THỐNG KÊ MÁY TÍNH & ỨNG DỤNG
Bài 1
GIỚI THIỆU XÁC SUẤT

Vũ Quốc Hoàng

(vqhoang@fit.hcmus.edu.vn)

FIT-HCMUS, 2018

Nội dung

- Thí nghiệm ngẫu nhiên
- Biến cố
- Xác suất
- Mô hình xác suất đơn giản
- Kỹ thuật đếm
- Công thức hợp xác suất

Thí nghiệm ngẫu nhiên

- Thí nghiệm ngẫu nhiên (random experiment) là quá trình:
  - không thể biết trước kết quả (outcome)
  - nhưng, có thể xác định trước tập các kết quả có thể
- Tập tất cả các kết quả có thể của một thí nghiệm được gọi là không gian mẫu (sample space), kí hiệu Ω
- Bước đầu tiên của việc khảo sát một thí nghiệm là xác định không gian mẫu:
  - Đúng
  - Đủ
  - Tiện lợi

Thí nghiệm ngẫu nhiên
Ví dụ

- TN1: tung đồng xu
  - Ω = {Mặt ngửa, Mặt sấp} = {N, S} = {Head, Tail} = {H, T} = {0, 1}
- TN2: học môn TKMT&UD
  - Ω = {Đậu, Rớt} = {Pass, Fail} = {1, 0}
  - Ω = {<4.5, 4.5, Trung Bình, Khá, Giỏi} (học lực)
  - Ω = {0, 0.5, 1, 1.5, …, 9.5, 10} (điểm)
- TN3: bỏ thi môn TKMT&UD
  - Ω = {Rớt}
- TN4: không nộp bài tập môn TKMT&UD
  - Ω = {Đậu, Rớt}

Thí nghiệm ngẫu nhiên
Ví dụ

- TN5: tung xúc xắc
  - Ω = {1, 2, 3, 4, 5, 6}
- TN6: tung đồng xu 3 lần
  - Ω = {HHH, HHT, …, TTT}
- TN7: tung 3 đồng xu cùng lúc
  - Ω = {HHH, HHT, …, TTT}
  - Ω = {{H, H, H}, {T, H, H}, …, {T, T, T}} = {0, 1, 2, 3} (số mặt sấp)
- TN8: tung xúc xắc 3 lần
  - $\Omega = \{(i, j, k)\colon\ i, j, k \in \{1, 2, 3, 4, 5, 6\}\} = \{1, 2, 3, 4, 5, 6\}^3$

Thí nghiệm ngẫu nhiên
Ví dụ

- TN9: tung đồng xu đến khi ra mặt sấp thì dừng
  - Ω = {T, HT, HHT, HHHT, HHHHT, …}
- TN10: đo nhiệt độ tại một địa điểm
  - $\Omega = \mathbb{R} = (-\infty, +\infty)$ (độ C)
  - $\Omega = [-273.15, +\infty)$ (độ C)
  - $\Omega = [-100, 3.6 \text{ tỉ}]$ (độ C)
  - $\Omega = [0, 1]$ (độ Planck)
- TN11: đo chiều cao của một người
  - $\Omega = \mathbb{R} = (-\infty, +\infty)$ (mét)
  - $\Omega = \mathbb{R} = [0, +\infty)$ (mét)
  - $\Omega = \mathbb{R} = [0.55, 2.51]$ (mét)

Thí nghiệm ngẫu nhiên

- Không gian mẫu rời rạc (discrete):
  - Hữu hạn (finite): $|\Omega| = n < \infty$
  - Vô hạn đếm được (countable): có tương ứng 1-1 giữa Ω và $\mathbb{N} = \{1, 2, \dots\}$
- Không gian mẫu liên tục (continuous): khoảng con của $\mathbb{R}$
- Ví dụ:
  - Hữu hạn: TN1 đến TN8
  - Vô hạn đếm được: TN9
  - Liên tục: TN10, TN11
  - Ở TN11, khi chiều cao được đo với độ chính xác đến cm thì sao?

Biến cố

- Nếu việc xảy ra hay không của một tình huống A được xác định hoàn toàn khi biết kết quả ω của một thí nghiệm T thì A được gọi là biến cố liên quan đến T
  - Nếu ω làm cho A xảy ra thì ω được gọi là kết quả thuận lợi cho A
  - Một biến cố được đặc trưng bởi tập các kết quả thuận lợi cho nó
- Biến cố (event) là tập con của không gian mẫu Ω
  - $A = \{\omega \in \Omega : \omega \text{ thuận lợi cho } A\}$
  - Với mỗi kết quả $\omega \in \Omega$ ta đồng nhất $\{\omega\}$ với ω, và gọi ω là biến cố sơ cấp
  - $\emptyset$ được gọi là biến cố không thể
  - Ω được gọi là biến cố chắc chắn

Biến cố
Ví dụ

- TN1: tung đồng xu, Ω = {H, T}, chỉ có 4 biến cố liên quan:
  - Biến cố không thể: {} = $\varnothing$ (được cả hai mặt, không ra mặt nào, …)
  - Biến cố sơ cấp "được mặt ngửa": {H}
  - Biến cố sơ cấp "được mặt sấp": {T}
  - Biến cố chắc chắn: {H, T} = Ω (được một trong hai mặt, được ít nhất một mặt, …)
- TN2: học môn TKMT&UD
  - Ω = {Đậu, Rớt}: biến cố "đậu" : {Đậu}; "được điểm cao" không là biến cố liên quan
  - Ω = {<4.5, 4.5, Trung Bình, Khá, Giỏi}: biến cố "đậu" : {4.5, TB, Khá, Giỏi}; biến cố "được điểm cao" : {Giỏi}

Biến cố
Ví dụ

- TN9: tung đồng xu đến khi ra mặt sấp thì dừng, Ω = {T, HT, HHT, HHHT, HHHHT, …}
  - Biến cố “tung không quá 5 lần”: {T, HT, HHT, HHHT, HHHHT}
  - Biến cố “có hai lần sấp”: $\emptyset$
  - Biến cố “có một lần sấp”: Ω
- TN10: đo nhiệt độ tại một địa điểm, $\Omega = \mathbb{R} = (-\infty, +\infty)$ (độ C)
  - Biến cố “nước có thể đóng băng”: $(-\infty, 0)$
- TN11: đo chiều cao của một người, $\Omega = \mathbb{R} = [0, +\infty)$ (mét)
  - Biến cố “chơi bóng rổ tốt”: $[1.8, +\infty)$

Biến cố

- "Lý thuyết biến cố" được hình thức hóa bằng "lý thuyết tập hợp". Xét thí nghiệm T với không gian mẫu Ω và các biến cố $A, B \subset \Omega$:
  - $A \subset B$: biến cố A kéo theo biến cố B; A xảy ra thì B xảy ra
  - $A = B$: biến cố A là biến cố B; A xảy ra khi và chỉ khi B xảy ra
  - Biến cố $A \setminus B$: biến cố A hiệu B; biến cố "A xảy ra nhưng B không xảy ra"
  - Biến cố $A^c = \Omega \setminus A$: biến cố đối của A; biến cố "A không xảy ra"
  - Biến cố $A \cup B$: biến cố A hợp B; biến cố "A xảy ra hoặc B xảy ra"
  - Biến cố $A \cap B$: biến cố A giao B; biến cố "A xảy ra và B xảy ra"
  - $A \cap B = \emptyset$: biến cố A và B rời nhau/xung khắc (disjoint/mutually exclusive); A và B không thể đồng thời xảy ra

Biến cố
Ví dụ

- TN5: tung xúc xắc, $\Omega = \{1, 2, 3, 4, 5, 6\}$
  - Biến cố "được mặt chẵn": $A = \{2, 4, 6\}$
  - Biến cố sơ cấp "được mặt 1": $B = \{1\}$
  - Biến cố "được mặt khác 1": $C = B^c = \Omega \backslash \{1\} = \{2, 3, 4, 5, 6\}$
  - Biến cố "được mặt lẻ": $D = \{1, 3, 5\}$
  - Biến cố A kéo theo biến cố C vì $A \subset C$
  - Biến cố "không được mặt chẵn" và "được mặt lẻ" là như nhau vì $A^c = D$
  - Biến cố "được mặt lẻ nhưng không là 1": $D \backslash B = \{3, 5\}$
  - Biến cố "được mặt chẵn hoặc 1": $A \cup B = \{1, 2, 4, 6\}$
  - Biến cố "được mặt lẻ và khác 1": $D \cap C = \{3, 5\}$
  - Biến cố "được mặt chẵn" và "được mặt 1" là xung khắc vì $A \cap B = \emptyset$

Xác suất

- Xét thí nghiệm T với không gian mẫu Ω. Một hàm P gắn mỗi biến cố $A \subset \Omega$ với số thực $P(A)$ được gọi là một độ đo xác suất (probability measure) trên Ω nếu P thỏa mãn 3 tiên đề:
  - TĐ1: Với mọi biến cố $A \subset \Omega$ ta có $0 \leq P(A) \leq 1$
  - TĐ2: $P(\Omega) = 1$
  - TĐ3: Với mọi dãy biến cố rời nhau $A_1, A_2, \dots$ (nghĩa là $A_i \cap A_j = \emptyset, i \neq j$) ta có:
  $$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$
  tức là: $P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots$
- $P(A)$ được gọi là xác suất (probability) của A và là số đo khả năng xảy ra của biến cố A khi không biết kết quả của thí nghiệm T

Xác suất
Các cách diễn giải (cách hiểu)

- Xác suất là tỉ lệ (proportion):
  - Khi khả năng xảy ra của các kết quả là như nhau
  - *Ví dụ*: lớp có 20 nữ và 30 nam, gọi ngẫu nhiên một sinh viên trong lớp, xác suất để "nữ được gọi" là 0.4, chính là tỉ lệ nữ của lớp (20/(20 + 30))
- Xác suất là tần suất (relative frequence):
  - Trong n lần thực hiện lặp lại thí nghiệm T có m lần biến cố A xảy ra thì tần suất xảy ra A là $f_A = m/n$. Khi n đủ lớn thì $P(A) \approx f_A$
  - *Ví dụ*: gieo rất nhiều lần một đồng xu không đồng chất thì thấy khoảng 70% số lần là mặt ngửa, vậy xác suất được mặt ngửa là 0.7
- Xác suất là niềm tin (belief):
  - $P(A)$ là mức độ tin tưởng (của ai đó) việc A sẽ xảy ra khi không biết kết quả của T
  - *Ví dụ*: theo tôi, xác suất để Việt Nam vô địch World Cup là gần như 0

Xác suất
Ví dụ

- Tung xúc xắc đồng chất, $\Omega = \{1, 2, 3, 4, 5, 6\}$:

$$\begin{cases} P(\{1\}) = P(\{2\}) = \cdots = P(\{6\}) \\ P(\Omega) = P(\{1\}) + P(\{2\}) + \cdots + P(\{6\}) \Rightarrow P(\{i\}) = \frac{1}{6}, i = 1..6 \\ P(\Omega) = 1 \end{cases}$$

  - Biến cố "được mặt 1": $A = \{1\}$

$$P(A) = P(\{1\}) = \frac{1}{6}$$

  - Biến cố "được mặt chẵn": $B = \{2, 4, 6\}$

$$P(B) = P(\{2\}) + P(\{4\}) + P(\{6\}) = \frac{3}{6}$$

Xác suất
Ví dụ

- Tung xúc xắc không đồng chất với khả năng ra mặt i tỉ lệ với i, $i = 1..6$:

$$\begin{cases} P(\{i\}) = c \times i, i = 1..6, c \geq 0 \\ P(\Omega) = \sum_{i=1}^{6} P(\{i\}) = \sum_{i=1}^{6} ci = c \sum_{i=1}^{6} i \Rightarrow 21c = 1 \Rightarrow P(\{i\}) = \frac{1}{21} i \\ P(\Omega) = 1 \end{cases}$$

  - Biến cố "được mặt 1": $A = \{1\}$

$$P(A) = P(\{1\}) = \frac{1}{21}$$

  - Biến cố "được mặt chẵn": $B = \{2, 4, 6\}$

$$P(B) = P(\{2\}) + P(\{4\}) + P(\{6\}) = \frac{2}{21} + \frac{4}{21} + \frac{6}{21} = \frac{12}{21}$$

Xác suất
Các tính chất

- $P(\emptyset) = 0$. *Chứng minh*: Vì $\Omega \cap \emptyset = \emptyset$ nên từ TĐ3 ta có:
$$P(\Omega) = P(\Omega \cup \emptyset) = P(\Omega) + P(\emptyset) \Longrightarrow P(\emptyset) = 0$$
- $P(A^c) = 1 - P(A)$
- Nếu $A \subset B$ thì $P(A) \leq P(B)$ và $P(B \backslash A) = P(B) - P(A)$
- Nếu A, B rời nhau thì $P(A \cup B) = P(A) + P(B)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $P(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i)$
- $P(\bigcap_{i=1}^{n} A_i) \geq 1 - \sum_{i=1}^{n} P(A_i^c)$

Mô hình xác suất đơn giản

- Khi không gian mẫu hữu hạn, $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$, độ đo xác suất được xác định bởi xác suất của các biến cố sơ cấp $p_i = P(\omega_i)$:
  - $p_i \geq 0, i = 1..n$
  - $\sum_{i=1}^{n} p_i = 1$
  - $P(A) = \sum_{\omega_i \in A} p_i$
- Khi không gian mẫu hữu hạn và các biến cố sơ cấp đồng khả năng, ta có mô hình xác suất đơn giản:
  - $p_i = \frac{1}{n} = \frac{1}{|\Omega|}, i = 1..n$
  - $P(A) = \frac{|A|}{|\Omega|}$, $|A|$ là số phần tử của A (số kết quả thuận lợi cho biến cố A)
  - Đếm

Mô hình xác suất đơn giản
Ví dụ

- Tung 3 đồng xu đồng chất:
  - $\Omega = \{HHH, HHT, \dots, TTT\}, |\Omega| = 8$
  - Các biến cố sơ cấp đồng khả năng
  - Biến cố "được đúng 2 mặt ngửa": $A = \{HHT, HTH, THH\}$

$$P(A) = \frac{|A|}{|\Omega|} = \frac{3}{8}$$

- Tung 3 đồng xu đồng chất:
  - $\Omega = \{0-\text{ngửa}, 1-\text{ngửa}, 2-\text{ngửa}, 3-\text{ngửa}\}, |\Omega| = 4$
  - Biến cố "được đúng 2 mặt ngửa": $A = \{2-\text{ngửa}\}$ là $P(A) = \frac{|A|}{|\Omega|} = \frac{1}{4}$
  - Lí luận sai: vì các biến cố sơ cấp không đồng khả năng nên không phải là mô hình xác suất đơn giản

Kỹ thuật đếm
Qui tắc nhân (Multiplication Rule)

- Nếu một việc T được thực hiện bằng 2 bước độc lập A, B; có m cách thực hiện A và n cách thực hiện B thì có $m \times n$ cách thực hiện T
  - $T = A \times B = \{(x, y): x \in A, y \in B\}$ thì $|T| = |A| \times |B|$
  - $|A_1 \times A_2 \times \cdots \times A_k| = |A_1| \times |A_2| \times \cdots \times |A_k|$
  - $\left|A^k\right| = |A|^k$
- *Ví dụ*: tung xúc xắc 3 lần, $\Omega = \left\{(i, j, k):\ i, j, k \in \{1, 2, 3, 4, 5, 6\}\right\} = \{1, 2, 3, 4, 5, 6\}^3$
  - $|\Omega| = 6^3 = 216$
  - Biến cố "được tổng cộng 4 nút": $A = \{(1,1,2), (1,2,1), (2,1,1)\}$

$$P(A) = \frac{|A|}{|\Omega|} = \frac{3}{216} = \frac{1}{72}$$

Kỹ thuật đếm
Qui tắc nhân

- Lấy mẫu có hoàn lại (sampling with replacement):
  - Từ tập A có n phần tử, chọn lần lượt k lần, mỗi lần một phần tử có hoàn lại. Số kết quả chọn, có kể đến thứ tự, là: n^k
- *Ví dụ*: Một hộp có 3 bi xanh và 2 bi đỏ. Bốc ngẫu nhiên 3 lần có hoàn lại.
  - $|\Omega| = 5^3$
  - Biến cố A: "Cả 3 lần bốc đều được bi đỏ" có xác suất:

$$P(A) = \frac{|A|}{|\Omega|} = \frac{2^3}{5^3} = 0.064$$

Kỹ thuật đếm
Hoán vị (Permutation)

- Lấy mẫu không hoàn lại (sampling without replacement):
  - Từ tập A có n phần tử, chọn ra lần lượt k phần tử không hoàn lại. Mỗi kết quả chọn, có kể đến thứ tự, được gọi là một chỉnh hợp chọn k của n phần tử
  - Số chỉnh hợp chọn chọn k của n phần tử là:

$$P_n^k = n(n-1)\dots(n-k+1) = \frac{n!}{(n-k)!}$$

  - Một chỉnh hợp chọn n của n phần tử được gọi là một hoán vị của n phần tử
  - Số hoán vị của n phần tử là: $P_n = n!$
- *Ví dụ*: Chọn ngẫu nhiên 3 người từ 10 người (trong đó có Bình) để trao giải nhất, nhì, ba.
  - $|\Omega| = P_{10}^3$, biến cố A: "Bình được giải nhất" có xác suất: $P(A) = \frac{|A|}{|\Omega|} = \frac{P_9^2}{P_{10}^3} = 0.1$

Kỹ thuật đếm
Tổ hợp (Combination)

- Tổ hợp (combination):
  - Từ tập A có n phần tử, chọn ra k phần tử không hoàn lại. Mỗi kết quả chọn, không kể thứ tự, được gọi là một tổ hợp chọn k của n phần tử
  - Số tổ hợp chọn k của n phần tử là:

$$C_n^k = \frac{n!}{k!\,(n-k)!}$$

- *Ví dụ*: tung đồng xu 10 lần, $\Omega = \{H,T\}^{10}, |\Omega| = 2^{10}$
  - Biến cố A: "được 3 lần ngửa"
  - Mỗi kết quả thuận lợi cho A sẽ có mặt ngửa trong 3 lần chọn từ 10 lần, như vậy: $|A| = C_{10}^3$
  - Xác suất của A: $P(A) = \frac{|A|}{|\Omega|} = \frac{C_{10}^3}{2^{10}} = 0.1172$

Kỹ thuật đếm
Ví dụ

- Một lớp có 15 nam và 30 nữ. Chọn ra ngẫu nhiên 10 học sinh đi lao động. Tính xác suất có 3 nam được chọn.
  - Kết quả của thí nghiệm là một tổ hợp chọn 10 của 45 học sinh: $|\Omega| = C_{45}^{10}$
  - Do chọn ngẫu nhiên nên ta có mô hình xác suất đơn giản
  - Gọi A là biến cố có 3 nam được chọn. Kết quả thuận lợi cho A là lựa chọn gồm 3 nam và 7 nữ
  - Có C_{15}^{3} cách chọn 3 nam từ 15 nam và C_{30}^{7} cách chọn 7 nữ từ 30 nữ
  - Theo qui tắc nhân ta có số kết quả thuận lợi cho A là: $|A| = C_{15}^{3} \times C_{30}^{7}$
  - Xác suất có 3 nam được chọn là:

$$P(A) = \frac{|A|}{|\Omega|} = \frac{C_{15}^{3} \times C_{30}^{7}}{C_{45}^{10}} = 0.2904$$

Công thức hợp xác suất

- Nếu $A_1, A_2, \dots, A_n$ rời nhau (nghĩa là $A_i \cap A_j = \emptyset, i \neq j$) thì:

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$$

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

- $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) - \dots + (-1)^{n+1} P(A_1 \cap A_2 \cap \dots \cap A_n)$

Công thức hợp xác suất
Ví dụ

- Một công ty kiểm tra chất lượng của 130 bóng đèn dựa trên 2 tiêu chí: kiểu dáng và độ sáng. Kết quả như sau:

		Kiểu dáng	
		Đạt	Không đạt
Độ sáng	Đạt	117	3
	Không đạt	8	2

- Chọn ngẫu nhiên 1 bóng đèn. Tính xác suất bóng đèn được chọn thỏa ít nhất một trong hai tiêu chí trên?
- Mô hình xác suất đơn giản, $|\Omega| = 130$

Công thức hợp xác suất
Ví dụ

		Kiểu dáng	
		Đạt	Không đạt
Độ sáng	Đạt	117	3
	Không đạt	8	2

- Đặt các biến cố:
  - A: "bóng đèn được chọn thỏa tiêu chí kiểu dáng"
  - B: "bóng đèn được chọn thỏa tiêu chí độ sáng"
- Ta có xác suất bóng đèn được chọn thỏa ít nhất một trong hai tiêu chí trên là:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{117 + 8}{130} + \frac{117 + 3}{130} - \frac{117}{130} = \frac{128}{130}$$

- Cách khác:

$$P(A \cup B) = 1 - P\big((A \cup B)^c\big) = 1 - P(A^c \cap B^c) = 1 - \frac{2}{130} = \frac{128}{130}$$