

Lý Thuyết Xác Suất (3): Phân Phối Xác Suất Đa Chiều

Khái quát

- Biến ngẫu nhiên và hàm phân phối xác suất đa chiều.
- Hiệp phương sai và đồng tương quan.

Phân Phối Xác Suất Đa Chiều

Biến Ngẫu Nhiên Đa Chiều (1)

- Trong thực tế, chúng ta sẽ gặp nhiều biến ngẫu nhiên có quan hệ với nhau. Các biến này thường được gọi là biến ngẫu nhiên đa chiều hay vector ngẫu nhiên.
- Ví dụ:
 - Thu thập số liệu của một cá nhân, các thông tin như chiều cao, cân nặng, học vấn, thu nhập.
 - Thu thập dữ liệu của một công ty, nhiều chiều dữ liệu có thể dùng để tìm hiểu: số lượng nhân viên, doanh thu, chi phí, chi phí R&D...

Biến Ngẫu Nhiên Đa Chiều (2)

- Ký hiệu biến ngẫu nhiên n chiều:

$$\mathbf{X} \equiv \{X_1, X_2, X_3 \dots X_n\}$$

$\mathbf{X}$ còn được gọi là vector n biến ngẫu nhiên.

- Với trường hợp 2 biến, ta có thể sử dụng kí hiệu X và Y . Ví dụ: Chọn ngẫu nhiên 1 người và hỏi họ về giới tính và thu nhập hằng tháng (triệu VND). Chúng ta có cặp biến ngẫu nhiên (X, Y) với X rời rạc, Y liên tục như sau:

$X \in (0; 1)$: giới tính (nam, nữ)

$Y \in [0, +\infty)$: thu nhập tính bằng triệu VNĐ

Biến Ngẫu Nhiên Hai Chiều (1)

- Cho một cặp biến ngẫu nhiên (X, Y) , chúng ta có thể biểu diễn hàm phân phối của chúng dưới dạng $P(X = x; Y = y) \equiv f(x; y)$:

X/Y	$Y = 1$	$Y = 2$	$Y = 3$
$X = 4$	0.18	0.22	0.16
$X = 6$	0.08	0.16	0.20

- Ví dụ: $P(X = 4; Y = 1) = f(4, 1) = 0.18$.
- Chúng ta có thể tính phân phối xác suất biên – xác suất không cần bất cứ thông tin gì của biến còn lại – của từng biến.

Biến Ngẫu Nhiên Hai Chiều (2)

- Cộng gộp từng cột và từng dòng:

$X \backslash Y$	$Y = 1$	$Y = 2$	$Y = 3$	$P(X)$
$X = 4$	0.18	0.22	0.16	0.54
$X = 6$	0.08	0.16	0.20	0.46
$P(Y)$	0.26	0.38	0.36	

- Hàm phân phối biên (the marginal probability function) cho từng biến:

Y	$Y = 1$	$Y = 2$	$Y = 3$
$P(Y = y) = f_Y(y)$	0.26	0.38	0.36

X	$X = 1$	$X = 2$
$P(X = x) = f_X(x)$	0.54	0.46

Biến Ngẫu Nhiên Hai Chiều (3)

- Xác suất có điều kiện $P(X = x|Y = y_j)$ – the conditional probability function – là xác suất xảy ra khi chúng ta **biết trước** rằng $Y = y_j$ đã xảy ra.
- Sử dụng định lý Bayes, chúng ta có:

$$P(X = x_i|Y = y_j) = \frac{P(X = x_i; Y = y_j)}{P(Y = y_j)}$$

$$P(Y = y_j|X = x_i) = \frac{P(X = x_i; Y = y_j)}{P(X = x_j)}$$

- Chú ý: chúng ta có thể kí hiệu $f_{X|Y}(x|y)$ hoặc $f_{Y|X}(y|x)$

Biến Ngẫu Nhiên Hai Chiều (4)

- Từ bảng:

$X \backslash Y$	$Y = 1$	$Y = 2$	$Y = 3$	$P(X)$
$X = 4$	0.18	0.22	0.16	0.54
$X = 6$	0.08	0.16	0.20	0.46
$P(Y)$	0.26	0.38	0.36	

- Chúng ta có thể tính

	$X = 4$	$X = 6$
$P(X Y = 1)$	0.69	0.31
$P(X Y = 2)$	0.58	0.42
$P(X Y = 3)$	0.44	0.56

$$P(X = 4|Y = 1) = \frac{P(X = 4; Y = 1)}{P(Y = 1)} = \frac{0.18}{0.26} = \frac{9}{13}$$

$$P(X = 6|Y = 1) = \frac{P(X = 6; Y = 1)}{P(Y = 1)} = \frac{0.08}{0.26} = \frac{4}{13}$$

Tổng Quát Hàm Xác Suất Biến Ngẫu Nhiên Đa Chiều

- Đối với hàm liên tục, chúng ta có viết **hàm phân phối tích lũy đồng thời (the joint c.d.f)** và **hàm phân phối xác suất đồng thời (the joint p.d.f.)** cho một vector biến ngẫu nhiên $X \equiv \{X_1, X_2, X_3 \dots X_n\}$:

$$F(x_1, x_2, x_3 \dots x_n);$$

$$f(x_1, x_2, x_3, \dots x_n)$$

- Các hàm phân phối biên và phân phối có điều kiện có thể được biểu diễn tương tự.

Các Đặc Trưng Cho Phân Phối Đa Biến

Giá Trị Kỳ Vọng Có Điều Kiện (1)

- Cho cặp biến X và Y , chúng ta có thể tính giá trị kỳ vọng biên và có điều kiện cho từng biến.
- Giá trị kỳ vọng có điều kiện (trường hợp rời rạc):

$$E(X|Y = y_j) = \sum_{i=1} x_i \frac{P(X = x_i; Y = y_j)}{P(Y = y_j)}$$

$$E(Y|X = x_i) = \sum_{j=1} y_j \frac{P(X = x_i; Y = y_j)}{P(X = x_i)}$$

- Giá trị kỳ vọng biên:

$$E(X) = \sum_{i=1} x_i P_X(X = x_i) = \sum_{i=1} x_i f_X(x_i)$$

$$E(Y) = \sum_{i=1} y_i P_Y(Y = y_i) = \sum_{i=1} y_i f_Y(y_i)$$

Giá Trị Kỳ Vọng Có Điều Kiện (2)

- Từ bảng:

$X \backslash Y$	$Y = 1$	$Y = 2$	$Y = 3$	$P(X)$
$X = 4$	0.18	0.22	0.16	0.54
$X = 6$	0.08	0.16	0.20	0.46
$P(Y)$	0.26	0.38	0.36	

	$X = 4$	$X = 6$
$P(X Y = 1)$	0.69	0.31
$P(X Y = 2)$	0.58	0.42
$P(X Y = 3)$	0.44	0.56

- Giá trị kỳ vọng có điều kiện $E(X|Y = y_i)$ là một hàm số phụ thuộc vào Y :

$$E(X|Y = 1) = 0.69 * 4 + 0.31 * 6 = 4.26$$

$$E(X|Y = 2) = 0.58 * 4 + 0.42 * 6 = 4.84$$

$$E(X|Y = 3) = 0.44 * 4 + 0.56 * 6 = 5.12$$

- Giá trị kỳ vọng biên cho X : $E(X) = 0.54 * 4 + 0.46 * 6 = 4.92$

Hiệp Phương Sai Và Đồng Tương Quan (1)

- Hiệp phương sai (covariance) của hai biến X, Y được tính như sau:

$$\begin{aligned}\text{cov}(X, Y) &= E([X - E(X)][Y - E(Y)]) \\ &= E(XY) - E(X)E(Y)\end{aligned}$$

$$= \sum_{\text{all}} \sum_{x,y} (x - E(X))(y - E(Y)) f(x; y)$$

- Hệ số tương quan (correlation) của hai biến được tính như sau:

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \cdot \sigma_Y}$$

Hiệp Phương Sai Và Đồng Tương Quan (2)

- $\text{cov}(X, Y) = \sum \sum (x - E(X)) (y - E(Y)) f(x; y)$

$X \backslash Y$	$Y = 1$	$Y = 2$	$Y = 3$	$P(X)$
$X = 4$	0.18	0.22	0.16	0.54
$X = 6$	0.08	0.16	0.20	0.46
$P(Y)$	0.26	0.38	0.36	

$f(x; y)$		y			$f_X(x)$
		1	2	3	
x	4	0.18	0.22	0.16	0.54
	6	0.08	0.16	0.20	0.46
$f_Y(y)$		0.26	0.38	0.36	

- $\text{cov}(X, Y) = 0.18 * (4 - 4.92) * (1 - 2.1) + 0.22 * (4 - 4.92) * (2 - 2.1) + 0.16 * (4 - 4.92) * (3 - 2.1) + 0.08 * (6 - 4.92) * (1 - 2.1) + 0.16 * (6 - 4.92) * (1 - 2.1) + 0.20 * (6 - 4.92) * (1 - 2.1) = -0.45$
- $\rho_{XY} = ?$

Tính Chất

- Giá trị của hệ số tương quan giới hạn trong khoảng -1 và 1 :

$$-1 \leq \rho_{XY} \leq 1$$

- Hai biến X, Y độc lập:

$$\text{cov}(X, Y) = 0$$

$$\rho_{XY} = 0$$

- Điều ngược lại không đúng.

Biến Ngẫu Nhiên Độc Lập (1)

- Hai biến ngẫu nhiên X và Y được gọi là độc lập (independent random variables) khi và chỉ khi với mọi giá trị x, y :

$$f(x; y) = f_X(x) f_Y(y)$$

- Khái niệm độc lập có thể được khái quát cho ba hay n biến ngẫu nhiên với điều kiện sau:

$$f(x_1, x_2, x_3, \dots, x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \dots f_{X_n}(x_n)$$

Biến Ngẫu Nhiên Độc Lập (2)

Khi hai biến X và Y độc lập, chúng ta có các tính chất sau đây:

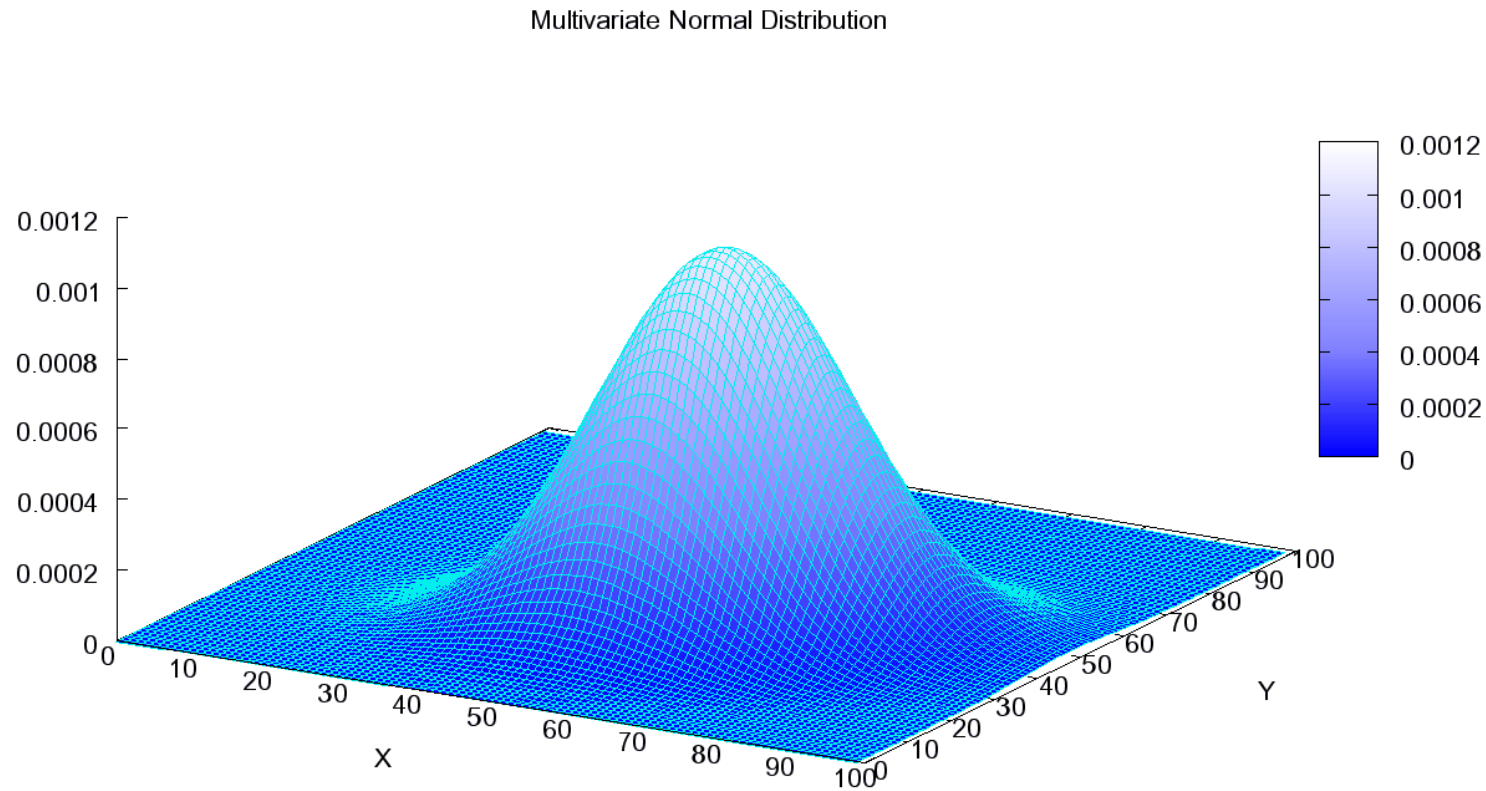
- Phân phối có điều kiện không phụ thuộc vào biến còn lại:

$$f_{X|Y}(x|Y = y_j) = f_X(x)$$

$$f_{Y|X}(y|X = x_i) = f_Y(y)$$

- Kỳ vọng có điều kiện $E(X|Y = y_j)$, $E(Y|X = x_i)$ không phụ thuộc vào giá trị của biến còn lại.

Biểu Diễn Hai Biến Phân Phối Chuẩn (Bivariate Normal Distribution)

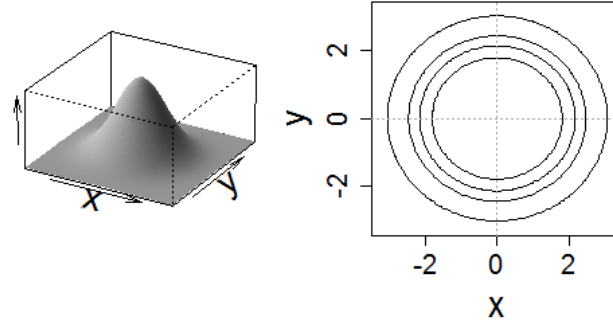


Source: https://en.wikipedia.org/wiki/Multivariate_normal_distribution

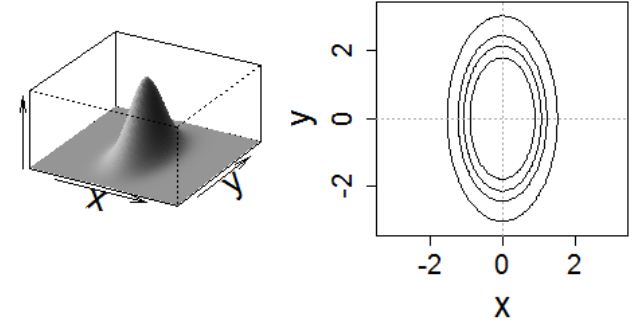
Minh họa đường đồng mức (contours) cho hàm phân phối xác suất đồng thời

- Đường đồng mức là cố định từng mức xác suất (cắt lát dọc hình ba chiều), và chọn tất cả giá trị (x, y) thành từng hình tròn hoặc elíp. Sau đó, chiếu từng hình đó xuống mặt phẳng hai chiều. Nhiều đường đồng mức sẽ hình thành các biểu đồ như bên phải.
- Từng hàm phân phối xác suất đồng thời khác nhau sẽ có các dạng đường đồng mức khác nhau.

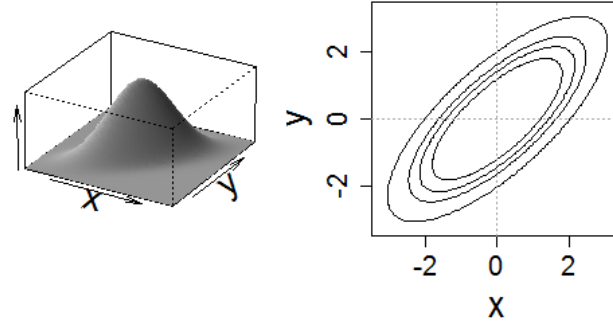
$$\sigma_x = \sigma_y, \rho = 0$$



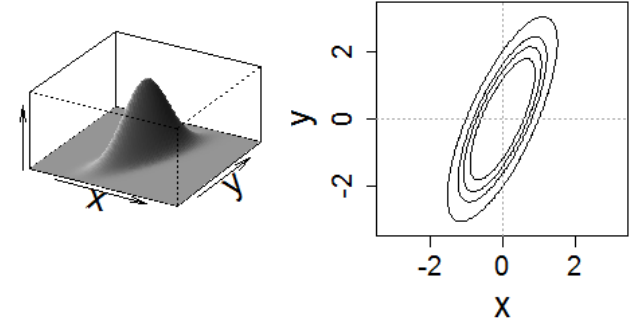
$$2\sigma_x = \sigma_y, \rho = 0$$



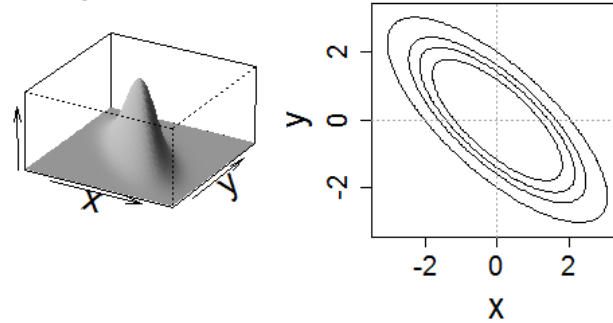
$$\sigma_x = \sigma_y, \rho = 0.75$$



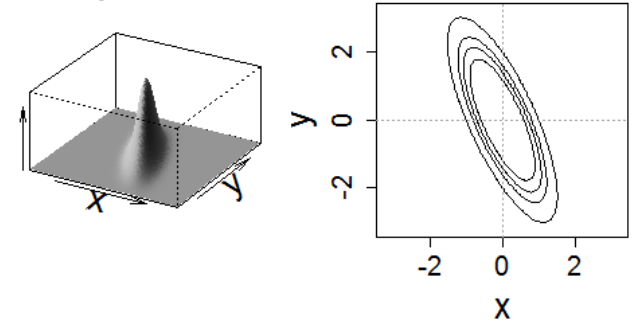
$$2\sigma_x = \sigma_y, \rho = 0.75$$



$$\sigma_x = \sigma_y, \rho = -0.75$$



$$2\sigma_x = \sigma_y, \rho = -0.75$$



Tổng Kết

Hàm phân phối xác suất đồng thời	Hàm phân phối xác suất biên	Hàm phân phối xác suất có điều kiện
$P(X = x; Y = y)$ $f(x; y)$	$P(X = x); f_X(x)$ $P(Y = y); f_Y(y)$	$P(X = x Y = y); f_{X Y}(x y)$ $P(Y = y X = x); f_{Y X}(x y)$
- Hàm tổng quát mô tả xác suất chung cho cặp biến X và Y. - Hàm này phải có giá trị cho cả X và Y.	- Hàm phân phối riêng cho X mà không quan tâm đến giá trị của Y. - Hàm này không phụ thuộc vào giá trị của biến còn lại. - Kì vọng (trung bình) biên – $E(X)$ hoặc $E(Y)$ – là một giá trị.	- Hàm phân phối của X cho biết trước các giá trị Y khác nhau đã xảy ra. - Hàm phân phối này phụ thuộc vào giá trị của cả biến, trừ khi hai biến này là hai biến độc lập. - Kì vọng (trung bình) có điều kiện – $E(X Y = y)$ hoặc $E(Y X = x)$ – là một hàm của biến còn lại (trừ trường hợp 2 biến độc lập hoặc liên hệ phi tuyến tính).

Lý Thuyết Xác Suất: Tổng Kết

- Các khái niệm cơ bản về xác suất: phép thử (experiment), biến cố (events), xác suất có điều kiện và định lý Bayes.
- Biến ngẫu nhiên, hàm xác suất (p.d.f. và c.d.f.): $P(X = x)$, $f(x)$, $F(x)$.
- Phân phối (rời rạc) binomial, phân phối (liên tục) chuẩn.
- Giá trị kì vọng, phương sai, độ lệch chuẩn, trung vị,...
- Biến ngẫu nhiên đa chiều và phân phối của nó (the joint p.d.f). Hàm phân phối biên và có điều kiện (the marginal and conditional distribution functions).
- Biến ngẫu nhiên độc lập.
- Hiệp phương sai (covariance) và hệ số tương quan (correlation).