Tianjin University of Commerce
Department of Statistics

Submitted by
Shikun Liu
Student Number 20183744

liushikun18@gmail.com

Bachelor Thesis

When mathematics meet statistics

Tianjin, 15th May 2019

When mathematics meet statistics

by

Shikun Liu

in partial fulfillment of the requirements for the degree of

Bachelor of Science

at Tianjin University of Commerce,
to be defended publicly on Friday July 1, 2022 at 10:00 AM.

Supervisor:	Prof. dr. ir. Z.Y. Lu	
Thesis committee:	Prof. dr. ir. Z.Y. Lu,	Department head
	Prof. dr. ir. J.Y. An,	Associate dean
	A.P. dr. ir. L. X,	Lecturer

This thesis is confidential and cannot be made public until December 31, 2022.

An electronic version of this thesis is available at http://gd.tjcu.edu.cn.

Preface

There is a saying: in the ultimate analysis, all knowledge is history; in the abstract sense, all science is mathematics; on the basis of reason, all judgments are statistics.

We all know that statistics originally belonged to mathematics. But a few years ago, statistics were separated and became the first-level disciplines like mathematics, belonging to the same category of science.

The relationship between mathematics and statistics is like the relationship between physics and mechanics, which is related but different.

We can look at these three questions:

Is Bayes a mathematician?

Is mathematics natural science?

Is human being an animal?

From here, it can be seen that studying statistics is not a thing that seems to be meaningless.

In recent years, there have been many opinions about whether statistics are mathematical or not. Therefore, this paper uses the 20th question of the 2018 college entrance examination to analyze the relationship between mathematics and statistics.

Shikun Liu
Tianjin, May 2019

Contents

1

The Introduction of Mathematics and Statistics

1.1. Domestic situation:

Statistics in the 2011 edition of the Ministry of Education's "Disciplinary and Personnel Training Discipline Catalogue" became the first-level subject for the first time, and can be taught in both science and economics. But in the past, statistics was a secondary subject under mathematics and applied economics. (available from https://yz.chsi.com.cn/kyzx/other/201104/20110411/185103201.html).

In the "National Standards for Disciplinary Classification and Code of the People's Republic of China" (that is "Subject Classification and Code", the latest edition is a revised edition in 2009), statistics are also listed as a first-level discipline, under the Humanities and Social Sciences category.

The teaching situation of domestic universities has not been carefully counted. The comprehensive, science, and normal colleges and universities, the undergraduate statistics majors are basically science colleges. And many universities are the same as Peking University, in the mathematics department/hospital. But the graduates are clearly classified. The Department of Mathematics has statistics, and other majors, such as economics, art, chemistry, geography, etc., have a bias toward statistical research.

1.2. International situation:

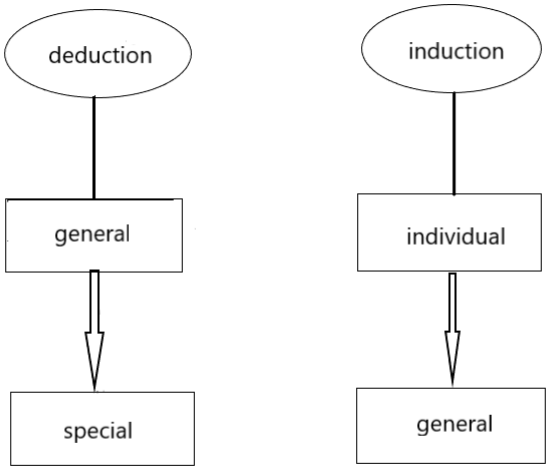
Two types of mainstream views:

1. Statistics is a branch of mathematics;
2. Statistics is another system that has some mathematical properties but is different from mathematics.

I think that these two viewpoints are all reasonable. The development of statistics will inevitably lead to a categorization of disciplines.This depends on which angle you look at.

1.3. A comparison of mathematics and statistics

First of all, we analyze these two disciplines from the source. Mathematics begins with deduction. Statistics begin with induction. Here I use legends to illustrate what is deduction (*the syllogism model, there is a major premise, corresponding to a conclusion; If there is a small premise to satisfy the big premise at this time, then you can draw a conclusion. Just as according to several axioms of mathematics, the present colorful world of mathematics is deduced*), and what is inductive (*inspecting individual things and introducing the laws of universal things*).



A sentence from *Encyclopedia Britannica* illustrates the statistics more vividly:

> Statistics is the science of collecting, analyzing, presenting, and interpreting data. Governmental needs for census data as well as information about a variety of economic activities provided much of the early impetus for the field of statistics. Currently the need to turn the large amounts of data available in many applied fields into useful information has stimulated both theoretical and practical developments in statistics.
>
> — Encyclopedia Britannica

Then we will separate them from the subject categories. As in the previous period, there are roughly five disciplines in mathematics (mathematical education is not included for the time being).

- Basic Mathematics;
- Operations Research and Cybernetics;
- Computational Mathematics;
- Probability Theory and Mathematical Statistics;
- Applied mathematics;

And the statistics are divided into

- Mathematical statistics;
- Applied statistics;

The above is a little difference between mathematics and statistics.

2

Exploring the College Entrance Examination

2.1. College entrance examination

2018 college entrance examination science mathematics (national paper I)

20.(12 points)

A certain product of a factory is packed in boxes, 200 pieces per box. Each box of products should be inspected before delivery to the user. If the non-conforming product is inspected, it will be replaced with a qualified product. Take 20 products for inspection, and then decide whether to test all the remaining products according to the inspection results. Let each product be a non-conforming product with a probability of $p(0 < p < 1)$. And whether each product is a non-conforming product is independent of each other.

(1) The probability of having 2 defective products in 20 products is $f(p)$, Find the maximum point p_0 of $f(p)$.

(2) Now, 20 pieces of products have been inspected. The result is exactly 2 pieces of non-conforming products. The p_0 determined in (1) is taken as the value of p. It is known that the inspection cost per product is 2 yuan. In the hands of the user, the factory must pay a compensation of 25 yuan for each non-conforming product.

(i) If the remaining products of the box are not inspected, the sum of the inspection cost and the compensation fee for this box product is recorded as X, seeking EX:

(ii) With the cost of inspection and compensation and the expected value as the basis for decision-making, should we test all the remaining products in this box?

The official answer is:

(1) the probability of 2 out of 20 products being unqualified is $f(p) = C_{20}^2 p^2(1-p)^{18}$. So

$$f'(p) = C_{20}^2[2p(1-p)^{18} - 18p^2(1-p)^{17}] = 2C_{20}^2 p(1-p)^{17}(1-10p).$$

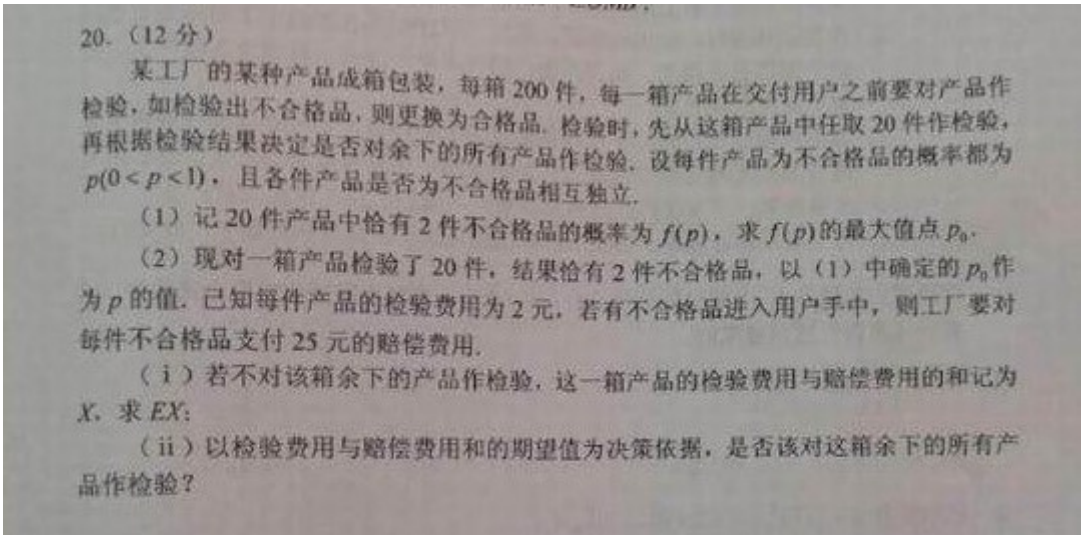
let $f'(p) = 0$, get $p = 0.1$. when $p \in (0, 0.1)$, $f'(p) > 0$; when $p \in (0.1, 1)$, $f'(p) < 0$.
so the maximum point of $f(p)$ is $p_0 = 0.1$.

(2) know by (1) $p = 0.1$.

(i) let Y denote the number of unqualified items in the remaining 180 products.
according to the question, $Y \sim B(180, 0.1)$, $X = 20\times2+25Y$, that is $X = 40+25Y$.
so $EX = E(40 + 25Y) = 40 + 25EY = 490$.

(ii) if the product is tested, the inspection fee for this box product is 400 yuan.
because of $EX > 400$,so the remaining products should be tested.

The above is the question and answer to this question. And the original version of this question is



20.（12 分）

某工厂的某种产品成箱包装，每箱 200 件，每一箱产品在交付用户之前要对产品作检验，如检验出不合格品，则更换为合格品．检验时，先从这箱产品中任取 20 件作检验，再根据检验结果决定是否对余下的所有产品作检验．设每件产品为不合格品的概率都为 $p(0<p<1)$，且各件产品是否为不合格品相互独立．

（1）记 20 件产品中恰有 2 件不合格品的概率为 $f(p)$，求 $f(p)$ 的最大值点 p_0．

（2）现对一箱产品检验了 20 件，结果恰有 2 件不合格品，以（1）中确定的 p_0 作为 p 的值．已知每件产品的检验费用为 2 元，若有不合格品进入用户手中，则工厂要对每件不合格品支付 25 元的赔偿费用．

（i）若不对该箱余下的产品作检验，这一箱产品的检验费用与赔偿费用的和记为 X，求 EX；

（ii）以检验费用与赔偿费用和的期望值为决策依据，是否该对这箱余下的所有产品作检验？

2.2. Analyze and reflect on this question

Now let's analyze this question.

The first question is to use the derivative to do, there is nothing to talk about.

For example, you can do this

$$f(p)=C_{20}^2p^2(1-p)^{18}=\frac{C_{20}^2}{81}(9p)^2(1-p)^{18}\leq\frac{C_{20}^2}{81}\left(\frac{2\times 9p+18\times(1-p)}{20}\right)^{20}=\frac{C_{20}^2}{81}\times\left(\frac{9}{10}\right)^{20}$$

if and only if

$$9p=(1-p)\Longleftrightarrow p=\frac{1}{10}$$

take the equal sign.

As for the answer to the second question.From the statistical point of view, this solution is no problem because it considers the actual demand. It is to test a large number of products, not just a box. But many many Box. So just compare the cost of "*full inspection*" and "*all non-inspection*" to determine whether a full inspection is needed.

The problem is that it is a topic in the "mathematical test paper", and it has its own "mathematical characteristics", which makes it difficult to find out whether it is to test "statistics" or to test "Math".[1]

The sub-question (1) in the title requires students to estimate the product's failure rate using the *maximum likelihood estimation*. There is no statistical term for "*maximum likelihood estimation*". The mathematical language is used completely. Find the local maximum point of the function, and the student naturally takes it as a math problem, finds the derivative, finds the zero point of the function, observes the sign of the derivative function, and confirms that $P_0=0.1$ is the absolute maximum point of the function $f(p)$.

The problem lies in the second (2) subtitle. The title says, "20 pieces of a box of products were inspected, and the result was exactly 2 pieces of non-conforming products. Take p_0 in (1) as the value of p." What is p? As mentioned, "The probability of each product being a non-conforming product is p", meaning that p is the failure rate. Fortunately, students do not learn more statistical knowledge, and there is no more time to think. Otherwise they will ask: Since "20 pieces have been inspected and 2 pieces have been found to be unqualified", then it is not possible to directly use the frequency $\frac{2}{20}=0.1$ to estimate the failure rate p. And to go around a circle to ask for the absolute maximum point of $f(p)$, Are you clearly trying to test our math?

In fact, the protagonists want to take the test is still statistical. Just because the students lack of statistical knowledge,can not make the problem clear. Because strictly speaking, statistics is not mathematics. It is an art of processing data. For different needs,the same batch of data can be processed differently. For example, the estimation of the failure rate can be either *frequency estimation* (also called *moment estimation*) or *maximum likelihood estimation*. Each has its own advantages and disadvantages, suitable for different occasions, to meet different needs. Coincidentally, in the case of this question, regardless of the number of pieces of defective products found in the 20 products sampled, the two estimated results are the same, both are $k/20$. Since the proposition is not willing to let the students use the frequency estimation, but must use the *maximum likelihood estimation*, then there is only one kind of understanding here. then it is still necessary to test mathematics. To assess the ability of students to find the maximum and minimum points of the function by derivation.[2]

Since this is a test of mathematics.Then consistent, the general question should be dealt with according to mathematical problems, At least others have done according to mathematical problems, can not be wrong.

For example, some people have answered question (2) (i):

> Set up the remaining 180 pieces of products and check the pieces$(0 \leq r \geq 180)$.
>
> If the sum of the inspection fee and the compensation fee at this time is recorded as Y, then there is
>
> $$E(Y) = 2r + \frac{25(180 - r)}{10} = 450 - \frac{r}{2}$$
>
> When $r = 180$, $E(Y)$ reaches the minimum, so the remaining products should be "full inspection".

This method of finding the absolute minimum is consistent with the spirit of sub-paragraph (1). Both use the maximum and minimum in mathematics as a solution to the problem. Since the maximum and minimum points available in (i) is the estimation of the failure rate, then (ii) can also choose the inspection plan according to the maximum and minimum.

What's more, here is also the question of how to negate the "*full name proposition*" in the logic. In any case, the negation of "*full inspection*" is "*not full inspection*", including "*partial inspection*" and "*all non-inspection*". Just as many middle school teachers pointed out, that this is repeatedly emphasized in the middle school mathematics teaching, and it is also the content to be assessed as the college entrance examination. It cannot reflect different requirements in the same test paper.

In fact, the trade-off criteria used in the official answer to (2) sub-question (i) are only known to those with product inspection experience, Because only when faced with large quantities of products, will only choose between "full inspection" and "all failures".[3]

3

Conclusion

Statistics have been separated from undergraduate majors in China. As a first-level discipline in parallel with mathematics, it is because “statistics is not mathematics.” A basic feature of mathematics is that “a problem has only one answer”, and statistics are not. For the same question, there can be different answers. For example, as mentioned above, for the same parameter, you can use frequency estimation or you can use maximum likelihood estimation. They tend to be different. In today’s big data fashion today,in the face of the same data, it is more common to draw different conclusions.

Since the middle school has put statistics into mathematics to learn, there is a question of how to teach how to learn how to assess. Especially when it comes to the college entrance examination questions, it is more necessary to consider. [4]

Because the probability is mathematical, the statistics are not. So be cautious and cautious when making probability statistics. Otherwise, there will be statistics encountering mathematics.

So although mathematics and statistics are separated, they are essentially closely related.

Although many people now lack statistical concepts. They are accustomed to looking at statistical issues from a mathematical perspective and treating statistical problems as mathematical problems. But I want to say *one is one, two is two, mathematics is mathematics, and statistics is statistics.*

Bibliography

[1] L. X and Y. L, *Analysis of the "probability and statistics" test questions in the national college entrance examination in 2018 and reflections on teaching,* Middle school mathematics magazine , 50 (2018).

[2] C. S, *Statistics encounter mathematics,* Middle school mathematics teaching , 1 (2018).

[3] Y. Z, H. Z, and Y. L, *A comparative study of probability and statistical content in the new and old curriculum standards of high school mathematics,* Teaching and management , 77 (2019).

[4] C. W, *Viewing the "core literacy of mathematics subjects" from the propositional trend of statistical probability,* Middle school mathematics , 72 (2019).