Mathematical Statistics and Data Analysis

Aleksei Zhukov

Data Scientist, Research Fellow

BI BRICS,
2020

Outline

1. Intro
   - Outline
   - Whatapp groups
2. Calculating probability
   - Multiplication principle
   - Permutations
   - Combinations
   - 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

Chats

- **Slides, homework, announcements**
  https://chat.whatsapp.com/C16VdVVWjktHqfA4QfyeNk
- **Discuss your questions regarding the course here**
  https://chat.whatsapp.com/LLNkbsCZeblGDt7vMeFL77

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 $C = c_1, c_2, ..., c_n$ 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. Examples

Examples 1. We have 5 lections that should be added into timetable one after another. How many ways can we do this?

Examples 2. We have 10 lections. But only 5 should be added into timetable one after another. How many ways can we do this?

Permutations. Examples

Examples 3. Driver license plate have 3 english letters followed by three numbers. How many combinations are possible?

Examples 4. What is the probability that the license plate for a new car will contain no duplicated letters or numbers

Permutations. Examples

Birthday problem

Suppose that a room contains n people. What is the probability that at least two of them have a common birthday?

Birthday problem

n	$P(A)$
4	.016
16	.284
23	.507
32	.753
40	.891
56	.988

Permutations. Examples

How many people must you ask to have a 50 / 50 chance of finding someone who shares your birthday?

Permutations. Examples

How many people must you ask to have a 50 / 50 chance of finding someone who shares your birthday?

$$P(A^c) = \frac{364^n}{365^n}$$

$$P(A) = 1 - \frac{364^n}{365^n}$$

From permutations to combinations

If r objects are taken from a set of n objects without replacement and disregarding order, how many different samples are possible?

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,

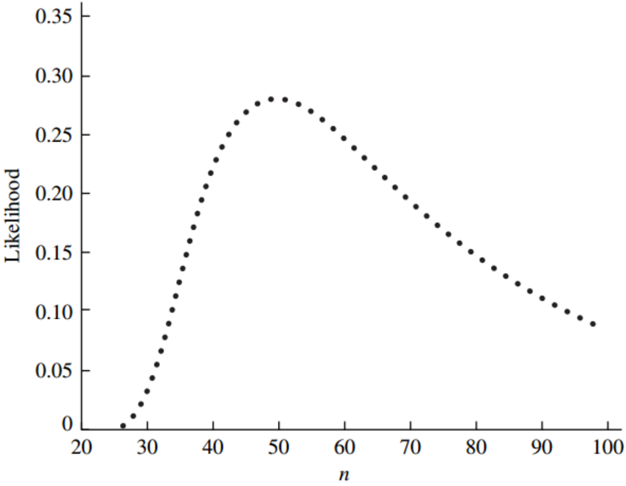
$$2^n = \sum_{k=0}^{n} \binom{n}{k}$$

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. ■

Capture/recapture method

The so-called capture/recapture method is sometimes used to estimate the size of a wildlife population. Suppose that 10 animals are captured, tagged, and released. On a later occasion, 20 animals are captured, and it is found that 4 of them are tagged. How large is the population?

Likelihood function



Combinations

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. ■

Combinations

A committee of seven members is to be divided into three subcommittees of size three, two, and two. This can be done in

$$\binom{7}{3\,2\,2} = \frac{7!}{3!2!2!} = 210$$

ways.

Combinations

In how many ways can the set of nucleotides $\{A, A, G, G, G, G, C, C, C\}$ be arranged in a sequence of nine letters?

Combinations

In how many ways can $n = 2m$ people be paired and assigned to m courts for the first round of a tennis tournament?

Combinations

The numbers $\binom{n}{n_1 n_2 \cdots n_r}$ are called **multinomial coefficients.** They occur in the expansion

$$(x_1 + x_2 + \cdots + x_r)^n = \sum \binom{n}{n_1 n_2 \cdots n_r} x_1^{n_1} x_2^{n_2} \cdots x_r^{n_r}$$

where the sum is over all nonnegative integers $n_1, n_2, \ldots, n_r$ such that $n_1 + n_2 + \cdots + n_r = n$.

Todays topics

1. Intro
   - Outline
   - Whatapp groups
2. Calculating probability
   - Multiplication principle
   - Permutations
   - Combinations
   - Conclusions

Questions

http://tovarish.ml/etc/StatisticsCourse/PracticalLesson1.zip

zhukovalex13@gmail.com

Questions

Time for your questions!

zhukovalex13@gmail.com