An Analysis and Practice of Teaching Reform on Conditional Probability

Xiaonan Xiao

Xiamen University Tan Kah Kee College, Zhangzhou, Fujian China
**Corresponding author. Email: xiaoxn@xujc.com*

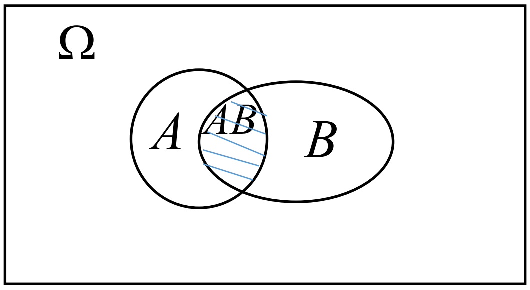
ABSTRACT
To begin with the exploration of the profound significance of conditional probability, the article investigates in depth the reduced sample space method and the reduced sample space cut method for calculating conditional probability. It aims to provide new ideas and methods for theoretical and practical teaching of probability statistics.
Keywords: *conditional probability, teaching reform, reducing sample space, section method, theoretical teaching, practical teaching*

1. THE MEANING OF CONDITIONAL PROBABILITY

Conditional probability is a very important concept in probability theory. Its significance can be explained from the following three aspects:

1.1. The Meaning of Geometric Intuition

We can use unit squares to represent the sample space Ω. The figure enclosed by any closed curve in the square represents the event whereas the area of the figure can be understood as the probability of the corresponding event. Assuming $A \subseteq \Omega, B \subseteq \Omega$



Unconditional probability (or absolute probability) $P(B)=\dfrac{P(B)}{P(\Omega)}$ (note $P(\Omega)=1$), on the aspect of geometrically intuitive, is equivalent to the proportion of B in space Ω. It can also be expressed as $P(B)=P(B \mid \Omega)$.

Conditional probability $P(B \mid A)=\dfrac{P(AB)}{P(A)}$ is actually limited to the scope of events A to examine the probability of events B. In terms of geometry, it is equivalent to the part B within A —— AB, the proportion in A [1-2].

1.2. The Perspective of Probability Space

Assuming a given probability space (Ω, F, P), and moreover $A \in F, P(A)>0$, then $P(\bullet \mid A)$ can be regarded as a newly introduced probability measure.

On the one hand, it can still be seen as changing the originally defined probability measure $P(\bullet)$ to $P(\bullet \mid A)$ in the original measurable space (Ω, F). Here, $P(\bullet \mid A)$ predicates any $B \in F$, and $P(B \mid A)=\dfrac{P(AB)}{P(A)}$. In this way, when the conditional probability $P(\bullet \mid A)$ is introduced, we see that the probability space changes from the original (Ω, F, P) to $(\Omega, F, P(\bullet \mid A))$.

On the other hand, we also hold that when the conditional probability $P(\bullet \mid A)$ is introduced, the probability space is reduced from the original (Ω, F, P) to $(\Omega_A^{'}, F_A^{'}, P_A^{'})$. Within there, $\Omega_A^{'}=\Omega \bigcap A$, $F_A^{'}=F \bigcap A$, and furthermore for any $B' \in F_A^{'}$, the definition is $P_A^{'}(B')=\dfrac{P(B')}{P(A)}$. In this way, the measurable space has been reduced from the original (Ω, F) to $(\Omega_A^{'}, F_A^{'})$ [3-4].

1.3. The Meaning of Probability Intuition

Conditional probability $P(B\mid A)$ and unconditional probability $P(B)$ can also be interpreted as posterior probability and prior probability. $P(B)$ can be interpreted as the understanding of the (absolute) possibility of the occurrence of the event B based on the data and experience accumulated in the past before the experiment. And now after the testing, we have obtained this new information that the event A (or result) has occurred. Then, this new information will require us and help us to re-examine or estimate the likelihood of the event B. It is after we have obtained the information that the event A has occurred that we re-understand the possibility of the occurrence of the event B. So it can be interpreted as the posterior probability [5-6].

To illustrate the above points, here are some examples:

Example 1 Suppose there are 100 new and old balls mixed in a box. Among the new ball, there are 59 white balls and 2 red balls. Among the old balls, there are one white ball and 38 red balls. Now we draw any of one ball from the box, so that A ——draws a white ball, B ——draws a new ball.

Suppose before the experiment (namely, before the ball is drawn), we have already known the number of various balls in the box. According to this known data, we can get the probability of "drawing any ball, hence a new ball" $P(B)=\frac{61}{100}$.

Now after the test, that is, after drawing any one of the ball from the box, assume that we have peeped through the fingers to find that the ball is white. After we have obtained this information, that is, "event A has already occurred", we will re-evaluate the probability of "drawing a new ball" (that is, the event B occurred) at this time. Because we already knew that out of 60 white balls, except for one old ball, all of the rest are new. Therefore, we can estimate that the possibility of drawing a new ball will greatly improve at this time. Hence, $P(B\mid A)=\frac{59}{60}$.

This is when we obtain the information that the event A has occurred, we try to re-understand the possibility of the occurrence of the event B. We can also presume that if we draw any one of the ball from the box, we have seen it through our fingers to determine the ball is red. After obtaining this information, that is, the "event has occurred," we can also reassess the probability of "drawing a new ball" (i.e., the occurrence of the event B) at this time. Because we know that out of 40 red balls, except for 2 new balls, the rest are all old balls. Therefore, it is estimated that the possibility of drawing a new ball should greatly reduce at this time. Hence, $P(B\mid \overline{A})=\frac{2}{40}$. That is a revised understanding of the possibility of an event B after being informed that the event $\overline{A}$ has occurred [7-8].

2. THE METHOD OF REDUCING SAMPLE SPACE

In general, the most common method for calculating conditional probability is the method of reducing sample space: in order to calculate conditional probability $P(B\mid A)$, the sample space Ω can be reduced to $\Omega'_A=\Omega\bigcap A$.And then on the reduced sample space Ω'_A, that is, under the circumstance of the additional condition "the event A has occurred," you can directly calculate the probability of the occurrence of the event B. That is $P(B\mid A)$.

As shown in the above example 1, under the condition that the event A ——"white ball drawn" occurs, the sample space is reduced from $\Omega=\{\textit{sixty white balls, forty red balls}\}$ to $\Omega'_A=\Omega\bigcap A=\{\textit{sixty white balls}\}$. On the aspect of Ω'_A, the probability of "new ball drawn" in the event B is $\frac{59}{60}$, that is $P(B\mid A)=\frac{59}{60}$.

Example 2 Suppose there are 4 products in the box, including 2 defective products and 2 genuine products. This time we will draw twice without putting them back, one at a time. Suppose the event A is "acquiring defective products for the first time" and the event B is "acquiring defective products for the second time", we get $P(B\mid A)$.

The dominant teaching method in the general textbooks is: suppose the sample space at the first sampling is Ω: assume $\Omega=\{\text{defective}_1, \text{defective}_2, \text{good}_1, \text{good}_2\}$, when A happens, that is, acquiring a defective product after the first sampling, the sample space is reduced from Ω to $\Omega'_A=\{\text{defective}_i, \text{good}_1, \text{good}_2\}$ (where i be one of 1 or 2). If we calculate the probability of "drawing the defective product at second time" in terms of Ω'_A, then $P(B\mid A)=\frac{1}{3}$.

I argue such a statement as often asserted is inappropriate for in this example, our test is a compound random test composed of two samplings. Its sample space Ω should have an equal sample point $4\times 3=12$. They can be expressed as follows:

$$e_1 = (\text{defective}_1, \text{defective}_2) \quad e_2 = (\text{defective}_1, \text{good}_1) \quad e_3 = (\text{defective}_1, \text{good}_2)$$
$$e_4 = (\text{defective}_2, \text{defective}_1) \quad e_5 = (\text{defective}_2, \text{good}_1) \quad e_6 = (\text{defective}_2, \text{good}_2)$$
$$e_7 = (\text{good}_1, \text{defective}_1) \quad e_8 = (\text{good}_1, \text{defective}_2) \quad e_9 = (\text{good}_1, \text{good}_2)$$
$$e_{10} = (\text{good}_2, \text{defective}_1) \quad e_{11} = (\text{good}_2, \text{defective}_2) \quad e_{12} = (\text{good}_2, \text{good}_1)$$

After the event A occurs, that is, when we have learned that a defective product is drawn for the first time, the sample space should be reduced from $\Omega = \{e_1, e_2, \cdots, e_{12}\}$ to $\Omega'_A = \{e_1, e_2, e_3, e_4, e_5, e_6\}$. However, when we investigate the situation of the occurrence of the event B in terms of Ω'_A (that is, "the second time we got the defective product"), it contains only two sample points e_1, e_4. Therefore, through the method of reducing sample space method we can obtain $P(B \mid A) = \frac{2}{6} = \frac{1}{3}$.

According to the abovementioned common method, we consider that the sample space is reduced from $\Omega = \{\text{defective}_1, \text{defective}_2, \text{good}_1, \text{good}_2\}$ to $\Omega'_A = \{\text{defective}_i, \text{good}_1, \text{good}_2\}$. Obviously, here Ω and Ω'_A are the sample spaces corresponding to two different random trials (i.e., the first and second sampling). In general, the method of reducing sample space should be adopted for the same random experiment. The occurrence of the events of A and B should be discussed in the same sample space. For two sample spaces corresponding to two different random trials, it is generally hard to say that one is a reduction of the other such as:

Example 3 Suppose there are 2 white balls and 1 red ball in the bag. Now we draw one from any of them, observe the color, put them back, and add a ball of the same color as the drawn ball. Then, we take any one from the bag. Suppose A——the first time to draw a white ball, B——the second time to draw a white ball, we get $P(B \mid A)$. Obviously, according to the usual way mentioned above, the sample space Ω during the first sampling is $\{\text{white}_1, \text{white}_2, \text{red}_1\}$. After A happens, that is, after the white ball is first drawn, a new white ball will be added (marked as white_3), at this time the sample space is "reduced" from $\Omega = \{\text{white}_1, \text{white}_2, \text{red}_1\}$ to $\Omega'_A = \{\text{white}_1, \text{white}_2, \text{white}_3, \text{red}_1\}$. Obviously, here Ω'_A is the "expansion" of Ω. How can it be seen as a "reduction"? Isn't this virtually a contradiction to the method of reducing sample space? So, how to explain the reduction of the sample space in this example? In fact, when applying the reduced sample space method, you must first understand what exactly the test and sample space are at issue. In this example, our test is a compound random test consisting of two ball draws. The sample space Ω has 3 × 4 = 12 equal sample points. We can arrange it into the following matrix form:

$$\Omega = \begin{bmatrix} (\text{white}_1, \text{red}_1) & (\text{white}_1, \text{white}_1) & (\text{white}_1, \text{white}_2) & (\text{white}_1, \text{white}_3) \\ (\text{white}_2, \text{red}_1) & (\text{white}_2, \text{white}_1) & (\text{white}_2, \text{white}_2) & (\text{white}_2, \text{white}_3) \\ (\text{red}_1, \text{red}_1) & (\text{red}_1, \text{white}_1) & (\text{red}_1, \text{white}_2) & (\text{red}_1, \text{red}_2) \end{bmatrix} (*)$$

Under the condition that "a white ball at the first drawing" (i.e., the event A occurs), the sample space Ω is reduced to be Ω'_A composed of 8 sample points in the first two rows of the above matrix. Thus according to the reduced sample space method, calculate the probability of "getting a white ball for the second time" on Ω'_A, we get $P(B \mid A) = \frac{6}{8} = \frac{3}{4}$.

To summarize, when we correctly select the test and sample space in the conditional probability problem, the result Ω'_A obtained through the reduced sample space method must be the reduction of the original sample space Ω. It must be Ω'_A is smaller than Ω [9-10].

3. SECTION METHOD FOR REDUCING SAMPLE SPACE

In Example 2, for the sake of $P(B \mid A)$, we consider that under the condition of "the first time drawing a defective product" (that is, in the case of the event A), at this time there is one defective product and two genuine products. The probability of "drawing another defective product" (i.e., the occurrence of the event B) is obviously $\frac{1}{3}$.

In Example 3, for the sake of $P(B \mid A)$, we consider that under the condition of "the first time a white ball is drawn" (that is, in the case of the event A), we put the white ball back at this time, and then add a new white ball. So there are 4 balls in the bag, of which there are 3 white balls and 1 red. The probability $P(B \mid A)$ of "getting a white ball after any drawing" (i.e. in the event of B) is obviously $\frac{3}{4}$.

It should be noted that this kind of approach commonly used in the calculation of conditional probabilities cannot be strictly called the reduced sample space method. For instance, in Example 3, if we regard the sample points in the sample space Ω as (x, y).

Herein $x \in \{\mathtt{white}_1, \mathtt{white}_2, \mathtt{red}_1\}$, $y \in \{\mathtt{white}_1, \mathtt{white}_2, \mathtt{white}_3, \mathtt{red}_1, \mathtt{red}_2\}$. (Note: $(\mathtt{white}_1, \mathtt{red}_2)$, $(\mathtt{white}_2, \mathtt{red}_2)$, $(\mathtt{red}_1, \mathtt{white}_3)$ do not belong to Ω). Then, Ω'_A corresponds to the first two rows of the matrix (*) listed in Example 3. The first row corresponds to $x = \mathtt{white}_1$, and the second row corresponds to $x = \mathtt{white}_2$. They can be regarded as two sections of Ω'_A respectively:

$$\Omega'_{A_1} = \{(\mathtt{white}_1, \mathtt{red}_1), (\mathtt{white}_1, \mathtt{white}_1), (\mathtt{white}_1, \mathtt{white}_2), (\mathtt{white}_1, \mathtt{white}_3)\}$$

and

$$\Omega'_{A_2} = \{(\mathtt{white}_2, \mathtt{red}_1), (\mathtt{white}_2, \mathtt{white}_1), (\mathtt{white}_2, \mathtt{white}_2), (\mathtt{white}_2, \mathtt{white}_3)\}$$

Apparently, $\Omega'_A = \Omega'_{A_1} + \Omega'_{A_2}$, $A = A_1 + A_2$. Here, A_1 and A_2 represent "the first time to draw No. 1 white ball ($\mathtt{white}_1$)" and "the first time to draw No. 2 white ball ($\mathtt{white}_2$)", respectively.

It can be seen from the symmetry that these two sections have the same probability structure with respect to event B (that is, the "white ball drawn for the second time"). So we can further reduce the sample space from Ω'_A to any section of Ω'_A, that is, on the aspect of Ω'_{A_1} or Ω'_{A_2}, it is easy to get $P(B \mid A) = \frac{3}{4}$. In addition, it can also be seen as combining the two sample points that correspond to the two sections of Ω'_{A_1} and Ω'_{A_2} on Ω'_A (that is, the two sample points corresponding to the first row and the second row of the matrix (*)). Thus we obtain a simpler sample space than Ω'_A, and we can calculate $P(B \mid A)$ from it.

$$\Omega''_A = \{(\mathtt{white}_1, \mathtt{red}_1), (\mathtt{white}_1, \mathtt{white}_1), (\mathtt{white}_1, \mathtt{white}_2), (\mathtt{white}_1, \mathtt{white}_3)\}$$

We may call this method the method of reducing the sample space. In fact, it is a simplification or "concentration" of the reduced sample space method. Therefore, this method is simpler than the reduced sample space method [11-12].

4. CONCLUSION

A lot of teaching practice proves that the key to further deepening the reform of teaching probability statistics is that teaching probability statistics must closely combine theory with practice. Only in this way can students' interest in probability statistics be greatly improved, and students' motivation and initiative in learning be cultivated. We can then promote the comprehensive and sustainable development in students. To push forward the teaching quality of probability statistics to a new higher level, we need to engage in many hard and meticulous work in the future.

REFERENCES

[1] Saint-Blancat C. Making Sense of Scientific Mobility: How Italian Scientists Loos Bake on Their Trajectories of Mobility in the EU [J].Higher Education Policy, 2017, 31(5):1-18.

[2] Bello Oyedele Adeshina, Bamiduro, Timothy Adebayo. Bootstrap Nonliner Regression Application in a Design of an Experi-ment Data for Fewer Sample Size [J]. International Journal of Research (Ijr), 2015, 2(2):1428-1441.

[3] Zhong Bing Lin. The reform of talent training mode is the core of the connotation construction of advanced education [J]. Chinese university teaching, 2013, (11)71-76.

[4] Gill S.Homecoming:an Investigation into the Effect that Studying Overseas had on Chinese Postgraduates' Life and Work on Their Return to China[J].Compare: A

journal of Comparative and International Education,2010,40(3):359-376.

[5] Lu Gen Shu. Classroom learning environment ,learning style and development of college students[J].Fudan university education BBS, 2012,(4):46-55

[6] Filippakou O, Tapper T.Quality Assurance in Higher Education: Thinking Beyond the English Experience [J].Higher Education Po; icy, 2007, 20(3):339-360.

[7] Harvey D I, Leybourne S J, Taylor A M R.Unit root tesing in practice:dealing with uncertainty over the trend andinitial condition[J].Econometric Theory,2009(25):587-636.

[8] Choi C, Moh Y. How useful are tests for unit-root in distinguishing unit-root process from stationary but non-linear process [J].Econometrics Journal, 2007(10):82-112.

[9] Filippakou O, Tapper T.Quality Assurance in Higher Education: Thinking Beyond the English Experience [J].Higher Education Po; icy, 2007, 20(3):339-360.

[10] M Geraci, M Bottai. Quantile regression for longitudinal data using the asymmetric Laplace distribution [J]. Biostatistics.2007 (8):140-154.

[11] Rodrigues P M. Properties of recursive trend-adjusterd unit root tests [J]. Economics Letters, 2006(91):413-419.

[12] Philip G A.Globalisation and the University: Myths and Realities in an Unequal World [J].Tertiary Education and Management, 2004, 10(1):3-2.