Tổng hợp công thức và ví dụ môn Xác suất Thống kê

PHẦN I: XÁC SUẤT

1. Không gian mẫu & biến cố

- Không gian mẫu: Ω – tập hợp tất cả kết quả có thể xảy ra.
- Biến cố A là tập con của Ω.

Ví dụ: Tung một con xúc xắc, Ω = {1, 2, 3, 4, 5, 6}, biến cố A = "ra số chẵn" = {2, 4, 6}.

2. Xác suất của biến cố

- P(A) = số phần tử của A / số phần tử của Ω
- Tính chất:
  - 0 ≤ P(A) ≤ 1
  - P(Ω) = 1
  - P(A^c) = 1 - P(A)
  - P(A ∪ B) = P(A) + P(B) - P(A ∩ B)

Ví dụ: Xác suất để ra số chẵn khi tung xúc xắc: P(A) = 3/6 = 0.5

3. Xác suất có điều kiện

- P(A|B) = P(A ∩ B) / P(B), với P(B) > 0

Ví dụ: Gọi A: "ra số chẵn", B: "ra số > 3". A ∩ B = {4, 6}, B = {4, 5, 6} ⇒ P(A|B) = 2/3

4. Định lý Bayes

- P(Bi|A) = P(Bi) * P(A|Bi) / Σ [P(Bj) * P(A|Bj)]

Ví dụ: Có 3 hộp A, B, C. Mỗi hộp có 10 viên bi. A: 7 đỏ, 3 xanh. B: 4 đỏ, 6 xanh. C: 2 đỏ, 8 xanh.
Chọn ngẫu nhiên 1 hộp rồi chọn 1 viên bi đỏ. Xác suất chọn từ hộp A là:
P(A|đỏ) = (1/3 * 7/10) / [(1/3 * 7/10) + (1/3 * 4/10) + (1/3 * 2/10)] = 7/13

5. Biến ngẫu nhiên, kỳ vọng & phương sai

- E(X) = Σ x_i * P(X = x_i)
- Var(X) = $E(X^2)$ - $(E(X))^2$

Ví dụ: X nhận các giá trị {1,2,3} với xác suất {0.2, 0.5, 0.3}. Khi đó:
E(X) = 1*0.2 + 2*0.5 + 3*0.3 = 2.1
$E(X^2) = 1^2*0.2 + 2^2*0.5 + 3^2*0.3 = 4.9$
⇒ Var(X) = 4.9 - 2.1^2 = 0.49

PHẦN II: THỐNG KÊ

1. Thống kê mô tả

- Trung bình mẫu: $\bar{x} = (1/n) * \Sigma x_i$
- Phương sai mẫu: $s^2 = (1/(n-1)) * \Sigma (x_i - \bar{x})^2$

Ví dụ: Dữ liệu: 2, 4, 6. Trung bình: $\bar{x} = 4$, $s^2 = ((2-4)^2 + (4-4)^2 + (6-4)^2)/2 = 4$

2. Ước lượng tham số

- Ước lượng điểm: $\mu \approx \bar{x}$, $\sigma^2 \approx s^2$
- Ước lượng khoảng cho μ:

  Nếu n lớn, biết σ: $\bar{x} \pm z_{\alpha/2} * (\sigma / \sqrt{n})$

  Nếu không biết σ: $\bar{x} \pm t_{\alpha/2,n-1} * (s / \sqrt{n})$

Ví dụ: $\bar{x} = 100$, $s = 10$, $n = 25$, $\alpha = 0.05 \Rightarrow t_{0.025, 24} \approx 2.064$

$\Rightarrow$ Khoảng tin cậy: $100 \pm 2.064*(10/\sqrt{25}) = [95.872, 104.128]$

3. Kiểm định giả thuyết

- Giả thuyết:

  $H_0: \mu = \mu_0$, $H_1: \mu \neq \mu_0$ (hai phía)
- Kiểm định:

  $Z = (\bar{x} - \mu_0) / (\sigma / \sqrt{n})$ nếu biết σ

  $T = (\bar{x} - \mu_0) / (s / \sqrt{n})$ nếu không biết σ

Ví dụ: $\bar{x} = 102$, $\mu_0 = 100$, $s = 8$, $n = 16 \Rightarrow T = (102-100)/(8/4) = 1.0$

4. Một số phân phối thường gặp

- Nhị thức: $P(X = k) = C(n, k) * p^k * (1-p)^{n-k}$
- Poisson: $P(X = k) = \lambda^k * e^{-\lambda} / k!$
- Chuẩn: $X \sim N(\mu, \sigma^2)$, chuyển chuẩn: $Z = (X - \mu) / \sigma$

Ví dụ: $X \sim N(100, 25)$, $P(X < 110) = P(Z < (110 - 100)/5) = P(Z < 2) \approx 0.9772$

Ví dụ mở rộng – PHẦN I: XÁC SUẤT

Ví dụ 1: Biến cố và Xác suất

Trong một lớp có 5 bạn nam và 7 bạn nữ. Chọn ngẫu nhiên một người.
- Gọi A: chọn được bạn nữ ⇒ P(A) = 7 / (5+7) = 7/12
- Gọi B: chọn được bạn nam ⇒ P(B) = 5/12
- Gọi C: chọn được người không phải nam ⇒ P(C) = 1 - P(B) = 7/12

Ví dụ 2: Xác suất có điều kiện

Một hộp có 4 bi đỏ và 6 bi xanh. Rút lần lượt 2 viên không hoàn lại.
Tính xác suất rút được 2 viên đỏ.
- $P(A_1)$: rút viên đỏ đầu tiên = 4/10
- $P(A_2|A_1)$: rút viên đỏ thứ hai = 3/9

⇒ $P(A_1 \cap A_2)$ = 4/10 * 3/9 = 2/15

Ví dụ 3: Biến ngẫu nhiên rời rạc

Tung 2 con xúc xắc. Gọi X là tổng điểm 2 mặt.
- Các giá trị của X: từ 2 đến 12.
- Ví dụ: P(X=7) = số cách ra 7 / 36 = 6/36 = 1/6

Ví dụ 4: Kỳ vọng và phương sai

Một trò chơi: tung đồng xu 1 lần, nếu ra sấp bạn được 5k, ngửa được 10k.
- X = {5, 10}, P(5) = 0.5, P(10) = 0.5
- E(X) = 5*0.5 + 10*0.5 = 7.5
- Var(X) = (5^2*0.5 + 10^2*0.5) - 7.5^2 = 6.25

Ví dụ mở rộng – PHẦN II: THỐNG KÊ

Ví dụ 5: Ước lượng khoảng

Một mẫu ngẫu nhiên kích thước n = 100 có trung bình mẫu $\bar{x}$ = 50, σ = 5.
- Muốn ước lượng khoảng tin cậy 95% cho μ:

⇒ z_{0.025} = 1.96

⇒ Khoảng = 50 ± 1.96 * (5/√100) = [49.02, 50.98]

Ví dụ 6: Kiểm định giả thuyết

Giả sử máy đóng gói có khối lượng trung bình chuẩn là 500g. Một mẫu 36 gói cho $\bar{x}$ = 495g, σ = 12.
- H_0: μ = 500, H_1: μ ≠ 500
- Z = (495 - 500)/(12/√36) = -5/2 = -2.5

⇒ Với α = 0.05, z_crit = ±1.96 → Bác bỏ H_0Ví dụ 7: Phân phối Poisson

Số cuộc gọi đến tổng đài trung bình là 3 cuộc/phút. Tính xác suất có đúng 5 cuộc gọi trong 1 phút.

- $\lambda = 3$, $k = 5$
- $P(X = 5) = (3^5 * e^{-3}) / 5! \approx 0.1008$

Chương 3: Một số quy luật phân phối xác suất

1. Phân phối siêu bội (Hypergeometric Distribution)

Biến ngẫu nhiên (BNN) X có phân phối siêu bội khi:
- Lấy mẫu không hoàn lại từ một tập hợp hữu hạn.
- Có tổng cộng N phần tử, trong đó có M phần tử thỏa mãn một tính chất nào đó.
- Lấy ngẫu nhiên n phần tử, X là số phần tử thỏa tính chất.

Ký hiệu: $X \sim H(N, M, n)$

Công thức xác suất:
$P(X = k) = [C(M, k) * C(N - M, n - k)] / C(N, n)$
Trong đó: $\max(0, n + M - N) \le k \le \min(n, M)$

Kỳ vọng và phương sai:
$E(X) = np$, với $p = M/N$
$V(X) = npq * (N - n) / (N - 1)$, với $q = 1 - p$

Ví dụ:

Một lô hàng có N = 50 bóng đèn, trong đó có 10 bóng hỏng. Lấy ngẫu nhiên 5 bóng để kiểm tra.
Gọi X là số bóng hỏng trong 5 bóng được chọn.

a) Tính $P(X = 2)$:
$P(X = 2) = [C(10, 2) * C(40, 3)] / C(50, 5)$

b) Tính kỳ vọng và phương sai:
$p = 10 / 50 = 0.2$
$E(X) = 5 * 0.2 = 1$
$V(X) = 5 * 0.2 * 0.8 * (45 / 49) \approx 0.7347$

2. Phân phối hình học (Geometric Distribution)

Phân phối hình học mô tả số lần thử độc lập cho đến khi thành công đầu tiên.

Ký hiệu: $X \sim G(p)$
Công thức xác suất: $P(X = k) = (1 - p)^{k - 1} * p$, với $k = 1, 2, 3, \ldots$

Kỳ vọng và phương sai:
$E(X) = 1/p$
$V(X) = (1 - p) / p^2$

3. Phân phối đều (Uniform Distribution)

a) Phân phối đều liên tục:
$X \sim U(a, b)$
$f(x) = 1 / (b - a), a \leq x \leq b$

$E(X) = (a + b) / 2$
$V(X) = (b - a)^2 / 12$

b) Phân phối đều rời rạc:
Nếu X nhận n giá trị rời rạc với xác suất bằng nhau:
$P(X = x_i) = 1 / n$

Ví dụ: Xe buýt đầu tiên qua trạm lúc 5h00 và cứ 15 phút thì có 1 chuyến đi qua. Nếu bạn A tới trạm vào một thời điểm ngẫu nhiên trong khoảng 5h00 – 5h30 thì xác suất để A phải đợi xe buýt trên 10 phút là bao nhiêu
Xe buýt qua trạm mỗi 15 phút từ 5h00 đến 5h30. Nếu bạn A đến trạm vào thời điểm ngẫu nhiên,
xác suất A phải chờ trên 10 phút là:
$P(X > 10) = (15 - 10) / 15 = 1/3$

4. Phân phối chuẩn (Normal Distribution)

$X \sim N(\mu, \sigma^2)$
Hàm mật độ: $f(x) = (1 / \sqrt{(2\pi\sigma^2)}) * e^{-((x - \mu)^2 / (2\sigma^2))}$

Chuẩn hóa:
$Z = (X - \mu) / \sigma \sim N(0, 1)$

Ví dụ:

Bài tập: Giả sử chiều cao của cây bạch đàn trong khu rừng sau 5 năm tròng là BNN có phân phối chuẩn với trung bình 7m và độ lệch chuẩn 1,5m. Chọn ngẫu nhiên một cây và đo chiều cao cây đó.

a) Tính xác suất để cây chọn được có chiều cao nhỏ hơn 8,5m
b) Chọn ngẫu nhiên 100 cây và đo chiều cao. Tính xác suất để có không quá 90 cây có chiều cao nhỏ hơn 8,5m. Nhiều khả năng nhất có bao nhiêu cây có chiều cao nhỏ hơn 8,5m trong 100 cây được chọn.
c) Tìm chiều cao t(m) tối thiểu sao cho tỉ lệ cây có chiều cao lớn hơn t không quá 1%.
   Chiều cao cây bạch đàn có phân phối chuẩn với $\mu = 7m$, $\sigma = 1.5m$

   a) $P(X < 8.5)$:
   $Z = (8.5 - 7)/1.5 = 1$
   $P(Z < 1) \approx 0.8413$

   b) Chọn 100 cây, $X \approx B(100, 0.8413) \approx N(84.13, 13.34)$
   $P(Y \leq 90) \approx P(Z \leq (90.5 - 84.13)/\sqrt{13.34}) \approx P(Z \leq 1.75) \approx 0.9599$

   c) Tìm t sao cho $P(X > t) = 0.01 \rightarrow P(Z < z) = 0.99 \rightarrow Z = 2.33$
   $t = 7 + 2.33 * 1.5 = 10.5m$

Tóm tắt Chương 4: Lý thuyết mẫu và lý thuyết ước lượng

1. Lý thuyết mẫu

- Tổng thể (population): Tập hợp tất cả các phần tử được nghiên cứu, ký hiệu kích thước là N.
- Mẫu (sample): Một phần tử được chọn từ tổng thể.
- Dấu hiệu nghiên cứu: Chia thành định tính và định lượng.

Đặc trưng của tổng thể:
- Biến ngẫu nhiên X mô hình hóa tổng thể.
- Trung bình tổng thể: $E(X) = \mu$
- Phương sai tổng thể: $V(X) = \sigma^2$
- Tỉ lệ tổng thể: p (nếu X có phân phối Bernoulli)

Đặc trưng của mẫu:
- Trung bình mẫu: $\bar{x} = (1/n)\sum xi$

- Phương sai mẫu: $s^2 = (1/n)\sum(xi - \bar{x})^2$
- Phương sai hiệu chỉnh: $s^2 = (1/(n-1))\sum(xi - \bar{x})^2$
- Độ lệch chuẩn mẫu: s
- Sai số chuẩn: $SE = s / \sqrt{n}$
- Hệ số biến thiên: $CV = (s / \bar{x}) * 100\%$
- Trung vị (median), yếu vị (mode), tứ phân vị (Q1, Q3), khoảng tứ phân vị: IQR = Q3 - Q1
- Điểm ngoại lai (outlier): nằm ngoài (Q1 - 1.5*IQR, Q3 + 1.5*IQR)

2. Phân phối xác suất của các đặc trưng mẫu

- Trung bình mẫu: Nếu $X \sim N(\mu, \sigma^2)$ hoặc $n \geq 30$ thì $\bar{x} \sim N(\mu, \sigma^2/n)$
- Tỉ lệ mẫu: $f \sim N(p, pq/n)$
- Phương sai mẫu: Nếu $X \sim N(\mu, \sigma^2)$, thì phân phối của phương sai mẫu theo phân phối Chi bình phương.

3. Lý thuyết ước lượng

Ước lượng điểm:
- Dùng giá trị thống kê từ mẫu để ước lượng tham số tổng thể.
- Các tính chất: Không chệch, vững, hiệu quả, hợp lý tối đa.

Ước lượng khoảng:
- Khoảng tin cậy: (θ_1, θ_2) với độ tin cậy $\gamma = 1 - \alpha$.
- Ước lượng tỉ lệ p:
  + Khoảng đối xứng: $p \in (F - \varepsilon, F + \varepsilon)$, $\varepsilon = z_{\alpha/2} * \sqrt{f(1-f)/n}$
  + Khoảng trái: $p \in (F - \varepsilon, \infty)$
  + Khoảng phải: $p \in (-\infty, F + \varepsilon)$

- Ước lượng trung bình tổng thể khi mẫu nhỏ và $X \sim N(\mu, \sigma^2)$:
  + Dùng phân phối t: $Q = (\bar{x} - \mu) / (s / \sqrt{n}) \sim t(n-1)$
  + Khoảng tin cậy: $(\bar{x} - \varepsilon, \bar{x} + \varepsilon)$, $\varepsilon = t_{\alpha/2}(n-1) * s / \sqrt{n}$

Giải bài tập

Bài 1

Số cá đánh dấu là 300. Trong lần bắt lại, bắt được 400 cá, trong đó có 60 cá có đánh dấu.
Tỉ lệ bắt được cá đánh dấu: 60/400 = 0.15
Gọi N là số cá trong hồ. Theo giả thiết: $300/N \approx 60/400 \rightarrow N \approx 300 * 400 / 60 = 2000$

Bài 2

n = 1000, số không nảy mầm = 140 → số nảy mầm = 860 → f = 860/1000 = 0.86
a) Với $\gamma = 98\%$, $z \approx 2.33$. Sai số: $\varepsilon = 2.33 * \text{sqrt}(0.86*0.14/1000) \approx 0.0264$
→ Khoảng: (0.86 - 0.0264 ; 0.86 + 0.0264) = (0.8336 ; 0.8864)
b) Khoảng một phía: [0.86 - 2.33 * sqrt(0.86*0.14/1000); 1] = [0.8336 ; 1]

Bài 3

n = 1600, số ủng hộ A = 960 → f = 960/1600 = 0.6
$\gamma = 97\%$ → $z \approx 2.17$
Sai số: $\varepsilon = 2.17 * \text{sqrt}(0.6*0.4/1600) \approx 0.026$
→ Khoảng tỉ lệ: (0.574 ; 0.626)
→ Suy ra số ủng hộ trong 4 triệu: (0.574 * 4,000,000 ; 0.626 * 4,000,000) = (2,296,000 ; 2,504,000)

Bài 4

a) Biết tỉ lệ ≈ 20%, $\gamma = 99\%$ → $z \approx 2.58$, $\varepsilon = 0.03$
$n = (z^2 * p * q) / \varepsilon^2 = (2.58^2 * 0.2 * 0.8)/(0.03^2) \approx 4742$
b) Nếu chưa biết p → dùng p = 0.5 để n lớn nhất → $n = (2.58^2 * 0.5 * 0.5)/(0.03^2) \approx 1856$

Bài 5

Ví dụ: đo lượng nước trong 25 quả cam. Giả sử $\bar{x} = 200$ml, s = 10ml
$\gamma = 95\%$, t (24 bậc tự do) ≈ 2.064
→ $\varepsilon = 2.064 * 10/\sqrt{25} = 4.128$
→ Khoảng tin cậy: (195.872ml ; 204.128ml)

Câu 1: (2 điểm) Để thành lập đội tuyển quốc gia về một môn học, người ta tổ chức một cuộc thi tuyển gồm 3 vòng. Vòng thứ nhất lấy 80% thí sinh; vòng thứ hai lấy 70% thí sinh đã qua vòng thứ nhất và vòng thứ ba lấy 45% thí sinh đã qua vòng thứ hai. Để vào được đội tuyển, thí sinh phải vượt qua được cả 3 vòng thi.

a) Được vào đội tuyển

b) Bị loại ở vòng thứ ba

<u>Giải</u>

Gọi V1 là biến cố thí sinh vượt qua vòng thứ nhất.

Gọi V2 là biến cố thí sinh vượt qua vòng thứ hai.

Gọi V3 là biến cố thí sinh vượt qua vòng thứ ba.

a) Gọi A là biến cố thí sinh được vào đội tuyển.

$$P(A) = P(V_1) \times P(V_2) \times P(V_3) = 0.8 \times 0.7 \times 0.45 = 0.252$$

b) Gọi B là biến cố thí sinh bị loại ở vòng thứ ba

$$P(A) = P(V_1) \times P(V_2) \times P(\overline{V_3}) = 0.8 \times 0.7 \times (1 - 0.45) = 0.308$$

Câu 2: (2 điểm) Việc tiêu dùng điện hàng tháng của các hộ gia đình ở TP Hồ Chí Minh là biến ngẫu nhiên có phân phối chuẩn với trung bình 200 (kWh) và độ lệch chuẩn là 40 (kWh). Tìm xác suất để chọn ngẫu nhiên một hộ gia đình thì hộ đó:

a) Có mức tiêu dùng điện hàng tháng trên 250 (kWh);

b) Có mức tiêu dùng điện hàng tháng dưới 180 (kWh)

<u>Giải</u>

a) Có mức tiêu dùng điện hàng tháng trên 250 (kWh);

$$P(X > 250) = 1 - P(X < 250) = 1 - \Phi\left(\frac{250 - 200}{40}\right) = 0.1056$$

b) Có mức tiêu dùng điện hàng tháng dưới 180 (kWh)

$$P(0 < X < 180) = \Phi\left(\frac{180 - 200}{40}\right) - \Phi\left(\frac{0 - 200}{40}\right) = 0.3085$$

Câu 3: (2 điểm) Giả sử ngày sinh của người dân trong một thành phố lớn có thể rơi ngẫu nhiên vào một ngày bất kỳ trong một năm (365) ngày. Chọn ngẫu nhiên 1095 người trong thành phố đó. Tính xác suất để :

a) **Có hai người có cùng ngày sinh đã cho;**
b) **Có không quá 7 người có cùng ngày sinh đã cho**

Giải:

Gọi X là BNN chỉ số người có cùng ngày sinh trong 1095 người: $X \sim B(1095; \frac{1}{365})$

a) Có hai người cùng ngày sinh đã cho.

$$P(X = 2) = C_{1095}^{2}\left(\frac{1}{365}\right)^{2}\left(\frac{364}{365}\right)^{1093}$$

b) Có không quá 7 người có cùng ngày sinh đã cho

$$P(0 \le X \le 7) = \sum_{k=0}^{7} C_{1095}^{k}\left(\frac{1}{365}\right)^{k}\left(\frac{364}{365}\right)^{1095-k}$$

Câu 4: (1 điểm) Một mẫu kích thước *n* được thành lập từ tổng thể tuân theo phân phối chuẩn với kỳ vọng μ và độ lệch chuẩn là 8. Hãy xác định *n* sao cho, với xác suất bằng 0,9524; trung bình mẫu nằm trong khoảng từ (μ - 4) đến (μ+ 4).

Giải:

$$P\left(\frac{\mu - 4 - \mu}{s}\sqrt{n} < Z < \frac{\mu + 4 - \mu}{s}\sqrt{n}\right) = 0.9524$$

Xác suất 0.9524 ứng với khoảng đối xứng mỗi bên $(1 - 0.9524)/2 = 0.0238$

$$P\left(Z < \frac{\overline{X} + \mu}{s}\sqrt{n}\right) = 0.9762$$

$$\frac{\mu - 4 + \mu}{s}\sqrt{n} = 1.98$$

$$\frac{4}{s}\sqrt{n} = 1.98$$

$$n = 16$$

Câu 5: (1 điểm) Số lỗi trên 1 mét vuông vải là một biến ngẫu nhiên tuân theo luật phân phối Poisson. Kiểm tra lô vải, người ta thấy 98% có lỗi. Vậy trung bình mỗi mét vuông vải có bao nhiêu lỗi?

Giải:

Gọi X là BNN chỉ số lỗi trên 1 mét vuông vải.

$P(X \geq 1) = 1 - P(X = 0) = 0.98$

$P(X = 0) = \dfrac{\lambda^0 e^{-\lambda}}{0!} = 0.02$

$\lambda = 3.912$

Câu 6: (2 điểm) Để nghiên cứu đường kính *X* (mm) của một loại sản phẩm do một xí nghiệp sản xuất, người ta đo ngẫu nhiên 100 sản phẩm của xí nghiệp và có kết quả cho trong bảng sau:

x_i	9,85	9,90	9,95	10,00	10,05	10,10	10,15
Tần số	8	12	20	30	14	10	6

Theo qui định, những sản phẩm có đường kính từ 9,9 mm đến 10,1 mm là những sản phẩm đạt tiêu chuẩn kỹ thuật.

a) Hãy ước lượng đường kính trung bình của những sản phẩm đạt tiêu chuẩn kỹ thuật bằng khoảng tin cậy 95%.

b) Hãy ước lượng tỷ lệ của những sản phẩm đạt tiêu chuẩn kỹ thuật bằng khoảng tin cậy 95%.

Giải:

a) Những sản phẩm đạt trung bình (Bỏ 9.85 và 10.15)

$\bar{X} = 9.994, s = 0.06$

Tra bảng được z = 1.96

Sai số $\varepsilon = 1.96 \times \frac{s}{\sqrt{n}} = 1.96 \times \frac{0.06}{\sqrt{86}} = 0.013$

Đường kính trung bình: $(9.994 \pm 0.013) => (9.981; 10.007)$

b) $f = \frac{86}{100} = 0.86$

$\varepsilon = z \times \sqrt{\dfrac{f(1-f)}{n}} = 1.96 \times \sqrt{\dfrac{0.86 \times (1 - 0.86)}{100}} = 0.068$

Tỉ lệ sản phẩm đạt tiêu chuẩn $(0.792; 0.928)$