Mathematical Statistics and Data Analysis

Aleksei Zhukov

Data Scientist, Research Fellow

BI BRICS,
2020

Outline

1. About course
   - Organization
   - Outline
   - Literature
   - Tips and recommendations
2. Probability theory introduction
   - Sample space
   - Probability measures and their properties
   - Calculating probabilities
     - Multiplication principle
     - Permutations and Combinations
   - Conclusions

Zhukov Aleksei
zhukovalex13@gmail.com

Aleksei Zhukov

Research Fellow

Professional interests

data science, machine learning, power systems, software design

Google Scholar:
https://scholar.google.ru/citations?user=Vc7NQuUAAAAJ

Course Organization

- 2 academic hours for lections and 2 for a practice (a week)
- Every lesson structure is 40 + 10 (timeout) + 40 minutes
- Chat for organization questions and announces
- Chat for discussion
- 2 big tests on theory and 2 on practical skills

Theoretical programme

Probability theory

- Introduction to Probablilty theory
- Random variables
- Joint distributions
- Expected Value of a Random Variable
- Limit theorems
- Distributions derived from Normal

Mathematical Statistics theory

- Survey Sampling
- Estimation of Parameters and Fitting of Probability Distributions
- Testing Hypotheses
- Comparing Two Samples
- Variance Analysis, Categorical Data Analysis, Regression Analysis

Literature

- Wackerly D. at al. Mathematical Statistics with Applications
- Rice J.A. Mathematical Statistics and Data Analysis
- Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning

Tips and recommendations

- Write your lecture notes by hand
- Examples are important
- Discuss questions with your classmates

Probability

Probability is the branch of mathematics concerning numerical descriptions of how likely an event is to occur or how likely it is that a proposition is true.

- Example 1.
- Example 2.

Probability and mathematical statistics

Probability and statistics are related areas of mathematics which concern themselves with analyzing the relative frequency of events. Still, there are fundamental differences in the way they see the world:

- Probability deals with predicting the likelihood of future events, while statistics involves the analysis of the frequency of past events.
- Probability is primarily a theoretical branch of mathematics, which studies the consequences of mathematical definitions. Statistics is primarily an applied branch of mathematics, which tries to make sense of observations in the real world.

Sample space: definition

Let the set of all possible outcomes is the **sample space** corresponding to an experiment. The sample space is denoted by Ω, and an element of Ω is denoted by ω. Where ω is a **simple event** such an event that cannot be decomposed. Each simple event corresponds to one and only one sample point.

Example 1. Students have 3 tests. Each test can be **p**asses or **f**ailed. The sample space is the set of all possible outcomes:

$$\Omega = \{ppp, ppf, pfp, pff, fpp, fpf, ffp, fff\}$$

Example 2. Number of rainy days per year $\Omega = \{1, 2, 3, \ldots, N\}$
Example 3. Time between some events $\Omega = \{t \mid t \geq 0\}$, for example lightnings or eathquakes.

Sample space: events and the set theory

Events is the particular subsets of Ω.
Remember Example 1.

$$\Omega = \{ppp, ppf, pfp, pff, fpp, fpf, ffp, fff\}$$

Event that first test was passed can be denoted as

$$A = \{ppp, ppf, pff, pfp\}$$

Event that third test was passed can be denoted as

$$B = \{ppp, pfp, pfp, ffp\}$$

Set operations

Union

The union of two events, $C = A \cup B$, is the event that A or B occur.

$$C = \{ppp, pfp, ppf, pff, pfp, ffp\}$$

Intersection

The intersection of two events, $C = A \cap B$, is the event that both A and B occur.

$$C = \{ppp, pfp\}$$

Complement

The complement of two events, A^C, is the event that A does not occur.

$$C = \{ppp, pfp\}$$

Set operations

Empty set

The empty set is the set with no elements (it is the event with no outcomes).
Identity :

$$A \cup \varnothing = A \qquad A \cup A^C = \Omega$$
$$A \cap \Omega = A \qquad A \cap A^C = \varnothing$$

Disjoint sets

Let $C = \{ppp, pfp\}$ and $D = \{fff, fpf\}$, is the event that both A and B occur. C and D have no outcomes in common, so these events are **disjoint** and we can write

$$A \cap C = \emptyset$$

Commutative Laws:

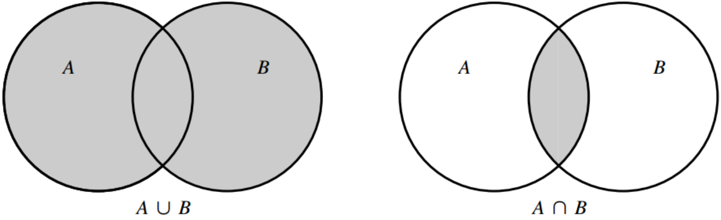
$$A \cup B = B \cup A$$
$$A \cap B = B \cap A$$

Associative Laws:

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

Distributive Laws:

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



Probability measure

A **probability measure** on Ω is a function P from subsets of Ω to the real numbers that satisfies the following axioms called **The Kolmogorov axioms**.

1. $P(\Omega) = 1.$
2. If $A \in \Omega$, then $P(A) \geq 0.$
3. If A_1 and A_2 are disjoint, then

$$P(A1 \cup A2) = P(A1) + P(A2).$$

Or more generally, if $A_1, A_2, ..., A_n, \ldots$ are mutually disjoint, then

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

Probability measure: properties A and B

Property A. $P(A^c) = 1 - P(A)$. This property follows since A and A^c are disjoint with $A \cup A^c = \Omega$ and thus, by the first and third axioms, $P(A) + P(A^c) = 1$.
Property B.$P(\emptyset) = 0$. This property follows from Property A since $\emptyset = \Omega^c$.

Probability measure: property C

Property C. If $A \in B$, then $P(A) \leq P(B)$. This property states that if B occurs whenever A occurs, then $P(A) \leq P(B)$. For example, if whenever it rains (A) it is cloudy (B), then the probability that it rains is less than or equal to the probability that it is cloudy.

Formally, it can be proved as follows: B can be expressed as the union of two disjoint sets:

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

Then, from the third axiom,

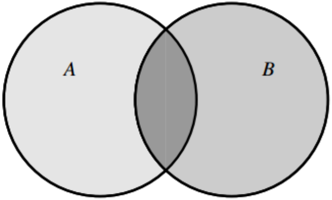
$$P(B) = P(A) + P(B \cap A^c)$$

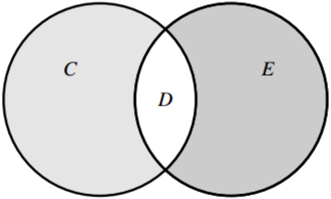
and thus

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

Property D Addition Law $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. This property is easy to see from the Venn diagram in Figure If $P(A)$ and $P(B)$ are added together, $P(A \cap B)$ is counted twice. To prove it, we decompose $A \cup B$ into three disjoint subsets, as shown in Figure :

$$C = A \cap B^c$$
$$D = A \cap B$$
$$E = A^c \cap B$$





We then have, from the third axiom,

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

Also, $A = C \cup D$, and C and D are disjoint; so $P(A) = P(C) + P(D)$. Similarly, $P(B) = P(D) + P(E)$. Putting these results together, we see that

$$\begin{aligned} P(A) + P(B) &= P(C) + P(E) + 2P(D) \\ &= P(A \cup B) + P(D) \end{aligned}$$

or

$$P(A \cup B) = P(A) + P(B) - P(D)$$

Probability measure: example

Suppose that a fair coin is thrown twice. Let A denote the event of heads on the first toss, and let B denote the event of heads on the second toss. The sample space is

$$\Omega = \{hh, ht, th, tt\}$$

We assume that each elementary outcome in Ω is equally likely and has probability $\frac{1}{4}$. $C = A \cup B$ is the event that heads comes up on the first toss or on the second toss.

What is $P(C)$?

Probability measure: example

Suppose that a fair coin is thrown twice. Let A denote the event of heads on the first toss, and let B denote the event of heads on the second toss. The sample space is

$$\Omega = \{hh, ht, th, tt\}$$

We assume that each elementary outcome in Ω is equally likely and has probability $\frac{1}{4}$. $C = A \cup B$ is the event that heads comes up on the first toss or on the second toss.

Clearly, $P(C) \neq P(A) + P(B) = 1$. Rather, since $A \cap B$ is the event that heads comes up on the first toss and on the second toss,

$$P(C) = P(A) + P(B) - P(A \cap B) = .5 + .5 - .25 = .75$$

Calculating probabilities

Probabilities are especially easy to compute for finite sample spaces. Suppose that $\Omega = \omega_1, \omega_2, ..., \omega_N$ and that $P(\omega_i) = p_i$. To find the probability of an event A, we simply add the probabilities of the ω_i that constitute A. Or if all the p_i are the same

$$P(A) = \frac{number of ways A can occur}{total number of outcomes}$$

Calculating probabilities: form formula to reality

The sample-point counting method has the following steps to find the probability of an event:

1. Define the experiment and clearly determine how to describe one simple event.
2. List the simple events associated with the experiment and test each to make certain that it cannot be decomposed. This defines the sample space Ω.
3. Assign reasonable probabilities to the sample points in Ω, making certain that $P(\omega_i) \geq 0$ and $\sum P(\omega_i) = 1$.
4. Define the event of interest, A, as a specific collection of sample points. (A sample point is in A if A occurs when the sample point occurs. Test all sample points in Ω to identify those in A.)
5. Find $P(A)$ by summing the probabilities of the sample points in A.

Multiplication principle

MULTIPLICATION PRINCIPLE

If one experiment has m outcomes and another experiment has n outcomes, then there are mn possible outcomes for the two experiments.

Proof

Denote the outcomes of the first experiment by $a_1, \ldots, a_m$ and the outcomes of the second experiment by $b_1, \ldots, b_n$. The outcomes for the two experiments are the ordered pairs (a_i, b_j). These pairs can be exhibited as the entries of an $m \times n$ rectangular array, in which the pair (a_i, b_j) is in the ith row and the jth column. There are mn entries in this array. ■

Extended multiplication principle

EXTENDED MULTIPLICATION PRINCIPLE

If there are p experiments and the first has n_1 possible outcomes, the second $n_2, \ldots,$ and the pth n_p possible outcomes, then there are a total of $n_1 \times n_2 \times \cdots \times n_p$ possible outcomes for the p experiments.

Proof

This principle can be proved from the multiplication principle by induction. We saw that it is true for $p = 2$. Assume that it is true for $p = q$—that is, that there are $n_1 \times n_2 \times \cdots \times n_q$ possible outcomes for the first q experiments. To complete the proof by induction, we must show that it follows that the property holds for $p = q + 1$. We apply the multiplication principle, regarding the first q experiments as a single experiment with $n_1 \times \cdots \times n_q$ outcomes, and conclude that there are $(n_1 \times \cdots \times n_q) \times n_{q+1}$ outcomes for the $q + 1$ experiments. ■

Permutations

A **permutation** is an ordered arrangement of objects. Suppose that from the set C = c1, c2, . . . , cn we choose **r** elements and list them in order. How many ways can we do this? The answer depends on whether we are allowed to duplicate items in the list. If no duplication is allowed, we are **sampling without replacement**. If duplication is allowed, we are **sampling with replacement**.

Permutations

PROPOSITION A

For a set of size n and a sample of size r, there are n^r different ordered samples with replacement and $n(n-1)(n-2)\cdots(n-r+1)$ different ordered samples without replacement. ■

COROLLARY A

The number of orderings of n elements is $n(n-1)(n-2)\cdots 1 = n!$. ■

Permutations

PROPOSITION **B**

The number of unordered samples of r objects selected from n objects without replacement is $\binom{n}{r}$.

The numbers $\binom{n}{k}$, called the **binomial coefficients,** occur in the expansion

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

In particular,

$$2^n = \sum_{k=0}^{n} \binom{n}{k}$$

This latter result can be interpreted as the number of subsets of a set of n objects. We just add the number of subsets of size 0 (with the usual convention that $0! = 1$), and the number of subsets of size 1, and the number of subsets of size 2, etc. ■

Permutations

PROPOSITION **C**

The number of ways that n objects can be grouped into r classes with n_i in the ith class, $i = 1, \ldots, r$, and $\sum_{i=1}^{r} n_i = n$ is

$$\binom{n}{n_1 n_2 \cdots n_r} = \frac{n!}{n_1! n_2! \cdots n_r!}$$

Proof

This can be seen by using Proposition B and the multiplication principle. (Note that Proposition B is the special case for which $r = 2$.) There are $\binom{n}{n_1}$ ways to choose the objects for the first class. Having done that, there are $\binom{n-n_1}{n_2}$ ways of choosing the objects for the second class. Continuing in this manner, there are

$$\frac{n!}{n_1!(n-n_1)!} \frac{(n-n_1)!}{(n-n_1-n_2)!n_2!} \cdots \frac{(n-n_1-n_2-\cdots-n_{r-1})!}{0!n_r!}$$

choices in all. After cancellation, this yields the desired result. ■

Todays topics

1. About course
   - Organization
   - Outline
   - Literature
   - Tips and recommendations
2. Probability theory introduction
   - Sample space
   - Probability measures and their properties
   - Calculating probabilities
     - Multiplication principle
     - Permutations and Combinations
   - Conclusions

Questions

Time for your questions!

zhukovalex13@gmail.com