Biến Ngẫu Nhiên (tiếp)

Khoa Công nghệ Thông tin và Truyền thông

CMC UNIVERSITY

Nội dung

Hợp phân bố của nhiều Biến ngẫu nhiên

Hợp phân bố

Định nghĩa 1

Cho X và Y là hai biến ngẫu nhiên.

(a) *Nếu X và Y đều rời rạc, thì*

$$p(x, y) = \mathrm{P}(X = x, Y = y)$$

được gọi là hàm hợp khối lượng xác suất (joint pmf) của X và Y.

(b) *Nếu X và Y đều liên tục, thì $f(x, y)$ được gọi là hàm hợp mật độ xác suất (joint pdf) của X và Y nếu*

$$\mathrm{P}(a \le X \le b, c \le Y \le d) = \int_c^d \int_a^b f(x, y)dxdy.$$

Ví dụ đơn giản

Gieo hai con xúc sắc cân. Gọi X, Y lần lượt là số chấm trên mặt của hai con xúc xắc này. Ta có bảng hợp phân bố xác suất của hai biến ngẫu nhiên X, Y như sau

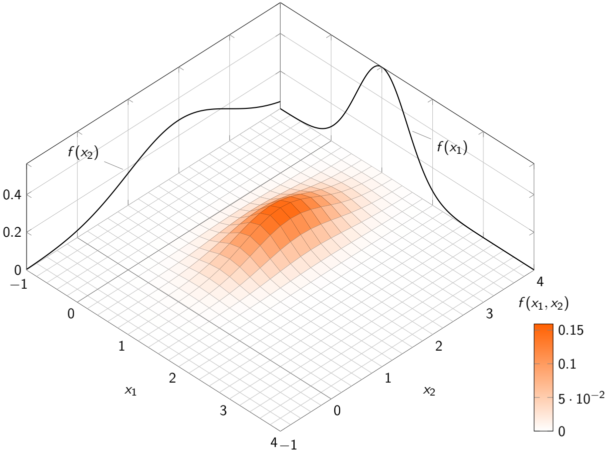
X \ Y	1	2	3	4	5	6
1	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$
2	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$
3	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$
4	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$
5	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$
6	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$	$1/36$

Quiz

Quiz

Trong ba-lô có 3 quả bóng đỏ, 2 quả bóng xanh, và 1 quả bóng trắng. Chọn ngẫu nhiên đồng thời ra hai quả. Gọi X là số bóng đỏ, Y là số bóng xanh trong hai quả bóng đã lấy ra. Lập bảng phân bố hợp xác suất của hai biến X và Y.

Ví dụ minh họa



Hình 1: Hợp phân bố của hai biến ngẫu nhiên Gauß.

Tính chất

Mệnh đề 1

Nếu X, Y là hai biến ngẫu nhiên có hàm hợp xác suất $p(x, y)$ (X, Y rời rạc) hoặc $f(x, y)$ (X, Y liên tục), thì

(a) $p(x, y), f(x, y) \geq 0$ *với mọi* x, y.

(b) *Nếu X, Y rời rạc, thì* $\sum_{x,y} p(x, y) = 1$. *Nếu X, Y liên tục, thì*

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)dxdy = 1.$$

Ví dụ

Ví dụ 1

Cho X, Y có hàm hợp mật độ

$$f(x, y) = cx, \quad 0 < x, y < 1.$$

(a) *Tìm c.*

(b) *Tính* $\mathrm{P}(X > 0.5)$.

Ví dụ

Ví dụ 1

Cho X, Y *có hàm hợp mật độ*

$$f(x,y) = cx, \quad 0 < x, y < 1.$$

(a) *Tìm* c.

(b) *Tính* $\mathrm{P}(X > 0.5)$.

Ví dụ 2

Cho X, Y *có hàm hợp mật độ*

$$f(x,y) = ce^{-(x+y)/5}, \quad 0 < x, y < \infty.$$

(a) *Tìm* c.

(b) *Tính* $\mathrm{P}(X < 5, Y < 5)$.

(c) *Tính* $\mathrm{P}(X < Y)$.

Hàm khối lượng/mật độ biên

Định nghĩa 2

Cho X, Y là hai biến ngẫu nhiên có hàm hợp xác suất $p(x, y)$ hoặc $f(x, y)$.

(a) *Nếu X, Y rời rạc, thì* **hàm khối lượng biên** *(marginal pmf) của X, ký hiệu là $p_X(x)$, được định nghĩa bởi*

$$p_X(x) = \sum_y p(x, y).$$

(b) *Nếu X, Y liên tục, thì* **hàm mật độ biên** *(marginal pdf) của X, ký hiệu là $f_X(x)$, được định nghĩa bởi*

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy.$$

Ví dụ

Ví dụ 3

Một chương trình gồm có hai khối chương trình con. Số lỗi X ở khối thứ nhất và số lỗi Y ở khối thứ hai có hợp phân bố cho bởi

$$\begin{aligned} p(0,0) = p(0,1) = p(1,0) &= 0.2, \\ p(1,1) = p(1,2) = p(1,3) &= 0.1, \\ p(0,2) = p(0,3) &= 0.05. \end{aligned}$$

(a) *Hãy xác định các hàm khối lượng $p_X(x)$ và $p_Y(y)$.*

(b) *Xác suất để không có lỗi ở khối thứ nhất.*

(c) *Xác định phân bố của tổng số lỗi ở cả hai khối.*

Ví dụ (tiếp)

Ví dụ 4

Cho hàm hợp mật độ của X và Y xác định bởi

$$f(x,y) = \frac{1}{y} e^{-\left(y + \frac{x}{y}\right)}, \quad 0 < x, y < \infty.$$

(a) *Hãy kiểm tra rằng $f(x,y)$ thực sự xác định một hàm mật độ.*

(b) *Xác định các hàm $f_X(x), f_Y(y)$.*

Ví dụ (tiếp)

Ví dụ 4

Cho hàm hợp mật độ của X và Y xác định bởi

$$f(x,y) = \frac{1}{y}e^{-\left(y+\frac{x}{y}\right)}, \quad 0 < x, y < \infty.$$

(a) *Hãy kiểm tra rằng $f(x,y)$ thực sự xác định một hàm mật độ.*

(b) *Xác định các hàm $f_X(x), f_Y(y)$.*

Ví dụ 5

Hai biến X, Y có hàm hợp mật độ
$f(x,y) = e^{-(x+y)}, 0 < x < y < \infty$. *Tính $f_X(x), f_Y(y)$.*

Xác suất điều kiện

Định nghĩa 3

Cho X, Y là hai biến ngẫu nhiên.

(a) *Nếu X, Y rời rạc, thì hàm khối lượng điều kiện của X biết rằng $Y = y$ được cho bởi*

$$p_{X|Y}(x|y) = \frac{p(x,y)}{p_Y(y)}.$$

(b) *Nếu X, Y liên tục, thì hàm mật độ điều kiện của X biết rằng $Y = y$ được cho bởi*

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}.$$

Ví dụ minh họa

Xét hàm phân bố $f(x, y) = x + y, \quad 0 < x, y < 1$. Để xác định hàm mật độ điều kiện của X với $Y = y_1 \in (0, 1)$, ta thực hiện

- Tìm hàm mật độ biên $f_Y(y)$. Ta có

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)dx = \int_0^1 (x + y)dx = y + \frac{1}{2}.$$

- Thay vào công thức trong định nghĩa ta được

$$f_{X|Y}(x|y_1) = \frac{x + y_1}{0.5 + y_1}.$$

Ví dụ

Ví dụ 6

Hai con xúc xắc cân được gieo. Gọi X là số mặt một chấm, Y là số mặt hai chấm.

1. *Lập bảng hợp phân bố xác suất của X và Y.*
2. *Xác định phân bố của X với điều kiện $Y = 1$.*

Ví dụ

Ví dụ 6

Hai con xúc xắc cân được gieo. Gọi X là số mặt một chấm, Y là số mặt hai chấm.

1. *Lập bảng hợp phân bố xác suất của X và Y.*
2. *Xác định phân bố của X với điều kiện $Y = 1$.*

Ví dụ 7

Cho hai biến rời rạc X, Y có hàm hợp khối lượng $p(x, y) = \frac{x+y+1}{c}$ với $x, y \in \{0, 1, 2\}$.

(a) *Tìm c.*

(b) *Tính $\mathrm{P}(X \leq 1, Y = 1)$.*

(c) *Tính $\mathrm{P}(Y = 1)$.*

(d) *Tính $\mathrm{P}(X \leq 1 | Y = 1)$.*

Ví dụ (tiếp)

Ví dụ 8

Cho X, Y có hàm hợp mật độ $f(x, y) = 6(1 - y), 0 < x < y < 1$. Xác định hàm mật độ điều kiện $f_{Y|X}(y|x)$.

Biến ngẫu nhiên Độc lập

Định nghĩa 4

Cho hai biến ngẫu nhiên X, Y. Ký hiệu các biến cố $E_a = \{X \le a\}$ và $F_b = \{Y \le b\}$. Chúng ta nói hai biến ngẫu nhiên X và Y là độc lập với nhau nếu các biến cố E_a và F_b độc lập với nhau với mọi cặp số thực (a, b). Nói cách khác,

$$\mathrm{P}(X \le a, Y \le b) = \mathrm{P}(X \le a) \cdot \mathrm{P}(Y \le b).$$

Tính chất

Mệnh đề 2

Cho hai biến ngẫu nhiên X, Y.

(a) *Nếu X, Y rời rạc thì chúng độc lập khi và chỉ khi*

$$p(x, y) = p_X(x) \cdot p_Y(y).$$

(b) *Nếu X, Y liên tục thì chúng độc lập khi và chỉ khi*

$$f(x, y) = f_X(x) \cdot f_Y(y).$$

Quiz

Quiz

Hai biến ngẫu nhiên X, Y có hàm hợp mật độ

$$f(x, y) = e^{-(x+y)}, \quad 0 < x, y < \infty.$$

X, Y có độc lập không?

Quiz

Quiz

Hai biến ngẫu nhiên X, Y *có hàm hợp mật độ*

$$f(x, y) = e^{-(x+y)}, \quad 0 < x, y < \infty.$$

X, Y *có độc lập không?*

Quiz

Hai biến ngẫu nhiên X, Y *có hàm hợp mật độ*

$$f(x, y) = e^{-(x+y)}, \quad 0 < x < y < \infty.$$

X, Y *có độc lập không?*

Kỳ vọng

Định nghĩa 5

Cho X, Y là hai biến ngẫu nhiên có hàm hợp xác suất $p(x, y)$ hoặc $f(x, y)$, và $g(x, y)$ là một hàm thực.

(a) *Nếu $\sum_{x,y} |g(x, y)|p(x, y) < \infty$, thì giá trị kỳ vọng của $g(X, Y)$ được cho bởi*

$$E\left[g(X, Y)\right] = \sum_{x,y} g(x, y)p(x, y).$$

- *Nếu $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |g(x, y)|f(x, y)dxdy < \infty$, thì giá trị kỳ vọng của $g(X, Y)$ được cho bởi*

$$E\left[g(X, Y)\right] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)f(x, y)dxdy.$$

Ví dụ

Ví dụ 9

Cho X, Y là hai biến Bernoulli với tham số $p_X = 0.2$ và $p_Y = 0.5$.

(a) *Xác định phân bố của $X + Y$.*

(b) *Tính $E[X + Y]$.*

Ví dụ 10

Cho $f(x, y) = 3x$, với $0 \leq y \leq x \leq 1$.

(a) *Tính $E[4X - 3Y]$.*

(b) *Tính $E[XY]$.*

Tính chất

Mệnh đề 3

Cho X, Y là hai biến ngẫu nhiên, a, b là hai hằng số.

(a) $E[aX + bY] = aE[X] + bE[Y]$.

(b) *Nếu X, Y là hai biến độc lập, thì*

$$E[X \cdot Y] = E[X] \cdot E[Y], \quad \mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).$$

Tính chất

Mệnh đề 3

Cho X, Y là hai biến ngẫu nhiên, a, b là hai hằng số.

(a) $E[aX + bY] = aE[X] + bE[Y]$.

(b) *Nếu X, Y là hai biến độc lập, thì*

$$E[X \cdot Y] = E[X] \cdot E[Y], \quad \mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).$$

Hệ quả

Nếu $X_1, \ldots, X_n$ là các biến ngẫu nhiên độc lập, có cùng kỳ vọng μ và phương sai σ^2, thì

(a) $E\left[\frac{X_1+\cdots+X_n}{n}\right] = \mu$.

(b) $\mathrm{Var}\left(\frac{X_1+\cdots+X_n}{n}\right) = \frac{\sigma^2}{n}$.

Áp dụng

Bài tập

Cho $U_1, \ldots, U_{100}$ *là các biến ngẫu nhiên độc lập, cùng có phân bố đều trên* $[-0.5, 0.5]$. *Xét biến ngẫu nhiên*

$$U = \frac{U_1 + \cdots + U_{100}}{100}.$$

Tính $E[U]$ *và* $\mathrm{Var}(U)$.

Kỳ vọng có điều kiện

Định nghĩa 6

Cho X, Y là hai biến ngẫu nhiên có hàm hợp xác suất $p(x, y)$ hoặc $f(x, y)$, và $g(x)$ là một hàm thực của x.

(a) *Nếu X, Y rời rạc, thì kỳ vọng có điều kiện của $g(X)$ với điều kiện $Y = y$ được cho bởi*

$$E\left[g(X)|Y = y\right] = \sum_x g(x)p_{X|Y}(x|y).$$

(b) *Nếu X, Y liên tục, thì kỳ vọng có điều kiện của $g(X)$ với điều kiện $Y = y$ được cho bởi*

$$E\left[g(X)|Y = y\right] = \int_{-\infty}^{\infty} g(x)f_{X|Y}(x|y)dx.$$

Ví dụ

Ví dụ 11

Cho X, Y là hai biến ngẫu nhiên có hàm hợp mật độ

$$f(x,y) = \begin{cases} x^2 + \frac{xy}{3}, & 0 \leq x \leq 1, \quad 0 \leq y \leq 2, \\ 0, & (x,y) \notin [0,1] \times [0,2]. \end{cases}$$

Tính kỳ vọng có điều kiện $E[X|Y = 0.5]$.

Hợp phương sai

Định nghĩa 7 (Hợp phương sai - Covariance)

Cho X và Y là hai biến ngẫu nhiên. Khi đó, hợp phương sai của X và Y, ký hiệu là $\mathrm{Cov}(X, Y)$, là một đại lượng cho bởi công thức

$$\mathrm{Cov}(X, Y) = E\big[(X - E[X])(Y - E[Y])\big].$$

Hợp phương sai

Định nghĩa 7 (Hợp phương sai - Covariance)

Cho X và Y là hai biến ngẫu nhiên. Khi đó, hợp phương sai của X và Y, ký hiệu là $\mathrm{Cov}(X, Y)$, là một đại lượng cho bởi công thức

$$\mathrm{Cov}(X, Y) = E\left[(X - E[X])(Y - E[Y])\right].$$

Mệnh đề 4

$$\mathrm{Cov}(X, Y) = E[XY] - E[X] \cdot E[Y].$$

Hợp phương sai

Định nghĩa 7 (Hợp phương sai - Covariance)

Cho X và Y là hai biến ngẫu nhiên. Khi đó, hợp phương sai của X và Y, ký hiệu là $\mathrm{Cov}(X, Y)$, là một đại lượng cho bởi công thức

$$\mathrm{Cov}(X, Y) = E\big[(X - E[X])(Y - E[Y])\big].$$

Mệnh đề 4

$$\mathrm{Cov}(X, Y) = E[XY] - E[X] \cdot E[Y].$$

Chú ý

Phương sai là trường hợp riêng của hợp phương sai.

Hệ số tương quan

Định nghĩa 8 (Hệ số tương quan - Correlation coefficient)

Cho X, Y là hai biến ngẫu nhiên. Khi đó, hệ số tương quan của X và Y, ký hiệu là $\rho = \rho(X, Y)$, được cho bởi

$$\rho = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)} \cdot \sqrt{\mathrm{Var}(Y)}}.$$

Ý nghĩa của hệ số tương quan

Hệ số phản ánh mức độ phụ thuộc tuyến tính giữa X và Y.

- Nếu $\rho(X, Y) = 0$, thì hai biến X, Y không có quan hệ tuyến tính.
- Nếu $|\rho(X, Y)|$ càng gần 1, mức độ phụ thuộc tuyến tính càng cao.

Tính chất

Với hai biến ngẫu nhiên X, Y,

- $-1 \leq \rho(X, Y) \leq 1$.
- Nếu $Y = aX + b$, với $a \neq 0$, thì $\rho(X, Y) = \pm 1$.
- Nếu X, Y độc lập, thì $\rho(X, Y) = \mathrm{Cov}(X, Y) = 0$.

Ví dụ

Ví dụ 12

Cho X, Y là hai biến ngẫu nhiên có bảng hợp phân bố

$X \backslash Y$	0	1	2
0	0.1	0	0
1	0.12	0.26	0
2	0.14	0.3	0.08

Tính $\rho(X, Y)$.

Ví dụ (tiếp)

Ví dụ 13

Xét hai biến X, Y có hàm hợp mật độ

$$f(x,y) = c(1-xy), \quad 0 \le x, y \le 1.$$

(a) *Tìm c.*

(b) *Tính $\rho(X,Y)$.*

Định lý Giới hạn Trung tâm

Chặn Markov

Mệnh đề 5

Nếu X là một biến ngẫu nhiên không âm, thì với mọi $a > 0$,

$$\mathrm{P}[X \geq a] \leq \frac{E[X]}{a}.$$

Hệ quả - BĐT Chebyshev

Thay X bởi $(X-\mu)^2$, a bởi k^2, với μ là kỳ vọng của biến X, và $k>0$, ta có hệ quả sau.

Hệ quả (BĐT Chebyshev)

Nếu X là biến ngẫu nhiên có kỳ vọng μ và phương sai σ^2, thì với mọi $k>0$,

$$\mathrm{P}\left[|X-\mu|\geq k\right]\leq\frac{\sigma^2}{k^2}.$$

Ý nghĩa

Chặn Markov và Bất đẳng thức Chebyshev cho phép đánh giá về xác suất của một biến ngẫu nhiên trong trường hợp chỉ thông tin về kỳ vọng và phương sai của X được biết.

Ví dụ

Ví dụ 14

Giả sử rằng số lượng sản phẩm của một cơ sở sản xuất trong một tuần là một biến ngẫu nhiên X với giá trị trung bình 500.

(a) *Có thể nói gì về xác suất để tuần này xí nghiệp đó sản xuất được ít nhất* 1000 *sản phẩm?*

(b) *Nếu X có phương sai* 100, *có thể nói gì về xác suất để tuần này số lượng sản phẩm của xí nghiệp nằm trong khoảng* 400 *đến* 600?

Định lý Giới hạn Trung tâm

Định lý 1 (Central Limit Theorem - CLT)

Cho $X_1, X_2, \ldots$ là một dãy các biến ngẫu nhiên độc lập, có cùng phân bố với kỳ vọng μ và phương sai σ^2. Khi đó, phân bố của

$$\frac{X_1 + \cdots + X_n - \mu n}{\sigma\sqrt{n}}$$

hội tụ tới phân bố chuẩn tắc khi $n \to \infty$. Nói cách khác,

$$\mathrm{P}\left(\frac{X_1 + \cdots + X_n - \mu n}{\sigma\sqrt{n}} \leq a\right) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{x^2}{2}} dx$$

khi $n \to \infty$.

CLT mạnh thế nào?

Các biến ngẫu nhiên X's trong CLT có thể có phân bố bất kỳ. Đặt $\overline{X}_n = \frac{X_1 + \cdots + X_n}{n}$.

- Nếu $X_i \sim U(-a, a)$, thì $E[\overline{X}_n] = E[X_i] = 0$ và $\overline{X}_n$ có độ lệch chuẩn $\sigma = a/\sqrt{3}$. Theo CLT,

$$\frac{\overline{X}_n}{a/\sqrt{3n}} \xrightarrow{n \to \infty} \mathcal{N}(0, 1).$$

- Nếu $X_i \sim \text{Bernoulli}(p)$, thì $E[\overline{X}_n] = E[X_i] = p$, và $\overline{X}_n$ có độ lệch chuẩn $\sigma = \sqrt{p(1-p)}$. Theo CLT,

$$\frac{\overline{X}_n - p}{\sqrt{p(1-p)/n}} \xrightarrow{n \to \infty} \mathcal{N}(0, 1).$$

Ví dụ

Ví dụ 15

Một loại pin A có thời gian sử dụng trung bình là 40 giờ với độ lệch chuẩn 20 giờ. Một hộp có 25 quả pin loại này, và được sử dụng lần lượt mỗi lần một quả. Giả sử thời gian sử dụng của mỗi quả pin độc lập với nhau, hãy ước lượng xác suất để hộp pin đó có thể dùng được trên 1100 giờ.

Ví dụ

Ví dụ 15

Một loại pin A có thời gian sử dụng trung bình là 40 giờ với độ lệch chuẩn 20 giờ. Một hộp có 25 quả pin loại này, và được sử dụng lần lượt mỗi lần một quả. Giả sử thời gian sử dụng của mỗi quả pin độc lập với nhau, hãy ước lượng xác suất để hộp pin đó có thể dùng được trên 1100 giờ.

Gọi X_i là thời gian sử dụng của quả pin thứ i và $X = X_1 + \cdots + X_{25}$. Ta cần tính $\mathrm{P}(X > 1100)$. Áp dụng Định lý giới hạn trung tâm, ta có

$$
\begin{aligned}
\mathrm{P}(X > 1100) &= \mathrm{P}\left(\frac{X - 25\cdot 40}{20\sqrt{25}} > \frac{1100 - 25\cdot 40}{20\sqrt{25}}\right) \\
&\approx \mathrm{P}(Z > 1) = 1 - \Phi(1) = 0.1587.
\end{aligned}
$$

Áp dụng

Bài tập

Một tài xế giao hàng có thời gian giao mỗi lần trung bình 20 *phút với độ lệch chuẩn* 15 *phút. Anh ta muốn tính số phút m sao cho tổng thời gian của* 100 *lần giao tiếp theo không vượt quá m với xác suất* 0.95. *Tìm m, biết rằng* $\Phi(1.645) = 0.95$.

Xấp xỉ Phân bố Nhị thức

Vấn đề

Cho X là biến ngẫu nhiên Nhị thức với tham số (n, p). *Trong trường hợp n lớn, giả sử* $n = 100$, *chúng ta tính giá trị* $p(k) = \mathrm{P}(X = k)$ *thế nào? Có cách nào nhanh chóng không?*

Quan sát

Chúng ta có các quan sát đơn giản sau

(a) Trong phát biểu của CLT, ta thấy

$$\mathrm{P}\left(\frac{X_1+\cdots+X_n-\mu n}{\sigma\sqrt{n}}\le a\right)\xrightarrow{n\to\infty}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a}e^{-\frac{x^2}{2}}dx.$$

Vế phải có giá trị bằng $\Phi(a)$.

(b) Biến Nhị thức X với tham số (n,p) có thể được biểu diễn bằng tổng của n biến Bernoulli $X_1,\ldots,X_n$ với tham số p. Theo CLT,

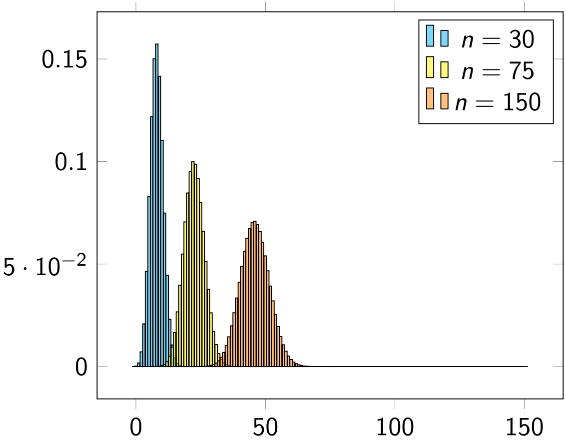
$$\frac{X-E[X]}{\sqrt{\mathrm{Var}(X)}}=\frac{X_1+\cdots+X_n-np}{\sqrt{p(1-p)}\sqrt{n}}\overset{CLT}{\approx}\mathcal{N}(0,1).$$

Xấp xỉ Phân bố Nhị thức

Trong trường hợp $np(1-p) \geq 10$, ta có thể xấp xỉ phân bố Nhị thức như sau

- $\mathrm{P}(X = k) \approx \Phi\left(\frac{k+0.5-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k-0.5-np}{\sqrt{np(1-p)}}\right).$
- $\mathrm{P}(k_1 \leq X) \approx 1 - \Phi\left(\frac{k_1-0.5-np}{\sqrt{np(1-p)}}\right).$
- $\mathrm{P}(X \leq k_2) \approx \Phi\left(\frac{k_2+0.5-np}{\sqrt{np(1-p)}}\right).$
- $\mathrm{P}(k_1 \leq X \leq k_2) \approx \Phi\left(\frac{k_2+0.5-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k_1-0.5-np}{\sqrt{np(1-p)}}\right).$

Đồ thị



Hình 2: Biến Nhị thức với $p = 0.3$.

Bảng giá trị của $\Phi(x)$

x	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441

Bảng 1: Một số giá trị của $\Phi(x)$.

Ví dụ

Ví dụ 16

Tung một đồng xu cân 100 lần. Tính xác suất để:

(a) *Có đúng 52 lần đồng xu cho mặt ngửa.*

(b) *Số lần cho mặt ngửa ít nhất là 52.*

(c) *Số lần cho mặt ngửa tối đa là 52.*

Áp dụng

Bài tập

Các số viết trong hệ thập phân thường được xấp xỉ bởi số nguyên gần nhất. Giả sử n số $X_1, \ldots, X_n$ được xấp xỉ bởi các số nguyên $J_1, \ldots, J_n$. Đặt $U_i = X_i - J_i$. Giả sử rằng U_i có phân bố đều trên $(-0.5, 0.5)$ và độc lập với nhau.

(a) *Chứng minh rằng* $\frac{\sum_{i=1}^n U_i}{\sqrt{n/12}} \sim \mathcal{N}(0,1)$ *khi* $n \to \infty$.

(b) *Tính* $\mathrm{P}\left(\frac{-5}{\sqrt{300/12}} \le \frac{\sum_{i=1}^{300} U_i}{\sqrt{300/12}} \le \frac{-5}{\sqrt{300/12}}\right)$.

(c) *Tìm a sao cho* $\mathrm{P}\left(-a \le \sum_{i=1}^{300} U_i \le a\right) = 0.95$.

(d) *Tìm a sao cho* $\mathrm{P}\left(-a \le \sum_{i=1}^{10^6} U_i \le a\right) = 0.99$.

Một số Phân bố thông dụng trong Thống kê

Phân bố χ^2

Định nghĩa 9

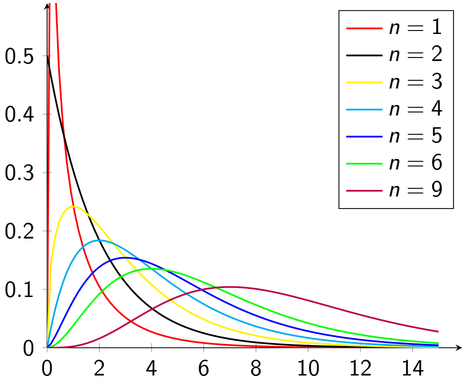
Cho $Z_1, \dots, Z_n$ là các biến ngẫu nhiên độc lập, cùng có phân bố chuẩn tắc. Khi đó, biến ngẫu nhiên $Z_1^2 + \cdots + Z_n^2$ được gọi là có phân bố Khi bình phương với n bậc tự do (degree(s) of freedom), ký hiệu là χ_n^2.

Mệnh đề 6

Hàm mật độ của biến χ_n^2 được cho bởi

$$f(x) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2}, \quad 0 < x < \infty.$$

χ^2



Hình 3: Phân bố χ^2.

Phân bố Student

Định nghĩa 10

Nếu Y và Z là các biến ngẫu nhiên độc lập, Y có phân bố χ^2 với n bậc tự do, $Z \sim \mathcal{N}(0,1)$, thì biến ngẫu nhiên

$$T = \frac{Z}{\sqrt{Y/n}}$$

*được gọi là có **phân bố Student** (hay phân bố t) với n bậc tự do. Phân bố Student được ký hiệu bởi $T \sim T_n$.*

Hàm mật độ

Mệnh đề 7

Hàm mật độ của biến T_n được cho bởi

$$f_n(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad -\infty < x < \infty.$$

Hàm mật độ

Mệnh đề 7

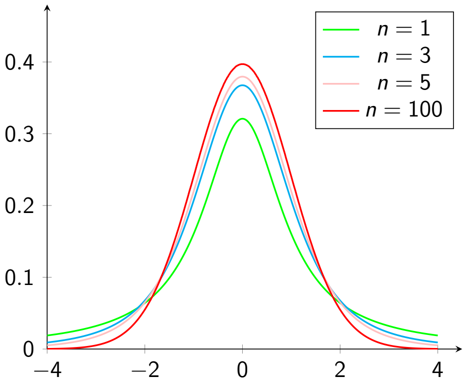
Hàm mật độ của biến T_n được cho bởi

$$f_n(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad -\infty < x < \infty.$$

Chú ý

$$\lim_{n\to\infty} f_n(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Student-t



Hình 4: Phân bố Student.

Phân bố Fisher-F

Định nghĩa 11

Cho U, V là hai biến ngẫu nhiên độc lập có phân bố χ^2 với bậc tự do lần lượt là m, n. Khi đó, biến ngẫu nhiên

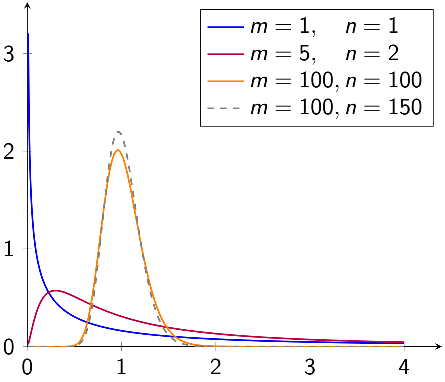
$$F = \frac{U/m}{V/n}$$

được gọi là có phân bố Fisher (hay phân bố F) với bậc tự do là m, n. Phân bố F được ký hiệu là $F \sim F(m, n)$.

Hàm mật độ của phân bố $F(m, n)$ là

$$f(x) = \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})}\left(\frac{m}{n}\right)^{\frac{m}{2}} x^{\frac{m}{2}-1}\left(1+\frac{m}{n}x\right)^{-\frac{m+n}{2}}, \quad 0 < x < \infty.$$

Fisher



Hình 5: Phân bố Fisher.

End of Chapter 2