Mathematical Statistics and Data Analysis: Lection 3. Conditional Probability

Aleksei Zhukov

Data Scientist, Research Fellow

BI BRICS,
2020

Outline

1. Intro
2. Conditional probability
   - Conditional probability by example
   - Conditional probability definition
   - Multiplication Law
   - Law of total probability
   - Bayes Rule
   - Bayesian Approach
   - Independent events
   - Conclusions

Theoretical programme

Probability theory

- **Introduction to Probablilty theory** $\leftarrow$ We are still here
- Random variables
- Joint distributions
- Expected Value of a Random Variable
- Limit theorems
- Distributions derived from Normal

Intro

Tversky and Kahneman (1974) presented their subjects with the following question:
"If Linda is a 31-year-old single woman who is outspoken on social issues such as disarmament and equal rights, which of the following statements is more likely to be true?

A) Linda is bank teller.

B) Linda is a bank teller and active in the feminist movement."

Digitalis toxicity example

$T+$ = high blood concentration (positive test)

$T-$ = low blood concentration (negative test)

$D+$ = toxicity (disease present)

$D-$ = no toxicity (disease absent)

	$D+$	$D-$	*Total*
$T+$	25	14	39
$T-$	18	78	96
Total	43	92	135

Conditional probability

DEFINITION

Let A and B be two events with $P(B) \neq 0$. The conditional probability of A given B is defined to be

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

Multiplication Law

In some cases, $P(A|B)$ and $P(B)$ can be found easier then $P(A \cap B)$, and we can use another variant of the same equation.

MULTIPLICATION LAW

Let A and B be events and assume $P(B) \neq 0$. Then

$$P(A \cap B) = P(A \mid B)P(B) \qquad \blacksquare$$

Occupation mobility example

An urn contains three red balls and one blue ball. Two balls are selected without replacement. What is the probability that they are both red?

Let R_1 and R_2 denote the events that a red ball is drawn on the first trial and on the second trial, respectively. From the multiplication law,

$$P(R_1 \cap R_2) = P(R_1)P(R_2 \mid R_1)$$

$P(R_1)$ is clearly $\frac{3}{4}$, and if a red ball has been removed on the first trial, there are two red balls and one blue ball left. Therefore, $P(R_2 \mid R_1) = \frac{2}{3}$. Thus, $P(R_1 \cap R_2) = \frac{1}{2}$.

Conditional probability example

Suppose that if it is cloudy (B), the probability that it is raining (A) is .3, and that the probability that it is cloudy is $P(B) = .2$ The probability that it is cloudy and raining is

$$P(A \cap B) = P(A \mid B)P(B) = .3 \times .2 = .06$$

Law of total probability

LAW OF TOTAL PROBABILITY

Let $B_1, B_2, \ldots, B_n$ be such that $\bigcup_{i=1}^n B_i = \Omega$ and $B_i \cap B_j = \emptyset$ for $i \neq j$, with $P(B_i) > 0$ for all i. Then, for any event A,

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

Proof

$$\begin{aligned} P(A) &= P(A \cap \Omega) \\ &= P\left(A \cap \left(\bigcup_{i=1}^{n} B_i\right)\right) \\ &= P\left(\bigcup_{i=1}^{n} (A \cap B_i)\right) \end{aligned}$$

Since the events $A \cap B_i$ are disjoint,

$$\begin{aligned} P\left(\bigcup_{i=1}^{n} (A \cap B_i)\right) &= \sum_{i=1}^{n} P(A \cap B_i) \\ &= \sum_{i=1}^{n} P(A \mid B_i) P(B_i) \end{aligned}$$

Law of total probability

An urn contains three red balls and one blue ball. Two balls are selected, what is the probability that a red ball is selected on the second draw?

The answer may or may not be intuitively obvious—that depends on your intuition. On the one hand, you could argue that it is "clear from symmetry" that $P(R_2) = P(R_1) = \frac{3}{4}$. On the other hand, you could say that it is obvious that a red ball is likely to be selected on the first draw, leaving fewer red balls for the second draw, so that $P(R_2) < P(R_1)$. The answer can be derived easily by using the law of total probability:

$$P(R_2) = P(R_2 \mid R_1)P(R_1) + P(R_2 \mid B_1)P(B_1)$$
$$= \frac{2}{3} \times \frac{3}{4} + 1 \times \frac{1}{4} = \frac{3}{4}$$

where B_1 denotes the event that a blue ball is drawn on the first trial.

Law of total probability

Suppose that occupations are grouped into upper (U), middle (M), and lower (L) levels. U_1 will denote the event that a father's occupation is upper-level; U_2 will denote the event that a son's occupation is upper-level, etc. (The subscripts index generations.) Glass and Hall (1954) compiled the following statistics on occupational mobility in England and Wales:

	U_2	M_2	L_2
U_1	.45	.48	.07
M_1	.05	.70	.25
L_1	.01	.50	.49

Bayes Rule

BAYES' RULE

Let A and $B_1, \ldots, B_n$ be events where the B_i are disjoint, $\bigcup_{i=1}^{n} B_i = \Omega$, and $P(B_i) > 0$ for all i. Then

$$P(B_j \mid A) = \frac{P(A \mid B_j)P(B_j)}{\sum_{i=1}^{n} P(A \mid B_i)P(B_i)}$$

Polygraph test example

Polygraph tests (lie-detector tests) are often routinely administered to employees or prospective employees in sensitive positions.

$+$ or $-$ denote polygraph results ($+$ means lie)
L/T is the event that subject is lying/telling the truth

$$P(+|L) = 0.88$$
$$P(-|T) = 0.86$$

Suppose only 1% of the employees are lying

$$P(L) = 0.01$$

$$P(T) = 0.99$$

Bayesian Approach

Bayes' rule is the fundamental mathematical ingredient of a subjective, or "Bayesian," approach to epistemology, theories of evidence, and theories of learning. According to this point of view, an individual's beliefs about the world can be coded in probabilities. According to Bayesian theory, our beliefs are modified as we are confronted with evidence.

Frequentist vs Bayesian Approach

There are two approach to the probability: frequentist approach and the Bayesian approach. Lets look at coin toss problem from two poinys of view.

Frequentist approach: the statement means that if the experiment were repeated many times, the long-run average number of heads would tend to 0.5.

Bayesian approach: the statement is a quantification of the speaker's uncertainty about the outcome of the experiment and thus is a personal or subjective notion; the probability that the coin will land heads up may be different for different speakers, depending on their experience and knowledge of the situation.

Independent events

Intuitively, we would say that two events, A and B, are independent if knowing that one had occurred gave us no information about whether the other had occurred; that is, $P(A \mid B) = P(A)$ and $P(B \mid A) = P(B)$. Now, if

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

then

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

We will use this last relation as the definition of independence. Note that it is symmetric in A and in B, and does not require the existence of a conditional probability, that is, $P(B)$ can be 0.

DEFINITION

A and B are said to be independent events if $P(A \cap B) = P(A)P(B)$. ■

Independent events examples

A system is designed so that it fails only if a unit and a backup unit both fail. Assuming that these failures are independent and that each unit fails with probability p, the system fails with probability p^2. If, for example, the probability that any unit fails during a given year is .1, then the probability that the system fails is .01, which represents a considerable improvement in reliability.

A fair coin is tossed twice. Let A denote the event of heads on the first toss, B the event of heads on the second toss, and C the event that exactly one head is thrown. A and B are clearly independent, and $P(A) = P(B) = P(C) = .5$. To see that A and C are independent, we observe that $P(C \mid A) = .5$. But

$$P(A \cap B \cap C) = 0 \neq P(A)P(B)P(C)$$

Todays topics

1. Intro
2. Conditional probability
   - Conditional probability by example
   - Conditional probability definition
   - Multiplication Law
   - Law of total probability
   - Bayes Rule
   - Bayesian Approach
   - Independent events
   - Conclusions

Questions

Time for your questions!

zhukovalex13@gmail.com