

PHÂN PHỐI NHỊ THỨC

Trong phân phối chuẩn, biến ngẫu nhiên là biến số liên tục (cân nặng, chiều cao, trị số đường máu..), còn trong phân phối nhị thức (Binomial distribution) biến ngẫu nhiên là biến rời rạc (discrete) hoặc gọi là biến nhị phân, kết cục chỉ có 2 tình huống xảy ra như: sống/chết, có bệnh/không bệnh.. sẽ phân bố như thế nào.

Ví dụ. Thủ tung 1 đồng xu, đương nhiên mỗi lần tung có kết cục độc lập, có nghĩa là kết cục ra Sấp (S), Ngửa (N) lần sau không tùy thuộc vào kết quả S,N của lần tung trước, như vậy xác suất $p(N)=1/2$ và xác suất $p(S)=1/2$. Nếu ta tung 4 lần và muốn ra mặt N, thì ta có thể dự đoán xác suất được 2 lần N là cao nhất, sau đó là được 1N hoặc 3N và cuối cùng là cả 4 lần tung đều ra N hoặc không có mặt N nào có xác suất thấp nhất.

Như vậy làm sao để tính xác suất để có 0,1,2,3,4 mặt ngửa (N) sau 4 lần tung?

Cách tính 1:

Ta biết xác suất “thực” của mặt ngửa là $1/2$ ($p=0,5$) và mặt sấp cũng $1/2$ ($q=0,5$)

Gọi k là số lần được mặt ngửa, tính xác suất với $k=0,1,2,3,4$:

			SNNS	
			SNSN	
	SSSN	SSNN	NNNS	
	SSNS	NSSN	NNSN	
	SNSS	NSNS	NSNN	
	SSSS	NSSS	NNSS	SNNN NNNN
$k =$	0	1	2	3
				4

Biểu đồ 11.1 Các tình huống xảy ra sau 4 lần tung ($n=4$)

k=0: xác suất để không ($k=0$) được mặt N nào là: $SSSS = q \cdot q \cdot q \cdot q$ (qui tắc nhân xác suất) $= 0,5 \times 0,5 \times 0,5 \times 0,5 = 0,5^4 = 0,0625$

k=1: có 4 tình huống được 1 mặt N $\Rightarrow P(k=1) = 4 \cdot 0,5^4 = 0,25$

k=2: có 6 tình huống được 1 mặt N $\Rightarrow P(k=2) = 6 \cdot 0,5^4 = 0,375$

k=3: 4 tình huống được 3 mặt N $\Rightarrow P(k=3) = 4 \cdot 0,5^4 = 0,25$

k=4: chỉ có 1 tình huống duy nhất $\Rightarrow P(k=4) = 0,5^4 = 0,0625$

Cách tính 2:

Dùng nhị thức Newton:

$$(p+q)^4 = p^4 + 4p^3q + 6p^2q^2 + 4pq^3 + q^4$$

$$(0,5+0,5)^4 = 0,5^4 + 4 \times 0,5^4 + 6 \times 0,5^4 + 4 \times 0,5^4 + 0,5^4$$

k=0 tương ứng số hạng thứ 1: $0,5^4$

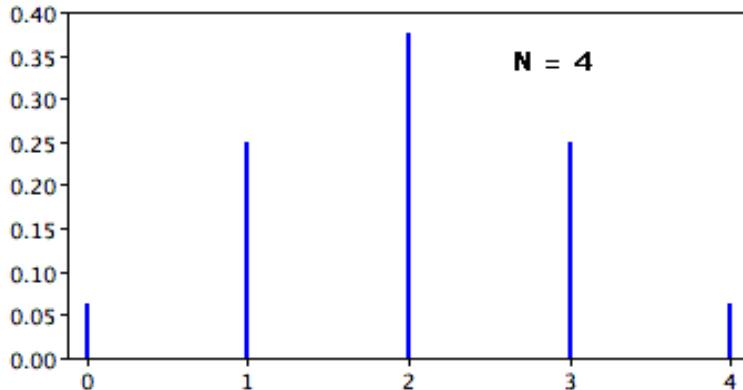
k=1 tương ứng số hạng thứ 2: $4 \times 0,5^4$

k=2 3: $6 \times 0,5^4$

k=3 4: $4 \times 0,5^4$

k=4 5: $0,5^4$

Bây giờ ta gọi X ($k=0,1,2,\dots$) là biến ngẫu nhiên trên trực hoành và p là tần số xác suất được mặt ngửa trên trực tung và vẽ biểu đồ phân phối (xem biểu đồ 11.2)



Biểu đồ 11.2 Phân phối nhị thức với $n=4$

11.1 Hàm phân phối nhị thức:

Tổng quát hóa hàm phân phối nhị thức được phát biểu như sau:

Một thử nghiệm với n lần thử độc lập

Gọi p là xác suất “thành công” (ví dụ. ra mặt ngửa)

Thì xác suất của k lần thử nghiêm thành công là:

$$P(k, n, p) = C_k^n p^k (1-p)^{n-k}$$

$$C_k^n = \frac{n!}{n!(n-k)!} \quad (\text{Tổ hợp } n \text{ chập } k) \quad \text{trong đó } k=0,1,2,\dots,n.$$

Nếu $n=12$, dùng công thức tính:

$$P(k=0; 12; 0,5) = \frac{12!}{12!(12!)} (0,5)^{12} = 0,000244$$

$$P(k=1; 12; 0,5) = \frac{12!}{12!(11!)} (0,5)^{11} = 0,00292$$

$$P(k=2; 12; 0,5) = \frac{12!}{12!(10!)} (0,5)^{10} = 0,01612$$

.....

$$P(k=6;12;0;5) = \frac{12!}{12!(6!)} (0,5)^6 = 0,2255$$

.....

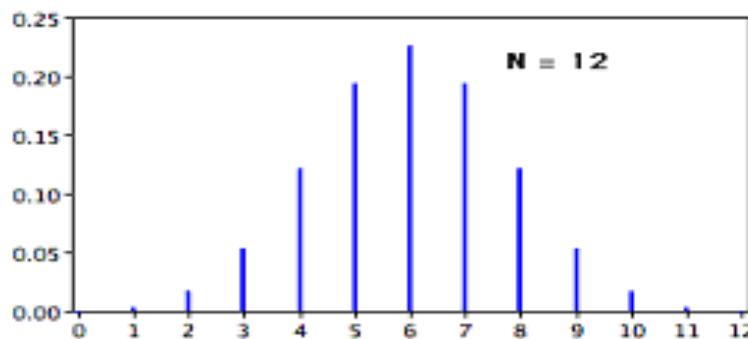
$$P(k=12;12;0;5) = \frac{12!}{12!(12!)} (0,5)^{12} = 0,000244$$

Cách tính 3.

Vào phần mềm R với lệnh đơn giản: **dbinom** (k,n,p)

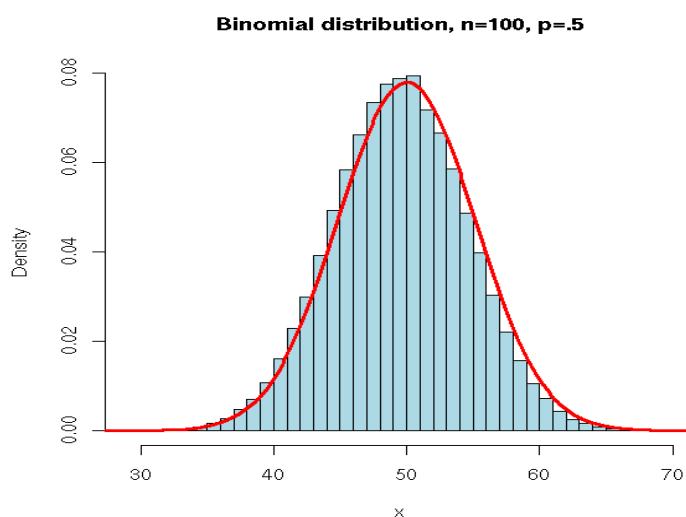
Ví dụ: tính xác suất với k=6, n=12 và p=0,5:

>**dbinom (6,12,0.5)** thì P= 0,2255



Biểu đồ 11.3 Phân phối nhị thức với n=12

Và khi n càng tăng dần, phân phối nhị thức có dạng phân phối chuẩn với trị số trung bình $\mu = np$ và phương sai $\sigma^2 = np(1-p)$ theo định luật giới hạn trung tâm (central limit theorem).



Biểu đồ 11.4 Phân phối nhị thức với n=100

Nhìn vào biểu đồ 4 với n lớn ($n=100$) chúng ta thấy phân phối nhị thức có phân phối chuẩn với trị trung bình là $\mu=np=100 \times 0,5=50$ và phương sai $\sigma^2=100 \times 0,5 \times 0,5=25$ hoặc độ lệch chuẩn $\sigma=5$.

11.2 Ứng dụng phân phối nhị thức:

Ví dụ 1. Biết tỉ lệ suy dinh dưỡng ở trẻ em dưới 5 tuổi là 20%. Nếu chúng ta khám 10 trẻ dưới 5 tuổi. Tính xác suất để có 2 em bị suy dinh dưỡng.

Như vậy $k=2$, $n=10$ và $p=0,20$

Xác suất có 2 trẻ bị suy dinh dưỡng.

$$P(k=2;10;0,2) = \frac{10!}{10!(8!)} (0,2)^2 \times (0,8)^8 = 0,30$$

hoặc tính trong R:

>**dbinom(2,10,0.2)**=0,30

Kết quả: Xác suất để có 2 em bị suy dinh dưỡng là 30%

Ví dụ 2. Tỉ lệ suy dinh dưỡng trẻ em 20% (như ví dụ trên). Tính xác suất để có ít hơn 2 em bị suy dinh dưỡng.

Như vậy tính tổng xác suất của: $P(k=0)+P(k=1)+P(k=2)$

Hoặc dùng hàm nhị thức tích lũy (cummulative binomial probability):

tính $P(X \leq 2)$:

>**pbinom(2,10,0.2)**= 0,67

Tài liệu tham khảo:

1. Nguyễn Văn Tuấn. Hàm phân phối nhị phân, trong Phân tích số liệu và tạo biểu đồ bằng R. Nhà xuất bản KH và KT, Thành phố HCM 2007, tr: 47-50.
2. Moore D.S. and McCabe G.P. Discrete random variables, in Introduction to the Practice of Statistics, 3rd Edition. W.H. Freeman and Company 1999 USA, pp: 313-317.