

**BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC MỎ-ĐỊA CHẤT**

BÁO CÁO HỌC THUẬT

**PHÂN PHỐI CHUẨN VÀ CÁC MOMENT CỦA
PHÂN PHỐI CHUẨN**

ThS Nguyễn Thu Hằng

Hà nội, tháng 12 năm 2020

**BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC MỎ-ĐỊA CHẤT**

BÁO CÁO HỌC THUẬT

**PHÂN PHỐI CHUẨN VÀ CÁC MOMENT CỦA
PHÂN PHỐI CHUẨN**

Xác nhận của bộ môn

Hà nội, tháng 12 năm 2020

MỤC LỤC

Lời giới thiệu	1
1. Lịch sử hình thành phân phối chuẩn	2
2. Định nghĩa và tính chất của phân phối chuẩn	3
3. Moment của phân phối chuẩn	7

LỜI GIỚI THIỆU

Phân phối chuẩn, còn gọi là phân phối Gauss hay (Hình chuông Gauss), là một phân phối xác suất cực kì quan trọng trong nhiều lĩnh vực. Nó là họ phân phối có dạng tổng quát giống nhau, chỉ khác giá trị kỳ vọng μ (Muy) còn được hiểu là giá trị trung bình, và độ lệch tiêu chuẩn σ (Sigma). Thế giới tự nhiên, cũng như nhiều các quy luật kinh tế xã hội tuân theo luật phân phối chuẩn này, điển hình như: Chỉ số thông minh IQ, chiều cao, cân nặng, chiều dài giác ngủ của con người, sự biến động giá trị cổ phiếu trên thị trường chứng khoán, hay mức thu nhập người lao động...

Phân phối chuẩn xuất hiện xuyên suốt trong chương trình xác suất thống kê đang được giảng dạy ở các trường đại học cũng như trong nhiều nghiên cứu chuyên ngành. Do đó, việc tìm hiểu cặn kẽ những kiến thức cơ bản về phân phối chuẩn là hết sức quan trọng. Chính vì vậy, tôi chọn làm báo cáo học thuật “Phân phối chuẩn và các moment của phân phối chuẩn”

Báo cáo học thuật chia làm ba phần

Phần 1: Giới thiệu sơ lược về lịch sử ra đời của phân phối chuẩn.

Phần 2: Trình bày về định nghĩa và các tính chất quan trọng của phân phối chuẩn.

Phần 3. Trình bày về moment của phân phối chuẩn.

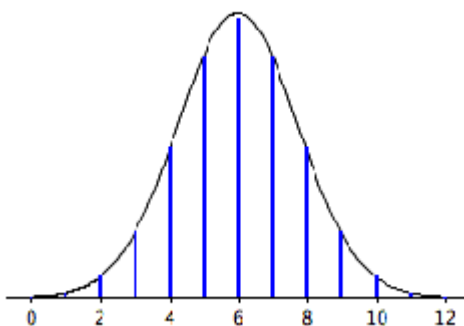
1. Lịch sử hình thành phân phối chuẩn

Trong chương về xác suất, chúng ta đã thấy rằng phân phối nhị thức có thể được sử dụng để giải các bài toán như "Tung đồng xu 100 lần thì xác suất có từ 60 lần trở xuất hiện mặt sấp là bao nhiêu?" Xác suất chính xác là

$$P = \sum_{k=60}^{100} C_{100}^k \left(\frac{1}{2}\right)^{100}$$

Ta sẽ phải tính xác suất có 60 lần, 61 lần, 62 lần... xuất hiện mặt sấp sau đó cộng tất cả các xác suất này. Hãy tưởng tượng phải mất bao lâu để tính toán xác suất theo công thức của phân phối nhị thức trước khi máy tính ra đời?

Abraham de Moivre (1667-1754), một nhà thống kê thế kỷ 18 và nhà tư vấn cho những người chơi cờ bạc. Ông thường được yêu cầu thực hiện những tính toán dài dòng kiểu này. De Moivre nhận thấy rằng khi số lần thực hiện phép thử (tung đồng xu) tăng lên, hình dạng của phân phối nhị thức tiến tới một đường cong rất mịn



De Moivre lý luận rằng nếu ông có thể tìm thấy một biểu thức toán học cho đường cong này, ông có thể giải các bài toán như tìm xác suất của những bài toán về số lần xuất hiện biến cố A trong dãy 100 hoặc hơn phép thử Bernoulli. Những lý luận này đã được ông trình bày trong bài báo năm 1734 (được in lại trong ấn bản lần 2 *The Doctrine of Chances*, 1738).

Kết quả được mở rộng bởi Laplace trong cuốn sách *Analytical Theory of probabilities* (1812), và bây giờ gọi là định lý Moivre-Laplace. Laplace đã dùng phân phối chuẩn để phân tích sai số của các thử nghiệm. Vào năm 1809, Gauss sử dụng đường cong chuẩn để phân tích dữ liệu thiên văn. Do đó, đường cong chuẩn còn thường được gọi là phân bố Gauss. Trong cuốn *Theorie Analytique des Probabilites*, Gauss phát triển các đặc điểm của luật phân phối chuẩn và chỉ ra rằng luật phân phối này phù hợp với các hiện tượng tự nhiên.

Tên gọi "đường cong chuông" do Joffret, người đầu tiên dùng thuật ngữ "bề mặt hình chuông" (năm 1872) cho phân phối chuẩn hai chiều với các thành phần độc lập. Tên gọi "phân phối chuẩn" được tạo ra bởi Charles S. Peirce, Francis Galton và Wilhelm Lexis khoảng năm 1875.

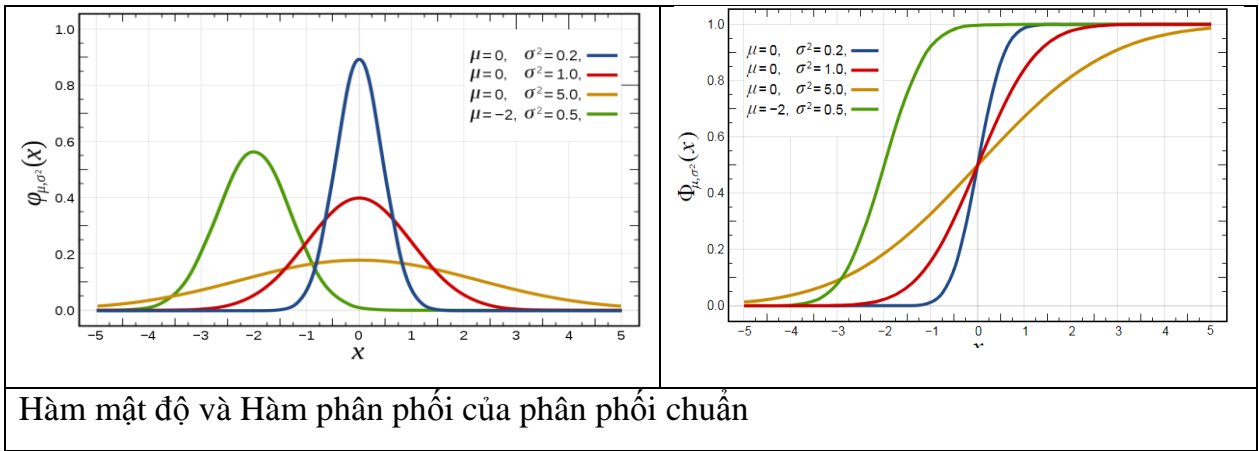
2. Định nghĩa và tính chất của phân phối chuẩn

2.1. Định nghĩa. Biến ngẫu nhiên X gọi là có phân phối chuẩn với tham số μ và σ^2 nếu hàm mật độ của nó có dạng:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Kí hiệu là $X \sim N(\mu, \sigma^2)$

Hàm phân phối $F(x) = \int_{-\infty}^x f(t)dt$.



Đặc biệt, khi $\mu = 0, \sigma = 1$, phân phối $N(0,1)$ được gọi là phân phối chuẩn tắc.

Khi đó, hàm mật độ $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ được gọi là hàm Gauss.

Hàm phân phối $F(x) = \int_{-\infty}^x \varphi(t)dt = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$. Do $\varphi(x)$ là hàm chẵn nên $F(x)$ là hàm lẻ. Ta chỉ cần xét trường hợp $x > 0$. Khi đó,

$$F(x) = \int_{-\infty}^0 \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt + \int_0^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.$$

Xét tích phân $I = \int_{-\infty}^{+\infty} e^{-x^2} dx$.

Ta có,

$$I^2 = \int_{-\infty}^{+\infty} e^{-x^2} dx \cdot \int_{-\infty}^{+\infty} e^{-y^2} dy = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-(x^2+y^2)} dx dy.$$

Áp dụng phương pháp đổi biến sang tọa độ cực để tính tích phân ta thu được

$$I^2 = \int_0^{2\pi} d\varphi \int_0^{+\infty} r \cdot e^{-r^2} dr = \pi.$$

Suy ra, $I = \sqrt{\pi}$. Kết hợp với $\varphi(x)$ là hàm chẵn ta được $\int_{-\infty}^0 \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = 0,5$. Hàm phân phối của phân phối chuẩn tắc trở thành

$$F(x) = P(X < x) = 0,5 + \int_0^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = 0,5 + \phi_0(x).$$

Hàm $\phi_0(x)$ được gọi là hàm Laplace. Giá trị của Hàm Laplace đã được các nhà toán học tính toán gần đúng. Thậm chí đã được tích hợp vào trong máy tính bỏ túi.

2.2. Tính chất.

❖ *Tính chất 1.* Nếu $X \sim N(\mu, \sigma^2)$ thì $EX = \mu, DX = \sigma^2$.

Chứng minh:

$$\text{Ta có } EX = \int_{-\infty}^{+\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Đặt $t = \frac{x-\mu}{\sigma}$, ta có $x = \sigma t + \mu$. Do vậy $dt = \frac{1}{\sigma} dx$. Suy ra

$$EX = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\sigma t + \mu) e^{-\frac{t^2}{2}} dt = \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} t e^{-\frac{t^2}{2}} dt + \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt$$

Tích phân $\int_{-\infty}^{+\infty} t e^{-\frac{t^2}{2}} dt$ là tích phân của một hàm lẻ trên khoảng đối xứng nên bằng 0. Tích

phân $\int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt$ là tích phân của hàm mật độ trên \mathbb{R} nên bằng 1.

Vậy $EX = \mu$.

Tương tự,

Vậy, $DX = EX^2 - (EX)^2 = \sigma^2$.

❖ *Tính chất 2.* Nếu $X \sim N(\mu, \sigma^2)$ thì $\frac{X-\mu}{\sigma} \sim N(0,1)$ và $aX + b \sim N(a\mu + b, (a\sigma)^2)$.

Do đó, ta có công thức tính xác suất của phân phối chuẩn

$$P(X < a) = 0,5 + \phi_0\left(\frac{a - \mu}{\sigma}\right)$$

$$P(X > a) = 0,5 - \phi_0\left(\frac{a - \mu}{\sigma}\right)$$

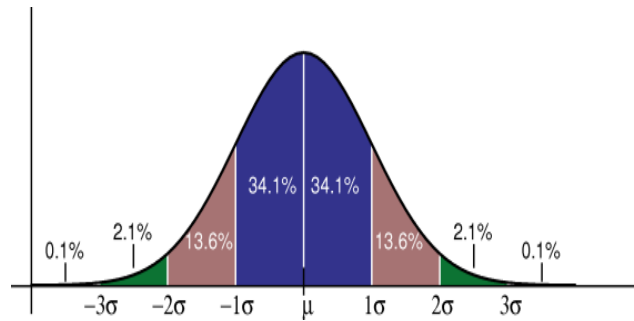
$$P(a < X < b) = \phi_0\left(\frac{b - \mu}{\sigma}\right) - \phi_0\left(\frac{a - \mu}{\sigma}\right)$$

Đặc biệt

$$P(|X - \mu| < \sigma) = P(\mu - \sigma < X < \mu + \sigma) = 2\phi_0(1) \approx 0,6827$$

$$P(|X - \mu| < 2\sigma) = 2\phi_0(2) \approx 0,9544$$

$$P(|X - \mu| < 3\sigma) = 2\phi_0(3) \approx 0,997$$



❖ *Tính chất 3* (Về hàm đặc trưng) Nếu $X \sim N(\mu, \sigma^2)$ thì hàm đặc trưng

$$\varphi_X(t) = E(e^{itX}) = \exp\left\{i\mu t - \frac{1}{2}\sigma^2 t^2\right\}, t \in \mathbb{R}.$$

Chứng minh. Giả sử $Y \sim N(0,1)$. Khi đó,

$$\varphi_Y(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{itx - \frac{1}{2}x^2} dx.$$

Đạo hàm $\varphi_Y(t)$ theo t ta được $\varphi_Y'(t) = -t \cdot \varphi_Y(t)$. Giải phương trình vi phân tuyến tính thuần nhất ta có

$$\varphi_Y(t) = C e^{\frac{-t^2}{2}}.$$

Nhưng do $\varphi_Y(0) = 1$ nên $C = 1$. Do đó, $\varphi_Y(t) = e^{-\frac{t^2}{2}}$.

Mặt khác, $X = \sigma Y + \mu$. Suy ra,

$$\varphi_X(t) = E(e^{itX}) = E(e^{it\sigma Y + it\mu}) = e^{it\mu} \cdot \varphi_Y(\sigma t) = e^{it\mu - \frac{\sigma^2 t^2}{2}}.$$

❖ *Tính chất 4.* (Về tổng của các biến ngẫu nhiên độc lập có cùng phân phối chuẩn)

- Giả sử $X_i \sim N(\mu, \sigma^2), i = 1, 2, \dots, n$. Ta có hàm đặc trưng

$$\varphi_{X_1 + \dots + X_n}(t) = \prod_{i=1}^n \varphi_{X_i}(t) = e^{itn\mu - \frac{n\sigma^2 t^2}{2}}, i = 1, 2, \dots, n.$$

Ta thấy $\varphi_{X_1 + \dots + X_n}(t) = \varphi_T(t)$, trong đó, $T \sim N(n\mu, n\sigma^2)$. Do tính chất duy nhất của hàm đặc trưng nên $\sum_{i=1}^n X_i \sim N(n\mu, n\sigma^2)$.

Tổng quát hơn, nếu $X_i \sim N(\mu_i, \sigma_i^2)$ thì $\sum_{i=1}^n X_i \sim N(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2)$.

- Ta cũng có nếu $X_1 \sim N(\mu_1, \sigma_1^2), X_2 \sim N(\mu_2, \sigma_2^2)$ thì $X_1 - X_2 \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$.
- Giả sử $X_i, i = 1, 2, \dots, n$ là các biến ngẫu nhiên độc lập có cùng phân phối chuẩn $N(0,1)$. Xét phân phối của thống kê $T(X) = \sum_{i=1}^n X_i^2$.

Ta có $P[X_i^2 < z] = 0$ nếu $z < 0$. Xét $z > 0$

$$\begin{aligned} P[X_i^2 < z] &= P[-\sqrt{z} < X_i < \sqrt{z}] = \int_{-\sqrt{z}}^{\sqrt{z}} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}u^2\right\} du \\ &= 2 \int_0^{\sqrt{z}} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}u^2\right\} du. \end{aligned}$$

Lấy đạo hàm 2 vế ta được hàm mật độ của X_i^2 là

$$f(x) = \frac{1}{\sqrt{2\pi}} x^{-\frac{1}{2}} e^{-\frac{x}{2}} I_{(0,+\infty)}(x).$$

Vì $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$ nên $X_i^2 \sim G\left(\frac{1}{2}, \frac{1}{2}\right) = \chi_1^2, i = 1, 2, \dots, n$.

Và từ đó suy ra $T = \sum_{i=1}^n X_i^2 \sim \chi_n^2$.

3. Moment của phân phối chuẩn

Định nghĩa. Cho X là một biến ngẫu nhiên và $k > 0$ là một số tự nhiên. Đại lượng $E(X^k)$ được gọi là moment bậc k của X .

- Moment của phân phối chuẩn tắc: Giả sử $Z \sim N(0,1)$. Theo tính chất về kỳ vọng và phương sai của phân phối chuẩn ta có $EZ = 0, EZ^2 = 1$.

Mệnh đề. $E(Z^{n+1}) = nE(Z^{n-1}), n \in \mathbb{N}^*$.

Chứng minh: Hàm mật độ của phân phối chuẩn tắc là hàm Gauss $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

Dễ thấy, $\varphi'(x) = -x\varphi(x)$.

$$\begin{aligned} E(Z^{n+1}) &= \int_{-\infty}^{+\infty} z^{n+1} \varphi(z) dz = \int_{-\infty}^{+\infty} z^n \cdot z \cdot \varphi(z) dz = - \int_{-\infty}^{+\infty} z^n d(\varphi(z)) \\ &= -z^n \cdot \varphi(z) \Big|_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} n \cdot z^{n-1} \varphi(z) dz = 0 + nE(Z^{n-1}) = nE(Z^{n-1}) \end{aligned}$$

Hệ quả. Với mọi $n \in \mathbb{N}^*$ ta có

$$E(Z^{2n}) = 1.3...(2n-1) = \frac{(2n)!}{2^n \cdot n!}$$

$$E(Z^{2n+1}) = 0.$$

- Moment của phân phối chuẩn tổng quát

Chú ý rằng, nếu $X \sim N(\mu, \sigma^2)$ thì $X = \mu + \sigma Z$, với $Z \sim N(0,1)$. Từ đó ta tính được

$$E(X^k) = E((\mu + \sigma Z)^k).$$

Chẳng hạn như

$$EX = \mu$$

$$EX^2 = \mu^2 + \sigma^2$$

$$EX^3 = \mu^3 + 3\mu\sigma^2$$

$$EX^4 = \mu^4 + 6\mu^2\sigma^2 + 4\sigma^4.$$

KẾT LUẬN

Phân phối chuẩn có thể coi là phân phối quan trọng bậc nhất trong xác suất và thống kê. Báo cáo đã trình bày sơ lược về lịch sử của phân phối chuẩn, định nghĩa của phân phối chuẩn cũng như một số tính chất quan trọng của phân phối chuẩn. Báo cáo cũng đưa ra một trong một số rất nhiều những phương pháp tính toán các moment của phân phối chuẩn. Tuy nhiên, phần kiến thức này chỉ là phần nội dung kiến thức cơ bản nhất về phân phối chuẩn. Bên cạnh đó, chúng ta còn có nhiều nội dung quan trọng khác liên quan đến phân phối chuẩn được giảng dạy trong học phần xác suất thống kê hay chuyên sâu hơn nữa, chẳng hạn như định lý giới hạn trung tâm, các phân phối liên quan đến đặc trưng mẫu hay tính toán các moment trung tâm của phân phối chuẩn... Trong tương lai, tác giả sẽ tiếp tục tìm hiểu về các tính chất cũng như các ứng dụng của phân phối chuẩn để phục vụ giảng dạy và các nghiên cứu chuyên ngành.

TÀI LIỆU THAM KHẢO

1. Nguyễn Viết Phú, Nguyễn Duy Tiến, *Cơ sở Lý thuyết Xác suất*, NXB ĐHQG Hà Nội.
2. https://vi.wikipedia.org/wiki/Ph%C3%A2n_ph%E1%BB%91i_chu%E1%BA%A9n
3. David M. Lane, *History of the Normal Distribution*, <http://onlinestatbook.com/>.