac{1}{k!}$ which tends to $\frac{1}{e}$ as $n \to \infty$.

Let $A_i = \{\sigma : \sigma(i) = i\}$. The $|A_i| = (n-1)!$. Moreover, for any finite subset $I \subseteq \{1, 2, \cdots, n\}$, $\cap_{i \in I} A_i$ has exactly $(n - |I|)!$ elements. Using the inclusion-exclusion principle, we see that a random permutation has a fixed point (i.e., in $\cup_{i \in \{1,2,\cdots,n\}} A_i$) with probability $\sum_{k=1}^{n}(-1)^{k+1} \frac{1}{k!}$. Thus the probability of derangements is $\sum_{k=0}^{n}(-1)^k \frac{1}{k!}$. This converges to $\frac{1}{e}$ as $n \to \infty$.

Continuity Property

Theorem

Let $A_1 \subseteq A_2 \subseteq \cdots \in \mathcal{F}$ and $A = \bigcup_{n=1}^{\infty} A_n$. Then

$$\mathbb{P}(A) = \lim_{n \to \infty} \mathbb{P}(A_n).$$

Continuity Property

IEOR IIT BOMBAY

Theorem

Let $A_1 \subseteq A_2 \subseteq \cdots \in \mathcal{F}$ and $A = \bigcup_{n=1}^{\infty} A_n$. Then

$$\mathbb{P}(A) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

Follows from

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus A_2) \cdots) \\ &= \mathbb{P}(A_1) + \mathbb{P}(A_2 \setminus A_1) + \mathbb{P}(A_3 \setminus A_2) + \cdots \\ &= \mathbb{P}(A_1) + (\mathbb{P}(A_2) - \mathbb{P}(A_1)) + (\mathbb{P}(A_3) - \mathbb{P}(A_1)) + \cdots = \lim_{n\to\infty} \mathbb{P}(A_n)\end{aligned}$$

The last equality follows from the absolute summability of the series.

Continuity Property

$\mathbb{P} : \mathcal{F} \to \mathbb{R}$ is said to be finitely additive probability measure if it satisfies

- $\mathbb{P}(A) \geq 0$ for every $A \in \mathcal{F}$.
- $\mathbb{P}(\Omega) = 1$.
- If $A_1, A_2, \cdots, A_n \in \mathcal{F}$ are pairwise disjoint, then

$$\mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_n) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \cdots + \mathbb{P}(A_n). \quad \text{(Finite Additivity)}$$

Continuity Property

Theorem

Let $\mathbb{P}$ be finitely additive probability measure, then $\mathbb{P}$ is countably additive if and only if the continuity property holds for $\mathbb{P}$.

Continuity Property

Theorem

Let $\mathbb{P}$ be finitely additive probability measure, then $\mathbb{P}$ is countably additive if and only if the continuity property holds for $\mathbb{P}$.

We have already proved that if $\mathbb{P}$ is countably additive, then the continuity holds. To prove the other way, let $A_1, A_2, \cdots,$ be disjoint sets in $\mathcal{F}$. Now consider $B_n = A_1 \cup A_2 \cup \cdots \cup A_n$ for $n \geq 1$. Then $B_1 \subseteq B_2 \subseteq \cdots$ and $\cup A_n = \cup B_n$. Therefore

$$\mathbb{P}(\bigcup_{n=1}^{\infty} A_n) = \mathbb{P}(\bigcup_{n=1}^{\infty} B_n) = \lim_{n\to\infty} \mathbb{P}(B_n) = \lim_{n\to\infty} \sum_{m=1}^{n} \mathbb{P}(A_m) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

Conditional Probability

Definition

Conditional probability of A given that the even B has occurred is given by

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

Theorem

If $A_1, A_2, \cdots, A_n \in \mathcal{F}$ and if $\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) > 0$, then

$$\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2 \mid A_1) \cdots \mathbb{P}(A_n \mid A_1 \cap A_2 \cap \cdots \cap A_{n-1}).$$

Note that

$$\mathbb{P}(A_1 \cap \cdots \cap A_{n-1}) \geq \mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) > 0.$$

Therefore

$$\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) = \mathbb{P}(A_1 \cap \cdots \cap A_{n-1})\mathbb{P}(A_n \mid A_1 \cap \cdots \cap A_{n-1}).$$

We can now complete the proof.

Theorem (Law of Total Probability)

Let A_n, $n \geq 1$ be a partition (finite or countable) of Ω. Then if $A \in \mathcal{F}$,

$$\mathbb{P}(A) = \sum_n \mathbb{P}(A \mid A_n)\mathbb{P}(A_n).$$

Note that

$$\mathbb{P}(A) = \mathbb{P}\Big(\bigcup_{n=1}^{\infty}(A \cap A_n)\Big) = \sum_{n=1}^{\infty} \mathbb{P}(A \cap A_n) = \sum_{n=1}^{\infty} \mathbb{P}(A \mid A_n)\mathbb{P}(A_n).$$

Bayes' Theorem

Let A_n, $n \geq 1$ be a partition (finite or countable) of Ω and suppose $\mathbb{P}(A) > 0$. Then

$$\mathbb{P}(A_n \mid A) = \frac{\mathbb{P}(A \mid A_n)\mathbb{P}(A_n)}{\sum_m \mathbb{P}(A \mid A_m)\mathbb{P}(A_m)}.$$

Using the multiplication rule, we have

$$\mathbb{P}(B \mid A)\mathbb{P}(A) = \mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\mathbb{P}(B)$$

Therefore

$$\mathbb{P}(B \mid A) = \frac{\mathbb{P}(A \mid B)\mathbb{P}(B)}{\mathbb{P}(A)}$$

Take $B = A_n$ and use the law of total probability.

Independence

Definition

Two events A and B are said to be independent if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

Consequences

Theorem

If A and B are independent, so also are A and B^c, A^c and B , and A^c and B^c.

Consequences

Theorem

A and B are independent if and only if $\mathbb{P}(A \mid B) = \mathbb{P}(A)$ (provided ....).

Consequences

Theorem

$A \mapsto \mathbb{P}(A \mid B)$ defines a new probability measure on $\mathcal{F}$, called the conditional probability measure given B (provided ...).

Some Examples

Boy or Girl paradox / Two Child Problem

IEOR IIT BOMBAY

Example

A family has two children, Jeet and Kiran. Kiran is a girl. What is the probability that both children are girls?

Compare this with the following:

Example

A family has two children, Jeet and Kiran. At least one of them is a girl. What is the probability that both children are girls?

Scandal of Arithmetic (The Paradox of the Chevalier De Méré)

Example

Which is more likely, obtaining at least one six in 4 tosses of a fair die (event A), or obtaining at least one double six in 24 tosses of a pair of dice (event B)?

- $\mathbb{P}(A) = \frac{6^4 - 5^4}{6^4} \approx 0.518$
- $\mathbb{P}(B) = \frac{36^{24} - 35^{24}}{36^{24}} \approx 0.491$
- Why is this paradox?

Problem of Points / Division of Stakes

- Two teams compete in a game of skill.
- They play until one wins six rounds.
- The winner will take the entire prize.
- How the should the prize be divided if the game is interrupted with the score 5 to 2?
- Luca Pacioli considered this problem in 1494 textbook **Summa de arithmetica, geometrica, proportioni et proportionalità**.
- Pacioli's solution is to divide the stakes in the ration 5:2.
- Mid 16th century, Niccolò Tartaglia noticed counterintuitive results (e.g., when the game is stopped after first round).
- Girolamo Cardano has provided one solution, which was also not adequate.

Problem of Points / Division of Stakes

- The problem was solved by Pierre de Fermat and Blaise Pascal in a series of letters.
- They never met in person.
- This communication gave birth to the Probability.
- This problem was brought to the attention of Pascal and Fermat around 1654 by the Chevalier de Méré, a famous gambler and noble man in Paris.

Problem of Points / Division of Stakes

- Two players play a game of chance.
- Each player puts up equal stakes.
- The first player who wins a certain number of rounds will collect the entire stake.
- Suppose the game is interrupted before either player has won.
- How do the players divide the stake?

Birthday Paradox

- What is the probability that, in a group of N people, two of them share the same birthday
- Consider N balls numbered 1 to N in an urn. We want to compute the probability that the numbers of the n balls picked with replacing are different.
- After two trials, the probability that the two numbers are different is $1 - \frac{1}{N}$.

Birthday Paradox

- After three trials, given that first two outcomes are different, the conditional probability is $\frac{2}{N}$ that the third outcome will be equal to one of the two outcomes. Thus the conditional probability is $1 - \frac{2}{N}$ that the third will differ from first two. Thus

$$P(\text{first three trials different}) = (1 - \frac{1}{N})(1 - \frac{2}{N}).$$

- Thus

$$P(\text{first } n \text{ trials different}) = (1 - \frac{1}{N})(1 - \frac{2}{N}) \cdots (1 - \frac{n}{N}).$$

Birthday Paradox

- Thus
$$P(\text{first } n \text{ trials different}) = (1 - \frac{1}{N})(1 - \frac{2}{N}) \cdots (1 - \frac{n-1}{N}).$$
- Taking logarithm,
$$\log(\text{probability}) = \log(1 - \frac{1}{N}) + \log(1 - \frac{2}{N}) + \cdots + \log(1 - \frac{n-1}{N})$$
- For $0 < x < 1$,
$$\log(1 - x) \leq -x.$$

Birthday Paradox

- Hence

$$\log\left(1-\frac{1}{N}\right)+\log\left(1-\frac{2}{N}\right)+\cdots+\log\left(1-\frac{n-1}{N}\right)\leq-\left(\frac{1}{N}+\frac{2}{N}+\cdots+\frac{n-1}{N}\right)$$

- Thus

$$\log P(n\text{ trials are different})\leq\frac{1}{2}\frac{n(n-1)}{N}$$

- This gives

$$P(n\text{ trials are different})\leq\exp\left(-\frac{n(n-1)}{2N}\right)$$

Birthday Paradox

- With 23 people, we can see that the probability will of two people having the same birthday will be more that 1/2. And with 42 people, it will be more than 0.9.
- If n is more than $\sqrt{N}$, the probability will be more than 1/2.

Monty Hall Problem

Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car, behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat.

He then says to you, "Do you want to pick door No. 2?"

Is it to your advantage to switch your choice ?

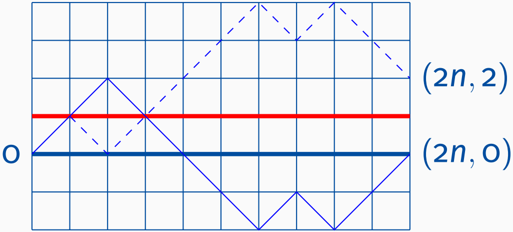
Change Problem

There are $2n$ people in a waiting line at a box office. n of them have ₹10 coins while the others have ₹5 coins. A ticket costs ₹5. Every one in the line needs exactly one ticket. The cashbox is empty initially. What is the probability that no person will wait for change?

Change Problem

- Assume ticket buyers are located at points $1, 2, \cdots, 2n$ on the horizontal axis.
- The empty cashbox is at origin.
- if a person has ₹10, the cash line goes one step up, otherwise it goes one step down.
- At both ends, the line has zero y-coordinate.
- Favourable lines are those which do not go above x-axis.

Change Problem



- Let us count the number of lines crossing or touching the line $y = 1$.
- These are the only lines which are favourable to the opposite event, when someone will wait for the change.

Change Problem

- For each of these lines draw a dummy line. It coincides with the original line till the first hit of the line $y = 1$, then it mirrors the original line.
- Any dummy line then starts at the origin and ends at the point $(2n, 2)$.
- It consists of $n + 1$ steps up and $n - 1$ steps down.
- Thus there are $\binom{2n}{n-1}$ dummy lines.
- Finally there are $\binom{2n}{n} - \binom{2n}{n-1}$ lines favourable to our event, and the probability is

$$p = \frac{\binom{2n}{n} - \binom{2n}{n-1}}{\binom{2n}{n}} = \frac{1}{n+1}.$$

Urn Problems

There are a blue balls and b red balls in an urn. Two balls are taken one by one. What's the probability that the first ball is red and the second ball is blue?

Let A be the event that the first ball is red and let B be the event that the second ball is blue.

$$\mathbb{P}(A) = \frac{b}{a+b}.$$

Now

$$\mathbb{P}(B|A) = \frac{a}{a+b-1}.$$

Finally

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B|A) = \frac{ab}{(a+b)(a+b-1)}.$$

Another way to solve this:

Let the balls be numbered through 1 to a be blue while the others are red. Then

$$\Omega = \{(1,2),(1,3),\cdots,(a+b,a+b-1)\}$$

where (x,y) denotes the outcome where x denotes the number on the first ball and y is number on the second ball. Therefore

$$A \cap B = \{(a+1,1),(a+1,2),\cdots,(a+b,a)\}$$

Note that $|\Omega| = (a+b)(a+b-1)$ and $|A \cap B| = ab$, solving the problem.

Surprises with Conditional Probability

Consider a rare disease X that affects one in a million people. A medical test is used to test for the presence of the disease. The test is 99% accurate in the sense that if a person has no disease, the chance that the test shows positive is 1% and if the person has disease, the chance that the test shows negative is also 1%.

Suppose a person is tested for the disease and the test result is positive. What is the chance that the person has the disease X?

Surprises with Conditional Probability

Let A be the event that the person has the disease X.

Let B be the event that the test shows positive.

Therefore $\mathbb{P}(A) = 10^{-6}$, $\mathbb{P}(B \mid A) = 0.99$ and $\mathbb{P}(B \mid A^c) = 0.01$.

Need to find $\mathbb{P}(A \mid B)$.

Baye's rule implies

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\mathbb{P}(B \mid A)\mathbb{P}(A) + \mathbb{P}(B \mid A^c)\mathbb{P}(A^c)} = 0.000099$$

Surprises with Conditional Probability

The test is quite an accurate one, but the person tested positive has a really low chance of actually having the disease! Of course, one should observe that the chance of having disease is now approximately 10^{-4} which is considerably higher than 10^{-6}.

A calculation-free understanding of this surprising looking phenomenon can be achieved as follows: Let everyone in the population undergo the test. If there are 10^9 people in the population, then there are only 10^3 people with the disease. The number of true positives is approximately $10^3 \times 0.99 \approx 10^3$ while the number of false positives is $(10^9 - 10^3) \times 0.01 \approx 10^7$. In other words, among all positives, the false positives are way more numerous than true positives.

Surprises with Conditional Probability

The surprise here comes from not taking into account the relative sizes of the sub-populations with and without the disease. Here is another manifestation of exactly the same fallacious reasoning.

A person X is introverted, very systematic in thinking and somewhat absent-minded. You are told that he is a doctor or a mathematician. What would be your guess - doctor or mathematician?

A common answer is “mathematician”. Even accepting the stereotype that a mathematician is more likely to have all these qualities than a doctor, this answer ignores the fact that there are perhaps a hundred times more doctors in the world than mathematicians! In fact, the situation is identical to the one in the example above, and the mistake is in confusing $\mathbb{P}(A \mid B)$ and $\mathbb{P}(B \mid A)$.

Random Variables

Discrete Random Variables

- $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.
- We assume that Ω is discrete (finite set or infinite and countable set).
- Since Ω is discrete, $\mathcal{F}$ can be the entire powerset of Ω i.e., 2^{Ω}.
- $X : \Omega \to \mathbb{R}$ is a discrete random variable provided $\{\omega \in \Omega : X(\omega) = x\} \in \mathcal{F}$ for each $x \in \mathbb{R}$.
- Since Ω is a discrete set, X takes only discretely many values.
- The smallest σ-field $\mathcal{F}_X \subseteq \mathcal{F}$ containing the events $\{\omega \in \Omega : X(\omega) = x\}$ for each $x \in \mathbb{R}$ is called the σ-field generated by X (or information space given by X).