TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**NGÔ CHÍ THUẬN – 523H0102**

**BÁO CÁO CUỐI KỲ**

**XÁC SUẤT VÀ THỐNG KÊ**

**CHO CÔNG NGHỆ THÔNG TIN**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2024**

TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**NGÔ CHÍ THUẬN – 523H0102**

**BÁO CÁO CUỐI KỲ**

**XÁC SUẤT VÀ THỐNG KÊ**

**CHO CÔNG NGHỆ THÔNG TIN**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2024**

**BÁO CÁO CUỐI KỲ XÁC SUẤT THỐNG KÊ**

**TÓM TẮT**

Bài báo cáo này sẽ trình bày chi tiết về lời giải, công thức cũng như cách thức giải từng câu hỏi đã được yêu cầu.

**MỤC LỤC**

[Câu 1. Bài toán tung xúc xắc 3](#_Toc186315262)

[Câu 2. Bài toán ngày sinh nhật 3](#_Toc186315263)

[Câu 3. Bài toán 2 hộp bóng xanh và đỏ 5](#_Toc186315264)

[Câu 4. Bài toán một túi đựng bóng 6](#_Toc186315265)

[Câu 5. Bài toán xạ thủ 7](#_Toc186315266)

[Câu 6. Bài toán độ tin cậy 8](#_Toc186315267)

[Câu 7. Bài toán chi tiêu mua sắm của các hộ gia đình 10](#_Toc186315268)

[Câu 8. Bài toán phương pháp học của 12 học sinh 12](#_Toc186315269)

# Bài toán tung xúc xắc

Trò chơi về 1 viên xúc xắc, xúc xắc là một biến cố ngẫu nhiên. Trong đó, mỗi mặt của xúc xắc có khả năng xuất hiện bằng nhau và với xác suất .

Do đó, kết quả của mỗi lần tung sẽ không chắc chắn và không đoán trước được. Nhưng qua rất nhiều lần tung xúc xắc, sẽ thấy được trung bình phần thưởng (đô la) sẽ tiềm cận với giá trị kỳ vọng .

Công thức tính giá trị kỳ vọng của bài toán này sẽ là:

Trong đó:

* là xác suất xuất hiện của mỗi mặt xúc xắc (đều bằng )
* là giá trị phần thưởng tương ứng với mỗi mặt xúc xắc (1, 2, 3, 4, 5, 6).

Áp dụng công thức trên ta có:

Vậy số tiền hợp lý mà người chơi cần trả để tham gia trò chơi là 3.5 đô la (đảm bào rằng mình (người tổ chức) sẽ không lời hoặc lỗ nếu trò chơi diễn ra nhiều lần), thu phí 3.5 đô la trờ lên nếu muốn bạn là người gần như luôn lời trong trường hợp chơi, tổ chức dài hạn

# Bài toán ngày sinh nhật

1. Xác suất để có ít nhất 2 người trong buổi tiệc có cùng ngày sinh

Gọi là xác suất có ít nhất 2 người trùng sinh nhật trong buổi tiệc

Gọi là xác suất không có 2 người cũng trùng sinh nhật trong buổi tiệc

Trước tiên ta sẽ tính xác suất không có 2 người cũng trùng sinh nhật trong buổi tiệc với một năm có ngày.

Do mỗi người đều có 365 ngày (giả thuyết) để chọn làm ngày sinh trong năm nên ta có công thức của 13 người là:

Số cách chọn ngày sinh mà không trùng sinh nhật với những người trước là:

Như vậy, xác suất không có 2 người cũng trùng sinh nhật trong buổi tiệc với 13 người là:

Xác suất có ít nhất 2 người trùng sinh nhật trong buổi tiệc này (với 13 người) là:

Vậy xác suất có ít nhất 2 người trùng sinh nhật trong buổi tiệc này (với 13 người) là hay

1. Xác suất để có ít nhất một người trùng ngày sinh với bạn

Gọi là xác suất để có ít nhất một người trùng ngày sinh với bạn

Gọi là xác suất không người nào trùng ngày sinh với bạn

Với số người trong buổi tiệc (không bao gồm bạn) là 13 người, xác suất để mỗi người có ngày sinh không trùng với bạn là . Khi có 13 người khác ngoài bạn, xác suất không có ai cùng ngày sinh với bạn sẽ là:

Vậy, xác suất để có ít nhất một người trùng ngày sinh với bạn được tính như sau:

# Bài toán 2 hộp bóng xanh và đỏ

Gọi là quả bóng đỏ lấy được ở lần thứ i

Gọi là quả bóng xanh lấy được ở lần thứ i

1. Xác suất bóng đầu tiên lấy ra có màu đỏ

Do lần lấy bóng đầu tiên luôn đến từ hộp đỏ nên xác suất lần đầu tiên lấy bóng ra có màu đỏ cũng là xác suất lấy được một bóng đỏ trong hộp màu đỏ:

Vậy xác suất quả bóng đầu tiên là màu đỏ là hay 0.8

1. Xác suất quả bóng thứ 2 lấy ra có màu đỏ

Do quả bóng tiếp theo được lấy từ hộp màu nào là do màu của quả bóng trước lấy được quyết định lấy được nên chúng ta phải áp dụng xác suất có điều kiện với dạng:

Trong đó:

B là sự kiện xảy ra trước và A là sự kiện xảy ra sau và bị quyết định, có ảnh hưởng do sự kiện B

Trong bài toán này, do sau khi lấy ra quả bóng đầu tiên, quả bóng thứ 2 được lấy ra từ hộp nào là do màu quả bóng trước đỏ (quả bóng thứ nhất) quyết định nên ta sẽ có 2 trường hợp:

* Xác suất lấy được bóng đỏ ở hộp đỏ ở lần thứ 2 do lần đầu lấy được bóng đỏ:
* Xác suất lấy được bóng đỏ ở hộp xanh ở lần thứ 2 do lần đầu lấy được bóng xanh:

Xác suất lấy được bóng xanh ở lần đầu tiên là:

Xác suất lấy được bóng đỏ ở lần 2 là:

Vậy xác suất quả bóng thứ 2 là màu đỏ là hay 0.715

# Bài toán một túi đựng bóng

Gọi X là số bóng đỏ lấy được

Gọi Y là số bóng xanh lấy được

Qua bài toán ta có các điều kiện sau:

* Số bóng xanh không được vượt quá số bóng xanh trong túi :
* Số bóng đỏ không được vượt quá số bóng đỏ trong túi :
* Lấy ngẫu nhiên 10 quả bóng: .

Qua các điều kiện trên ta có được miền giá trị như sau:

Từ đó ta có thể biểu diễn hàm Joint Probability Mass (PMF):

# Bài toán xạ thủ

Gọi là số viên đạn bắn trúng. X là một biến ngẫu nhiên tuân theo phân phối nhị thức với các tham số (số lần bắn) và (Xác suất một viên đạn bắn trúng).

Do đó, ta có:

Gọi điểm mà xạ thủ đạt được sau khi bắn 25 viên đận là D

Điểm của xạ thủ với nguyên tắc bắn trúng thì được cộng một điểm còn hụt thì trừ một điểm:

Chúng ta cần tính xác suất để tổng điểm nằm trong khoảng từ 4 đến 6 điểm, được viết như sau:

Hay:

Giải bất phương trình trên:

Do (số viên đạn bắn trúng) là số nguyên không âm nên ta có:

Và bởi là biến ngẫu nhiên tuân theo phân phối nhị thức nên ta có công thức phân phối nhị thức cho một biến ngẫu nhiên có dạng:

Trong đó:

* là tổng số lần bắn
* là số lần bắn trúng
* là xác suất bắn trúng của mỗi lần
* là số tổ hợp chọn trong và được tính theo công thức;

Thay số vào công thức ta có:

# Bài toán độ tin cậy

Qua đề bài, ta có biến ngẫu nhiên X tuân theo phân phối nhị thức, , trong đó:

* là tổng số lần thử
* là xác suất thành công trong mỗi lần thử

Quan sát mẫu ngẫu nhiên gồm các giá trị

Số mẫu:

Tổng số lần thử trên toàn bộ mẫu ngẫu nhiên:

Hàm mật độ xác suất: Phân phối nhị thức có hàm xác suất:

Với tổng số lần thử , ta có:

Hàm hợp lý (Likelihood function): Với mẫu quan sát , hàm hợp lý là tích của xác suất:

Trong đó:

Do các hệ số tổ hợp không phụ thuộc vào , ta chỉ cần xét phần phụ thuộc vào :

Lấy log của để đơn giản hóa:

Tìm điểm cực đại, để tối đa hóa , lấy đạo hàm theo , đặt bằng 0 và giải phương trình:

Với mẫu , tổng các giá trị là:

Thay số vào biểu thức trước đó ta có:

Vậy ước lượng hợp lý cực đại (MLE) cho tham số của bài toán là .

# Bài toán chi tiêu mua sắm của các hộ gia đình

1. Tìm khoảng tin cậy 95% cho trung bình chi tiêu mỗi tuần

Với mẫu ngẫu nhiên kích thước , nếu mẫu được chọn độc lập và đủ lớn, trung bình mẫu sẽ xấp xỉ phân phối chuẩn theo định lý giới hạn trung tâm

Công thức khoảng tin cậy 95% cho trung bình:

Trong đó:

* là trung bình mẫu
* là giá trị tới hạn với mức tin cậy là 95%
* là độ lệch chuẩn mẫu
* là kích thước mẫu .

Trung bình mẫu được tính như sau:

Với mức tin cậy 95%, 95% tương ứng với vùng nằm giữa 2 điểm và trên phân phối chuẩn (). Phần còn lại (5%) sẽ được chia đều cho 2 đuôi của phân phối, mỗi đuôi chiếm:

Giá trị là điểm mà xác suất tích lũy (cumulative probability) từ đến bằng 97.5% (95% + 2.5%)

Tra bản phân phối chuẩn ta có được:

Độ lệch chuẩn mẫu có công thức:

Khoảng tin cậy 95% ta có:

Biên độ sai số là:

Khoảng tin cậy sẽ là:

Hay:

Vậy, với mức tin cậy là 95%, trung bình chi tiêu hằng tuần của các gia đình sẽ trong khoảng từ 81.97 USD đến 84.49 USD.

1. Giải thích nghĩa “mức tin cậy 95%”

“Mức tin cậy 95%” nghĩa là nếu chúng ta lấy nhiều mẫu ngẫu nhiên từ cùng một tổng thể và tính toán khoảng tin cậy cho mỗi mẫu, thì 95% trong số các khoảng tin cậy này sẽ bao gồm giá trị trung bình thực sự của tổng thể.

1. Thay đổi độ rộng của khoảng tin cậy khi tăng mức tin cậy và ví dụ

Khi mức tin cậy tăng, giá trị tới hạn cũng sẽ tăng, từ đó làm cho biên độ sai số tăng. Điều này làm cho khoảng tin cậy trở nên lớn hơn.

Ví dụ:

Khi tăng mức tin cậy lên 99%, giá trị tới hạn sẽ có giá trị , biên độ sai số sẽ là:

Khoảng tin cậy lúc này sẽ là:

Khoảng tin cậy với mức tin cậy là 99% lớn hơn khoảng tin cậy với mức tin cậy là 95% trước đó).

1. Nếu phân phối của bài toán này không phải phân phối chuẩn

Nếu không phải phân phối chuẩn thì khoảng tin cậy vẫn có thể chấp nhận được, hợp lệ nhờ định lý giới hạn trung tâm. Do kích thước mẫu ngẫu nhiên đủ lớn (100 mẫu) nên phân phối của trung bình mẫu cũng sẽ xấp xỉ về chuẩn.

# Bài toán phương pháp học của 12 học sinh

1. Kiểm định giả thuyết rằng điểm trung bình thực tế của sinh viên dùng phương pháp học mới nhỏ hơn 80 điểm với mức ý nghĩa 5%.

Giả thuyết thứ nhất (): Cho rằng điểm trung bình của sinh viên dùng phương pháp học mới sẽ có điểm lớn hơn hoặc bằng 80

Giả thuyết thứ 2 (): Cho rằng điểm trung bình của sinh viên dùng phương pháp học mới sẽ có điểm thấp hơn 80.

Trung bình mẫu:

Độ lệch chuẩn mẫu:

Giá trị :

Giá trị tới hạn (critical t-value):

Với (mức ý nghĩa 5%) và bậc tự do

Dựa vào bảng phân phối với bậc tự do và mức ý nghĩa cho kiểm định một đuôi, ta nhận được giá tị tới hạn

Vậy giá trị nên ta bác bỏ giả thuyết một (), bác bỏ giả thuyết cho rằng điểm trung bình của sinh viên dùng phương pháp học mới sẽ có điểm lớn hơn hoặc bằng 80.