Chương 3:

Một biến ngẫu nhiên - Mở đầu

Nguyễn Linh Trung
Trần Thị Thúy Quỳnh

Đại học Công nghệ, ĐHQGHN

Nội dung

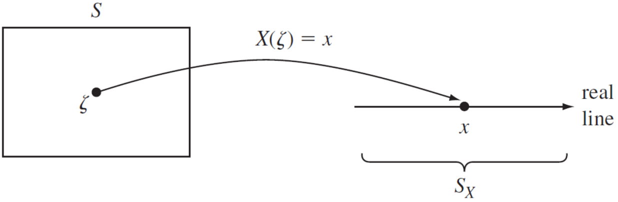
Định nghĩa, ý nghĩa của biến ngẫu nhiên

- **Định nghĩa**:
  Một biến ngẫu nhiên (random variable RV) X là một hàm $X(\zeta)$ ánh xạ một/nhiều kết quả ζ (outcome) thành một số thực x.

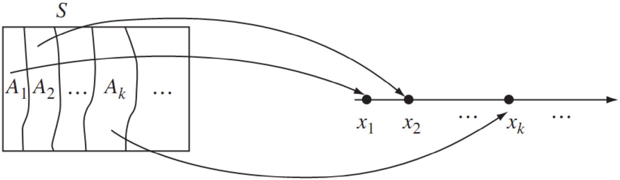
$$\boxed{\begin{aligned} X : S &\longrightarrow S_X \subset \mathbb{R} \\ \zeta &\mapsto x = X(\zeta) \end{aligned}}$$

S được gọi là "domain" của biến ngẫu nhiên X.
S_X được gọi là "range" của biến ngẫu nhiên X.

- **Ánh xạ**:
  - Ánh xạ một - một: một kết quả đơn ζ ánh xạ thành x.



  - Ánh xạ nhiều - một: nhiều kết quả trong tập con A_k thuộc S ánh xạ thành x_k.



- **Ý nghĩa:**
  - Các mô hình xác suất khác nhau chứa các đối tượng vật lý khác nhau (chọn hai bóng, tung đồng xu,...) nhưng không gian mẫu có cùng tính chất.
  - Một biến ngẫu nhiên được dùng để biểu diễn các kết quả của các không gian mẫu này bởi một biến số, để phối hợp tốt hơn với việc xác định các xác suất của các vấn đề khác nhau chỉ với một biến số chung.
  - Tính toán bằng công thức dễ hơn mô tả bằng lời.

Ví dụ

Tung một đồng xu ba lần và ghi lại mặt sấp/mặt ngửa.

- Không gian mẫu là:
  $S = \{\text{HHH,HHT,HTH,HTT,THH,THT,TTH,TTT}\}$
- Đặt X là số mặt ngửa sau ba lần tung thì:
  $S_X = \{0, 1, 2, 3\}$
- Đặt Y là số tiền tương ứng mà người chơi nhận được tương ứng với số mặt ngửa thì:
  $S_Y = \{0, 1, 8\}$

ζ:	HHH	HHT	HTH	THH	HTT	THT	TTH	TTT
$X(\zeta)$:	3	2	2	2	1	1	1	0
$Y(\zeta)$:	8	1	1	1	0	0	0	0

- Câu hỏi: Chúng ta có thể ánh xạ S bởi X' sao cho $S_{X'} = \{0, 0.1, 1, 10\}$ không?

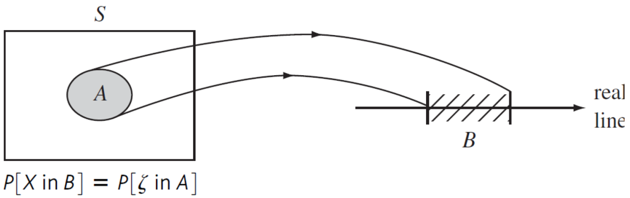
Phân loại biến ngẫu nhiên

- **Biến ngẫu nhiên rời rạc**: là biến ngẫu nhiên có giá trị thuộc tập có thể đếm được.
  Ví dụ: Gọi X là số lần gói tin cần được phát lại đến khi được nhận đúng.
  $S_X = \{1, 2, 3, \dots\}$
- **Biến ngẫu nhiên liên tục**: là biến ngẫu nhiên nhận một số vô hạn các giá trị có thể.
  Ví dụ: X là khoảng thời gian trước khi nhận được cuộc gọi tiếp theo.
- **Biến ngẫu nhiên hỗn hợp**: là biến ngẫu nhiên có một phần nhận các giá trị như biến ngẫu nhiên liên tục và phần khác nhận các giá trị như biến ngẫu nhiên liên tục.

Nội dung

Các thước đo xác suất

- Làm sao có thể tính xác suất của một biến cố $B \subset S_X$?



Tìm biến cố $A \subset S$ tương đương với biến cố $B \subset S_X$: A xuất hiện khi và chỉ khi B xuất hiện. Do đó, A chứa tất cả các kết quả ζ mà được ánh xạ vào B:

$$A = \{\zeta : X(\zeta) \in B\}$$

Do đó

$$P[B] = P[A] = P[\{\zeta : X(\zeta) \in B\}]$$

- **Hàm phân bố tích lũy**: cdf (cumulative distribution function)

$$F_X(x) = P[X \leq x] \tag{1}$$

- **Hàm mật độ xác suất**: pdf (probability density function) với biến ngẫu nhiên liên tục

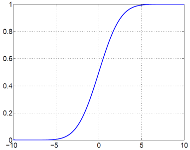
$$f_X(x) = \frac{d}{dx} F_X(x) \tag{2}$$

- **Hàm khối xác suất**: pmf (probability mass function) với biến ngẫu nhiên rời rạc

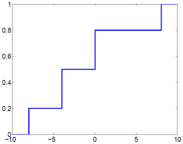
$$p_X(x) = P[X = x] \tag{3}$$

Tính chất

- **Hàm phân bố tích lũy**
  1. $0 \leq F_X(x) \leq 1$
  2. $F_X(x) \to 1$ khi $x \to +\infty$
  3. $F_X(x) \to 0$ khi $x \to -\infty$
  4. $F_X(b) - F_X(a) = \int_a^b f_X(x)dx$ với biến ngẫu nhiên liên tục



  5. $F_X(x) = \sum_{x \leq a} p_X(x)$ với biến ngẫu nhiên rời rạc



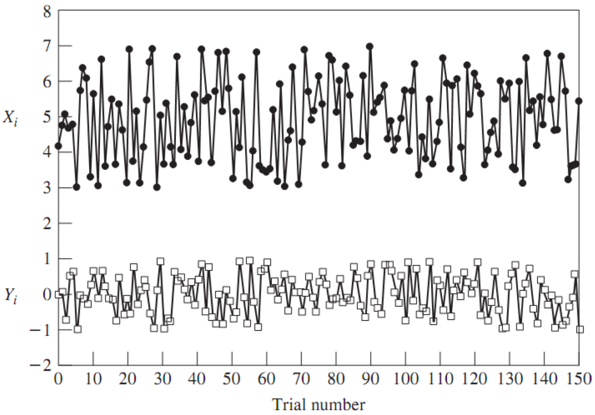
- **Hàm mật độ xác suất**
  1. $f_X(x) \geq 0$
  2. $P[a \leq X \leq b] = \int_a^b f_X(x)dx$
  3. $F_X(x) = \int_{-\infty}^x f_X(u)du$
- **Hàm khối xác suất**
  1. $p_X(x_k) \geq 0$ với tất cả các x
  2. $\sum_{x \in S_X} p_X(x) = 1$ với $S_X = \{x_1, x_2, \ldots, x_n\}$
  3. Với $B \subset S_X$, thì: $P[X \in B] = \sum_{x \in B} p_X(x)$.

Bài tập

Ví dụ 3.5, Hình 3.4(a) với $p = 1/2$.

Nội dung

Các giá trị kỳ vọng



Biểu đồ biểu diễn 150 lần lặp lại thực nghiệm đối với cả X và Y. Ta thấy rằng X tập trung xung quanh giá trị 5 trong khi Y tập trung xung quanh giá trị 0. Ngoài ra độ trải của X lớn hơn độ trải của Y.

- **Trung bình** hay **Kỳ vọng bậc 1** của biến ngẫu nhiên

$$m_X = E[X] = \begin{cases} \displaystyle\sum_{x \in S_X} x p_X(x) & \text{với} \quad X \text{ là RV rời rạc} \\ \displaystyle\int_{-\infty}^{+\infty} x f_X(x) dx & \text{với} \quad X \text{ là RV liên tục} \end{cases} \tag{4}$$

  - Giá trị kỳ vọng của RV rời rạc không tồn tại nếu tổng trên không hội tụ.
  - Theo Chương 1, giá trị kỳ vọng của RV rời rạc ứng với trung bình theo thời gian (trung bình mẫu) sau n lần lặp lại thực nghiệm:

$$\langle X \rangle_n = \sum_k x_k f_k(n) \xrightarrow{n \to \infty} \sum_k x_k p_X(k) = E[X]$$

  - Giá trị kỳ vọng biểu diễn trung tâm khối của pmf.

- Giá trị kỳ vọng của **hàm của biến ngẫu nhiên** ($g(X)$):

$$E[g(X)] = \begin{cases} \sum_{x \in S_X} g(x) p_X(x) & \text{với} \quad X \text{ là RV rời rạc} \\ \int_{-\infty}^{+\infty} g(x) f_X(x) dx & \text{với} \quad X \text{ là RV liên tục} \end{cases} \tag{5}$$

- Tính chất tuyến tính của giá trị kỳ vọng:
  1. $E[g(X) + h(X)] = E[g(X)] + E[h(X)]$
  2. $E[ag(X)] = aE[g(X)]$
  3. $E[X + c] = E[X] + c$
  4. $E[c] = c$

- **Phương sai**

$$\sigma_X^2 = \mathrm{VAR}[X] = E[(X - m_X)^2] \tag{6}$$

- σ^2 đo độ lệch của X so với giá trị trung bình m_X.
- Tính chất của giá trị phương sai:
  1. $\mathrm{VAR}[X] = E[X^2] - m_X^2$
  2. $\mathrm{VAR}[X + c] = \mathrm{VAR}[X]$

3. $\mathrm{VAR}[cX] = c^2\,\mathrm{VAR}[X]$
4. $\mathrm{VAR}[c] = 0$

- Giá trị **Độ lệch chuẩn**:

$$\sigma_X = \mathrm{STD}[X] = \sqrt{\mathrm{VAR}[X]} \tag{7}$$

- **Moment bậc** n

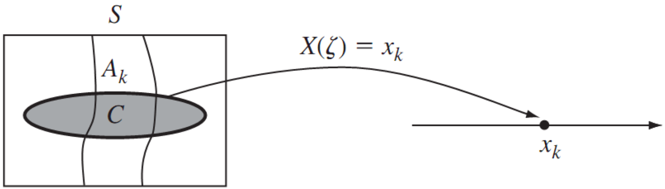
$$E[X^n] \tag{8}$$

Nội dung

PMF có điều kiện

- **PMF có điều kiện** của $X = x$ với điều kiện C là:

$$p_X(x|C) = p_X(\{X = x\}|C) = \frac{P[\{X = x\} \cap C]}{P[C]} \qquad (9)$$



Conditional pmf of X given event C.

$$p_X(x_k|C) = \begin{cases} \frac{p_X(x_k)}{P[C]}, & x_k \in C \\ 0, & x_k \notin C \end{cases} \qquad (10)$$

- Ví dụ 3.23

- Cho trước các mảnh $\{B_1, \ldots, B_n\}$ thuộc S (hay S_x), theo định lý xác suất tổng cộng, ta có:

$$p_X(x) = \sum_{i=1}^{n} p_X(x|B_i)P[B_i]$$

- **Kì vọng có điều kiện** của X với điều kiện C là:

$$m_{X|C} = E[X|C] = \sum_{x \in S_X} x p_X(x|C) = \sum_k x_k p_X(x_k|C) \tag{11}$$

- Cho trước các mảng $\{B_1, \ldots, B_n\}$ thuộc S (or S_x), biểu diễn $E[X]$ dưới dạng $E[X|B_i]$ ($E[X]$ và $E[X]$ được gọi là Giá trị kì vọng tổng cộng.

$$E[X] = \sum_k E[X|B_i]P[B_i] \tag{12}$$

- **Phương sai có điều kiện** của X với điều kiện C:

$$\boxed{\mathrm{VAR}[X|C] = E[(X - m_{X|C})^2|C]}$$

  - Ví dụ 3.26

Một số biến ngẫu nhiên quan trọng

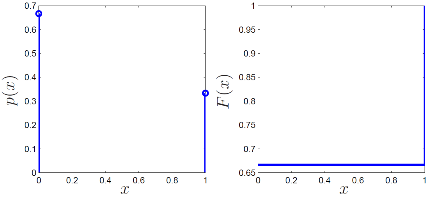
- **RV rời rạc**:
  - Bernoulli
  - Binomial
  - Geometric
  - Negative binomial
  - Poisson
  - Uniform
- **RV liên tục**:
  - Uniform
  - Exponential
  - Gaussian (Normal)
  - Gamma
  - Rayleigh
  - Cauchy
  - Laplacian

Biến ngẫu nhiên Bernoulli

- Biến ngẫu nhiên Bernoulli X được định nghĩa $X = 1$ nếu biến cố A xuất hiện và $X = 0$ nếu biến cố A không xuất hiện.

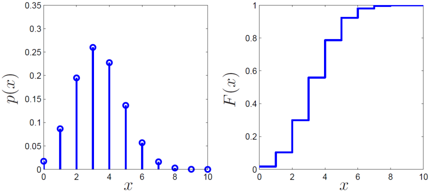
$$X(\zeta) = \begin{cases} 1 & \text{nếu} \quad X \in A \\ 0 & \text{nếu} \quad X \notin A \end{cases} \tag{13}$$

- **PMF**: $p(X = 1) = p$; $p(X = 0) = 1 - p$
- **Giá trị trung bình**: $E[X] = p$
- **Giá trị phương sai**:
  $VAR[X] = E[X^2] - m_X^2 = p - p^2 = p(1 - p)$
- Ví dụ: với $p = 1/3$



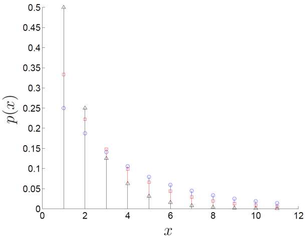
Biến ngẫu nhiên Binomial

- Biến ngẫu nhiên Binomial X được định nghĩa là số lần thành công trong chuỗi n phép thử độc lập (mỗi phép thử có xác suất thành công là p). Khi đó, $S_X = \{0, 1, \ldots, n\}$
- Ví dụ: Trong truyền dẫn nhị phân, X là số lần truyền đúng trong n lần truyền
- **PMF**: $p_X(k) = P(X = k) = C_n^k p^k (1-p)^{n-k}$ với $k = 0, 1, \ldots, n$
- **Giá trị trung bình**: $E[X] = np$
- **Giá trị phương sai**: $VAR[X] = np(1-p)$
- Ví dụ: với $p = 1/3$, $n = 10$



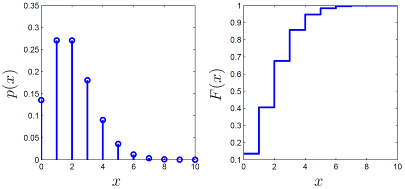
Biến ngẫu nhiên Geometric

- Phép thử Bernulli với xác suất thành công là p.
- X là số phép thử Bernoulli được thực hiện cho đến lần thành công đầu tiên. Khi đó, $S_X = \{0, 1, 2, \ldots, \}$
- Ví dụ: Trong truyền dẫn nhị phân, X là số lần gói tin được truyền lại cho đến khi nhận đúng
- **PMF**: $p_X(k) = P(X = k) = (1-p)^{k-1}p$ với $k = 0, 1, \ldots, n$
- **Giá trị trung bình**: $E[X] = \frac{1}{p}$
- **Giá trị phương sai**: $VAR[X] = \frac{1-p}{p^2}$
- Ví dụ: với $p = 1/4, 1/3$ và $1/2$



Biến ngẫu nhiên Poisson

- X là số biến cố xuất hiện trong một khoảng thời gian nhất định. Khoảng thời gian giữa hai biến cố có phân bố mũ với giá trị trung bình bằng $1/\alpha$..
- Khi đó, $S_X = \{0, 1, 2, \ldots, \}$
- Ví dụ:
  - Số lượng câu hỏi đến trung tâm chăm sóc khách hàng trong khoảng thời gian t.
  - Số lượng gói tin đến bộ ghép kênh trong khoảng thời gian t.
- **PMF**: $p_X(k) = P(X = k) = \frac{\alpha^k}{k!} e^{-\alpha}$ với $k = 0, 1, \ldots, n$ và $\alpha > 0$
- **Giá trị trung bình**: $E[X] = \alpha$
- **Giá trị phương sai**: $VAR[X] = \alpha$
- Ví dụ: với $\alpha = 2$



Biến ngẫu nhiên Uniform

- **Uniform RV rời rạc**:
  - X nhận n giá trị $\{x_1, x_2, \ldots, x_n\}$ với khả năng bằng nhau.
  - **PMF**:

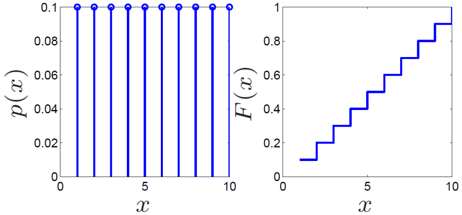
$$p_X(x) = \begin{cases} \frac{1}{n} & \text{nếu} \quad x \in \{x_1, x_2, \ldots, x_n\} \\ 0 & \text{nếu} \quad x \notin \{x_1, x_2, \ldots, x_n\} \end{cases} \tag{14}$$

- **Uniform RV liên tục**:
  - X nhận giá trị bất kỳ trong khoảng $[a, b]$ với khả năng bằng nhau.
  - **PMF**:

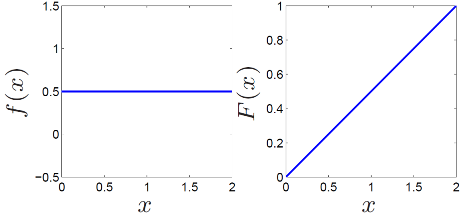
$$p_X(x) = \begin{cases} \frac{1}{(b-a)} & \text{nếu} \quad x \in [a, b] \\ 0 & \text{nếu} \quad x \notin [a, b] \end{cases} \tag{15}$$

  - **Giá trị trung bình**: $E[X] = (a + b)/2$
  - **Giá trị phương sai**: $VAR[X] = (b - a)^2/12$

- Ví dụ:
  - **Uniform RV rời rạc**: với $x = \{0, 1, 2, \ldots, 10\}$



  - **Uniform RV liên tục**: với $x \in [0, 2]$



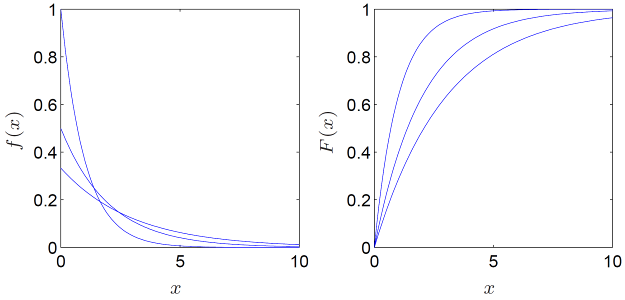
Biến ngẫu nhiên số mũ

- Hình thành khi mô tả thời gian xuất hiện giữa hai biến cố.
- Ví dụ:
  - Thời gian giữa hai yêu cầu của khách hàng để kết nối cuộc gọi.
  - Thời gian để nhân viên ngân hàng phục vụ khách hàng.
- λ là tỷ lệ biến cố nào xuất hiện.
- **PDF**:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{nếu} \quad x \geq 0 \\ 0 & \text{nếu} \quad x < 0 \end{cases} \tag{16}$$

- **Giá trị trung bình**: $E[X] = \frac{1}{\lambda}$
- **Giá trị phương sai**: $VAR[X] = \frac{1}{\lambda^2}$

- Ví dụ: với $\lambda = 1, 1/2, 1/3$



Biến ngẫu nhiên Gauss (Normal)

- Tổng của một số lớn các RV có phân bố xấp xỉ phân bố chuẩn.
- Gọi X là biến ngẫu nhiên Gauss ($X \sim \mathcal{N}(\mu, \sigma^2)$) có giá trị trung bình μ và phương sai σ^2.
- **PDF**:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \frac{(x-\mu)^2}{2\sigma^2}, -\infty < x < \infty \tag{17}$$

- **Giá trị trung bình**: $E[X] = \mu$
- **Giá trị phương sai**: $VAR[X] = \sigma^2$

- Ví dụ: với $\mu = 1; \sigma = 1, 2, 3$