Random Variables

Đặng Thanh Hải (Ph.D)
School of Engineering and Technology, VNUH
Homepage: uet.vnu.edu.vn/~hai.dang
Email: haidt82@yahoo.com

Acknowledgement

Hayder Radha

Associate Professor

Michigan State University

Department of Electrical & Computer Engineering

Content

- Definition of random variables
  Định nghĩa về biến ngẫu nhiên
- The Cumulative Distribution Function (cdf)
  Hàm phân bố tích luỹ (cdf)
- The Probability Density Function (pdf)
  Hàm mật độ xác suất (pdf)
- Moments of random variables
  Mô men của biến ngẫu nhiên
- Examples of discrete and continuous
  Các ví dụ về biến ngẫu nhiên rời rạc và liên tục
- Functions of RVs
  Hàm của các biến ngẫu nhiên

Random Variables

- **A random variable X(s) is a mapping <u>from</u> an "outcome" s (of a random experiment with sample space S) <u>to</u> a real number:**

  Một biến ngẫu nhiên X(s) là một ánh xạ từ một "Kết quả" s (của một thí nghiệm ngẫu nhiên với không gian mẫu S) tới một số thực

$$X(s):\ S \rightarrow S_X \subset R$$

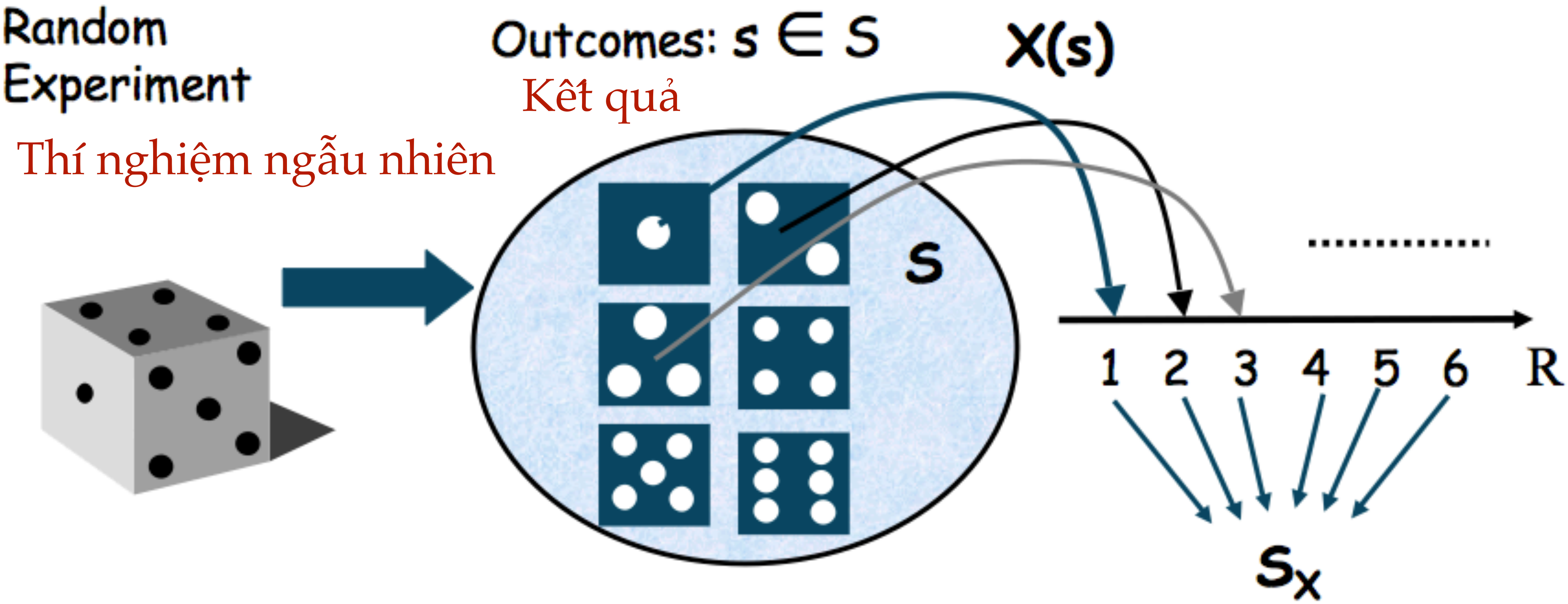
Domain　　Range

Miền xác định　　Miền giá trị

Random Variable

- A random variable $X(s): S \rightarrow S_X \subset R$

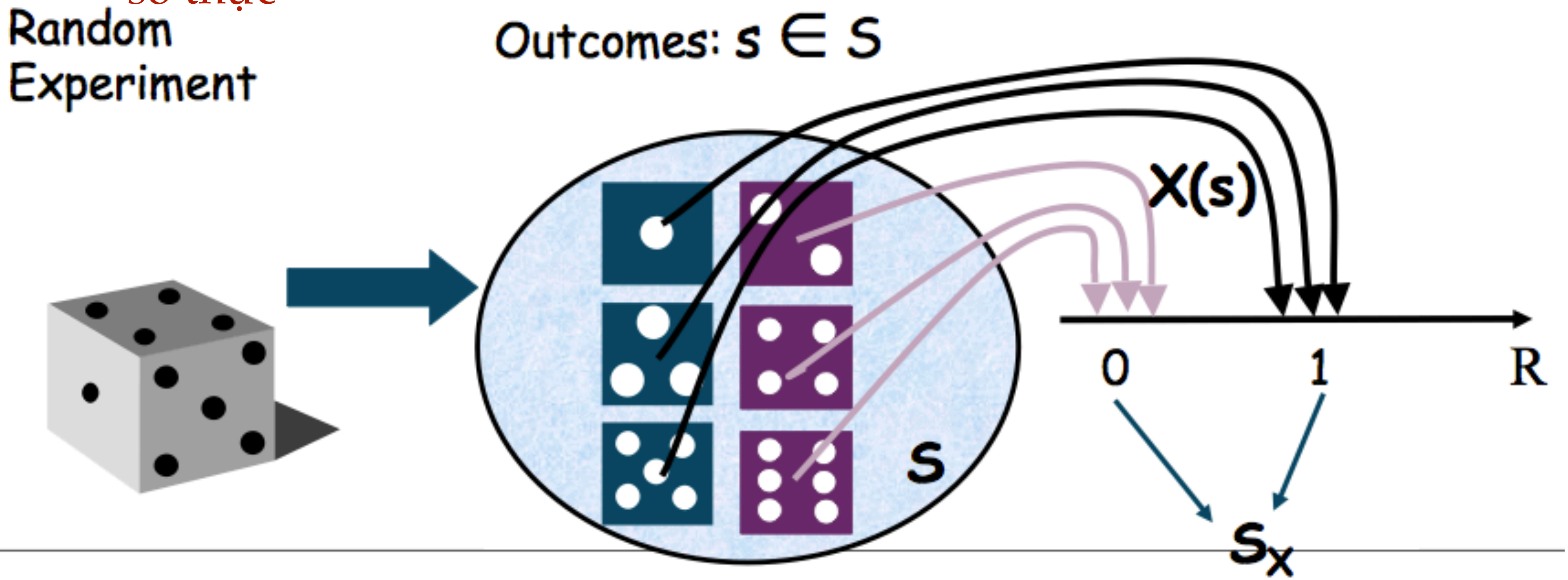
  Một biến ngẫu nhiên



Random Variable

- **Each outcome is mapped into a single real-number** Mỗi kết quả được ánh xạ vào một số thực
- **However, more than one outcome can be mapped into the same real-number**

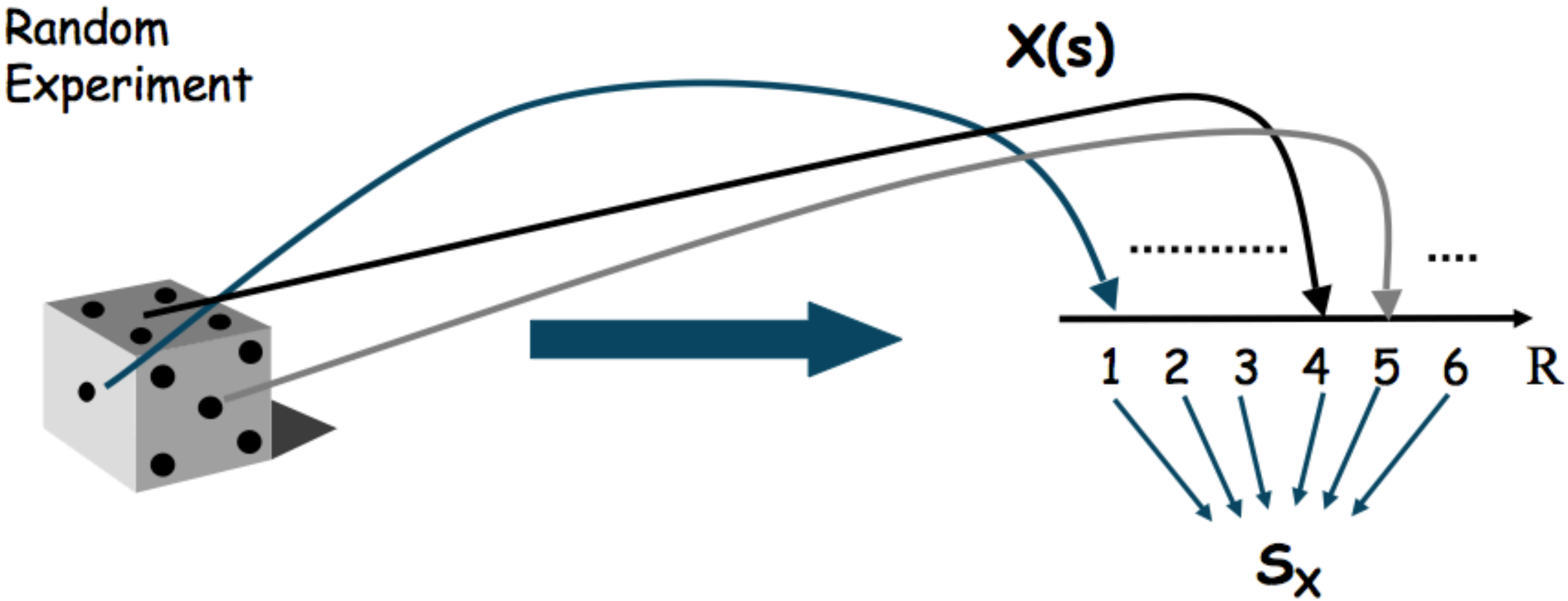
  Tuy nhiên, Nhiều hơn một kết quả có thể được ánh xạ vào cùng một số thực



Random Variables

- **Performing direct mapping**

  Thực hiện việc ánh xạ trực tiếp



Types of Random Variables

Các loại biến ngẫu nhiên

- **Discrete random variables:** Biến ngẫu nhiên rời rạc

  S_X is a countable set (finite or infinite)
  - countable finite set $S_X = \{1, 2, 3, 4, 5, 6\}$
  - countable infinite set $S_X=\{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$

  Sx là một tập đếm được (hữu hạn hoặc vô hạn)
- **Continuous random variables** Biến ngẫu nhiên liên tục

  S_X is an uncountable set with a continuous cdf
  - $S_X=(0, 1]$; $S_X=[-1, 1]$; $S_X= R$

  Sx là một tập không đếm được với một hàm cdf liên tục

Types of Random Variables

- **Hybrid discrete-continuous (mixed) random variable** Biến ngẫu nhiên lai liên tục - rời rạc (trộn)
- **S_X can be expressed as the union of two M.E. sets (or events):**

Sx có thể được biểu diễn như là hợp của 2 tập rời nhau (hay các sự kiện) $S_X = S_{XD} \cup S_{XC}$ where $S_{XD} \cap S_{XC} = \varnothing$

- $S_X=\{0\} \cup (0, 1]$; $S_X=\{0, 1\} \cup \{ (-\infty, 0) \cup (1, +\infty) \}$

Cumulative Distribution Function (cdf)

Hàm phân bố tích luỹ

- cdf $F_X(a)$ of a random variable X is defined as "the probability that X has a value smaller than <u>or</u> equal to a":

  cdf Fx(a) của một biến ngẫu nhiên X được định nghĩa là "Xác suất để X nhận giá trị nhỏ hơn hoặc bằng a"

$$F_X(a) = P[X \leq a] \qquad \forall \; -\infty \leq a \leq +\infty$$

Properties of cdf $F_X(a)=P[X \le a]$

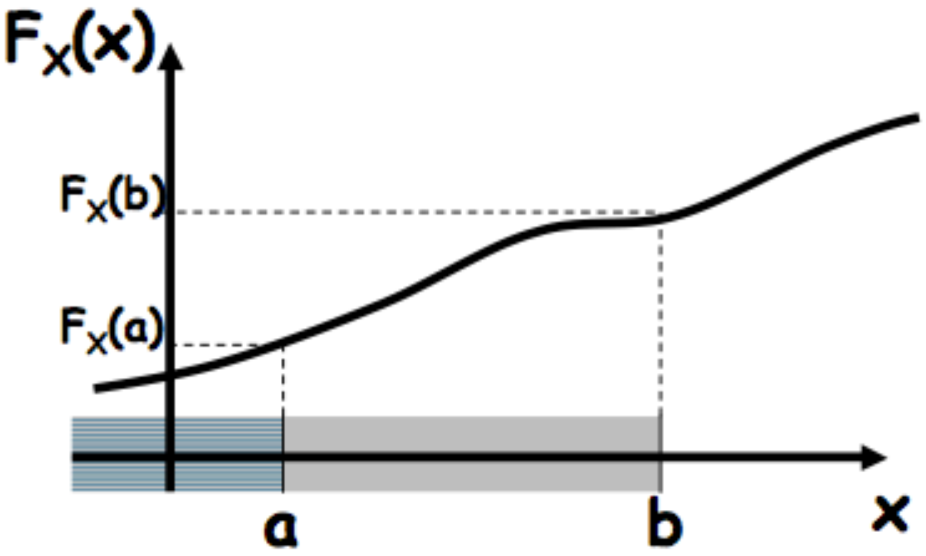
Các đặc trưng của cdf

- If $a \le b \Rightarrow F_X(a) \le F_X(b)$

  Nếu a <= b => Fx(a) <= Fx(b)

  i.e. F_X is a non-decreasing function

  Nghĩa là Fx là một hàm không giảm



Properties of cdf $F_X(a)=P[X \leq a]$

- $0 \leq F_X(a) \leq 1$
- $\lim_{a\to\infty} F_X(a) = 1$
- $\lim_{a\to-\infty} F_X(a) = 0$

(Dis)Continuity Properties of cdf

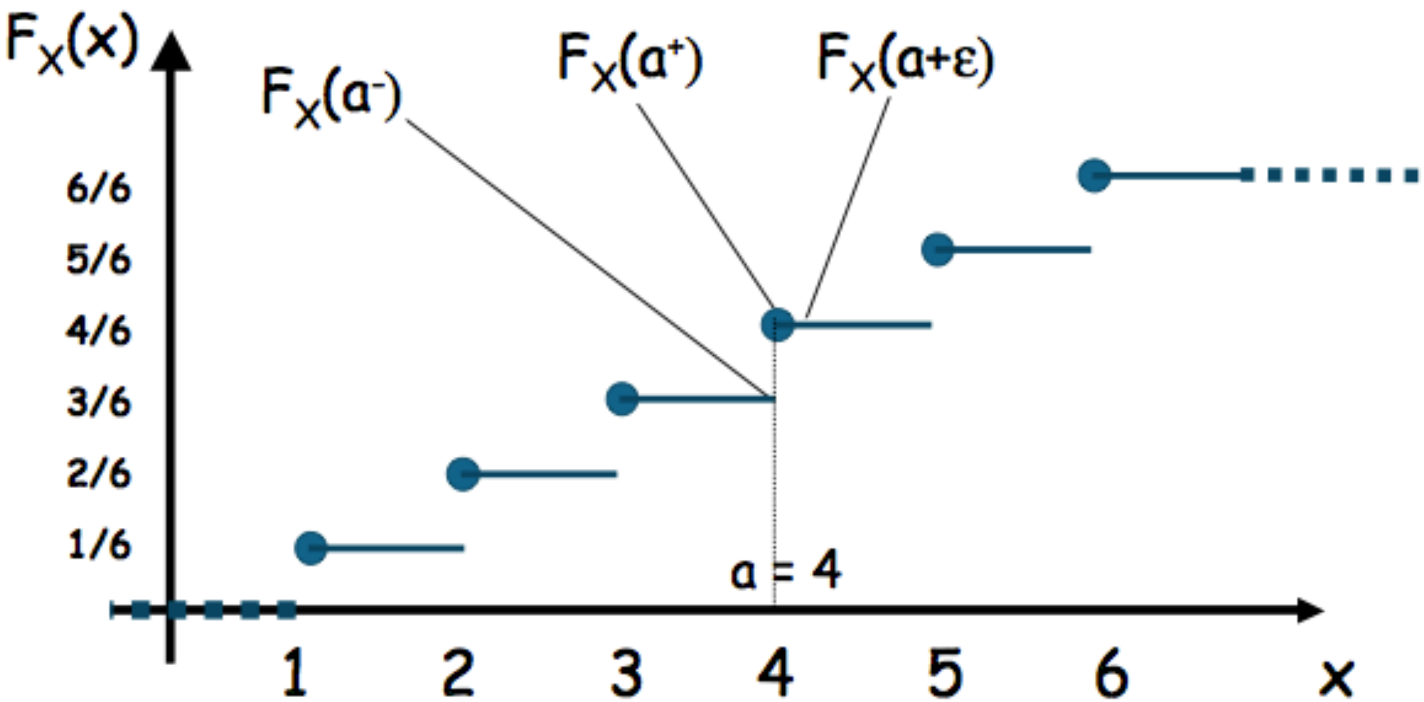
Tính (không) liên tục của cdf

- For any positive ε, $\lim_{\varepsilon \to 0} F_X(a+\varepsilon) = F_X(a^+)$

  Với bất kỳ số dương

- F_X is "continuous from the right"

  Fx "liên tục từ bên phải"

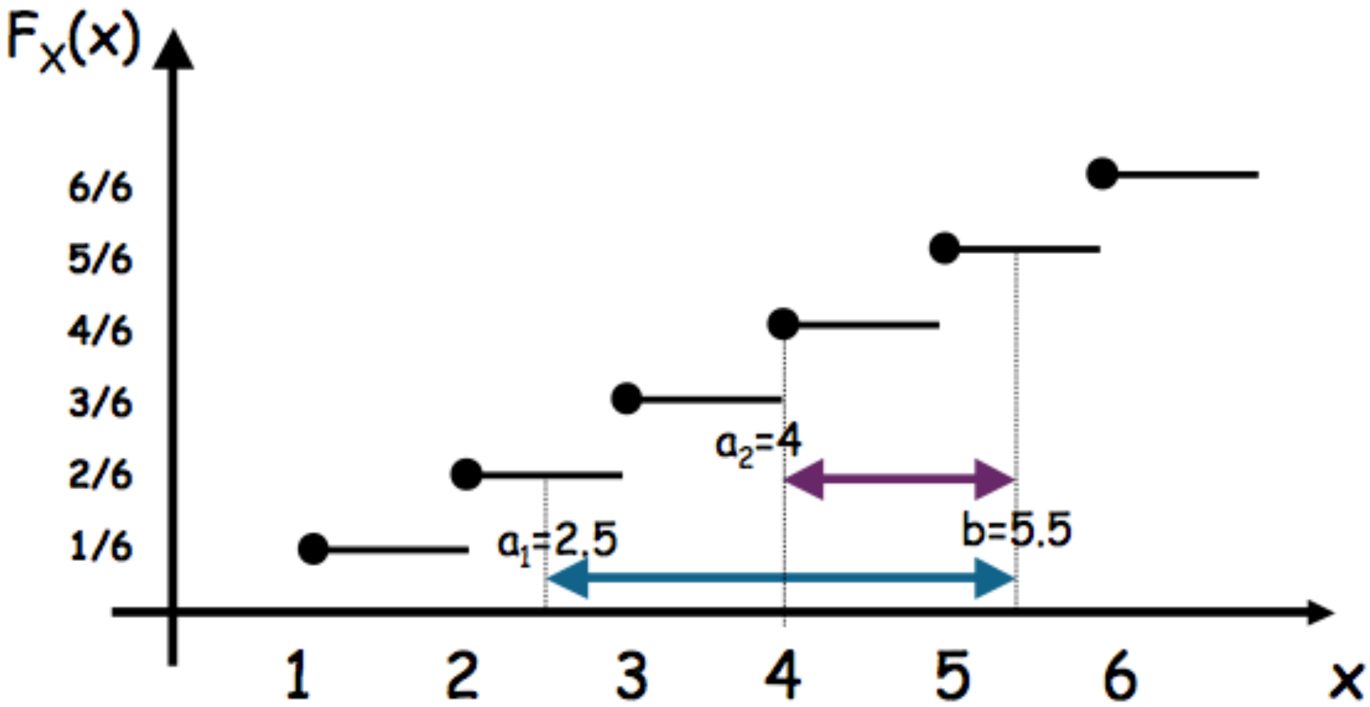


Đặc trưng của cdf Properties of cdf

- $P[a < X \leq b] = F_X(b) - F_X(a)$

$$P[2.5 = a_1 < X \leq b = 5.5] = (5/6) - (2/6) = 3/6$$

$$P[4 = a_2 < X \leq b = 5.5] = (5/6) - (4/6) = 1/6$$

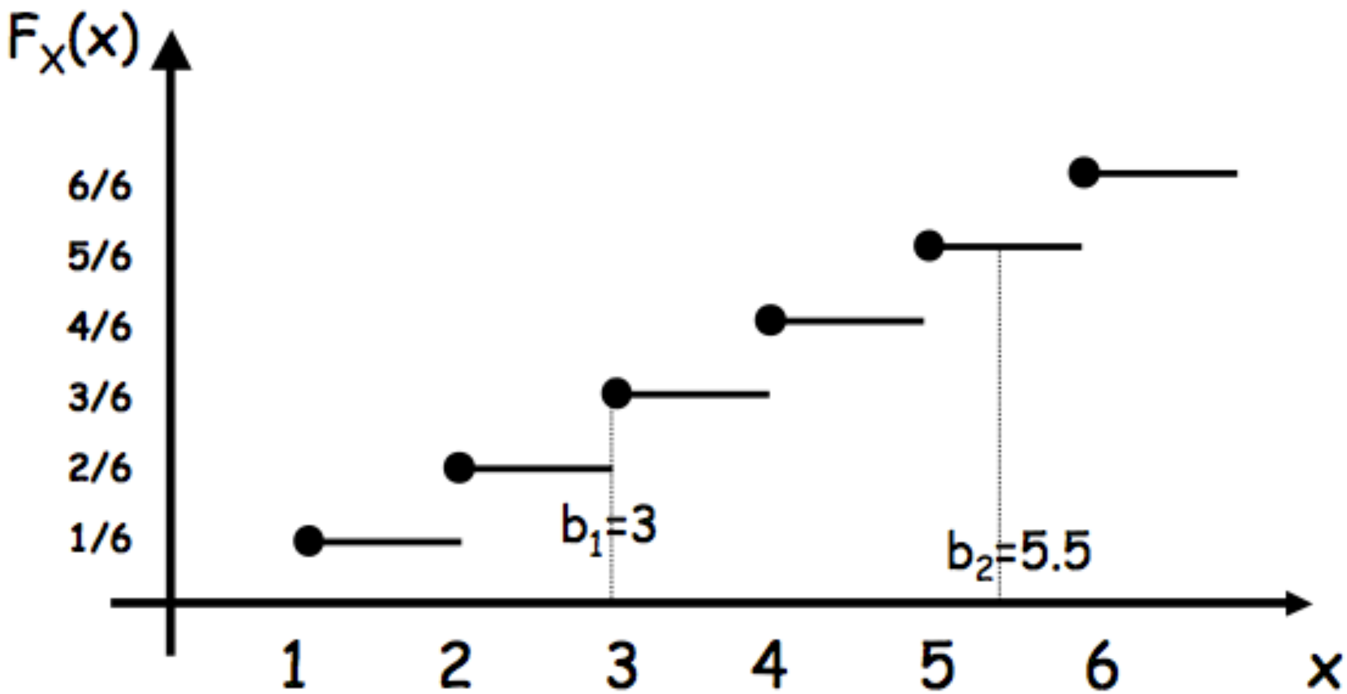


Properties of cdf Đặc trưng của cdf

- $P[X=b] = F_X(b) - F_X(b^-)$

$P[X=b_1=3] = (3/6) - (2/6) = 1/6$

$P[X=b_2=5.5] = (5/6) - (5/6) = 0$

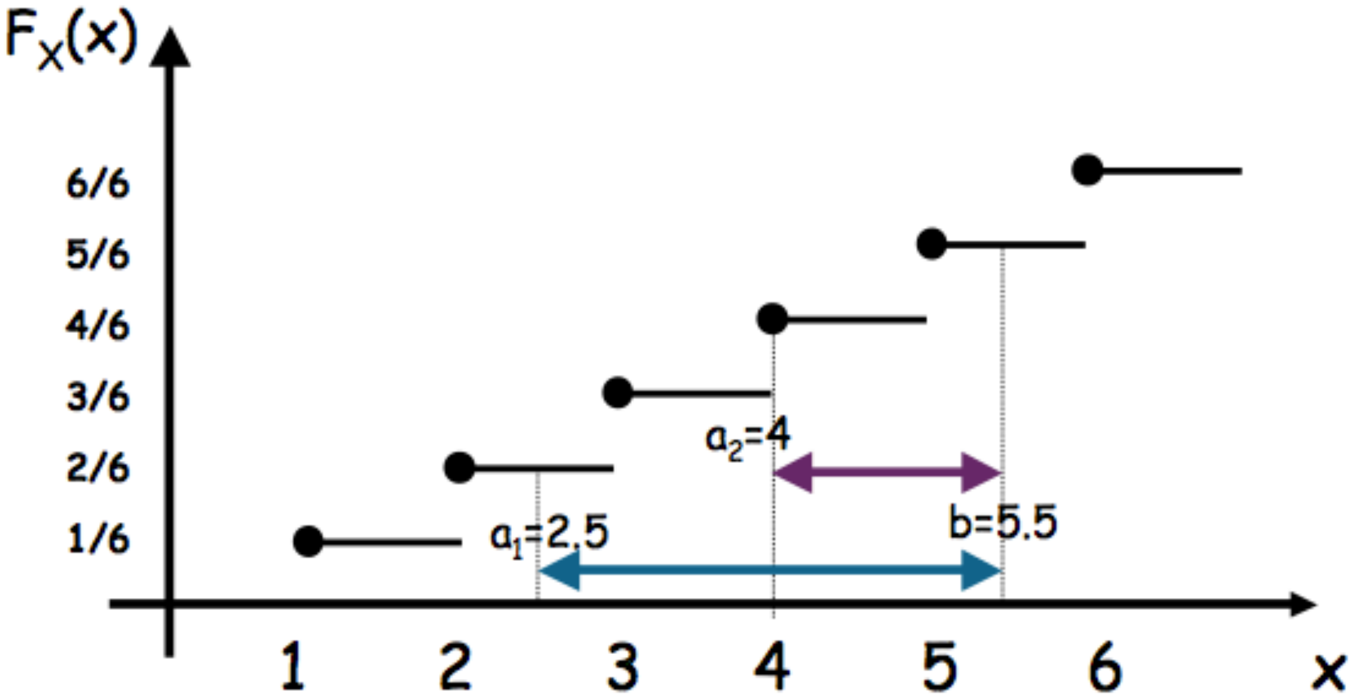


Properties of cdf Đặc trưng của cdf

- $P[a \le X \le b] = F_X(b) - F_X(a^-)$

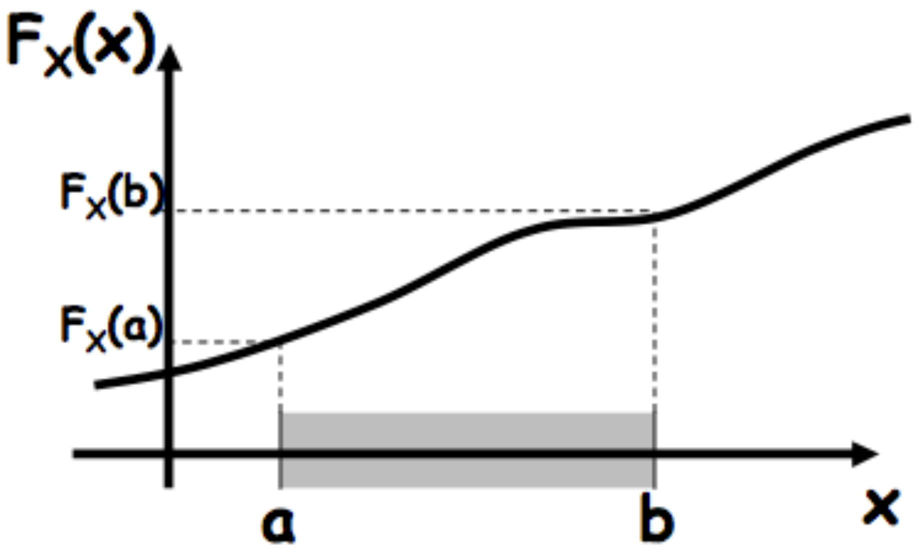
$$P[2.5 = a_1 \le X \le b = 5.5] = (5/6) - (2/6) = 3/6$$

$$P[4 = a_2 \le X \le b = 5.5] = (5/6) - (3/6) = 2/6$$



Properties of cdf Đặc trưng của cdf

- **For a continuous random variable X, the following probabilities are equivalent:**
  Cho một biến ngẫu nghiên liên tục X, các xác suất sau là bằng nhau
  - $P[a \le X \le b]$
  - $P[a < X \le b]$
  - $P[a \le X < b]$
  - $P[a < X < b]$



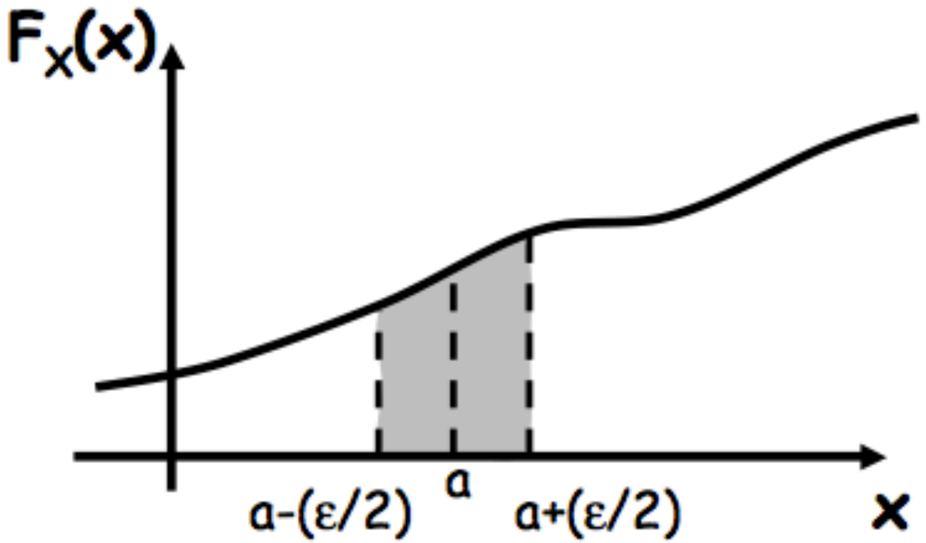
Probability Density Function pdf

Hàm mật độ xác suất pdf

- **We are interested in the probability that a random variable X takes on a value in the "vicinity" (ε neighborhood) of "a"**

Chúng ta quan tâm đến xác suất để một biến ngẫu nhiên X nhận một giá trị nằm trong vùng lân cận (lân cận epsion) của a

$$P[a-(\varepsilon/2) < X \leq a+(\varepsilon/2)]$$



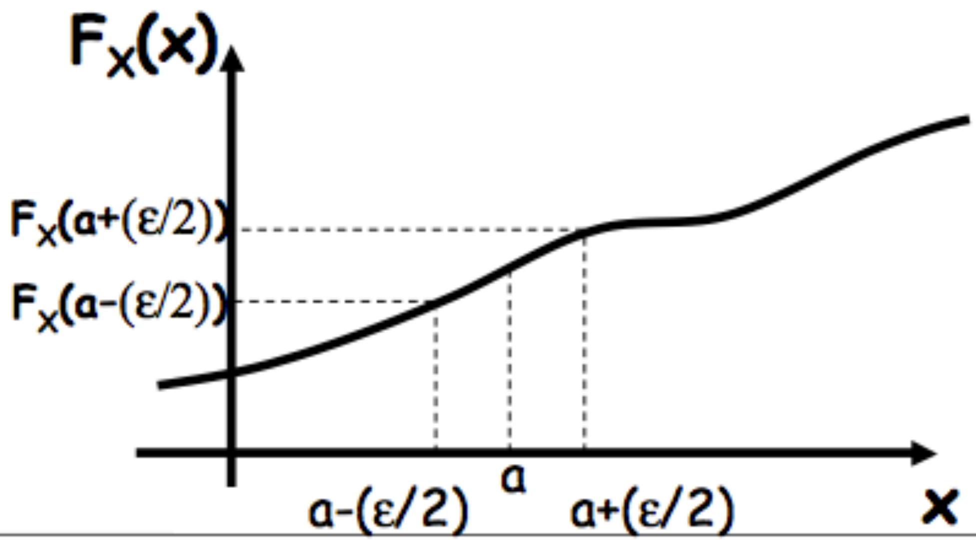
Probability Density Function pdf

Hàm mật độ xác suất pdf

- **Since $P[a < X \le b] = F_X(b) - F_X(a) \Rightarrow$**

  Vì ......

$$P[a-(\varepsilon/2) < X \le a+(\varepsilon/2)] = F_X(a+(\varepsilon/2)) - F_X(a-(\varepsilon/2))$$



Probability Density Function pdf

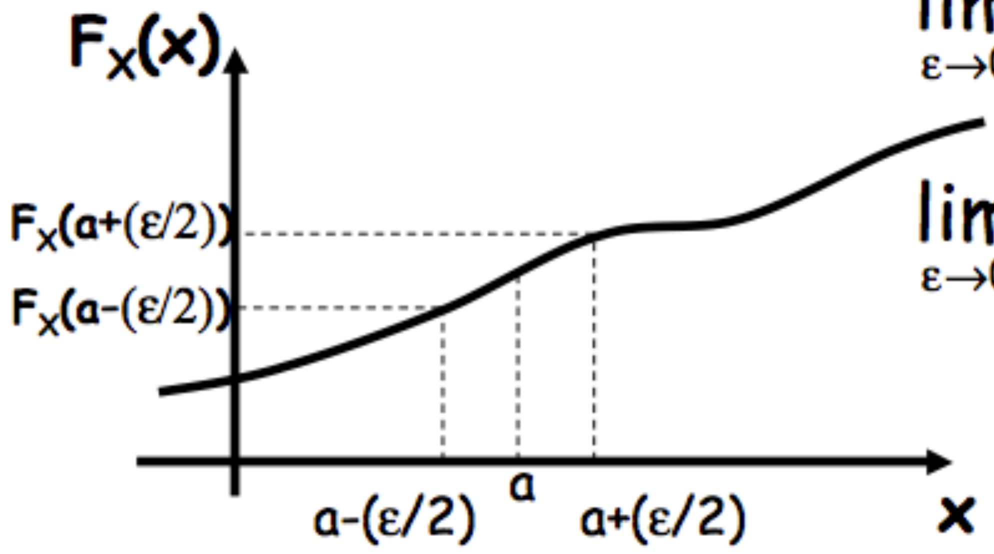
Hàm mật độ xác suất pdf

- **Therefore,**

Do đó,

$$P[a-(\varepsilon/2) < X \le a+(\varepsilon/2)] = F_X(a+(\varepsilon/2)) - F_X(a-(\varepsilon/2))$$

$$\Rightarrow$$



$$\lim_{\varepsilon\to 0}\frac{P[(a-(\varepsilon/2)) < X \le (a+(\varepsilon/2))]}{\varepsilon} =$$

$$\lim_{\varepsilon\to 0}\frac{F_X(a+(\varepsilon/2)) - F_X(a-(\varepsilon/2))}{\varepsilon}$$

$$\left.\frac{dF_X}{dx}\right|_{x=a} = \frac{dF_X(a)}{dx}$$

Probability Density Function pdf

Hàm mật độ xác suất pdf

$$\frac{dF_X(a)}{dx} \equiv f_X(a)$$

$$F_X(a) = \int_{-\infty}^{a} f_X(x)\,dx$$

$$\lim_{a\to\infty} F_X(a) = \lim_{a\to\infty} \int_{-\infty}^{a} f_X(x)\,dx \Rightarrow \boxed{1 = \int_{-\infty}^{+\infty} f_X(x)\,dx}$$

Probability Density Function pdf

Hàm mật độ xác suất pdf

Với một biến ngẫu nhiên liên tục X,

- For a continuous random variable X,

$$P[a \le x \le b] = P[a < x \le b] = F_X(b) - F_X(a)$$

using $\quad F_X(a) = \int_{-\infty}^{a} f_X(x)\,dx$

Sử dụng

$$F_X(b) - F_X(a) = \int_{-\infty}^{b} f_X(x)\,dx - \int_{-\infty}^{a} f_X(x)\,dx$$

$$F_X(b) - F_X(a) = \cancel{\int_{-\infty}^{a} f_X(x)\,dx} + \int_{a}^{b} f_X(x)\,dx - \cancel{\int_{-\infty}^{a} f_X(x)\,dx}$$

Probability Density Function pdf

Hàm mật độ xác suất pdf

- **Probability that a continuous random variable X takes on values between a and b:**

Xác suất để một biến ngẫu nhiên liên tục X nhận giá trị giữa a và b:

$$\boxed{P[a \le X \le b] = \int_a^b f_X(x)\,dx}$$

Probability Mass Function pmf

Hàm khối lượng xác suất pmf

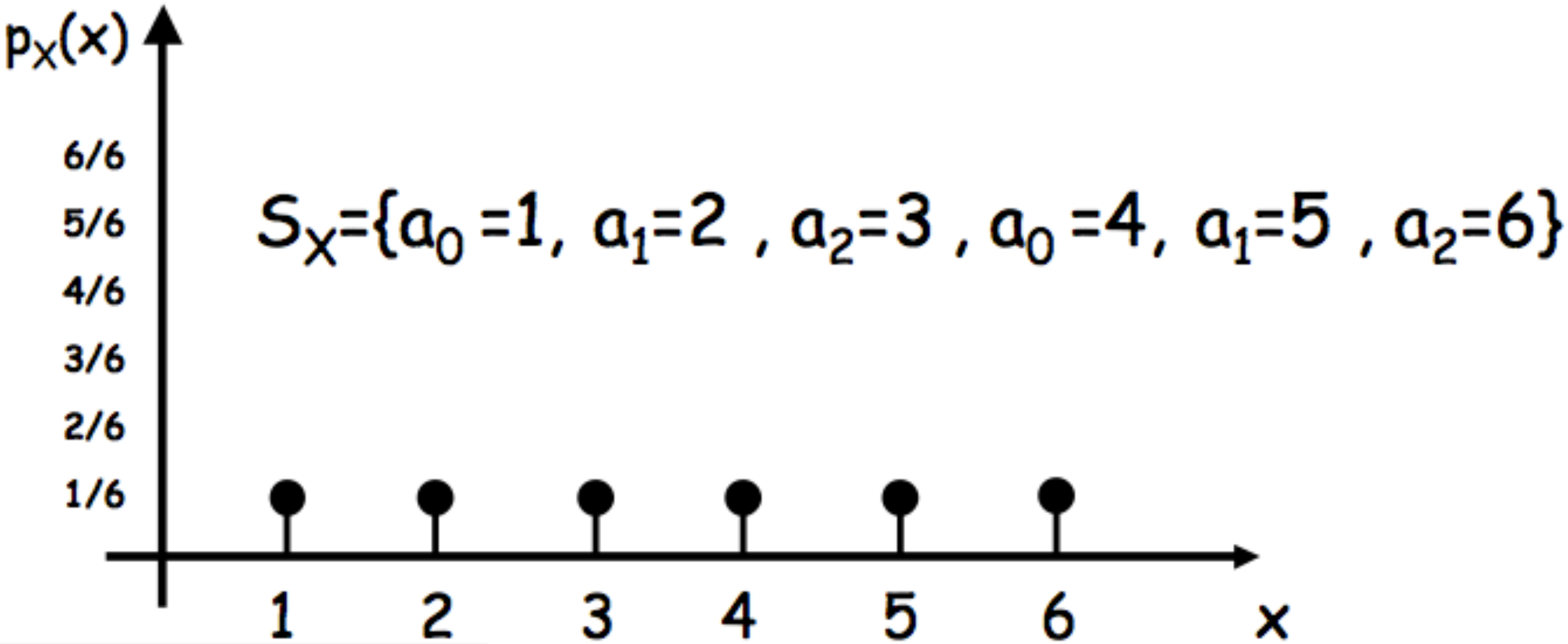
- For a discrete random variable X, we have a countable set $S_X=\{a_0 , a_1 , a_2 , ...\}$

Với một biến ngẫu nhiên rời rạc X, chúng ta có một tập đếm được

- pmf is the set of probabilities:

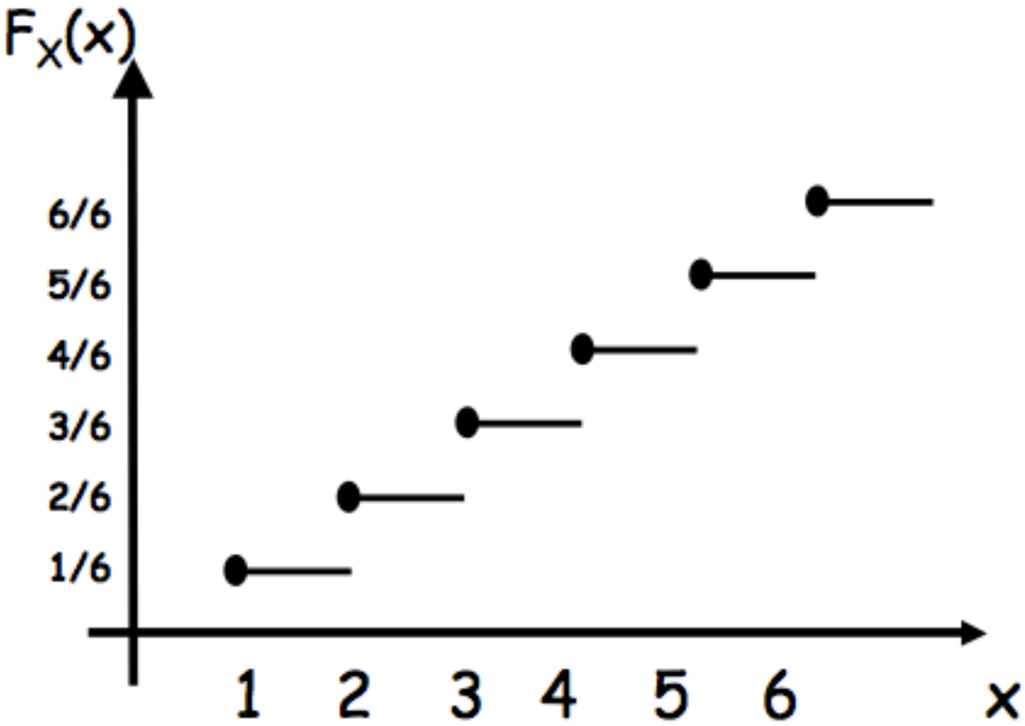
pmf là một tập các giá trị xác suất:

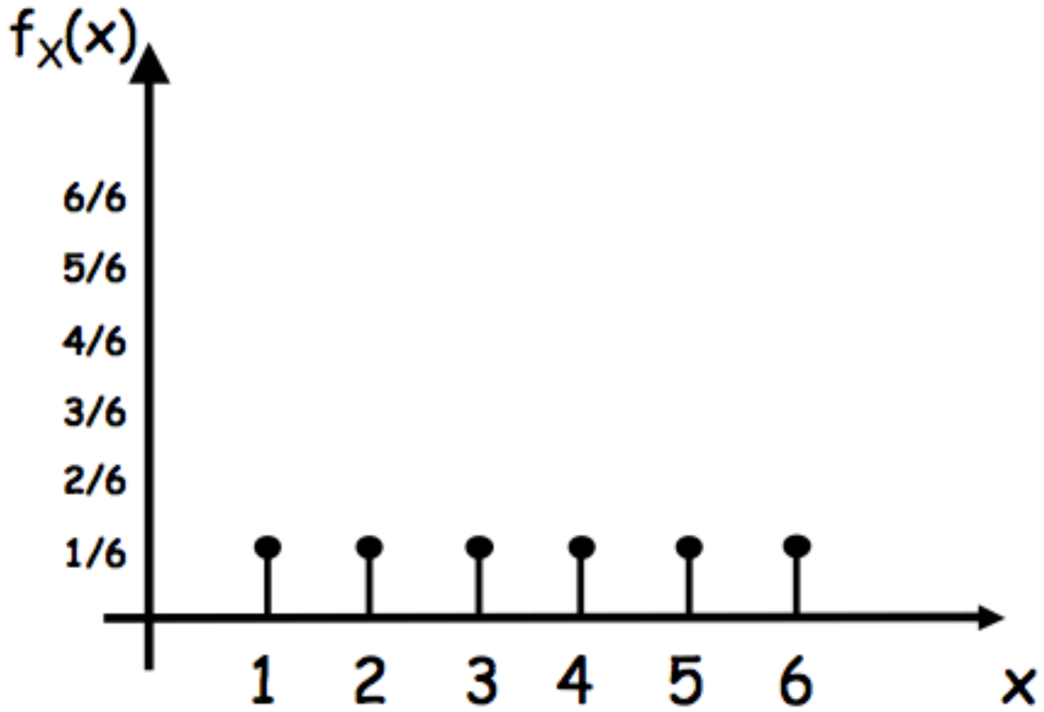
$$P[X=a_k]=p_X(a_k) , k = 0, 1, 2, \ldots$$



pmf, cdf, & pdf for discrete RVs

$$F_X(a) = \sum_k p_X(a_k)\, u(a - a_k) \qquad f_X(a) = \sum_k p_X(a_k)\, \delta(a - a_k)$$





Moments of random variables

Mô men của biến ngẫu nhiên

- There are several parameters, known as moments, that can characterize the behavior of a RV

Có một vài thông số, gọi là mô men), đặc trưng cho một biến ngẫu nhiên

- A general form of moments is:

Công thức tổng quát của mô ment là: $E[(X-a)^n]$

- For continuous RVs,

Với các biến ngẫu nhiên liên tục

$$E[(X-a)^n] = \int_{-\infty}^{+\infty} (x-a)^n f_X(x)\,dx$$

Moments of random variables

- **Special cases of $E[(X-a)^n]$:**

  Các trường hợp đặc biệt của ....

  - The mean: $E[X] = m$, $a = 0$, $n = 1$

    Kỳ vọng:

  - The second moment: $E[X^2]$, $a = 0$, $n = 2$

    Mô men bậc 2

  - The central moments: $E[(X-m)^n]$, $a = m = E[X]$

    Mô men trung tâm

  - The variance: $E[(X-m)^2]$, $a = m = E[X]$, $n = 2$

    Phương sai

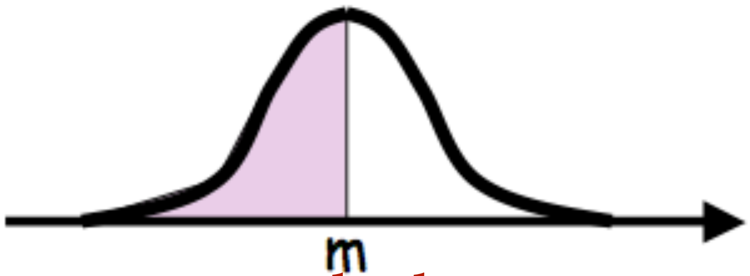
The Mean

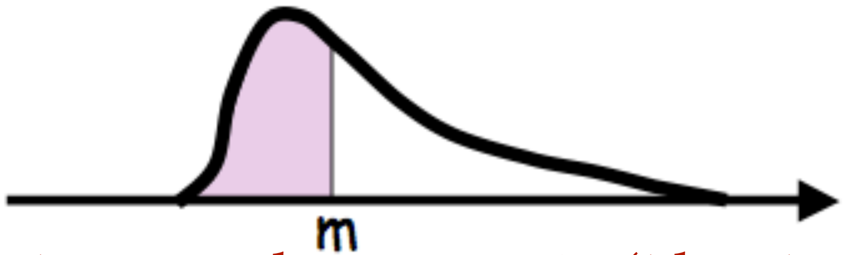
Kỳ vọng

- The mean is the "expected value" or the "average value" of a RV

Kỳ vọng là "giá trị thật" hay "giá trị trung bình" của một biến ngẫu nhiên

$$E[X] = \int_{-\infty}^{+\infty} x\, f_X(x)\,dx$$





Kỳ vọng thể hiện "trung tâm của trọng lực" của hàm mật độ (khối lượng) xác suất

The mean represents the "center of gravity" for the probability (mass) density function

The Second Moment

Mô men bậc 2

- The second moment $E[X^2]$ is the (total) power of the random variable X

Mô men bậc 2 E[X2] là điện năng (tổng) của một biến ngẫu nhiên X

$$E[X^2] = \int_{-\infty}^{+\infty} x^2 f_X(x)dx$$

Mô men bậc 2 thể hiện "mô men của gia tốc so với giá trị gốc" của hàm mật độ (khối lượng) xác suất

The second moment $E[X^2]$ represents the "moment of inertia with respect to the origin" for the probability (mass) density function

The Second Moment

- **The second moment $E[X^2]$ can be expressed as:**

  Mô men bậc 2 E[X2] có thể được biểu diễn như sau:

$$E[X^2] = \int_{-\infty}^{+\infty} x^2 f_X(x)dx = \int_{-\infty}^{+\infty} (x-m+m)^2 f_X(x)dx$$

$$= m^2 \cancelto{1}{\int_{-\infty}^{+\infty} f_X(x)dx} + \int_{-\infty}^{+\infty} (x-m)^2 f_X(x)dx + 2m \cancelto{0}{\int_{-\infty}^{+\infty} (x-m) f_X(x)dx}$$

$$= m^2 + \int_{-\infty}^{+\infty} (x-m)^2 f_X(x)dx$$

The Second Moment

- Therefore, the second moment $E[X^2]$ is the sum of two "power" terms:

Do đó, mô men bậc 2 E[X2] là tổng của hai loại "điện năng"

$$E[X^2] = \underbrace{m^2}_{\text{"DC" Power}} + \underbrace{\int_{-\infty}^{+\infty} (x-m)^2 f_X(x)\,dx}_{\text{"AC" Power}}$$

Điện năng "một chiêu"

Điện năng "xoay chiêu"

The Variance

Phương sai

- **The variance σ^2 is the second central moment:**

  Phương sai ... là mô men trung tâm bậc 2

$$\sigma^2 = E[(X-m)^2] = \int_{-\infty}^{+\infty} (x-m)^2 f_X(x)\,dx$$

Phương sai thể hiện mô men trung tâm của gia tốc

The variance represents the “central moment of inertia”

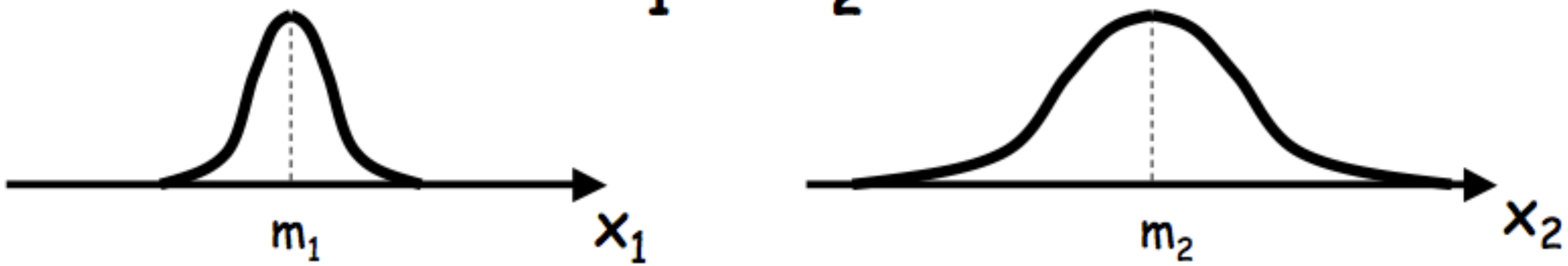
The Variance

Phương sai ... là thước đo sự biến thiên giá trị của một biến ngẫu nhiên xung quanh giá trị kỳ vọng của nó.

- **The variance σ^2 is a measure of "the amount of variation of a RV around its mean"**
- **σ is defined as the "standard deviation"**

.... được gọi là "độ lệch chuẩn"

$$\sigma_1 < \sigma_2$$



The Variance

- **The variance can be computed from the second moment (total power) and the first moment ("DC" power):**

  Phương sai có thể được tính từ mô men bậc 2 (điện năng tổng) và mô men bậc 1 (Điện năng một chiêu)

$$\sigma^2 = E[(X - m)^2] = E[X^2] - m^2$$

Back to Moments $E[X^n]$

Trở lại với mô men E[X^n]

- **Can the mean and the variance completely specify a random variable (e.g. its pdf)?**

  Liệu kỳ vọng và phương sai có thể là mô tả được đây đủ một biến ngẫu nhiên (ví dụ như hàm pdf của nó)

  In general, NO

  Về cơ bản là KHÔNG

- **Knowledge of all moments $E[X^n]$, $n=1, 2, \ldots,$ (if they exist) can be used to uniquely determine the pdf of the RV X**

  Thông tin về tất cả mô men E[X^n], n=1, 2, ... (nếu chúng tồn tại) mới có thể xác định được duy nhất hàm pdf của biến ngẫu nhiên X

Functions of Random Variables

Hàm của các biên ngẫu nhiên

- Functions of random variables are also random variables

  Hàm của [các] biên ngẫu nhiên cũng là biến ngẫu nhiên

- The pdf of the random variable $Y=g(X)$ depends on:

  Hàm pdf của biến ngẫu nhiên Y = g(X) phụ thuộc vào:

  - The function $g(x)$

    Hàm g(x)

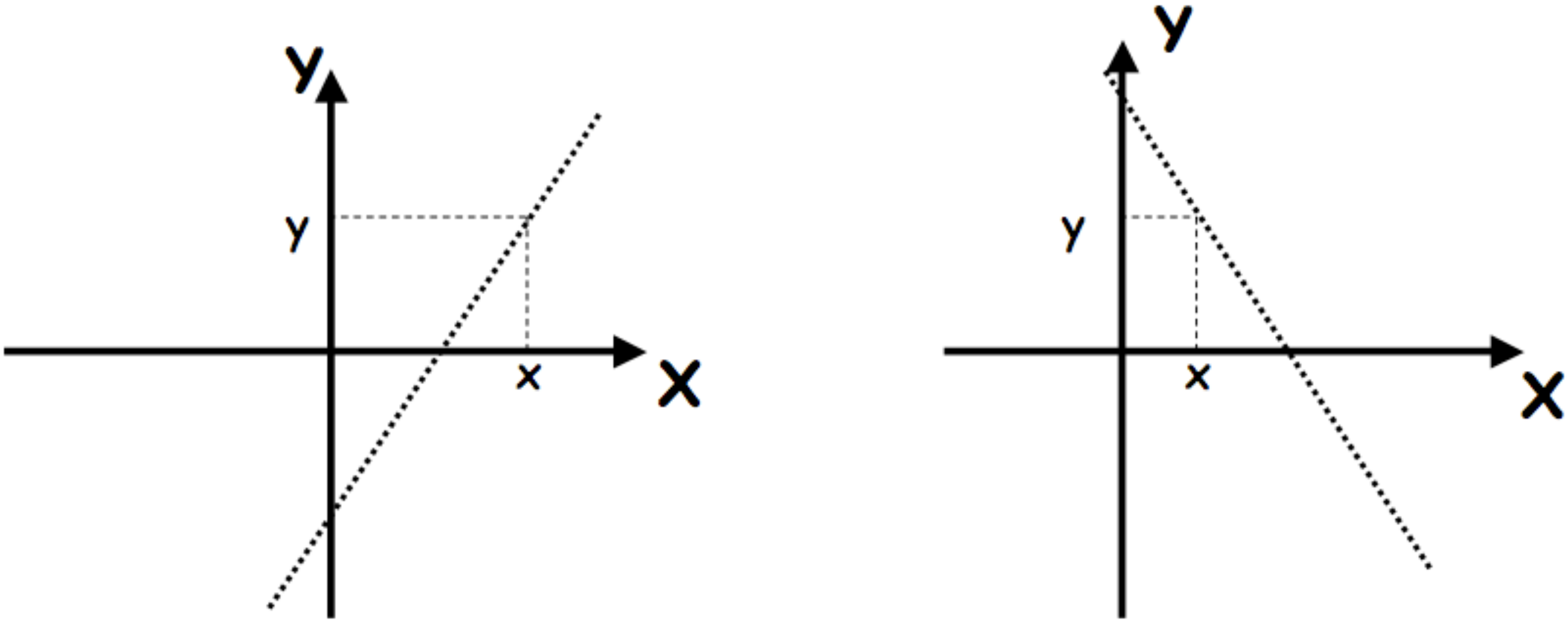
  - The pdf of the random variable X: $f_X(x)$

    Hàm pdf của biến ngẫu nhiên X: fX(x)

Example I.4

Cho Y=aX+b, a và b là các hằng khác 0 và X là một biến ngẫu nhiên với cdf FX(x). Hãy tìm cdf và pdf của Y.

- Let $Y=aX+b$, where a and b are some non-zero constants and X is a RV with cdf $F_X(x)$. Find the cdf and pdf for X.



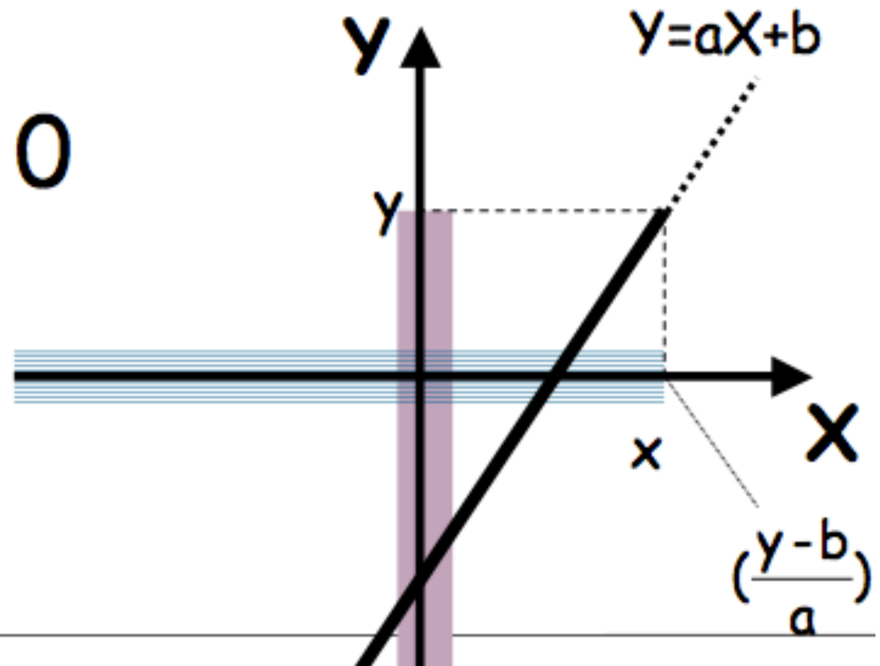
Example I.4

- $F_Y(y) = P[Y \le y]$

$$P[Y \le y] = P[(aX + b) \le y]$$

$$P[Y \le y] = P[X \le (\frac{y-b}{a})] \quad \text{if } a > 0$$

$$P[Y \le y] = F_X(\frac{y-b}{a}) \quad \text{if } a > 0$$



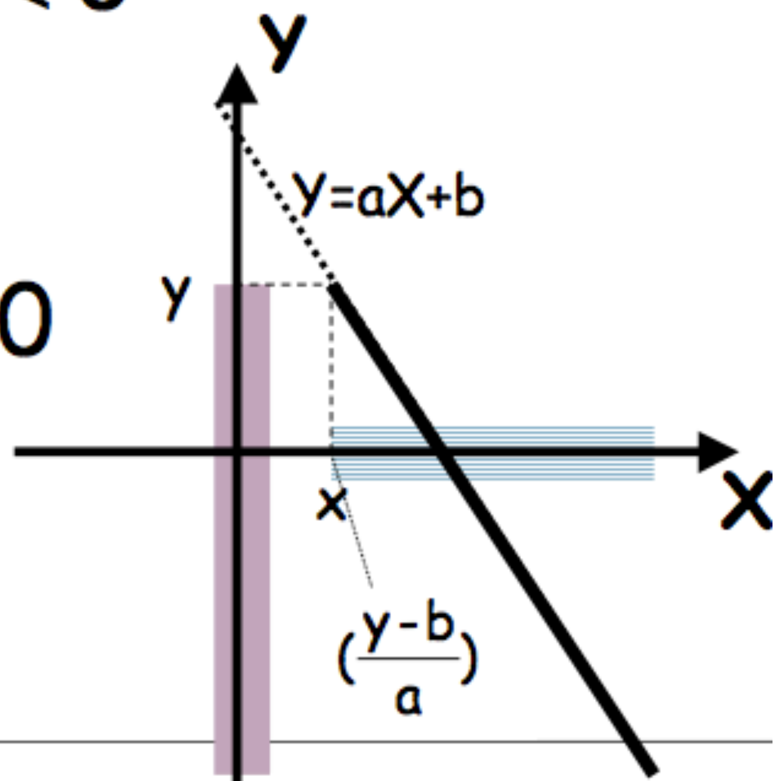
Example I.4

- $F_Y(y) = P[Y \leq y]$

$$P[Y \leq y] = P[(aX + b) \leq y]$$

$$P[Y \leq y] = P[X \geq (\frac{y-b}{a})] \quad \text{if } a < 0$$

$$P[Y \leq y] = 1 - F_X(\frac{y-b}{a}) \quad \text{if } a < 0$$



Example I.4

- **Therefore, for Y=aX+b**

  Do đó, với Y=aX+b

$$P[Y \le y] = F_Y(y) = F_X\left(\frac{y-b}{a}\right) \quad \text{if } a > 0$$

$$P[Y \le y] = F_Y(y) = 1 - F_X\left(\frac{y-b}{a}\right) \quad \text{if } a < 0$$

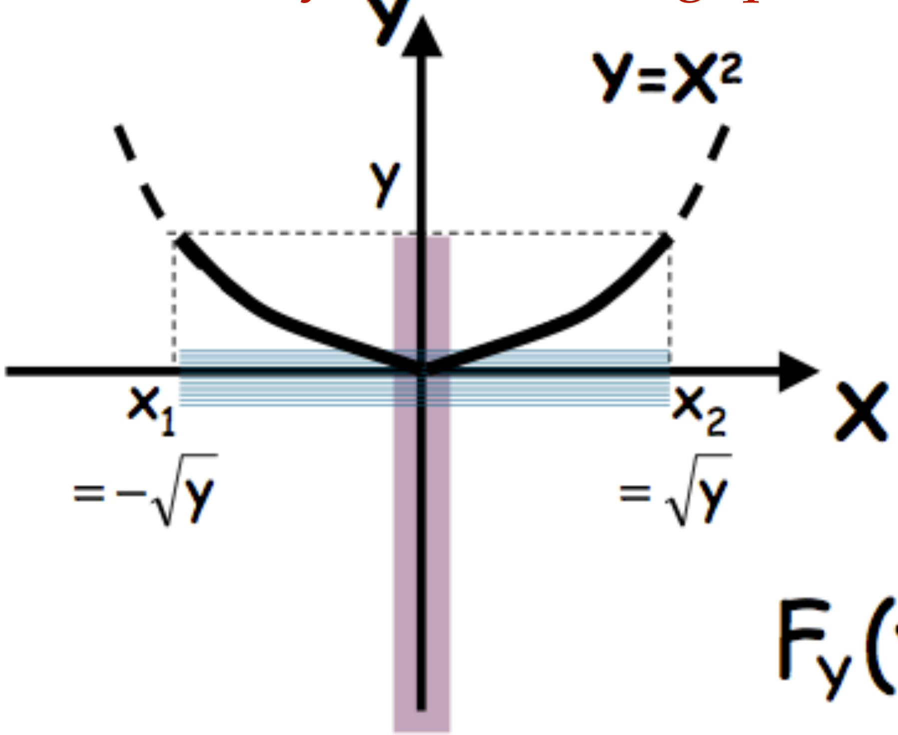
- **Derive $f_Y(y)$ is terms of f_X**

  Tìm fY(y) theo fX

Example I.5

- **Let $Y=X^2$, where X is a RV with cdf F_X. Derive F_Y in terms of F_X .**

Cho Y=X^2, trong đó X là một biến ngẫu nhiên với cdf FX. Hãy tìm FY thông qua FX.



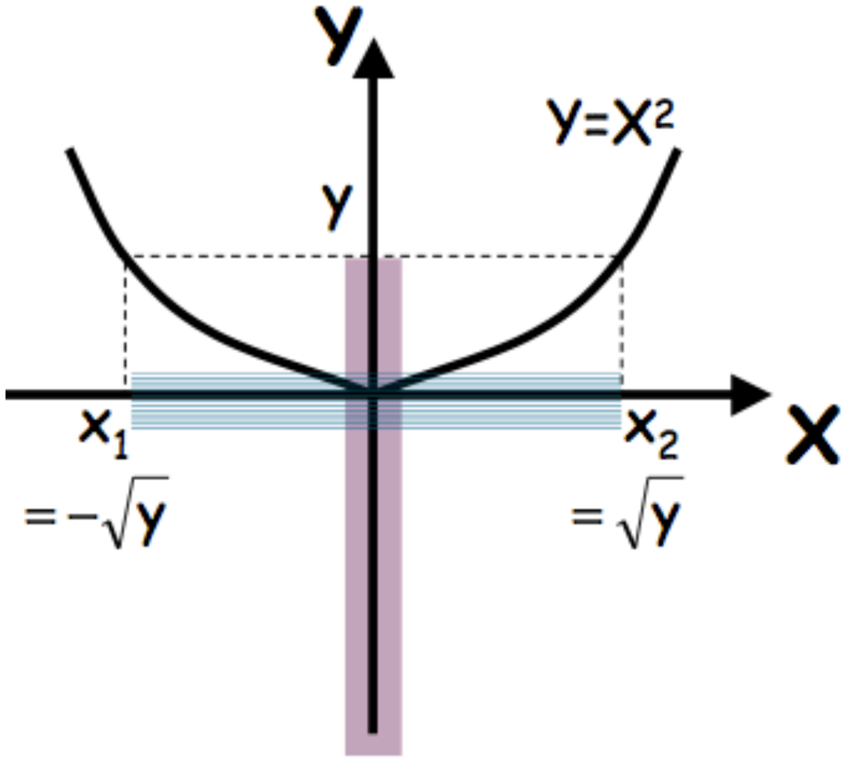
$$F_Y(y) = P[Y \le y]$$

$$Y \le y \quad \Leftrightarrow \quad x_1 \le X \le x_2$$

$$F_Y(y) = P[Y \le y] = P[x_1 \le X \le x_2]$$

Example I.5

$$F_Y(y) = P[Y \le y] = P[x_1 \le X \le x_2]$$



Remember, in general:

Nhớ lại rằng, một cách tổng quát ta có

$$P[x_1 \le X \le x_2] =$$

$$F_X(x_2) - F_X(x_1^-)$$

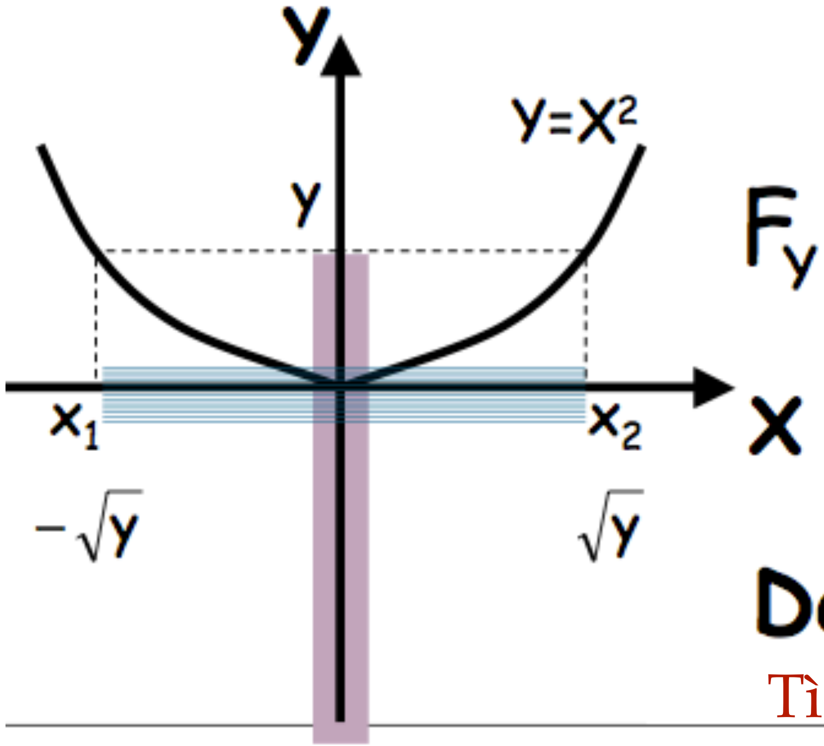
Example I.5

Với một biến ngẫu nhiên có hàm cdf liên tục:

For a RV with a continuous cdf:

$$P[x_1 \le X \le x_2] = F_X(x_2) - F_X(x_1^-) = F_X(x_2) - F_X(x_1)$$

$$F_Y(y) = F_X(x_2) - F_X(x_1)$$



$$F_Y(y) = F_X(\sqrt{y}) - F_X(-\sqrt{y}) \quad \forall\, y \ge 0$$

Derive $f_Y(y)$ is terms of f_X

Tìm fY(y) thông qua fX

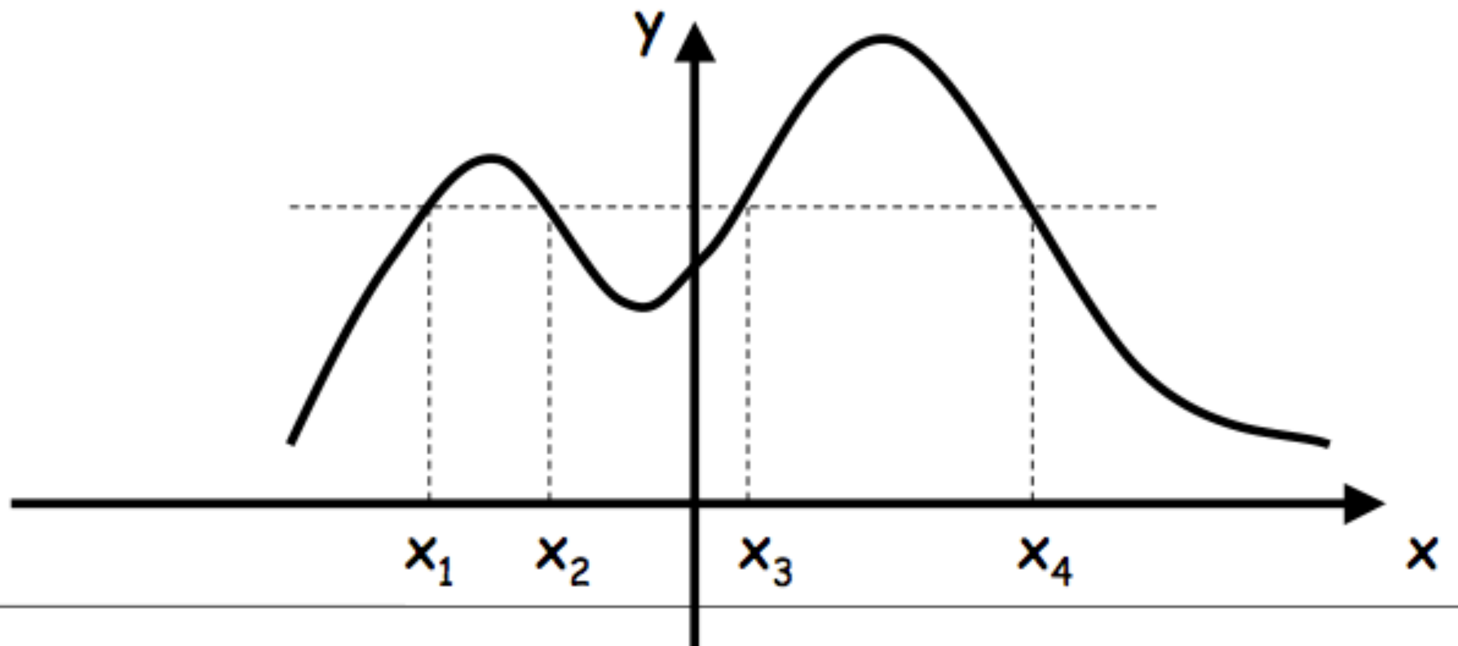
Fundamental Theorem

Định lý cơ sở

- If x_1 , x_2 , x_3 , …. x_n , …. are solutions to the equation y=g(x), then the pdf of the RV Y can be expressed as:

Nếu x1, x2, x3, ..., xn là các nghiệm của phương trình Y=g(X), khi đó hàm pdf của biến ngẫu nhiên Y có thể được biểu diễn như sau:

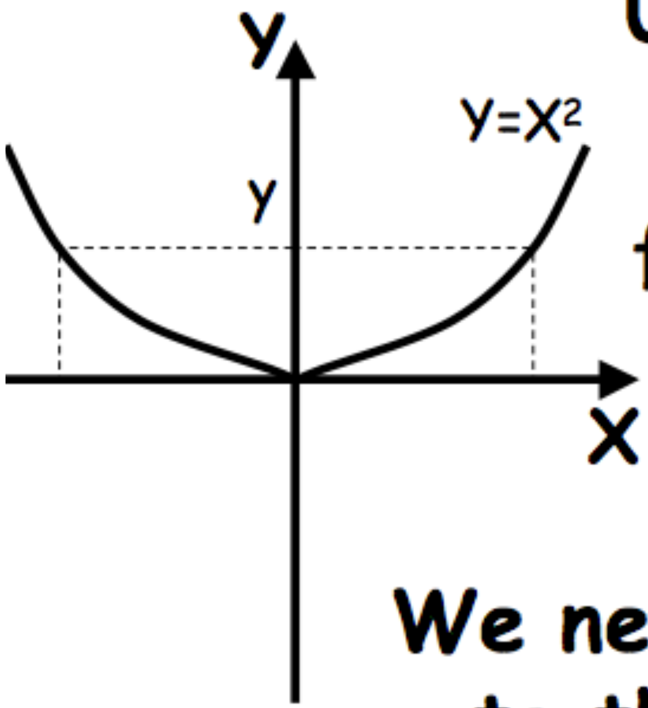
$$f_Y(y) = \frac{f_X(x_1)}{|dg(x_1)/dx|} + \frac{f_X(x_2)}{|dg(x_2)/dx|} + \cdots$$



Example I.6

- Let $Y=X^2$, where X is a RV with pdf f_X. Derive f_Y in terms of f_X.

Cho Y=X^2, X là một biến NN với pdf là fX. Tìm fY thông qua fX



Using the fundamental theorem,

Sử dụng định lý cơ sở

$$f_Y(y) = \frac{f_X(x_1)}{|dg(x_1)/dx|} + \frac{f_X(x_2)}{|dg(x_2)/dx|} + \cdots$$

We need to find the solutions x_1, x_2, x_3 ,... to the equation $y=g(x)=x^2$

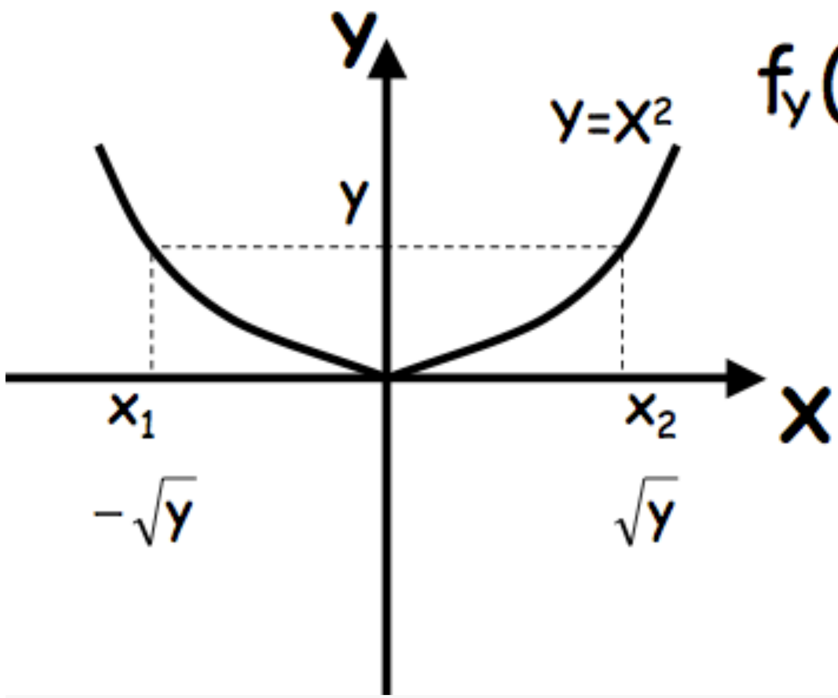
Chúng ta cần tìm các nghiệm x1, x2, x3, ... của phương trình Y=g(X)=X^2

Example I.6

The equation $y=g(x)=x^2$, has two solutions:

Phương trình Y=g(X)=X^2, có hai nghiệm

$$x_1 = -\sqrt{y} \quad \text{and} \quad x_2 = \sqrt{y} \quad \forall\, y \geq 0$$



$$f_Y(y) = \frac{f_X(x_1)}{|dg(x_1)/dx|} + \frac{f_X(x_2)}{|dg(x_2)/dx|} + \cdots$$

$$f_Y(y) = \frac{f_X(-\sqrt{y})}{2\sqrt{y}} + \frac{f_X(\sqrt{y})}{2\sqrt{y}}$$

Expected Value of Functions of RVs

Kỳ vọng của hàm của biến ngẫu nhiên

- Expected value of $g(x)$: $E[g(x)]$

  Giá trị kỳ vọng của g(X): E[g(X)]

- For a continuous random variable:

  Với một biến ngẫu nhiên liên tục:

$$E[g(x)] = \int_{-\infty}^{+\infty} g(x)\, f_X(x)dx$$

$g(x)=ax+b$; $g(x)=ax^2+bx+c$; $g(x)=e^x$; $g(x)=\cos(x)$

$g(x)=x \Rightarrow E[g(x)]=m$ is the mean of X

m là kỳ giá trị trung bình của X

$g(x)=(x-m)^2 \Rightarrow E[g(x)]=\sigma^2$ is the variance of X

m là kỳ phương sai của X