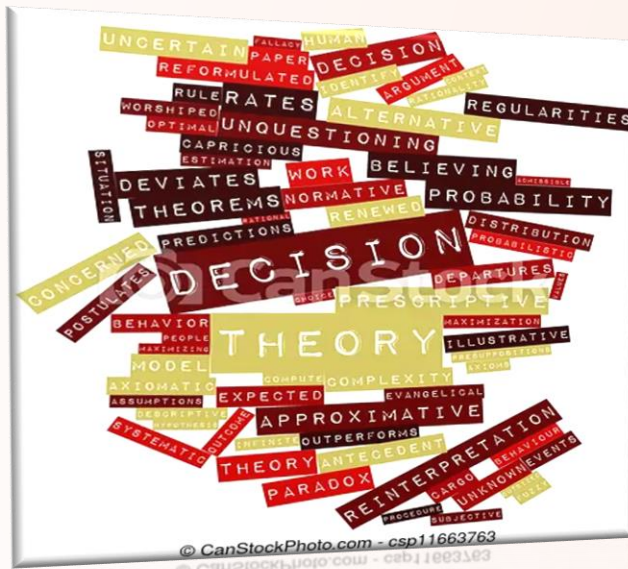


LÝ THUYẾT QUYẾT ĐỊNH

(Decision theory)



Biên soạn: ThS. Phạm Đình Tài
pdtai@ntt.edu.vn
0985.73.39.39

LUẬT NHÂN QUẢ & LÝ THUYẾT QUYẾT ĐỊNH CÓ BẰNG CHỨNG

- *Nhân quả và lý thuyết quyết định có bằng chứng*
- *Bài toán Newcomb*
- *Lý thuyết quyết định nhân quả*
- *Lý thuyết quyết định có bằng chứng*

1. Nhân quả

- *Luật nhân quả là một phép tắc, được quy định theo sự hoạt động tự nhiên của vũ trụ, để quân bình trật tự an toàn của các hành tinh trong không gian.*
- Nếu không có luật nhân quả quy định trật tự thì các hành tinh trong không gian này sẽ bị đảo lộn, và → muôn vật sống trên các hành tinh sẽ bị tiêu diệt.
- Trong một số vấn đề quyết định, niềm tin về các quá trình nhân quả đóng một vai trò quan trọng đối với những gì chúng ta trực giác nghĩ rằng nó là hợp lý để làm.
- Tuy nhiên, rất khó để đưa ra một giải trình thuyết phục về vai trò của niềm tin về quá trình nhân quả.

1. Nhân quả

❖ Logic ngầm ẩn của cách tiếp cận chính thống đối với suy luận nhân quả

- Các yếu tố của bất kỳ luận giải nhân quả nào, từ điểm xuất phát của mô hình chính thống, với hàm ý **C** gây ra **E** (viết tắt là **$C \Rightarrow E$**) và sau đó là một đồng biến được quan sát (sự kết hợp cố định) giữa **$C(t)$** và **$E(t+)$** , trong đó C và E xếp theo trật tự thời gian.
- Sự luận giải như thế rất hạn hẹp trong khuôn khổ của thuyết nội sinh, nếu suy luận nhân quả là xác thực trong mọi phương diện.
- Hơn nữa, một đồng biến có thể bị ẩn dấu nếu tương quan giả dương và âm của 2 biến số C và E tồn tại.
- *Như vậy, suy luận nhân quả ở khắp mọi nơi luôn là vị trí tạm thời, có trước của một sự vật có trước mang tính nhân quả chung và chưa được kiểm định (Goldthorpe, 2000).*

1. Nhân quả

❖ Một Tiếp cận cho Suy luận Nhân quả

- Các nghiên cứu trường hợp, khi một nghiên cứu có thể liên hệ với một hay một vài trường hợp của $C(t)$ và $E(t+1)$ có thể đóng bất kỳ vai trò có tính hệ thống nào trong suy luận nhân quả.

Bất kỳ phân tích nào về suy luận nhân quả trong các trường hợp nghiên cứu đều cần có các đặc trưng sau đây:

- *Có khả năng phân tích thuyết nhân quả tương tác phức hợp.*
- *Là ngẫu nhiên về bản chất,*
- *Sẽ cho thấy các cơ chế của các hành động (tương tác) tạo ra liên kết nhân quả đơn như thế nào.*

1. Nhân quả

❖ Ví dụ:

- Phương trình toán học đơn giản $y = f(x)$, trong đó:
- x là một biến số mà giá trị đã được xác định, nó đại biểu cho những việc đã xảy ra, những bài học thu nhận được v.v.
- Hàm $f()$ tượng trưng cho một hành động, hàm này nhận x làm tham số đầu vào.
- y là kết quả của biểu thức $f(x)$, và cũng là kết quả tạo thành sau khi hành động kể trên đã diễn ra.

1. Nhân quả

❖ Ví dụ:

- Nếu gọi đây là “**phương trình Luật Nhân Quả**”, \rightarrow thì nó có thể được phát biểu như sau:
 - ✓ Dựa vào kinh nghiệm và nền tảng trong quá khứ, một cá nhân làm một hành động gì đó trong hiện tại và lập tức những diễn biến trong tương lai được sinh ra.
 - ✓ Thay đổi giá trị của **x** sẽ làm thay đổi giá trị của **y**. Đưa hàm **f1()** thế vào chỗ của hàm **f()** cũng sẽ tạo ra một y mới. Cả **x**, **y**, **f()** đều tồn tại song song và có thể được quan sát cùng lúc.

1. Nhân quả

- Như vậy, khi một ai đó vừa thực hiện một hành động, thì hành động ấy sẽ được ghi nhận lại, và hậu quả của hành động ấy cũng như toàn bộ cuộc đời từ thời điểm đó về sau cũng được sinh ra và tồn tại sẵn giống như một bức tranh đã được vẽ xong.
- *Chẳng hạn một người nọ đánh người kia chính là hành động thể hiện giá trị của x ; tùy theo tính cách, mối quan hệ giữa hai người mà ta có được hàm $f()$ cho kết quả nhận được là y .*

Quyết định có bằng chứng

❖ Cốt lõi:

- Trong khi nhiều người xem xét các sự kiện và dữ liệu khi họ đưa quyết định, điều khác biệt giữa quyết định dựa trên bằng chứng với các phong cách quyết định khác là sự phụ thuộc vào bằng chứng lịch sử.
- Điều quan trọng là phải xem xét các tình huống mà quyết định bạn đang cân nhắc → thành công.

Quyết định có bằng chứng

❖ Các nguồn để lấy bằng chứng:

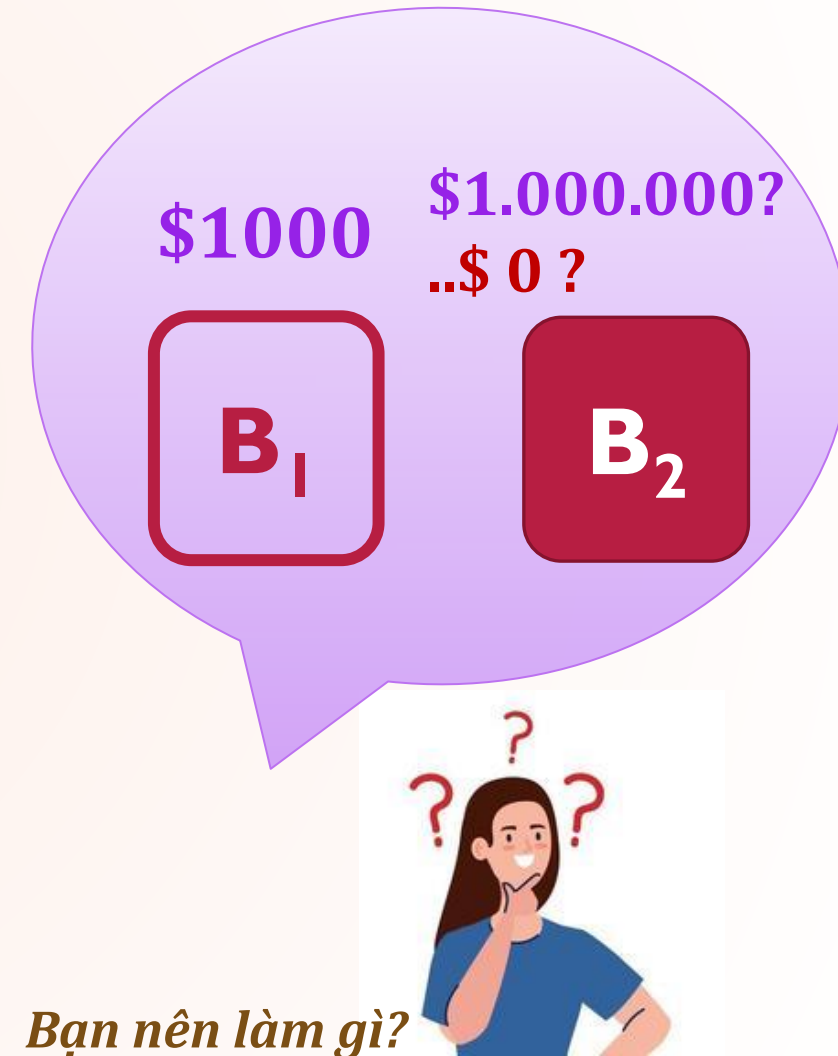
- **Dữ liệu thực nghiệm:** thu thập được thông qua các công cụ như báo cáo và hệ thống theo dõi
- **Ý kiến chuyên môn từ các chuyên gia:** quan sát và ý tưởng mới từ các chuyên gia lĩnh vực được tìm thấy tại các hội nghị, trong các ấn phẩm, trong các cuộc thảo luận của hội đồng hoặc từ các tổ chức chuyên nghiệp
- **Nghiên cứu từ các nguồn đáng tin cậy:** được kiểm duyệt xuất bản trên các tạp chí học thuật hoặc được thực hiện bởi các học giả có uy tín
- **Suy nghĩ và cảm nhận của các bên liên quan:** Sở thích của những người liên quan, như nhân viên và đối tác, mà họ có thể thể hiện bằng lời nói, văn bản hoặc thông qua các câu trả lời khảo sát

2. Bài toán Newcomb

Ví dụ:

Một người Dự đoán rất giỏi trong việc dự đoán, 99% tất cả các dự đoán được đưa ra từ trước đến nay đều đúng.

Bạn được cung cấp một sự lựa chọn giữa hai hộp, B_1 và B_2 . Hộp B_1 chứa 1.000\$ và bạn biết điều này(vì nó trong suốt). Hộp B_2 chứa 1 triệu \$ hoặc không có gì. Hộp này không trong suốt, vì vậy bạn không thể nhìn thấy nội dung của nó.

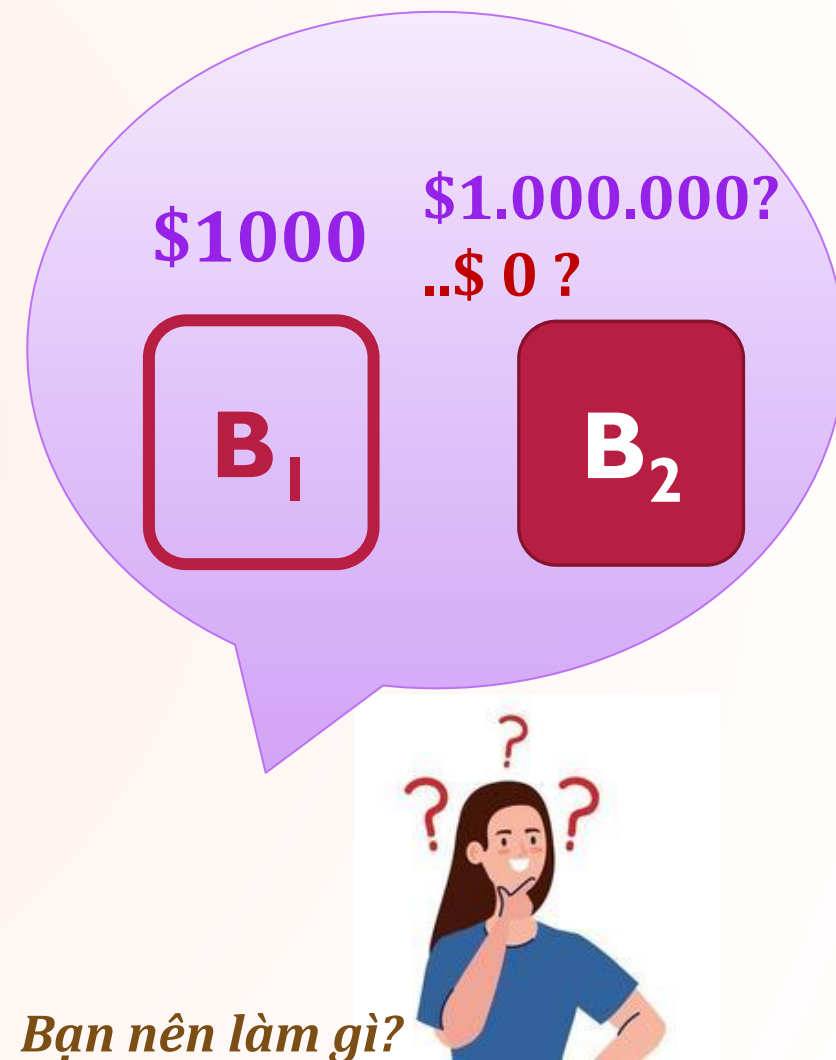


2. Bài toán Newcomb

Ví dụ:

Bây giờ bạn được mời lựa chọn giữa các cặp thay thế sau:

- Bạn lấy những gì có trong cả hai hộp hoặc chỉ lấy những gì trong hộp B_2
- Bạn được thông báo rằng người dự đoán sẽ bỏ 1 triệu \$ vào hộp B_2 nếu và chỉ khi người dự đoán cho rằng bạn sẽ chỉ lấy hộp B_2 và không có gì nếu là không, Người dự đoán biết rằng bạn biết điều này.
- Như vậy: Đầu tiên, đưa ra dự đoán của người đó, có thể đặt 1 triệu \$ hoặc không có gì vào B_2 , theo dự đoán của người dự đoán và sau đó bạn đưa ra lựa chọn của mình.



2. Bài toán Newcomb

Bài toán quyết định này lần đầu tiên được đề xuất bởi nhà vật lý William Newcomb vào những năm 1960. (*Nghịch lý dự đoán*)

Phương án 1	Lấy hộp B_1 (1.000 \$) và hộp B_2 (0 \$ hoặc 1 triệu \$).
Phương án 1	Chỉ lấy hộp B_2 (0\$ hoặc 1 triệu\$).

Bạn chọn phương án nào???

2. Bài toán Newcomb

Ứng dụng đơn giản của nguyên tắc tối đa hóa tiện ích mong đợi

	Hộp thứ hai chứa 1 triệu \$	Hộp thứ hai trống
Chỉ lấy hộp thứ hai	1 triệu \$ (xác suất 0,99)	\$ 0 (xác suất 0,01)
Lấy cả hai hộp	1 triệu \$ + 1.000 \$ (xác suất 0,01)	1.000 \$(tỷ lệ 0,99)

Giả định rằng tiện ích tiền tệ của bạn là tuyến tính, tức là giả định rằng đô la có thể được chuyển trực tiếp thành tiện ích. → tiện ích mong đợi của việc chỉ lấy hộp thứ hai là như sau:

$$\text{Lấy hộp 2: } 0,99 \times u(1 \text{ triệu } \$) + 0,01 \times u(0 \$) = 0,99 \times 1.000.000 + 0,01 \times 0 = 990.000.$$

Tuy nhiên, công dụng mong đợi của việc dùng cả hai hộp thấp hơn nhiều:

$$\text{Lấy hộp 1 và 2: } 0,01 \times u(1 \text{ triệu } \$) + 0,99 \times u(0 \$) = 0,01 \times 1.000.000 + 0,99 \times 0 = 10.000.$$

2. Bài toán Newcomb

- Rõ ràng, vì $990.000 > 10.000$, nguyên tắc tối đa hóa tiện ích mong đợi cho bạn biết rằng chỉ lấy hộp thứ hai là hợp lý.
- Do đó, hai nguyên tắc cơ bản nhất của lý thuyết quyết định
 - *nguyên tắc thống trị*
 - *nguyên tắc tối đa hóa tiện ích mong đợi*đưa ra các khuyến nghị trái ngược nhau.

Vậy một người ra quyết định hợp lý phải làm gì?

2. Bài toán Newcomb

- Gợi ý mà người ta không thể biết chắc chắn rằng xác suất là 0,99 rằng người dự đoán đã đưa ra dự đoán chính xác vì những người dự đoán tốt như vậy không tồn tại trong thế giới thực.
- Thoạt nhìn, đây có thể là một phản ứng hợp lý.
- Tuy nhiên, ai nghĩ xa hơn một chút sẽ thấy điểm này không liên quan.
- Các nhà lý thuyết quyết định tìm cách thiết lập các nguyên tắc chung nhất về tính hợp lý, và chắc chắn chúng ta có thể hình dung ra một sinh vật có sức mạnh tiên đoán rất ấn tượng

3. Lý thuyết quyết định nhân quả

- **Lý thuyết quyết định nhân quả** là quan điểm cho rằng một người ra quyết định hợp lý nên giữ cố định tất cả niềm tin của mình về các quá trình nhân quả trong quá trình ra quyết định và luôn chọn một giải pháp thay thế tối ưu theo những niềm tin này.
- Những người ra quyết định hợp lý nên làm bất cứ điều gì có khả năng mang lại kết quả tốt nhất như mong đợi, trong khi vẫn giữ cố định tất cả các quan điểm về cấu trúc nhân quả có thể có của thế giới?

Lý thuyết quyết định nhân quả

Ví dụ:

- Có một số khiếm khuyết di truyền được biết là gây ra cả ung thư phổi và thói quen hút thuốc, trái ngược với những gì mà hầu hết các nhà khoa học lý thuyết quyết định, bởi vì:
 - (i) một người có khiếm khuyết di truyền đó hoặc không?*
 - (ii) có một chút thích thú với việc hút thuốc*
 - (iii) xác suất ung thư phổi không bị ảnh hưởng bởi những lựa chọn.*
- Kết luận này phụ thuộc nhiều vào một giả định có phần kỳ quặc về cấu trúc nhân quả của thế giới, nhưng dường như không có gì sai với logic cơ bản. Nếu ung thư phổi và thói quen hút thuốc thực sự có cùng một nguyên nhân, thì hành động do nhà lý thuyết quyết định nhân quả chỉ định có vẻ hợp lý.

Lý thuyết quyết định nhân quả

Ví dụ:

- Gọi $X \rightarrow Y$ viết tắt mệnh đề. Nếu người ra quyết định làm X , thì Y sẽ là trường hợp, và đặt $p(X \rightarrow Y)$ thì biểu thị xác suất $X \rightarrow Y$ là đúng.
- Ví dụ: Trong trường hợp hút thuốc, các nhà lý thuyết quyết định nhân quả cho rằng $p(\text{Khói thuốc} \rightarrow \text{Ung thư})$ bằng $p(\neg \text{Khói thuốc} \rightarrow \text{Ung thư})$.
- Do đó, tốt hơn là “nên hút thuốc?”, vì một người sẽ nhận được thêm một khoản tiền thưởng nhỏ.
- Tất nhiên, có một mối tương quan thống kê chặt chẽ giữa hút thuốc và ung thư, nhưng theo giả định của ví dụ thì việc hút thuốc gây ra ung thư là sai.

4. Lý thuyết quyết định có bằng chứng

- Một số nhà lý thuyết quyết định cho rằng có những trường hợp lý thuyết quyết định nhân quả đưa ra các khuyến nghị phản trực giác.

Ví dụ:

Paul được cho biết rằng số lượng kẻ thái hóa nhân cách trên thế giới khá thấp. Kịch bản sau đây sẽ gây nghi ngờ về phân tích nhân quả. (*Egan 2007*)

Paul đang xác định cho dù đến nay đã định sẵn nút kiểm soát tất cả những kẻ thái hóa nhân cách. Anh ấy nghĩ sẽ tốt hơn nhiều khi được sống trong một thế giới không có kẻ tâm thần.

Thật không may, Paul khá tự tin rằng chỉ có một kẻ tâm thần mới nhấn nút như vậy.

- *Paul rất thích sống trong một thế giới có những kẻ tâm thần đến chết.*
- *Paul có nên bấm nút không?*

4. Lý thuyết quyết định có bằng chứng

- Theo trực giác, có vẻ hợp lý khi cho rằng Paul không nên nhấn nút.
- Nếu anh ta ép, có khả năng anh ta sẽ chết, vì gần như tất cả những người gây áp lực đều là một kẻ tâm thần.
- Tuy nhiên, lý thuyết nhân quả quyết định không chính xác ngụ ý rằng Paul nên nhấn nút
- Để biết lý do tại sao, hãy xem rằng $p(\text{nhấn nút} \rightarrow \text{chết})$ thấp hơn nhiều so với $p(\text{nhấn nút} \rightarrow \text{sống trong một thế giới không có kẻ thái hóa nhân cách})$.
- Điều này là do Paul có hoặc không phải là một kẻ thái hóa nhân cách, và xác suất của hai khả năng xảy ra không phụ thuộc vào những gì anh ta quyết định làm.

4. Lý thuyết quyết định có bằng chứng

Nhấn nút	$p(\text{nhấn nút} \rightarrow \text{chết}) \cdot u(\text{chết}) + p(\text{nhấn nút} \rightarrow \text{sống}) \cdot u(\text{sống}) = (0,001 \cdot (-100)) + (0,999 \cdot 1) = 0,899$
Không nhấn nút	$p(\text{không nhấn nút} \rightarrow \text{sống}) \cdot u(\text{sống}) = 1 \cdot 0 = 0$

- Những người theo thuyết quyết định hiển nhiên đồng ý với những người theo thuyết quyết định nhân quả rằng việc nhấn nút không gây ra bất kỳ bệnh tâm thần nào, nhưng nếu bạn nhấn nút, bạn sẽ gián tiếp biết được điều gì đó về bản thân mà bạn chưa biết, cụ thể là bạn là một kẻ thái nhân cách.
- Phần thông tin bổ sung này không thể được tính theo bất kỳ cách hợp lý nào bởi nhà lý thuyết quyết định nhân quả.

4. Lý thuyết quyết định có bằng chứng

- Về lý thuyết quyết định hiển nhiên, nó có thể được định nghĩa là khẳng định rằng không phải các xác suất như $p(X \rightarrow Y)$ sẽ hướng dẫn các quyết định đó, mà là các xác suất chẳng hạn như $p((X \rightarrow Y) | X)$.
- Nghĩa là, thay vì tự hỏi bản thân, xác suất nếu làm X, thì Y sẽ là trường hợp nào ?, một người ra quyết định hợp lý nên hỏi, xác suất mà nếu làm X, thì Y sẽ là bao nhiêu?

4. Lý thuyết quyết định có bằng chứng

- Việc gán các xác suất cho các lựa chọn của chính mình lại là không mạch lạc?
- Đầu tiên, nếu các xác suất được coi là khách quan thì có vẻ như rất khó để dung hòa ý tưởng của một người ra quyết định đưa ra lựa chọn tự do với suy nghĩ rằng các lựa chọn của bạn bằng cách nào đó bị chi phối bởi các quá trình xác suất. Làm thế nào một cái gì đó có thể là kết quả của một quá trình cân nhắc và một sự kiện ngẫu nhiên?
- Thứ hai, các vấn đề tương tự cũng phát sinh nếu các xác suất được coi là chủ quan và được xác định theo sở thích đối với các cược.

4. Lý thuyết quyết định có bằng chứng

- Với khả năng tự dự đoán không phải là mối đe dọa thực sự đối với lý thuyết quyết định hiển nhiên.
 - i) *nhà lý thuyết ra quyết định hiển nhiên có thể tìm ra cách thể hiện quan điểm không cho rằng người ra quyết định có thể gán bất kỳ xác suất nào cho các lựa chọn của chính mình.*
 - ii) *cũng có thể lập luận rằng khó khăn với việc tự dự đoán các xác suất không quá nghiêm trọng như nó vẫn thường xảy ra.*

THANK YOU !

