

Đánh giá định lượng nguy cơ vi sinh vật Dự án SAFEGRO 2024

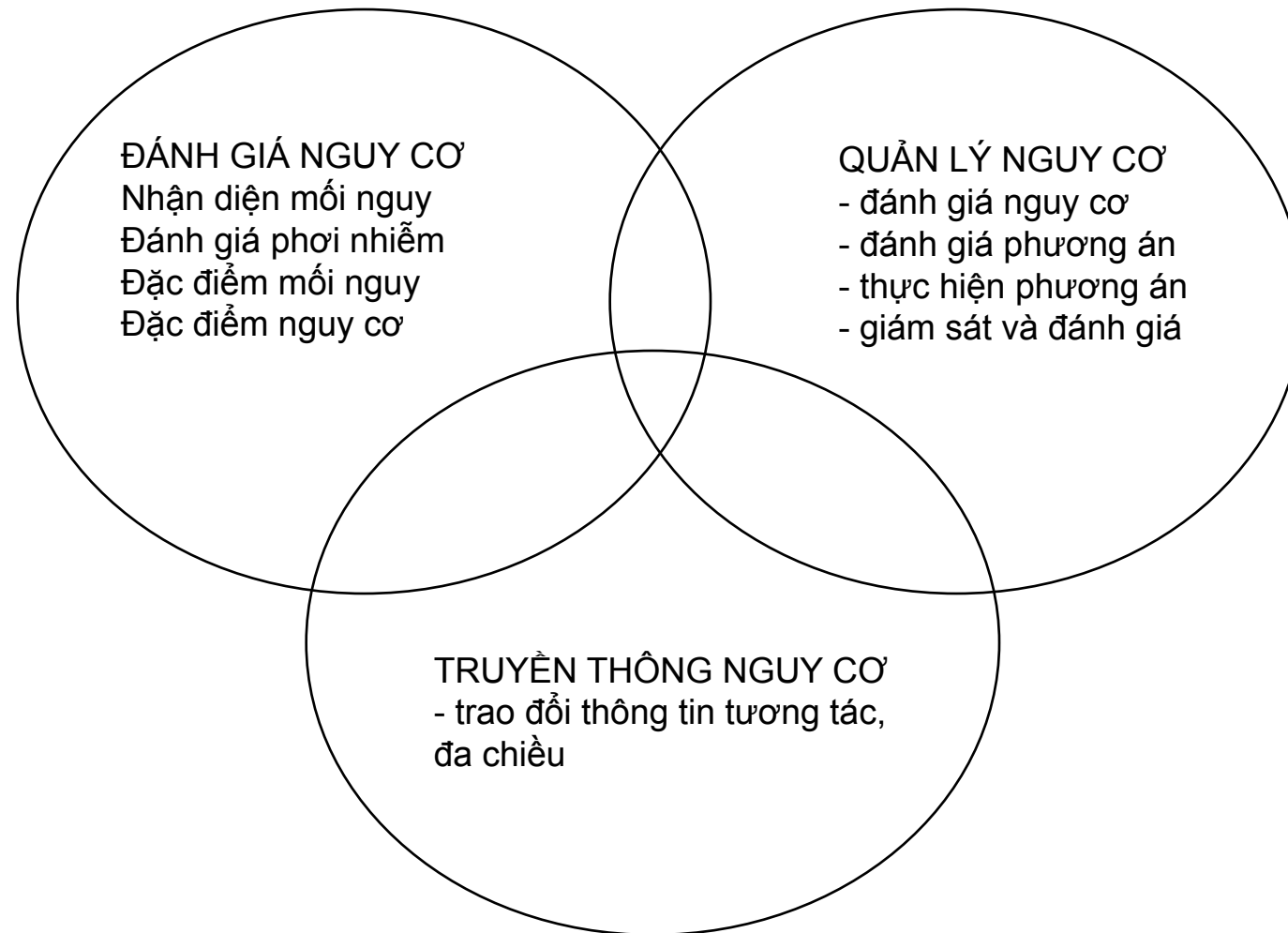
GREG PAOLI

RISK SCIENCES INTERNATIONAL
(VIỆN NGHIÊN CỨU NGUY CƠ QUỐC TẾ)
OTTAWA, CANADA

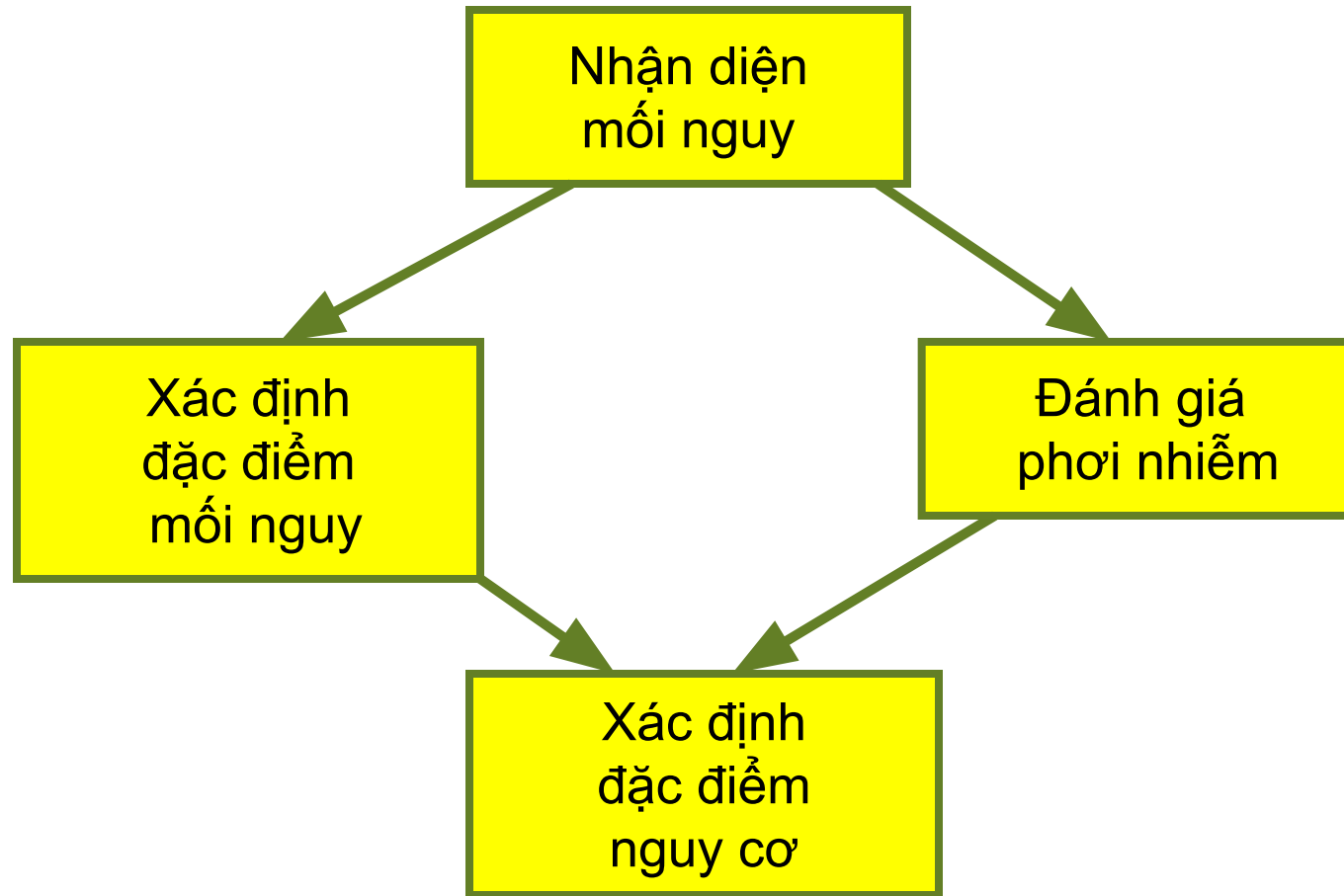
HÀ NỘI, NGÀY 28-31 THÁNG 10 NĂM 2024



Phân tích nguy cơ: Cấu trúc quản lý nguy cơ



Hệ thống ủy ban Codex Alimentarius



Đánh giá nguy cơ

MÔ HÌNH TẮT ĐỊNH

Mô hình tất định là gì?

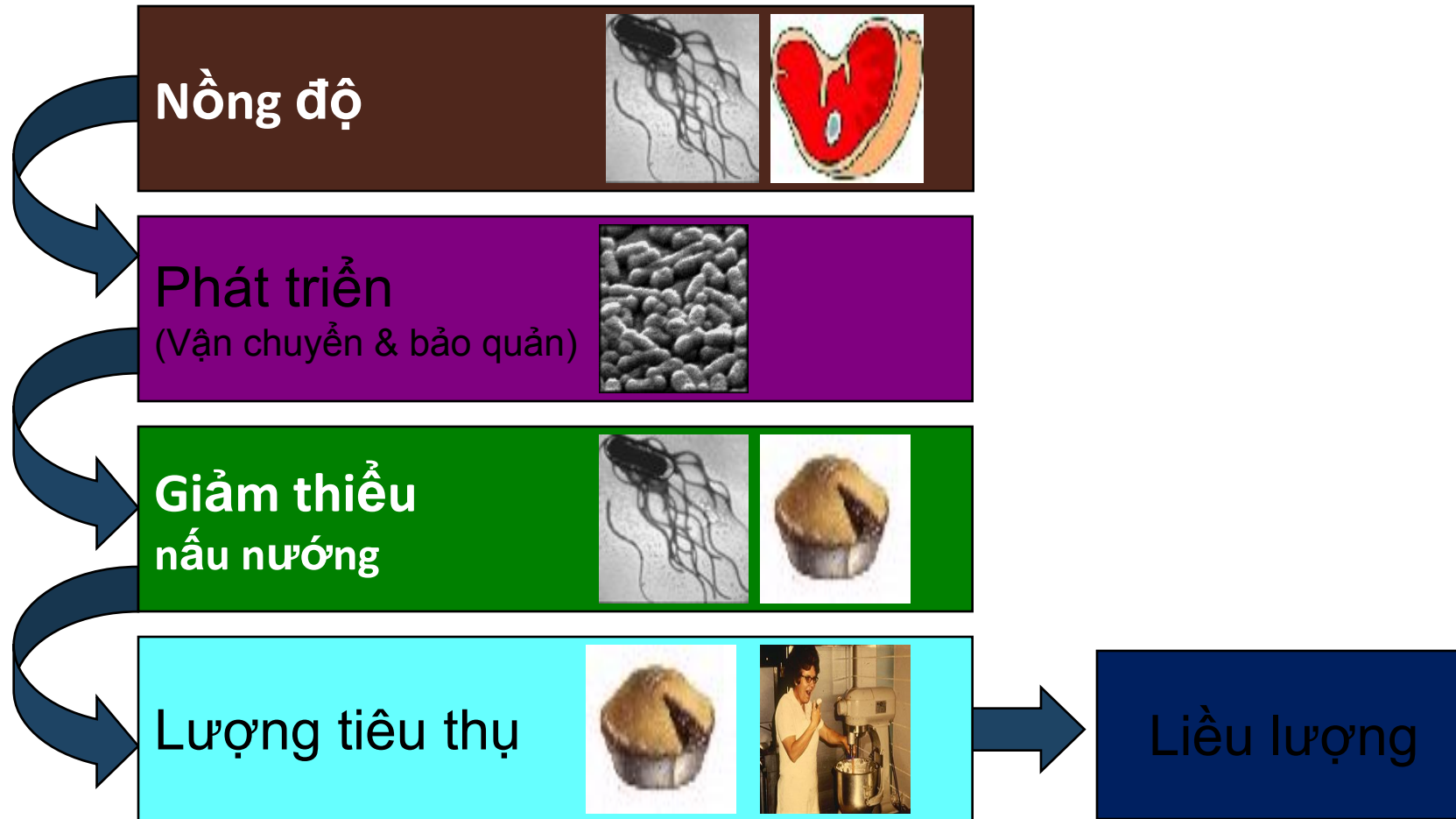
- Trong một mô hình tất định, các kết quả được xác định chính xác thông qua các mối quan hệ đã biết giữa các tham số mô hình
- Một đầu vào nhất định sẽ luôn tạo ra cùng một đầu ra
- Không xem xét bất kỳ biến thể ngẫu nhiên nào trong hệ thống
- Mô hình có thể được xây dựng bằng cách sử dụng các giá trị dự kiến, ước tính trường hợp xấu nhất, v.v...

Kịch bản Đánh giá Nguy cơ Định lượng Vi sinh (QMRA): Xây dựng Mô hình tất định

- Vi sinh vật “X” xuất hiện trong Meat Pies (bánh nhân thịt)
- Vi sinh vật “X” có thể phát triển trong Meat Pies
- Vi sinh vật “X” có thể bị vô hiệu hóa bằng cách nấu

- Sinh viên đại học tiêu thụ Meat Pies
- Sinh viên đại học đôi khi bảo quản Meat Pies không đúng cách
- Sinh viên đại học đôi khi không nấu Meat Pies đủ kỹ.

Kịch bản ví dụ



Kịch bản ví dụ

- Cách tiếp cận thứ nhất
 - Ước tính liều lượng bằng cách sử dụng các giá trị trung bình
- Cách tiếp cận thứ hai
 - Ước tính liều bằng cách sử dụng trường hợp xấu nhất

Kịch bản ví dụ

Giá trị trung bình

- Nồng độ vi sinh vật “X” = 2,0 log CFU/g
- Tăng trưởng vi sinh vật “X” = 1,5 log (nhân tử không đơn vị)
- Bất hoạt vi sinh vật “X” = 3,6 log (nhân tử không đơn vị)
- Kích cỡ khẩu phần = 53,33 g

Kịch bản ví dụ

Trường hợp xấu nhất (giới hạn trên)

- Nồng độ vi sinh vật “X” = 4.0 log CFU/g
- Tăng trưởng vi sinh vật “X” = 1,85 (nhân tử không đơn vị)
- Bất hoạt vi sinh vật “X” = 2,6 log (nhân tử không đơn vị)
- Kích cỡ khẩu phần = 85,00 g

Ước lượng điểm

- Giá trị trung bình

- Liều lượng ước tính được sử dụng
- Khoảng 36 sinh vật

Tính toán

$$(10 [2 + 1,5 - 3,6] \times 53,33)$$

- Giá trị bảo toàn

- Liều lượng ước tính được sử dụng
- Xấp xỉ 152.000 sinh vật

Tính toán

$$(10 [4 + 1,85 - 2,6] \times 85,00)$$

Thảo luận

- Nếu bệnh rất khó xảy ra với liều dưới 1.000 sinh vật, nhưng tăng trên 1.000, bánh nướng nhân thịt có phải là "Thực phẩm an toàn" không?
- Tại sao bạn có thể cho rằng chúng KHÔNG an toàn?
- Tại sao bạn có thể cho rằng chúng an toàn?

Giải thích ước tính điểm

- Nếu ước tính điểm bảo toàn giảm xuống dưới mức nguy cơ tối đa có thể chấp nhận được thì chúng ta biết rằng nguy cơ đó thực sự có thể chấp nhận được
 - ... nhưng mức độ bảo vệ quá mức vẫn chưa được biết
- Nếu ước tính điểm bảo toàn nằm trên mức nguy cơ tối đa có thể chấp nhận được, thì chúng ta không biết liệu nguy cơ có thực sự không thể chấp nhận được hay đó là result of propagated conservatism.
[kết quả của việc ước tính quá mức]

Burmester 1995

Đánh giá nguy cơ

NGHIÊN CỨU ĐIỂN HÌNH:
CRONOBACTER SAKAZAKII
SỮA CÔNG THỨC CHO TRẺ SƠ SINH

Nghiên cứu điển hình: Mô hình tất định phức tạp

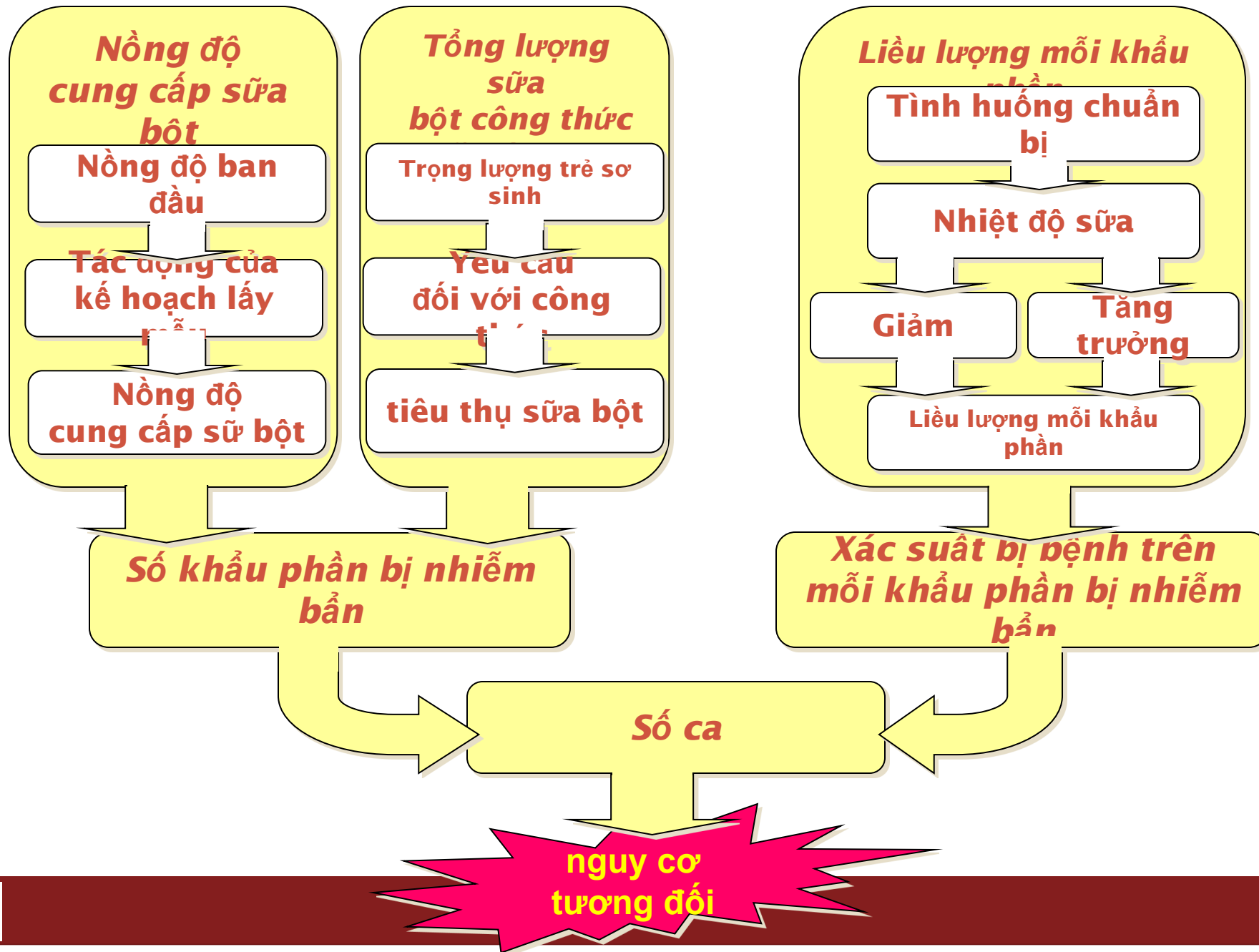
- *Cronobacter sakazakii* trong sữa bột cho trẻ sơ sinh
- Dựa trên một công cụ đánh giá nguy cơ thực tế có sẵn công khai
- Minh họa một ví dụ thực tế về Mô hình tất định được sử dụng trong việc ra quyết định dựa trên nguy cơ
- Kết hợp các tính năng phổ biến để áp dụng đánh giá nguy cơ vi sinh trong nhiều lĩnh vực
 - Ứng phó với vi sinh học dự đoán

C. *Sakazakii* trong sữa công thức dạng bột cho trẻ sơ sinh

- Sữa bột cho trẻ sơ sinh (PIF) đáp ứng các tiêu chuẩn quốc tế/Codex hiện có có liên quan đến các trường hợp mắc bệnh với *C. sakazakii*
- Do đó, Codex bắt đầu quá trình sửa đổi mã
 - Quy tắc thực hành vệ sinh quốc tế về thực phẩm cho trẻ sơ sinh và trẻ em được đề xuất
- Theo yêu cầu của FAO/WHO, công cụ đánh giá nguy cơ này đã được hoàn thành
 - Cung cấp tư vấn khoa học dựa trên nguy cơ cho Codex và các nhà quản lý nguy cơ khác, về vấn đề *C. sakazakii* trong PIF
 - Dành cho các nhà quản lý nguy cơ sử dụng ‘trực tiếp’ khi tham khảo ý kiến của các nhóm làm việc khoa học

Tóm tắt ngắn gọn về Đánh giá nguy cơ

- Mô hình ước tính liều lượng *C. sakazakii* trong PIF đã điều chế khi tiêu thụ, và sau đó là nguy cơ
- Đặc điểm kỹ thuật của các kịch bản làm cơ sở cho dự đoán liều lượng trong công thức điều chế khi tiêu thụ
- Kết quả đầu ra là về sự thay đổi nguy cơ tương đối trong các kịch bản



Mô tả các tình huống chuẩn bị

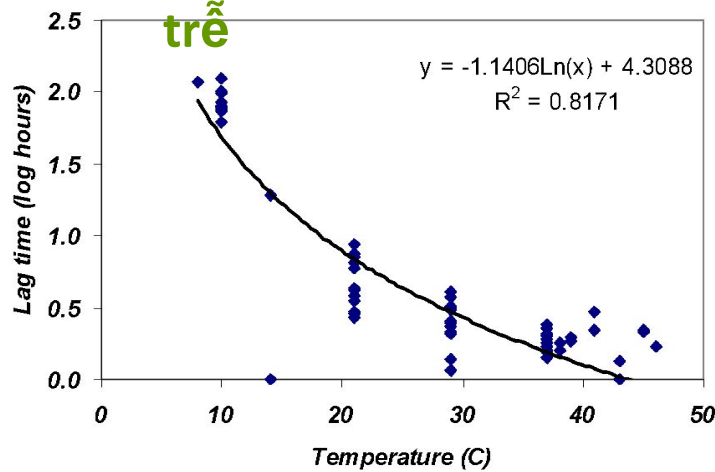
- Các kịch bản được xác định về mặt chuẩn bị, làm mát/bảo quản, hâm nóng và sử dụng
- Các tình huống cần xem xét:
 - Nhiệt độ của chất lỏng tái hydrat hóa
 - Tình huống chuẩn bị (chai đơn, bình chứa 1 lít..)
 - Nhiệt độ để làm mát/bảo quản
 - Nhiệt độ phòng để sử dụng
 - Thời gian của từng giai đoạn chuẩn bị
- Mô hình dự đoán nhiệt độ của sữa công thức trong toàn bộ thời gian từ tái hydrat hóa đến lúc cho ăn

Xem xét kỹ hơn về việc chuẩn bị và xử lý

Tình huống chuẩn bị

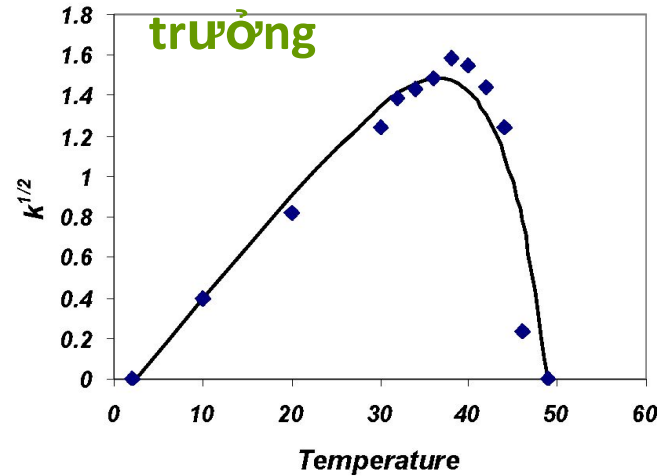
Nhiệt độ của sữa

Thời gian giai đoạn
trễ



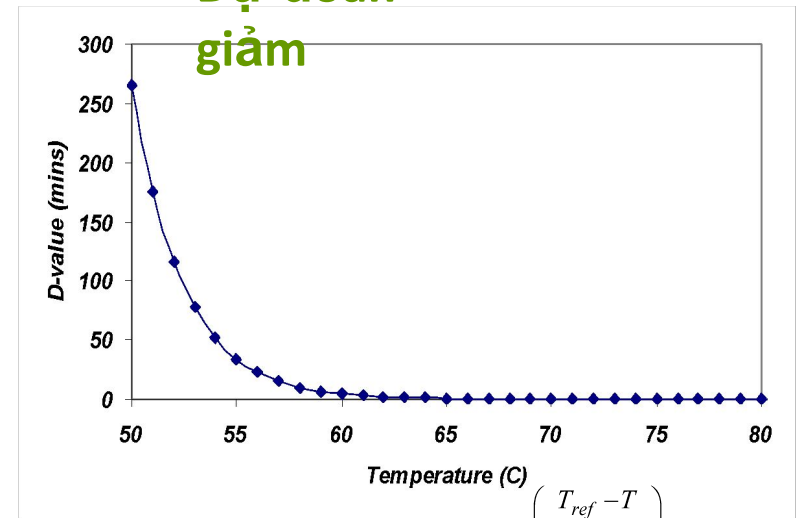
$$\text{Log}_{10}(\lambda) = c_L \ln(T) + b_L$$

Dự đoán tăng
trở



$$\sqrt{k} = b_G (T - T_{\min}) \{1 - \exp(c_G (T - T_{\max}))\}$$

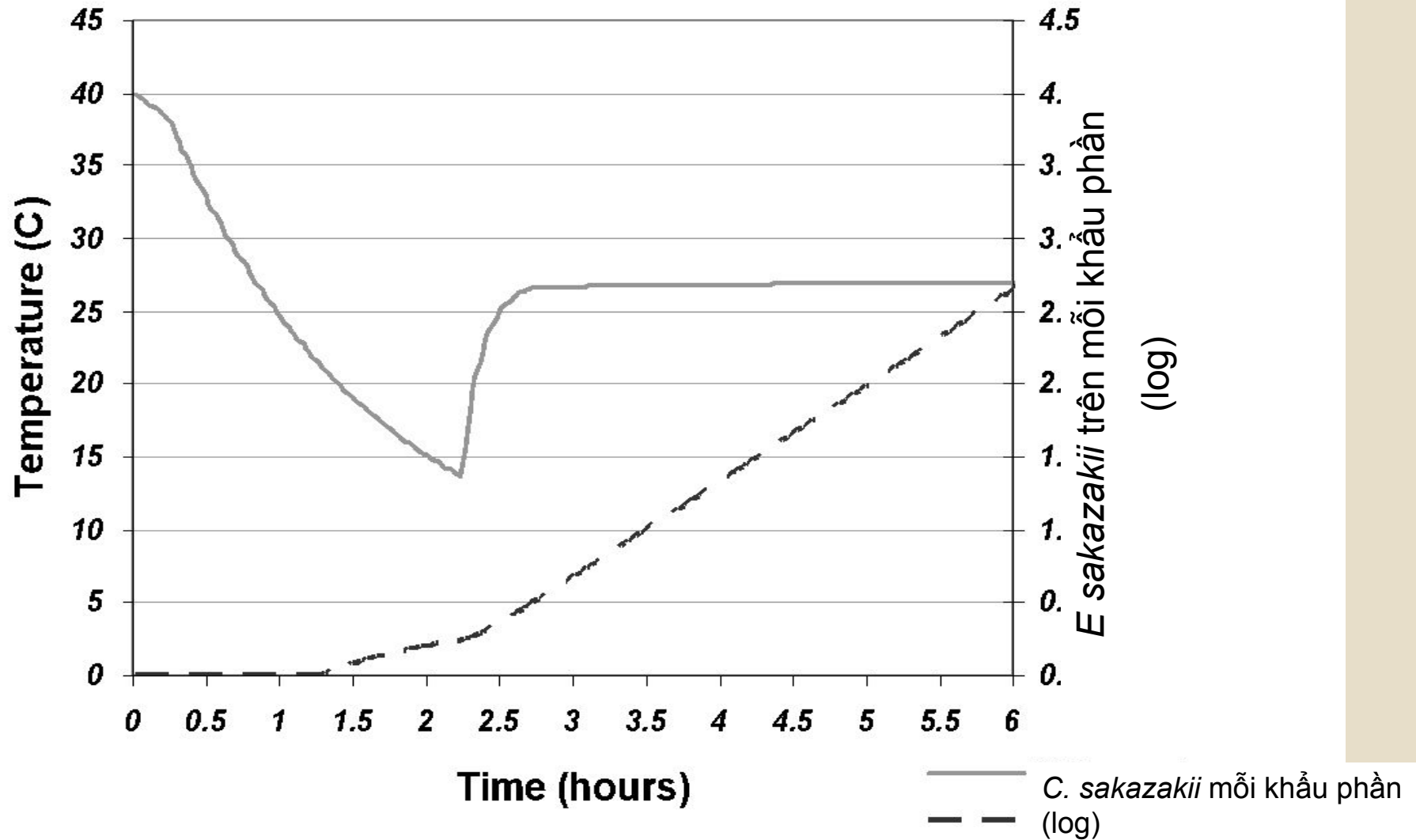
Dự đoán
giảm



$$D_T = D_{ref} - 10^{\left(\frac{T_{ref} - T}{Z\text{-value}}\right)}$$

Liều lượng mỗi khẩu phần

biến động quần thể theo thời gian



So sánh các tình huống chuẩn bị & xử lý

- FAO/WHO triệu tập cuộc họp chuyên gia
 - Tháng 01/2006, Rome
- Các kịch bản được tạo ra bởi một nhóm làm việc tại cuộc họp
- Bảng câu hỏi đã được gửi đến các bệnh viện trên khắp thế giới
- Một danh sách đầy đủ các tình huống đã được tìm hiểu
 - ví dụ: nhiệt độ/thời gian tủ lạnh, nhiệt độ phòng...
- Kết quả được tạo ra tại cuộc họp và được giải thích bởi các nhóm làm việc
 - Báo cáo đầy đủ có sẵn trên trang web của JEMRA
 - <http://www.who.int/foodsafety/publications/micro/mra10.pdf>

Các tình huống cơ bản

- Tám kịch bản cơ bản đã được điều tra
- Các điều kiện được chỉ định cho nhiệt độ phòng mát, ấm và rất ấm
- Các tình huống bao gồm sự kết hợp của:
 - Làm mát bằng cách làm lạnh (4C) hoặc giữ ở nhiệt độ phòng
 - Bao gồm hoặc loại trừ một hành động làm ấm lại rõ ràng
 - Thời gian cho ăn ngắn hoặc dài
- Mỗi kịch bản này được thực hiện ở một loạt các nhiệt độ hoàn nguyên khác nhau
 - 10, 20, 30, 40, 50, 60 và 70C
 - So sánh 168 tình huống chuẩn bị khác nhau

Đầu ra ví dụ: Các tình huống cơ bản

Table 11. Relative risk of different preparation, storage and handling practices for formula prepared and used at a warm ambient room temperature (Room temperature = 30°C) (+ X means an increase in risk of X fold, - X means a decrease in risk of X fold)

Preparation, storage and feeding scenarios	Relative increase or decrease in risk compared to the baseline scenario of 1 at different temperatures of rehydration of PIF						
	10 °C	20 °C	30 °C	40 °C	50 °C	60 °C	70 °C
Refrigeration, re-warming, extended feeding period	+ 2	+ 34	+ 8	+ 27	+ 83	+ 1.8	> - 100,000
Refrigeration, re-warming, short feeding period	1	1	1	1	+ 2.6	- 1.3	> - 100,000
Refrigerated storage, no re-warming, extended feeding period	1	1	1	1	+ 2.7	- 1.3	> - 100,000
Refrigeration, no re-warming, short feeding period	1	1	1	1	1	- 1.3	> - 100,000
No refrigeration, re-warming, extended feeding period	+ 3	+ 6	+ 15	+ 55	+ 161	1	> - 100,000
No refrigeration, re-warming, short feeding period	1	1	1	+ 1.7	+ 5	- 1.3	> - 100,000
No refrigeration, no re-warming, extended feeding period	1	1	+ 2.8	+ 22	+ 97	- 1.3	> - 100,000
No refrigeration, no re-warming, short feeding period	1	1	1 (Base line)	1	+ 3	- 1.3	> - 100,000

Sản phẩm mẫu

Table 18. Comparison of the relative risk related to holding time in /out of refrigeration before extended feeding (2 hour) for scenarios conducted at a warm ambient room temperature. (+ X means an increase in risk of X fold).

	Time between preparation and feeding	Relative increase in risk compared to the baseline scenario of 1 at different temperatures of rehydration of PIF						
		10°C	20°C	30°C	40°C	50°C	60°C	70°C
Refrigeration	2 hour (baseline)	1 (Baseline)	1 (Baseline)	1 (Baseline)	1 (Baseline)	1 (Baseline)	1 (Baseline)	1 (Baseline)
	4	+ 1.06	+ 1.09	+ 1.09	+ 1.12	+ 1.17	+ 1.19	1
	6	+ 1.08	+ 1.11	+ 1.11	+ 1.14	+ 1.19	+ 1.22	1
	8	+ 1.11	+ 1.11	+ 1.13	+ 1.16	+ 1.19	+ 1.24	1
No Refrigeration	2 hour (baseline)	1 (Baseline)	1 (Baseline)	1 (Baseline)	1 (Baseline)	1 (Baseline)	1 (Baseline)	1 (Baseline)
	4	+ 11	+ 16	+ 32	+ 46	+ 53	+ 63	1
	6	+ 354	+ 600	+ 1,189	+ 1,377	+ 680	+ 2,745	1
	8	+ 12,721	+ 18,548	+ 15,268	+ 2,793	+ 705	+ 65,502	

Đưa ra lời khuyên

- Sau khi sử dụng đánh giá nguy cơ, cuộc họp kết luận rằng:
 - Một số hướng dẫn hiện tại trên nhãn sản phẩm PIF và những hướng dẫn được cơ quan y tế khuyến nghị, có thể dẫn đến tăng nguy cơ nhiễm khuẩn *C. sakazakii* và những hướng dẫn này nên được xem xét dựa trên kết quả đánh giá nguy cơ
- Đánh giá đã được FAO/WHO sử dụng để xây dựng hướng dẫn và những hướng dẫn này được công bố công khai
 - Hướng dẫn chuẩn bị, bảo quản và xử lý an toàn sữa bột cho trẻ sơ sinh
 - <http://www.who.int/foodsafety/publications/micro/pif2007/en/>
 - Công cụ có sẵn miễn phí trực tuyến tại www.fstools.org

Đánh giá nguy cơ định lượng vi sinh vật

**XÁC ĐỊNH ĐẶC ĐIỂM MỐI NGUY VÀ NGUY CƠ: MÔ HÌNH ĐÁP ỨNG -
LIỀU LƯỢNG VI SINH VẬT VÀ ƯỚC TÍNH SỐ LƯỢNG BỆNH TẬT**

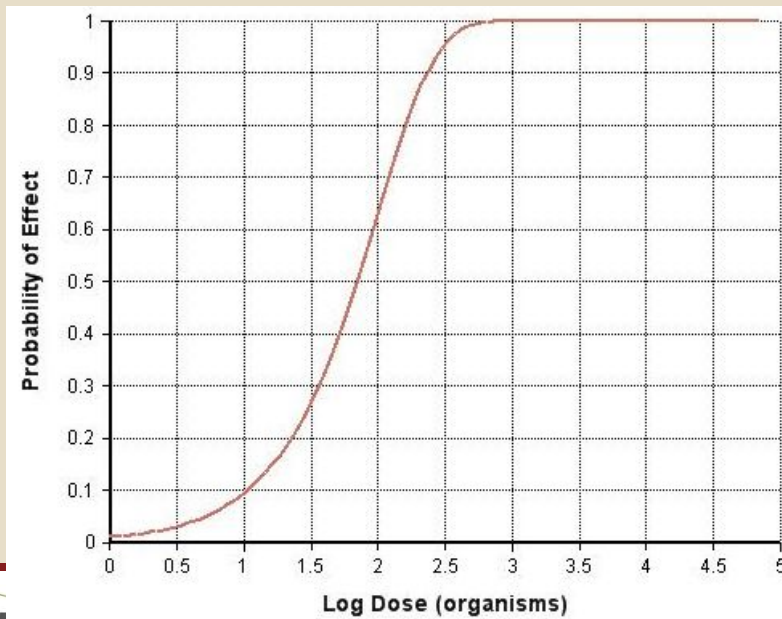
Xác định đặc điểm mỗi nguy trong đánh giá nguy cơ vi sinh

- Đánh giá định tính(?) và/hoặc định lượng về bản chất của các tác động bất lợi đến sức khỏe liên quan đến tác nhân sinh học
- Nên xem xét rõ ràng sự phức tạp của tương tác (bao gồm cả di chứng, di tật) giữa con người và tác nhân sau khi tiếp xúc cũng như khả năng lây lan thêm
 - Nên thực hiện đánh giá đáp ứng - liều lượng

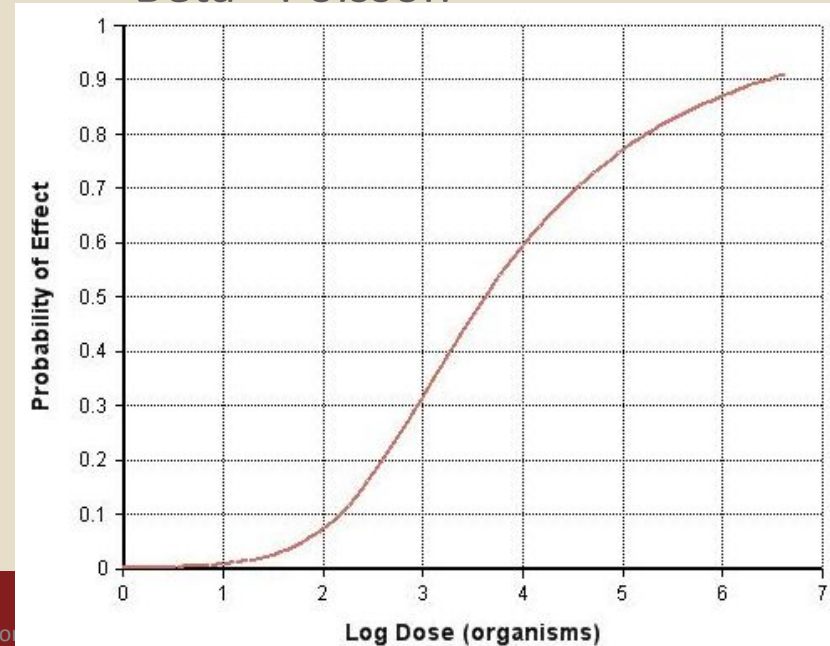
Đánh giá liều lượng - đáp ứng

- Các mô hình đáp ứng - liều lượng là các hàm toán học mô tả mối quan hệ đáp ứng liều - lượng đối với các mầm bệnh, đường lây truyền và vật chủ cụ thể
- Ước tính nguy cơ đáp ứng (ví dụ: nhiễm trùng, bệnh tật hoặc tử vong) với liều lượng mầm bệnh đã biết

• Số mũ



Beta - Poisson



Mô hình đáp ứng liều lượng vi sinh vật

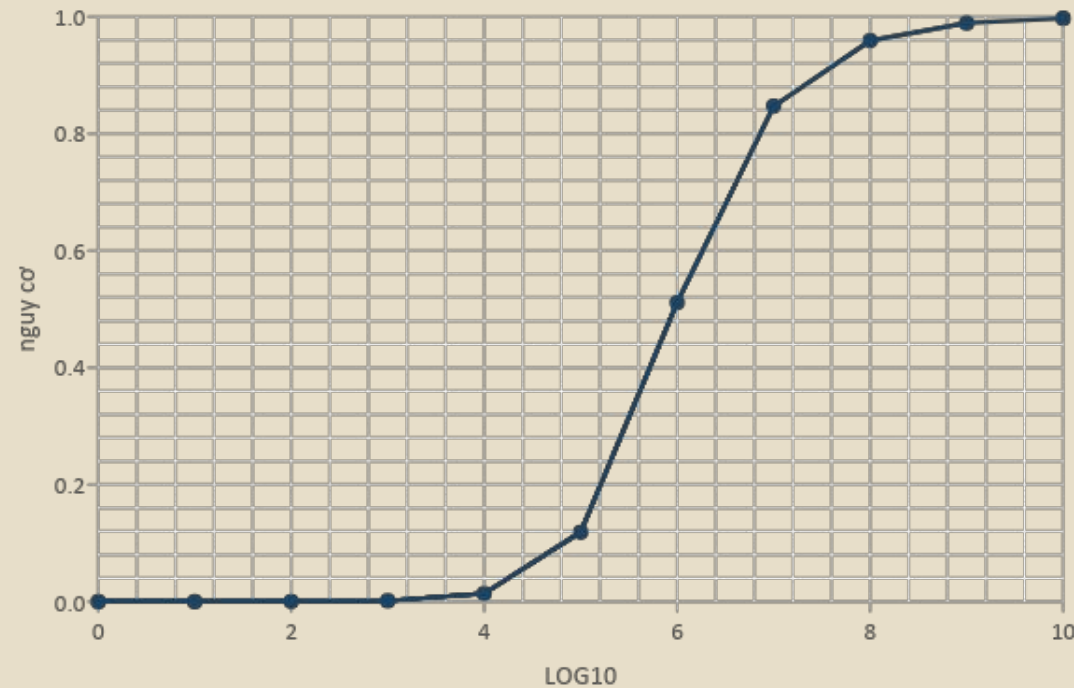
- Luôn được coi là phơi nhiễm cấp tính
- 1 CFU có khả năng gây nhiễm trùng
 - Lý thuyết về liều lây nhiễm tối thiểu (of minimum infectious dose - MID) không còn được chấp nhận
- Có thể dựa trên các nghiên cứu về thức ăn hoặc dữ liệu về ổ dịch

Hoàn thiện mô hình bánh nhân thịt

- Giả sử Bug X tuân theo mối quan hệ liều Beta-Poisson và thêm nó vào Mô hình tất định

- $P_{ill} = 1 - \left(\frac{1+d}{\beta}\right)^{-\alpha}$

- $\alpha = 0,581, \beta = 4,11 \times 10^5$



Tài liệu chính cho các mô hình đáp ứng liều lượng vi sinh vật

- Bản tóm tắt các thí nghiệm và mô hình được biên soạn:

- <https://qmrawiki.org/framework/dose-response/experiments>

- Xem thêm, <https://www.who.int/publications/i/item/9789240024892>

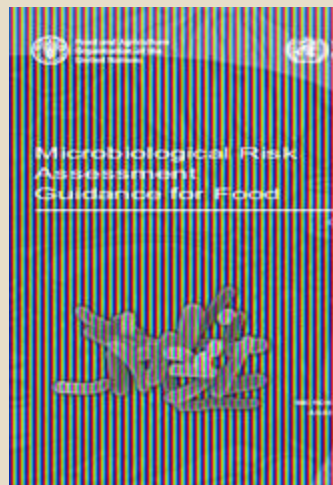


TABLE 7. Dose-response models and parameter estimates commonly used in QMRA

Organism	Reference	Model	Parameters	Lower bound (Percentile)	Upper bound (Percentile)
<i>Salmonella</i> spp.	FAO/WHO (2002a)	Beta-Poisson	$\alpha=0.1324$ $\beta=51.43$	0.0940 (2.5th) 43.75 (2.5th)	0.1817 (97.5th) 56.39 (97.5th)
<i>Listeria monocytogenes</i> ^a	FAO/WHO (2004)	Exponential (susceptible) Exponential (healthy)	$r=1.06 \times 10^{-12}$ $r=2.37 \times 10^{-14}$	2.47×10^{-13} (5th) 3.55×10^{-15} (5th)	9.32×10^{-12} (95th) 2.70×10^{-13} (95th)
<i>Campylobacter</i> spp. ^b	FAO/WHO (2009d)	Beta-Poisson	$\alpha=0.21$ $\beta=59.95$		
<i>Shigella dysenteriae</i> / <i>E. coli</i> O157	Cassin et al. (1998)	Beta-binomial	$\alpha=0.267$ $\beta=\text{Lognormal}$ (5.435, 2.47 ²)		
<i>Vibrio vulnificus</i>	FAO/WHO (2005)		$\alpha=9.3 \times 10^{-6}$ $\beta=110\ 000$		

^aFor *L. monocytogenes*, newer animal model data (Roulo et al., 2014; Smith et al., 2003, 2008; Williams et al., 2007, 2009) and outbreak data (Poilliot et al., 2016) suggest much higher *r*-values and hence lower ID_{50} values than predicted by this model which was based on the method of Buchanan et al. (1997) of matching expected loads of *L. monocytogenes* across the food supply to the total annual cases in a community, and which relies on many untested assumptions.

^bThe dose-response relation is for infection. The conditional probability of disease following infection was 33 percent (29/89) and can be described by a beta(30,61) distribution.

Xác định đặc điểm nguy cơ: Tìm hiểu về số lượng bệnh

- Phương trình cơ bản

- Xác suất mắc bệnh: $P_{ill} = G(d)$
- $G(d)$ là một hàm chuyển đổi liều lượng thành xác suất mắc bệnh cho một khẩu phần ăn duy nhất chứa liều lượng, d .

- Khi liều lượng thay đổi:

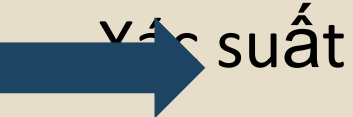
- Xác suất mắc bệnh: $MeanP_{ill} = \int_{d_{min}}^{d_{max}} G(d)f(d)dd$
- Số bệnh: $N_{ill} = N_{servings} \times MeanP_{ill}$
- $G(d)$ là một hàm chuyển đổi liều lượng thành xác suất mắc bệnh với liều lượng đó.
- $f(d)$ là sự phân bố xác suất của sự thay đổi liều lượng trong khẩu phần ăn
- phân bố $f(d)$ và tích phân thường được tạo ra bằng cách sử dụng Mô phỏng Monte Carlo

Đánh giá nguy cơ

RÀ SOÁT XÁC SUẤT VÀ GIỚI THIỆU VỀ MÔ HÌNH HÓA XÁC SUẤT

Định nghĩa xác suất

- Thử nghiệm ngẫu nhiên có thể tiết lộ một mô hình
 - Mô hình mặt sấp và mặt ngửa của một đồng xu
 - Mô hình chiều cao khác nhau của con người
- Lặp lại thử nghiệm nhiều lần để tìm hiểu thêm về mô hình
- Tần suất tương đối của kết quả cụ thể so với các kết quả khác có xu hướng hướng tới giá trị không đổi

Xác suất

Ví dụ

- Xác suất là một thước đo giữa 0 và 1
- Một quả bóng được chọn ngẫu nhiên từ một túi chứa 3 quả bóng đỏ và 7 quả bóng trắng
- Xác suất chọn một quả bóng màu đỏ sẽ là $3/10$

- Ngẫu nhiên – có nghĩa là mỗi quả bóng trong số 10 quả bóng có cùng xác suất (cơ hội) được chọn
 - Tất cả 10 kết quả đều có khả năng như nhau
- Xác suất mà một quả bóng trắng được chọn là $7/10$
 - Tổng hai xác suất là 1 – không có kết quả nào khác có thể xảy ra – bóng có màu trắng hoặc đỏ

Không gian mẫu

- Các giá trị mà kết quả của các thí nghiệm ngẫu nhiên có thể lấy
 - Mặt sấp hoặc mặt ngửa khi tung đồng xu
 - Tất cả chiều cao có thể của mọi người trong một căn phòng
- Tập hợp con các giá trị
 - Chiều cao từ 150 cm đến 155 cm



Nội dung thi đấu

Thử nghiệm ngẫu nhiên

- Quy trình đạt được thông tin
 - kết quả của việc tung đồng xu
 - chiều cao của người tiếp theo vào phòng
 - thời gian cần có để chiếu mỗi slide
- Nếu ngẫu nhiên thì không chắc chắn về kết quả
 - Bạn sẽ có tung được mặt sấp hay mặt ngửa?
 - Không chắc chắn về chiều cao trước khi nhìn thấy (và đo lường) người
 - Một số slide sẽ nhanh hơn những slide khác

Xác suất có điều kiện

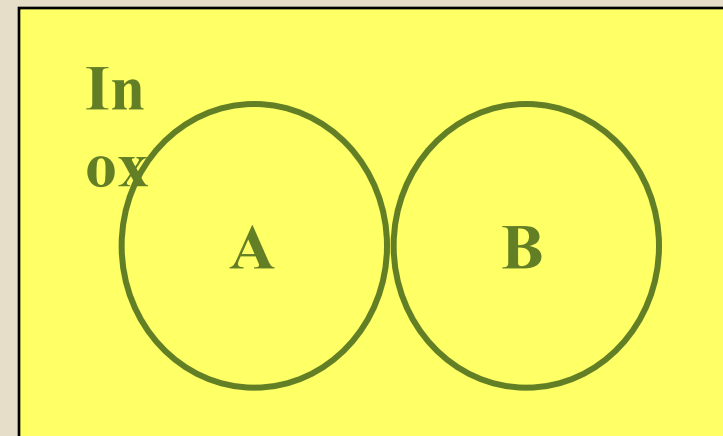
- A và B là kết quả của một thí nghiệm ngẫu nhiên
 - $P(A)$ - Xác suất A xảy ra
 - $P(\text{không phải } A)$ - Xác suất A không xảy ra
 - $P(A \cap B)$ - Xác suất A và B xảy ra
 - $P(A \cup B)$ - Xác suất xảy ra A hoặc B hoặc xảy ra cả hai
 - $P(A | B)$ - Xác suất A xảy ra khi B đã xảy ra

Đại diện cho xác suất

- Giúp xem xét vấn đề một cách trực quan
- Thường tránh những sai lầm đơn giản
 - Sơ đồ Venn
 - Cây sự kiện

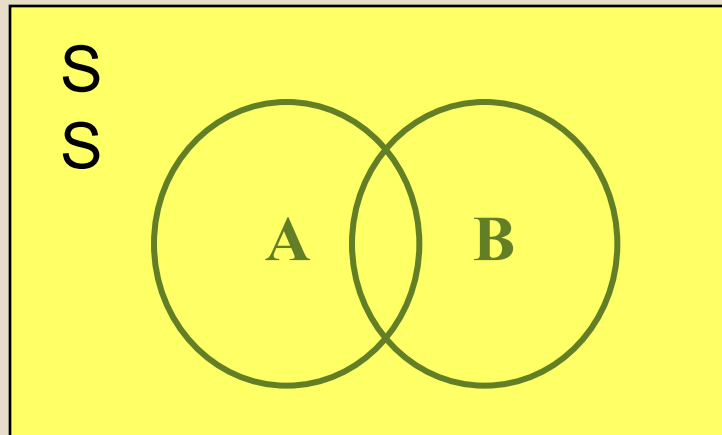
Sơ đồ Venn

- Sơ đồ Venn cho thấy không gian mẫu được chia thành các sự kiện như thế nào
- Bình phương là tổng không gian mẫu = 1
 - $P(\text{không phải } A) = 1 - P(A)$
 - $P(\text{không phải } B) = 1 - P(B)$

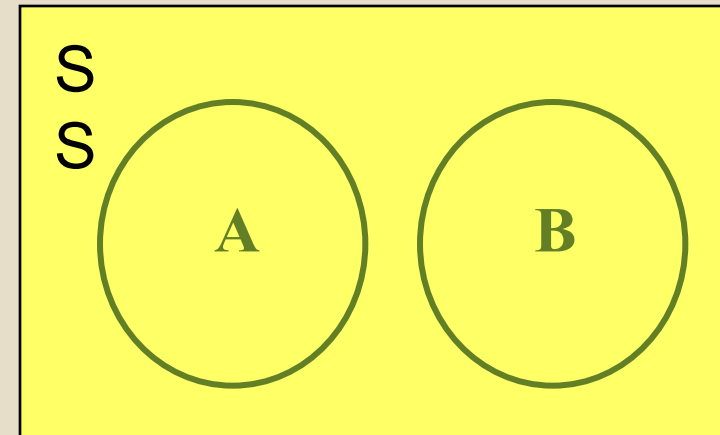


Sơ đồ Venn

- các sự kiện loại trừ lẫn nhau không thể xảy ra cùng một lúc

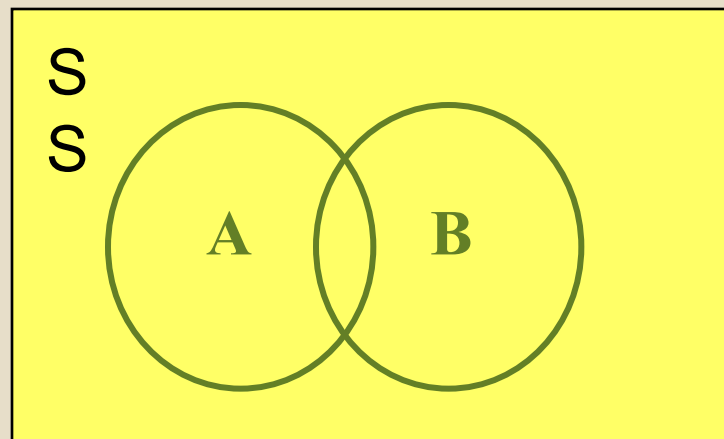


- A và B không loại trừ lẫn nhau



- A và B loại trừ lẫn nhau

Tính toán

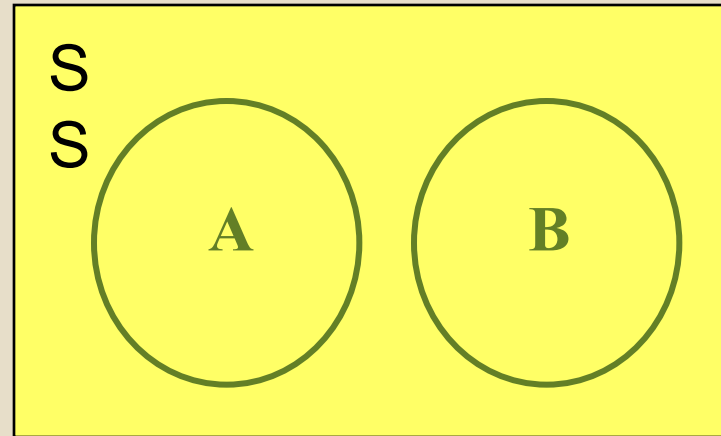


Nếu A và B **không** loại trừ lẫn nhau

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

$$\square P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Tính toán



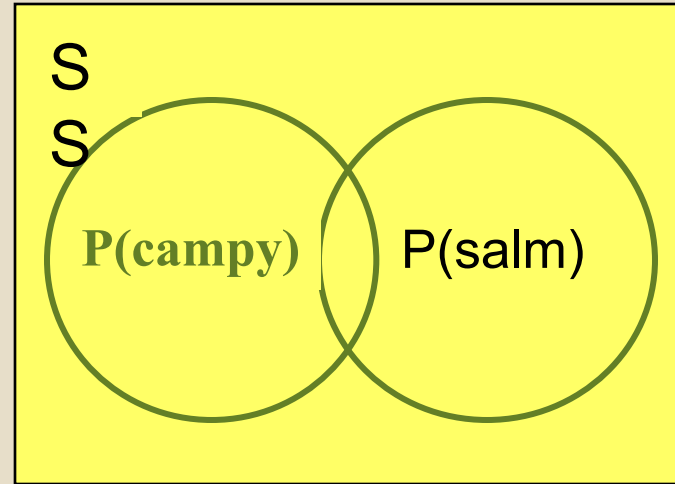
Nếu A và B loại trừ lẫn nhau

- $P(A \cap B) = P(B|A) \times P(A) = 0$ (tại sao không?)
- $P(A \cup B) = P(A) + P(B)$

Biểu đồ ven

- Một chiếc bánh gà có 20% cơ hội nhiễm khuẩn campylobacter và 10% cơ hội nhiễm salmonella. Xác suất mà một chiếc bánh gà có campylobacter hoặc salmonella, hoặc cả hai là bao nhiêu?
- $P(\text{campy}) = 0,2$
- $P(\text{salm}) = 0,1$
- $P(\text{campy} \ \& \ \text{salm}) = 0,2 + 0,1 - (0,2 * 0,1) = 0,28$

Ví dụ

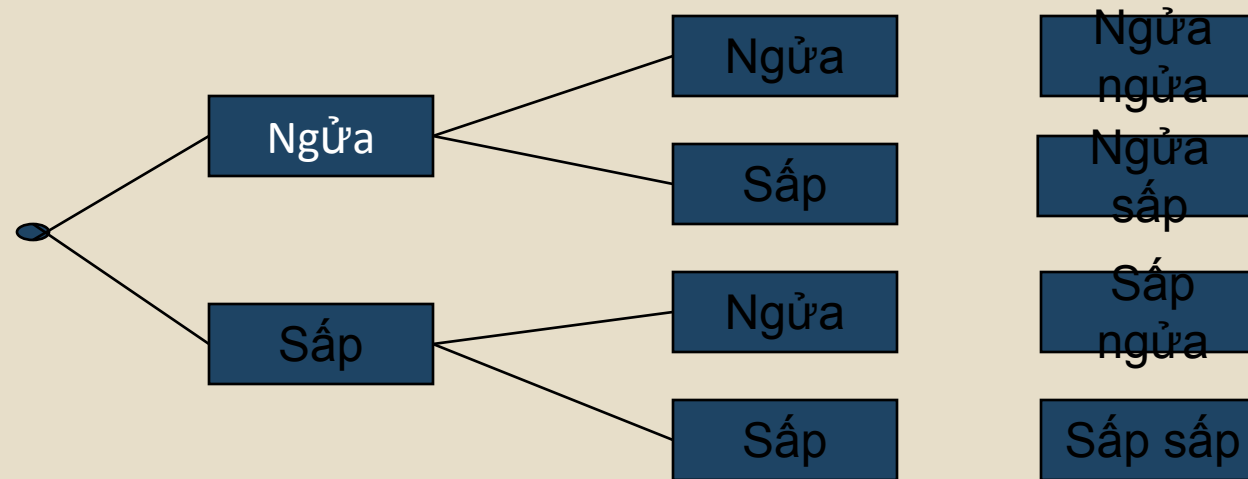


$$P(\text{campy}) = 0,2$$

$$P(\text{salm}) = 0,1$$

$$P(\text{campy} \cup \text{salm}) = 0,2 + 0,1 - (0,2 * 0,1) = 0,28$$

Cây sự kiện (hai lần tung đồng xu)



- ❑ $P(0 \text{ Ngửa}) = (0,5 \cdot 0,5) = 0,25$
 - ❑ $P(1 \text{ Ngửa}) = (0,5 \cdot 0,5) + (0,5 \cdot 0,5) = 0,5$
 - ❑ $P(2 \text{ Ngửa}) = (0,5 \cdot 0,5) = 0,25$
- ▣ $p(x=0,1,2) = \{0.25, 0.5, 0.25\}$

7 PHÂN BỐ XÁC SUẤT

phân bố xác suất là gì?

- Một hàm mô tả tất cả các giá trị mà một biến ngẫu nhiên có thể nhận được và xác suất liên quan đến mỗi biến
- Biến ngẫu nhiên – phải lấy một và chỉ một giá trị từ không gian mẫu bất cứ lúc nào
- Các giá trị trong không gian mẫu loại trừ lẫn nhau
- Xác suất trong tổng phân bố đến 1

phân bố xác suất có thể là ...

- Rời rạc
- Liên tục
- Tham số
- Phi tham số

phân bố rời rạc

- Một hàm có thể lấy một số giá trị riêng biệt (không nhất thiết là hữu hạn).
- Mỗi giá trị (x) có xác suất xuất hiện chính xác
- Tổng xác suất bằng 1

$$\sum_j P_j = 1$$

- Đây thường là các số nguyên không âm
- Thường được gọi là hàm khối lượng xác suất

Ví dụ về các biến rời rạc

- Ví dụ về các biến rời rạc là
 - Kết quả của việc tung đồng xu (Ngửa hoặc Sấp)
 - Giới tính (Nam hoặc Nữ)
 - Số lượng ô tô trong bãi đỗ xe (số nguyên)
 - Kết quả mẫu dương tính hoặc âm tính
 - Số lượng sinh vật (0,1, 2, ...)
 - Tình trạng sức khỏe (suy giảm miễn dịch, chuẩn)
 - Ngày trong tuần mà một sự kiện xảy ra

phân bố liên tục

- Được xác định cho một số điểm vô hạn trong một khoảng thời gian liên tục
- Diện tích dưới đường cong bằng sự thống nhất
- Xác suất cho bất kỳ giá trị cụ thể nào là 0
- Xác suất để x nằm giữa hai điểm a và b là
$$P[M \leq x \leq b] = \int_a^b f(x) dx$$
- Thường được gọi là hàm mật độ xác suất

Ví dụ về biến liên tục

- Ví dụ về các biến liên tục là
 - Thời gian, Thời lượng và Cường độ mưa
 - Chiều dài, Trọng lượng, Chiều cao
 - Vị trí xảy ra sự cố (vỡ đường ống)
 - Nhiệt độ
 - Nồng độ, Thể tích, Tốc độ
 - Quãng đường xe cứu thương phải đi để cứu hộ

TRƯỜNG hợp biên

- Một số đại lượng vốn đã *rời rạc*, nhưng được đặc trưng là *liên tục* để thuận tiện cho tính toán:
 - Số lượng lớn mầm bệnh
 - Quy mô dân số phơi nhiễm
 - Số lượng cuộc gọi điện thoại được xử lý bởi nhân viên điều phối khẩn cấp trong một tuần
- Tác động của lựa chọn này có thể từ thường xuyên đến nghiêm trọng (hiếm gặp, nhưng quan trọng).
Tìm hiểu ngay!

Trường hợp biên

- Một số đại lượng là liên tục, nhưng có thể được đặc trưng là rời rạc:
 - Số đo được báo cáo (làm tròn)
 - Nhiệt độ
 - Chiều cao? Độ tuổi?
 - Chiều cao tòa nhà (số tầng)
- Nếu cần, chúng ta có thể sử dụng nhiều phương pháp khác nhau để tái tạo phân bố liên tục từ dữ liệu rời rạc
 - Có tính đến bản chất dự kiến của hiện tượng liên tục, hình thức phân biệt (ví dụ: làm tròn), thiên vị dự kiến

Họ cũng có thể là.....

- Giới hạn
 - bị giới hạn bởi 2 giới hạn
- Không giới hạn"
 - kéo dài từ \pm vô cực
- Giới hạn một phần
 - hạn chế ở một cực

BIỂU DIỄN BẰNG ĐỒ THỊ VỀ XÁC SUẤT

Biến ngẫu nhiên rời rạc

Ví dụ: biến ngẫu nhiên liên tục

phân bố tích lũy (cdf)

- Cdf là xác suất biến nhận giá trị nhỏ hơn hoặc bằng x :

phân bố liên tục

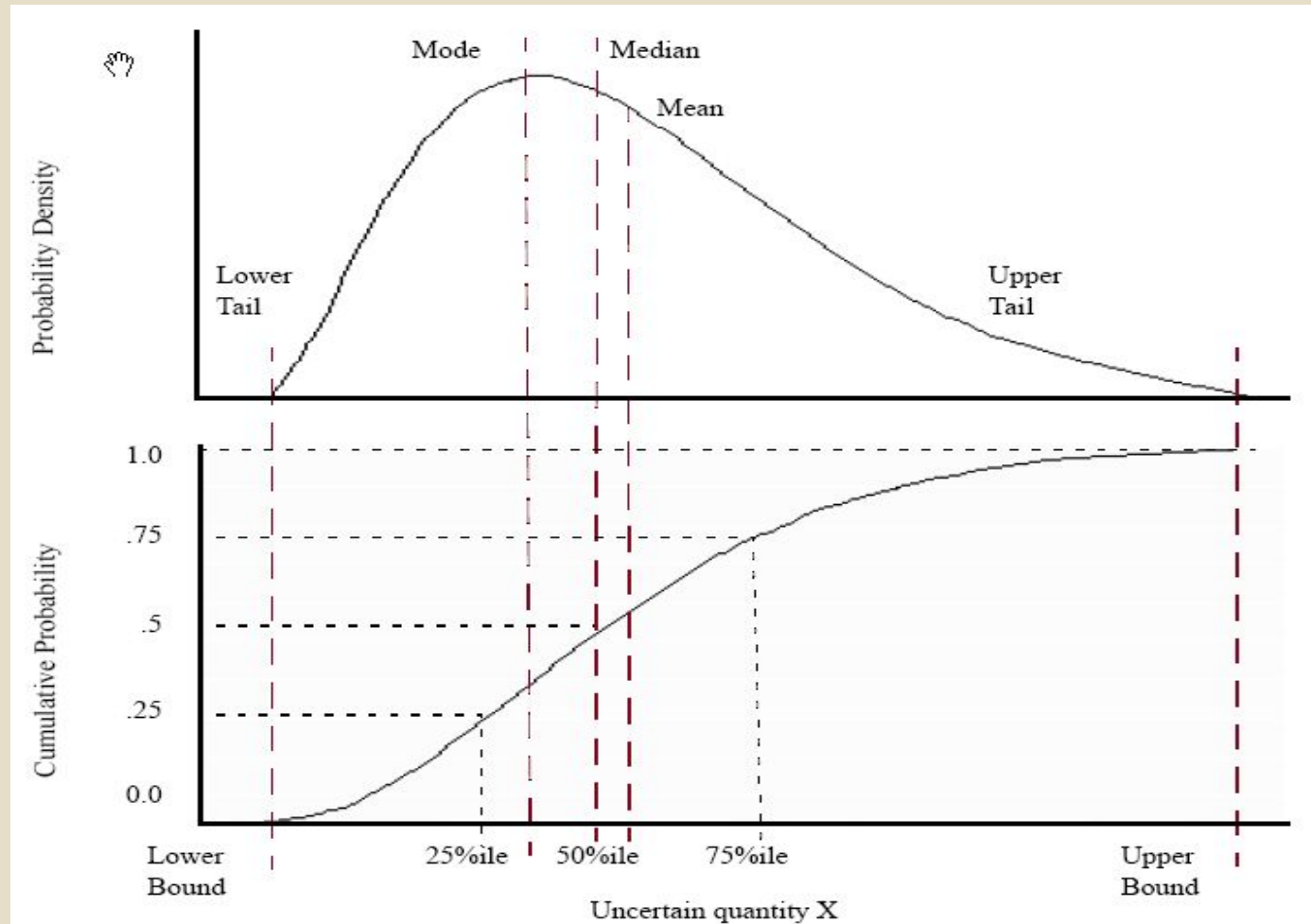
$$F(x) = \Pr[X \leq x]$$

$$F(x) = \int_{-\infty}^x f(\mu) d\mu$$

- phân bố rời rạc

$$F(x) = \sum_{i=-\infty}^x f(i)$$

PDFs và CDFs



Rà soát các biện pháp thống kê

- Các biện pháp đo lường xu hướng trung tâm
 - Trung bình
 - Trung vị
 - Mốt
- Các biện pháp phân tán
 - Phạm vi
 - Phương sai, độ lệch chuẩn
- Bách phân vị

Thước đo xu hướng trung tâm: Mean hoặc average (giá trị trung bình = m)

Rời rạc

$$m = \sum_{i=1}^n x_i p(x_i)$$

- n = số kết quả có thể xảy ra trong SS
- x_i = giá trị của kết quả i
- $p(x_i)$ = xác suất kết quả i xảy ra

Liên tục

$$m = \int_{-\infty}^{\infty} x \cdot f(x) dx$$

- x = giá trị kết quả
- $f(x)$ = hàm mật độ xác suất

Các biện pháp đo lường xu hướng trung tâm

- Giá trị trung bình
- Tung một đồng xu hai lần, có bao nhiêu mặt ngửa?
- phân bố rời rạc
- không gian mẫu $SS=\{0,1,2\}$
- phân bố xác suất $p(x)=\{0.25,0.5,0.25\}$
- Trung bình = $0 \cdot 0,25 + 1 \cdot 0,5 + 2 \cdot 0,25 = 1$

Các biện pháp đo lường xu hướng trung tâm

- Trung vị

- giá trị mà 50% phân bố ở trên và 50% phân bố ở dưới

- Mốt

- quan sát hoặc giá trị thường xuyên nhất với xác suất xảy ra cao nhất (giá trị có khả năng xảy ra cao nhất)

Đo lường độ phân tán

- Phạm vi

- chênh lệch giữa giá trị tối thiểu và tối đa
- ví dụ: phạm vi trọng lượng nai con là $69,3 - 25,8 = 43,5\text{kg}$

- Phương sai

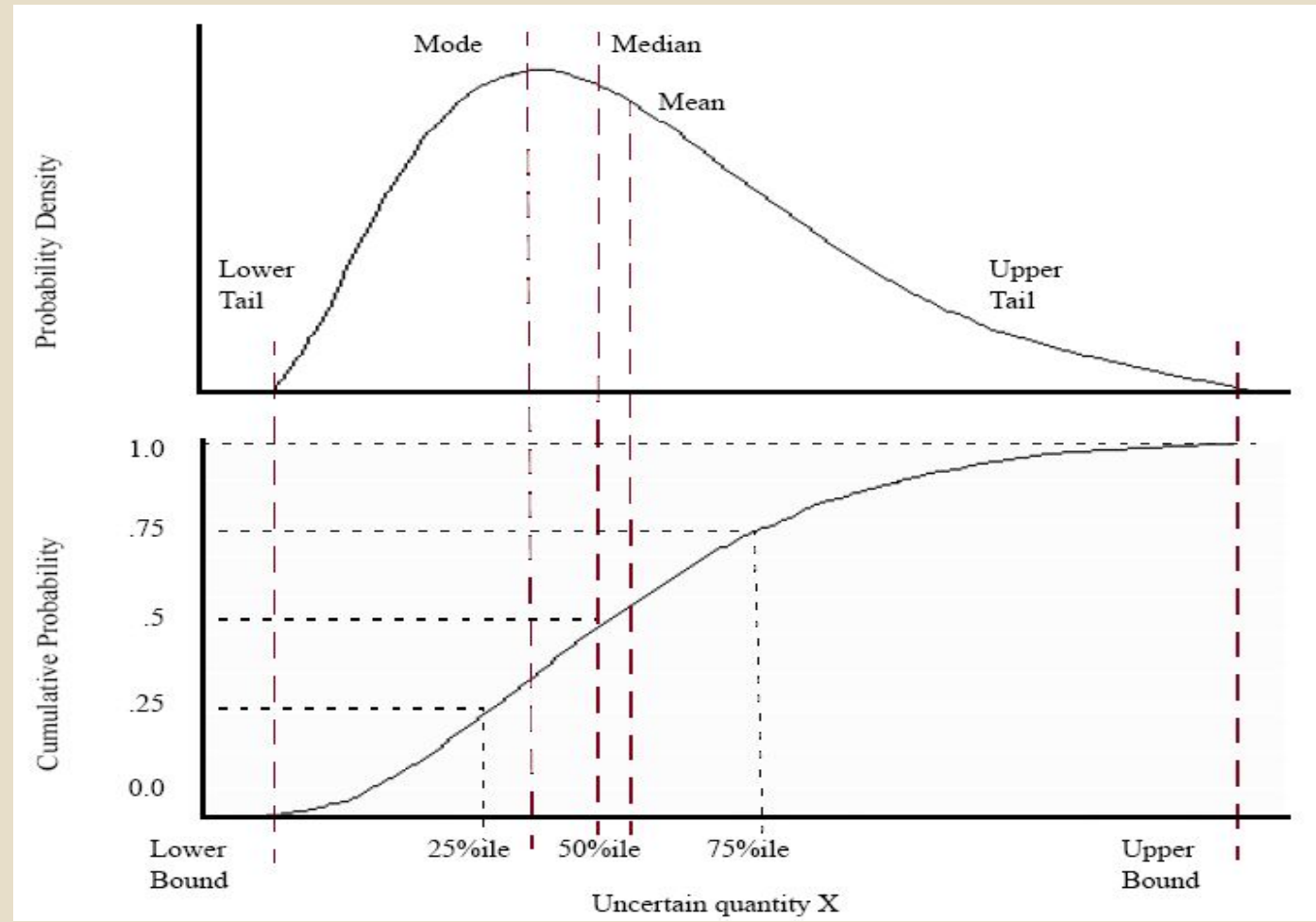
- Trung bình của độ lệch bình phương so với giá trị trung bình

Đo lường độ phân tán

● Bách phân vị

- Bách phân vị thứ x là giá trị mà $x\%$ dữ liệu có giá trị thấp hơn
- cũng được nghĩ đến về "độ chắc chắn"
 - 95% chắc chắn rằng số lượng cừu trong một đàn ít hơn 300
 - khoảng tin cậy (khoảng ước tính)
- Bách phân vị thứ 50 còn được gọi là "trung vị"

Đo lường độ phân tán



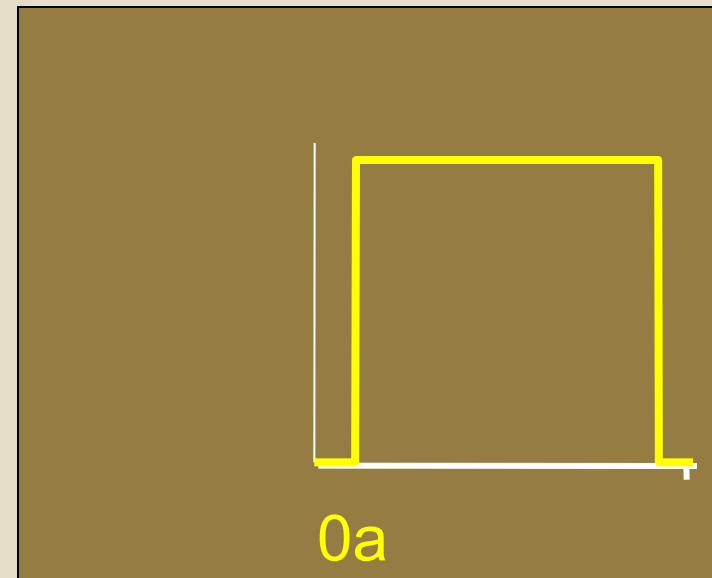
Đánh giá nguy cơ

MỘT PHÂN BỐ XÁC SUẤT ĐƠN GIẢN:

ĐỒNG NHẤT

Đồng nhất (a,c)

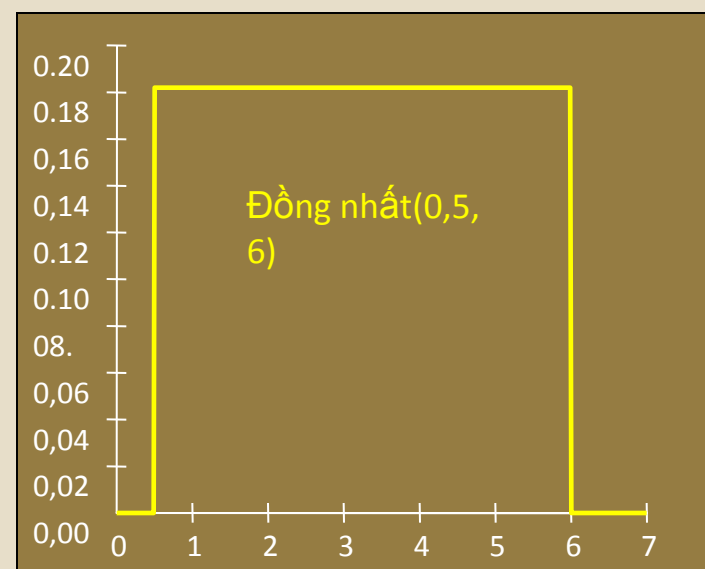
- Thông số: tối thiểu (**a**); tối đa (**c**)
- Giả sử *tất cả* các giá trị từ **a** đến **c** đều có khả năng xảy ra như nhau
- Thường được sử dụng để đại diện cho sự không chắc chắn (sự thiếu thông tin)
- liên tục
- Giới hạn
 - Không gian mẫu: $(a \leq x \leq c)$
- Trung bình = $(a+c)/2$



Ví dụ về phân bố đồng nhất

- Thời gian chờ điều trị trong phòng cấp cứu (Emergency Room = ER) không được biết, nhưng có thể từ tối thiểu 0,5 đến tối đa 6 giờ
- phân bố = Đồng nhất(0,5,6)
- Trung bình = 3,25

=RiskUniform(a,c)



Đánh giá nguy cơ

GIỚI THIỆU VỀ MÔ PHỎNG MONTE CARLO

- Phương pháp mô phỏng Monte Carlo

- **Mô phỏng** - bất kỳ phương pháp phân tích nào nhằm tạo ra một phiên bản đơn giản hóa của một hệ thống, quá trình hoặc hiện tượng phức tạp để nghiên cứu, phân tích và dự đoán hành vi của hệ thống đó trong các điều kiện khác nhau, đặc biệt là khi các phân tích khác quá phức tạp về mặt toán học hoặc quá khó hoặc quá tốn kém để làm.
- Đối với mỗi biến xác suất, xác định các giá trị có thể có với một phân bố.
- Phân tích Monte Carlo
 - xây dựng một cách có hệ thống phân bố xác suất của các biến đầu ra, bằng cách chọn ngẫu nhiên các giá trị cho các biến đầu vào *theo phân bố xác suất của chúng*.

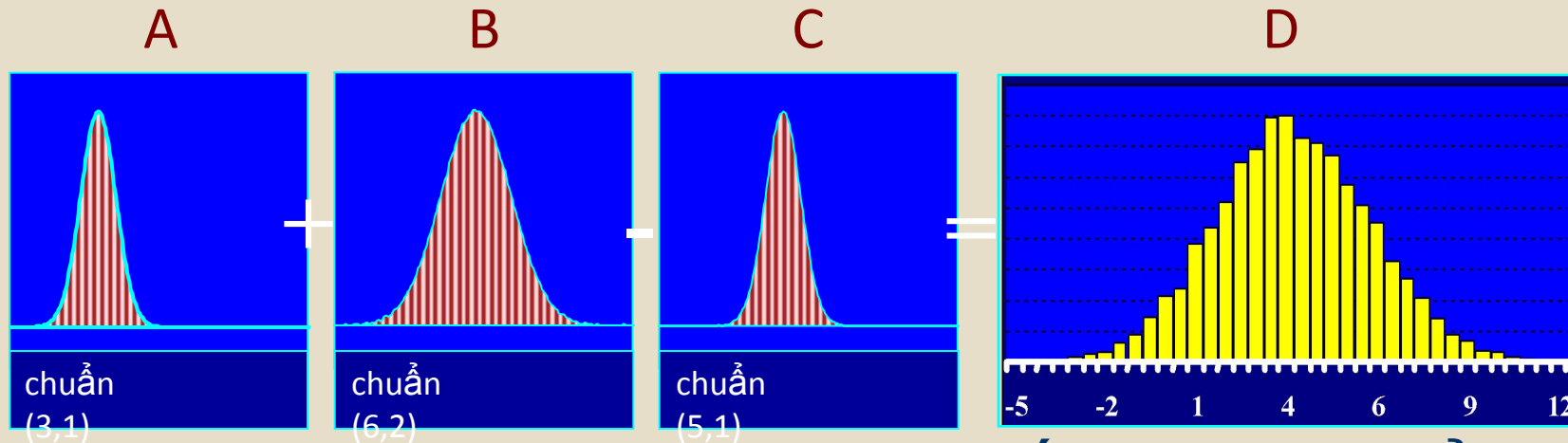
- Phương pháp mô phỏng Monte Carlo

- Quá trình lựa chọn ngẫu nhiên được lặp lại nhiều lần
 - nhiều tình huống
- Mỗi giá trị đại diện cho một kịch bản có thể xảy ra
- Cùng với nhau, các kịch bản này đưa ra một loạt các giải pháp khả thi
- Một số giải pháp có nhiều khả năng xảy ra hơn và một số ít khả năng xảy ra hơn – phân bố xác suất

- Phương pháp mô phỏng Monte Carlo

- Phân tích Monte Carlo cho phép chúng tôi mô phỏng sự thay đổi và không chắc chắn trong các giá trị

$$\text{Ví dụ : } D = A + B - C$$



- Phạm vi các giá trị cho "D" và xác suất xảy ra có thể được xác định.

- Phương pháp mô phỏng Monte Carlo

- Khi được lặp lại nhiều lần, giải pháp trung bình sẽ đưa ra câu trả lời gần đúng cho vấn đề
- Độ chính xác của câu trả lời này có thể được cải thiện bằng cách mô phỏng nhiều tình huống hơn.
 - Sẽ nói chi tiết sau

Phần mềm mô phỏng: @Risk

- Minh họa
- Cách @Risk hoạt động với Excel
- Xây dựng mô hình mô phỏng đơn giản sử dụng @Risk
- Tìm hiểu phân bố xác suất bằng cách sử dụng @Risk
- Tìm hiểu đầu ra của @Risk

Xúc xắc: Quá trình ngẫu nhiên

- Chơi trò xúc xắc 'Craps': một ví dụ đơn giản
- Xác định phân bố xác suất cho tổng (S) của hai con xúc xắc, X và Y
- Cách phân tích

$$\Pr[S = s] = \sum_x \Pr[X = x] \Pr[Y = s - x]$$

Quá trình ngẫu nhiên

- Tung xúc xắc và ghi lại kết quả

Lần	Lần		Lần	Tổng
lặp 1	1 2	+	26	= 8
2	4	+	5	= 9
3	2	+	2	= 4
4	4	+	3	= 7
5	4	+	6	= 10

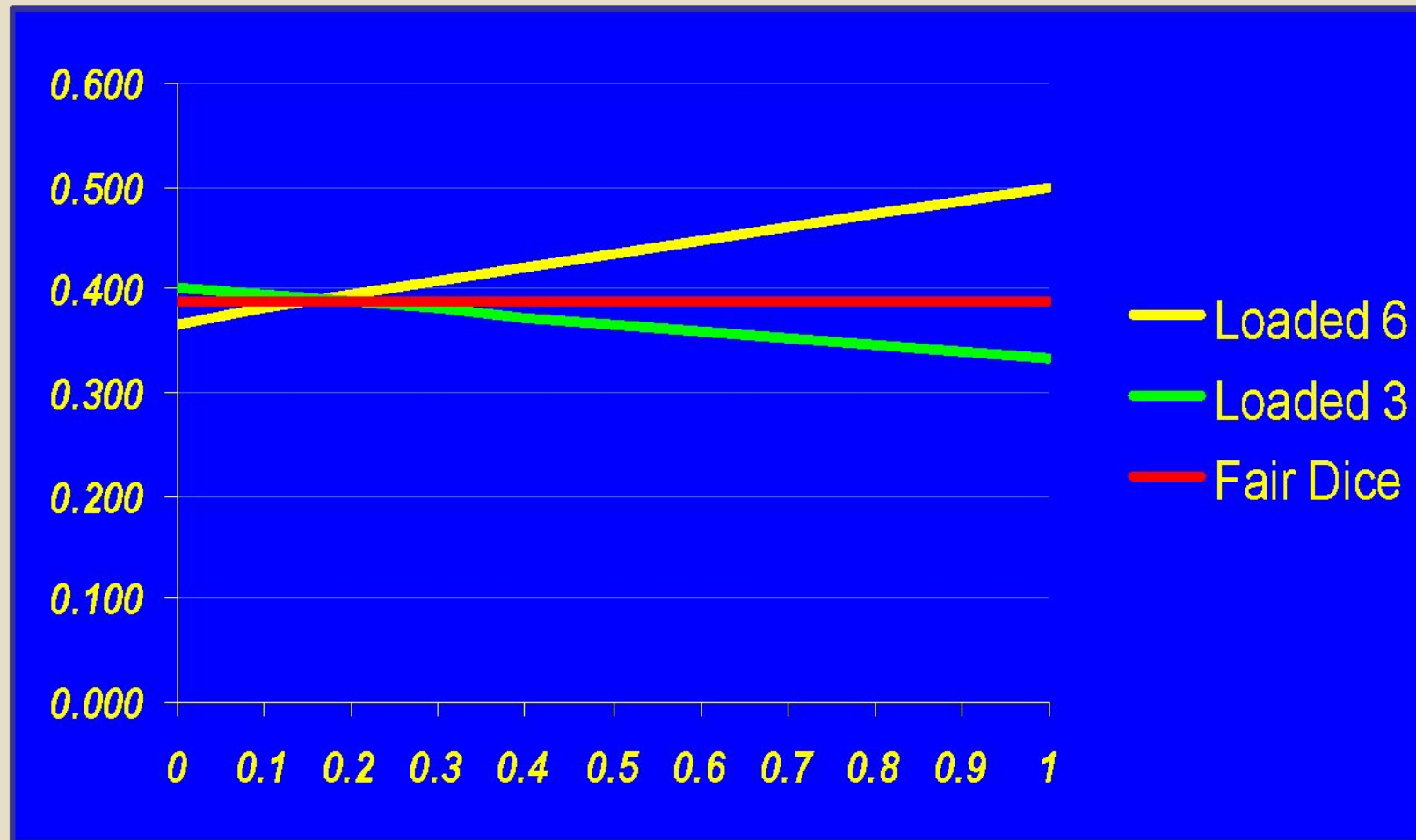
Chiến thắng: 7, 11 hoặc Doubles

	1	2	3	4	5	6
1	D					7
2		D			7	
3			D	7		
4			7	D		
5		7			D	11
6	7				11	D

Tỷ lệ thắng

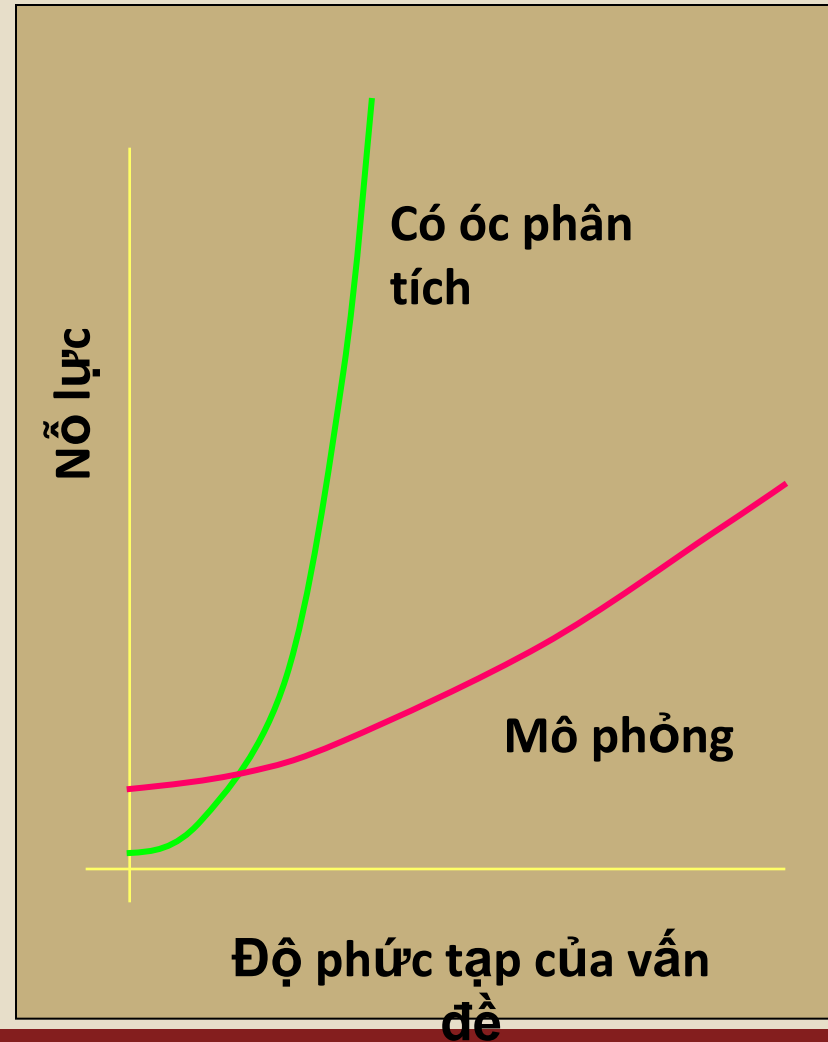
- Đối với xúc xắc công bằng (Fair Dice): 38,88%
- Với xúc xắc gian lận (Loaded dice) (một số mặt nhất định sẽ xuất hiện nhiều hơn các mặt còn lại):
 - Mặt 3: 25% thời gian 38,3%
 - Mặt 3: 35% thời gian
 - Mặt 6: 25% thời gian 40,0%
 - Mặt 6: 35% thời gian 41.3%

Giải pháp phân tích



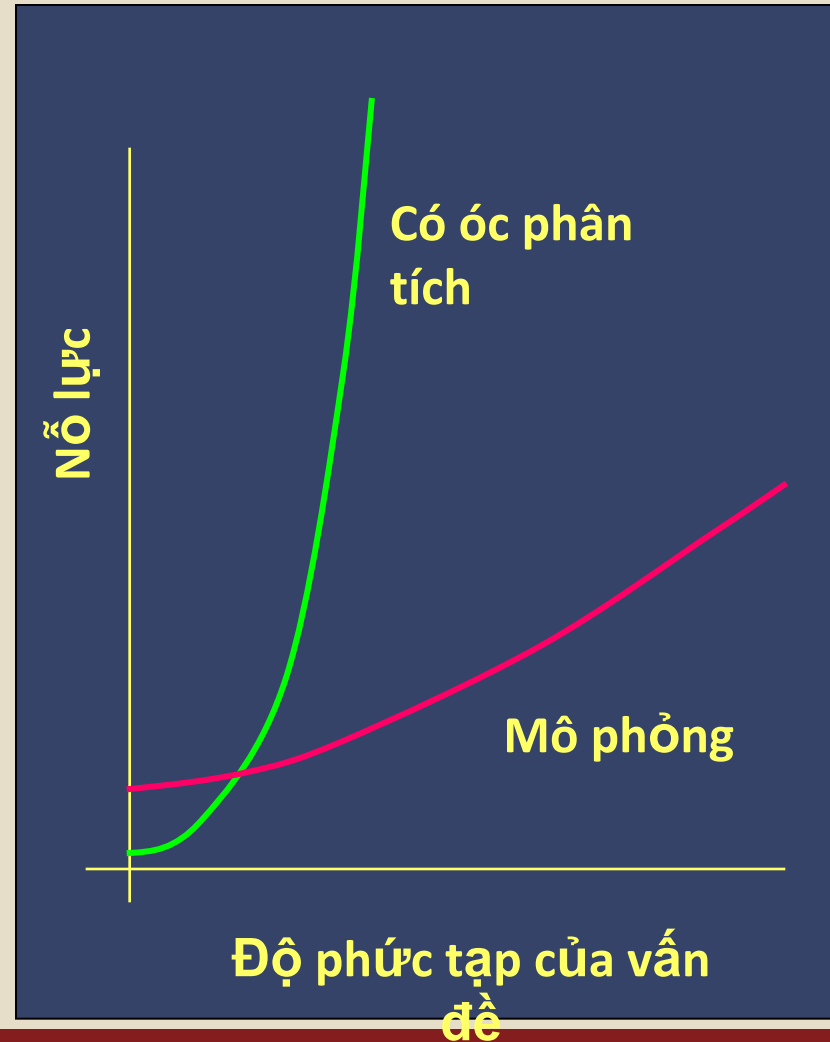
Phân tích so với mô phỏng

- Các giải pháp phân tích là chính xác, trực quan và có thể giải thích.
- Tuy nhiên, chúng đòi hỏi nỗ lực rất lớn trong các vấn đề trong thế giới thực.
- Nguồn nhân lực cần thiết thường không có sẵn.



Phân tích so với mô phỏng

- Các giải pháp phân tích là chính xác, trực quan và có thể giải thích.
- Tuy nhiên, chúng đòi hỏi nỗ lực rất lớn trong các vấn đề trong thế giới thực.
- Nguồn nhân lực cần thiết thường không có sẵn.



Phần mềm mô phỏng Monte Carlo

- Analytica™

- Phiên bản miễn phí (dành cho các mô hình có độ phức tạp từ nhỏ đến trung bình)
- Giấy phép: <https://analytica.com/products/free-edition/>

- Gói phần mềm thống kê R™

- Miễn phí và không giới hạn
- Thường được sử dụng với R Studio (phiên bản miễn phí và phiên bản thương mại)

- Risk™ (bổ trợ cho Microsoft Excel)

- Dùng thử miễn phí 15 ngày: <https://lumivero.com/resources/free-trial/atrisk/>
- Phí: 2125 USD mỗi năm

Đánh giá nguy cơ

MỘT VÀI PHÂN BỐ XÁC SUẤT ĐƠN GIẢN HƠN:

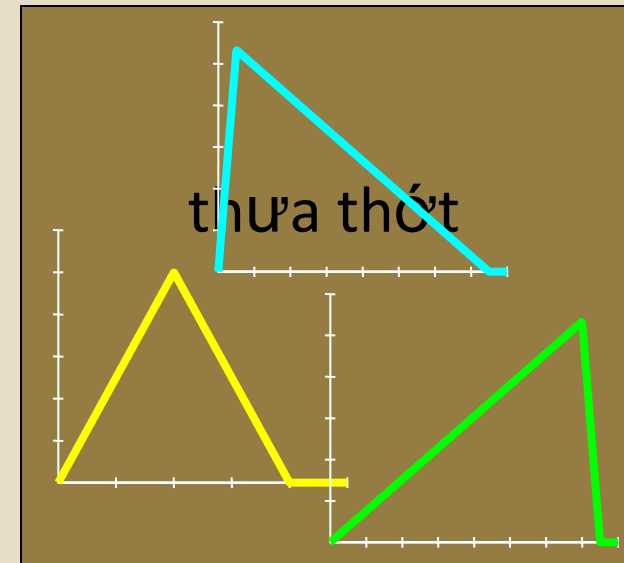
TRIANGULAR, PERT, BETA

phân bố trong đánh giá nguy cơ

- Nhiều phân bố được sử dụng trong mô hình đánh giá nguy cơ cho sức khỏe cộng đồng
- Bây giờ chúng ta sẽ xem xét thêm 3 bản phân bố để đưa ra một ví dụ nhanh về mô hình hóa nguy cơ
- Có nhiều sách thống kê về chủ đề này
- Nguồn thông tin:
 - https://en.wikipedia.org/wiki/Probability_distribution
 - <https://mathworld.wolfram.com/topics/ProbabilityandStatistics.html>

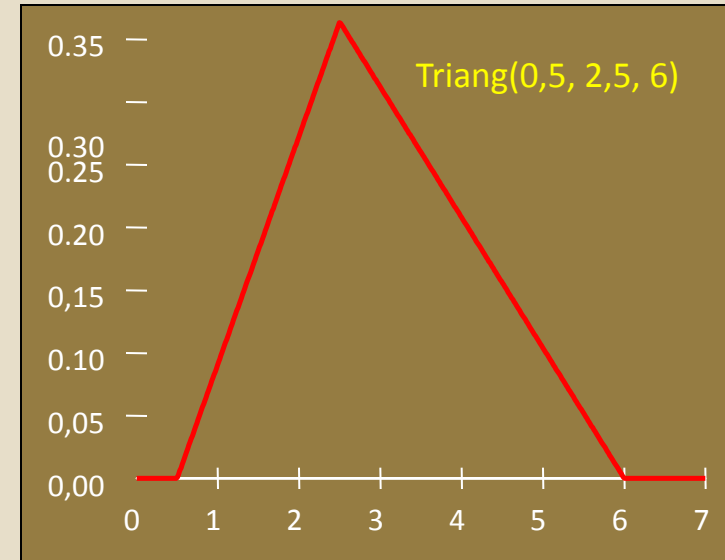
Tam giác (a,b,c)

- Thông số: tối thiểu (**a**); một (**b**); tối đa (**c**)
- Liên kết các điểm
- Liên tục
- Giới hạn
 - Miền: $(a \leq x \leq b)$
- Trung bình = $(a+b+c)/3$
- Thường được sử dụng khi dữ liệu
 - “Mô hình thô”



Ví dụ về phân bố tam giác

- Một cuộc khảo sát bệnh nhân cho thấy thời gian chờ đợi nhiều khả năng nhất trong phòng cấp cứu (Emergency Room - ER là 2,5 giờ (tối thiểu là 0,5 và tối đa là 6 giờ))
- Tam giác(0,5,2,5,6)
- Trung bình= $(a+b+c)/3$
3 giờ
=RiskTriang(a,b,c)

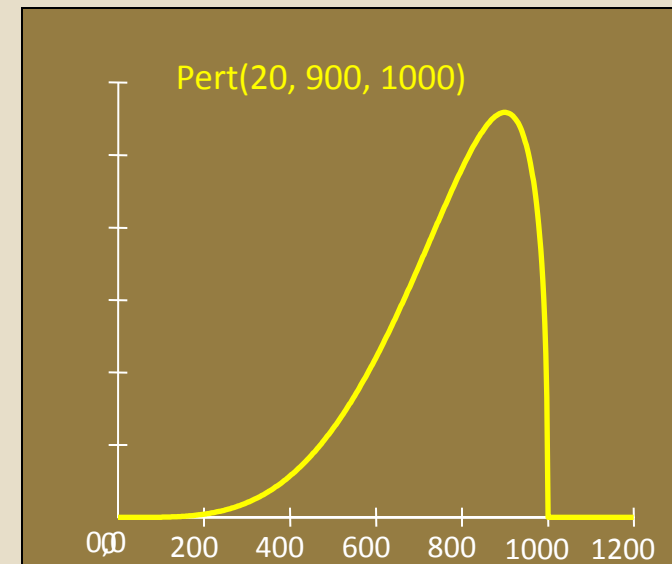


(A) (B) (C)

- Thông số: tối thiểu (**a**); một (**b**); tối đa (**c**)
- Liên kết các điểm trong hình dạng giống như “quả chuông”
- Liên tục
- Giới hạn
 - Miền: $(a \leq x \leq b)$
- Trung bình = $(a + 4 * b + c) / 6$
- Cũng được sử dụng khi dữ liệu thưa thớt –
 - “Mô hình thô”
 - Tương tự, nhưng trợn tru hơn so với phân bố hình tam giác.

Ví dụ về phân bố Pert

- Nồng độ trung bình của salmonella trong trứng sống bị ô nhiễm vẫn chưa được biết. Nó được cho là có tối thiểu 20cfu, tối đa 1000 cfu và mức rất có thể là 900cfu. Mức độ cho mỗi quả trứng là bao nhiêu?
- PERT (20.900.1000)
- **Trung bình** = $(a+4*b+c)/6$
= $(20+4*900+1000)/6$
= 770 cfu mỗi trứng
= RiskPert(a,b,c)



Beta(α , β)

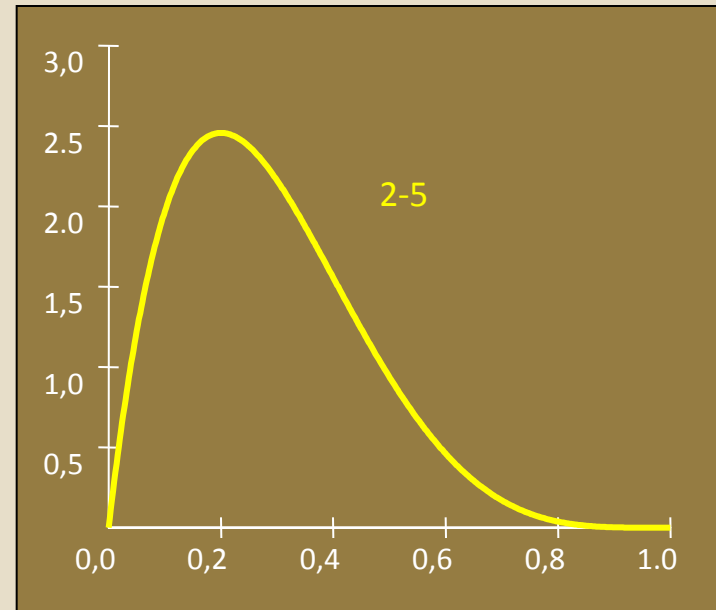
- Mô tả xác suất thành công (p) biết rằng thành công (s) xảy ra trong n thử nghiệm
- Liên tục
- Giới hạn
 - Tên miền: $0 < x < 1$
- Trung bình = $\alpha / (\alpha + \beta)$
- Có thể được sử dụng để biểu hiện sự không chắc chắn về tỷ lệ mắc bệnh với kết quả xét nghiệm dương tính s và âm tính $n-s$ mà không biết trước ($\alpha = s + 1$, $\beta = n - s + 1$)

Ví dụ: phân bố beta

- Dữ liệu khảo sát cho thấy 1 xét nghiệm máu dương tính và 4 xét nghiệm máu âm tính (vì vậy ... $s=1$, $n=5$).
- $Beta(1,1)$ đại diện cho “uncertainty distribution” (phân bố không chắc chắn)
- Chúng tôi thêm kết quả khảo sát vào phân bố không chắc chắn

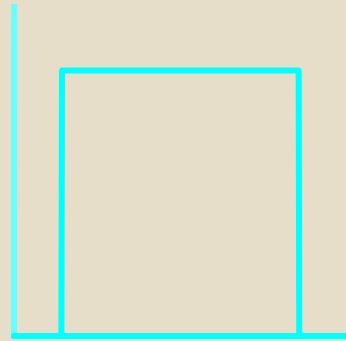
- $Prev = Beta(s+1, n-s+1)$
 $Beta(1+1, 5-1+1)$
 $Beta(2, 5)$

$=RiskBeta(s+1, n-s+1)$

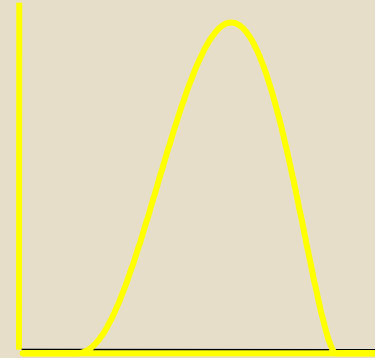


Tóm tắt: 4 phân bố

Đồng nhất (min,max)



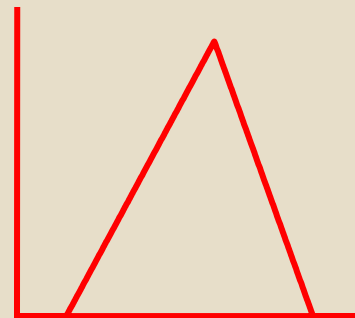
Beta(s+1,n-s+1)



Pert(min,ml, max)



Tam giác(min,ml,max)



Đánh giá nguy cơ

SO SÁNH PHƯƠNG PHÁP TẮT ĐỊNH VÀ XÁC SUẤT

So sánh các giải pháp tất định và xác suất

- Hãy nhớ kịch bản Bánh nướng nhân thịt
 - mô hình tất định
- Điều gì sẽ xảy ra nếu chúng ta đưa biến thiên vào hệ thống?
- Câu hỏi: Chúng ta có thể muốn biến thiên ở đâu trong mô hình?

Xem lại kịch bản



Kịch bản ví dụ

Giá trị trung bình

- Nồng độ vi sinh vật “X” = 2,0 log CFU/g
- Tăng trưởng vi sinh vật “X” = 1,5 log (nhân tử không đơn vị)
- Bất hoạt vi sinh vật “X” = 3,6 log (nhân tử không đơn vị)
- Kích cỡ khẩu phần = 53,33 g

Kịch bản ví dụ

Trường hợp xấu nhất (giới hạn trên)

- Nồng độ vi sinh vật “X” = 4.0 log CFU/g
- Tăng trưởng vi sinh vật “X” = 1,85 (nhân tử không đơn vị)
- Bất hoạt vi sinh vật “X” = 2,6 log (nhân tử không đơn vị)
- Kích cỡ khẩu phần = 85,00 g

Bao gồm khả năng biến đổi thông qua phân bố xác suất

Thay thế ước tính điểm bằng phân bố

- Nồng độ vi sinh vật “X” = Đồng nhất (2.0, 4.0) log CFU/g
- Tăng trưởng vi sinh vật “X” = Thay đổi nhật ký tam giác (1, 1,5, 2)
- Bất hoạt vi sinh vật “X” = Thay đổi nhật ký tam giác (2,5, 3, 5)
- Kích thước khẩu phần = Tam giác (10, 50, 100) gram

Hãy nhớ rằng để ước tính điểm...

- Nếu ước tính điểm bảo toàn giảm xuống dưới mức nguy cơ tối đa có thể chấp nhận được, thì chúng ta biết rằng nguy cơ đó thực sự có thể chấp nhận được (Không xác định được mức độ bảo vệ quá mức)
- Nếu ước tính điểm bảo toàn nằm trên mức nguy cơ tối đa có thể chấp nhận được, thì chúng ta không biết liệu nguy cơ có thực sự không thể chấp nhận được hay là kết quả của propagated conservatism (việc ước tính quá mức).

Burmester 1995

Xác suất so với ước tính điểm

- Sử dụng giá trị trung bình:

- hoàn toàn có khả năng xảy ra - thực tế
- liều cao hơn mức này thường xảy ra - không bảo toàn

- Sử dụng các ước tính thận trọng

- không có khả năng xảy ra - không thực tế
- liều cao hơn mức này hiếm khi xảy ra – “bảo toàn”
- Tuy nhiên, có thể chưa đủ bảo toàn
 - Độ tin cậy 95% có phải là thay thế cho ‘an toàn’ không?

Xác suất so với ước tính điểm

- Ước tính điểm

- Xác suất của một sự kiện xảy ra không được xem xét
- Thể hiện sự mất mát thông tin đáng kể.
- Các quyết định Quản lý nguy cơ được đưa ra với rất ít thông tin.
- Các đánh giá có thể quá bảo toàn hoặc bảo vệ không đầy đủ, tùy thuộc vào ứng dụng.

Xác suất so với ước tính điểm

- Lựa chọn ước tính bảo toàn (thận trọng) là một vấn đề gây tranh cãi:
 - Thận trọng đến mức nào?
 - Kích bản trường hợp xấu nhất (sự sáng tạo có thể là giới hạn duy nhất cho điều này)
 - Hướng dẫn quy định mặc định
 - **Nhân lên** các ước tính thận trọng thông qua kết quả đánh giá trong các ước tính nguy cơ không có bối cảnh xác suất
 - Giảm độ tin cậy của đánh giá
 - Các quyết định Quản lý nguy cơ không “dựa trên cơ sở khoa học”

Giới thiệu về Phân tích mô hình

- Đánh giá kết quả đầu ra từ @Risk
 - Tập mô phỏng
 - Đầu ra và Báo cáo Đồ họa
- Phân tích độ nhạy (Excel và @Risk)
- Phân tích tầm quan trọng (@Risk)
- Chạy nhiều tình huống

Đánh giá nguy cơ định lượng: Tập trung vào mô phỏng và đánh giá phơi nhiễm

TÌM HIỂU MÔ PHỎNG MONTE CARLO

Nhớ lại...

- **Mô phỏng** - bất kỳ phương pháp phân tích nào nhằm bắt chước một hệ thống thực tế, đặc biệt là khi các phân tích khác quá phức tạp về mặt toán học hoặc quá khó để tái tạo.
- Đối với mỗi biến xác suất, xác định các giá trị có thể có với một phân bố.
- Phân tích Monte Carlo
 - xây dựng một cách có hệ thống phân bố xác suất của các biến đầu ra, bằng cách chọn ngẫu nhiên các giá trị cho các biến đầu vào *theo phân bố xác suất của chúng*.

Và ...

- Quá trình lựa chọn ngẫu nhiên được lặp lại nhiều lần
 - nhiều tình huống ngẫu nhiên, thường được gọi là "lặp lại"
- Một số giá trị đầu ra (kết hợp đầu vào) được tạo ra thường xuyên hơn các giá trị khác – phân bố tần số này tiếp cận phân bố xác suất thực khi số lần lặp lại tăng lên (nếu chúng ta thực sự có thể biết nó bằng cách phân tích).

Tạo phân bố

- Dựa trên các số ngẫu nhiên giả
- Phương pháp ví dụ:
- Đảo ngược phân tích
- Nếu u được phân bố đều trên $(0,1)$ và Y có phân bố tích lũy F_Y thì $F_Y^{-1}(u)$ có cdf F_Y
- Phương thức
- Tạo u , xác định $x=F^{-1}(u)$, trả về x

Ví dụ

- Tạo phân bố theo cấp số nhân
- Cdf là $F(x)=1-\exp(-\lambda x)$
- Cho $u=1-\exp(-\lambda x)$
- Do đó $x=-(1/\lambda)\ln(u)$
- Thuật toán:
 - Tạo u
 - Trả về $x=-(1/\lambda)\ln(u)$

Phương pháp lấy mẫu: Monte Carlo

Lấy mẫu ngẫu nhiên đơn giản (hoặc, “Monte Carlo đơn giản”)

- Phương pháp lấy mẫu đơn giản nhất
- Mẫu $U(0,1)$ có thay thế
- Yêu cầu kích thước mẫu tương đối lớn để tạo thống kê đầu ra chính xác khi các mô hình phức tạp được mô phỏng

Phương pháp lấy mẫu: Latin Hypercube

Lấy mẫu siêu lập phương Latinh

- Phương pháp ít phổ biến hơn (nhưng phổ biến hơn trong phân tích nguy cơ do tính khả dụng trong phần mềm có sẵn)
- Diện tích dưới đường cong phân bố được phân tách theo cỡ mẫu quy định (gọi tắt là lần lặp lại)
- Các mẫu ngẫu nhiên một lần trong mỗi 'khu vực'

Hội tụ

- Khi số lần lặp lại tăng lên, số liệu thống kê (giá trị trung bình, phương sai) của phân bố đầu ra mô phỏng sẽ hội tụ về phía giải pháp phân tích chính xác.
- Độ chính xác của câu trả lời này có thể được cải thiện bằng cách mô phỏng nhiều tình huống hơn.

Cài đặt tối ưu

- Không có quy trình áp dụng chung nào để xác định cài đặt mô phỏng tối ưu
- Có một số hướng dẫn chung mà nhà phát triển mô hình có thể áp dụng để giúp đảm bảo số lần lặp lại trong mô phỏng mô hình có độ lớn thích hợp

Tối ưu là gì?

- Khi số lần lặp lại tăng lên, sự thể hiện của các phân bố đầu vào và đầu ra được cải thiện
- Khi các mô hình bao gồm các phân bố lệch, các phương trình phi tuyến tính cao hoặc các sự kiện hiếm gặp, số lần lặp lại cần thiết để đạt được sự thể hiện tốt về phân bố đầu ra sẽ cao hơn các mô hình không có các thuộc tính này
- Mục đích là để xác định điểm mà " tại đó những nỗ lực bổ sung để đạt được độ chính xác sẽ vượt quá lợi ích thu được"

Phương án thứ nhất

- Xác định “giá trị trung bình thực” ở một số lần lặp lại rất lớn, đó là một con số đủ vượt quá điểm hội tụ dự kiến của mô hình và xem xét sự thay đổi của giá trị trung bình chạy từ “giá trị trung bình thực” này và chấp nhận nó nếu nó nằm trong một số phạm vi
 - Ví dụ $\pm 1\%$

Phương án thứ hai

- Quan sát sự thay đổi trong thống kê qua việc tăng số lần lặp lại và chấp nhận kết quả khi thống kê không còn thay đổi nhiều hơn một số mức chấp nhận được (ví dụ $\pm 1\%$).
 - Không có “giá trị trung bình thực” được xác định thay vào đó, giả định được đưa ra rằng tại thời điểm thu được độ ổn định đại diện cho “giá trị trung bình thực”.
 - Thận trọng cần thiết như một mô hình có thể xuất hiện để ổn định nhưng trong các lần lặp tiếp theo khác nhau rất nhiều.
 - Các mô hình phi tuyến tính (hàm mũ, ngược) và các sự kiện hiếm gặp.
 - Sử dụng "giá trị trung bình thực sự" như một tiêu chí sẽ tránh được vấn đề này.

Các bước thực hiện nhiệm vụ

- Xác định đầu ra đại diện cho một mô hình ổn định (hoặc hội tụ)
- Xác định (các) số liệu thống kê cần theo dõi
- Xác định các tiêu chí để ước tính ổn định (hội tụ)
- Chạy mô hình nhiều lần với các cỡ mẫu khác nhau (số lần lặp lại)

Các bước để thực hiện (tiếp theo)

- Kiểm tra kết quả và xác định điểm mà mô hình hội tụ có thể chấp nhận được
- Sự hội tụ của mô hình nên được xem xét lại bất cứ khi nào có sự thay đổi đối với mô hình trong các phân bố được sử dụng để mô tả các biến trong mô hình hoặc thay đổi trong chính các phương trình mô hình.
- Kiểm tra là nếu nhiều lần chạy cho kết quả xấp xỉ nhau
 - Different number seeds (Số mỗi khác nhau; số mỗi nghĩa là số để tạo ra các kết quả tự động trong công thức)
 - Cho phép sự ngẫu nhiên!

Đánh giá nguy cơ định lượng: Tập trung vào mô phỏng và đánh giá phơi nhiễm

**TÌM HIỂU SÂU VỀ VỀ CÁC PHÂN BỐ
XÁC SUẤT THƯỜNG ĐƯỢC SỬ DỤNG**

phân bố có nguồn gốc từ quy trình

- Hình dạng của phân bố xuất phát từ toán học mô tả một hiện tượng lý thuyết
 - Còn được gọi là ‘cơ học’
- Đòi hỏi sự hiểu biết về quá trình ngẫu nhiên cơ bản và bất kỳ sự ngẫu nhiên nào là một phần của việc quan sát nó.
- Cơ sở lý thuyết cho một phân bố cụ thể có thể được sử dụng để ‘ghi đè’ sự tốt đẹp của số liệu thống kê phù hợp cho thấy các phân bố khác dường như là thích hợp hơn.

Bản phân bố 'thực nghiệm'

- Được sử dụng cho quy trình cơ bản không xác định, dữ liệu hỗn hợp hoặc khi hiện tượng quá phức tạp để gán cho một lớp lý thuyết
- Được sử dụng để nắm bắt các phán đoán chủ quan (ví dụ: niềm tin trước đây, niềm tin của chuyên gia)
- Ngoài ra, áp dụng cho các bản phân bố dựa trực tiếp trên dữ liệu, bất kể quy trình

Có nguồn gốc từ quy trình so với thực nghiệm

Có nguồn gốc từ quy trình

- Hai phần tử, Hai phần tử âm
- Số mũ
- Gamma
- Hình học, Siêu hình học
- Gumbel
- Chuẩn, log chuẩn
- Siméon-Denis Poisson
- Weibull

Thực nghiệm

- Beta *, Beta-Pert **
- Đồng nhất
- Tam giác
- PDF thực nghiệm, CDF dựa trực tiếp trên dữ liệu
- * Thường đóng vai trò chính thức trong Cập nhật Bayes
- ** Thường được sử dụng để đánh giá chuyên gia hoặc dự đoán đầu tiên

Một cách khác để phân loại các phân bố

- "Không giới hạn"
 - kéo dài từ âm đến dương vô cực
- Ràng buộc một phần
 - bị hạn chế ở một cực (thường bằng không)
- Ranh giới hai mặt
 - bị giới hạn trong giới hạn dưới và giới hạn trên
- Miền cố định
 - Có thể nhận một số lượng giá trị cố định

CÁC PHÂN BỐ CHUẨN VÀ LOGNORMAL VÀ ĐỊNH LÝ GIỚI HẠN TRUNG TÂM

phân bố chuẩn

- Đường cong đối xứng hình chuông
- chuẩn(μ, σ)
- μ là giá trị trung bình và σ là độ lệch chuẩn
- Liên tục
- "Không giới hạn"
 - Tên miền: $-\infty < x < \infty$
 - (Kurtosis = 3)

phân bố Log-Normal (Lognormal)

- Liên quan đến phân bố chuẩn:
 - Khi dữ liệu phân bố log-normally được chuyển đổi log, chúng tuân theo phân bố chuẩn
- Các tham số khác nhau được sử dụng (hãy thận trọng):
 - $\text{LogNormal}(\log\mu, \log\sigma)$
 - $\text{LogNormal}(\text{trung vị}, \text{gsd})$
- Liên tục
- Giới hạn từ dưới
 - Tên miền: $0 < x < \infty$

Định lý giới hạn trung tâm

- Nói một cách đơn giản, phân bố tổng của một số lượng đủ lớn các biến ngẫu nhiên độc lập sẽ hội tụ về phía phân bố chuẩn khi số lượng biến tăng lên.

Sản phẩm của RVs Lognormal

- Còn *sản phẩm* của một số lượng đủ lớn các biến ngẫu nhiên thì sao?

Hãy nhớ rằng:

Nếu $\text{ProdX} = X_1 * X_2 * X_3 * \dots * X_n$, sau đó

$$\text{Log}(\text{ProdX}) = \text{Log}(X_1) + \text{Log}(X_2) + \text{Log}(X_3) + \dots + \text{Log}(X_n)$$

- Vì một sản phẩm luôn có thể được viết lại dưới dạng tổng của các biến ngẫu nhiên được biến đổi theo log, CLT dự đoán rằng log của sản phẩm này sẽ được phân bố chuẩn.
 - Do đó, sản phẩm phải được phân bố chuẩn.

Định lý giới hạn trung tâm

- Lấy tổng của n biến ngẫu nhiên, mỗi biến được phân bố dưới dạng phân bố đồng đều giữa 0 và 1: Đồng nhất(0,1)
- Hãy xem xét sự phân bố của giá trị trung bình với sự gia tăng n

2

8

4

10

6

.

.

.

.

50

Tổng 50 $U(0,1)$

Kurtosis ≈ 3

**CÁC PHÂN BỐ CÓ NGUỒN GỐC TỪ QUY TRÌNH QUAN TRỌNG ĐỂ MÔ TẢ
CÁC HỆ THỐNG VẬT LÝ
VÀ QUAN SÁT CỦA CHÚNG**

phân bố tham số

Ba quy trình ngẫu nhiên riêng biệt chính được sử dụng trong đánh giá nguy cơ là:

1. Binomial (với Beta)
2. Poisson (với số mũ và Gamma)
3. Siêu hình học (ít phổ biến hơn)

Quy trình nhị thức (Bernoulli)

- Đưa ra một số thử nghiệm độc lập (n)
- Hai kết quả có thể xảy ra của mỗi thử nghiệm - thành công hay thất bại
- Một biến ngẫu nhiên nhị thức đếm số lần thành công trong số n thử nghiệm.
- Xác suất thành công = p
- Xác suất thất bại = $1-p$

Quy trình nhị thức!

Ví dụ về quy trình nhị thức

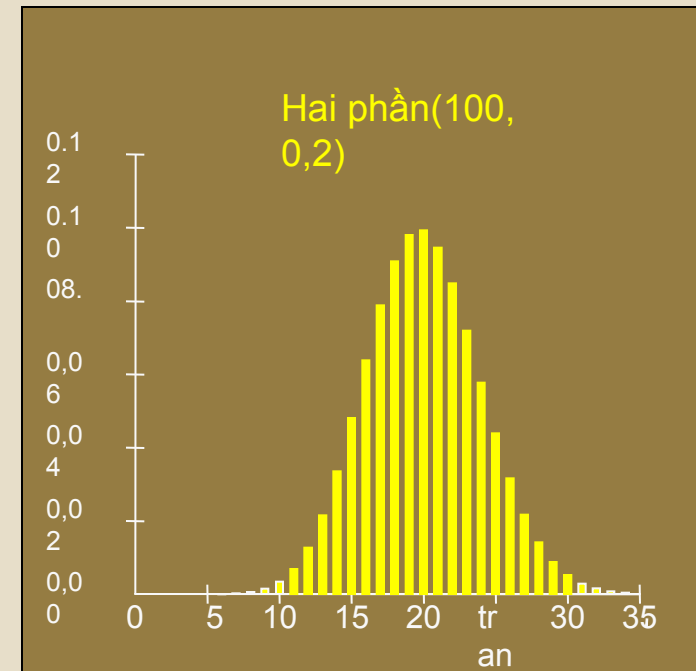
- Điều hiển nhiên... lật một đồng xu!
 - Tung đồng xu n lần sẽ có số lần s lật (coi một lật là một thành công!)
 - Có một xác suất p để có được một lật
 - Có xác suất $1-p$ để có được một sấp
- Chọn người từ đám đông - sẽ là nam hoặc nữ
- Số lượng động vật mắc “Bệnh X” được chọn từ một đàn
 - bị bệnh hoặc không

Nhị thức(n,p)

- Mô tả số lượng thành công được đưa ra trong n thử nghiệm, mỗi thử nghiệm có xác suất p thành công
- Rời rạc
- Giới hạn
 - Miền: $(0 \leq x \leq n)$
- Trung bình = np
- Bernoulli là một trường hợp đặc biệt của nhị thức với $n=1$

phân bố nhị thức (Binomial) - ví dụ

- Công tắc TV có 0,2 khả năng bị lỗi. Có bao nhiêu lỗi trong một lô 100 ngẫu nhiên?
 - phân bố nhị thức(100,0.2)
 - Trung bình=20



g
25

Tính toán nhị thức

- Xác suất một người bị dị ứng với mè là 0,3. Xác suất của ít nhất một trong một nhóm 50 người, được chọn ngẫu nhiên, bị dị ứng là bao nhiêu?
- Xác suất một người không bị dị ứng
 - $(1-p) = (1-0.3)$ hoặc 0.7
- Xác suất tất cả những người trong nhóm không bị dị ứng
 - $(1-p)^n = (1-0.3)^{50}$ hoặc $0,7^{50}$
- Xác suất ít nhất một người trong nhóm bị dị ứng
 - $1 - (1-p)^n = 1 - (1-0.3)^{50} = 0.999999982$

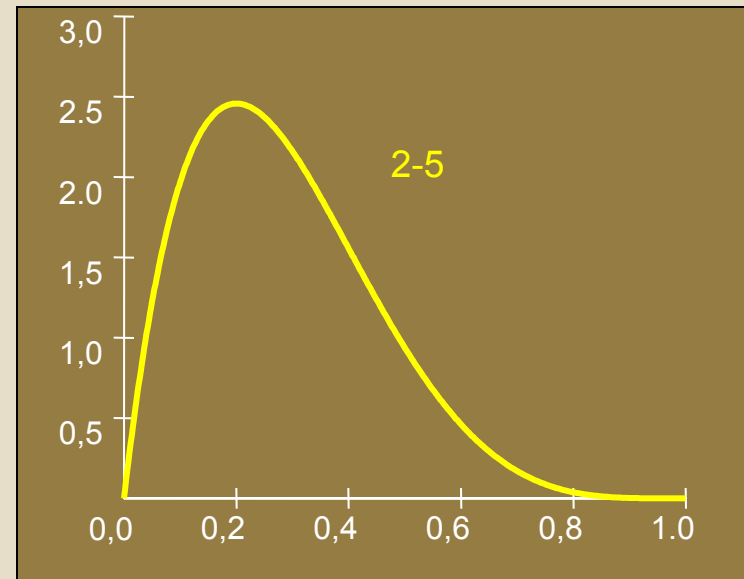
Beta(α , β)

- Xem lại bản phân bố Beta
- Chúng ta đã nói về Beta($s+1, n-s+1$)
- Nhưng nó thực sự nên được coi là:
- Beta(α , β)
 - $\alpha = s+a$
 - $\beta = n-s+b$
 - a và b phụ thuộc vào trước, trong đó trước là Beta(a,b).

Thông thường kiến thức có từ trước là 'ignorance' (sự không biết trước) được phản ánh bởi Beta(1,1) mang lại $\alpha=s+1$ $\beta= n-s +1$ sau khi quan sát n thử nghiệm.

Cập nhật phiên bản Beta với thông tin mới

- Bắt đầu với Beta(1,1) trước đó, tức là ignorance (sự chưa biết trước)
- Dữ liệu khảo sát cho thấy 1 xét nghiệm máu dương tính và 4 xét nghiệm máu âm tính, thêm chúng vào ignorance distribution để có được
 $\text{Beta}(2,5)$



Beta: Nhiều dữ liệu dẫn đến phân bố chặt chẽ hơn

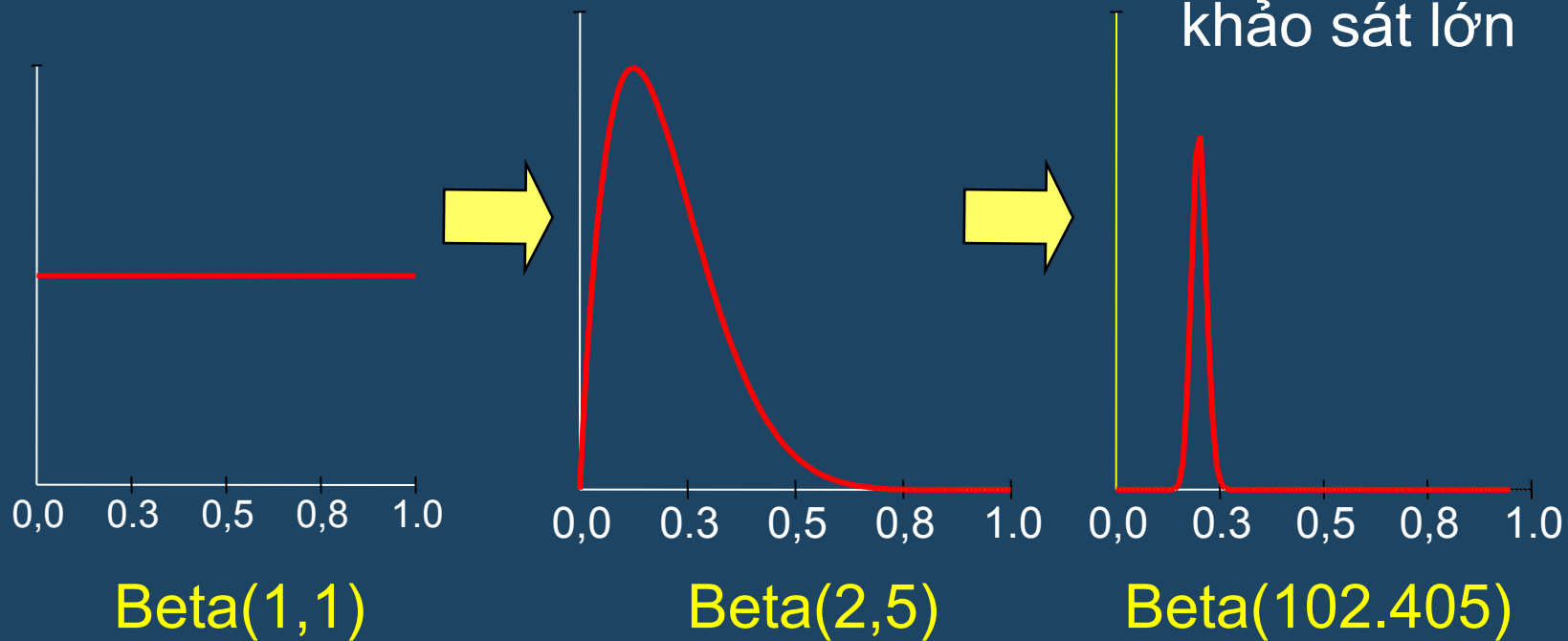
- Nếu một xét nghiệm khác được thực hiện trên 500 trang trại, và 100 là dương tính.
 - Ghi nhớ Beta($s+a$, $n-s+b$)
- Từ thử nghiệm trước $a=2$, $b=5$
- Do đó, phân bố không chắc chắn thích hợp sẽ là Beta(102.405).
 - Giá trị trung bình vẫn còn $\sim 20\%$, nhưng phân bố được phân bố hẹp, phản ánh sự tự tin tăng lên.

'Cập nhật Bayes'

Ignorance

01 mẫu nhỏ.

khảo sát lớn

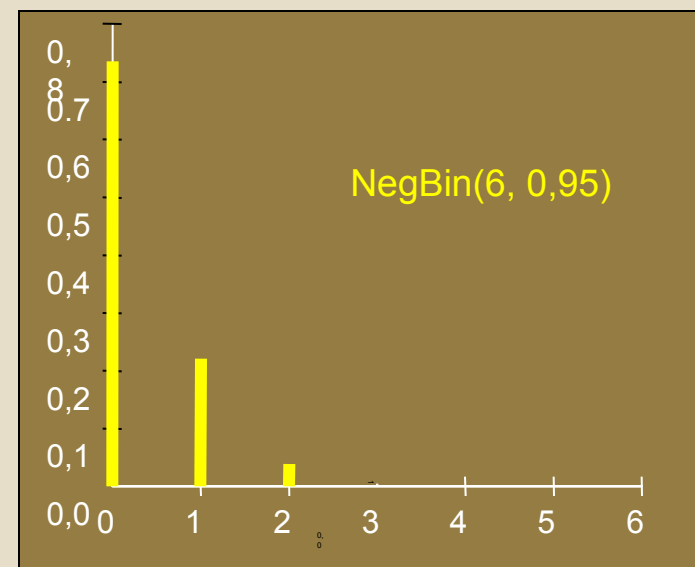


Nhị thức âm (s,p)

- Mô tả số lần thất bại trong một quá trình rời rạc với xác suất thành công p , cho đến khi thành công, mỗi quá trình dừng lại ở thành công cuối cùng
- Rời rạc
- Giới hạn ở 0
 - Tên miền: $\{0,1,2,3,\dots\}$
- Cũng có thể được sử dụng để phản ánh 'sự phân tán quá mức' trong quy trình Poisson (được thảo luận sau).

Nhị thức âm - Ví dụ

- Bệnh nhân được xét nghiệm bằng xét nghiệm với độ nhạy 95%, xét nghiệm dừng lại ở kết quả dương tính thứ 6
- Có bao nhiêu kết quả dương tính mà chúng ta có thể đã chẩn đoán sai? $\text{NegBin}(6, 0,95)$
 - Trung bình = 0,315
 - $P(\text{bỏ lỡ ít nhất một}) = 36,5\%$



Quá trình Poisson

- Quá trình Poisson là một quá trình trong đó các sự kiện xảy ra ngẫu nhiên trong một số cửa sổ cơ hội.
- phân bố Poisson đếm số lượng quan sát của quá trình trong một cửa sổ nhất định.
- Xảy ra trong một chuỗi các cơ hội
- Quan sát quá trình được mô tả bởi
 - Quá trình quan sát phụ thuộc vào cả cường độ của quá trình và mức độ quan sát (thời gian, khoảng cách, v.v...)

Thông số Poisson ()

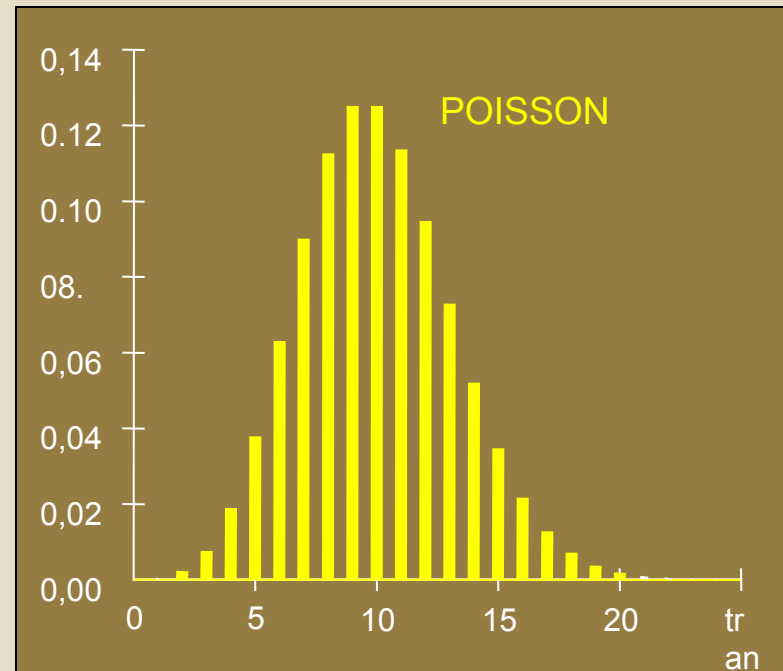
- Một quy trình Poisson có một tham số duy nhất, trong đó = **w**
- là cường độ trung bình của quá trình
 - Điều này có thể là theo thời gian hoặc không gian hoặc đơn vị đo lường khác
 - Ví dụ: # sự kiện trên một đơn vị thời gian hoặc # sự kiện trên một đơn vị không gian
- “**w**” là kích thước của cửa sổ quan sát
 - Đơn vị thời gian, diện tích, thể tích, v.v...

Quy trình Poisson

- 3 phân bố chính
 1. Poisson (λ)
 2. Gamma (α, β)
 3. Số mũ (β)

Poisson(λ)

- Mô tả số lượng sự kiện (α) xảy ra đã cho w (tức là)
- Rời rạc
- Giới hạn ở 0
 - Tên miền: $\{0,1,2,\dots\}$
- Trung bình = λ



g
25

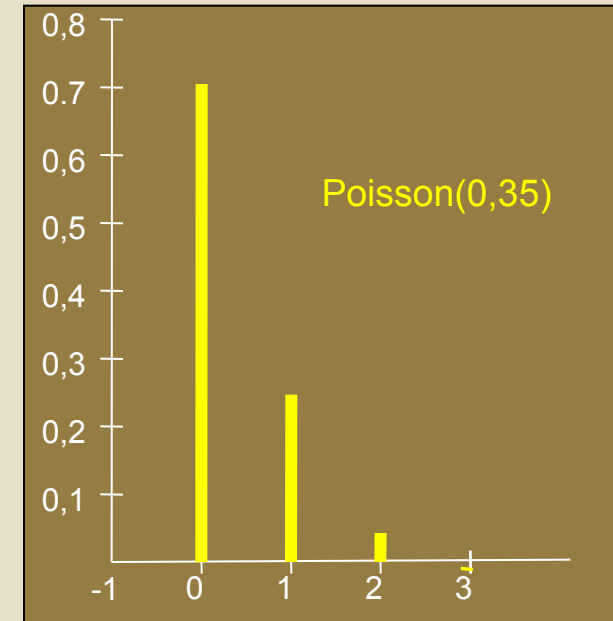
Poisson - Ví dụ

- Tai nạn xảy ra ở mức trung bình 1 trên 100 km mỗi năm ($\mu=0,01$ trên km-năm).
- Nếu chúng ta quan sát đoạn đường 1000 km trong một năm ($w = 1000*1$), chúng ta có thể mô hình hóa sự phân bố số vụ tai nạn mỗi năm được quan sát như sau:

RiskPoisson($0,01*1000$) hoặc
RiskPoisson(10)

Poisson – Ví dụ (2)

- Tác nhân gây bệnh được phân bố ngẫu nhiên trong toàn bộ sản phẩm thực phẩm đồng nhất. Nồng độ được cho là 1 CFU trên 100 g (=0,01 CFU/g). Người tiêu dùng ăn 35 g sản phẩm ($w=35$)
- Liều lượng tiêu thụ có thể được mô hình hóa như:
RiskPoisson (0,01x35)
hoặc RiskPoisson (0,35)



Tính toán Poisson

- Xác suất của bất kỳ số lượng cụ thể nào:

$$P(x) = \lambda^x e^{-\lambda} / x!$$

$$x! = x(x-1)(x-2)\dots(2)(1)$$

- Kết quả quan trọng:

- Xác suất quan sát bằng không:

$$P(x=0) = e^{-\lambda}$$

- Xác suất của ít nhất một:

$$P(x>0) = 1 - e^{-\lambda}$$

Ví dụ

- Ở quốc gia Y, số ca mắc bệnh Creutzfeldt-Jakob trung bình hàng năm là 7.
- Xác suất sẽ có 0 trường hợp trong năm tới là bao nhiêu?
 - $P(x=0) = e^{-7} = 0,000912$
- Xác suất sẽ có ít nhất một là bao nhiêu?
 - $P(x>0) = 1 - e^{-7} = 1 - 0,000912 = 0,999088$

**XẤP XỈ CỦA MỘT
PHÂN BỐ BỞI MỘT PHÂN BỐ KHÁC**

Xấp xỉ

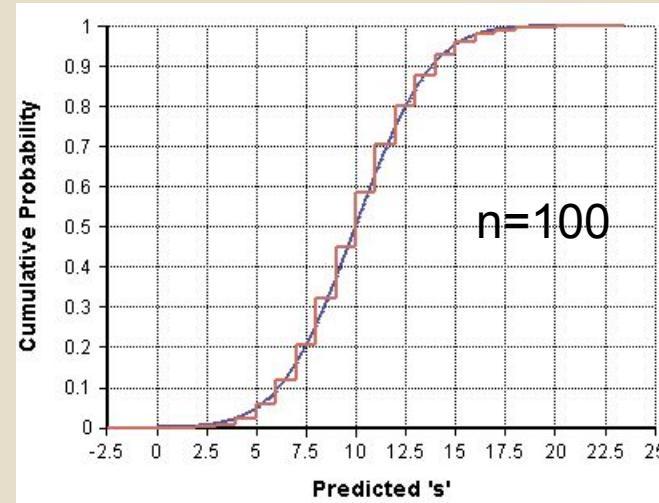
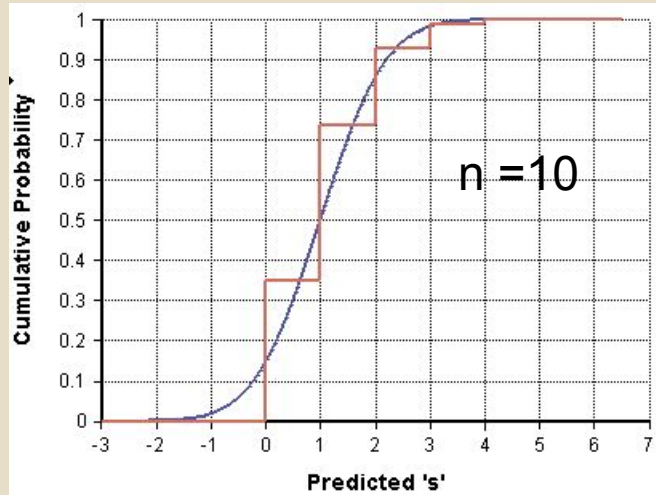
- Trong một số trường hợp, việc sử dụng xấp xỉ cho phân bố là thuận tiện
 - Nếu một tình huống yêu cầu tính toán số lượng lớn, hoặc giai thừa của số lượng lớn
 - phân bố nhị thức hoặc Poisson
- Có thể áp dụng xấp xỉ với các điều kiện nhất định được đáp ứng

phân bố nhị thức

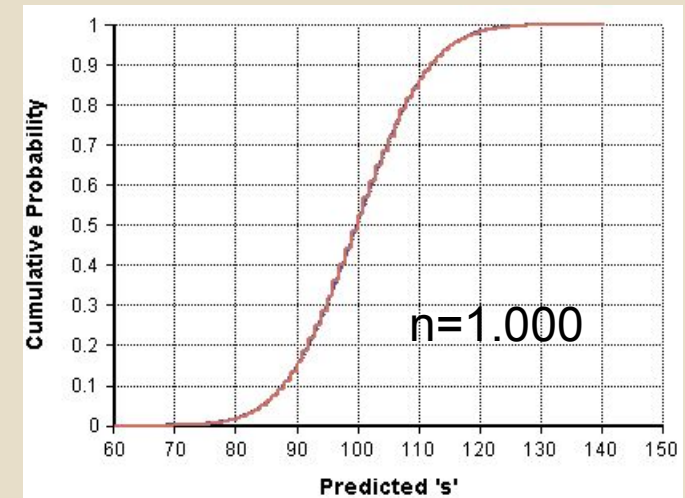
- Pmf được đưa ra bởi $f(x) = \binom{n}{x} p^x (1-p)^{n-x}$
- Liên quan đến tính toán giai thừa
- Điều gì sẽ xảy ra nếu tôi tung đồng xu 1 triệu lần...yêu cầu tính toán các giai thừa lên đến 1 triệu!
- Có thể được xấp xỉ bằng phân bố chuẩn

- Binomial(n,p) \approx Normal(np, (npq)^{0.5})
 - q=(1-p)
 - Một tiêu chí có thể sử dụng n 0,31p>0,47

Xấp xỉ chuẩn với nhị thức, $p=0,1$



— Binomial
— normal



Phương pháp đánh giá nguy cơ định lượng

MÔ HÌNH HÓA SỰ KHÔNG CHẮC CHẮN VÀ BIẾN ĐỔI (CÁC KHÁI NIỆM NÂNG CAO TRONG MÔ PHỎNG)

KHông chắc chắn

- Sự không chắc chắn được sử dụng để mô tả thực tế rằng chúng ta có kiến thức chưa đầy đủ.
- Sự không chắc chắn có thể được xử lý:
 - Một cách chính thức (ví dụ: lỗi lấy mẫu)
 - gần như chính thức (ví dụ: gợi ý của chuyên gia)
 - không chính thức (ví dụ: phán xét)

Độ biến thiên

- Sự thay đổi đề cập đến thực tế là các hiện tượng tự nhiên có sự phân tán vốn có.
- Loại phân tán này không thể giảm thông qua lấy mẫu hoặc nghiên cứu
- Giảm sự phân tán không phải là một sự cải thiện về kiến thức...
nó sẽ phản ánh sự mất mát thông tin.

Sự không chắc chắn và biến thiên

- Hãy tưởng tượng bạn đo chiều cao 10% của một lớp đại học
- Dữ liệu thể hiện sự thay đổi
- Nhưng không phải là một mẫu hoàn chỉnh, vì vậy cũng có sự không chắc chắn
- Bạn lấy mẫu nhiều hơn – vẫn ít không chắc chắn hơn
- Lấy mẫu 100% và bạn có kiến thức hoàn hảo về sự phân bố biến thiên
- Nhưng nếu bạn không thể tăng việc lấy mẫu thì sao?.....

Ví dụ về sự không chắc chắn và biến đổi

Không chắc chắn

- Hiệu quả của việc luộc trứng trong 7 phút.
- Xác suất của bất kỳ trang trại đơn lẻ nào là dương tính
- Tỷ lệ người tiêu dùng ăn sản phẩm sống

Độ biến thiên

- Thời gian người tiêu dùng luộc trứng
- Sự thay đổi về quy mô đàn gia súc
- Kích thước khẩu phần tiêu thụ

Một cách (không hoàn hảo) để phân biệt

- Xây dựng phát biểu sau đây liên quan đến tham số phân tán, P:
 - "Với thông tin hoàn hảo, P có thể được giảm xuống một giá trị duy nhất."
- Nghe có vẻ hợp lý đấy. Sự bất bình
- Nếu điều đó nghe có vẻ không phù hợp ... Độ biến thiên
- Hầu hết các hiện tượng được mô hình hóa với cả U & V
 - Thường thì sự không chắc chắn quan trọng nhất của chúng ta là mức độ biến thiên

Monte Carlo hai giai đoạn (hoặc 2-D)

- Monte Carlo hai giai đoạn thường được ứng dụng để tách biệt sự không chắc chắn (U) và sự biến thiên (V).
 - Về mặt khái niệm, cần phân biệt các thước đo về ignorance (sự chưa biết trước) và các thước đo về sự thay đổi thực sự.
 - Trên thực tế, rất khó để thực hiện hoàn toàn.
- Hai giai đoạn mô phỏng
 - Mô phỏng các giá trị cho các biến ngẫu nhiên không chắc chắn
 - Sử dụng các biến ngẫu nhiên không chắc chắn để thúc đẩy một loạt các mô phỏng chỉ tìm hiểu sự thay đổi.
 - Phân tích sự thay đổi và không chắc chắn một cách riêng biệt.
- FDA-iRISK hỗ trợ đầy đủ mô phỏng Monte Carlo 2-D
- Gói R™ mc2d cũng hỗ trợ mô phỏng Monte Carlo 2-D