



DOI:10.22144/ctujos.2024.391

## MÔ PHỎNG XÁC SUẤT CÓ ĐIỀU KIỆN BẰNG PHẦN MỀM R

Bùi Anh Kiệt<sup>1\*</sup> và Trịnh Phùng Chí<sup>2</sup><sup>1</sup>Khoa Sư phạm, Trường Đại học Cần Thơ<sup>2</sup>Trường THPT Thiệu Văn Chôi, huyện Kế Sách, tỉnh Sóc Trăng

\*Tác giả liên hệ (Corresponding author): bakiet@ctu.edu.vn

### Thông tin chung (Article Information)

Nhận bài (Received): 09/12/2023

Sửa bài (Revised): 29/01/2024y

Duyệt đăng (Accepted): 05/03/2024

**Title:** Simulation of conditional probability using R software**Author(s):** Bui Anh Kiet<sup>1\*</sup> and Trinh Phung Chi<sup>2</sup>**Affiliation(s):** <sup>1</sup>Can Tho University;  
<sup>2</sup>Thieu Van Choi High School, Ke Sach district, Soc Trang province

### TÓM TẮT

Xác suất có điều kiện là một nội dung mới được bổ sung vào lớp 12 trong Chương trình Giáo dục phổ thông môn Toán 2018. Đây là một khái niệm hoàn toàn mới đối với học sinh phổ thông, đòi hỏi giáo viên phải có sự chuẩn bị kỹ về kế hoạch dạy học, phương pháp giảng dạy và phương tiện giảng dạy. Bài viết trình bày việc sử dụng mô phỏng trong giải bài toán Xác suất có điều kiện, đề xuất quy trình sử dụng phần mềm R trong giải bài toán Xác suất có điều kiện.

**Từ khóa:** Giáo dục phổ thông môn Toán, mô phỏng, phần mềm R, Xác suất có điều kiện

### ABSTRACT

Conditional probability is a new content added to grade 12 in the 2018 Mathematics General Education Program. This is a completely new concept for high school students, requiring teachers to be highly qualified and thorough in teaching plans, teaching methods and teaching aids. The article presents the use of simulation in solving Conditional Probability problems, and proposed procedures for using R software in solving Conditional Probability problems.

**Keywords:** Conditional probability, Mathematics General Education, R software, simulation

### 1. ĐẶT VẤN ĐỀ

Trong Chương trình Giáo dục phổ thông môn Toán 2018, khi được thực hành trong phòng máy tính, học sinh phải sử dụng được phần mềm để hỗ trợ việc học các kiến thức thống kê và xác suất. Một yêu cầu được đặt ra khi đó là phải sử dụng các phần mềm hiệu quả, linh hoạt và sáng tạo để tăng hứng thú, phát huy tính tích cực và phát triển năng lực cho học sinh. Việc sử dụng phần mềm toán học động như Geogebra, Geometer's Sketchpad (miễn phí), Cabri II Plus, Cabri 3D, Yenka (có phí) hỗ trợ học sinh rất nhiều trong quá trình tìm hiểu một số khái niệm liên quan đến hình học và các bài toán về đồ

thị. Tuy nhiên, ở nội dung xác suất, khi cần đến các mẫu số liệu lớn thì các phần mềm này chưa đáp ứng được. Trong khi đó, phần mềm R lại đáp ứng được yêu cầu này. Mô phỏng bằng phần mềm R là một mô hình dạy học mà trong đó học sinh sử dụng phần mềm này để mô phỏng các phép thử ngẫu nhiên và tính được tần suất thực nghiệm của biến cố. Từ đó giúp học sinh dễ dàng tiếp cận định nghĩa thống kê của xác suất và sử dụng mô phỏng để dự đoán kết quả của bài toán tính xác suất. Vì vậy, việc sử dụng mô phỏng trong dạy học vừa giúp học sinh nắm vững tri thức, vừa giúp hình thành các kỹ năng tham gia thực hành xã hội, đáp ứng các yêu cầu mà xã hội đặt ra cho ngành giáo dục hiện nay.

Theo Anh (2020), phần mềm R là một phần mềm hữu dụng trong thống kê và phân tích dữ liệu. Phần mềm R cài đặt được trên bất kỳ hệ điều hành nào, là phần mềm miễn phí mã nguồn mở, rất dễ cài đặt, dễ sử dụng và rất dễ học. Phần mềm R được sử dụng miễn phí nhưng nó mang đầy đủ những tính năng ưu việt của các phần mềm thương mại khác hiện có như SPSS, AMOS, STATA hay EViews. Ngoài ra, phần mềm R cho phép người dùng thêm các tính năng bổ sung bằng cách xây dựng các hàm mới nên R có những “package” thống kê phong phú, toàn diện nhất với công nghệ mới nhất, những ý tưởng mới thường xuất hiện đầu tiên trên R. Theo Batanero và Chernoff (2018), xã hội thế kỷ 21 cần những giáo viên được chuẩn bị tốt sẵn sàng giải quyết một nhiệm vụ phức tạp như dạy xác suất ở trường từ góc độ tiếp cận tần suất. Nếu giáo viên không được chuẩn bị tốt, việc dạy xác suất và thống kê ở trường sẽ gặp khó khăn. Để đạt được điều đó, mô phỏng bằng máy tính là công cụ được sử dụng.

Mô phỏng là một trong những kỹ thuật được sử dụng phổ biến nhất để thu thập thông tin về các hệ thống phức tạp, nhưng thuật ngữ mô phỏng được sử dụng để truyền đạt nhiều ý nghĩa khác nhau. Theo Từ điển Trực tuyến Merriam-Webster, mô phỏng là “sự trình bày bất chước hoạt động của một hệ thống hoặc quy trình bằng cách vận hành của một hệ thống hoặc quy trình khác”. Theo từ điển tiếng Việt, mô phỏng là “phỏng theo, lấy mẫu để tạo ra cái tương tự”.

Theo Batanero et al. (2021), các nhà phát triển chương trình giảng dạy nhấn mạnh rằng mô phỏng xác suất của các tình huống ngẫu nhiên giúp học sinh phát triển khả năng hình thành và phân tích các mô hình toán học. Trên thực tế, sự hiểu biết về xác suất và mô phỏng xác suất sẽ giúp họ trở thành những người hiểu biết hơn về toán học và thống kê.

Các tác giả theo trường phái giáo dục kiến tạo (constructive pedagogy) đã mô tả mô phỏng như là “kịch bản tái tạo” của cuộc sống thực trên máy tính, trong đó người học đóng vai trò chính trong việc tạo ra những nhiệm vụ phức tạp. Mô phỏng sẽ phản ánh “tổ hợp” của thế giới thực mà người học phải học các kỹ năng nhận thức ở mức cao hơn như điều tra (kỹ năng được đánh giá là chủ yếu cho nghiên cứu khoa học). Các mô phỏng này tạo cho người học có môi trường chỉ đạo một số nhiệm vụ tích hợp để qua đó học được các kỹ năng phức tạp ở các vấn đề xác thực hoặc thông qua các hướng dẫn (Tùng & Hạnh, 2013). Burrill and Ben-Zvi (2019) nhấn mạnh vai trò của mô phỏng các bài toán xác suất trong việc

dạy học thống kê và xác suất trong việc nâng cao hiệu quả dạy học nội dung này.

Theo Heumann and Schomaker (2022), có rất nhiều thiết kế nghiên cứu và cách thu thập dữ liệu khác nhau: thông qua thí nghiệm ngẫu nhiên, tìm kiếm trên web, đăng ký y tế, khảo sát, một cuộc điều tra dân số và nhiều cách khác. Tuy nhiên cách đơn giản và ít tốn kém hơn cả là sử dụng mô phỏng trên phần mềm máy tính.

Theo Kiệt và Tài (2021), mô phỏng là một công cụ rất tốt giúp kiểm chứng lại các kết quả tính toán bằng công thức xác suất cổ điển.

Theo Batanero et al. (2021), cơ sở lý thuyết của mô phỏng các bài toán xác suất là tiếp cận tần suất (hay định nghĩa thống kê của xác suất).

## 2. PHẦN MỀM R

R là một phần mềm miễn phí về phân tích và xử lý dữ liệu thống kê. Cộng đồng sử dụng nó rất đông đảo và ngày càng lớn mạnh. R có hơn 10000 thư viện lệnh với nhiều tính năng nổi trội trong phân tích dữ liệu nên hiện nay được nhiều trường đại học trên thế giới đưa vào giảng dạy ở các học phần liên quan đến thống kê. Ngoài ra, R còn hỗ trợ rất tốt trong mô phỏng các phép thử ngẫu nhiên, cỡ mẫu lên đến 100.000.000 với tốc độ khá nhanh.

### *Đặc điểm phần mềm R*

R dựa trên ngôn ngữ máy tính S, được phát triển bởi John Chambers và những người khác tại Phòng thí nghiệm Bell vào năm 1976. Đầu những năm 1990, Ross Ihaka và Robert Gentleman đã thử nghiệm thành công ngôn ngữ R ứng với chữ cái đầu tiên trong tên của họ. Theo Ihaka and Gentleman (1996), R là một môi trường lập trình miễn phí với những tính toán, thống kê, đồ họa và cho phép xử lý, lưu trữ dữ liệu. Do đó, R được sử dụng rộng rãi để phân tích thống kê trong nhiều lĩnh vực, có thể sử dụng cho nhiều mục tiêu khác nhau: từ các tính toán đơn giản, tính toán trên ma trận, đến các phân tích thống kê phức tạp và có thể sử dụng trong giảng dạy xác suất và thống kê và có một số lượng lớn các gói tiện ích (bổ trợ) Add-on mở rộng hệ thống và kho package phong phú và rất mạnh mẽ, cùng một lượng người dùng rất lớn.

Theo Bui and Lagrange (2017), không giống như với Excel, phần mềm R là một ngôn ngữ lập trình. Việc sử dụng các ngôn ngữ lập trình có thể là một thách thức đối với giáo viên và học sinh nhưng nó cũng có lợi ích cho việc dạy học. R là một phần mềm thống kê chuyên nghiệp và do đó việc sử dụng nó giúp học sinh chuẩn bị cho những ứng dụng mới

trong cuộc sống chuyên nghiệp của họ sau này. Ngoài ra, do R là một ngôn ngữ có cấu trúc nên học sinh có thể sử dụng một phương pháp tiếp cận mô-đun trong việc xây dựng một mô hình. So với Excel, một mô phỏng trong R có thể được thực hiện trên các mẫu rất lớn, lên đến mười triệu và nhiều hơn nữa nên có thể giúp học sinh hình dung quá trình dãy tần suất tiến về xác suất lý thuyết của biến cố khi kích thước mẫu tăng đến một số rất lớn.

### Ưu điểm của phần mềm R

R là một ngôn ngữ lập trình với nhiều gói tiện ích Add-on nên giúp người dùng tiếp cận các vấn đề của họ một cách dễ dàng và hoàn toàn miễn phí.

R cũng rất mạnh trong mô phỏng với nhiều chức năng hỗ trợ, thử nghiệm với cỡ mẫu lớn làm cho sự xấp xỉ xác suất chính xác hơn.

R cũng có đồ thị đẹp với rất nhiều loại khác nhau để giúp người dùng trực quan hóa dữ liệu hoàn toàn.

Dell'omodarme and Valle (2006) đã xác nhận rằng R như một gói (package) để giảng dạy thống kê với sự sẵn có của nhiều thư viện mở rộng, cho phép chia sẻ kiến thức cực kỳ hữu ích giữa các giảng viên thống kê làm việc trong các lĩnh vực khác nhau và ở các cấp độ khác nhau.

Theo Garfield et al. (2000), R được nâng cấp lên từ phiên bản Splus, nhưng cung cấp các kỹ thuật đồ họa mẫu để khám phá các bộ dữ liệu phức tạp hơn. Theo Barr and Scott (2011), R có các đặc điểm sự phạm đáng chú ý như tạo ra một môi trường trong đó học sinh có thể thực hiện sáng tạo, giải quyết vấn đề bằng cách nhập một chuỗi các lệnh khá tương tự với các bước được sử dụng để giải quyết vấn đề bằng thủ công nên học sinh dễ tiếp cận các thao tác lập trình dù trình độ công nghệ thông tin còn hạn chế.

R có thể chạy được với cỡ mẫu rất lớn; có thể lên tới 100.000.000 (đối với một số bài toán) với tốc độ khá nhanh, điều này cho phép học sinh thấy được một cách trực quan tần suất của biến cố ngẫu nhiên tiến về rất gần giá trị xác suất lý thuyết khi số phép thử tăng dần.

Với các đặc điểm trên, ta có thể khai thác sử dụng phần mềm R có hiệu quả ở nhiều khâu trong dạy học xác suất như: gợi động cơ, hướng đích, làm việc với nội dung mới, củng cố, kiểm tra đánh giá theo hướng tích cực hóa hoạt động học tập của học sinh.

### Hạn chế của phần mềm R

Theo Morishita et al. (2001), R là hoàn toàn mới cho tất cả các học sinh trong lớp nên họ cần thời gian

để làm quen với phần mềm này. R là một ngôn ngữ lập trình dựa trên các lệnh có cú pháp lập trình riêng và nghiêm ngặt. Mỗi câu lệnh trong R đều kết thúc bằng phím Enter nên khi viết chương trình chỉ cần sai một dòng là phải làm lại từ đầu. Vì thế, nhằm tránh xảy ra các lỗi trong cú pháp lệnh khi mô phỏng bằng phần mềm R thì đòi hỏi người học cần một số kỹ năng về công nghệ thông tin để sử dụng và hiểu rõ các hàm lệnh trong R.

So với Excel thì R không có giao diện đồ họa trực quan và không có các bảng tính quan sát dữ liệu nhằm đánh đổi khả năng xử lý dữ liệu với cỡ mẫu lớn trong quá trình mô phỏng.

Tuy nhiên, quá trình mô phỏng bài toán xác suất thống kê sử dụng R không đòi hỏi một chương trình với những câu lệnh phức tạp. Bước đầu, giáo viên sẽ giới thiệu R cho học sinh để họ dần tiếp cận và nắm được các hàm lệnh và thao tác mô phỏng với các bài toán đơn giản và nâng cấp dần về sau khi HS đã làm quen với R.

## 3. QUY TRÌNH MÔ PHỎNG BÀI TOÁN XÁC SUẤT

Qua nghiên cứu các tài liệu kết hợp mô phỏng thí nghiệm các bài toán xác suất, bài viết trình bày việc mô phỏng một bài toán xác suất theo một quy trình như sau:

**Bước 1:** Xác định nội dung bài toán bao gồm giả thiết và kết luận của bài toán.

**Bước 2:** Phân tích bài toán để lựa chọn câu lệnh thích hợp để mô phỏng không gian mẫu, biến cố cần tính xác suất.

**Bước 3:** Viết chương trình bằng phần mềm R dựa trên các câu lệnh ở Bước 2.

**Bước 4:** Thực hiện mô phỏng với số lần cho trước.

**Bước 5:** Kiểm tra kết quả, hiệu chỉnh và hoàn thiện chương trình (có thể quay lại Bước 2 để hiệu chỉnh).

## 4. MỘT SỐ KIẾN THỨC XÁC SUẤT CẦN THIẾT

### 4.1. Định nghĩa

**Tiếp cận cổ điển:** Theo Batanero et al. (2021), Laplace định nghĩa xác suất của một biến cố là “tỉ số của số trường hợp thuận lợi với số tất cả các trường hợp có thể xảy ra” (tr. 3).

$$P(A) = \frac{m}{n}$$

với  $m$  là số trường hợp thuận lợi của biến cố  $A$ ,  $n$  là số trường hợp có thể xảy ra, tức là số phần tử của không gian mẫu.

**Tiếp cận thống kê:** Theo Batanero et al. (2021), Bernoulli đề nghị ước lượng tần suất cho khái niệm xác suất căn cứ tính ổn định của thống kê. Trong cách tiếp cận này, xác suất biến cố  $A$  là một giá trị mà tần suất của sự kiện xuất hiện  $A$  dao động quanh giá trị ấy khi thực hiện một số lượng đủ lớn các phép thử. Trong thực hành ta có thể lấy tần suất (với  $n$  đủ lớn) xấp xỉ cho giá trị  $P(A)$ :

$$P(A) \approx \frac{n(A)}{n}$$

Với  $n$  là tổng số các trường hợp được khảo sát và  $n(A)$  là số lần xuất hiện biến cố  $A$  trong  $n$  trường hợp đó ( $n$  đủ lớn).

**4.2. Các công thức xác suất**

**Công thức cộng xác suất**

Cho  $A$  và  $B$  là hai biến cố bất kỳ trong cùng một phép thử. Ta có

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

Nếu  $A$  và  $B$  là hai biến cố xung khắc thì

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

**Công thức xác suất có điều kiện**

Xác suất của biến cố  $A$  khi biết rằng một biến cố  $B$  khác xảy ra kí hiệu là  $P(A/B)$ . Khi đó

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

với  $P(B) > 0$ .

**Công thức nhân xác suất**

Từ công thức xác suất có điều kiện

$$P(A/B) = \frac{P(A \cap B)}{P(B)} \text{ và } P(B/A) = \frac{P(A \cap B)}{P(A)}$$

Ta suy ra công thức nhân xác suất

$$P(A \cap B) = P(B) \cdot P(A/B) = P(A) \cdot P(B/A)$$

với  $P(A) > 0, P(B) > 0$ .

**Công thức xác suất toàn phần**

Giả sử  $\{B_1, B_2, \dots, B_n\}$  là hệ đầy đủ các biến cố với  $P(B_i) > 0, \forall i = \overline{1, n}$ . Khi đó

$$P(A) = \sum_{i=1}^n P(B_i) \cdot P(A/B_i)$$

**Công thức Bayes**

Giả sử  $P(A) > 0$  và  $\{B_1, B_2, \dots, B_n\}$  là hệ đầy đủ các biến cố với  $P(B_i) > 0, \forall i = \overline{1, n}$ . Khi đó:

$$\begin{aligned} P(B_k/A) &= \frac{P(B_k \cap A)}{P(A)} = \frac{P(B_k)P(A/B_k)}{P(A)} \\ &= \frac{P(B_k) \cdot P(A/B_k)}{\sum_{i=1}^n P(B_i) \cdot P(A/B_i)} \end{aligned}$$

**5. MÔ PHỎNG MỘT SỐ BÀI TOÁN BẰNG PHẦN MỀM R**

Dựa trên các bước ở quy trình trên, bài viết tiến hành nghiên cứu mô phỏng một số bài toán xác suất có điều kiện.

**Bài toán 1:** Gieo một con xúc xắc cân đối và đồng chất hai lần. Tính xác suất để lần đầu gieo được mặt một chấm biết tổng số chấm trong hai lần gieo không vượt quá 3.

**Giải**

Không gian mẫu là  $\Omega = \{(i; j) : 1 \leq i, j \leq 6\}$ .

Trong đó, cặp số  $(i; j)$  thể hiện việc lần đầu gieo xuất hiện mặt  $i$  chấm, lần sau gieo xuất hiện mặt  $j$  chấm.

Gọi  $A$  là biến cố: “Lần đầu gieo xuất hiện mặt 1 chấm”

$B$  là biến cố: “Tổng số chấm trong hai lần gieo không vượt quá 3”. Chúng ta dễ dàng liệt kê được các phần tử thuận lợi cho từng biến cố là

$$\begin{aligned} A &= \{(1;1), (1;2), (1;3), (1;4), (1;5), (1;6)\} \\ B &= \{(1;1), (1;2), (2;1)\} \\ A \cap B &= \{(1,1), (1,2)\}. \end{aligned}$$

Theo định nghĩa cổ điển của xác suất thì ta có

$$P(A) = \frac{6}{36}$$

$$P(B) = \frac{3}{36}$$

$$P(A \cap B) = \frac{2}{36}$$

Nếu biết rằng B đã xảy ra thì A xảy ra khi một trong hai kết cục (1;1) và (1;2) xảy ra. Do đó, xác suất của A trong điều kiện B là

$$P(A/B) = \frac{2}{3}$$

Nhận xét rằng:

$$\frac{2}{3} = \frac{\frac{2}{36}}{\frac{3}{36}} = \frac{P(A \cap B)}{P(B)} \approx 0.6667.$$

Hay chính là  $P(A/B) = \frac{P(A \cap B)}{P(B)}$ .

**Quy trình mô phỏng**

**Bước 1:** Xác định nội dung bài toán

Giả thiết: Gieo một con xúc xắc cân đối và đồng chất hai lần.

Kết luận: Xác suất của biến cố “Lần đầu nhận được mặt một chấm biết tổng số chấm trong hai lần gieo không vượt quá 3”.

**Bước 2:** Phân tích bài toán

Gieo một con xúc xắc hai lần: sử dụng hàm `a=sample(1:6,1)` và `b=sample(1:6,1)`.

Điều kiện lần đầu được mặt một chấm: (`a==1`) và điều kiện tổng số chấm trong hai lần gieo không vượt quá 3: (`a+b<=3`).

**Bước 3:** Viết chương trình

```
Function: lan_dau_mat_mot_cham (GlobalEnv)
1 function(n) {
2   d1=0
3   d2=0
4   for (i in 1:n){
5     a=sample(1:6,1)
6     b=sample(1:6,1)
7     if (a+b<=3){
8       d1=d1+1
9       if (a==1) d2=d2+1
10    }
11  }
12  d2/d1
13 }
```

**Bước 4:** Thực hiện mô phỏng với số lần: 10, 100, 1.000, 10.000, 100.000, 1.000.000, 10.000.000.

```
Console Terminal x Background Jobs x
R 4.3.2 ~ / ↻
>
> lan_dau_mat_mot_cham(10)
[1] 1
> lan_dau_mat_mot_cham(100)
[1] 0.875
> lan_dau_mat_mot_cham(1000)
[1] 0.6666667
> lan_dau_mat_mot_cham(10000)
[1] 0.6734694
> lan_dau_mat_mot_cham(100000)
[1] 0.6688088
> lan_dau_mat_mot_cham(1000000)
[1] 0.6670612
> lan_dau_mat_mot_cham(10000000)
[1] 0.6667891
```

**Bước 5:** Kiểm tra kết quả: Ta thấy khi cỡ mẫu lớn dần thì sai số so với kết quả giải bằng xác suất cổ điển nhỏ dần và rất gần với số không.

**Bài toán 2:** Gieo đồng thời ba con xúc xắc cân đối đồng chất. Tính xác suất để tổng số chấm xuất hiện trên ba con xúc xắc bằng 8 biết rằng ít nhất có một con xuất hiện mặt 5 chấm.

**Giải**

Không gian mẫu gồm các phần tử  $\Omega = \{(i, j, k) : 1 \leq i, j, k \leq 6\}$ .

Trong đó bộ số  $(i, j, k)$  kí hiệu cho việc “con xúc xắc thứ nhất xuất hiện mặt  $i$  chấm, con xúc xắc thứ hai xuất hiện mặt  $j$  chấm và con xúc xắc thứ ba xuất hiện mặt  $k$  chấm”.

Gọi A là biến cố: “Tổng số chấm xuất hiện trên ba con xúc xắc bằng 8”, B là biến cố: “Ít nhất một con xúc xắc xuất hiện mặt 5 chấm”. Ta có

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

Vì B là biến cố: “Ít nhất một con xúc xắc xuất hiện mặt 5 chấm” nên  $\bar{B}$  là biến cố: “Không có con xúc xắc nào xuất hiện mặt 5 chấm, do đó

$$\bar{B} = \{(i, j, k) : 1 \leq i, j, k \leq 6; i, j, k \neq 5\}.$$

$$\text{Suy ra } P(\overline{B}) = \frac{n(\overline{B})}{n(\Omega)} = \frac{5^3}{6^3}$$

$$\Rightarrow P(B) = 1 - P(\overline{B}) = 1 - \frac{5^3}{6^3} = \frac{91}{216}$$

Ta thấy  $A \cap B$  là biến cố: “Tổng số chấm xuất hiện trên ba con xúc xắc bằng 8 và ít nhất một con xúc xắc xuất hiện mặt 5 chấm”, do đó

$$A \cap B = \{(1; 2; 5), (1; 5; 2), (2; 1; 5), (2; 5; 1), (5; 1; 2), (5; 2; 1)\}$$

$$\text{Suy ra } P(A \cap B) = \frac{n(A \cap B)}{n(\Omega)} = \frac{6}{216}$$

Vậy xác suất cần tìm là

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{6}{216}}{\frac{91}{216}} = \frac{6}{91} \approx 0.0659.$$

**Quy trình mô phỏng**

**Bước 1:** Xác định nội dung bài toán

Giả thiết: Gieo ba con xúc xắc cân đối và đồng chất.

Kết luận: Xác suất của biến cố “Tổng số chấm xuất hiện trên ba con xúc xắc bằng 8 biết rằng ít nhất có một con xuất hiện mặt 5 chấm”.

**Bước 2:** Phân tích bài toán

Gieo ba con xúc xắc: sử dụng hàm `a=sample(1:6,3,replace=T)`.

Điều kiện có ít nhất một con xuất hiện mặt 5 chấm:  $(a[1]==5 \mid a[2]==5 \mid a[3]==5)$  và điều kiện tổng số chấm xuất hiện trên ba con xúc xắc bằng 8:  $(a[1]+a[2]+a[3]==8)$  hoặc  $(\text{sum}(a)==8)$ .

**Bước 3:** Viết chương trình

```
Function: Tong_bang_8 (.GlobalEnv)
1 function(n){
2   d1 = 0
3   d2 = 0
4   for (i in 1:n){
5     a = sample(1:6,3,replace=T)
6     if (a[1]==5 | a[2]==5 | a[3]==5){
7       d1=d1+1
8       if (a[1]+a[2]+a[3]==8) d2=d2+1
9     }
10  }
11  d2/d1
```

**Bước 4:** Thực hiện mô phỏng với số lần: 10, 100, 1.000, 10.000, 100.000, 1.000.000, 10.000.000.

```
Console Terminal x Background Jobs x
R 4.3.2 ~/>
> Tong_bang_8(10)
[1] 0
> Tong_bang_8(100)
[1] 0.04761905
> Tong_bang_8(1000)
[1] 0.08163265
> Tong_bang_8(10000)
[1] 0.06077087
> Tong_bang_8(100000)
[1] 0.06577264
> Tong_bang_8(1000000)
[1] 0.06627473
> Tong_bang_8(10000000)
[1] 0.06577233
```

**Bước 5:** Kiểm tra kết quả: Ta thấy khi cỡ mẫu lớn dần thì sai số so với kết quả giải bằng xác suất cổ điển nhỏ dần.

**Bài toán 3:** Một gia đình có 2 đứa con. Biết rằng có ít nhất 1 đứa bé gái. Hỏi xác suất có 2 đứa bé gái là bao nhiêu? Biết rằng xác suất sinh trai và gái như nhau.

**Giải**

Chúng ta có các nhận xét sau:

Xác suất để một đứa bé là trai hoặc gái là bằng nhau và bằng  $\frac{1}{2}$ .

Giới tính cả hai đứa bé là ngẫu nhiên và không liên quan nhau.

Do gia đình có 2 đứa bé nên có thể có 4 khả năng: (trai; trai), (gái; gái), (gái; trai), (trai; gái).

Gọi A là biến cố “Cả hai đứa trẻ đều là con gái” và B là biến cố “Có ít nhất một đứa trẻ là con gái” thì ta có

$$P(A) = \frac{1}{4}$$

$$P(B) = \frac{3}{4}$$

Do nếu A xảy ra thì đương nhiên sẽ xảy ra B nên ta có

$$P(A \cap B) = P(A) = \frac{1}{4}$$

Suy ra xác suất để cả hai đứa trẻ đều là con gái khi biết ít nhất có một đứa trẻ là gái là

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3} \approx 0.3333.$$

**Quy trình mô phỏng**

**Bước 1:** Xác định nội dung bài toán

Giả thiết: một gia đình có 2 đứa con.

Kết luận: xác suất của biến cố “Gia đình có 2 đứa bé gái biết rằng có ít nhất 1 đứa bé gái”.

**Bước 2:** Phân tích bài toán

Một gia đình có 2 đứa con: sử dụng hàm `a=sample(1:2,2,replace=T)` với quy ước 1 là bé gái, 2 là bé trai.

Điều kiện gia đình có ít nhất một bé gái:  $(a[1]==1 \mid a[2]==1)$  và điều kiện gia đình có hai bé gái:  $(a[1]+a[2]==2)$  hoặc  $(a[1]==1 \ \& \ a[2]==1)$ .

**Bước 3:** Viết chương trình

```
Function: Hai_be_gai (.GlobalEnv)
1 - function(n){
2   d1 = 0
3   d2 = 0
4 - for (i in 1:n){
5     a = sample(1:2,2,replace=T)
6 -     if (a[1]==1 | a[2]==1){
7       d1=d1+1
8       if (a[1]+a[2]==2) d2=d2+1
9     }
10  }
11  d2/d1
12 }
```

**Bước 4:** Thực hiện mô phỏng với số lần: 10, 100, 1.000, 10.000, 100.000, 1.000.000, 10.000.000.

```
Console Terminal x Background Jobs x
R R 4.3.2 · ~/
> Hai_be_gai(10)
[1] 0.4444444
> Hai_be_gai(100)
[1] 0.3552632
> Hai_be_gai(1000)
[1] 0.3267457
> Hai_be_gai(10000)
[1] 0.3337316
> Hai_be_gai(100000)
[1] 0.3333156
> Hai_be_gai(1000000)
[1] 0.3328833
> Hai_be_gai(10000000)
[1] 0.3331659
```

**Bước 5:** Kiểm tra kết quả: Ta thấy khi cỡ mẫu lớn dần thì sai số so với kết quả giải bằng xác suất cổ điển nhỏ dần.

**Bài toán 4:** Một hộp đựng 3 bi đỏ và 2 bi trắng. Lấy ngẫu nhiên đồng thời 3 bi từ hộp nói trên. Sau đó lấy tiếp 1 bi từ 3 bi vừa lấy ra. Giả sử bi lấy lần sau có màu đỏ. Tính xác suất để trong 3 bi lấy lần đầu có 2 bi đỏ và 1 bi trắng. Biết rằng các bi trong hộp là cân đối và kích thước giống nhau.

**Giải**

Gọi  $A$  là biến cố: “bi lấy lần sau có màu đỏ”

$B_1$  là biến cố: “3 bi lấy lần đầu có 3 bi đỏ”

$B_2$  là biến cố: “3 bi lấy lần đầu có 2 bi đỏ và 1 bi trắng”

$B_3$  là biến cố: “3 bi lấy lần đầu có 1 bi đỏ và 2 bi trắng”

Ta có  $n(\Omega) = C_5^3$

$$P(B_1) = \frac{C_3^3}{C_5^3} = \frac{1}{10}$$

$$P(B_2) = \frac{C_3^2 \cdot C_2^1}{C_5^3} = \frac{6}{10}$$

$$P(B_3) = \frac{C_3^1 \cdot C_2^2}{C_5^3} = \frac{3}{10}$$

$$P(A) = P(B_1)P(A/B_1) + P(B_2)P(A/B_2) + P(B_3)P(A/B_3)$$

$$= \frac{1}{10} \cdot \frac{3}{3} + \frac{6}{10} \cdot \frac{2}{3} + \frac{3}{10} \cdot \frac{1}{3} = \frac{18}{30} = \frac{3}{5}$$

Khi đó

$$P(B_2/A) = \frac{P(B_2 \cap A)}{P(A)} = \frac{P(B_2) \cdot P(A/B_2)}{P(A)}$$

$$= \frac{\frac{6}{10} \cdot \frac{2}{3}}{\frac{3}{5}} = \frac{2}{3} \approx 0,6667$$

**Quy trình mô phỏng**

**Bước 1:** Xác định nội dung bài toán

Giả thiết: Một hộp đựng 3 bi đỏ và 2 bi trắng. Lấy ngẫu nhiên đồng thời 3 bi từ hộp nói trên. Sau đó lấy tiếp 1 bi từ 3 bi vừa lấy ra.

Kết luận: Xác suất của biến cố “3 bi lấy lần đầu có 2 bi đỏ và 1 bi trắng biết rằng bi lấy lần sau có màu đỏ”.

Function: Hai\_do\_mot\_trang (GlobalEnv)

```

1 - function(n){
2   d1=0
3   d2=0
4 - for (i in 1:n){
5     a=sample (1:5,3)
6     b=sample (1:3,1)
7     red=0
8     for (j in 1:3) {if (a[j]<=3) red=red+1}
9 -   if (a[b]<=3){
10    d1=d1+1
11    if (red==2) d2=d2+1
12 -   }
13 - }
14   d2/d1
15 - }
```

**Bước 2:** Phân tích bài toán

Lấy ngẫu nhiên đồng thời 3 bi từ hộp đựng 3 bi đỏ và 2 bi trắng: sử dụng hàm a=sample(1:5,3) với quy ước 1, 2, 3 là bi đỏ; 4, 5 là bi trắng.

Lấy tiếp 1 bi từ 3 bi vừa lấy ra: sử dụng hàm b=sample(1:3,1).

**TÀI LIỆU THAM KHẢO**

Anh, V. T. L. (2020). *Ứng dụng phần mềm R trong dạy học xác suất, thống kê ở trường trung học phổ thông*. Luận văn thạc sĩ Sư phạm Toán, Trường Đại học Giáo dục - Đại học Quốc gia Hà Nội.

Batanero, C. & Chernoff, E. Editors (2018). *Teaching and Learning Stochastics*. Springer

Viết dòng lệnh tính số bi đỏ trong 3 bi được lấy ra lần đầu với biến red.

Điều kiện 3 bi lấy lần đầu có 2 bi đỏ và 1 bi trắng: (red==2) và điều kiện bi lấy lần sau có màu đỏ: (a[b]<=3).

**Bước 3:** Viết chương trình

**Bước 4:** Thực hiện mô phỏng với số lần: 10, 100, 1.000, 10.000, 100.000, 1.000.000, 10.000.000.

```

Console Terminal x Background Jobs x
R 4.3.2 · ~/
> Hai_do_mot_trang(10)
[1] 0.5714286
> Hai_do_mot_trang(100)
[1] 0.7
> Hai_do_mot_trang(1000)
[1] 0.6677909
> Hai_do_mot_trang(10000)
[1] 0.6703901
> Hai_do_mot_trang(100000)
[1] 0.6667166
> Hai_do_mot_trang(1000000)
[1] 0.6667632
> Hai_do_mot_trang(10000000)
[1] 0.6667868
```

**Bước 5:** Kiểm tra kết quả: Ta thấy khi cỡ mẫu lớn dần thì sai số so với kết quả giải bằng xác suất cổ điển nhỏ dần.

**6. KẾT LUẬN**

Mô phỏng bằng ngôn ngữ lập trình R giúp người học hiểu được bản chất của khái niệm xác suất (trong đó có khái niệm xác suất có điều kiện) một cách trực quan dựa trên tiếp cận thống kê (tiếp cận tần suất). Khái niệm xác suất được hình thành một cách hết sức tự nhiên tạo cho người học sự hứng thú, tích cực. Đồng thời, người học cũng trở nên linh hoạt, sáng tạo, trong sử dụng các công cụ, phương tiện học toán để khám phá và giải quyết vấn đề toán học. Bài viết cũng đưa ra được quy trình mô phỏng bài toán xác suất nhằm hỗ trợ cho giáo viên trong quá trình giảng dạy. Bên cạnh đó, nó còn là công cụ kiểm chứng lại các kết quả của bài toán xác suất.

Publisher. <https://doi.org/10.1007/978-3-319-72871-1>

Batanero, C., Begué, N., Álvarez-Arroyo, R., Valenzuela-Ruiz, S.M. (2021). Prospective Mathematics Teachers Understanding of Classical and Frequentist Probability.



- Mathematics*, 9, 2526.  
<https://doi.org/10.3390/math9192526>
- Barr, G. D., & Scott, L. (2011). Teaching Statistics in a Spreadsheet Environment Using Simulation. *Spreadsheets in Education (eJSiE)*, 4(3), 1-16.
- Bui, A. K., & Lagrange, J-B. (2017). Using Simulation by Excel and R in the Teaching of Probability and Statistics for Non-Math Major Students. *Journal of Mathematics and Statistical Science*, 7, 210 - 220.  
<http://www.ss-pub.org/wp-content/uploads/2017/07/JMSS17031901.pdf>
- Burrill, G., & Ben-Zvi, D. (2019). *Topics and Trends in Current Statistics Education Research International Perspectives*. Springer Publisher.
- Dell'Omodarme, M., & Valle, G. (2006). Teaching statistics with Excel and R. *ArXiv Physics e-prints*, ref. physics/0601083.  
<http://arxiv.org/pdf/physics/0601083.pdf>.
- Garfield, J., Chance, B. L., & Snell, J. L. (2000). Technology in college statistics courses. In D. Holton et al. (Eds.). *The teaching and learning of mathematics at university level: An ICMI study* (pp. 357–370). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Heumann, C., & Schomaker, M. (2022). *Introduction to Statistics and Data Analysis with Exercises, Solutions and Applications in R (Second Edition)*. Springer Publisher. <https://doi.org/10.1007/978-3-031-11833-3>
- Ihaka, R., & Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5(3), 299-314.  
<https://doi.org/10.1080/10618600.1996.10474713>
- Kiệt, B. A., & Tài, V. V. (2021). *Giáo trình Dạy học Xác suất Thống kê*. NXB Đại học Cần Thơ.
- Kiệt, B. A., & Thảo, T. T. T. (2020). Sử dụng phần mềm R để mô phỏng một số bài toán xác suất lớp 11. *Tạp chí Thiết Bị Giáo dục*, 227, 4-6.
- Morishita, E., Iwata, Y., Yoshida K. Y., & Yoshida H. (2001). Spreadsheet fluid dynamics for aeronautical course problems. *International Journal of Engineering Education*, 17(3), 294-311.  
<https://tdl.libra.titech.ac.jp/journaldocs/en/recordID/article.bib-02/ZR000000009558?hit=3&caller=xc-search>
- Tùng, L. H., & Hạnh, L. T. (2013). Ứng dụng kỹ thuật mô phỏng trong dạy học. *Tạp chí Giáo dục*, 316, 37-38.