Như vậy, mặc dù từng BNN độc lập có thể nhận giá trị khác nhiều so với kì vọng của chúng, song trung bình số học của một số lớn các BNN lại nhận giá trị gần bằng kì vọng của chúng với xác suất rất lớn. Sự kiện này cho phép ta ước lượng kì vọng bằng trung bình cộng các kết quả đo đạc độc lập của BNN có kì vọng đó. Chẳng hạn, gieo một con xúc xắc cân đối. Giả sử X là số nốt xuất hiện ở mặt trên con xúc xắc. Ta có $E(X) = 3,5$. Một nhà thống kê đã gieo một con xúc xắc cân đối 1 triệu lần (nhờ sự trợ giúp của máy vi tính) và ghi lại số nốt xuất hiện ở mặt trên con xúc xắc. Số trung bình của 1 triệu lần gieo được tìm thấy là

$$\frac{x_1 + \cdots + x_{10^6}}{10^6} \approx 3,500876.$$

Luật số lớn Chebyschev có ứng dụng rộng rãi trong nhiều lĩnh vực, chẳng hạn nó chính là cơ sở cho phương pháp đo lường trong vật lí. Để xác định giá trị của một đại lượng vật lí nào đó người ta thường tiến hành đo n lần độc lập và lấy trung bình số học của các kết quả đo làm giá trị thực của đại lượng cần đo. Thật vậy, giả sử xem kết quả của n lần đo là các BNN $X_1, \cdots, X_n$. Ta thấy rằng các BNN này độc lập, có cùng kì vọng bằng chính giá trị thực của đại lượng vật lí (giả sử không có sai số hệ thống), các phương sai của chúng đều bị chặn trên bởi bình phương của độ chính xác của thiết bị đo. Do đó theo Định lí 2.3 ta có thể cho rằng trung bình số học của các kết quả đo sẽ sai lệch rất ít so với giá trị thực của đại lượng vật lí với xác suất gần như bằng một.

2.3. MỘT SỐ PHÂN PHỐI XÁC SUẤT THƯỜNG GẶP

2.3.1. Phân phối nhị thức

a) Phân phối Bernoulli

Định nghĩa 2.11. BNN X được gọi là tuân theo luật phân phối Bernoulli, kí hiệu $X \sim \mathcal{B}(1; p)$, nếu X có bảng phân phối xác suất như sau

X	0	1
$p(x)$	p	$1-p$

Ta thấy mọi phép thử chỉ có hai kết cục đều có thể mô hình hóa bằng phân phối này. Chẳng hạn một phép thử chỉ có kết quả A với xác suất p và $\overline{A}$ với xác suất $q = 1 - p$. Xây dựng BNN X sao cho $P(X = 1) = P(A) = p$ và $P(X = 0) = P(\overline{A}) = q$.

Từ bảng phân phối của X ta có các khẳng định sau:

(i) $E(X) = 0.q + 1.p = p$,

(ii) $D(X) = 0^2.q + 1^2.p - p^2 = p.q$.

Trong lí thuyết thống kê, BNN có phân phối Bernoulli thường được dùng để đặc trưng cho các dấu hiệu nghiên cứu có tính định tính trong đó mỗi cá thể của tổng thể có dấu hiệu này hoặc không có dấu hiệu này. Chẳng hạn khi muốn nghiên cứu giới tính của khách hàng ta có thể đặc trưng cho giới tính bằng BNN với 2 giá trị bằng 1 (Nam) và bằng 0 (Nữ); trong bài toán bầu cử, nếu cử tri nào sẽ bỏ phiếu cho ứng cử viên A ta cho nhận giá trị 1, ngược lại ta cho nhận giá trị 0; để xác định tỷ lệ phế phẩm của lô hàng ta gán cho mỗi sản phẩm một trong hai giá trị 0 và 1, nếu sản phẩm là phế phẩm ta cho nhận giá trị 1 và ngược lại cho nhận giá trị 0... Đó là các BNN có phân phối Bernoulli.

b) Phân phối nhị thức

Đây là một trong các phân phối rất hay dùng trong thống kê hiện đại. Ở Chương I ta đã làm quen với lược đồ Bernoulli khi xét dãy n phép thử độc lập, giống nhau, trong mỗi phép thử sự kiện A xuất hiện với xác suất p. Nếu gọi X là số lần xuất hiện A trong dãy n phép thử đó, ta đã biết X có các giá trị từ 0 đến n với các xác suất tương ứng được xác định bởi

$$p(x) = P_n(x;p) = C_n^x p^x (1-p)^{n-x}, \quad x = \overline{0,n}. \tag{2.8}$$

Định nghĩa 2.12. BNN X được gọi là tuân theo luật phân phối nhị thức, kí hiệu $X \sim \mathcal{B}(n;p)$, nếu hàm xác suất của nó có dạng *(2.8)*.

Bảng 2.1: Bảng phân phối xác suất của BNN $\mathbf{X} \sim \mathcal{B}(\mathbf{n};\mathbf{p})$

X	0	1	$\cdots$	k	$\cdots$	n
$p(x)$	$(1-p)^n$	$C_n^1 p(1-p)^{n-1}$	$\cdots$	$C_n^k p^k (1-p)^{n-k}$	$\cdots$	p^n

Thực hiện n phép thử Bernoulli với xác suất thành công của sự kiện A trong mỗi lần thử là p. Với mỗi $i = \overline{1,n}$, nếu ở lần thử thứ i, sự kiện A xuất hiện thì X_i nhận giá trị 1, nếu sự kiện A không xuất hiện thì X_i nhận giá trị 0. Như vậy $X_i, i = \overline{1,n}$ là BNN có phân phối Bernoulli. Gọi X là số lần xuất hiện sự kiện A đó trong dãy n phép thử Bernoulli. Khi đó

$$X = X_1 + X_2 + \cdots + X_n \sim \mathcal{B}(n;p).$$

Do các $X_i (i = \overline{1,n})$ độc lập, mặt khác $E(X_i) = p, D(X_i) = p.q, \forall i = \overline{1,n}$, $q = 1-p$, nên ta có các kết quả sau:

(i) $E(X) = E\left(\sum\limits_{i=1}^{n} X_i\right) = \sum\limits_{i=1}^{n} E(X_i) = n.p,$

(ii) $D(X) = D\left(\sum\limits_{i=1}^{n} X_i\right) = \sum\limits_{i=1}^{n} D(X_i) = n.p.q,$

(iii) Từ Định lí 1.2 suy ra Mod$X = m$ thỏa mãn $(n+1)p - 1 \le m \le (n+1)p$.

Ví dụ 2.19. Theo một điều tra xã hội học cho thấy tỉ lệ sinh viên học không đúng với ngành nghề mà họ yêu thích là 34%. Một lớp gồm 60 sinh viên. Gọi X là số sinh viên không theo đúng ngành nghề yêu thích trong 60 sinh viên này.

a) Hãy mô tả quy luật phân phối của X.

b) Về trung bình thì trong 60 sinh viên sẽ có bao nhiêu sinh viên không thích ngành đang học?

c) Có bao nhiêu sinh viên không thích ngành đang học là có khả năng nhất?

Lời giải.

a) Lớp có 60 sinh viên được coi như ta đã chọn ngẫu nhiên ra 60 sinh viên từ tập tất cả sinh viên với tỉ lệ không thích ngành đang học là 34%. Việc chọn 60 sinh viên tương ứng với 60 phép thử Bernoulli với xác suất không đúng ngành nghề yêu thích là $p = 0,34$. Như vậy BNN $X \sim \mathcal{B}(60; 0,34)$.

b) Về trung bình trong 60 sinh viên sẽ có $E(X) = 60.0,34 = 20,4$ sinh viên không thích ngành đang học.

c) Ta có $(n+1).p - 1 = (60+1).0,34 - 1 = 19,74$ nên $m = 20$. Trong 60 sinh viên thì tình huống có 20 sinh viên không thích ngành đang học là có khả năng xảy ra cao nhất.

2.3.2. Phân phối Poisson

Khi n đủ lớn, việc tính giá trị hàm xác suất của phân phối nhị thức hết sức cồng kềnh. Nhà toán học và vật lí học người Pháp Siméon Denis Poisson (1781-1840) đã đưa ra một phân phối mang tên ông để khắc phục hạn chế này.

Ví dụ 2.20. BNN "số vụ tai nạn giao thông xảy ra trong một ngày" ở một vùng nào đó có thể được mô hình hóa như sau. Giả sử các tai nạn giao thông xảy ra một cách ngẫu nhiên, độc lập với nhau, và trung bình mỗi ngày có λ vụ tai nạn. Ta chia 24 tiếng đồng hồ trong ngày thành n khoảng thời gian (n có thể là một số rất lớn) sao cho trong mỗi khoảng thời gian có nhiều nhất 1 vụ tai nạn giao thông, và khả năng xảy ra tai nạn giao thông trong mỗi khoảng thời gian bằng λ/n. Khi đó tổng số tai nạn xảy ra trong ngày có phân phối $\mathcal{B}\left(n; \dfrac{\lambda}{n}\right)$. Ta có

$$P(X = k) = C_n^k \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \to \frac{\lambda^k}{k!} e^{-\lambda}, \quad khi\ n \to \infty.$$

Như vậy, khi $n \to \infty$ thì X trở thành phân phối mới được gọi là phân phối Poisson. Tất nhiên phân phối Poisson không thể là phân phối xác suất chính xác của vấn đề (vì số người là hữu hạn, và số tai nạn bị chặn trên bởi số người chứ không lớn tuỳ ý được), nhưng nó là phân phối gần đúng thuận tiện cho việc tính toán.

Định nghĩa 2.13. BNN X được gọi là tuân theo luật phân phối Poisson tham số $\lambda > 0$, kí hiệu là $X \sim \mathcal{P}(\lambda)$, nếu hàm xác suất của nó có dạng

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad với\ x = 0, 1, 2, \cdots,$$

trong đó, λ là trung bình số lần xuất hiện sự kiện ta quan tâm trong một khoảng xác định (khoảng thời gian hoặc một khoảng đơn vị tính nào đó).

Phân phối Poisson có nhiều ứng dụng trong lí thuyết phục vụ đám đông, kiểm tra chất lượng sản phẩm... Chẳng hạn số cuộc gọi điện thoại của một tổng đài trong khoảng thời gian nào đó (1 giờ, 1 ngày), số lượng khách hàng đến một địa điểm trong 1 giờ, số phương tiện giao thông qua một ngã tư, số hạt α phát ra từ các hạt phóng xạ trong một chu kì, số lỗi in sai trong một trang (hoặc một số trang) của một quyển sách, số transitor bị hỏng trong ngày đầu tiên sử dụng... đều là các BNN có phân phối Poisson.

Ví dụ 2.21. Một tổng đài chuyển điện trong khoảng thời gian 10^{-5} giây. Trong quá trình chuyển điện, số tín hiệu ồn ngẫu nhiên trong 1 giây là 10^4. Nếu trong thời gian truyền tín hiệu, có đúng một tín hiệu ồn ngẫu nhiên thì tổng đài sẽ ngừng làm việc. Tính xác suất để cho việc truyền tính hiệu bị gián đoạn. Biết rằng số tín hiệu ồn ngẫu nhiên rơi vào trong khoảng thời gian truyền tín hiệu tuân theo luật phân phối Poisson.

Lời giải. Gọi X là BNN chỉ số các tín hiệu ồn trong khoảng thời gian 10^{-5} giây truyền tin. Khi đó

$$X \sim \mathcal{P}(10^4.10^{-5}) \Leftrightarrow X \sim \mathcal{P}(0,1).$$

Do đó, xác suất việc truyền tin bị gián đoạn là

$$P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-0,1}\frac{(0,1)^0}{0!} = 0,0952.$$

Ví dụ 2.22. Người ta vận chuyển 5000 chai rượu vào kho với xác suất vỡ của mỗi chai là 4.10^{-4}. Hỏi xác suất để khi vận chuyển có không quá 1 chai vỡ?

Lời giải. Có thể dùng lược đồ Bernoulli (phân phối nhị thức), nhưng $n = 5000$ rất lớn, còn $p = 0,0004$ nhỏ. Nếu gọi X là số chai bị vỡ khi vận chuyển, có thể coi phân phối của X xấp xỉ với phân phối Poisson với $\lambda = n.p = 2$. Khi đó

$$P(0 \leq X \leq 1) = e^{-2}.\frac{2^0}{0!} + e^{-2}.\frac{2^1}{1!} = \frac{3}{e^2} \approx 0,406.$$

Các đặc số của BNN $X \sim \mathcal{P}(\lambda)$ gồm: $E(X) = D(X) = \lambda$ và $\lambda - 1 \leq \text{Mod}X \leq \lambda$.

2.3.3. Phân phối mũ

Định nghĩa 2.14. BNN X được gọi là có phân phối mũ tham số $\lambda > 0$ nếu có hàm mật độ

$$f(x) = \begin{cases} 0, & \text{nếu } x \leq 0, \\ \lambda e^{-\lambda x}, & \text{nếu } x > 0. \end{cases}$$

Phân phối mũ thường xuất hiện trong các bài toán về thời gian sống của một loài sinh vật, tuổi thọ của thiết bị... hoặc khoảng cách giữa hai lần xuất hiện của một sự kiện A nào đó mà số lần xuất hiện của A tuân theo luật phân phối Poisson như: khoảng thời gian giữa 2 ca cấp cứu ở một bệnh viện, giữa 2 lần hỏng hóc của một cái máy; khoảng cách giữa 2 gen đột biến kế tiếp trên một dải ADN,...

Các đặc số của phân phối mũ là: $E(X) = \dfrac{1}{\lambda}$, $D(X) = \dfrac{1}{\lambda^2}$.

Ví dụ 2.23. Giả sử tuổi thọ (tính bằng năm) của một mạch điện tử trong máy vi tính là một BNN có phân phối mũ với kì vọng là 6,25. Thời gian bảo hành của mạch điện tử này là 5 năm. Hỏi có bao nhiêu phần trăm mạch điện tử bán ra phải thay thế trong thời gian bảo hành?

Lời giải. Gọi X là tuổi thọ của mạch điện tử, vì X có phân phối mũ nên $\lambda = 1/E(X) = 1/6,25$. Khi đó

$$P(X \leq 5) = \frac{1}{6,25}\int_0^5 e^{-x/6,25}dx = 1 - e^{-0,8} \approx 1 - 0,449 = 0,551.$$

Có khoảng 55% mạch điện tử bán ra phải thay thế trong thời gian bảo hành.

2.3.4. Phân phối đều

a) Phân phối đều rời rạc

Định nghĩa 2.15. BNN X được gọi là tuân theo luật phân phối đều rời rạc với tham số n, kí hiệu là $X \sim \mathcal{U}(n)$, nếu X có bảng phân phối xác suất

X	1	2	$\cdots$	n
$p(x)$	$\frac{1}{n}$	$\frac{1}{n}$	$\cdots$	$\frac{1}{n}$

Như vậy hàm xác suất sẽ có dạng $p(i) = \dfrac{1}{n}, i = \overline{1, n}$. Người ta còn mở rộng khái niệm phân phối đều cho BNN X nhận giá trị trên một tập hữu hạn bất kì có n phần tử $\{x_1, x_2, \cdots, x_n\}$, khi đó $p(x_i) = \dfrac{1}{n}, i = \overline{1, n}$.

Các tham số đặc trưng của phân phối đều rời rạc là:

$$E(X) = \frac{n+1}{2} \quad \text{và} \quad D(X) = \frac{n^2-1}{12}.$$

b) Phân phối đều liên tục

Định nghĩa 2.16. BNN X được gọi là tuân theo luật phân phối đều liên tục trên $[a;b]$, kí hiệu là $X \sim \mathcal{U}([a;b])$, nếu X có hàm mật độ xác suất

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & \text{nếu } x \in [a;b], \\ 0, & \text{nếu } x \notin [a;b]. \end{cases}$$

Nếu biết rằng BNN X nhận giá trị nào đó trong khoảng $(a;b)$ mà không biết thêm thông tin gì khác về X thì có thể xem mỗi giá trị có thể của X trong $(a;b)$ là đồng khả năng. Nói cách khác, X có phân phối đều trên $(a;b)$.

Các đặc số của phân phối đều liên tục là: $E(X) = \dfrac{a+b}{2}$, $D(X) = \dfrac{(b-a)^2}{12}$, $\text{Mod}X = m, \forall m \in (a;b)$, $\text{Med}X = \dfrac{a+b}{2}$.

Ví dụ 2.24. Lịch chạy xe buýt tại một trạm xe buýt được bố trí như sau: chuyến đầu tiên trong ngày sẽ khởi hành từ trạm này lúc 7h00, cứ sau mỗi 15 phút sẽ có một xe khác đến trạm. Giả sử một hành khách đến trạm vào khoảng thời gian từ 7h00 đến 7h30. Tìm xác suất để hành khách này phải chờ

a) Ít hơn 5 phút.

b) Ít nhất 12 phút.

Lời giải. Gọi X là số phút sau 7h00 mà hành khách đến trạm xe buýt thì X là BNN có phân phối đều trong khoảng $(0;30)$.

a) Hành khách sẽ chờ ít hơn 5 phút nếu đến trạm xe buýt trong khoảng từ 7h10 đến 7h15 hoặc từ 7h25 đến 7h30. Do đó xác suất cần tìm là

$$P(10 < X < 15) + P(25 < X < 30) = \frac{1}{30}\left(\int_{10}^{15} dx + \int_{25}^{30} dx\right) = \frac{1}{3}.$$

b) Hành khách chờ ít nhất 12 phút nếu đến trạm từ 7h00 đến 7h03 hoặc từ 7h15 đến 7h18. Xác suất cần tìm là

$$P(0 < X < 3) + P(15 < X < 18) = \frac{1}{30}\left(\int_{0}^{3} dx + \int_{15}^{18} dx\right) = \frac{1}{5}.$$

2.3.5. Phân phối chuẩn

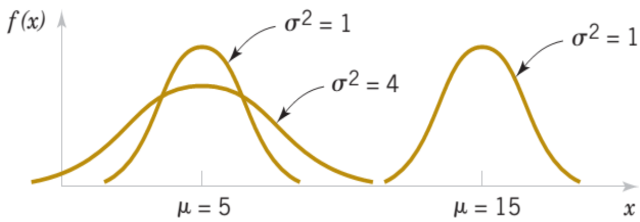
Phân phối chuẩn là phân phối thường gặp trong tự nhiên cũng như trong kĩ thuật, nó đóng vai trò rất quan trọng trong lí thuyết xác suất cũng như các ứng dụng của phân phối này trong xử lí số liệu. Phân phối chuẩn còn có tên gọi là *phân phối Gauss*.

a) Phân phối chuẩn và phân phối chuẩn tắc

Định nghĩa 2.17. BNN X được gọi là tuân theo luật phân phối chuẩn, kí hiệu là $X \sim \mathcal{N}(\mu, \sigma^2)$, nếu hàm mật độ của nó có dạng

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

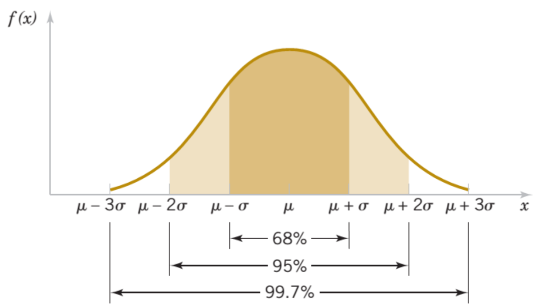
Đồ thị của hàm mật độ của phân phối chuẩn có hình cái chuông, và bởi vậy phân phối này còn được gọi là *phân phối hình chuông*. Trung điểm của cái chuông này chính là điểm $x = \mu$, và độ cao của chuông chính bằng $\dfrac{1}{\sigma\sqrt{2\pi}}$. Nếu σ càng nhỏ thì chuông càng cao và càng "hẹp", ngược lại σ càng lớn thì chuông càng thấp và càng rộng ra (H.2.2).



Hình 2.2: Đồ thị hàm mật độ phân phối chuẩn

Hình 2.3 cho thấy hầu hết xác suất của một phân phối chuẩn nằm trong đoạn $[\mu - 3\sigma, \mu + 3\sigma]$. Cụ thể, nếu $X \sim \mathcal{N}(\mu, \sigma^2)$, thì với xác suất 99,7% ta có thể tin rằng giá trị của X nằm trong $[\mu - 3\sigma, \mu + 3\sigma]$.

Phân phối chuẩn là một trong những phân phối xác suất quan trọng nhất, vì nhiều phân phối xác suất gặp trong thực tế có dáng điệu khá giống phân phối chuẩn, ví dụ như phân phối của chiều cao của đàn ông, phân phối của chỉ số IQ (chỉ số trí tuệ), phân phối của giá chứng khoán trong tương lai... Định lí giới hạn trung tâm mà chúng ta đề cập đến ở phần **c)** sẽ cho chúng ta cơ sở lí thuyết để hiểu tại sao có nhiều phân phối xác suất trong thực tế trông giống phân phối chuẩn.



Hình 2.3: Quy tắc $2\sigma, 3\sigma$

Các tham số đặc trưng của phân phối chuẩn gồm:

$$E(X) = \mu,\ D(X) = \sigma^2 \text{ và } \mathrm{Mod}X = \mathrm{Med}X = \mu.$$

Ví dụ 2.25. Độ dài một chi tiết máy giả sử tuân theo luật phân phối chuẩn với giá trị trung bình 20cm và độ lệch chuẩn là 0,5cm. Hãy tính xác suất khi chọn ngẫu nhiên ra một chi tiết có độ dài:
a) lớn hơn 20cm,
b) bé hơn 19,5cm,
c) lớn hơn 21,5cm.

Lời giải. Gọi X là độ dài của chi tiết máy chọn ra, rõ ràng $X \sim \mathcal{N}(20; 0,5)$.
a) Do $\mathrm{Med}X = 20$ nên $P(X > 20) = 1 - P(X \le 20) = 0,5$.
b) Từ Hình 2.3, ta có

$$P(19,5 \le X \le 20,5) = P(\mu - \sigma \le X \le \mu + \sigma) = 0,68.$$

Do tính đối xứng nên $P(X < 19,5) = \dfrac{0,32}{2} = 0,16$ (và cũng bằng $P(X > 20,5)$).
c) Tương tự câu b) ta được

$$P(18,5 \le X \le 21,5) = P(\mu - 3\sigma \le X \le \mu + 3\sigma) = 0,99.$$

Từ đây suy ra $P(X > 21,5) = \dfrac{0,01}{2} = 0,005$.

Định nghĩa 2.18. Nếu BNN X có phân phối chuẩn với kì vọng $\mu = 0$ và phương sai $\sigma^2 = 1$ thì X được gọi là BNN có phân phối chuẩn tắc.

Hàm mật độ của phân phối chuẩn tắc kí hiệu là $\varphi(x)$ cho bởi

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad \forall x \in \mathbb{R}.$$

Hàm phân phối của phân phối chuẩn tắc kí hiệu là $\Phi(x)$ có biểu thức

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt, \quad \forall x \in \mathbb{R}.$$

Phụ lục Bảng I, II cung cấp bảng tính sẵn các giá trị của $\varphi(x)$ và $\Phi(x)$. Cần chú ý rằng một số tài liệu cung cấp bảng tính giá trị hàm $\Phi_0(x) = \dfrac{1}{\sqrt{2\pi}} \int\limits_0^x e^{-\frac{t^2}{2}} dt, x \ge 0$ (còn gọi là hàm Laplace). Hơn nữa phân phối chuẩn tắc có kì vọng $\mu = 0$ nên $\Phi(0) = \dfrac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{0} e^{-\frac{t^2}{2}} dt = \dfrac{1}{2}$. Khi đó

$$\Phi(x) = \Phi_0(x) + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{-\frac{t^2}{2}} dt = \Phi_0(x) + 0,5, \quad \forall x \ge 0.$$

Hàm phân phối của phân phối chuẩn tắc có các tính chất sau:
(i) $\Phi(-x) = 1 - \Phi(x)$;
(ii) Nếu $X \sim \mathcal{N}(0; 1)$ thì với mọi $a > 0$, ta có

$$P(|X| > a) = 2\big(1 - \Phi(a)\big) \quad \text{và} \quad P(|X| < a) = 2\Phi(a) - 1.$$

Định nghĩa 2.19. Giá trị U_α được gọi là giá trị tới hạn mức α của phân phối chuẩn tắc nếu $\Phi(U_\alpha) = 1 - \alpha$.

Từ định nghĩa trên, nếu $X \sim \mathcal{N}(0,1)$ và với mọi $\alpha \in (0;1)$ thì

$$P(X > U_\alpha) = P(|X| > U_{\alpha/2}) = \alpha \quad \text{và} \quad P(|X| < U_{\alpha/2}) = 1 - \alpha. \qquad (2.9)$$

Có thể chứng minh được rằng: nếu $X \sim \mathcal{N}(\mu;\sigma^2)$ thì $\dfrac{X-\mu}{\sigma} \sim \mathcal{N}(0;1)$. Do đó

$$P(X \le a) = P\left(\frac{X-\mu}{\sigma} \le \frac{a-\mu}{\sigma}\right) = \Phi\left(\frac{a-\mu}{\sigma}\right),$$

$$P(\alpha \le X \le \beta) = P\left(\frac{\alpha-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{\beta-\mu}{\sigma}\right) = \Phi\left(\frac{\beta-\mu}{\sigma}\right) - \Phi\left(\frac{\alpha-\mu}{\sigma}\right),$$

$$P(|X-\mu| < \varepsilon) = P\left(\left|\frac{X-\mu}{\sigma}\right| < \frac{\varepsilon}{\sigma}\right) = 2\Phi\left(\frac{\varepsilon}{\sigma}\right) - 1.$$

Ví dụ 2.26. Gọi X là chỉ số thông minh (IQ) của học sinh trung học cơ sở. Giả sử $X \sim \mathcal{N}(85;25)$.

a) Hỏi chỉ số IQ trung bình của học sinh trong lứa tuổi này là bao nhiêu?

b) Tính xác suất chọn được học sinh rất thông minh ($X \ge 90$).

c) Tính tỉ lệ học sinh trong lứa tuổi này có chỉ số IQ thuộc $(80;95)$.

d) Gọi Y là số học sinh có IQ thuộc $(80;95)$ trong lớp 50 học sinh. Hãy chỉ rõ luật phân phối xác suất của Y.

e) Trong một lớp gồm 50 học sinh thì trung bình có bao nhiêu em rất thông minh ($X \ge 90$)? Con số trung bình tìm được có phải là số có khả năng xảy ra cao nhất hay không? Vì sao?

Lời giải. BNN $X \sim \mathcal{N}(85;25)$ nên $E(X) = 85$, $D(X) = 25$.

a) Chỉ số IQ trung bình chính là kì vọng và bằng 85.

b) Xác suất chọn được học sinh rất thông minh là

$$\begin{aligned} P(90 \le X < +\infty) &= \Phi(+\infty) - \Phi\left(\frac{90-85}{5}\right) \\ &= 1 - \Phi(1) = 1 - 0,8413 = 0,1587. \end{aligned}$$

c) Tỉ lệ học sinh trong lứa tuổi này có chỉ số IQ thuộc $(80;95)$ là

$$\begin{aligned} P(80 < X < 95) &= \Phi\left(\frac{95-85}{5}\right) - \Phi\left(\frac{80-85}{5}\right) \\ &= \Phi(2) - \Phi(-1) = \Phi(2) + \Phi(1) - 1 \\ &= 0,9773 + 0,8413 - 1 = 0,8168 \approx 82\%. \end{aligned}$$

d) Một lớp gồm 50 học sinh được chọn từ tập học sinh với tỉ lệ

$$p = P(80 < X < 95) \approx 82\%$$

được xem như 50 phép thử Bernoulli với xác suất 82%. Như vậy $Y \sim \mathcal{B}(50;0,82)$.

e) Tương tự câu d), ta lại có 50 phép thử Bernoulli với $p = 0,1587$ nên trung bình trong lớp sẽ có $n.p = 50.0,1587 = 7,935$ (gần 8) học sinh rất thông minh. Hơn nữa số trung bình trên không phải là số có khả năng cao nhất. Số học sinh rất thông minh với xác suất lớn nhất là

$$(n+1).p - 1 \le m \le (n+1).p \Leftrightarrow m = 8.$$