Giới thiệu về thống kê Bayes

I. Ôn tập một số khái niệm

- Xác suất.
- Định lý Bayes.

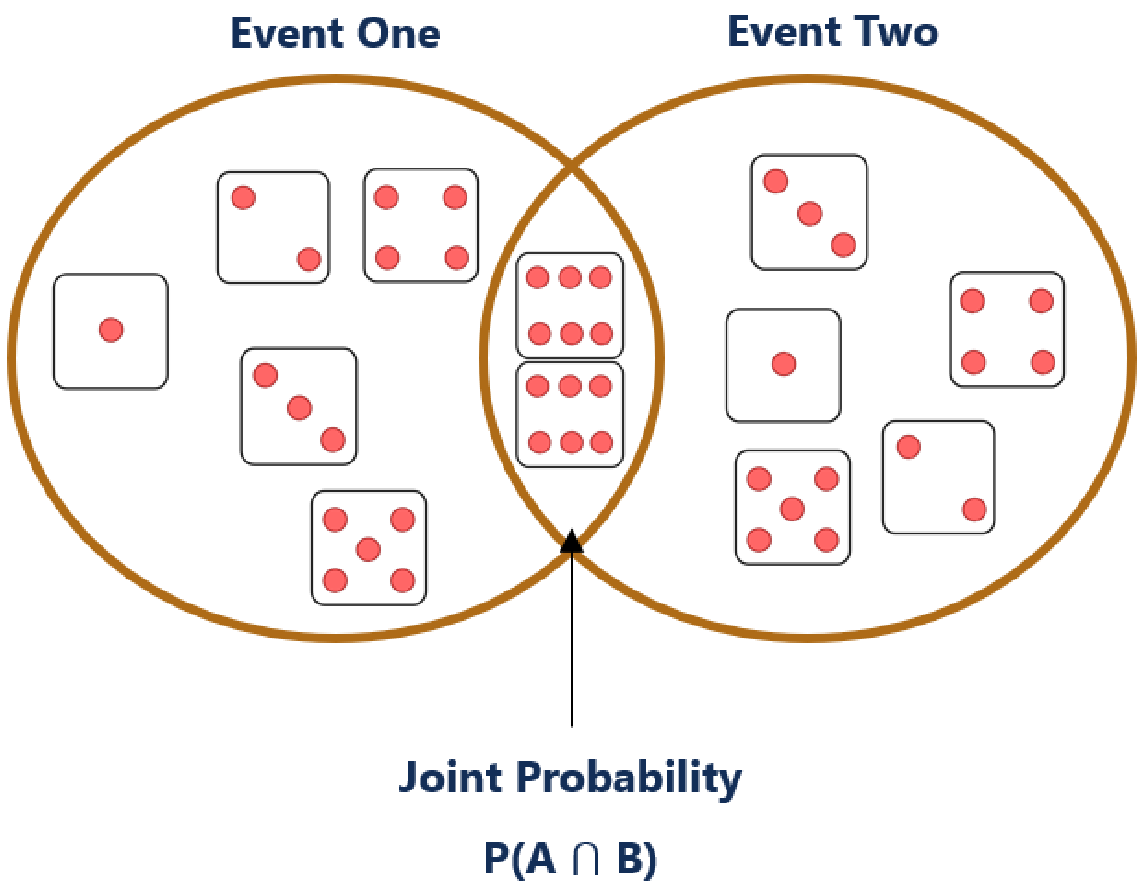
1. Nhắc lại về xác suất:

- Định nghĩa không gian mẫu: Tập hợp tất cả các kết quả có thể có của một thử nghiệm xác suất tạo thành một tập hợp được gọi là **không gian mẫu**. Thường được kí hiệu là Ω.
- Định nghĩa: **Biến ngẫu nhiên** là một biến có giá trị không xác định hoặc một hàm số gán giá trị cho từng kết quả của một thí nghiệm. Các biến ngẫu nhiên thường được chỉ định bằng các chữ cái và có thể được phân loại là biến rời rạc (các biến có giá trị cụ thể) hoặc liên tục (các biến có thể có bất kì giá trị nào trong phạm vi liên tục).
- P(X) : tổng quát. P(X(w)=x) = P(x) (w thuộc không gian mẫu)

Xác suất hợp:

- Định nghĩa: **Xác suất hợp** (*Joint probability*) là xác suất của hai biến cố cùng xảy ra. Xác suất hợp của *A* và *B* được ký hiệu $\mathrm{P(A,B)}$.
- Liệu $\mathrm{P(A,B)} = P(A).P(B)$? Không phải lúc nào cũng có.
- Biểu diễn sự xuất hiện cùng lúc của các sự kiện.

Xác suất hợp:



Xác suất có điều kiện:

- Định nghĩa: **Xác suất có điều kiện** (*Conditional probability*) là xác suất của một biến cố *A* nào đó, biết rằng một biến cố *B* khác xảy ra. Ký hiệu *P*(*A*|*B*), và đọc là "xác suất của *A*, biết *B*".
- Công thức xác suất có điều kiện : Cho 2 biến cố A và B (hoặc biến ngẫu nhiên), khi đó:

$$\mathrm{P(A|B)} = \frac{\mathrm{P(A,B)}}{\mathrm{P(B)}} \Leftrightarrow P(A,B) = P(A|B).P(B)\ (1)$$

Luật xích (chain rule):

- Luật xích (chain rule): Cho n biến cố (biến ngẫu nhiên) $A_1, A_2, \dots, A_n$, mở rộng từ (1), ta được:

$$P(A_1, A_2, \dots, A_n) = P(A_1, A_2, \dots, A_{n-1}|A_n).P(A_1, A_2, \dots, A_{n-2}|A_{n-1}, A_n) \dots P(A_1|A_2, \dots, A_n)$$

- Luật xích rất quan trọng vì nó cho ta tìm ra phân phối hợp cho toàn bộ mô hình.

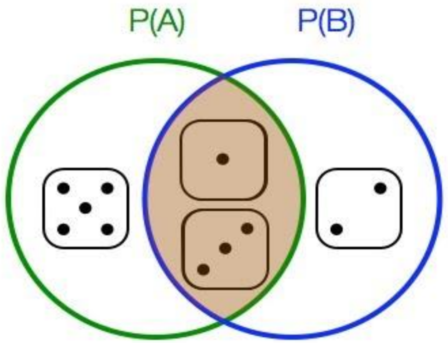
Conditional Probability

What is the Probability of

rolling a dice and it's value is less than 4

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

knowing that the value is an odd number



Xác suất biên:

- **Xác suất biên** (*Marginal probability*) là xác suất của một biến cố mà không quan tâm đến các biến cố khác.. Xác suất biên của *A* được ký hiệu là *P*(*A*), còn xác suất biên của *B* được ký hiệu là *P*(*B*).
- Công thức tính xác suất biên: Cho A và B là hai biến ngẫu nhiên, $val(A) = \{a_1, a_2, \dots, a_n), \mathrm{val(B)} = \{\mathrm{b_1}, \mathrm{b_2}, \dots, \mathrm{b_m}\}$, khi đó, ta có:

$$P(A = a_i) = \sum_{j=1}^{m} P(A = a_i, B = b_j)$$

Bài tập:

- Đặt S là biến ngẫu nhiên thể thao, G là biến ngẫu nhiên giới tính. Tính xác suất biên của các biến trong hình dưới:

	Male	Female
Football	0.24	0.15
Rugby	0.2	0.05
Other	0.1	0.26

- Tính $P(S = Football),\ P(S = other),\ P(G = Male)$.

Bài tập:

Subjects are asked if they would vote for a qualified woman for President. Results are broken down by gender.

- Task: Fill in the marginal probabilities.

Vote for Female?

Gender	Yes	No	Total
Male	.41	.07	
Female	.47	.05	
Total			

Biến cố tổng:

- *Biến cố xung khắc*: A và B là hai biến cố xung khắc khi chúng không đồng thời
  cùng xuất hiện trong phép thử, nghĩa là $A.B = \emptyset$.
- *Biến cố tổng* của hai biến cố A và B (ký hiệu $A \cup B$): Là biến cố xuất hiện khi
  A hoặc B xuất hiện trong phép thử, nghĩa là khi có ít nhất một trong các biến
  cố A, B xuất hiện.
  Chú ý: Nếu A và B là hai biến cố xung khắc khi đó ta ký hiệu biến cố tổng đơn
  giản là $A + B$.

Sự độc lập-phụ thuộc:

- Định nghĩa: Hai biến cố A và B được gọi là **độc lập** nếu $\mathrm{P(A, B)} = \mathrm{P(A). P(B)}$.
- Định nghĩa trên tương đương: A và B độc lập nếu $\mathrm{P(A|B)} = \mathrm{P(A)}$ và $\mathrm{P(B|A)} = \mathrm{P(B)}$.
- **Định lý Bayes**: Đưa ra công thức tính xác suất có điều kiện.
- Mô tả mối liên hệ giữa xác suất hợp và xác suất có điều kiện.
- Phát biểu: cho A và B là hai biến cố, khi đó:

$$\frac{\mathrm{P}(A|B)}{P(A)} = \frac{\mathrm{P}(B|A)}{\mathrm{P(B)}} \Leftrightarrow P(A|B) = P(A).\frac{P(B|A)}{P(B)}$$

Tính chất – bài tập:

- Cho A và B là hai sự kiện độc lập, khi đó các mệnh đề sau là tương đương:

1) *Sự kiện A và B độc lập.*
2) *Sự kiện A và $\overline{B}$ độc lập.*
3) *Sự kiện $\overline{A}$ và B độc lập.*
4) *Sự kiện $\overline{A}$ và $\overline{B}$ độc lập.*

- Chứng minh tính chất trên.

Bài giải:

- 1) => 2)
  Vì A và B độc lập nên $\mathrm{P(A)} - \mathrm{P(A}\bar{B}) = P(AB) = P(A).P(B) = P(A)\left(1 - P(\bar{B})\right) = P(A) - P(A)P(\bar{B})$. Suy ra $\mathrm{P(A}\bar{B}) = P(A)P(\bar{B})$.
- 2) => 3)
  Vì A và $\bar{B}$ độc lập nên: $\mathrm{P(A}\bar{B}) = P(A)P(\bar{B}) = \left(1 - P(B)\right)\left(1 - P(\bar{A})\right) = 1 - P(B) - P(\bar{A}) + P(B)P(\bar{A})$.
  Mà $\mathrm{P(A}\bar{B}) = 1 - P((\mathrm{A}\bar{B})^c) = 1 - P(\bar{A} \cup B) = 1 - P(B) - P(\bar{A}) + P(\bar{A}B)$. Suy ra $P(\bar{A}B) = P(B)P(\bar{A})$.

Bài giải:

- 3) => 4)
  Từ $P(\bar{A}B) = P(B)P(\bar{A})$, ta được :
  $\mathrm{P}(\bar{A}) - \mathrm{P}(\bar{A}\bar{B}) = P(\bar{A}B) = P(B)P(\bar{A}) = (1 - \mathrm{P}(\bar{B}))\mathrm{P}(\bar{A}) = \mathrm{P}(\bar{A}) - \mathrm{P}(\bar{A})\mathrm{P}(\bar{B})$. Suy ra $\mathrm{P}(\bar{A}\bar{B}) = \mathrm{P}(\bar{A})\mathrm{P}(\bar{B})$.
- d) Tương tự 2) => 3)

Công thức xác suất toàn phần:

- ***Công thức xác suất toàn phần*** *Cho* $A_1, A_2, \dots, A_n$, *là những biến cố xung khắc từng*
  đội một $(P(A_i A_j) = 0;\ i \neq j)$ *và chúng tạo nên một nhóm đủ, nghĩa là* $\sum_{i=1}^{n} A_i = \Omega$,
  và B là một biến cố với $P(B) \neq 0$ *, ta sẽ có:*

  $$P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \cdots + P(A_n)P(B|A_n).$$

Định lý Bayes mở rộng:

- Cho n biến ngẫu nhiên $A_1, A_2, \dots, A_n$ và B:
  Theo công thức trên ta có
  $$P(A_k \cap B) = P(A_k)P(B|A_k) = P(B)P(A_k|B),$$
  Từ đó với $P(B) \neq 0$, ta sẽ thu được:
  $$P(A_k|B) = \frac{P(A_k)P(B|A_k)}{P(B)}$$
  $$= \frac{P(A_k)P(B|A_k)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \cdots + P(A_n)P(B|A_n)}$$
  (công thức Bayes mở rộng)

Bài tập:

Chứng minh các tính chất sau của xác suất có điều kiện, khi $P(A) > 0$*:*

1) $P(B|A) \geq 0.$
2) $P(A|A) = 1.$
3) $P(B + C|A) = P(B|A) + P(C|A).$
4) $P(\sum_{k=1}^{\infty} B_k\,|A) = \sum_{k=1}^{\infty} P(B_k|A).$

Lời giải:

- 1) Hiển nhiên vì xác suất của một sự kiện luôn lớn hơn hoặc bằng 0.
- 2) $P(A|A) = \frac{P(A\cap A)}{P(A)} = \frac{P(A)}{P(A)} = 1.$
- 3) $P(C+B|A) = \frac{P(C+B,A)}{P(A)} = \frac{P\left((C\cup B)\cap A\right)}{P(A)} = \frac{P\left((C\cap A)\cup(B\cap A)\right)}{P(A)}$ (1)

  vì $C \cap B = \emptyset$ nên $(C \cap A) \cap (B \cap A) = \emptyset$, suy ra hai biến cố này xung khắc và $P\left((C \cap A) \cup (B \cap A)\right) = P\left((C \cap A)\right) + P\left((B \cap A)\right)$.

Thay vào (1), ta được$P(C + B|A) = \frac{P\left((C\cap A)\cup(B\cap A)\right)}{P(A)} =$
$\frac{P\left((C\cap A)\right)+P\left((B\cap A)\right)}{P(A)} = \frac{P(C\cap A)}{P(A)} + \frac{P(B\cap A)}{P(A)} = P(C|A) + P(B|A).$

Kì vọng:

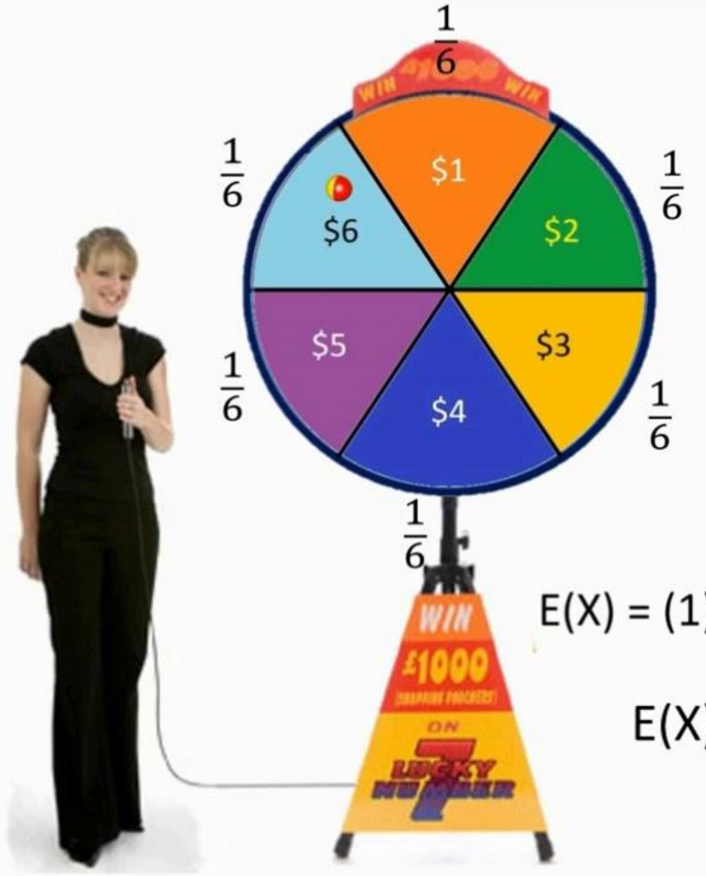
- Cho X là một biến ngẫu nhiên trên (Ω, P), **giá trị kì vọng** của X được định nghĩa:
  Trong trường hợp X là biến ngẫu nhiên rời rạc:
$$E[X] = \sum_{i=1}^{n} x_i p_i$$
- Trong đó, $Val(X) = \{x_1, \dots, x_n\}$, $p_i = P(X = x_i)$, $\sum_{i=1}^{n} p_i = 1$.

Kì vọng:

- Trong trường hợp X là một biến ngẫu nhiên liên tục, và X lấy giá trị: $a < X < b$. Khi đó kì vọng của X được tính như sau:

$$E[X] = \int_{-\infty}^{+\infty} xf(x) = \int_{a}^{b} xf(x)dx$$

- Trong đó: f là hàm phân phối xác suất liên tục trên khoảng (a, b).



Kỳ vọng là mong điều gì đó xảy ra, ví dụ kỳ vọng vào một ngày mai tươi đẹp, trong xác suất thống kê kỳ vọng là dự đoán điều gì sẽ xẩy ra và cho ra những số liệu cụ thể.

Kết quả (Outcomes)	1	2	3	4	5	6
Xác suất (Probability)	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$$E(X) = (1)(\frac{1}{6})+(2)(\frac{1}{6})+(3)(\frac{1}{6})+(4)(\frac{1}{6})+(5)(\frac{1}{6})+(6)(\frac{1}{6})$$

$$E(X) = \frac{1}{6} + \frac{2}{6} + \frac{3}{6} + \frac{4}{6} + \frac{5}{6} + \frac{6}{6} = \frac{21}{6}$$

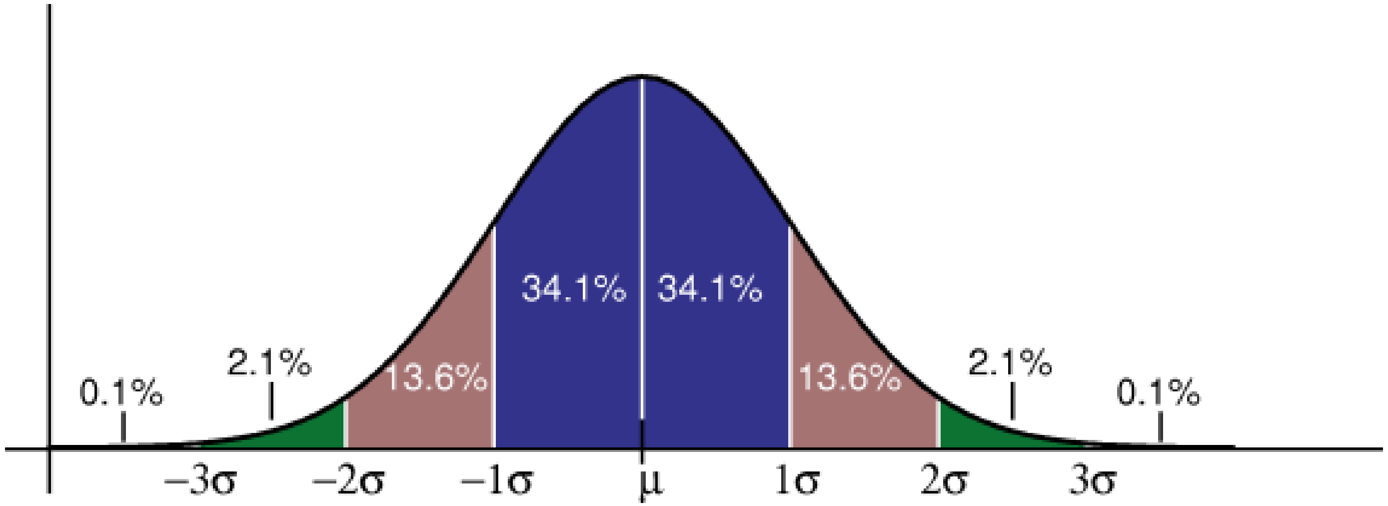
Phương sai:

- Cho X là một biến ngẫu nhiên trên (Ω, P), $\mu = \mathrm{E}[\mathrm{X}]$ là kì vọng của X, khi đó **phương sai của X**, kí hiệu σ, có công thức:

$$\mathrm{var}(\mathrm{X}) = \mathrm{E}((\mathrm{X}-\mu)^2)$$

- Công thức trên tương đương với:

$$\mathrm{var}(\mathrm{X}) = \mathrm{E}(\mathrm{X}^2) - \left(\mathrm{E}(\mathrm{X})\right)^2.$$



II. Thống kê Bayes:

- Ý tưởng của thống kê cổ điển.
- Phương pháp tiếp cận của thống kê Bayes.

1. Ý tưởng tiếp cận của thống kê cổ điển:

a) Ý tưởng:

- Thống kê cổ điển tính toán dựa trên xác suất một cách rất trực tiếp và không để ý đến sự thay đổi của tham số.
- Khi bắt đầu bài toán thống kê. Tham số luôn được cố định và không đổi.
- Xét một bài toán gồm dữ liệu D và tham số θ.
- Khi đó, tham số đươc cố định và ta sẽ làm việc với dữ liệu đã cho.

Ý tưởng:

- Công thức thường thấy trong thống kê cổ điển sẽ là:

$$P(D|\theta)$$

- Khi xác suất $P(D|\theta)$ gần bằng 0, ta sẽ đưa ra kết luận là tham số không phù hợp với mô hình.
- Tiếp tục thay đổi tham số để tính toán.

Ý tưởng:

- Cách tìm tham số: Sử dụng phương pháp lấy mẫu.
- Sử dụng công thức xác suất cơ bản:

$$P(\text{sự kiện } A) = \frac{\text{số lần xuất hiện sự kiện } A \text{ trên } n \text{ phép thử}}{n}$$

- Sau khi thử nghiệm hoàn thành, ta sẽ lấy kết quả trên để áp vào tham số theta.

Ý tưởng:

- Khi càng tăng n và tăng số lần thử, ta sẽ dần đi đến kết quả chính xác hơn bằng định luật số lớn.
- Định luật số lớn:

$$P(A) = \lim_{n \to \infty} \frac{\{số\ lần\ biến\ cố\ A\ xuất\ hiện\}}{n}$$

b) Ví du:

- Cho một đồng xu có hai mặt sấp ngửa. Ta kí hiệu biến cố ra mặt ngửa là H và mặt sấp là T.
- Gọi $\mathrm{X_i}$ là biến ngẫu nhiên cho phép thử tung đồng xu lần thứ i.
- Như vậy, ta có $\mathrm{val}(\mathrm{X_i}) = \{\mathrm{H}, \mathrm{T}\}$.
- Đặt $\mathrm{P}(\mathrm{X_i} = \mathrm{H}) = \theta_{\mathrm{i}}$.

Tiếp cận đơn giản:

- Ta sẽ tiếp cận bài toán này một cách đơn giản.
- Lấy mẫu tung đồng xu 100 lần, ta được 35 ra mặt ngửa.
- Khi đó $\mathrm{P(H)} = \theta = \frac{35}{100} = 0{,}35.$
- Theo định luật số lớn, việc lấy mẫu n lần với n tiến về vô cùng sẽ cho chúng ta đi tới $\hat{\theta}$ cần tìm.
- Trong trường hợp này 0,35 cũng có thể coi là rất gần so với $\hat{\theta}$.

Tiếp cận đơn giản:

- Giả sử ta có một chuỗi tung đồng xu:

$$\mathrm{A} = \mathrm{H} - \mathrm{T} - \mathrm{T} - \mathrm{H} - \mathrm{H}$$

- Vì xác suất mỗi lần thảy độc lập, nên ta có:

$$\mathrm{P}(\mathrm{A}|\theta) = \theta^3(1-\theta)^2$$

Tiếp cận đơn giản:

- Như vậy, ta sẽ tìm $\mathrm{P}(\mathrm{A}|\theta)$ bằng ước lượng hợp lí cực đại maximum likelihood.
- Nói cách khác, ta sẽ tìm:

$$\mathrm{argmax}_{\theta}\ \mathrm{P}(\mathrm{A}|\theta) = \mathrm{argmax}_{\theta}\theta^3(1-\theta)^2$$

Ở đây $0 < \theta < 1$.

Tiếp cận đơn giản:

- Theo phương pháp trên, giá trị θ nào càng khiến cho xác suất $\mathrm{p}(\mathrm{A}|\theta)$ càng lớn thì θ đó càng tiến gần đến giá trị ta mong muốn.
- Như vậy phương pháp likelihood cho ta biết θ sao cho θ phù hợp với xác suất tiên nghiệm nhất.

Trường hợp tổng quát:

- Từ trường hợp cụ thể ở trên, ta sẽ xây dựng một mô hình tổng quát cho phương pháp MLE cho bài toán tung đồng xu.
- Gọi $\mathrm{M}[1]$ là số lần tung đồng xu ra mặt ngửa.
- Gọi $\mathrm{M}[0]$ là số lần tung đồng xu ra mặt sấp.
- Khi đó hàm Likelihood sẽ là:

$$\mathrm{P}(\mathrm{D}|\theta) = \mathrm{L}(\theta : \mathrm{D}) = \theta^{M[1]}(1-\theta)^{\mathrm{M}[0]}$$

Trường hợp tổng quát:

- Lấy log của likelihood, kí hiệu $\mathrm{Log\ L}(\theta\colon \mathrm{D}) = \mathrm{l}(\theta\colon \mathrm{D})$, khi đó ta có:

$$\mathrm{l}(\theta\colon \mathrm{D}) = \mathrm{M}[1]\ln\theta + \mathrm{M}[0]\ln(1-\theta)$$

- Lấy đạo hàm hai vế phải theo θ và cho đạo hàm đó bằng 0, ta có:

$$0 = \mathrm{M}[1].\frac{1}{\theta} + \mathrm{M}[0].\frac{1}{1-\theta}$$

- Lưu ý: ở trên chính là phương pháp tìm cực trị mà ta đã được học từ THPT.
- Như vậy, ta được:

$$\hat{\theta} = \mathrm{argmax}_{\theta}\, \mathrm{l}(\theta : \mathrm{D}) = \frac{\mathrm{M}[1]}{\mathrm{M}[1] + \mathrm{M}[0]} = \frac{M[1]}{n}$$

- Như vậy ta có thể thấy trong bài toán đơn giản này, tham số tốt nhất mà ta nhận được lại chính là phép tính xác suất bằng phép đếm cơ bản.
- Như vậy, nếu ta tung đồng xu ra mặt ngửa 3 lần trong 10 lần tung thì tham số tốt nhất ta được chính là:

$$\hat{\theta} = \frac{3}{10} = 0,3$$

- Như vậy, với tham số đã tìm được, ta có thể dự đoán cho những trường hợp với số lần tung đồng xu lớn hơn.
- Ví dụ: nếu thảy 1000 lần thì ta có thể đoán rằng số lần ra mặt ngửa là 300.
- Tất nhiên con số có thể không chính xác như vậy nhưng đáp án cũng sẽ xấp xỉ gần bằng (theo ý tưởng của chúng ta).

Một vài kết luận:

- Rõ ràng trong ví dụ này, phương pháp cổ điển bị ảnh hưởng rất nhiều vào việc lấy mẫu.
- Định luật số lớn: lấy mẫu càng nhiều thì xác suất $\mathrm{P(H)} = \theta$ càng tiến về giới hạn. Vậy nên ta sẽ chọn một số n cụ thể cho phép thử này, với $\mathrm{M[1] + M[0] = n}$.
- Chỉ với n càng lớn thì $\frac{\mathrm{M[1]}}{\mathrm{n}}$ càng chính xác.

c) Vấn đề:

- Phương pháp truyền thống (phương pháp frequentist) dựa nhiều vào việc cố định các tham số để tính toán.
- Đòi hỏi thử sai nhiều.
- Phương pháp trên xây dựng dựa trên cơ sở của định luật số lớn.
- Định luật số lớn là kết quả đẹp về toán học nhưng có những vấn đề mà định luật này không thể giải quyết.

- Đôi khi ta gặp những bài toán mà không thể sử dụng sự thử sai.
- Ví dụ: Dự báo thời tiết cho ngày mai. Chúng ta chỉ duy nhất một "ngày mai", không thể nào có nhiều ngày mai để thử sai được.
- Trong những tình huống như vậy, ta sẽ không thể áp dụng định luật số lớn.
- Ta cần một phương pháp tiếp cận khác cho vấn đề tham số và tính toán xác suất.

2. Thống kê Bayes:

- Quan điểm của thống kê Bayes.
- Sự khác biệt giữa hai phép tiếp cận.
- So sánh ưu nhược điểm của hai phương pháp.
- Các tính toán liên quan đến thống kê Bayes.

a) Ý tưởng

- Sử dụng quan điểm tính toán theo phương pháp của Bayes.
- Trái ngược với phương pháp cổ điển, phương pháp của những nhà thống kê theo trường phái Bayes coi mọi thứ đều có khả năng của nó, kể cả giá trị tham số.
- Thay vì tìm $P(D|\theta)$, phương pháp Bayes còn xác định:

$$P(\theta) \text{ và } P(\theta|D)$$

- Thống kê Bayes coi tham số là một đại lượng có phân phối.
- Phân phối của tham số ảnh hưởng trực tiếp đến bài toán xác suất và phân phối của bài toán.
- Điều này giúp ta có thể chọn tham số để tính toán với công thức:

$$P(D|\theta) = P(\theta|D).\frac{P(\theta)}{P(D)} \infty P(\theta|D).P(\theta)$$

- Ta có thể thấy rằng đại lượng $P(D)$ là cố định, và dựa vào phân phối của theta mà có thể tính $P(\theta|D)$ và $P(\theta)$.
- Như vậy ta có thể nhận được một kết quả giống với phương pháp truyền thống, nhưng:

❖ Không phụ thuộc vào định luật số lớn.

❖ Dựa trên những tính toán và dự đoán về thay đổi của tham số.

Ví dụ:

- Tiếp tục với ví dụ về bài toán tung đồng xu

Ước lượng tham số bằng phương pháp Bayes:

- Theo ví dụ về bài toán tung đồng xu, ta thấy nếu lấy một mẫu $\mathrm{m} = 10$ và ta được 3 trường hợp mặt ngửa thì khi đó tham số $\theta = 0{,}3$.
- Tuy nhiên có thực sự là chúng ta sẽ lấy tham số $\theta = 0{,}3$?
- Trên thực tế, ta đã thử với các trường hợp số mẫu khác nhau, và nhận thấy rằng mỗi trường hợp cho ta số lần mặt ngửa rất khác.
- Vì vậy tính toán cố định như trên không phải là một ý tưởng tiếp cận chính xác.

Ước lượng tham số Bayes:

- Như vậy phương pháp cổ điển không thể cho chúng ta thấy sự khác biệt giữa 10 lần tung và 1000 lần tung.
- Mấu chốt nằm ở chỗ: ta luôn coi θ là một con số cố định sau mỗi lần tung.
- Vấn đề: θ hoàn toàn bị ảnh hưởng bởi kinh nghiệm đi trước, tức là θ là một tham số tiên nghiệm.
- Như thế θ có thể được lấy mẫu rất khác nhau, và hoàn toàn được coi như là một biến ngẫu nhiên.

- M = 10

T-T-T-T-T-T-T-H-H-H

- M = 11

T-T-T-T-T-T-T-H-H-H-H hoặc T-T-T-T-T-T-T-H-H-H-T

$$P\left(\theta = \frac{7}{11}\middle|D\right) = P\left(\theta = \frac{8}{11}\middle|D\right) = 0.5$$

$$P\left(D\middle|\theta = \frac{7}{11}\right) = \left(\frac{7}{11}\right)^7 . \left(\frac{4}{11}\right)^4 \Rightarrow P\left(\theta = \frac{7}{11}\right) = \frac{P\left(D\middle|\theta = \frac{7}{11}\right)}{P\left(\theta = \frac{7}{11}\middle|D\right)} =$$

- M = 5, D = T-T-H-T-T

$$P\left(\theta = \frac{7}{11}\right) = \alpha, P\left(D \middle| \theta = \frac{7}{11}\right) = \left(\frac{7}{11}\right)^4 . \frac{4}{11} \Rightarrow P(\theta|D)$$

- Xét xác suất hợp:

$$\begin{aligned}\mathrm{P}(X_1 = \mathrm{x}[1], \ldots, X_M = \mathrm{x}[\mathrm{M}], \theta) &= \mathrm{P}(\mathrm{x}[1], \ldots, \mathrm{x}[\mathrm{M}]|\theta).\,\mathrm{P}(\theta) \\ &= \mathrm{P}(\theta)\prod_{\mathrm{i}=1}^{\mathrm{M}} \mathrm{P}(\mathrm{x}[\mathrm{m}]|\theta) \\ &= \mathrm{P}(\theta)\theta^{M[1]}(1-\theta)^{\mathrm{M}[0]}\end{aligned}$$

Vẫn có thể chuyển về: $P(\theta|D)$

- Như vậy trong trường hợp này, ta tính cả θ và $\mathrm{P}(\theta)$

Phân phối Beta:

- Với $\mathrm{P}(\theta) = 1$, đây là trường hợp đã xét ở phần trước.
- Chúng ta sẽ sử dụng phân phối Beta cho tính toán xác suất $\mathrm{P}(\theta)$ trong trường hợp này.
- Phân phối Beta gồm có hai siêu tham số (hyper parameter) α_0 và α_1.
- Đây cũng chính là phần không thể thấy (phần ảo) của xác suất tung đồng xu ra mặt sấp và ngửa tương ứng.

Phân phối Beta:

- Ta đặt phân phối Beta bằng:

  $\theta \sim \mathrm{Beta}(\alpha_1, \alpha_0)$ nếu $\mathrm{p}(\theta) = \gamma\theta^{\alpha_1 - 1}(1-\theta)^{\alpha_0 - 1}$
- $\theta^{\alpha_1 - 1}(1-\theta)^{\alpha_0 - 1}$ chinh là xác xuất $P(\theta|D)$ với $\alpha_1 - 1$ lần tung mặt ngửa và $\alpha_0 - 1$ lần tung ra mặt sấp. Coi $M = \alpha_1 + \alpha_0 - 2$ là số phép thử tung đồng xu trng quá khứ.
- Ví dụ trước: $M = 10 = 8 + 2 - 2 = 8 - 1 + 4 - 1$

ở đây, ta có $\gamma = \frac{\Gamma(\alpha_1 + \alpha_0)}{\Gamma(\alpha_1)\Gamma(\alpha_0)}$ với $\Gamma(\mathrm{x}) = \int_0^{+\infty} \mathrm{t}^{\mathrm{x}-1}\mathrm{e}^{-\mathrm{t}}\mathrm{dt}$ là hàm gamma.

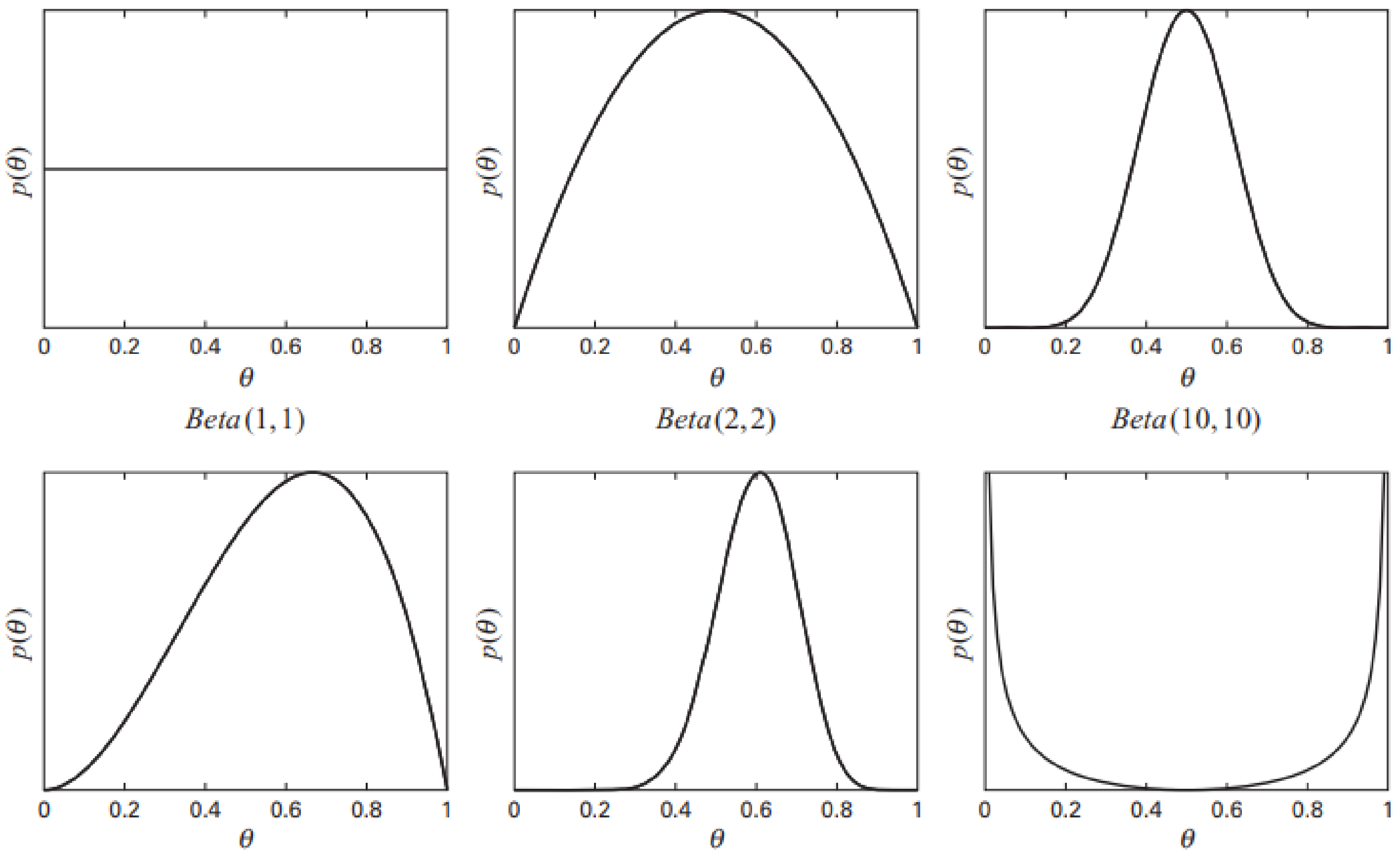
Phân phối beta:

- Với n là một số nguyên dương, ta có:

$$\Gamma(\mathrm{n}+1)=\mathrm{n}!\ \text{với n rất lớn}$$

- Trong trường hợp này, siêu tham số tương ứng với số lần tung đồng xu trước khi vào quá trình thử nghiệm.
- Cụ thể, α_1 tương ứng với xác suất ra mặt ngửa và α_0 tương ứng với xác suất ra mặt sấp.
- $y=\frac{\Gamma(\alpha_1+\alpha_0)}{\Gamma(\alpha_1)\Gamma(\alpha_0)}=\frac{(\alpha_1+\alpha_0-1)!}{(\alpha_1-1)!(\alpha_0-1)!}\infty\frac{(\alpha_1+\alpha_0-2)!}{(\alpha_1-1)!(\alpha_0-1)!}=C_{\alpha_1+\alpha_0-2}^{\alpha_1-1}$

Ví dụ:



- Xác suất biên dựa trên $\mathrm{P}(\theta) \sim \mathrm{beta}(\alpha_1, \alpha_0)$

$$\mathrm{P}\left(\mathrm{X}[\mathrm{i}] = \mathrm{x}^{\mathrm{i}}\right) = \int_0^1 \mathrm{P}\left(\mathrm{X}[\mathrm{i}] = \mathrm{x}^{\mathrm{i}} \middle| \theta\right). \mathrm{P}(\theta) \mathrm{d}\theta = \int_0^1 \theta \mathrm{P}(\theta) \mathrm{d}\theta = \frac{\alpha_1}{\alpha_1 + \alpha_0}$$

- Giả sử ta có nhiều lần lấy mẫu hơn, khi đó:

$$\mathrm{P}(\theta|\mathrm{D}) \propto \mathrm{P}(\mathrm{D}|\theta)\mathrm{P}(\theta) = \theta^{\alpha_1 + \mathrm{M}[1] - 1}(1-\theta)^{\alpha_0 + \mathrm{M}[0] - 1}$$

chính là $\mathrm{beta}(\alpha_1 + \mathrm{M}[1], \alpha_0 + \mathrm{M}[0])$

- Dự đoán dữ liệu $\mathrm{X}[\mathrm{M}+1]$:

$$\mathrm{P}(\mathrm{X}[\mathrm{M}+1]=\mathrm{x}^1|\mathrm{D})=\frac{\mathrm{M}[1]+\alpha_1}{\mathrm{M}[1]+\mathrm{M}[0]+\alpha_1+\alpha_0}$$

Cho ta biết xác suất trong $\alpha_1+\mathrm{M}[1]$ ra mặt ngửa và $\alpha_0+\mathrm{M}[0]$ ra mặt sấp.

- Với trường hợp $\mathrm{P}(\theta) = 1$, ta lại có:

$$\mathrm{P}(\mathrm{X}[\mathrm{M}+1] = \mathrm{x}^1|\mathrm{D}) = \frac{\mathrm{M}[1]+1}{\mathrm{M}[1]+\mathrm{M}[0]+2}$$

- Rõ ràng việc chọn xác suất tiên nghiệm có ảnh hưởng rất lớn đến tính toán phân phối xác suất.

b) Sự khác biệt giữa hai trường phái:

- Hai trường phái cổ điển và Bayes khác nhau về nhiều mặt.
- Mục tiêu:

❖ Trường phái cổ điển (hay còn được gọi là trường phái thường xuyên) có mục tiêu tạo ra những kết quả có thể áp dụng lâu dài, thường xuyên, không đổi.

❖ Trường phái Bayes tạo ra công cụ và mô hình có thể cập nhật "niềm tin" (tham số) một cách thường xuyên.

❖ Cả hai trường phái không đối lập nhau, đôi khi ta sẽ thấy tính chất "thường xuyên" của thống kê Bayes.

- Cách thức sử dụng các phương pháp tính xác suất có điều kiện:

❖Phương pháp cổ điển dùng xác suất có điều kiện để tìm ra những kết quả mang tính thường xuyên.

❖Phương pháp Bayes dùng xác suất có điều kiện để cập nhật những thay đổi về “niềm” (tham số).

Bảng so sánh:

	Bayesian	Frequentist
Probability	subjective degree of belief	limiting frequency
Goal	analyze beliefs	create procedures with frequency guarantees
θ	random variable	fixed
X	random variable	random variable
Use Bayes' theorem?	Yes. To update beliefs.	Yes, if it leads to procedure with good frequentist behavior. Otherwise no.

Ví dụ:

- Để thấy rõ sự khác biệt giữa hai quan điểm trên, ta có một vài ví dụ sau.
- Trong một hộp kẹo có rất nhiều viên kẹo đỏ, và chỉ có một số rất ít viên kẹo xanh bị trộn lẫn trong đó. Liên tục lấy kẹo ra ngoài, ta hãy tự ước lượng xác suất lấy được viên kẹo xanh
- Trong quá trình lấy kẹo ra ngoài, những người theo trường phái cổ điển và những người theo trường phái bayes có cách suy nghĩ khác biệt.

Ví dụ:

- Trường phái cổ điển: Khi lấy ra được càng nhiều kẹo màu đỏ, thì xác suất lần tiếp theo lấy được kẹo màu đỏ càng giảm vì số kẹo đỏ giảm
- Trường phái Bayes: Khi lấy ra được càng nhiều kẹo màu đỏ, thì niềm tin lấy được kẹo đỏ càng tang vì khi đó “tham số” $\theta = P(Kẹo\ \text{đỏ})$ sẽ rất cao.
- Bạn thuộc trường hợp nào ?