Xác suất có điều kiện - Công thức Bayes - Xác suất tiền, hậu nghiệm

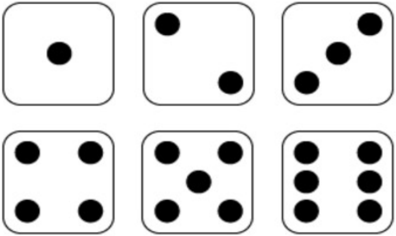
Trương Mạnh Tuấn
Nguyễn Nguyễn
Linh Trần

Ngày 5 tháng 8 năm 2019

Mục lục

Giới thiệu xác suất có điều kiện

Trong thực tế, việc xét các biến cố hoàn toàn độc lập với nhau là **lý tưởng** và rất khó đạt được.
Nếu ta bỏ qua sự cần thiết về tính phụ thuộc giữa các biến cố thì các kết quả sẽ bị ảnh hưởng thế nào?



Hình: Thí nghiệm tung xúc xắc

Xác suất của một biến cố sẽ bị ảnh hưởng thế nào nếu như một biến cố khác đã xảy ra?

Phép thử có điều kiện

Phép thử 1. Tung xúc xắc. $\Omega = \{1, 2, 3, 4, 5, 6\}$. Đặt 2 biến cố:

- $A = \{2, 3, 5\}$ (Ra được mặt là số nguyên tố). $\mathrm{P}(A) = 1/2$.
- $B = \{2, 4, 6\}$ (Ra được mặt chia hết cho 2). $\mathrm{P}(B) = 1/2$.

Phép thử có điều kiện

Phép thử 1. Tung xúc xắc. $\Omega = \{1, 2, 3, 4, 5, 6\}$. Đặt 2 biến cố:

- $A = \{2, 3, 5\}$ (Ra được mặt là số nguyên tố). $\mathrm{P}(A) = 1/2$.
- $B = \{2, 4, 6\}$ (Ra được mặt chia hết cho 2). $\mathrm{P}(B) = 1/2$.

Khi đó $AB = \{2\}$ và $\mathrm{P}(AB) = 1/6$.

Phép thử có điều kiện

Phép thử 1. Tung xúc xắc. $\Omega = \{1, 2, 3, 4, 5, 6\}$. Đặt 2 biến cố:

- $A = \{2, 3, 5\}$ (Ra được mặt là số nguyên tố). $\mathrm{P}(A) = 1/2$.
- $B = \{2, 4, 6\}$ (Ra được mặt chia hết cho 2). $\mathrm{P}(B) = 1/2$.

Khi đó $AB = \{2\}$ và $\mathrm{P}(AB) = 1/6$.

Lặp lại phép thử này nhiều lần: có khoảng $1/2$ số lần xảy ra A.

Phép thử có điều kiện

Phép thử 2. Tung 1 con xúc xắc.

- Ghi lại kết quả nếu kết quả là số chẵn (biến cố B).
- Bỏ kết quả đi nếu kết quả là số lẻ.

Phép thử có điều kiện

Phép thử 2. Tung 1 con xúc xắc.

- Ghi lại kết quả nếu kết quả là số chẵn (biến cố B).
- Bỏ kết quả đi nếu kết quả là số lẻ.

Phép thử 2 là *phép thử con* của phép thử 1, với *điều kiện* B.

Phép thử có điều kiện

Phép thử 2. Tung 1 con xúc xắc.

- Ghi lại kết quả nếu kết quả là số chẵn (biến cố B).
- Bỏ kết quả đi nếu kết quả là số lẻ.

Phép thử 2 là *phép thử con* của phép thử 1, với *điều kiện* B.

Lặp lại phép thử này nhiều lần: có khoảng $1/2$ số lần kết quả được ghi lại nhờ B xảy ra.

Phép thử có điều kiện

Phép thử 2. Tung 1 con xúc xắc.

- Ghi lại kết quả nếu kết quả là số chẵn (biến cố B).
- Bỏ kết quả đi nếu kết quả là số lẻ.

Phép thử 2 là *phép thử con* của phép thử 1, với *điều kiện* B.

Lặp lại phép thử này nhiều lần: có khoảng $1/2$ số lần kết quả được ghi lại nhờ B xảy ra. Trong đó, mỗi mặt 2, 4, và 6 ra được khoảng $1/3$ số lần.

$\Longrightarrow$ Khi áp đặt điều kiện B, xác suất A xảy ra (tức ra mặt 2) trở thành $1/3$.

Không gian mẫu con

Phép thử 2 có không gian mẫu $\{2, 4, 6\}$, chính là B.

Có thể xem B như *không gian mẫu con* của Ω.

Không gian mẫu con

Phép thử 2 có không gian mẫu $\{2, 4, 6\}$, chính là B.

Có thể xem B như *không gian mẫu con* của Ω.

Xác suất xảy ra biến cố A với điều kiện B là xác suất xảy ra A trên không gian mẫu con này. **Ký hiệu.** $\mathrm{P}(A \mid B)$.

Không gian mẫu con

Phép thử 2 có không gian mẫu $\{2, 4, 6\}$, chính là B.

Có thể xem B như *không gian mẫu con* của Ω.

Xác suất xảy ra biến cố A với điều kiện B là xác suất xảy ra A trên không gian mẫu con này. **Ký hiệu.** $\mathrm{P}(A \mid B)$.

Ở ví dụ này, $\mathrm{P}(A \mid B) = 1/3$ trong khi $\mathrm{P}(A) = 1/2$.

Xác suất có điều kiện

Cho một phép thử ngẫu nhiên và 2 biến cố A, B. Xác suất của A với điều kiện B, ký hiệu $\mathrm{P}(A \mid B)$ (còn gọi là xác suất của A, biết trước B) đo khả năng A xảy ra trong các tình huống có B xảy ra.

Công thức Bayes

Công thức Bayes

Cho một phép thử và 2 biến cố A, B, sao cho $\mathrm{P}(B) > 0$. Khi đó:

$$\mathrm{P}(A \mid B) = \frac{\mathrm{P}(AB)}{\mathrm{P}(B)}.$$

Công thức Bayes

Cho một phép thử và 2 biến cố A, B, sao cho $\mathrm{P}(B) > 0$. Khi đó:

$$\mathrm{P}(A \mid B) = \frac{\mathrm{P}(AB)}{\mathrm{P}(B)}.$$

Đẳng thức sau đúng với mọi A, B mà không cần điều kiện $\mathrm{P}(B) > 0$:

$$\mathrm{P}(A \mid B)\mathrm{P}(B) = \mathrm{P}(B \mid A)\mathrm{P}(A) = \mathrm{P}(AB).$$

Công thức Bayes - Ví dụ

Phép thử 1. Tung 1 con xúc xắc. $\Omega = \{1, 2, 3, 4, 5, 6\}$ với $A = \{2, 3, 5\}$ và $B = \{2, 4, 6\}$.

Công thức Bayes - Ví dụ

Phép thử 1. Tung 1 con xúc xắc. $\Omega = \{1, 2, 3, 4, 5, 6\}$ với $A = \{2, 3, 5\}$ và $B = \{2, 4, 6\}$. Khi đó:

- $\mathrm{P}(B) = 1/2$.
- $\mathrm{P}(A \mid B) = 1/3$.
- $AB = \{2\}$ và $\mathrm{P}(AB) = 1/6$.

Công thức Bayes - Ví dụ

Phép thử 1. Tung 1 con xúc xắc. $\Omega = \{1, 2, 3, 4, 5, 6\}$ với $A = \{2, 3, 5\}$ và $B = \{2, 4, 6\}$. Khi đó:

- $\mathrm{P}(B) = 1/2$.
- $\mathrm{P}(A \mid B) = 1/3$.
- $AB = \{2\}$ và $\mathrm{P}(AB) = 1/6$.

$$1/3 = \frac{1/6}{1/2} \implies \mathrm{P}(A \mid B) = \frac{\mathrm{P}(AB)}{\mathrm{P}(B)}.$$

Công thức Bayes - Ví dụ

Phép thử. Rút ngẫu nhiên 1 trong n lá thăm: $\Omega = \{1, 2, \ldots, n\}$.
A, B là 2 biến cố với $|A| = a, |B| = b$ và $|AB| = c \le \min\{a, b\}$.
Khi đó

Công thức Bayes - Ví dụ

Phép thử. Rút ngẫu nhiên 1 trong n lá thăm: $\Omega = \{1, 2, \ldots, n\}$.
A, B là 2 biến cố với $|A| = a, |B| = b$ và $|AB| = c \leq \min\{a, b\}$.
Khi đó

- $\mathrm{P}(B) = b/n$ và $\mathrm{P}(AB) = c/n$.
- Trong các trường hợp B xảy ra: việc A xảy ra $\iff$ AB xảy ra. Mỗi kết quả trong B đều có khả năng xảy ra như nhau

Công thức Bayes - Ví dụ

Phép thử. Rút ngẫu nhiên 1 trong n lá thăm: $\Omega = \{1, 2, \ldots, n\}$.
A, B là 2 biến cố với $|A| = a, |B| = b$ và $|AB| = c \leq \min\{a, b\}$.
Khi đó

- $\mathrm{P}(B) = b/n$ và $\mathrm{P}(AB) = c/n$.
- Trong các trường hợp B xảy ra: việc A xảy ra $\iff$ AB xảy ra. Mỗi kết quả trong B đều có khả năng xảy ra như nhau $\implies$ Xác suất xảy ra AB là c/b.

Công thức Bayes - Ví dụ

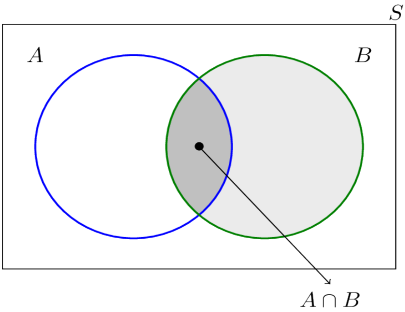
Phép thử. Rút ngẫu nhiên 1 trong n lá thăm: $\Omega = \{1, 2, \ldots, n\}$.
A, B là 2 biến cố với $|A| = a, |B| = b$ và $|AB| = c \leq \min\{a, b\}$.
Khi đó

- $\mathrm{P}(B) = b/n$ và $\mathrm{P}(AB) = c/n$.
- Trong các trường hợp B xảy ra: việc A xảy ra $\iff$ AB xảy ra.
  Mỗi kết quả trong B đều có khả năng xảy ra như nhau
  $\implies$ Xác suất xảy ra AB là c/b.

Do đó $\mathrm{P}(A \mid B) = \dfrac{c}{b} = \dfrac{c/n}{b/n} = \dfrac{\mathrm{P}(AB)}{\mathrm{P}(B)}$.

Công thức Bayes - Giải thích

Có thể dùng sơ đồ sau để giải thích ví dụ trên:



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

Hình: Biểu đồ Venn thể hiện định lý Bayes

Công thức xác suất đầy đủ

Cho $B_1, B_2, \ldots, B_n$ là một hệ biến cố đầy đủ, tức là:

1. $B_1 \vee B_2 \vee \ldots \vee B_n = \Omega$.
2. $B_i B_j = \varnothing$ với mọi $i \neq j$.

Công thức xác suất đầy đủ

Cho $B_1, B_2, \ldots, B_n$ là một hệ biến cố đầy đủ, tức là:

1. $B_1 \vee B_2 \vee \ldots \vee B_n = \Omega$.
2. $B_i B_j = \varnothing$ với mọi $i \neq j$.

A là một biến cố bất kì. Khi đó:

$$\mathrm{P}(A) = \mathrm{P}(A \mid B_1).\mathrm{P}(B_1) + \mathrm{P}(A \mid B_2).\mathrm{P}(B_2) + \cdots + \mathrm{P}(A \mid B_n).\mathrm{P}(B_n).$$

Công thức xác suất đầy đủ

Cho $B_1, B_2, \ldots, B_n$ là một hệ biến cố đầy đủ, tức là:

1. $B_1 \vee B_2 \vee \ldots \vee B_n = \Omega$.
2. $B_i B_j = \varnothing$ với mọi $i \neq j$.

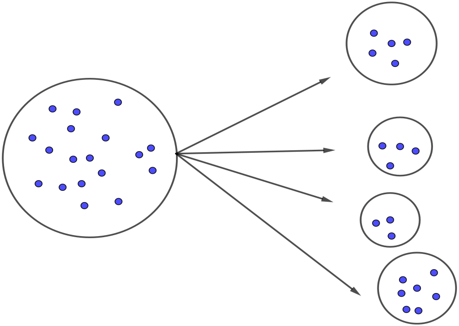
A là một biến cố bất kì. Khi đó:

$$\mathrm{P}(A) = \mathrm{P}(A \mid B_1).\mathrm{P}(B_1) + \mathrm{P}(A \mid B_2).\mathrm{P}(B_2) + \cdots + \mathrm{P}(A \mid B_n).\mathrm{P}(B_n).$$

$\Longrightarrow$ Nếu biết trước $\mathrm{P}(A \mid B_i)$ và $\mathrm{P}(B_i)$,
có thể tính $\mathrm{P}(A)$ ngay lập tức.

Ý nghĩa

Công thức Bayes có nhiều ý nghĩa về mặt xử lý tính phức tạp của dữ liệu thống kê.



Hình: Phân hoạch không gian mẫu

Bốc thăm

Có 3 chiếc thăm trong đó có 2 chiếc thăm có dấu X. Ba người chơi lần lượt sẽ bốc từng lá thăm (người đầu bốc rồi tới người thứ hai rồi thứ ba). Hỏi liệu khả năng bốc trúng lá thăm có dấu X của ba người là như nhau?

Bốc thăm

Gọi A_i lần lượt là biến cố bốc trúng lá thăm X của người bốc thứ i.

Bốc thăm

Gọi A_i lần lượt là biến cố bốc trúng lá thăm X của người bốc thứ i.

1. $P(A_1) = \frac{2}{3}$.

Bốc thăm

Gọi A_i lần lượt là biến cố bốc trúng lá thăm X của người bốc thứ i.

1. $P(A_1) = \frac{2}{3}$.
2. $P(A_2) = P(A_1)P(A_2|A_1) + P(\overline{A_1})P(A_2|\overline{A_1}) = \frac{2}{3}.\frac{1}{2} + \frac{1}{3}.1 = \frac{2}{3}$.

Bốc thăm

Gọi A_i lần lượt là biến cố bốc trúng lá thăm X của người bốc thứ i.

1. $P(A_1) = \frac{2}{3}$.
2. $P(A_2) = P(A_1)P(A_2|A_1) + P(\overline{A_1})P(A_2|\overline{A_1}) = \frac{2}{3}.\frac{1}{2} + \frac{1}{3}.1 = \frac{2}{3}$.
3. $P(A_3) = P(A_1\overline{A_2}) + P(\overline{A_1}) = P(A_1).P(\overline{A_2}|A_1) + P(\overline{A_1}) = \frac{2}{3}.\frac{1}{2} + \frac{1}{3} = \frac{2}{3}$

Bốc thăm

Gọi A_i lần lượt là biến cố bốc trúng lá thăm X của người bốc thứ i.

1. $P(A_1) = \frac{2}{3}$.
2. $P(A_2) = P(A_1)P(A_2|A_1) + P(\overline{A_1})P(A_2|\overline{A_1}) = \frac{2}{3}.\frac{1}{2} + \frac{1}{3}.1 = \frac{2}{3}$.
3. $P(A_3) = P(A_1\overline{A_2}) + P(\overline{A_1}) = P(A_1).P(\overline{A_2}|A_1) + P(\overline{A_1}) = \frac{2}{3}.\frac{1}{2} + \frac{1}{3} = \frac{2}{3}$

Vậy kết quả bốc thăm là **công bằng**. Nếu mở rộng, ta có thể thấy quá trình bốc thăm là công bằng vì nó không phân biệt thứ tự bốc.

Hỗn hợp hạt giống

Ví dụ. Anh Linh trộn 3 loại giống táo A, B, C với tỉ lệ nảy mầm lần lượt là $90\%, 80\%, 60\%$ vào hỗn hợp tỉ lệ tương ứng là $A : B : C = 8 : 3 : 1$ để bán. Anh Linh muốn tìm tỉ lệ nảy mầm của 1 hạt ngẫu nhiên trong hỗn hợp để ghi lên bao bì.

Hỗn hợp hạt giống

Ví dụ. Anh Linh trộn 3 loại giống táo A, B, C với tỉ lệ nảy mầm lần lượt là $90\%, 80\%, 60\%$ vào hỗn hợp tỉ lệ tương ứng là $A : B : C = 8 : 3 : 1$ để bán. Anh Linh muốn tìm tỉ lệ nảy mầm của 1 hạt ngẫu nhiên trong hỗn hợp để ghi lên bao bì.

Lời giải. Xác suất một hạt giống x nảy mầm:

$$\begin{aligned} \mathrm{P}(X) &= \mathrm{P}(X \mid A)\mathrm{P}(A) + \mathrm{P}(X \mid B)\mathrm{P}(B) + \mathrm{P}(X \mid C)\mathrm{P}(C) \\ &= 0.9 \times \frac{8}{12} + 0.8 \times \frac{3}{12} + 0.6 \times \frac{1}{12} = \frac{10.2}{12} = 0.85 \end{aligned}$$

Chú thích. $X =$"x nảy mầm", $A =$"x thuộc loại A".
Tương tự với B và C.

Xét nghiệm HIV

Máy xét nghiệm HIV có tỉ lệ chẩn đoán chính xác:

- Nếu bệnh nhân bị HIV: **99.7%** (xác suất dương tính là **99.7%**).
- Nếu bệnh nhân không bị HIV: **99%** (xác suất âm tính là **99%**).

Xét nghiệm HIV

Máy xét nghiệm HIV có tỉ lệ chẩn đoán chính xác:

- Nếu bệnh nhân bị HIV: **99.7%** (xác suất dương tính là **99.7%**).
- Nếu bệnh nhân không bị HIV: **99%** (xác suất âm tính là **99%**).

Tỉ lệ một người ngẫu nhiên bị bệnh HIV là **5%**.

Xét nghiệm HIV

Máy xét nghiệm HIV có tỉ lệ chẩn đoán chính xác:

- Nếu bệnh nhân bị HIV: **99.7**% (xác suất dương tính là **99.7**%).
- Nếu bệnh nhân không bị HIV: **99**% (xác suất âm tính là **99**%).

Tỉ lệ một người ngẫu nhiên bị bệnh HIV là **5**%.

1. Trung xét nghiệm 1 lần: **dương tính**. Xác suất Trung bị HIV?

Xét nghiệm HIV

Máy xét nghiệm HIV có tỉ lệ chẩn đoán chính xác:

- Nếu bệnh nhân bị HIV: **99.7%** (xác suất dương tính là **99.7%**).
- Nếu bệnh nhân không bị HIV: **99%** (xác suất âm tính là **99%**).

Tỉ lệ một người ngẫu nhiên bị bệnh HIV là **5%**.

1. Trung xét nghiệm 1 lần: **dương tính**. Xác suất Trung bị HIV?
2. Sau kết quả dương tính lần đầu, Trung không tin và xét nghiệm thêm 3 lần nữa: **dương tính - âm tính - dương tính**. Biết kết quả xét nghiệm các lần độc lập nhau, xác suất Trung bị HIV?

Xét nghiệm HIV

Máy xét nghiệm HIV có tỉ lệ chẩn đoán chính xác:

- Nếu bệnh nhân bị HIV: **99.7%** (xác suất dương tính là **99.7%**).
- Nếu bệnh nhân không bị HIV: **99%** (xác suất âm tính là **99%**).

Tỉ lệ một người ngẫu nhiên bị bệnh HIV là **5%**.

1. Trung xét nghiệm 1 lần: **dương tính**. Xác suất Trung bị HIV?
2. Sau kết quả dương tính lần đầu, Trung không tin và xét nghiệm thêm 3 lần nữa: **dương tính - âm tính - dương tính**. Biết kết quả xét nghiệm các lần độc lập nhau, xác suất Trung bị HIV?
3. Giả sử Trung làm xét nghiệm n lần, trong đó có k lần âm tính và $n-k$ lần dương tính. Tìm giá trị nhỏ nhất của k để Trung yên tâm rằng xác suất mình bị bệnh dưới 0.1%.

Xét nghiệm HIV

Ký hiệu:

- A là biến cố anh Trung bị HIV.
- B_1, B_2, B_3, B_4 là biến cố máy chẩn đoán anh Trung bị HIV (xét nghiệm 4 lần).
- $p = \mathrm{P}(B_i|A) = 0.997$ - xác suất chẩn đoán đúng nếu bị bệnh.
- $q = \mathrm{P}(B_i|\neg A) = 0.01$ - xác suất chẩn đoán sai nếu không bị bệnh.
- $r = \mathrm{P}(A) = 0.05$ - xác suất anh Trung bị bệnh khi không biết kết quả chẩn đoán.

Xét nghiệm HIV

Trung xét nghiệm 1 lần: **dương tính**. Tìm xác suất Trung bị HIV.

Xét nghiệm HIV

Trung xét nghiệm 1 lần: **dương tính**. Tìm xác suất Trung bị HIV.

Đầu tiên,

$$\mathrm{P}(B_1) = \mathrm{P}(B_1|A)\mathrm{P}(A) + \mathrm{P}(B_1|\neg A)\mathrm{P}(\neg A) = pr + q(1-r).$$

Như vậy

$$\mathrm{P}(A|B_1) = \frac{\mathrm{P}(B_1|A)\mathrm{P}(A)}{\mathrm{P}(B_1)} = \frac{pr}{pr + q(1-r)} \approx 0.8399$$

Xét nghiệm HIV

Sau kết quả dương tính lần đầu, Trung không tin và xét nghiệm thêm 3 lần nữa: **dương tính - âm tính - dương tính**. Biết kết quả xét nghiệm các lần là độc lập với nhau, tìm xác suất Trung bị HIV.

Xét nghiệm HIV

Sau kết quả dương tính lần đầu, Trung không tin và xét nghiệm thêm 3 lần nữa: **dương tính - âm tính - dương tính**. Biết kết quả xét nghiệm các lần là độc lập với nhau, tìm xác suất Trung bị HIV.

Đầu tiên,

$$\begin{aligned} \mathrm{P}(B_1B_2(\neg B_3)B_4) &= \mathrm{P}(B_1B_2(\neg B_3)B_4|A)\mathrm{P}(A) + \mathrm{P}(B_1B_2(\neg B_3)B_4|\neg A)\mathrm{P}(\neg \\ &= p^3(1-p)r + q^3(1-q)(1-r). \end{aligned}$$

Như vậy,

$$\mathrm{P}(A|B_1B_2(\neg B_3)B_4) = \frac{p^3(1-p)r}{p^3(1-p)r + q^3(1-q)(1-r)} \approx 0.9937$$

Xét nghiệm HIV

Giả sử Trung làm xét nghiệm n lần, trong đó có k lần âm tính và $n-k$ lần dương tính. Tìm giá trị nhỏ nhất của k để Trung yên tâm rằng xác suất mình bị bệnh dưới 0.1%.

Xét nghiệm HIV

Giả sử Trung làm xét nghiệm n lần, trong đó có k lần âm tính và $n-k$ lần dương tính. Tìm giá trị nhỏ nhất của k để Trung yên tâm rằng xác suất mình bị bệnh dưới 0.1%.

KMTTQ, giả sử k lần đầu ra âm tính và $n-k$ lần sau ra dương tính. Đặt $\vec{B} = (\neg B_1)(\neg B_2)\cdots(\neg B_k)B_{k+1}\cdots B_n$. Khi đó

$$\begin{aligned}\mathrm{P}(\vec{B}) &= \mathrm{P}(\vec{B} \mid A)\mathrm{P}(A) + \mathrm{P}(\vec{B} \mid \neg A)\mathrm{P}(\neg A)\\ &= p^{n-k}(1-p)^k + q^{n-k}(1-q)^k(1-r).\end{aligned}$$

$$\mathrm{P}(A \mid \vec{B}) = \frac{p^{n-k}(1-p)^k}{p^{n-k}(1-p)^k + q^{n-k}(1-q)^k(1-r)}$$

Công thức Bayes - Ứng dụng

$$\mathrm{P}(A \mid \vec{B}) = \frac{p^{n-k}(1-p)^k}{p^{n-k}(1-p)^k + q^{n-k}(1-q)^k(1-r)}$$

$$= \frac{1}{1 + \frac{q^{n-k}(1-q)^k(1-r)}{p^{n-k}(1-p)^k}} = \frac{1}{1 + TU^nV^k}$$

Với $T = \frac{1-r}{r} = 19$, $U = \frac{q}{p} = \frac{1}{99.7}$, $V = \frac{(1-q)p}{(1-p)q} = 32901.$

Như vậy

$$\mathrm{P}(A \mid \vec{B}) < 0.1\% \iff 19\left(\frac{1}{99.7}\right)^n 32901^k > 999$$

$$\iff k > n\log_{32901}(99.7) + \log_{32901}\left(\frac{999}{19}\right) \approx \boxed{0.44n + 0.38}$$

Tình huống

1. Tòa án liên bang xem xét một vụ án của anh K vì tội sàm sỡ bé gái trong thang máy. Tòa án tin rằng anh K có 85% khả năng có tội. Hỏi sự tin cậy của tòa án vào khả năng có tội của anh K sẽ như thế nào nếu camera trong thang máy cho thấy anh thực sự sàm sỡ bé gái?

Tình huống

1. Tòa án liên bang xem xét một vụ án của anh K vì tội sàm sỡ bé gái trong thang máy. Tòa án tin rằng anh K có 85% khả năng có tội. Hỏi sự tin cậy của tòa án vào khả năng có tội của anh K sẽ như thế nào nếu camera trong thang máy cho thấy anh thực sự sàm sỡ bé gái?
2. Có giả thuyết cho rằng có 40% dân số ở nước T mang trong mình nhóm máu O. Một vị bác sĩ địa phương tin rằng giả thuyết này có 80% khả năng đúng. Nhưng sau đó ông khám cho 10 bệnh nhân và thấy 8 bệnh nhân có nhóm máu O. Khi đó sự tin tưởng sẽ thay đổi thế nào?

Giả thuyết và quan sát

Công thức Bayes: $\mathrm{P}(A \mid B) = \dfrac{\mathrm{P}(B \mid A)\mathrm{P}(A)}{\mathrm{P}(B)}$.

Giả thuyết và quan sát

Công thức Bayes: $\mathrm{P}(A \mid B) = \dfrac{\mathrm{P}(B \mid A)\mathrm{P}(A)}{\mathrm{P}(B)}$.

- A là một **giả thuyết** về một thống kê nào đó. **Ví dụ:** Hơn 80% số trại sinh PiMA học giỏi toán.
- B là một **kết quả quan sát thực tế** về thống kê đó. **Ví dụ:** Trong 5 trại sinh được chọn ngẫu nhiên, có 3 bạn giỏi toán.

Xác suất tiền nghiệm, hậu nghiệm

Trong công thức Bayes:

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

Khi A là giả thuyết và B là kết quả quan sát được

Xác suất tiền nghiệm, hậu nghiệm

Trong công thức Bayes:

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

Khi A là giả thuyết và B là kết quả quan sát được , ta có:

- $\mathrm{P}(A)$: **Xác suất tiền nghiệm** (trước thử nghiệm) (**Prior probability**) của A: *niềm tin ban đầu* của người quan sát về khả năng A đúng.
- $\mathrm{P}(A \mid B)$: **Xác suất hậu nghiệm** (sau thử nghiệm) (**Posterior probability**) của A: *niềm tin sau khi quan sát* về khả năng A đúng.

Ví dụ: Trại sinh giỏi toán

Giả thuyết A: ít nhất 80% học sinh PiMA 2018 giỏi toán, tương đương 23/28 bạn.

Trung ban đầu tin rằng khả năng A xảy ra là 90%.

Ví dụ: Trại sinh giỏi toán

Giả thuyết A: ít nhất 80% học sinh PiMA 2018 giỏi toán, tương đương 23/28 bạn.

Trung ban đầu tin rằng khả năng A xảy ra là 90%.

Trung hỏi 5 bạn trại sinh ngẫu nhiên và đánh giá có 3 bạn giỏi toán.

Ví dụ: Trại sinh giỏi toán

Giả thuyết A: ít nhất 80% học sinh PiMA 2018 giỏi toán, tương đương 23/28 bạn.

Trung ban đầu tin rằng khả năng A xảy ra là 90%.

Trung hỏi 5 bạn trại sinh ngẫu nhiên và đánh giá có 3 bạn giỏi toán. Gọi biến cố B là "Có không quá 3 bạn giỏi toán trong 5 bạn được chọn", ta có:

Ví dụ: Trại sinh giỏi toán

Giả thuyết A: ít nhất 80% học sinh PiMA 2018 giỏi toán, tương đương 23/28 bạn.

Trung ban đầu tin rằng khả năng A xảy ra là 90%.

Trung hỏi 5 bạn trại sinh ngẫu nhiên và đánh giá có 3 bạn giỏi toán. Gọi biến cố B là "Có không quá 3 bạn giỏi toán trong 5 bạn được chọn", ta có:

$$\mathrm{P}(B \mid A) \leq \frac{\sum_{i=0}^{3} \mathrm{C}_{23}^{i} \mathrm{C}_{5}^{5-i}}{\mathrm{C}_{28}^{5}} \approx 0.207.$$

Ví dụ: Trại sinh giỏi toán

Nếu kết quả khảo sát các năm trước cho thấy $\mathrm{P}(B) \approx 0.4$, ta có:

$$\mathrm{P}(A \mid B) = \frac{\mathrm{P}(B \mid A)\mathrm{P}(A)}{\mathrm{P}(B)} \leq \frac{0.207 \cdot 0.9}{0.4} \approx 0.466$$

Như vậy sau khi quan sát, niềm tin của Trung đã giảm khoảng một nửa và thấp hơn 50%.

Định nghĩa

Với 2 biến ngẫu nhiên X, Y trong đó Y phụ thuộc vào X. Ta biết trước:

1. Phân phối $p_{Y|X=x}$ với mỗi giá trị x.
2. Phân phối p_X.

Định nghĩa

Nếu X và Y là rời rạc, ta có công thức:

$$\mathrm{P}(X = x \mid Y = y) = \frac{\mathrm{P}(Y = y \mid X = x)\mathrm{P}(X = x)}{\mathrm{P}(Y = y)}, \quad \forall x, y$$

Định nghĩa

Nếu X và Y là rời rạc, ta có công thức:

$$\mathrm{P}(X = x \mid Y = y) = \frac{\mathrm{P}(Y = y \mid X = x)\mathrm{P}(X = x)}{\mathrm{P}(Y = y)}, \quad \forall x, y$$

hay

$$p_{X|Y}(x \mid y) = \frac{p_{Y|X}(y \mid x)p_X(x)}{p_Y(y)}, \quad \forall x, y,$$

Định nghĩa

Nếu X và Y là rời rạc, ta có công thức:

$$\mathrm{P}(X = x \mid Y = y) = \frac{\mathrm{P}(Y = y \mid X = x)\mathrm{P}(X = x)}{\mathrm{P}(Y = y)}, \quad \forall x, y$$

hay

$$p_{X|Y}(x \mid y) = \frac{p_{Y|X}(y \mid x)p_X(x)}{p_Y(y)}, \quad \forall x, y,$$

Trong đó, $p_Y(y)$ có thể được tính như sau:

$$p_Y(y) = \sum_x p_{Y|X}(y \mid x)p_X(x) \quad \forall y$$

Định nghĩa

Nếu X và Y là liên tục, ta có:

$$p_{X|Y}(x \mid y) = \frac{p_{Y|X}(y \mid x) p_X(x)}{p_Y(y)}, \quad \forall x, y,$$

Định nghĩa

Nếu X và Y là liên tục, ta có:

$$p_{X|Y}(x \mid y) = \frac{p_{Y|X}(y \mid x)p_X(x)}{p_Y(y)}, \quad \forall x, y,$$

trong đó các hàm $p_{X|Y}, p_{Y|X}, p_Y, p_X$ lần lượt là hàm mật độ xác suất biên của X (biết Y), hàm mật độ xác suất biên của Y biết X, hàm mật độ xác suất của Y và hàm mật độ xác suất của X.

Định nghĩa

Nếu X và Y là liên tục, ta có:

$$p_{X|Y}(x \mid y) = \frac{p_{Y|X}(y \mid x) p_X(x)}{p_Y(y)}, \quad \forall x, y,$$

trong đó các hàm $p_{X|Y}, p_{Y|X}, p_Y, p_X$ lần lượt là hàm mật độ xác suất biên của X (biết Y), hàm mật độ xác suất biên của Y biết X, hàm mật độ xác suất của Y và hàm mật độ xác suất của X.

$p_Y(y)$ có thể được tính như sau:

$$p_Y(y) = \int_x p_{Y|X}(y \mid x) p_X(x) \quad \forall y$$

Một số khái niệm

Phân phối ứng với p_X là **Phân phối tiền nghiệm** (**Prior distribution**) của X, biểu thị *niềm tin ban đầu* về phân phối của X khi chưa có quan sát gì.

Phân phối ứng với $p_{X|Y}$ là **Phân phối hậu nghiệm** (**Posterior distribution**) của Y, biểu thị *niềm tin* về phân phối của X *sau khi quan sát* được Y.

Ví dụ: trại sinh PiMA giỏi toán

Anh Trung tin rằng xác suất để một bạn trại sinh bất kì giỏi toán là λ, với $\lambda \geq 0.8$, với phân phối tiền nghiệm của λ là $\mathrm{U}(0,1)$. Gọi số bạn giỏi toán trong trại là x,

[1]https://en.wikipedia.org/wiki/Beta_function

Ví dụ: trại sinh PiMA giỏi toán

Anh Trung tin rằng xác suất để một bạn trại sinh bất kì giỏi toán là λ, với $\lambda \geq 0.8$, với phân phối tiền nghiệm của λ là $U(0,1)$. Gọi số bạn giỏi toán trong trại là x, khi đó

$$
\begin{aligned}
P(\lambda \mid x) &= \frac{P(x \mid \lambda)P(\lambda)}{\int_0^1 P(x \mid \lambda)P(\lambda)} = \frac{C_{28}^x \lambda^x (1-\lambda)^{28-x}}{\int_0^1 C_{28}^x \lambda^x (1-\lambda)^{28-x} d\lambda} \\
&= \frac{\lambda^x (1-\lambda)^{28-x}}{\int_0^1 \lambda^x (1-\lambda)^{28-x} d\lambda}
\end{aligned}
$$

[1] https://en.wikipedia.org/wiki/Beta_function

Ví dụ: trại sinh PiMA giỏi toán

Anh Trung tin rằng xác suất để một bạn trại sinh bất kì giỏi toán là λ, với $\lambda \geq 0.8$, với phân phối tiền nghiệm của λ là $\mathrm{U}(0,1)$. Gọi số bạn giỏi toán trong trại là x, khi đó

$$P(\lambda \mid x) = \frac{P(x \mid \lambda)P(\lambda)}{\int_0^1 P(x \mid \lambda)P(\lambda)} = \frac{C_{28}^{x}\lambda^x(1-\lambda)^{28-x}}{\int_0^1 C_{28}^{x}\lambda^x(1-\lambda)^{28-x}d\lambda} = \frac{\lambda^x(1-\lambda)^{28-x}}{\int_0^1 \lambda^x(1-\lambda)^{28-x}d\lambda}$$

Hàm số ở mẫu số thực chất là hàm Beta[1] của x và $28-x$, ta có:

$$\int_0^1 \lambda^x(1-\lambda)^{28-x}d\lambda = \frac{1}{29C_{28}^{x}}$$

[1] https://en.wikipedia.org/wiki/Beta_function

Ví dụ: trại sinh PiMA giỏi toán

Do đó:

$$P(\lambda \mid x) = 29 C_{28}^{x} \lambda^{x} (1 - \lambda)^{28-x}$$

Ví dụ: trại sinh PiMA giỏi toán

Do đó:

$$P(\lambda \mid x) = 29 C_{28}^{x} \lambda^{x} (1-\lambda)^{28-x}$$

Giả sử thống kê cho thấy chỉ có 18 bạn giỏi toán, khi đó:

$$P(\lambda \mid x = 18) = 29 C_{28}^{18} \lambda^{18} (1-\lambda)^{10}$$

Ví dụ: trại sinh PiMA giỏi toán

Do đó:

$$P(\lambda \mid x) = 29C_{28}^{x}\lambda^{x}(1-\lambda)^{28-x}$$

Giả sử thống kê cho thấy chỉ có 18 bạn giỏi toán, khi đó:

$$P(\lambda \mid x = 18) = 29C_{28}^{18}\lambda^{18}(1-\lambda)^{10}$$

Xác suất để $\lambda \geq 0.8$ là:

$$P(\lambda \geq 0.8 \mid x = 18) = 29C_{28}^{18}\int_{0.8}^{1}\lambda^{18}(1-\lambda)^{10} \approx 0.02$$

Như vậy chỉ có khoảng 2% khả năng niềm tin của anh Trung là đúng.