The Analysis Of Two Kinds Of Error-Prone Problems In Probability Theory And Mathematical Statistics

Yunjuan[1,a] Li, Xueshuang Fan[1,b,]* and Fengjun Yan[1,c]

[1]Xi 'an siyuan university Xi 'an, Shanxi, China

[a]liyunjuan22@163com, [b]fanxueshuang11@163.com, [c]yanfj1979@163.com

*corresponding author

Keywords: Conditional Probability; Full Probability; Interval Estimation; Hypothesis Test

Abstract: This paper from the probability theory and mathematical statistics in the course of two types of easy error, analysis of the causes of errors, examples of relevant topics further explained how to correctly distinguish and choose the correct solution, through the example of the correct solution to strengthen, so as to deepen students' understanding of the relevant concepts and methods.

1. Introduction

Probability theory and mathematical statistics is a mathematical discipline that studies the statistical regularity of random phenomena and the processing technology of random data. It is widely used in economics, management, education, psychology and many other disciplines [1-5].This course as a common course in colleges and universities set up a foundation, has the very important status in the college course. Many students are learning this class often feels tired, mainly is the course and students study the higher mathematics in the analysis question, solve the problem of way of thinking have different, students often confuse together [6], the knowledge problem solving error caused by this article will give two confusing easily wrong, and wrong solutions analysis, give the correct solution, explaining the difference, give the relevant type questions in order to make clear and strengthen the correct solution.

2. Confusion Between Conditional Probability And Full Probability Formula

Example 1: There are 10 transistors in a box, 4 of which are defective and 6 of which are genuine. Take one of them twice each time and do not put it back into the sample.

(1) What is the probability of getting the real product the first time?

(2) What is the probability of getting the authentic product the second time?

Wrong solution: (2) Set A: The first time I got a defective transistor.

B: The second time I got a genuine transistor

$$P(B) = P(B|A) = \frac{5}{9}.$$

Analysis: the wrong solution to the problem of the second meaning understanding is flawed, because the first question is for the first time to get the real thing the probability of the transistor, when doing the second question, students can easily relate the first question and the second q, wrong considered under the condition of the first to ask for a second to ask, there are also some students because of all the applicable condition of generalized formula without understanding and led to the solution to this error.

Set A: The first time I got a defectivetransistor

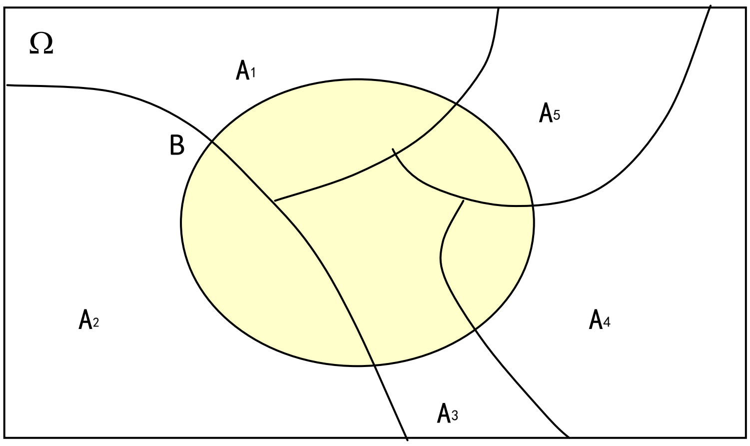
B: The second time I got a genuine transistor

$$P(A) = \frac{6}{10} = \frac{3}{5},\ P(\overline{A}) = \frac{4}{10} = \frac{2}{5},\ P(B|A) = \frac{5}{9},\ P(B|\overline{A}) = \frac{6}{9}.$$

$$P(B) = P(A)P(B|A) + P(\bar{A})P(B|\bar{A}) = \frac{3}{5}\times\frac{5}{9}+\frac{2}{5}\times\frac{6}{9} = \frac{9}{15}.$$

Note: The difference between the application of conditional probability and full probability formula: conditional probability is with A condition, which is usually directly or indirectly implied in the problem. Generally, it is the probability of B occurring in the case of A, and this B is unique and determined. The total probability formula for finding the probability of event B is also conditional, but the total probability formula generally comes from at least two conditions A_1, A_2, or more conditions. For example, when there is a relationship of figure 1between B events and A_i the full probability formula $P(B)=\sum_{i=1}^{\infty}P(A_i)P(B|A_i)$ can be used [7,8].

Fig 1. Schematic diagram of the total probability formula



The solutions to these two probabilistic problems are illustrated by example 2 below:

Example 2 7 out of the 10 lottery tickets have prizes. Now Party A and Party B have bought one lottery ticket in succession.

If A wins the lottery, the probability that B wins.

（1）The probability of B winning the lottery.

Analysis :(1)The question to be asked is obviously the probability with a condition, and the condition "in the case of A winning" is determined to be unique, so the conditional probability formula can be used to solve this problem.

(2) Find out the probability of B winning the lottery. B winning the lottery will happen under the two conditions of A winning the lottery and A not winning the lottery. The whole sample space Ω includes A winning the lottery A and A not winning the lottery $\bar{A}$.

Solution: A: A won the lottery, B: B won the lottery

$$P(B|A) = \frac{6}{99} = \frac{2}{33} \quad (1)$$

$$P(B) = p(A)P(B|A) + P(\bar{A})P(B|\bar{A}) = \frac{7}{100}\times\frac{6}{99}+\frac{93}{100}\times\frac{7}{99} = \frac{7}{100} \quad (2)$$

3. Confusion Between Interval Estimation And Parameter Hypothesis Testing

The following question is a final exam in our school, and some students have made the following wrong solutions

Example 3. The diameter of buttons produced in a workshop of a certain factory is known as μ_0=26, and it follows normal distribution $N(\mu, 5.2^2)$. Now samples with capacity n=100 are taken, and the sample mean is 26.56. Is the production in this workshop normal? (Let $z_{0.025}=1.96$, α =0.05 and keep the decimal to 0.01)

The wrong solution:. $\sigma = \sqrt{5.2^2} = 5.2$, $n = 100$, $\bar{x} = 26.56$, $P\left\{\left|\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}\right| < Z_{0.025}\right\} = 95\%$

The confidence interval of satisfying the above conditions is $(\bar{X} - Z_{0.025}\frac{\sigma}{\sqrt{n}}, \bar{X} + Z_{0.025}\frac{\sigma}{\sqrt{n}})$

$$\bar{x} - \frac{\sigma}{\sqrt{n}} Z_{\alpha/2} = 26.56 - \frac{5.2}{\sqrt{100}} \times 1.96 \approx 24.54, \quad \bar{x} + \frac{\sigma}{\sqrt{n}} Z_{\alpha/2} = 26.56 + \frac{5.2}{\sqrt{100}} \times 1.96 \approx 26.58$$

Because $26 \in (24.54, 26.56)$, the production in this workshop is normal.

Wrong solutions analysis: When workshop production is normal μ_0 = 26, we take the sample size of 100 samples and get some of the sample data, sample to determine whether the production is normal, is about to see whether sample expectations μ is 26, we can assume that the $\mu = \mu_0$ = 26, then construct the test in the small probability event, that is to say, the application of hypothesis test to do, and this problem students confusion of interval estimation and hypothesis testing parameters are made into interval estimation.

Truth: H_0: $\mu = \mu_0 = 26$; $H_1: \mu \neq \mu_0$ Statistics $Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$ are selected [9,10] .

Let $\alpha = 0.05$, $P\{|Z| > Z_{0.025}\} = 0.05$. For the observed value of $|z_0| = \left|\frac{26.56-26}{5.2/\sqrt{100}}\right| \approx 1.08$, $1.08 < 1.96$, it is accepted H_0 that production is considered normal at the significance level $\alpha = 0.05$.

Note: the interval estimation is a type of parameter estimation, it is through the sample some of the data to find the parameters of a general scope, makes the parameter have higher probability in this range, here will construct big probability event, through a big probability event, the scope of the higher probability is called confidence level, the range is called a confidence interval.

Parameter hypothesis testing is a type of hypothesis testing. It usually puts forward assumptions about population parameters and then constructs small probability events for inference. If a small probability event occurs in an experiment, the null hypothesis is rejected and the alternative hypothesis is considered to be valid; otherwise, the null hypothesis is accepted.

When doing the problem, if the confidence interval is explicitly calculated, the interval estimation method can be used. If you ask whether the production is normal, or check whether the value of a parameter is true, or whether there is a significant difference between a parameter and a value, etc., you can use the method of parameter hypothesis testing to solve the problem.

The following is an example 4 to illustrate the difference between interval estimation and parameter hypothesis testing when solving the problem:

Example 4. Given the length of a certain part produced by a factory follows a normal distribution $X \sim N(\mu, 0.06)$, 6 pieces are randomly selected from a batch of parts produced on a certain day, and the measured diameter data (unit :mm) is

14.6, 15.1, 14.9, 14.8, 15.2, 15.1

To take the confidence interval in degree of confidence0.95.(Where $z_{0.025} = 1.96$)

Analysis: The confidence interval of the length of this batch of parts is 0.95. That is the interval estimation problem.

Answer: $\sigma = \sqrt{0.06}$, $n = 6$, $\bar{x} = 14.95$, $P\left\{\left|\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}\right| < Z_{0.025}\right\} = 95\%$

The confidence interval of satisfying the above conditions is $(\bar{X} - Z_{0.025}\frac{\sigma}{\sqrt{n}}, \bar{X} + Z_{0.025}\frac{\sigma}{\sqrt{n}})$

$$\bar{x} - \frac{\sigma}{\sqrt{n}} Z_{\alpha/2} = 14.95 - \frac{\sqrt{0.06}}{\sqrt{6}} \times 1.96 = 14.754, \quad \bar{x} + \frac{\sigma}{\sqrt{n}} Z_{\alpha/2} = 14.95 + \frac{\sqrt{0.06}}{\sqrt{6}} \times 1.96 = 15.146$$

The confidence interval is $(14.754, 15.146)$.

Conclusion

In this paper, the conditional probability and the generalized formula, interval estimation and hypothesis testing parameters two easy wrong topic, for example, to the wrong solution, this paper analyzes the reasons of wrong solutions, and then gives the correct method and the application of how to distinguish the matters needing attention are given, finally illustrates the subject further illustrates how to correctly distinguish the correct solution to hope to play the role of the topic.

Acknowledgements

Thanks for the support of the school-level project probability theory and Mathematical Statistics of xi 'an Siyuan University in 2020.

References

[1] Sun Ye-qiang, Wang Na, a brief history of the development of probability theory and its applications in life, Journal of Jilin Normal University of Engineering and Technology, 2019 12.

[2] Zhao Wenxuan, introduction to the development and application of probability theory, mathematical physics problem solving research, 16, 2019.

[3] Wang Qi, Li Zhong, application of probability theory in geological advance prediction of new Luona tunnel, railway construction technology, no.6, 2000.

[4] Si Shoukui, Sun Zhaoliang, Mathematical Modeling Algorithm and Application, National Defense Industry Press, 2017.

[5] Li Zhiwei, Statistics, China Financial and Economic Press, 2016.

[6] Teng Jihong, Lu Zhibo, Huang Xiaoying. Teaching in Probability Theory and Mathematical Statistics

Reflections on the Teaching of Advanced Mathematics in Practice [J]. Research in Advanced Mathematics, 2017(1):117-120.

[7] Li Xieling, Tan Weiming, Probability Theory and Mathematical Statistics, Aviation Industry Press, 2017.

[8] Li Changqing, Zhang Yefang, Probability Theory and Mathematical Statistics, Tongji University Press, 2015.

[9] Hu Min, Probability theory and Mathematical Statistics, Shanghai Jiao Tong University Press, 2012.

[10] Miao Establishment, Probability Theory and Mathematical Statistics, East China Normal University Press, 2003.