

Chương IV: MỘT SỐ QUY LUẬT PHÂN PHỐI XÁC SUẤT THÔNG DỤNG

IV.1. Một số quy luật phân phối xác suất thông dụng

IV.1.1 Phân phối chuẩn

IV.1.2 Phân phối Bernoulli (PP không – một)

IV.1.3 Phân phối Nhị thức

IV.1.4 Phân phối Siêu bội

IV.1.5 Phân phối Poisson

IV.1.6 Phân phối Hình học

IV.1.7 Phân phối đều

IV.1.8 Phân phối lũy thừa

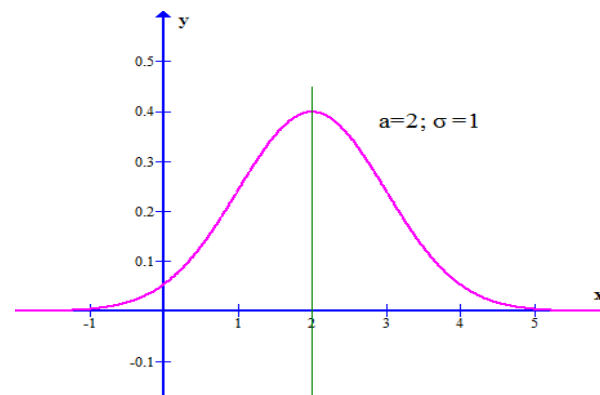
IV.1.9 Phân phối Student; PP Chi Bình Phương; PP Fisher.

IV.2. Các định lý giới hạn

IV.1.1 Phân phối chuẩn (hay là pp bình thường).

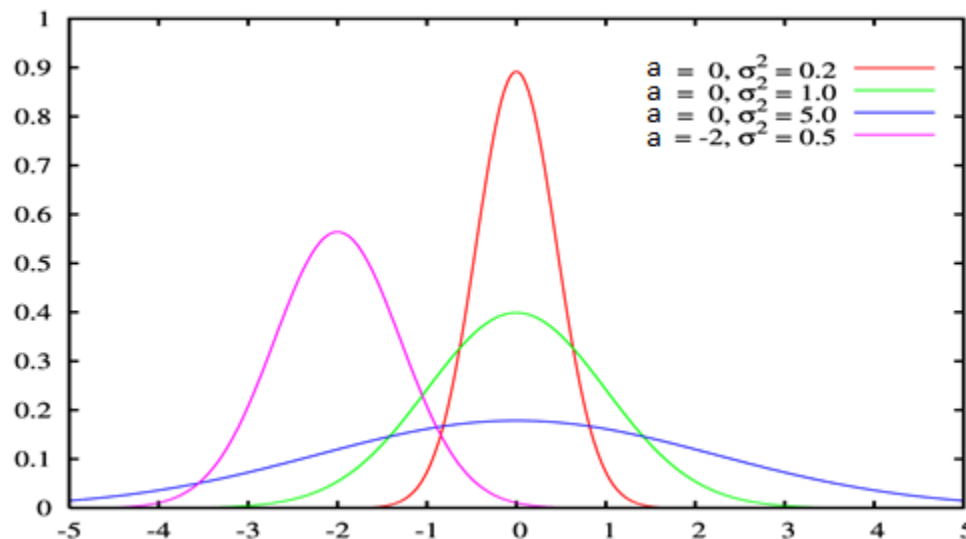
Định nghĩa: BNN X được gọi là có **phân phối chuẩn** (*Normal Distribution*), ký hiệu $X \sim N(a, \sigma^2)$, nếu hàm mật độ xác suất của X có dạng:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}, \sigma > 0, x \in \mathbb{R}$$



Tính chất:

- Nếu $X \sim N(a, \sigma^2)$
thì $E(X) = a, D(X) = \sigma^2$



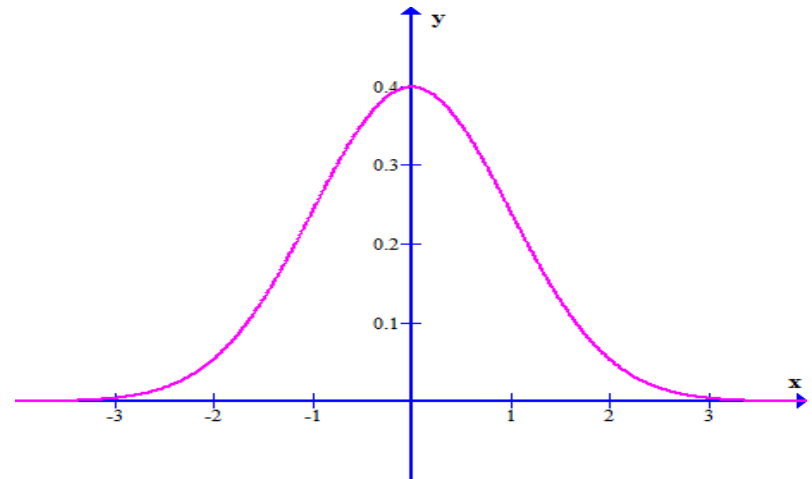
IV.1.1.a Hàm chuẩn chuẩn tắc

Trường hợp riêng: $a = 0$ và $\sigma = 1$ thì $X \sim N(0, 1)$ còn gọi là phân phối (chuẩn) chuẩn tắc. Hàm mật độ của phân phối chuẩn tắc được gọi là **hàm mật độ Gauss**.

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

- Nếu $X \sim N(a, \sigma^2)$ thì

$$Y = \frac{X-a}{\sigma} \sim N(0; 1)$$

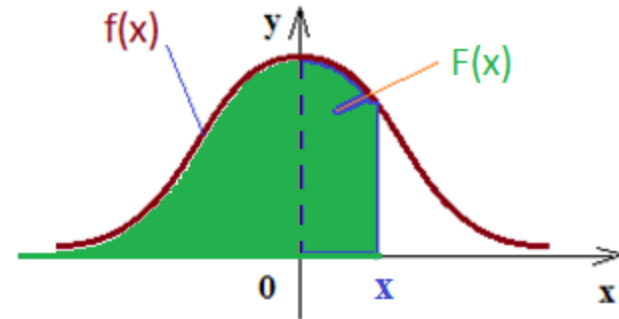


⇒ Nhờ phép đổi biến, ta có thể tìm các tính chất của hàm chuẩn thông qua việc khảo sát hàm Gauss.

- Hàm mật độ Gauss là hàm chẵn.
- Khi $|x| > 3$, hàm Gauss nhận các giá trị xấp xỉ 0.

- **Hàm phân phối** xác suất của phân phối **chuẩn tắc**:

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



Vì hàm $f(x)$ không có nguyên hàm,

chúng ta có thể tìm giá trị gần đúng của $F(x)$ bằng nhiều cách:

- Nhập trực tiếp công thức vào MTBT
- Lấy giá trị hàm $P(t)$ qua chức năng STAT của MTBT (slide 8, 9)
- Tra bảng số.
-

VD: $F(1,24) = P(1,24) = 0,89251$

$$F(-\infty) = 0$$

$$F(-1,24) = P(-1,24) = 0,10749$$

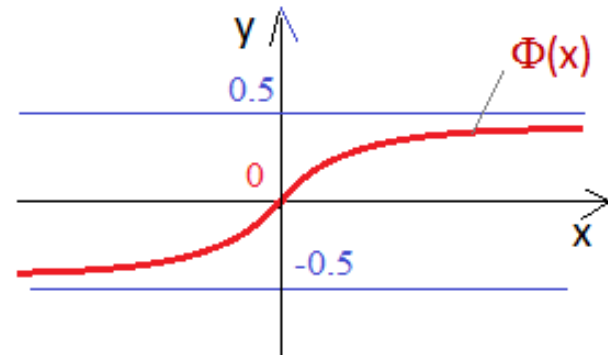
$$F(+\infty) = 1$$

- **Hàm hàm tích phân Laplace:**

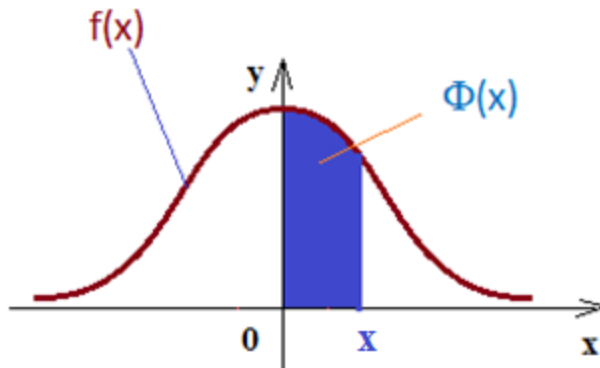
$$\Phi(x) = \int_0^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

- **Các tính chất:**

- **Hàm Φ lẻ**, $\Phi(-x) = -\Phi(x)$, $\forall x$
- $\Phi(x) = F(x) - 0,5$



- Φ đơn điệu tăng trên \mathbb{R} .
- $\Phi(-\infty) = -0,5$; $\Phi(+\infty) = 0,5$



$$\text{VD: } \Phi(1,24) = Q(1,24) \\ \approx 0,39251$$

$$\Phi(-1,24) = -Q(1,24) = -Q(-1,24) \\ \approx -0,39251$$

Ta có thể tìm giá trị $\Phi(x)$ theo các cách:

* Cách 1: Dùng bảng II để tìm gần đúng giá trị của hàm.

Ví dụ: $\Phi(1,24) \approx 0,39251$. $\Phi(-1,24) \approx -0,39251$

Tra dòng 1,2; cột 4; thêm kí hiệu 0,... trước kết quả tra.

* Cách 2: Theo công thức hàm $\Phi(x)$.

$$\int_0^{1,24} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \approx 0,39251$$

* Cách 3: Sử dụng MTBT CASIO fx 570 ES PLUS để tính $|\Phi(x)|$:

Vào chế độ thống kê 1 biến **MODE -- 3 (STAT) -- 1 (1-VAR)**

Nhấn phím **AC** để bỏ qua bước nhập số liệu.

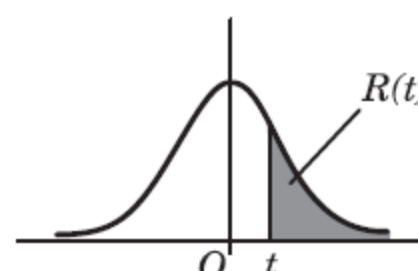
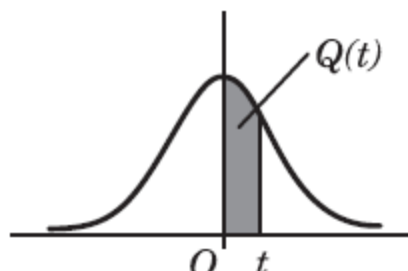
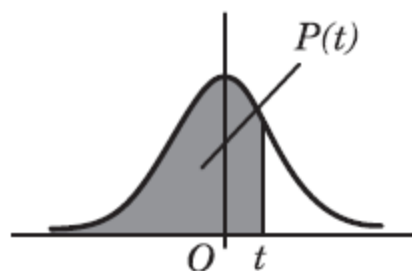
Bấm **SHIFT -- 1 (STAT) -- 5 (Distr) -- 2 (Q()) -- x (Nhập x) -- =**

Nếu $x > 0$ thì $\Phi(x) = Q(x) > 0$; nếu $x < 0$ thì $\Phi(x) = -Q(x) < 0$

Menu phụ Distr (SHIFT **1** STAT **5** (Distr))

1 P(**2** Q(**3** R(**4** **▶** t

Menu này giúp chúng ta tính các giá trị hàm phân phối chuẩn chuẩn tắc. Biến t được tính từ giá trị trung bình \bar{x} và độ lệch tiêu chuẩn σ_n nhận được từ dữ liệu ở màn hình nhập



$$X \triangleright t = \frac{X - \bar{x}}{\sigma_n}$$

$$P(t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

$$R(t) = \int_t^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

$$Q(t) = \left| \int_0^t \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \right|$$

IV.1.1.b Các tính chất của phân phối chuẩn:

* **Nếu $X \sim N(a, \sigma^2)$ thì:**

Cm:

$$P(\alpha < X < \beta)$$

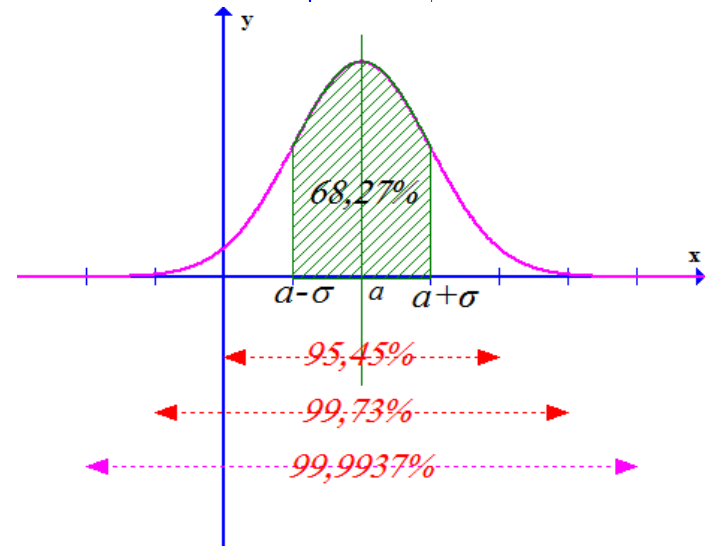
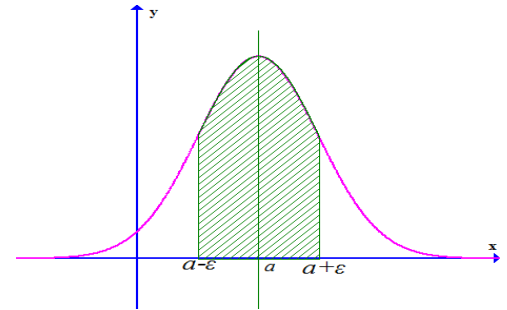
$$\begin{aligned} &= \int_{\alpha}^{\beta} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} dx \stackrel{db}{=} \int_{\frac{\alpha-a}{\sigma}}^{\frac{\beta-a}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy, \quad y = \frac{x-a}{\sigma} \\ &= \int_{\frac{\alpha-a}{\sigma}}^{\frac{\beta-a}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = - \int_{\frac{\alpha-a}{\sigma}}^{\frac{\beta-a}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \\ &= \Phi\left(\frac{\beta-a}{\sigma}\right) - \Phi\left(\frac{\alpha-a}{\sigma}\right) \stackrel{hay}{=} F\left(\frac{\beta-a}{\sigma}\right) - F\left(\frac{\alpha-a}{\sigma}\right) \end{aligned}$$

$$\begin{aligned} P(\alpha < X < \beta) &\stackrel{ct1}{=} \int_{\alpha}^{\beta} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} dx \\ &\stackrel{ct2}{=} \Phi\left(\frac{\beta-a}{\sigma}\right) - \Phi\left(\frac{\alpha-a}{\sigma}\right) \\ &\stackrel{ct3}{=} F\left(\frac{\beta-a}{\sigma}\right) - F\left(\frac{\alpha-a}{\sigma}\right) \end{aligned}$$

$$* P(|X - a| < \varepsilon) = 2 \cdot \Phi\left(\frac{\varepsilon}{\sigma}\right) \quad \varepsilon > 0.$$

+ Quy tắc **2- sigma**: $P(|X - a| < 2\sigma) \approx 95,45 \%$

+ Quy tắc **3- sigma**: $P(|X - a| < 3\sigma) \approx 99,73 \%$



Ví dụ :

Một công ty thực phẩm đang chuẩn bị đưa một loại bánh mới ra thị trường.

Người ta nhận thấy rằng số ngày sử dụng tốt tối đa của mỗi chiếc bánh trong điều kiện khuyến cáo là bnn có phân phối chuẩn mà trung bình

là 20 ngày và phương sai $1,3 \text{ ngày}^2$. (dùng kiểm định thống kê)

Công ty nên công bố thời hạn sử dụng của loại bánh này là bao nhiêu ngày?

$$(a - 3\sigma; a + 3\sigma) = (20 - 3 \times \sqrt{1,3}; 20 + 3 \times \sqrt{1,3}) \approx (16,6; 23,4)$$

* **Nếu $X \sim N(a_1, \sigma_1^2)$; $Y \sim N(a_2, \sigma_2^2)$ và X, Y độc lập**
thì biến nn $Z = aX + bY$ cũng tuân theo phân phối chuẩn
và $E(Z) = a.E(X) + b.E(Y)$; $D(Z) = a^2.D(X) + b^2.D(Y)$, $\forall a, b \in R$

* Phân phối chuẩn là 1 quy luật phân phối rất thường gặp vì có nhiều phân bố xác suất trong tự nhiên và trong thực tế đời sống có hình dáng khá giống phân phối chuẩn.

Trong công nghiệp, người ta đã xác định được rằng kích thước của các chi tiết do các nhà máy sản xuất ra sẽ có phân phối chuẩn nếu quá trình sản xuất diễn ra bình thường. Trong nông nghiệp, năng suất của cùng một loại cây trồng tại các thửa ruộng khác nhau cũng có phân phối chuẩn. .. Một số chỉ số về thể lực và trí tuệ con người cũng tuân theo phân phối chuẩn...

* Tham khảo phương pháp kiểm tra 1 tập dữ liệu có tuân theo pp chuẩn hay không, xem tài liệu (7), từ trang 163 – 169.

Ví dụ 1 Tìm giá trị k thích hợp trong các biểu thức xác định hàm mật độ XS của bnn X sau:

$$a) f(x) = k.e^{-\frac{(x-1)^2}{3}} \quad b) f(x) = k.e^{-x^2+4x} \quad c) f(x) = \begin{cases} 0 & x < 0 \\ ke^{-x^2} & x \geq 0 \end{cases}$$

Ví dụ 2 Một máy tự động cung cấp trung bình mỗi lần 200ml nước sạch vào một cốc nước. Giả sử lượng nước rót vào cốc tuân theo pp. chuẩn với độ lệch chuẩn 6 ml.

- a) Tìm xác suất nhận được 1 cốc nước chứa từ 195 đến 205 ml nước.
- b) Có khoảng bao nhiêu cốc nước có thể bị tràn nếu chúng ta sử dụng 1000 cốc có dung tích đựng 210 ml?
- c) Tìm một mức nước mà có 70% số lần máy cung cấp lượng nước ít hơn.



Hướng dẫn VD2:

Gọi X là lượng nước rót vào cốc trong 1 lần lấy nước ngẫu nhiên. $X \sim N(a=200 \text{ ml}; \sigma^2 = (6 \text{ ml})^2)$.

$$\begin{aligned} \text{a)} \quad P(195 < X < 205) &= \int_{195}^{205} \frac{1}{6 \times \sqrt{2\pi}} e^{-\frac{(x-200)^2}{2 \times 6^2}} dx \\ &= \Phi\left(\frac{205-200}{6}\right) - \Phi\left(\frac{195-200}{6}\right) = 0,59534 \end{aligned}$$

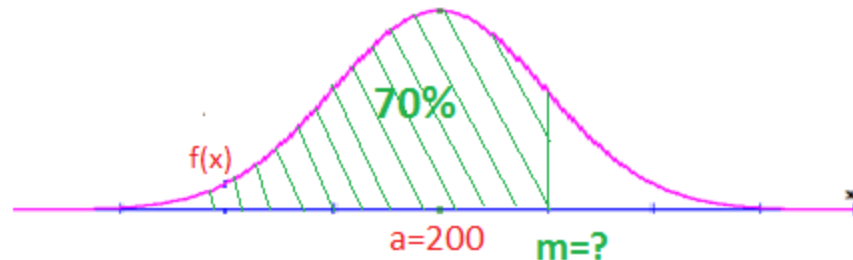
$$\begin{aligned} \text{b) XS 1 cốc nước bị tràn} &= P(X > 210) \\ &= \Phi(+\infty) - \Phi\left(\frac{210-200}{6}\right) = 0,5 - \Phi\left(\frac{210-200}{6}\right) = 0,04779 \end{aligned}$$

Dự đoán số cốc nước bị tràn trong 1000 cốc là :

$$1000 \times 0,04779 \approx 48 \text{ cốc.}$$

c) $X \sim N(a=200 \text{ ml}; \sigma^2 = (6 \text{ ml})^2)$.

Tìm một mức nước mà có 70% số lần máy cung cấp lượng nước ít hơn.



Gọi m là mức nước cần tìm.

$$P(X < m) = P(0 < X < m) = \Phi\left(\frac{m-200}{6}\right) - (-0,5) = 0,7$$

$$\Rightarrow \Phi\left(\frac{m-200}{6}\right) = 0,2 \quad \Rightarrow \frac{m-200}{6} = \Phi^{-1}(0,2) = 0,5244$$

$$\Rightarrow m = 203,1464 \text{ ml}$$

↓
Tra ngược bảng 2
hoặc bấm mò trong STAT
hoặc giải phương trình ...

Ví dụ 3

Một nhà sản xuất cần mua số lượng lớn 1 loại gioăng cao su có độ dày từ 0,118 cm đến 0,122 cm. Có 2 công ty cùng chào hàng loại gioăng này, độ dày của chúng là các biến ngẫu nhiên phân phối chuẩn với các số liệu sau:

	Độ dày trung bình	Độ lệch chuẩn	Giá bán/ hộp (1000 cái)
Công ty A	0,12	0,001	3 usd
Công ty B	0,12	0,0015	2,7 usd



Hỏi nhà sản xuất nên chọn mua gioăng của công ty nào để có lợi hơn về chi phí ?

Hướng dẫn:

Gọi X_A (cm) là độ dày của 1 gioăng cao su do công ty A sản xuất.

Gọi X_B (cm) là độ dày của 1 gioăng cao su do công ty B sản xuất.

Tỉ lệ gioăng sử dụng được cho nhà máy của mỗi công ty lần lượt như sau:

$$p_A = P(0,118 < X_A < 0,122) = \Phi\left(\frac{0,122 - 0,12}{0,001}\right) - \Phi\left(\frac{0,118 - 0,12}{0,001}\right) = 0,9545$$

$$p_B = P(0,118 < X_B < 0,122) = \Phi\left(\frac{0,122 - 0,12}{0,0015}\right) - \Phi\left(\frac{0,118 - 0,12}{0,0015}\right) = 0,817579$$

Dự đoán số gioăng trung bình đáp ứng được yêu cầu trong 1 hộp sản phẩm của 2 công ty A là $1000 \times p_A$; và $1000 \times p_B$.

Suy ra giá trung bình 1 cái gioăng dùng được của mỗi công ty:

$$t_A = 3 \text{usd} / 1000 \times p_A \approx 0,00314 \text{ usd}$$

$$t_B = 2,7 \text{usd} / 1000 \times p_B \approx 0,003302 \text{ usd}. \Rightarrow \text{Chọn công ty A.}$$

Ví dụ 4

Tuổi thọ của một loại sản phẩm là biến ngẫu nhiên phân phối chuẩn với trung bình 11 năm và độ lệch chuẩn là 2 năm.

a) Nếu quy định thời gian bảo hành sản phẩm là 10 năm thì tỉ lệ sản phẩm phải bảo hành là bao nhiêu?

b) Nếu muốn tỉ lệ sản phẩm phải bảo hành chỉ là 10% thì người ta cần quy định thời gian bảo hành là bao lâu ?

c) Nếu một sản phẩm đã hoạt động tốt qua thời gian bảo hành là 10 năm (câu a) thì xác suất nó vẫn hoạt động tốt trong 3 năm tiếp theo là bao nhiêu ?

Tuổi thọ sản phẩm trong bài này được quy ước là khoảng thời gian liên tục từ khi người dùng mua sản phẩm cho đến khi sản phẩm cần đem đến bảo hành hoặc sửa chữa.



Gọi X là tuổi thọ của sản phẩm loại này. $X \sim N(11, (2 \text{ năm})^2)$.

a) $P(X \leq 10) = P(0 \leq X \leq 10) =$

$$= \Phi\left(\frac{10-11}{2}\right) - \Phi\left(\frac{0-11}{2}\right) = \Phi(-0,5) + 0,5 = -0,1915 + 0,5 = 0,3085$$

b) Gọi A là thời hạn bảo hành cần tìm theo yêu cầu.

Từ giả thiết suy ra $P(X \leq A) = 10\%$, hay $\Phi\left(\frac{A-11}{2}\right) + 0,5 = 0,1$

$$\Rightarrow \Phi\left(\frac{A-11}{2}\right) = -0,4 = \Phi(-1,28) \Rightarrow \frac{A-11}{2} = -1,28 \quad (\text{do hàm } \Phi \text{ đơn điệu}).$$

Suy ra $A = 8,4$ năm.

c) A là b/c sản phẩm đã hoạt động tốt qua thời gian bảo hành,
B là biến cố sản phẩm vẫn hoạt động tốt trong 3 năm tiếp.

Xác suất cần tìm:

$$P(B/A) = \frac{P(AB)}{P(A)} = \frac{P(X > 13)}{P(X > 10)} = \frac{0,5 - \Phi\left(\frac{13-11}{2}\right)}{0,5 - \Phi\left(\frac{10-11}{2}\right)} = 0,2295$$

Ví dụ 5

Một nhà máy có 2 dây chuyền tự động đóng gói các bao xi măng. Trọng lượng các bao xi măng trên mỗi dây chuyền là biến ngẫu nhiên phân phối chuẩn, chúng có cùng kỳ vọng 50 kg , còn phương sai là $0,04 \text{ kg}^2$ trên dây chuyền A và $0,01 \text{ kg}^2$ trên dây chuyền B.

Giả thiết rằng trong kho của nhà máy chứa tỷ lệ bao xi măng được đóng gói từ dây chuyền A và B tương ứng 5:7 .

Hãy tìm tỉ lệ bao có trọng lượng đạt từ 49,95 kg trở lên trong kho.



Hướng dẫn

Gọi X_A ; X_B lần lượt là các biến ngẫu nhiên chỉ trọng lượng một bao xi măng được sản xuất trên các dây chuyền A và B.

$$X_A \sim N(a_1 = 50; \sigma_1^2 = 0,04)$$

$$X_B \sim N(a_2 = 50; \sigma_2^2 = 0,01)$$

Lấy ngẫu nhiên 1 bao xi măng từ trong kho của nhà máy.

H_A : b.cố bao xi măng đó được sản xuất trên dây chuyền A,

H_B : b.cố bao xi măng đó được sản xuất trên dây chuyền B.

Gọi F là biến cố bao xm có trọng lượng từ 49,95 kg trở lên.

Theo công thức xác suất đầy đủ:

$$\begin{aligned} P(F) &= P(H_A)P(F/H_A) + P(H_B)P(F/H_B) \\ &= P(H_A)P(X_A > 49,95) + P(H_B)P(X_B > 49,95) \\ &= \frac{5}{12} \times \left[0,5 - \Phi\left(\frac{49,95 - 50}{0,2}\right) \right] + \frac{7}{12} \times \left[0,5 - \Phi\left(\frac{49,95 - 50}{0,1}\right) \right] \approx 0,6528 \end{aligned}$$

Ví dụ 6

Giả sử rằng chiều cao (cm) của sinh viên nam khoa KHMT và khoa KTXD là các biến ngẫu nhiên độc lập, có dạng phân phối chuẩn tương ứng lần lượt là $N(\mu_1=171 ; \sigma_1^2= 4,8^2)$ và $N(\mu_2=170,5 ; \sigma_2^2= 5^2)$.

Chọn ngẫu nhiên một sinh viên nam khoa KHMT và một sinh viên nam khoa KTXD.

Tìm xác suất bạn sinh viên khoa KHMT có chiều cao cao hơn.



Hướng dẫn:

Gọi X là chiều cao của bạn sinh viên nam khoa KHMT.

Gọi Y là chiều cao của bạn sinh viên nam khoa KTXD.

Xác suất cần tìm là $P(X > Y) = P(X - Y > 0)$

Đặt $Z = X - Y \Rightarrow Z$ có phân phối chuẩn $N(\mu_1 - \mu_2; \sigma_1^2 + \sigma_2^2)$

Ví dụ 7

Một nhân viên nam từ thứ 2 đến thứ 6 hàng tuần đều rời khỏi nhà lúc 8g30 để đến văn phòng làm việc. Thời gian di chuyển đến văn phòng là biến ngẫu nhiên có phân phối chuẩn với trung bình 24 phút và độ lệch chuẩn 3,6 phút. Văn phòng làm việc từ 9g và cafe sáng tại văn phòng được phục vụ miễn phí từ 8g45' đến 8g55'.

- a) Hãy tính XS trong 1 tuần anh ta có ít nhất 1 ngày đi trễ?
- b) Giả sử anh nhân viên không bị trễ giờ làm, tìm xác suất anh ta bị trễ giờ cafe sáng.
- b) Có 1 chuyến tàu hàng ngày đi ngang qua văn phòng vào 1 giờ cố định. Anh nhân viên nói rằng trung bình 10 ngày chỉ có 4 ngày anh ta tới đủ sớm để có thể ngắm được đoàn tàu. Hãy dự đoán giờ mà con tàu chạy ngang qua văn phòng.



IV.1.2 Phân phối Bernoulli. (tham khảo)

Định nghĩa: BNN rời rạc **X** gọi là có **phân phối Bernoulli**, (*Bernoulli Distribution*), hay là *phân phối Không – Một*, ký hiệu $X \sim A(p)$ hay $X \sim B(1, p)$, nếu X có bảng phân phối XS như sau:

X	0	1
P_i	$1-p$	p

Tính chất:

- Nếu $X \sim B(1, p)$ thì $E(X) = p$ và $D(X) = pq$; $q=1-p$.
- Phân phối Bernoulli thường dùng để đặc trưng cho các dấu hiệu định tính có 2 phạm trù luân phiên như giới tính....
- Nếu $X_1; X_2; \dots; X_n \sim B(1, p)$ thì biến ngẫu nhiên $X = X_1 + X_2 + \dots + X_n$ có phân phối Nhị thức $B(n, p)$ (xem mục tiếp sau).

IV.1.3 Phân phối nhị thức:

Định nghĩa: BNN rời rạc X gọi là có **phân phối nhị thức** (*Binomial Distribution*), kí hiệu **$X \sim B(n, p)$** , với 2 tham số $n \in \mathbb{N}$; $p \in (0,1)$; ($q=1-p$) nếu X có bảng PPXS dạng:

X	0	1	\dots	k	\dots	n
P_i	q^n	$n \cdot p^1 \cdot q^{n-1}$	\dots	$C_n^k p^k q^{n-k}$	\dots	p^n

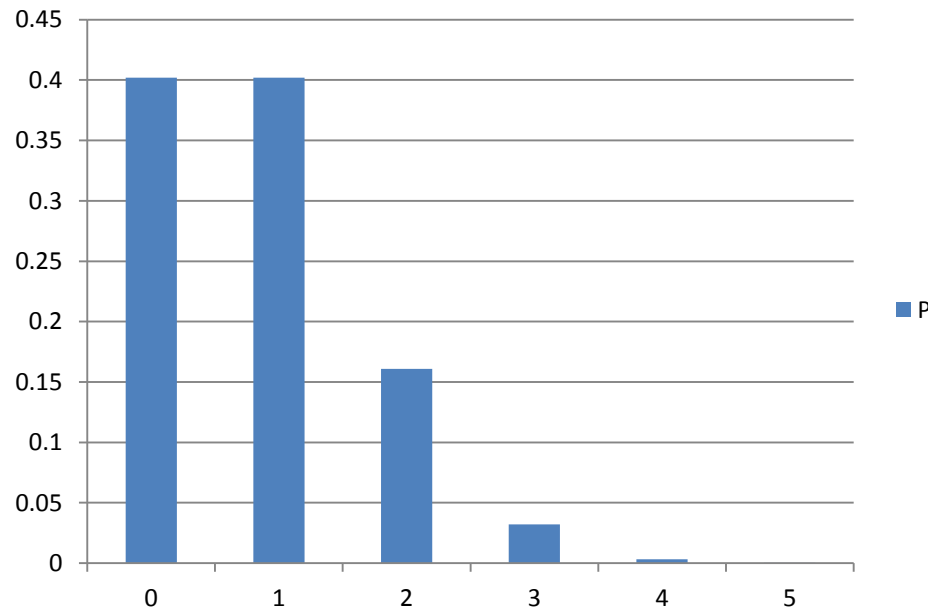
Tính chất:

- * Xét dãy n phép thử Bernoulli với xác suất “thành công” trong mỗi phép thử là p . Ký hiệu X là số lần “thành công” xuất hiện trong dãy n phép thử thì $X \sim B(n, p)$.
- * Nếu $X \sim B(n, p)$ thì $E(X) = np$ và $D(X) = npq$, với $q=1-p$.
- * $\text{Mod}(X)$ chính là số lần “thành công” có khả năng nhất.

Ví dụ 8₁

Gọi X là số lần nhận được mặt 6 chấm trong 5 lần tung một con xúc xắc. Lập bảng phân phối xác suất của X ; Tìm $E(X)$; $D(X)$.

X	0	1	2	3	4	5
P	0,4019	0,4019	0,1608	0,0322	0,0032	0,0001

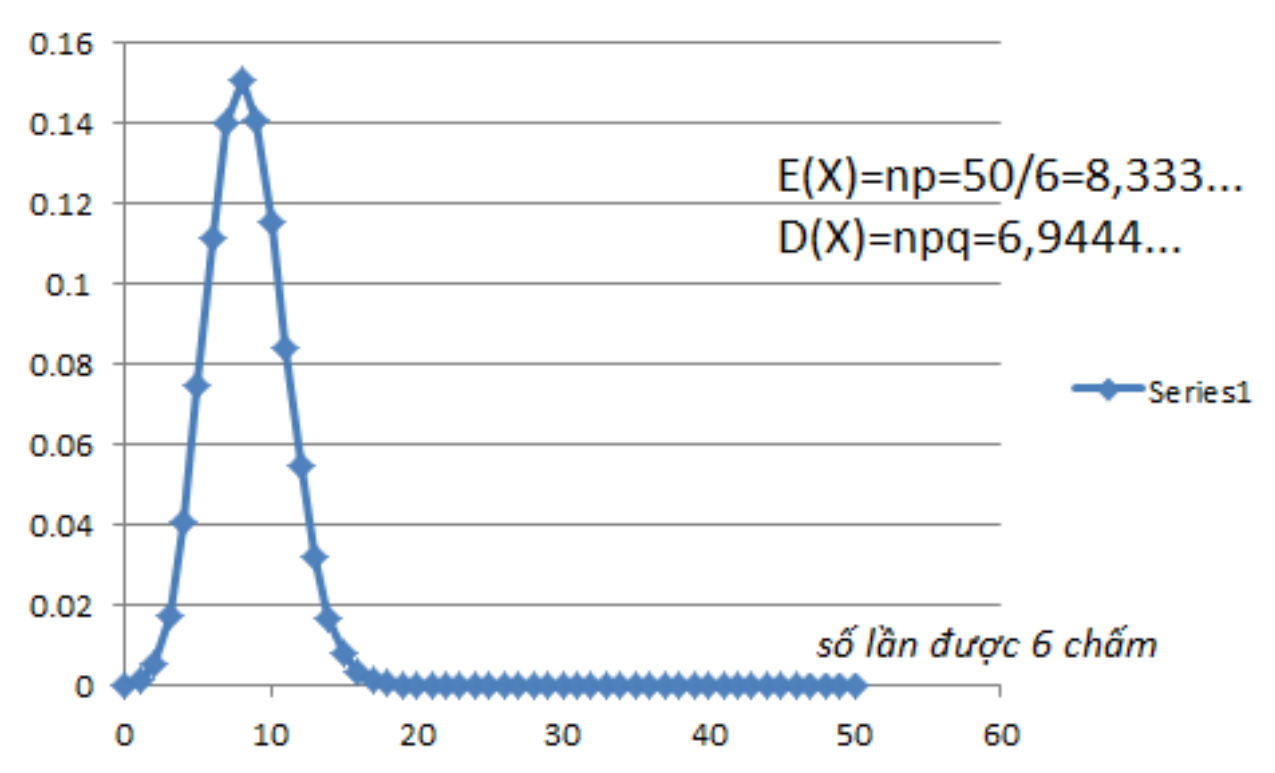


- Công thức tính p_i ?
- Kiểm tra lại $E(X)$; $D(X)$
theo công thức của
phân phối nhị thức?



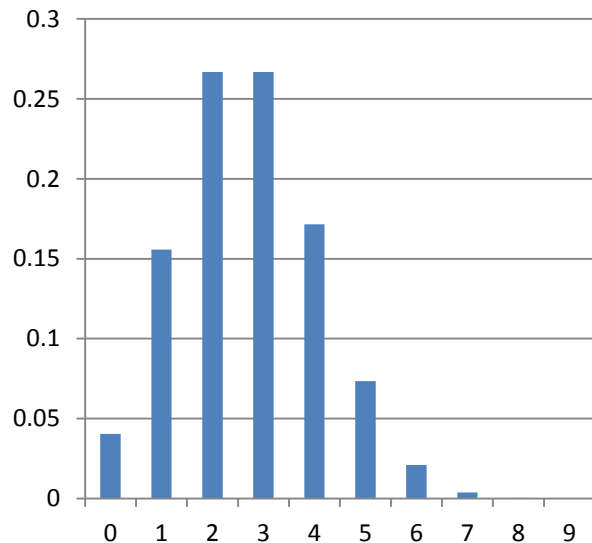
Ví dụ 8₂

Gọi X là số lần nhận được mặt 6 chấm trong 50 lần tung một con xúc xắc.

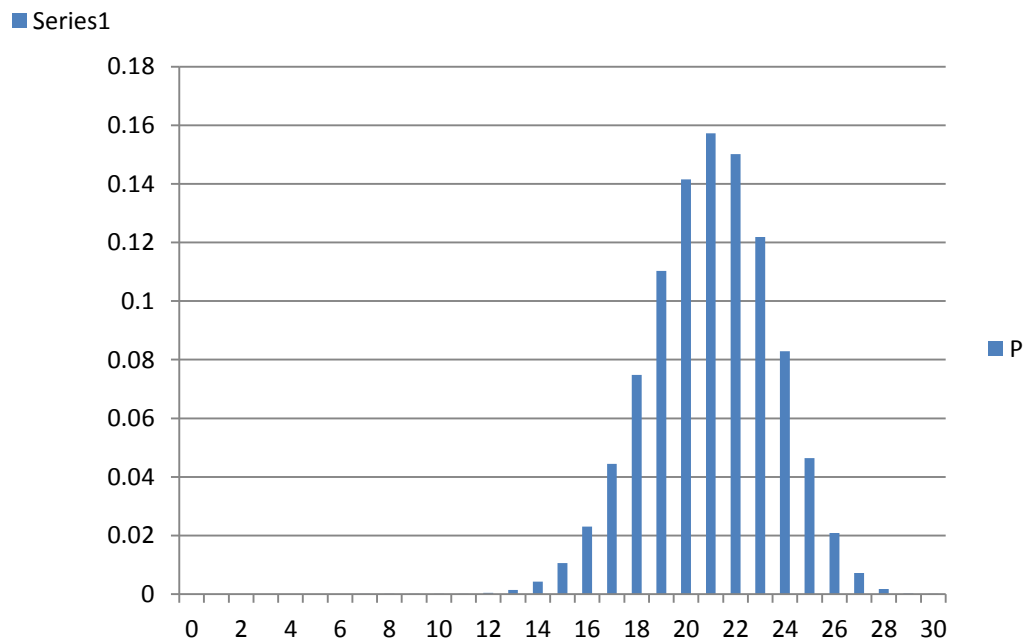


Minh họa biểu đồ phân phối nhị thức

$$X \sim B(n=9, p=0.3)$$



$$X \sim B(n=30, p=0.7)$$



TH1: Khi số n khá lớn, p không quá gần 0 hay quá gần 1, biến ngẫu nhiên X có *phân phối Nhị thức $B(n,p)$ được xem như xấp xỉ với phân phối Chuẩn $N(a=np, \sigma^2=npq)$* , cụ thể:

$$+ P(X = k) = C_n^k p^k q^{n-k} \approx \frac{1}{\sqrt{npq} \cdot \sqrt{2\pi}} e^{-\frac{(k-np)^2}{2npq}} \quad (1)$$

$$+ P(k_1 \leq X \leq k_2) = \sum_{k=k_1}^{k=k_2} C_n^k p^k q^{n-k} \approx \Phi\left(\frac{k_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{k_1 - np}{\sqrt{npq}}\right) \quad (2)$$

$$\Phi(x) = \int_0^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \quad \text{là hàm tích phân Laplace.}$$

TH2: Khi số n khá lớn và p rất gần với 0 hoặc 1, biến ngẫu nhiên X có *phân phối Nhị thức $B(n,p)$ được xem như xấp xỉ với phân phối Poisson $P(\lambda=np)$ (xem mục II.4.5)*, cụ thể:

$$+ P(X = k) = C_n^k p^k q^{n-k} \approx \frac{e^{-np} (np)^k}{k!}, \quad (3)$$

Một số lưu ý thêm:

- * Xấp xỉ ở công thức (2) là tốt nhất nếu n lớn và $np > 5$; $nq > 5$ hoặc $npq > 20$. Công thức (2) là kết quả định lý giới hạn Moivre – Laplace.
- * Công thức (2) có thể hiệu chỉnh để có kết quả tốt hơn bằng cách tăng k_2 lên nửa đơn vị và k_1 giảm đi nửa đơn vị, khi tính x_1, x_2 . Cụ thể:

$$\sum_{k=k_1}^{k=k_2} C_n^k p^k q^{n-k} \approx \Phi(x_2^*) - \Phi(x_1^*)$$
$$x_1^* = \frac{k_1 - \frac{1}{2} - np}{\sqrt{npq}} \qquad x_2^* = \frac{k_2 + \frac{1}{2} - np}{\sqrt{npq}}$$

- * Công thức (3) xấp xỉ tốt khi $n > 20$ và $p < 5\%$. (Kết quả xấp xỉ rất chính xác khi $n > 100$ và $\lambda = np < 10$). Vì p nhỏ nên người ta còn gọi nó là công thức của định luật số hiếm.

Ví dụ 9

Một tổng đài nội bộ của một cơ quan phục vụ 100 máy điện thoại, các máy gọi đến tổng đài là độc lập với nhau. Xác suất để trong một phút mỗi máy điện thoại gọi đến tổng đài là 0,02.



- a) Gọi X là BNN chỉ số máy điện thoại gọi đến tổng đài trong 1 phút. Hãy cho biết X có phân phối gì?
- b) Tìm số máy gọi đến tổng đài trung bình trong một phút.
- c) Tìm xác suất trong một phút có từ 3 đến 10 máy gọi đến tổng đài (*tính bằng các công thức xấp xỉ rồi so sánh với cách tính trực tiếp và nhận xét*).

Ví dụ 10 Tỷ lệ nảy mầm của một loại hạt giống là 75%.
Nếu gieo ngẫu nhiên 120 hạt giống thì xác suất có
được từ 80 hạt nảy mầm trở lên là bao nhiêu?



HD: VD 9 Đây là bài toán Bernoulli với $n=100$; $p=0,02$.

a) $X \sim B(n=100; p=0,02)$ b) $E(X) = np = 2$

c) * Tính trực tiếp: $\sum_{k=3}^{10} C_{100}^k \times 0,02^k 0,98^{100-k} \approx 0,3233087$

* Xấp xỉ pp Chuẩn: $\Phi\left(\frac{10 - 100 \times 0,02}{\sqrt{100 \times 0,02 \times 0,98}}\right) - \Phi\left(\frac{3 - 100 \times 0,02}{\sqrt{100 \times 0,02 \times 0,98}}\right) \approx 0,237525$

* Xấp xỉ pp Poisson: $\sum_{k=3}^{10} e^{-2} \frac{2^k}{k!} \approx 0,3233153$; $\lambda = 100 \times 0,02 = 2$

HD: VD 10

$$\sum_{k=80}^{120} C_{120}^k (0.75)^k (0.25)^{120-k} \approx \Phi\left(\frac{120 - 90}{\sqrt{90 \times 0,25}}\right) - \Phi\left(\frac{80 - 90}{\sqrt{90 \times 0,25}}\right) \approx 98,249\%$$

Ví dụ 11

Xác suất 1 sản phẩm không được kiểm tra chất lượng là 8%.

a) Tính xác suất trong 900 sản phẩm có 70 sản phẩm không được kiểm tra chất lượng từ nhà máy .

b) Tính xác suất trong 9000 sản phẩm có từ 700 đến 800 sản phẩm không được kiểm tra chất lượng.



Ví dụ 12

Xác suất 1 sản phẩm không được kiểm tra chất lượng là 2%.

a) Tính xác suất trong 900 sản phẩm có 20 sản phẩm không được kiểm tra chất lượng.

b) Tính xác suất trong 900 sản phẩm có tối đa 20 sản phẩm không được kiểm tra chất lượng.

HD VD 11 *Xấp xỉ phân phối chuẩn $N(a=np; \sigma^2=npq)$*

$$a) C_{900}^{70} (0,08)^{70} (0,92)^{830} \approx \frac{1}{\sqrt{900 \times 0,08 \times 0,92}} \times \frac{1}{\sqrt{2\pi}} e^{-\frac{(70-900 \times 0,08)^2}{2 \times 900 \times 0,08 \times 0,92}} \approx 0,122868 \times 0,38708 \approx 0,0476$$

$$b) \Phi\left(\frac{800 - 9000 \times 0,08}{\sqrt{9000 \times 0,08 \times 0,92}}\right) - \Phi\left(\frac{700 - 9000 \times 0,08}{\sqrt{9000 \times 0,08 \times 0,92}}\right) \approx \Phi(3,1083) - \Phi(-0,7771) = 0,7805$$

HD VD 12 *Xấp xỉ phân phối Poisson $P(\lambda=np)$*

Gọi X là số sản phẩm không được kiểm tra trong 900 sản phẩm.

a) $n=900$; $p=0,02$; $k=20$; $np=18$.

$$P(X=20) = \frac{e^{-18} \times 18^{20}}{20!} = 0,0798$$

b) $n=900$; $p=0,02$; $k_1=0$; $k_2=20$.

$$P(0 \leq X \leq 20) = \sum_{k=0}^{k=20} \frac{e^{-18} \times 18^k}{k!} = 0,7307$$

Ví dụ 13

Trọng lượng sản phẩm X do một máy tự động sản xuất là một BNN tuân theo qui luật chuẩn với $E(X) = 100$ g và độ lệch chuẩn 1g. Sản phẩm được coi là đạt tiêu chuẩn kỹ thuật nếu trọng lượng của nó nằm từ 99 đến 101,2 g .

a) Tìm tỉ lệ sản phẩm đạt tiêu chuẩn do máy sản xuất.

b) Cho máy sản xuất liên tiếp 200 sản phẩm. Tìm XS có được ít nhất 100 sản phẩm đạt tiêu chuẩn.

c) Giả sử chiều dài Y (cm) cũng của sản phẩm này là BNN có phân phối chuẩn với kỳ vọng 20 cm và phương sai là $0,2 \text{ cm}^2$. Tìm tỉ lệ sản phẩm loại 1, là những sản phẩm đạt chuẩn và có độ dài trong khoảng $20\text{cm} \pm 0,2 \text{ cm}$. (giả thiết thêm là X,Y độc lập)



Hướng dẫn

$$a) P(99 < X < 101,2) = \Phi\left(\frac{101,2 - 100}{1}\right) - \Phi\left(\frac{99 - 100}{1}\right) = 0,38493 + 0,34134 = 72,63\%$$

b) Gọi Y là số sản phẩm đạt tiêu chuẩn trong 200 sản phẩm.

$Y \sim B(n=200; p=0,7263) \sim N(np, npq)$. Xác suất cần tìm:

$$P(100 \leq Y \leq 200) \approx \Phi\left(\frac{200 - 200 \times 0,7263}{\sqrt{200 \times 0,7263 \times 0,2737}}\right) - \Phi\left(\frac{100 - 200 \times 0,7263}{\sqrt{200 \times 0,7263 \times 0,2737}}\right) = 0,5 + 0,5 = 1$$

c) Xác suất cần tìm:

$$= P(\text{SP đạt chuẩn trọng lượng}) \times P(\text{SP đạt chuẩn độ dài})$$

$$= 72,63\% \times P(|Y - 20| < 0,2)$$

$$= 72,63\% \times 2 \times \Phi\left(\frac{0,2}{\sqrt{0,2}}\right) \approx 72,63\% \times 2 \times 0,17264 \approx 0,2508$$

IV.1.4 Phân phối siêu bội.

Định nghĩa: BNN X gọi là có phân phối Siêu bội (*Hypergeometric Distribution*), kí hiệu $X \sim H(N, M, n)$, với tham số là các số tự nhiên n, N, M , $0 < n \leq M \leq N$; $0 < n \leq N - M \leq N$; nếu bảng PPXS của X có dạng:

X	0	1	\dots	k	\dots	n
P_i	$\frac{C_{N-M}^n}{C_N^n}$	$\frac{C_M^1 C_{N-M}^{n-1}}{C_N^n}$	\dots	$\frac{C_M^k C_{N-M}^{n-k}}{C_N^n}$	\dots	$\frac{C_M^n}{C_N^n}$

- Định nghĩa tổng quát với điều kiện tham số: $0 < n; M \leq N$:

$$P(X=k) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}; \quad \max\{0; n + M - N\} \leq k \leq \min\{n; M\}$$

Tính chất:

* Nếu $X \sim H(N, M, n)$ thì $E(X) = np$
 và $D(X) = npq \frac{N-n}{N-1}$ với $p = \frac{M}{N}$ và $q = 1-p$.

Ví dụ 14

Một lô hàng có $N = 50$ bóng đèn, trong đó lẫn 10 bóng hỏng. Lấy ngẫu nhiên 5 bóng để kiểm tra.

Gọi X là BNN chỉ số bóng hỏng trong 5 bóng được lấy ra.

a) Tính $P(X=2)$ và lập bảng PPXS của X . X có phân phối gì ?

b) Tính số bóng hỏng trung bình trong các bóng được lấy ra và phương sai của X .

a)

X	0	1	2	3	4	5
P	$\frac{C_{10}^0 \times C_{40}^5}{C_{50}^5}$	$\frac{C_{10}^1 \times C_{40}^4}{C_{50}^5}$	$\frac{C_{10}^2 \times C_{40}^3}{C_{50}^5}$	$\frac{C_{10}^3 \times C_{40}^2}{C_{50}^5}$	$\frac{C_{10}^4 \times C_{40}^1}{C_{50}^5}$	$\frac{C_{10}^5 \times C_{40}^0}{C_{50}^5}$

c)
$$E(X) = np = 5 \times \frac{10}{50} = 1$$

$$D(X) = npq \frac{N-n}{N-1} = 5 \times \frac{10}{50} \times \frac{40}{50} \times \frac{50-5}{50-1} = \frac{36}{49}$$



Lưu ý : Nếu đổi giả thiết là “Một lô hàng có 50 bóng đèn, mỗi bóng đèn có tỉ lệ hỏng là 20% .. \Rightarrow **X có PP Nhị thức.**

- * Khi $n \ll N$ (n rất nhỏ so với N) thì người ta nhận thấy *phân phối Siêu bội xấp xỉ với phân phối Nhị thức $B(n,p)$.*

VD \Rightarrow Nếu đổi giả thiết $N=500$, ta thấy $n=5 \ll N=500$ nên

$$P(X=2) = \frac{{}^H C_{10}^2 C_{490}^3}{C_{500}^5} = 0.003436 \approx {}^{B(n,p)} C_5^2 \left(\frac{10}{500} \right)^2 \left(\frac{490}{500} \right)^3 = 0.003765$$

Quy luật siêu bội để tính xác suất xuất hiện k lần kết quả khi lấy ngẫu nhiên n đơn vị từ một tập hợp nào đó theo phương thức không hoàn lại. Khi $n \ll N$ thì việc xấp xỉ phân phối siêu bội bởi phân phối nhị thức cho phép ta tính toán như trong trường hợp lấy theo phương thức có hoàn lại.

IV.4.5 Phân phối Hình học (tham khảo)

Định nghĩa: BNN rời rạc X gọi là có **phân phối Hình học** (*Geometric Distribution*), kí hiệu $X \sim G(p)$ với tham số $p \in (0;1)$ nếu bảng phân phối xác suất của X có dạng:

X	1	2	3	...	k	...
P_i	p	$q^1.p$	$q^2.p$...	$(1-p)^{k-1}.p$...

Tính chất:

* Nếu $X \sim P(\lambda)$ thì:

$$E(X) = \frac{1}{p}; \quad D(X) = \frac{1}{p^2} - \frac{1}{p}$$

* X được coi là số thử nghiệm Bernoulli cần thiết để có được sự thành công.

* Trong Ví dụ 2, BNN Y có phân phối hình học.

Ví dụ 15:

Một băng chuyền tự động sản xuất từng sản phẩm với xác suất phế phẩm là p và được dừng ngay để điều chỉnh khi xuất hiện *một* phế phẩm. Tìm kỳ vọng của số sản phẩm được sản xuất giữa 2 lần điều chỉnh kề nhau.

ĐS: $1/p$.



Ví dụ 16:

Hai cầu thủ lần luân phiên ném bóng vào rổ cho đến khi có người ném lọt mới thôi. Xác suất ném trúng rổ trong mỗi lần ném của người thứ nhất là a và của người thứ 2 là b . Tìm số lần ném trung bình của mỗi người và tìm phương sai của số lần mỗi người ném.

IV.4.6 Phân phối Poisson

Định nghĩa: Biến ngẫu nhiên rời rạc X gọi là có **phân phối Poisson** (*Poisson Distribution*), kí hiệu $X \sim P(\lambda)$; với tham số $\lambda > 0$, nếu bảng phân phối XS của X có dạng:

X	0	1	2	\dots	k	\dots
P_i	$e^{-\lambda}$	$e^{-\lambda} \cdot \lambda$	$\frac{e^{-\lambda} \cdot \lambda^2}{2!}$	\dots	$\frac{e^{-\lambda} \lambda^k}{k!}$	\dots

Tính chất:

$$X \sim P(\lambda) \text{ thì } E(X) = D(X) = \lambda$$

* $\lambda - 1 \leq \text{Mode } X \leq \lambda$; $\text{Mode } X \in$

Phân phối Poisson cho ta biết số kết quả sẽ xảy ra (hay gọi là số lần *thành công*) trong một độ dài thời gian quy ước hay là trong 1 độ đo của không gian.

Có nhiều BNN tuân theo luật phân phối Poisson, chẳng hạn như: số người vào 1 trạm phục vụ công cộng trong mỗi một giờ; số lượt truy cập vào trang web Bkel trong mỗi phút; số cuộc gọi khẩn cấp tới tổng đài 115 trong 30 giây; số lỗi trên mỗi trang của một cuốn tiểu thuyết; số hoa nở trong ngày của một cây hoa; số lỗi trên 1 lốp xe cao su; số hạt nảy mầm trên 1 dm^2 ; số lượng ngôi sao trong 1 thể tích không gian vũ trụ...

Có 1 số đặc điểm của dạng phân phối Poisson:

- Số lượng kết quả xảy ra trong các khoảng thời gian khác nhau cùng độ dài (hay trong các không gian khác nhau cùng độ đo) là độc lập.
- Xác suất một kết quả đơn lẻ sẽ xảy ra trong một khoảng

thời gian rất ngắn (hay 1 không gian rất nhỏ) tương ứng với độ dài khoảng thời gian (hay độ đo của không gian) và không phụ thuộc vào số kết quả xuất hiện bên ngoài khoảng thời gian (hay không gian) này.

- Xác suất có nhiều hơn 1 kết quả xuất hiện trong 1 khoảng thời gian rất ngắn (hay 1 miền rất nhỏ) như vậy là không đáng kể.

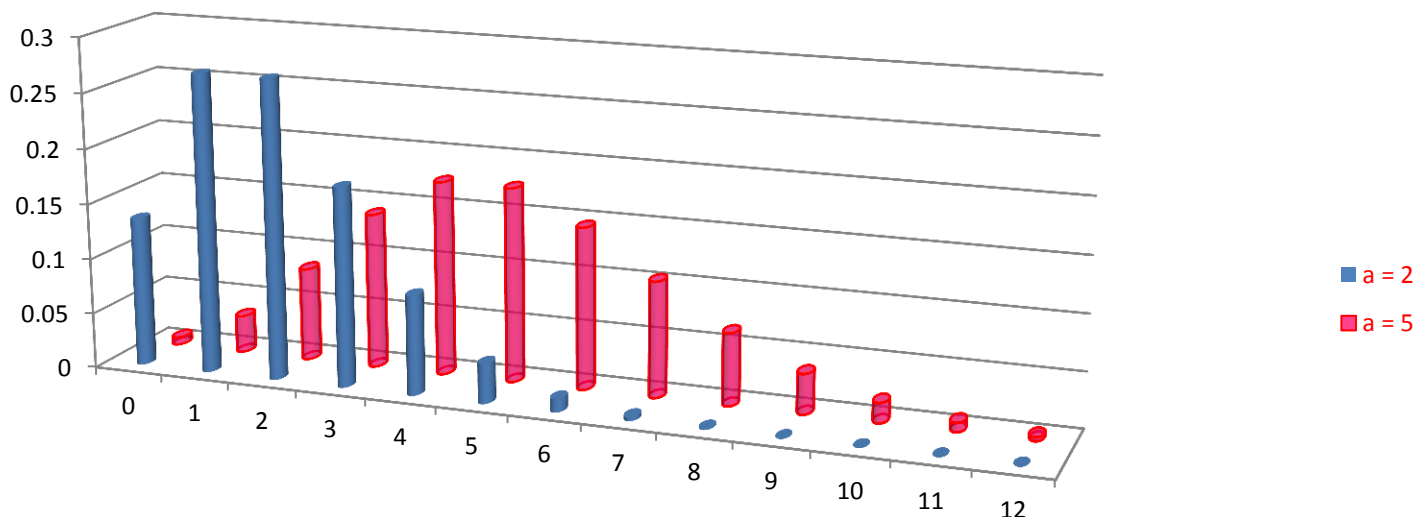
Sự hội tụ của về quy luật chuẩn:

Khi $\lambda > 20$; phân phối Poisson hội tụ về phân phối chuẩn.

\Rightarrow nếu $X \sim P(\lambda)$ và lớn, $\lambda > 20$ thì ta coi như $X \sim N(a = \lambda; \sigma^2 = \lambda)$.

Dưới đây là biểu đồ minh họa cho phân phối Poisson, khi λ (hay a)=2 ; và khi $\lambda = 5$.

k=	0	1	2	3	4	5	6	7	8	9	10	11	12
$\lambda = 2$	0.1353	0.2707	0.2707	0.1804	0.0902	0.0361	0.0120	0.0034	0.0009	0.0002	3.82E-05	6.944E-06	1.157E-06
$\lambda = 5$	0.0067	0.0337	0.0842	0.1404	0.1755	0.1755	0.1462	0.1044	0.0653	0.0363	0.0181	0.0082	0.0034



Ví dụ 17

Biết rằng số người vào một đại lý bưu điện trong một khoảng thời gian xác định là một BNN có phân phối Poisson.

Quan sát thấy cứ 5 phút có 12 người ghé vào một ĐLBĐ.

Tìm xác suất trong một phút có 3 người vào đại lý bưu điện đó?



Hướng dẫn: Gọi X là số người vào một đại lý bưu điện trong khoảng thời gian 1 phút. Khi đó $X \sim P(\lambda)$.

Từ giả thiết ta thấy số người trung bình vào ĐLBĐ này trong 1 phút là $12/5 = 2,4 \Rightarrow E(X) = 2,4 \Rightarrow \lambda = 2,4$.

X	0	1	2	3	4	k	...
P	...	$\frac{e^{-2,4} \cdot 2,4}{1!}$...	?	$\frac{e^{-2,4} \cdot 2,4^k}{k!}$...

$$\text{Vậy xác suất cần tìm là } P(X=3) = e^{-2,4} \frac{2,4^3}{3!} \approx 0,2090$$

Ví dụ 18

Ở một bến cảng, trung bình một ngày có 10 tàu chở dầu cập bến. Năng lực của cảng chỉ có thể đáp ứng được tối đa 15 tàu trong một ngày. Tìm xác suất trong một ngày có tàu phải quay trở ra?

Hướng dẫn:

Gọi X là số tàu chở dầu cập cảng trong 1 ngày.

Do $E(X) = 10$ nên $X \sim P(\lambda=10)$.

Vậy xác suất cần tìm là $P(X > 15) = 1 - P(0 \leq X \leq 15)$

$$= 1 - \sum_{k=0}^{15} \frac{e^{-10} \times 10^k}{k!} = 1 - 0,9513 = 4,87\%$$



Ví dụ 19 (tham khảo)

Giả sử một siêu thị nọ có trung bình 5 khách ra quầy tính tiền mỗi 2 phút (và mỗi khách thường tính tiền xong trong khoảng 2 phút).

Câu hỏi đặt ra là siêu thị nên lập bao nhiêu quầy tính tiền, để có thể cân đối giữa việc tiết kiệm chi phí với việc làm hài lòng khách hàng?

Gọi X là số khách đến quầy tính tiền trong mỗi khoảng thời gian 2 phút thì $X \sim P(\lambda=5)$.

Tham khảo:

<https://www.ecoblader.com/2016/11/bom-roi-o-london-phan-phoi-poisson-va-vu-khi-toi-thuong-cua-dan-kinh-doanh/>

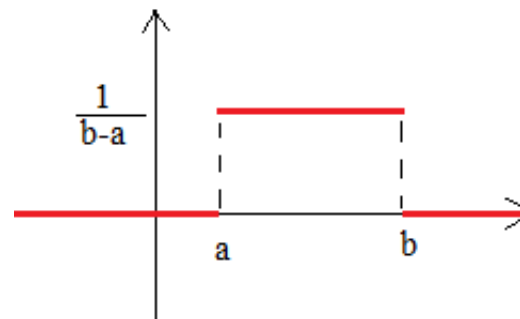


II.4.7 Phân phối đều

a) Phân phối đều liên tục:

Định nghĩa: Biến ngẫu nhiên X được gọi là có **phân phối đều** (*Uniform Distribution*) trên đoạn $[a, b]$, kí hiệu **$X \sim U(a, b)$** nếu hàm mật độ của nó có dạng:

$$f(x) = \begin{cases} \frac{1}{b-a} & x \in [a, b] \\ 0 & x \notin [a, b] \end{cases}$$



Tính chất:

- Nếu $X \sim U(a, b)$ thì **$E(X) = \frac{a+b}{2}$** và **$D(X) = \frac{(b-a)^2}{12}$**

b) Phân phối đều rời rạc: Nếu X chỉ nhận n giá trị với xác suất như nhau thì ta nói X có phân phối đều rời rạc.

- Quy ước: nếu ta không biết gì về quy luật PPXS của BNN X thì coi như mỗi giá trị có thể có của X là đồng khả năng. Dẫn đến việc coi X như là một BNN tuân theo quy luật phân phối đều.

Ví dụ 20

Xe buýt đầu tiên qua trạm lúc 5g00 và cứ 15 phút có 1 chuyến đi qua. Nếu An tới trạm vào một thời điểm nn trong khoảng 5g00 - 5g30 thì XS An phải đợi buýt trên 10 phút là bao nhiêu?

HD: Gọi X là thời điểm A có mặt ở bến xe buýt.
Coi như $X \sim U [5g00; 5g30]$, X có hàm mật độ:

$$f(x) = \begin{cases} \frac{1}{30} & x \in [5g00; 5g30] \\ 0 & x \notin [5g00; 5g30] \end{cases}$$

XS cần tìm = $P(5g00 < X < 5g05) + P(5g15 < X < 5g20) = 1/3$.



Ví dụ 21

Một đoạn thẳng AB dài 12 cm được chia thành 2 đoạn bởi một điểm M lấy ngẫu nhiên trên AB. Người ta dùng 2 đoạn AM và BM để làm 2 cạnh của 1 hình chữ nhật.

- a) Tính diện tích trung bình của hình chữ nhật đó.
- b) Tìm xác suất hình chữ nhật đó có diện tích lớn hơn 27 cm^2 .



Hướng dẫn:

Gọi X là độ dài đoạn $AM \Rightarrow X$ có phân phối đều trên $(0; 12)$.

X có hàm mật độ xác suất là:

$$f(x) = \begin{cases} \frac{1}{12} & x \in [0, 12] \\ 0 & x \notin [0, 12] \end{cases}$$

a) Diện tích hình chữ nhật $S = X(12-X) = 12X - X^2$.

Diện tích trung bình $E(S) = E(12X - X^2)$

$$E(12X - X^2) = \int_{-\infty}^{+\infty} (12x - x^2) \times f(x) dx = \int_0^{12} (12x - x^2) \times \frac{1}{12} dx = 24$$

b) $P(S > 27) = P(12X - X^2 > 27) = P(X^2 - 12X + 27 < 0) = P(3 < X < 9)$

$$= \int_3^9 \frac{1}{12} dx = \frac{1}{2}$$

IV.4.8 Phân phối mũ.

Định nghĩa: Biến nn liên tục X được gọi là có **phân phối mũ** hay **pp. lũy thừa** (*Exponential Distribution*) với tham số $\lambda > 0$, kí hiệu $X \sim E(\lambda)$, nếu hàm mật độ của nó có dạng :

$$f(x) = \begin{cases} 0 & x < 0 \\ \lambda e^{-\lambda x} & x \geq 0 \end{cases}$$

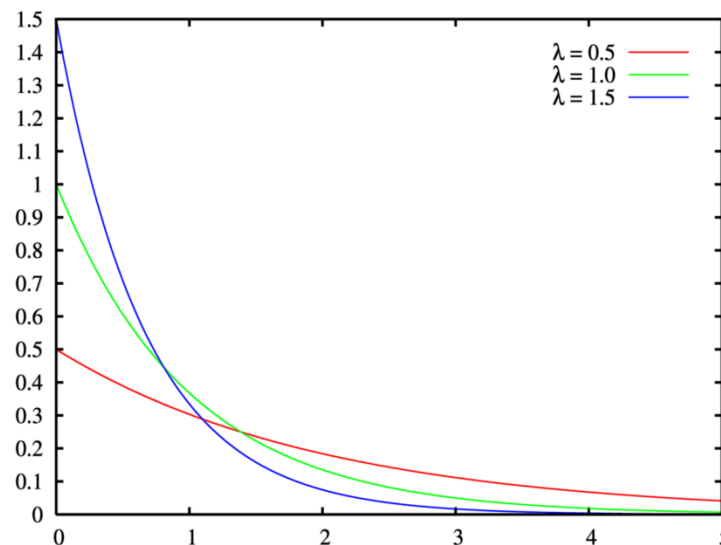
Tính chất:

* Nếu $X \sim E(\lambda)$ thì:

$$E(X) = \frac{1}{\lambda} \quad \text{và} \quad D(X) = \frac{1}{\lambda^2}$$

* Hàm phân phối XS của X :

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$$



- * Phân phối mũ có ứng dụng trong nhiều lĩnh vực, nó thường có mặt trong các quy trình sản xuất và cung cấp dịch vụ. Trong hệ thống kỹ thuật, thời gian làm việc liên tục của máy móc thiết bị giữa 2 lần sửa chữa cũng tuân theo phân pp. mũ.
- * Thời gian chờ đợi giữa 2 lần thành công kế tiếp nhau trong phân phối Poisson là biến ngẫu nhiên có phân phối mũ.
 VD: Trong 1 xưởng thuê, người ta thấy trung bình cứ 10 phút thì có 3 lần máy phải dừng để thay chỉ. Số lần dừng X để thay chỉ trong 10 phút là bnn có pp. Poison. Khoảng thời gian Y giữa 2 lần thay chỉ liên nhau là bnn có pp. mũ với trung bình $10/3$ phút $\Rightarrow \lambda = 3/10$.
- * Người ta nói phân phối mũ có tính chất không nhớ, tức là:

$$P(X > k) = P(X > k+a \mid X > a), \forall k > 0; a > 0$$
 VD: nếu chúng ta ngẫu nhiên bước vào xưởng thì thời gian ta nhìn thấy lần thay chỉ tiếp theo cũng là biến ngẫu nhiên Y , không phụ thuộc vào việc thay chỉ lần ngay trước đó xảy ra cách bao lâu.

Ví dụ 22 Thời gian chờ được phục vụ của một khách hàng ở một cửa hàng là 1 BNN X có hàm mật độ xác suất sau:

$$\varphi(x) = \begin{cases} Ae^{-5x} & \text{khi } x > 0 \\ 0 & \text{khi } x \leq 0 \end{cases} \quad (\text{đơn vị } x: \text{phút})$$



- Tìm A và xác suất một khách hàng phải chờ trong khoảng từ 0,4 phút đến 1 phút.
- Tìm thời gian trung bình khách chờ và phương sai của X .
- Tìm hàm phân phối xác suất của X .

HD: a) Để thấy $X \sim E(\lambda = 5) \Rightarrow A = 5$; $P(0,4 < X < 1) = \int_{0,4}^1 5e^{-5x} dx \approx 0,1286$

b) Thay vào công thức: $E(X) = \frac{1}{\lambda} = \frac{1}{5}$ $D(X) = \frac{1}{\lambda^2} = \frac{1}{25}$

c) $F_X(x) = \int_{-\infty}^x \varphi(x) dx = \begin{cases} 0 & x < 0 \\ 1 - e^{-5x} & x \geq 0 \end{cases}$

Ví dụ 23 Tuổi thọ của 1 loại linh kiện (đơn vị x : giờ) là 1 biến ngẫu nhiên X có hàm mật độ xác suất sau:

$$f(x) = \begin{cases} ke^{-\frac{x}{3000}} & \text{khi } x > 0 \\ 0 & \text{khi } x \leq 0 \end{cases}$$



- a) Tìm k và tuổi thọ trung bình của linh kiện đó.
- b) Một nhà máy mua 800 linh kiện loại này để sử dụng. Tính xác suất ít nhất phân nửa số linh kiện đó có thời gian hoạt động vượt qua tuổi thọ trung bình của loại linh kiện.

Hướng dẫn:

- a) $X \sim E(\lambda = 1/3000) \Rightarrow k = 1/3000$ và $E(X) = 3000$.
- b) Gọi Y là số linh kiện ở nhà máy có thời gian hoạt động vượt quá 3000 giờ.

$$Y \sim B(n = 800; p = P(X > 3000)),$$

$$\text{ở đây } p = 1 - \int_0^{3000} \frac{1}{3000} e^{-\frac{x}{3000}} dx \approx 0,3679$$

Yêu cầu bài toán dẫn đến việc tính $P(400 \leq Y \leq 800)$

$$\text{Công thức tính trực tiếp: } \sum_{k=400}^{800} C_{800}^k (0,3679)^k (1-0,3679)^{800-k}$$

Do n lớn nên ta tính gần đúng biểu thức trên bằng cách xem như Y xấp xỉ phân phối chuẩn $N(a=np; \sigma^2 = npq)$.

$$\text{Cụ thể : } a = np = 294,3036$$

$$\sigma = \sqrt{npq} = \sqrt{186,03533} \approx 13,6395$$

Dùng công thức tính xấp xỉ (công thức 2, mục PP nhị thức):

$$P(400 \leq Y \leq 800) = \Phi\left(\frac{800-a}{\sigma}\right) - \Phi\left(\frac{400-a}{\sigma}\right) = 0,5 - \Phi\left(\frac{400-294,3036}{13,6395}\right) = 0$$

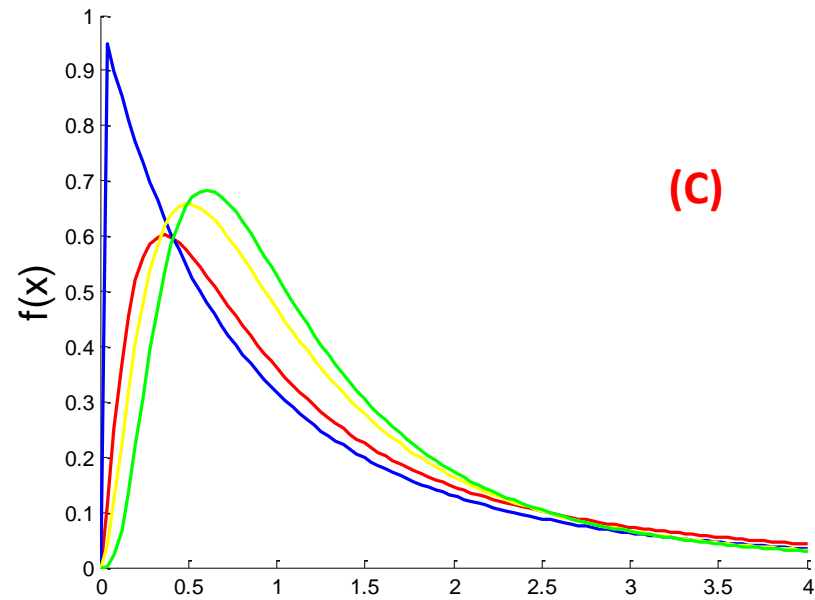
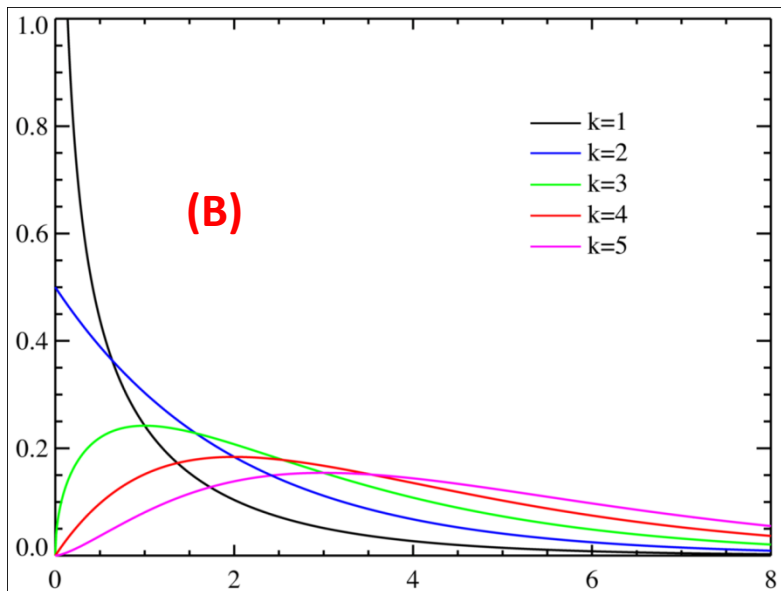
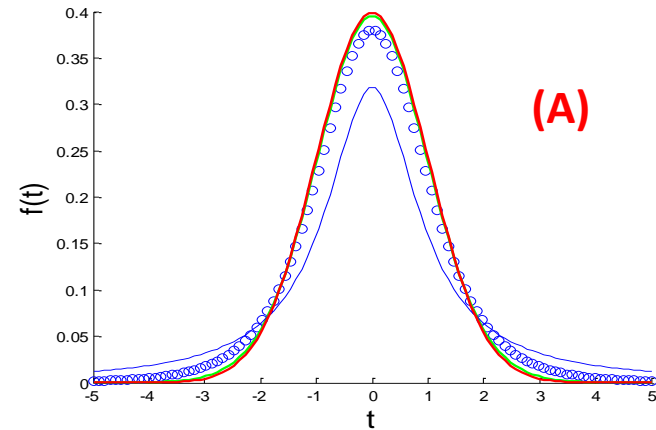
SV đọc giáo trình để tìm hiểu thêm về các dạng phân phối sau:

IV.4.9 Phân phối Student (A)

Khi $n \geq 30$, PP Student coi như xấp xỉ PP Chuẩn chuẩn tắc.

IV.4.10 Phân phối Chi bình phương (B) (hay Khi bình phương)

IV.4.11 Phân phối Fisher (C)



Trích từ giáo trình...

4.2.4 Phân phối “Chi bình phương”

Đại lượng ngẫu nhiên \mathcal{X}^2 gọi là có phân phối “Chi bình phương” n bậc tự do nếu

$$\mathcal{X}^2 = X_1^2 + X_2^2 + \dots + X_n^2$$

trong đó X_1, X_2, \dots, X_n là các đại lượng ngẫu nhiên độc lập có phân phối chuẩn chuẩn tắc.

Trong trường hợp này ta ký hiệu: $\mathcal{X}^2 \sim \mathcal{X}^2(n)$

Ký hiệu $\Gamma(x)$ là hàm gama $\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt$

Định lý 4.11 Cho $X \sim \mathcal{X}^2(n)$. Khi đó

$$(i) \text{ Hàm mật độ của } \mathcal{X}^2 \text{ là } f_n(x) = \begin{cases} \frac{e^{-\frac{x}{2}} x^{\frac{n}{2}-1}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} & \text{nếu } x > 0 \\ 0 & \text{nếu } x \leq 0 \end{cases}$$

$$(ii) \quad E(\mathcal{X}^2) = n; \quad D(\mathcal{X}^2) = 2n$$

4.2.5 Phân phối Student

Đại lượng ngẫu nhiên T gọi là có phân phối Student n bậc tự do nếu $T = \frac{U}{\sqrt{V/n}}$

trong đó: $U \sim N(0, 1)$ và $V \sim \mathcal{X}^2(n)$.

Trong trường hợp này ta ký hiệu: $T \sim T(n)$.

Định lý 4.12 Cho $T \sim T(n)$. Khi đó

(i) Hàm mật độ của T là
$$f_n(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi} \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

(ii) $E(T) = 0$; $D(T) = \frac{n}{n-2}$

IV.2 CÁC ĐỊNH LÝ GIỚI HẠN

II.5.1 Bất đẳng thức Chebyshev:

Nếu X là biến ngẫu nhiên có kỳ vọng toán và phương sai hữu hạn thì với mọi số dương ε tùy ý ta luôn có:

$$P(|X - E(X)| < \varepsilon) \geq 1 - \frac{D(X)}{\varepsilon^2}$$

Ví dụ 24:

Thu nhập trung bình hàng năm của dân cư ở 1 vùng là 18 triệu và độ lệch chuẩn là 3,2 triệu đồng. Hãy tìm một khoảng thu nhập hàng năm xung quanh giá trị trung bình của ít nhất 95% cư dân vùng đó.

HD: Gọi X là BNN chỉ mức thu nhập hàng năm của dân cư trong vùng. Ta chưa biết phân phối xác suất của X chỉ biết $E(X)=18$ triệu, $D(X)= (3,2 \text{ triệu đồng})^2$. Do vậy kết quả tìm được sau đây là tương đối.

Theo bất Chebyshev thì : $P(|X - 18| < \varepsilon) \geq 1 - \frac{3,2^2}{\varepsilon^2}$

Cho vế phải $=0,95 \Rightarrow \varepsilon =14,3108$.

Khoảng cần tìm (18 -14,3108; 18+14,3108).

II.5.2 Định lý Chebyshev:

- Nếu các BNN X_1, X_2, \dots, X_n độc lập từng đôi, có các kỳ vọng toán hữu hạn và các phương sai đều bị chặn trên bởi hằng số C thì với mọi số dương ε tùy ý ta luôn có:

$$\lim_{n \rightarrow +\infty} P\left(\left|\frac{X_1 + X_2 + \dots + X_n}{n} - \frac{E(X_1) + E(X_2) + \dots + E(X_n)}{n}\right| < \varepsilon\right) = 1$$

- Trường hợp riêng, nếu các BNN X_1, X_2, \dots, X_n độc lập từng đôi, có cùng kỳ vọng $E(X_i) = m, i=1,2,\dots,n$; và các phương sai cùng bị chặn trên thì với mọi số dương ε tùy ý ta luôn có:

$$\lim_{n \rightarrow +\infty} P\left(\left|\frac{X_1 + X_2 + \dots + X_n}{n} - m\right| < \varepsilon\right) = 1$$

Định lý này còn gọi là luật số lớn của Chebyshev.

- Định lý Chebyshev chứng minh sự hội tụ theo xác suất của trung bình số học của 1 số lớn BNN về trung bình số học của các kỳ vọng toán tương ứng, mặc dù từng BNN độc lập có thể nhận giá trị khác nhiều so với kỳ vọng toán của chúng.
- Định lý Chebyshev có nhiều ứng dụng rộng rãi trong các lĩnh vực khác nhau. Ví dụ:
 - + Trong việc đo lường 1 đại lượng vật lý, người ta thường tiến hành đo nhiều lần và lấy trung bình số học của các kết quả đo làm giá trị thực của đại lượng cần đo. Định lý Chebyshev chỉ ra rằng trung bình số học của các kết quả đo sẽ sai lệch rất ít so với giá trị thực của đại lượng cần đo.
 - + Định lý Chebyshev cho phép dự đoán giá trị của trung bình số học các BNN. Nó là cơ sở của phương pháp mẫu trong thống kê: dựa vào mẫu ngẫu nhiên khá nhỏ có thể kết luận về toàn bộ tập hợp tổng quát của các đối tượng được nghiên cứu.

II.5.3 Định lý Bernoulli:

Nếu $f_n = \frac{m}{n}$ là tần suất xuất hiện biến cố A trong n phép thử độc lập và p là xác suất xuất hiện biến cố đó trong mỗi phép thử thì với mọi số dương ε tùy ý ta luôn có:

$$\lim_{n \rightarrow +\infty} P(|f_n - p| < \varepsilon) = 1$$

- Định lý này còn gọi là luật số lớn của Bernoulli, được xem là cơ sở toán học của định nghĩa xác suất theo thống kê.
- Ở đây sự hội tụ theo xác suất của tần suất $f = m/n \rightarrow p$ khác với sự hội tụ theo nghĩa giải tích cổ điển. Theo nghĩa giải tích, với $\varepsilon > 0$ cho trước, luôn tồn tại số tự nhiên N để với mọi $n > N$ thì $|m/n - p| < \varepsilon$. Sự hội tụ hiểu theo nghĩa xác suất ở chỗ dù n lớn bao nhiêu đi nữa thì vẫn có thể xảy ra trường hợp cá biệt mà biểu thức $|m/n - p| < \varepsilon$ không được thỏa mãn.

II.5.4 Định lý Giới hạn trung tâm:

(Trường hợp riêng)

Giả sử X_1, X_2, \dots, X_n là các BNN độc lập cùng tuân theo một quy luật phân phối xác suất nào đó. Kí hiệu $E(X_i) = a$ và $D(X_i) = \sigma^2$, $\forall i$. Khi $n \rightarrow \infty$, chúng ta có sự hội tụ theo xs của các BNN sau:

a) Bnn $Z = X_1 + X_2 + \dots + X_n$ hội tụ về phân phối chuẩn $N(n.a, n.\sigma^2)$

b) Bnn $\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$ hội tụ về phân phối chuẩn $N(a, \frac{\sigma^2}{n})$,

hay Bnn $U = \frac{\bar{X} - a}{\frac{\sigma}{\sqrt{n}}}$ hội tụ về phân phối chuẩn tắc $N(0,1)$.

Nói cách khác, khi $n \rightarrow \infty$, ta có: $P(U < x) \approx \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$

Trong phần thống kê, khi $n > 30$ ta sử dụng công thức xấp xỉ này.

Ví dụ 25:

Tung 1 con xúc xắc 200 lần. Tính xác suất tổng số chấm thu được trong các lần tung nhận giá trị từ 300 đến 650.

Gọi X_i là số chấm xuất hiện trên con xúc xắc ở lần tung thứ i , $i=1,2,\dots,200$. Các biến ngẫu nhiên X_i độc lập.

Ta tính được $E(X_i) = a = 3,5$ và $D(X_i) = \sigma^2 \approx 2,9167$.

Đặt $Z = X_1 + X_2 + \dots + X_{200}$.

$E(Z) = 200 \times 3,5 = 700$; $D(Z) = 200 \times 2,9167 = 583,3333$.

Theo định lý giới hạn trung tâm: $Z \sim N(700; 583,3333)$

Do đó xác suất cần tìm:

$$P(300 \leq Z \leq 650) \approx \Phi\left(\frac{650-700}{\sqrt{583,3333}}\right) - \Phi\left(\frac{300-700}{\sqrt{583,3333}}\right) = 1,92\%$$

Ví dụ 26:

Chọn ngẫu nhiên 500 số trong đoạn $[1; 2]$. Tính xác suất giá trị trung bình của các số đó nằm trong khoảng $(1,45; 1,55)$.

Gọi bnn X_i là giá trị số thứ i được chọn, $i=1,2,\dots,500$.

Ta xem như các bnn X_i có phân phối đều liên tục trên $[1; 2]$.

Các X_i độc lập và $E(X_i) = a = 1,5$; $D(X_i) = \sigma^2 = 1/12 \approx 0,0833$.

$$\text{Đặt: } \bar{X} = \frac{X_1 + X_2 + \dots + X_{500}}{500}$$

Theo định lý giới hạn trung tâm: $\bar{X} \sim N\left(1,5; \frac{0,08333}{500} = 0,000167\right)$

Do đó xác suất cần tìm:

$$P(1,45 \leq \bar{X} \leq 1,55) \approx \Phi\left(\frac{1,55 - 1,5}{\sqrt{0,000167}}\right) - \Phi\left(\frac{1,45 - 1,5}{\sqrt{0,000167}}\right) = 99,99\%$$

Ví dụ 27:

Tung 1 đồng xu 1000 lần.

Tìm xác suất để độ lệch giữa tần số xuất hiện mặt sấp và xác suất xuất hiện của mặt sấp bé hơn 0,1.

Hướng dẫn:

Gọi X là số lần được mặt sấp trong 1000 lần tung thì

$$X \sim B(n=1000, p=0,5)$$

Khi đó tần suất xuất hiện mặt sấp trong n lần tung là $X/1000$.

Xác suất cần tìm là $P(|X/1000 - 1/2| < 0.1)$

Biến đổi dẫn đến biểu thức tương đương $P(400 < X < 600)$
và dùng công thức xấp xỉ chuẩn cho X .