

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM

Độc Lập – Tự Do – Hạnh Phúc

TRƯỜNG ĐẠI HỌC SÀI GÒN

MÔN MÁY HỌC

BÁO CÁO ĐỒ ÁN

MẠNG MÔ HÌNH BAYES SỬ DỤNG THƯ VIỆN BRMS

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM

Độc Lập – Tự Do – Hạnh Phúc

TRƯỜNG ĐẠI HỌC SÀI GÒN

MÔN MÁY HỌC

BÁO CÁO ĐỒ ÁN

MẠNG MÔ HÌNH BAYES SỬ DỤNG THƯ VIỆN BRMS

Học Viên: Dương Ngọc Vọng

MSHV: CH11191008

Giáo Viên: TS Trịnh Tấn Đạt

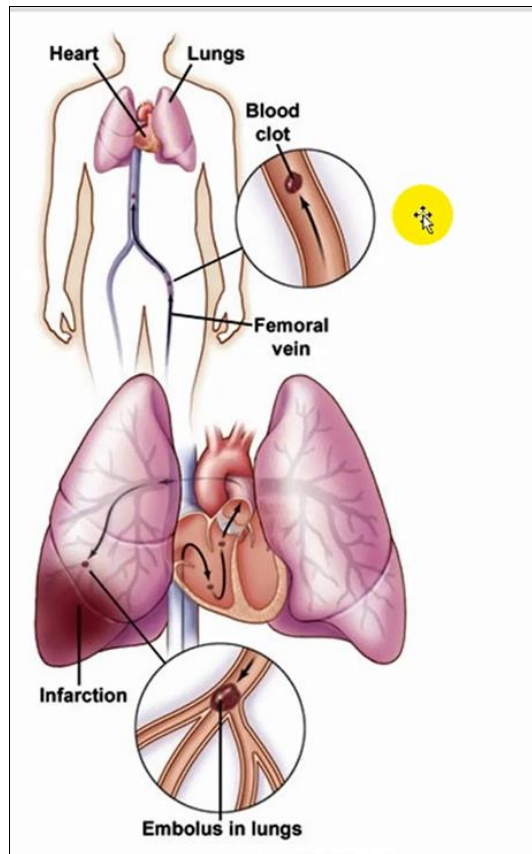
MỤC LỤC

I. Mô Tả Bài Toán.	2
1. Mở đầu.	2
2. Phát biểu bài toán.	4
3. Mô Tả Dữ Liệu.....	4
II. Giải Pháp Đề Xuất.	7
1. Định lý Bayes.....	7
1.1. Phát biểu định lý:	7
1.2. Định lý Bayes với hàm mật độ xác suất	8
1.3. Áp dụng mô hình mạng Bayesian Network (BN)	10
1.4. Lý thuyết xác suất.....	10
1.5. Mạng Bayesian	11
1.6. Suy diễn Bayesian.....	12
1.7. Ứng dụng	14
III. Tổng Kết.....	29
IV. Tài Liệu Tham Khảo.....	30

I. Mô Tả Bài Toán.

1. Mở đầu.

- Theo tài liệu về y khoa thì **thuyên tắc phổi** là tình trạng tắc động mạch phổi thường do huyết khối từ hệ tĩnh mạch sâu chi dưới.
- Huyết khối đã bị vỡ và trôi nổi tự do trong mạch máu có thể di chuyển đến một vùng khác của cơ thể và gây tắc nghẽn mạch máu tại đó. Có thể có một hoặc nhiều huyết khối.



Hình 1: Trạng thái bị thuyên tắc phổi

- Bình thường tất cả tĩnh mạch trong cơ thể dẫn máu đi vào tĩnh mạch lớn hơn, rồi dẫn máu đến tim phải và tiếp tục vào

động mạch phổi. Nếu có huyết khối trong hệ tĩnh mạch, huyết khối này sẽ di chuyển từ các tĩnh mạch sang tim phải rồi lại từ tim phải đi vào động mạch phổi chính và có thể bị mắc kẹt ở đó hoặc tiếp tục di chuyển vào một trong hai phổi.

- Khi huyết khối nằm trong động mạch phổi sẽ chặn lưu lượng máu đến phổi để nhận lấy oxy. Nếu không có đủ máu để nhận được oxy và di chuyển sang tim trái, nồng độ oxy trong cơ thể giảm xuống một cách nguy hiểm và có thể gây ra tổn thương cho tất cả các cơ quan trong cơ thể, kể cả não, thận và tim.
- Ngoài ra, do tắc nghẽn tại phổi làm tăng áp suất lên tim phải. Tim phải có thể bị phình to và co bóp nặng nề hơn, thậm chí chèn ép làm ảnh hưởng đến tim trái. Nếu tim trái không thể bơm đủ máu, huyết áp cũng giảm xuống. Tất cả những tác động này có thể dẫn đến tử vong đột ngột hoặc sau khi thuyên tắc phổi đã xảy ra một thời gian ngắn mà không được chữa trị.
- Thuyên tắc phổi là bệnh thường gặp và gây tử vong với tỉ lệ tử vong khoảng 30% nếu không được điều trị. Tuy nhiên tử suất có thể giảm đáng kể nhờ chẩn đoán và điều trị sớm.

2. Phát biểu bài toán.

- Bài toán cần giải quyết với việc xét nghiệm ddmer hoặc khám lâm sàng bằng cách sử dụng thang điểm lâm sàng để đánh giá, tiên lượng nguy cơ từ đó có thể đưa ra quyết định có áp dụng xét nghiệm về xạ hình động mạch phổi hay không.
- Việc đưa ra quyết định chẩn đoán sớm sẽ giúp bác sỹ can thiệp kịp thời nhằm cứu sống bệnh nhân.

3. Mô Tả Dữ Liệu.

```
> data(PE)
> PE%>%head()
```

	wells	pregnant	pe	angio	d.dimer	treat	death
1	3	No	No	Negative	206	No	No
2	2	No	No	Positive	180	Yes	No
3	3	No	No	Negative	197	No	No
4	3	No	No	Negative	192	Yes	No
5	2	No	No	Negative	281	Yes	No
6	6	No	Yes	Positive	318	Yes	No

Hình 2: Dữ liệu mẫu được trích xuất từ tập dataset PE thuộc thư viện Hydenet

- Bộ dữ liệu PE là một dataset với 1000 trường hợp (bệnh nhân), với 7 cột giá trị thuộc thư viện Hydenet được mô tả lần lượt như sau:

1. **wells** là 1 biến số rời rạc, trình bày điểm lâm sàng theo tiêu chí Wells.

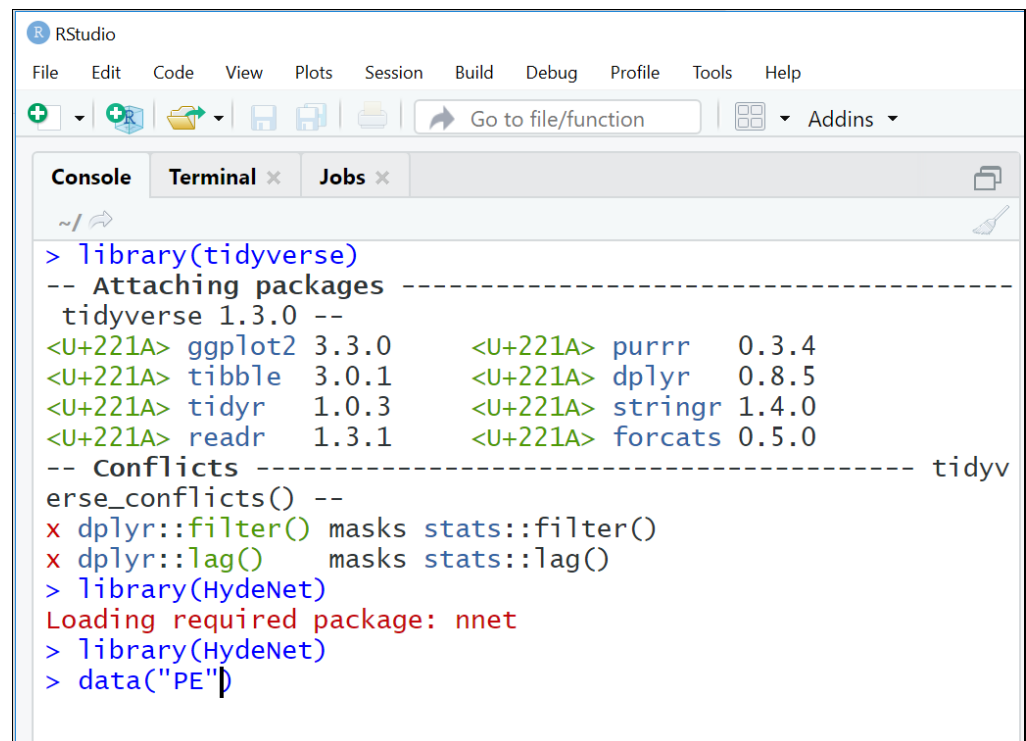
Wells' Criteria for Pulmonary Embolism ☆

Objectifies risk of pulmonary embolism.

When to Use ▾	Pearls/Pitfalls ▾	Why Use ▾
Clinical signs and symptoms of DVT	No 0	Yes +3
PE is #1 diagnosis OR equally likely	No 0	Yes +3
Heart rate > 100	No 0	Yes +1.5
Immobilization at least 3 days OR surgery in the previous 4 weeks	No 0	Yes +1.5
Previous, objectively diagnosed PE or DVT	No 0	Yes +1.5
Hemoptysis	No 0	Yes +1
Malignancy w/ treatment within 6 months or palliative	No 0	Yes +1

2. **pregnant** là 1 biến nhị phân, chỉ tình trạng Có/không khi bệnh nhân là nữ mang thai.
3. **pe** là 1 biến nhị phân, cho biết chẩn đoán của bệnh thuyên tắc phổi.
4. **angio** là 1 biến nhị phân, trình bày kết quả xét nghiệm xạ hình động mạch phổi với thuốc cản quang, với 2 giá trị Positive/Negative.
5. **d.dimer** là biến số liên tục, trình bày kết quả xét nghiệm định lượng D.dimer trong máu.

6. **treat** là 1 biến nhị phân, là quyết định điều trị : Có/không.
7. **death** là 1 biến nhị phân, cho biết kết cục của bệnh nhân: **No**=sống sót, **Yes**=Tử vong



```
> library(tidyverse)
-- Attaching packages -----
tidyverse 1.3.0 --
<U+221A> ggplot2 3.3.0      <U+221A> purrr 0.3.4
<U+221A> tibble 3.0.1      <U+221A> dplyr 0.8.5
<U+221A> tidyr 1.0.3       <U+221A> stringr 1.4.0
<U+221A> readr 1.3.1      <U+221A> forcats 0.5.0
-- Conflicts ----- tidyv
erse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag() masks stats::lag()
> library(HydeNet)
Loading required package: nnet
> library(HydeNet)
> data("PE")
```

Hình 3: Load thư viện HydeNet có tích hợp bộ dữ liệu PE

II. Giải Pháp Đề Xuất.

1. Định lý Bayes

1.1. Phát biểu định lý:

Định lý Bayes cho phép tính xác suất xảy ra của một sự kiện ngẫu nhiên A khi biết sự kiện liên quan B đã xảy ra. Xác suất này được ký hiệu là $P(A|B)$, và đọc là "xác suất của A nếu có B". Đại lượng này được gọi xác suất có điều kiện hay xác suất hậu nghiệm vì nó được rút ra từ giá trị được cho của B hoặc phụ thuộc vào giá trị đó.

Theo định lý Bayes, xác suất xảy ra A khi biết B sẽ phụ thuộc vào 3 yếu tố:

- Xác suất xảy ra A của riêng nó, không quan tâm đến B. Ký hiệu là $P(A)$ và đọc là xác suất của A. Đây được gọi là xác suất biên duyên hay xác suất tiên nghiệm, nó là "tiên nghiệm" theo nghĩa rằng nó không quan tâm đến bất kỳ thông tin nào về B.
- Xác suất xảy ra B của riêng nó, không quan tâm đến A. Ký hiệu là $P(B)$ và đọc là "xác suất của B". Đại lượng này còn gọi là hằng số chuẩn hóa (normalising constant), vì nó luôn giống nhau, không phụ thuộc vào sự kiện A đang muốn biết.
- Xác suất xảy ra B khi biết A xảy ra. Ký hiệu là $P(B|A)$ và đọc là "xác suất của B nếu có A". Đại lượng này gọi là khả năng (likelihood) xảy ra B khi biết A đã xảy ra. Chú ý không nhầm lẫn giữa khả năng xảy ra B khi biết A và xác suất xảy ra A khi biết B.

Khi biết ba đại lượng này, xác suất của A khi biết B cho bởi công thức:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{\text{likelihood} * \text{prior}}{\text{normalizing_constant}}$$

Từ đó dẫn tới

$$P(A|B)P(B) = P(A \cap B) = P(B|A)P(A)$$

- Các dạng khác của định lý Bayes
Định lý Bayes thường cũng thường được viết dưới dạng

$$P(B) = P(A, B) + P(A^C, B) = P(B|A)P(A) + P(B|A^C)P(A^C)$$

hay

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^C)P(A^C)},$$

trong đó A^C là biến cố bù của biến cố A (thường được gọi là "không A "). Tổng quát hơn, với $\{A_i\}$ tạo thành một phân hoạch của không gian các biến cố,

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_j P(B|A_j)P(A_j)},$$

với mọi A_i trong phân hoạch.

Công thức này còn được biết dưới tên công thức xác suất đầy đủ.

1.2. Định lý Bayes với hàm mật độ xác suất

Cũng có một dạng của định lý Bayes cho các phân bố liên tục. Đối với chúng, thay cho các xác suất trong định lý Bayes ta dùng hàm mật độ xác suất. Như vậy ta có các công thức tương tự định nghĩa xác suất điều kiện:

$$f(x|y) = \frac{f(y|x) f(x)}{f(y)}$$

và công thức tương tự công thức xác suất đầy đủ:

$$f(x|y) = \frac{f(y|x) f(x)}{\int_{-\infty}^{\infty} f(y|x') f(x') dx'}.$$

Ý nghĩa của các thành phần trong các công thức trên là $f(x, y)$ là mật độ phân phối của phân phối đồng thời của các biến ngẫu nhiên X và Y , $f(x|y)$ là mật độ phân phối xác suất hậu nghiệm của X với điều kiện $Y=y$, $f(y|x) = L(x|y)$ là (một hàm của x) hàm khả năng của X với điều kiện $Y=y$, và $f(x)$ và $f(y)$ là các mật độ phân phối của X và Y tách biệt nhau, với $f(x)$ là mật độ phân phối tiên nghiệm của X .

Điều kiện mặc định trong các công thức là hàm f khả vi và các tích phân công thức tồn tại.

Ứng dụng của định lý Bayes thường dựa trên một giả thiết có tính triết học Bayesian probability ngầm định rằng độ bất định và kỳ vọng có thể tính toán được giống như là xác suất. Định lý Bayes được đặt theo tên của Reverend Thomas Bayes (1702—1761), người nghiên cứu cách tính một phân bố với tham số là một phân bố nhị phân. Người bạn của ông, Richard Price, chỉnh sửa và giới thiệu công trình năm 1763, sau khi Bayes mất, với tựa đề *An Essay towards solving a Problem in the Doctrine of Chances*. Pierre-Simon Laplace mở rộng kết quả trong bài luận năm 1774.

1.3. Áp dụng mô hình mạng Bayesian Network (BN)

Mạng Bayesian là một mô hình xác suất đồ thị sử dụng suy diễn Bayesian để tính toán xác suất. Mục đích của mạng Bayesian là mô hình hóa các phụ thuộc có điều kiện của các biến ngẫu nhiên bằng cách biểu diễn các phụ thuộc đó dưới dạng các cạnh của đồ thị. Đồ thị của mạng Bayesian biểu diễn các mối quan hệ về xác suất và ta có thể sử dụng các quan hệ này để suy diễn và tính toán các biến ngẫu nhiên của đồ thị một cách hiệu quả.

1.4. Lý thuyết xác suất

Nhắc lại kiến thức về xác suất một chút, ta biết một phân phối xác suất chung của một tập các biến ngẫu nhiên **A1, A2, ..., An**, ký hiệu là **P(A1, A2, ..., An)** được tính bằng công thức:

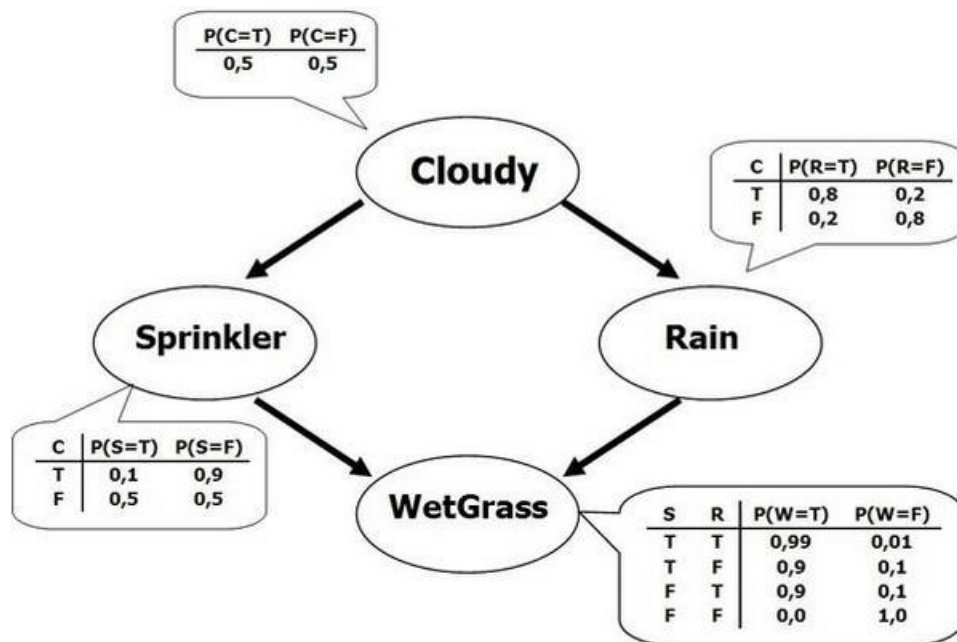
- $P(A1, A2, \dots, An) = P(A1 \mid A2, \dots, An) * P(A2 \mid A3, \dots, An) * \dots * P(An)$
theo quy tắc chuỗi trong xác suất. Tiếp theo, ta có: Mối quan hệ độc lập có điều kiện (conditional independence) giữa 2 biến ngẫu nhiên A và B khi có một biến ngẫu nhiên thứ 3 là C, được biểu diễn bằng công thức:
- $P(A, B \mid C) = P(A \mid C) * P(B \mid C)$

Ta nói, 2 biến ngẫu nhiên A và B độc lập với nhau với điều kiện biến C xảy ra. Khi ta biết biến C xảy ra, những thông tin về khi nào biến A xảy ra không cung cấp bất kỳ thông tin gì về biến B và ngược lại. Một công thức khác để biểu diễn quan hệ độc lập có điều kiện như sau:

- $P(A \mid B, C) = P(A \mid C)$

1.5. Mạng Bayesian

- Sử dụng các mối quan hệ xác suất trong mạng Bayesian, ta có thể biểu diễn các phân phối xác suất chung giữa các biến ngẫu nhiên một cách trực quan qua việc sử dụng các mối quan hệ độc lập có điều kiện. Dưới đây là một ví dụ của mạng Bayesian:



Hình 3: mô hình mạng BN

- Một mạng Bayesian là một đồ thị có chu kỳ với mỗi nút của mạng biểu diễn một biến ngẫu nhiên và mỗi cạnh biểu diễn mối quan hệ phụ thuộc giữa các biến. Một cạnh (A, B) trong đồ thị biểu diễn xác suất $P(B | A)$ và xác suất này là một hệ số trong phân phối xác suất chung của các biến.
- Mạng Bayesian thỏa mãn tính chất cục bộ của Markov (local Markov): Một nút trong đồ thị là độc lập có điều kiện với các nút không phải con của nút đó khi biết nút cha. Ở mạng trên,

ta có $P(\text{Sprinkler} \mid \text{Cloudy}, \text{Rain}) = P(\text{Sprinkler} \mid \text{Cloudy})$ vì biến ngẫu nhiên Sprinkler độc lập có điều kiện với biến ngẫu nhiên Rain - là nút không phải nút con của Sprinkler khi có Cloudy. Tính chất này khiến việc tính toán phân phối xác suất chung trở nên đơn giản hơn rất nhiều. Sau khi đơn giản hóa, phân phối xác suất chung được tính là tích của các phân phối $P(X \mid \text{Cha}(X))$ như sau:

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i \mid X_1, \dots, X_{i-1}) = \prod_{i=1}^n P(X_i \mid \text{Parents}(X_i))$$

với X_i là một biến ngẫu nhiên trong đồ thị. Với những mạng có kích thước lớn, tính chất này sẽ giảm thiểu chi phí tính toán đi rất nhiều.

1.6. Suy diễn Bayesian

Suy diễn trên một mạng Bayesian có thể thực hiện theo 2 dạng:

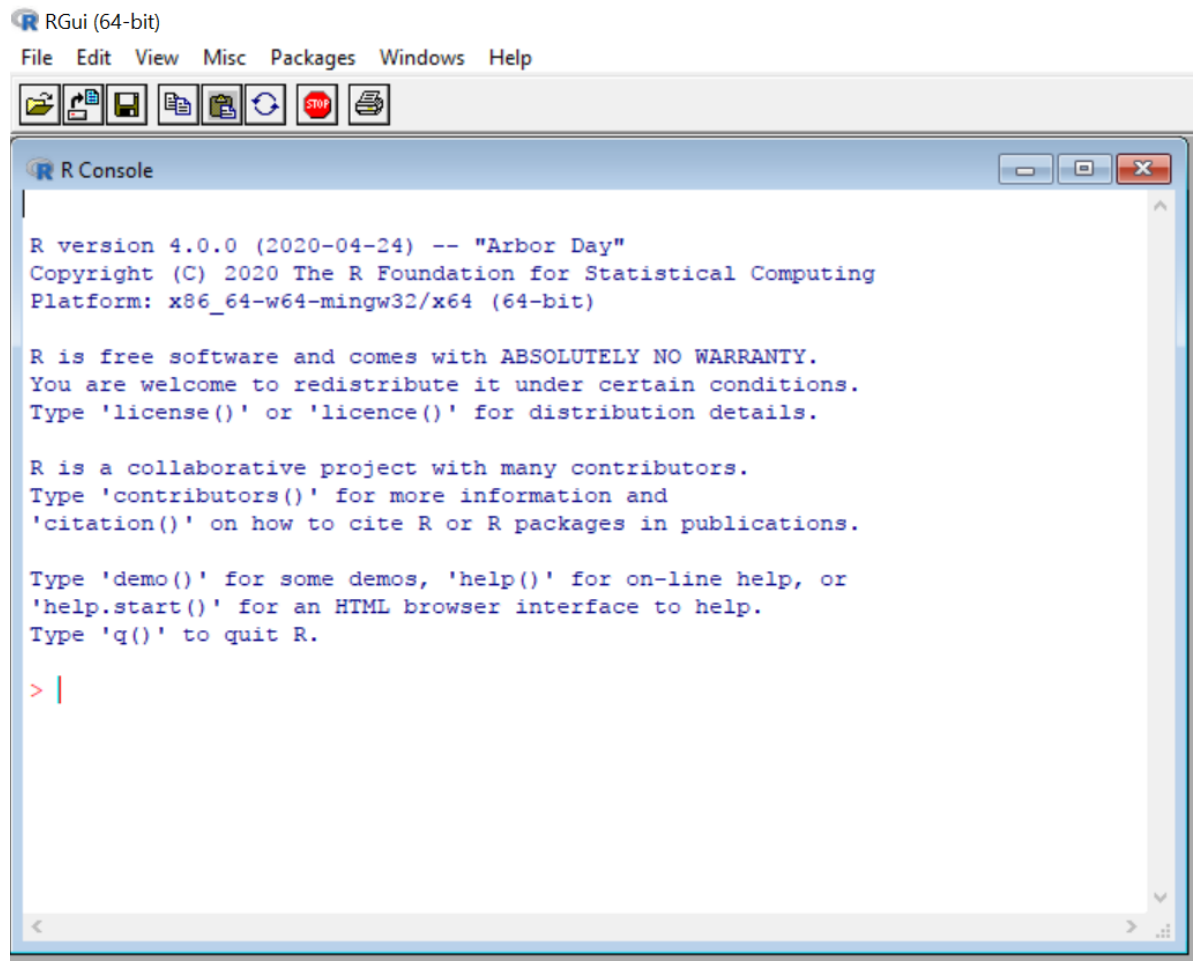
- Tính toán xác suất chung cho một mảng của đồ thị. Dựa theo công thức trên ta có thể tính toán xác suất này một cách dễ dàng dựa vào các xác suất có điều kiện. Nếu chỉ quan tâm tới một vài biến ngẫu nhiên nhất định, ta cần loại bỏ những biến không quan tâm đến.
- Tìm xác suất $P(X \mid E)$. xác suất xảy ra của (các) biến ngẫu nhiên X khi (các) biến ngẫu nhiên E xảy ra. Với mạng Bayesian như trên, ta có thể suy diễn xác suất $P(\text{Sprinkler}, \text{WetGrass} \mid \text{Cloudy})$ với $\{\text{Sprinkler}, \text{WetGrass}\}$ là X còn $\{\text{Cloudy}\}$ là E . Để tính xác suất này, ta sử dụng tính chất $P(X \mid E) = P(X, E) / P(E) = \alpha P(X, E)$ với α là một hằng số chuẩn hóa. Để tính xác suất $P(X \mid E)$, ta phải tính xác suất

chung trên các biến ngẫu nhiên không xuất hiện ở X và E .
Ta ký hiệu tập biến ngẫu nhiên đó là Y .

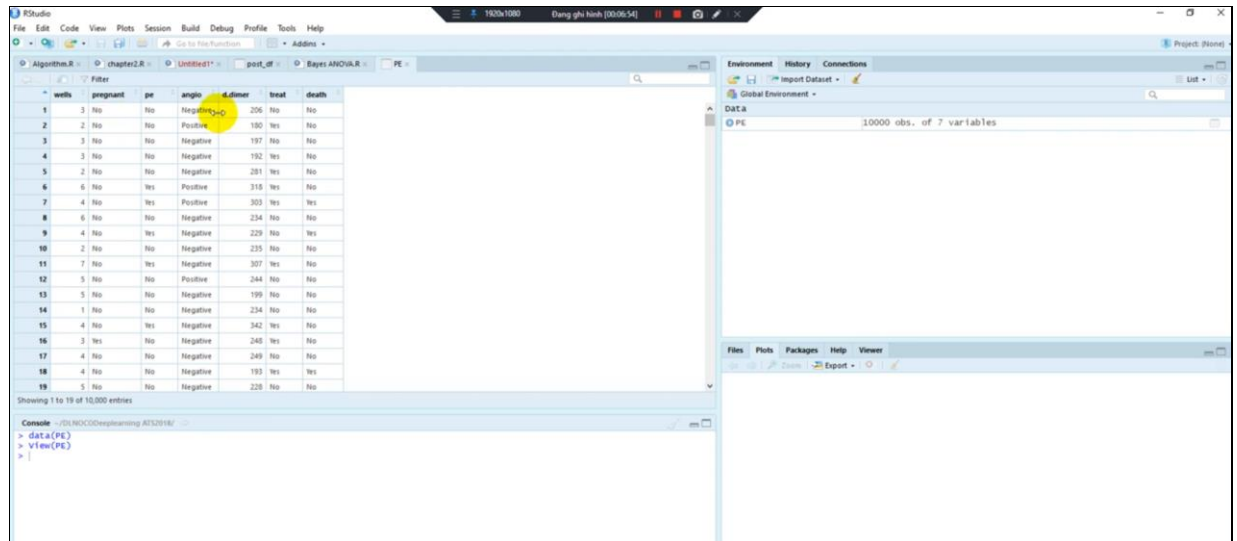
$$P(x|e) = \alpha \sum_{\forall y \in Y} P(x, e, Y)$$

1.7. Ứng dụng

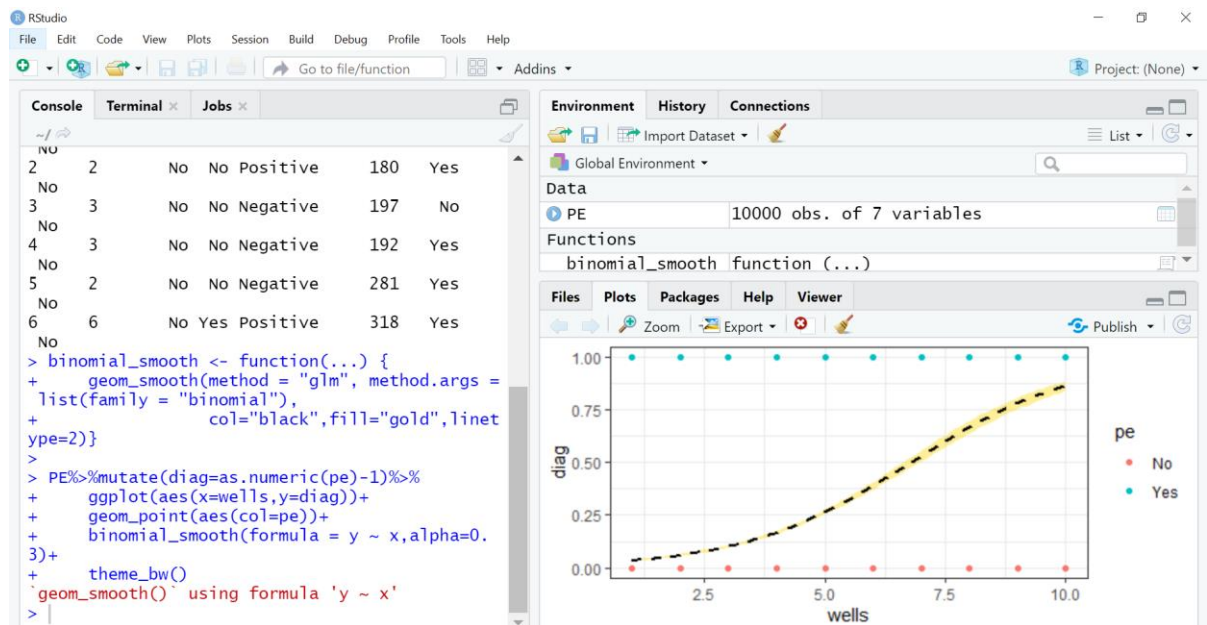
Bước 0: Cài đặt ngôn ngữ R và công cụ Rstudio và các packages cho phần mô phỏng.



Hình 4: Giao diện R



Hình 5: Hiển thị bộ dữ liệu bằng lệnh `>view(PE)`



Bước 1: Thiết kế cấu trúc mạng mô hình

- Với dữ liệu hiện thời, rất nhiều hướng nghiên cứu có thể đặt ra, với các giả thuyết/câu hỏi khác nhau. Thí dụ: Ta có thể dựng một mô hình nhằm tiên lượng nguy cơ thuyên tắc phổi (kết quả = biến `pe`) dựa vào các thông tin lâm sàng (điểm `wells`, kết quả xét nghiệm `d.dimer`, `angiography`); ta cũng có thể nhắm đến kết cục “tử vong” của bệnh nhân và khảo sát mối liên hệ giữa kết cục này và chẩn đoán (`pe`) và chỉ định

điều trị (treat); Cũng có thể ta muốn khảo sát sự ảnh hưởng của kết quả các xét nghiệm và điểm wells với quyết định điều trị của bác sĩ...

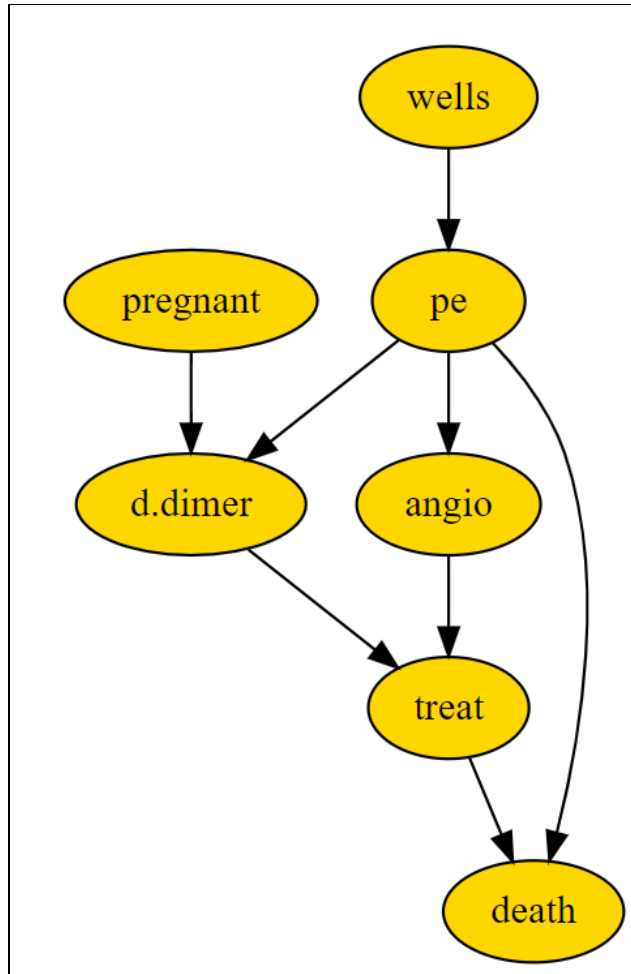
- Tuy nhiên, trong bài thực hành hôm nay ta sẽ đặt ra một cách tiếp cận vấn đề toàn diện và phức tạp hơn, với mục tiêu/ý tưởng:
 - Dựng mô hình cho phép khảo sát được toàn bộ (mạng lưới) những mối liên hệ / ảnh hưởng khả dĩ giữa các yếu tố: Bệnh lý (pe), kết quả xét nghiệm cận lâm sàng (d.dimer, angio), cảm nhận chủ quan (wells), quyết định điều trị (treat), và kết cục lâm sàng (nguy cơ tử vong) của bệnh nhân. Kết quả mô hình bao gồm chiều hướng/kích thước của mỗi hiệu ứng /liên hệ.
 - Mục tiêu này có thể hiểu như: tôi muốn dùng phân tích thống kê để suy diễn nhân/quả: giữa các yếu tố, và tôi muốn kể một câu chuyện có ý nghĩa dựa trên dữ liệu này.
 - Khác với những phương pháp phân tích mô hình đa biến quy ước - vốn thường quá chú trọng về thống kê, thí dụ chọn lọc mô hình / biến số dựa vào các chỉ số BIC, AIC (Stepwise) hoặc BMA, các phương pháp rút gọn mô hình như Lasso..., trong bài toán này chúng ta sẽ ưu tiên xây dựng mạng lưới tương tác giữa các yếu tố một cách chủ quan, dựa vào kiến thức y học, kinh nghiệm và quy luật hợp lý khác.
- Ta có thể bắt đầu việc này như sau:
 - Trước hết, mạng lưới mô hình của chúng ta có 7 nút , trong đó 2 nút quan trọng nhất là Chẩn đoán (pe) và Điều

trị (treat), nút tận cùng là kết cục: sống sót hay tử vong (death).

- Ở thời điểm nhập viện, chưa có xét nghiệm nào được làm, bệnh nhân được khai thác bệnh sử và khám bởi bác sĩ, do đó điểm khởi đầu của mạng lưới chính là nút wells. Như ta biết, Wells là một thang điểm lâm sàng tiên lượng nguy cơ thuyên tắc phổi, dựa vào các tiêu chí như bệnh sử (bất động kéo dài, phẫu thuật, bệnh lý tĩnh mạch, điều trị ung thư), triệu chứng (dấu hiệu huyết khối tĩnh mạch, ho ra máu, nhịp tim nhanh...). Ta có thể nói chẩn đoán lâm sàng (pe) phụ thuộc thang điểm Wells, từ đó vạch ra liên kết đầu tiên trong mạng lưới: pe | wells. Ta có thể trình bày quan hệ điều kiện này bằng một mô hình logistic đơn biến.
- Tiếp theo, ta xét nút D.dimer như biến kết quả. Có thể nói giá trị d.dimer phụ thuộc vào tình trạng có thuyên tắc phổi hay không, do đó có quan hệ giữa d.dimer và pe; tuy nhiên kết quả xét nghiệm d.dimer có thể bị sai lệch khi có thai (pregnant = Yes, với nguy cơ dương tính giả, do đó cả pe và pregnant cùng tham gia vào mô hình tiên lượng d.dimer.
- Kết quả xét nghiệm Angiography chỉ phụ thuộc vào sự hiện diện thuyên tắc phổi (pe) hay không, nên chỉ có 1 liên kết từ pe đến angio.
- Tiếp theo, ta xét đến nút: chỉ định điều trị (treat): Bệnh nhân chỉ được điều trị khi có xét nghiệm d.dimer và/hoặc angio dương tính, xác định chẩn đoán; do đó ta có thể

dùng một mô hình logistic khác với 2 biến là angio và d.dimer.

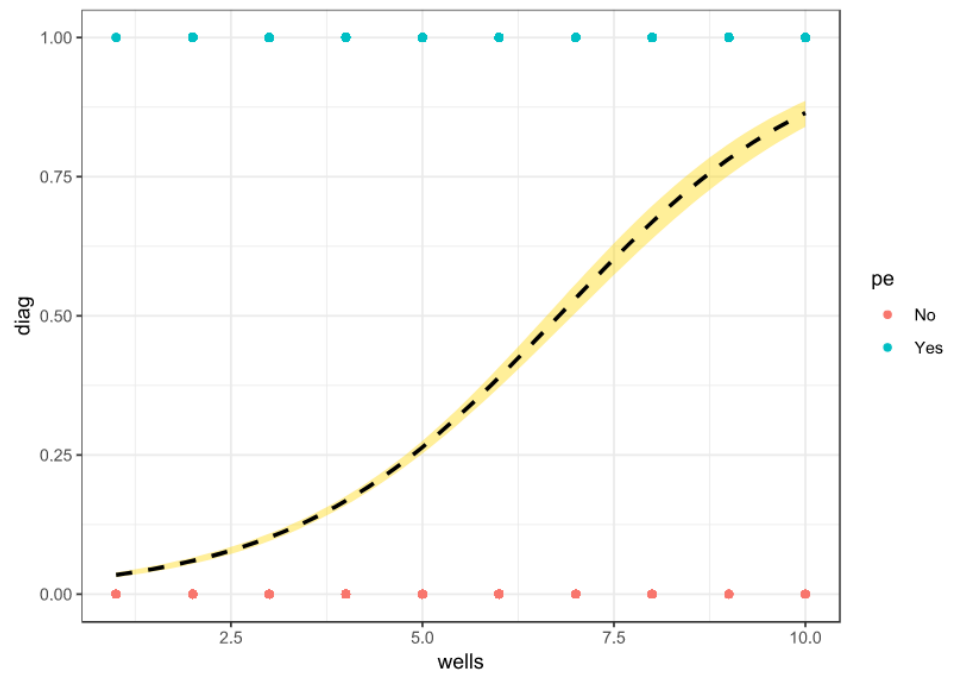
- Cuối cùng, nút tận cùng của mạng lưới là kết cục tử vong của bệnh nhân, đây là kết quả của sự tương tác giữa 2 điều kiện: bệnh lý thuyên tắc phổi (pe) và chỉ định điều trị (treat). Thí dụ: ở bệnh nhân có thuyên tắc phổi , nguy cơ tử vong có thể thấp hơn nếu được điều trị kịp thời, ngược lại nguy cơ tử vong cao nếu không được điều trị.
- Sau khi phác thảo mạng mô hình trên giấy, Nhi dùng package HydeNet để thiết kế cấu trúc của mạng lưới mô hình và vẽ sơ đồ :



Hình 6: Cấu trúc mạng mô hình

Bước 2: Chuyển sơ đồ mạng thành mạng mô hình Bayes

- Dựa vào sơ đồ mạng mô hình như trên, ta có thể hình dung về nó như một tập hợp 5 mô hình tương ứng với 5 nút kết quả: **pe**, **angio**, **d.dimer**, **treat** và **death**. Như vậy nếu ta xác định thành công 5 mô hình con này, xem như mục tiêu đã đạt được.
- Phân tích hồi quy với hàm GLM
 - Mô hình cho liên kết Wells - PE có dạng một mô hình logistic đơn biến, với giả định kết quả xác suất PE được ước tính bằng phân bố binomial, link function là hàm logit



term <chr>	OR <dbl>	LL <dbl>	UL <dbl>	p.value <dbl>
(Intercept)	0.02017027	0.01713912	0.0237375	0.000000e+00
wells	1.77837178	1.71825968	1.8405868	3.653918e-236

2 rows

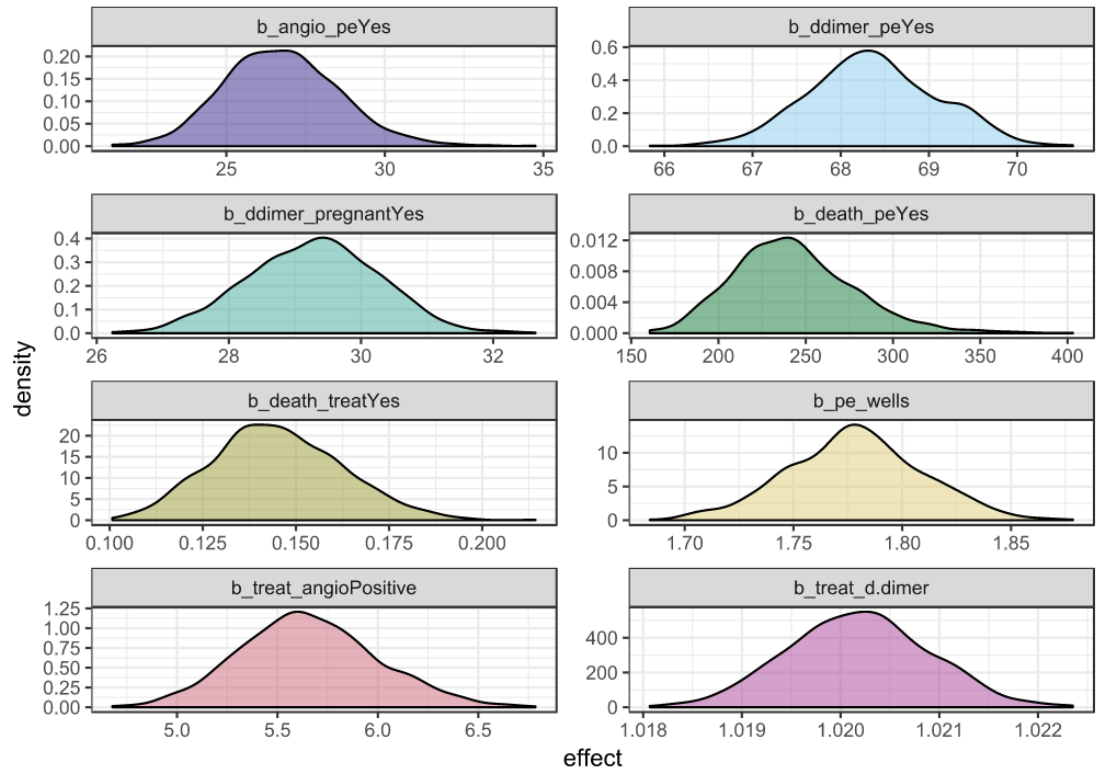
- Mạng mô hình Bayes sử dụng thư viện brms
 - Package brms là một công cụ rất mạnh mẽ, linh hoạt cho phân tích Bayes. Một trong những tính năng mà brms hỗ trợ là dựng mô hình mạng lưới cấu trúc (có thể gọi phương pháp này bằng nhiều cái tên: SEM, DAG, Latent variable analysis hay Bayesian network..., nhưng cách thực hiện cũng như nhau). Tính năng này có bản chất là ước lượng (MCMC sampling) đồng thời cho hàng loạt mô hình nằm trong một mạng lưới cấu trúc phức tạp. brms cho phép áp dụng priors, family và công thức riêng biệt

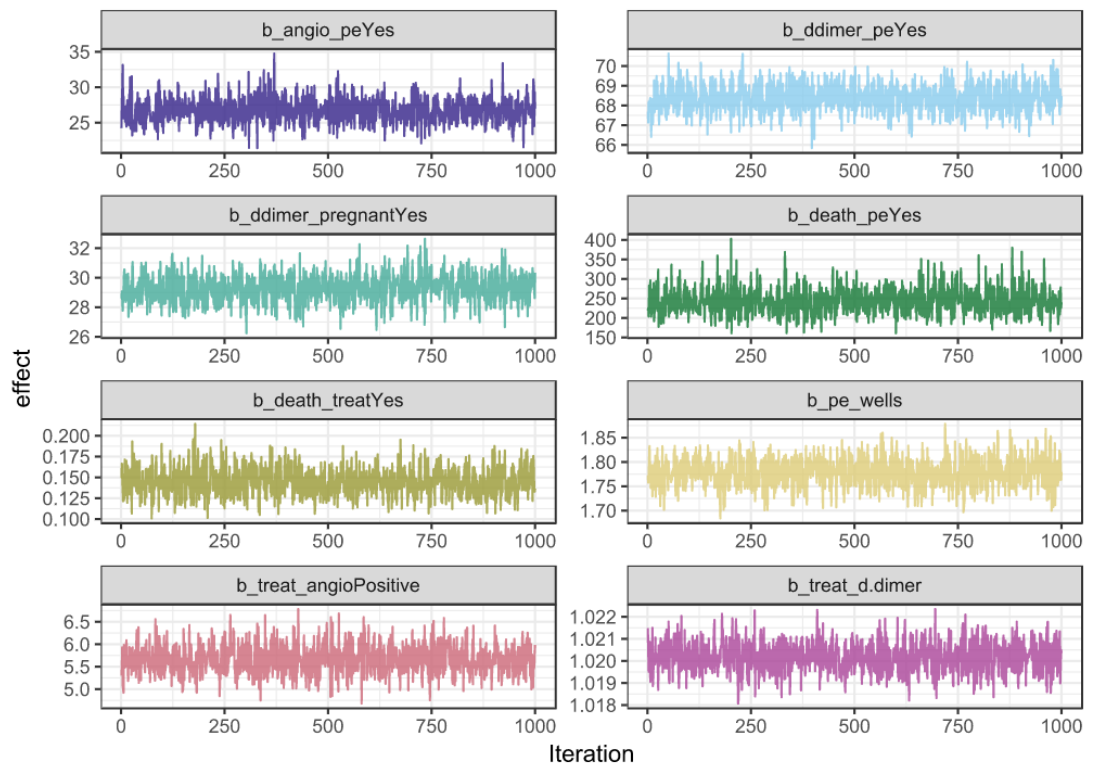
cho mỗi mô hình bên trong mạng lưới, do đó hết sức linh hoạt khi đối diện với các bài toán phức tạp.

Bước 3: Khảo sát nội dung mạng mô hình Bayes

- Trong mô phỏng này, để đơn giản vấn đề bằng cách chỉ áp dụng 2 kiểu mô hình: Logistic và Gaussian, và không tùy chỉnh sâu về priors. Trước hết, ta viết công thức cho 5 mô hình theo cú pháp brms:
- Kế đến, ta dùng hàm brm để đóng gói tất cả mô hình trong cùng 1 chế độ lấy mẫu MCMC: với 1 chuỗi, 1000 lượt khởi động, 2000 lượt lấy mẫu, sau đó rút gọn 50%.

- BRMS sẽ compile mô hình thành STAN code, sau đó thi hành. Tiến trình tương đối chậm, tùy vào cấu hình của máy tính mạnh hay yếu.

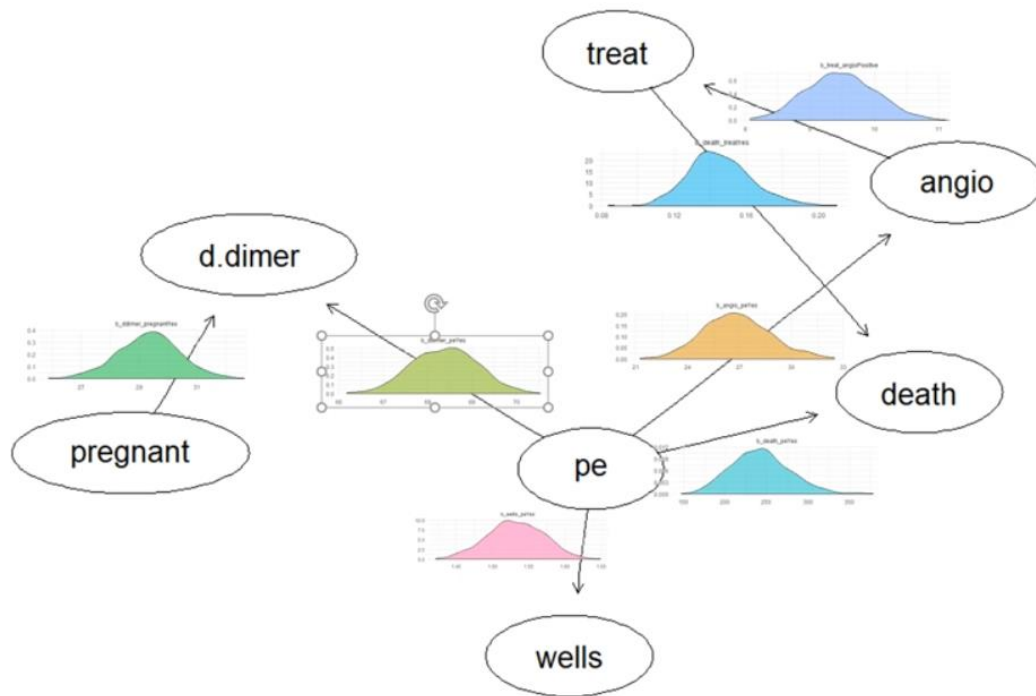




- Phân phối hậu nghiệm cho các điều kiện/hiệu ứng này hoàn toàn hợp lý, và có thể diễn giải như thường lệ, thí dụ: bệnh nhân bị thuyên tắc phổi có ddimer tăng trung bình 68 đơn vị, nguy cơ tử vong cao đến 240 lần. Việc điều trị làm giảm nguy cơ tử vong với Odds-ratio = 0.14, ...

Variable <chr>	Median <dbl>	SD <dbl>	LL <dbl>	UL <dbl>
b_angio_peYes	26.6753616	1.818773e+00	23.4351507	30.5190610
b_ddimer_peYes	68.3483742	7.250515e-01	66.9847062	69.7721663
b_ddimer_pregnantYes	29.3051568	9.874206e-01	27.3561911	31.1607824
b_death_peYes	239.7556499	3.477132e+01	184.3294615	320.6601411
b_death_treatYes	0.1442334	1.732946e-02	0.1139976	0.1808722
b_pe_wells	1.7778360	3.082602e-02	1.7154245	1.8387196
b_treat_angioPositive	5.6363352	3.398282e-01	5.0083865	6.3779393
b_treat_d.dimer	1.0201716	7.134562e-04	1.0188450	1.0215459
8 rows				

- Để trình bày nội dung mô hình mạng dưới dạng một sơ đồ DAG có kèm Odds-ratio(và hiệu ứng của mỗi mô hình:



Kết quả của việc ứng dụng mạng mô hình Bayes (Bayesian Network)

P/S : Khái niệm Odds Ratio và Risk(yếu tố nguy cơ)

1- Odds Ratio (OR)

- Nhiều công trình nghiên cứu lâm sàng đối chứng ngẫu nhiên (randomized controlled trial - RCT) thường có xu hướng báo cáo kết quả qua chỉ số **RR**, nhưng cũng có khi **OR** được sử dụng để mô tả ảnh hưởng của một thuật điều trị hay mối liên hệ giữa hai yếu tố. Sự lựa chọn này dẫn đến hiểu lầm rằng hai chỉ số này giống nhau, và sự hiểu lầm xảy ra ở ngay cả những nhà nghiên cứu có kinh nghiệm. Tuy nhiên, **OR** không có cùng ý nghĩa với **RR**.

Nói ngắn gọn, OR là một ước số của RR. Trong điều kiện tần số mắc bệnh thấp hay rất thấp (dưới 1%) thì OR và RR tương đương nhau, nhưng khi tần số mắc bệnh cao hơn 20% thì OR có xu hướng ước tính RR cao hơn thực tế.

2- Khái niệm nguy cơ (risk) và odds

- Trong y khoa, nguy cơ mắc bệnh thực chất là xác suất. Xác suất, như chúng ta biết, là một biến số giữa 0 và 1. Xác suất thực chất là tỉ lệ, tỉ số, và phần trăm. Do đó, thuật ngữ RISK trong y khoa có thể có nghĩa là xác suất, tỉ lệ lưu hành, hay tỉ lệ phát sinh.
- Cụm từ nguy cơ, dịch từ chữ risk trong tiếng Anh, có rất nhiều nghĩa trong y khoa. Cần phải phân biệt nguy cơ mắc bệnh và bệnh. Khi nói đến ung thư, chúng ta muốn nói đến một sự kiện cho một cá nhân; nhưng khi nói đến nguy cơ ung thư hay cancer risk, chúng ta nói đến nguy cơ xảy ra, nguy cơ phát sinh cho một cá nhân hay một quần thể. Xin nhắc lại, sự kiện khác với nguy cơ sự kiện. Do đó, ung thư khác với nguy cơ ung thư, vì ung thư là một sự kiện mang tính khẳng định (certainty), còn nguy cơ ung thư là một biến số liên tục mang tính bất định (uncertainty). Tất cả chúng ta trong bất cứ thời điểm nào đều có nguy cơ bị bệnh; nhưng có người có nguy cơ cao, có người có nguy cơ thấp.
- Trong tiếng Anh còn có một chữ nữa mà các ngôn ngữ khác như Pháp, Tây Ban Nha, Đức, và ngay cả tiếng Việt cũng không có: đó là chữ odds. Nếu nguy cơ bệnh nhân mắc

bệnh là p , thì có một cách nói khác rằng odds mà bệnh nhân đó mắc bệnh so với không mắc bệnh.

- Ví dụ: nếu nguy cơ bệnh nhân bị ung thư trong vòng 5 năm tới là 0.10 (tức 10%) thì odds mà bệnh nhân bị ung thư là $0.1 / (1 - 0.1) = 0.11$. Theo định nghĩa này odds không phải là nguy cơ hay risk.

3- OR và RR: cơ chế tính toán

- OR và RR là hai chỉ số thống kê rất phổ biến và có ích trong nghiên cứu lâm sàng, vì cả hai chỉ số kiểm định mối liên hệ giữa một yếu tố nguy cơ và bệnh tật - một mục tiêu gần như căn bản của nghiên cứu y học hiện đại. Cơ chế tính toán của hai chỉ số này cực kì đơn giản.
- Hãy tưởng tượng một công trình nghiên cứu RCT với 2 nhóm: nhóm được điều trị tích cực với một loại thuốc gồm n_1 bệnh nhân, và một nhóm chứng (placebo) gồm n_2 bệnh nhân. Sau một thời gian điều trị, có k_1 bệnh nhân trong nhóm được điều trị mắc bệnh, và k_2 bệnh nhân trong nhóm chứng mắc bệnh. Như vậy, tỉ lệ mắc bệnh của nhóm điều trị (kí hiệu p_1) và nhóm chứng (p_2) được ước tính như sau:
 - Nếu $RR > 1$ (hay $p_1 > p_2$), chúng ta có thể phát biểu rằng yếu tố nguy cơ làm tăng khả năng mắc bệnh; nếu $RR = 1$ (tức là $p_1 = p_2$), chúng ta có thể nói rằng không có mối liên hệ nào giữa yếu tố nguy cơ và khả năng mắc bệnh; và nếu $RR < 1$ (tức $p_1 < p_2$), chúng ta có bằng chứng để thể phát biểu rằng yếu tố nguy cơ có thể làm giảm khả năng mắc bệnh.

Tóm lại :

- Odd Ratio của một biến cố là tỉ số giữa số lần biến cố đó xảy ra và số lần biến cố đó không xảy ra.
- Cả hai *RR* và *OR* đều là những chỉ số phản ánh độ tương quan giữa một yếu tố nguy cơ và bệnh; nhưng *RR* mới là chỉ số chúng ta cần biết (còn *OR* chỉ là ước số của *RR*). Cần phải xác định rằng *odds* không phải là *risk* hay nguy cơ. Do đó, ý nghĩa của *OR* rất khó diễn giải. Đây chính là lí do mà một số nhà nghiên cứu đòi "tẩy chay" *OR* [1,2]. Nhưng vì tính nhất quán của *OR* so với *RR* nên việc sử dụng *OR* cần phải đặt vào bối cảnh nghiên cứu [3]. Trong nghiên cứu cắt ngang hay nghiên cứu theo thời gian, và khi tỉ lệ lưu hành hay tỉ lệ phát sinh bệnh cao thì nên tránh sử dụng *OR*
- Việc chọn *OR* và *RR* tùy theo mô hình nghiên cứu [5-7]. *OR* có thể sử dụng cho tất cả các nghiên cứu bệnh chứng (case-control study), cắt ngang (cross-sectional study), nghiên cứu theo dõi bệnh nhân theo thời gian (prospective study) kể cả nghiên cứu lâm sàng đối chứng ngẫu nhiên (RCT). *RR* chỉ có thể sử dụng cho các nghiên cứu theo dõi bệnh nhân theo thời gian và nghiên cứu lâm sàng đối chứng ngẫu nhiên. Đối với các nghiên cứu cắt ngang, *PR* thường được sử dụng để khắc phục những khó khăn trong diễn giải *OR*.
- Về mặt tính toán, không có gì sai khi một nghiên cứu cắt ngang hay theo thời gian sử dụng *OR*. Nhưng cần phải hết sức cẩn thận khi diễn giải *OR* trong các nghiên cứu cắt

ngang, vì OR tùy thuộc vào nguy cơ mắc bệnh (và khi nguy cơ mắc bệnh cao - như trên 10% - thì OR thường cao hơn so với thực tế). Do đó, các nghiên cứu cắt ngang ngày nay thường sử dụng prevalence ratio.

III. Tổng Kết.

- chúng ta vừa thử nghiệm tính năng mô hình cấu trúc đa biến của thư viện brms. Tính năng mang lại nhiều hiệu quả thiết thực, vì nó cho phép chúng ta lai ghép giữa thống kê Bayes và những phương pháp mô hình cấu trúc khác như SEM, DAG, Latent var. analysis, network ... với tính linh hoạt rất lớn, thậm chí hơn cả những thư viện sử dụng JAGs sampler.
- Về giá trị thực tiễn, phương pháp mạng mô hình Bayes (Bayesian network) mở ra nhiều ứng dụng tiềm năng cho nghiên cứu Y học lâm sàng, vì cho phép chúng ta kể một câu chuyện hấp dẫn và hợp lý với dữ liệu hiện có, suy diễn nhân quả, biện luận quy trình đưa ra quyết định chẩn đoán, phác đồ điều trị...bằng kết quả phân tích đa biến,dựa vào lý thuyết y học và tất cả những ưu thế mà trường phái Bayes có thể mang lại.

-

IV. Tài Liệu Tham Khảo.

1. https://rstudio-pubs-static.s3.amazonaws.com/406775_f181a39d745e4b1b80f45b87243c3abe.html<https://noron.vn/post/gioi-thieu-ve-mang-bayesian-5279034884347959>
2. <https://www.jw.org/vi/an-pham/tap-chi/g201507/b%E1%BB%87nh-s%E1%BB%91t-r%C3%A9t-l%C3%A0-g%C3%AC/>
3. <https://www.vinmec.com/vi/benh/thuyen-tac-phoi-4442/>
4. <https://dominhhai.github.io/vi/2018/04/nn-intro/>
5. https://vi.wikipedia.org/wiki/%C4%90%E1%BB%8Bnh_l%C3%BD_Bayes
6. <https://suckhoenoi tiet.vn/phan-biet-odds-ratio-relative-risk-trong-nghien-cuu-y-hoc.html>