LỚP PHÂN PHỐI (a,b,0)

THE (a,b,0) CLASS OF DISTRIBUTION

ThS. Nguyễn Đức Khiêm
Khoa Khoa học Cơ bản - Trường ĐHXD Miền Tây
Email: ndkhiem@mtu.edu.vn
Điện thoại: 0980 720 737

Ngày nhận bài: 25/8/2022
Ngày gửi phản biện: 14/9/2022
Ngày chấp nhận đăng: 27/9/2022

Tóm tắt:

Bài báo này nghiên cứu việc thiết lập một lớp các phân phối của biến ngẫu nhiên rời rạc, cụ thể là các biến ngẫu nhiên có phân phối đếm. Phân phối đếm là phân phối rời rạc với xác suất trên giá trị nguyên không âm. Phân phối này thường được ứng dụng trong các mô hình tổn thất trong bảo hiểm [1].

Từ khóa: *biến ngẫu nhiên rời rạc, hàm xác suất, hàm sinh xác suất, hàm phân phối, kỳ vọng, phương sai.*

Abstract:

This paper studies the establishment of the class of distributions of discrete random variables, namely, random variables with a counting distribution. The counting distribution is a discrete distribution with probabilities over non-negative integer values. This distribution is commonly used in insurance loss models.

Keywords: *random variable, probability function, probability generating function, distribution function, mean, variance.*

1. Một số khái niệm

• **Hàm phân phối xác suất** [2] của biến ngẫu nhiên X

$$F(x) = P(X \leq x), x \in \mathbb{R}. \quad (1.1)$$

• **Hàm xác suất** p_k là xác suất của biến cố (chẳng hạn như tổn thất hay yêu cầu) thứ k xảy ra. Đặt X là biến ngẫu nhiên biểu diễn số biến cố. Khi đó

$$p_x = P(X = x), x = 0,1,2, \ldots \quad (1.2)$$

• **Hàm sinh xác suất** của biến ngẫu nhiên rời rạc X với hàm xác suất p_k là

$$P(t) = E(t^X) = \sum_{x=0}^{\infty} p_k t^x, \quad (1.3)$$

đặc biệt

$$P'(1) = E(X) \text{ và } P''(1) = E[X(X-1)].$$

Phần tiếp theo của bài báo này là thiết lập lớp phân phối của một số phân phối xác suất của biến ngẫu nhiên rời rạc.

2. Một số phân phối thông dụng của biến ngẫu nhiên rời rạc

2.1. Phân phối nhị thức

Phân phối nhị thức (Binomial Distribution): Khi ta tiến hành lặp một phép thử n lần một cách độc lập và quan sát một biến cố A nào đó có xảy ra (thành công) hay không. Trong mỗi phép thử chỉ có thể xảy ra 1 trong 2 trường hợp là hoặc A xảy ra hoặc A không xảy ra và xác suất A xảy ra trong mỗi lần là $p = P(A)$ luôn không đổi. Gọi X là số lần xảy ra của biến cố A trong n phép thử" thì X là rời rạc và nó có thể nhận 1 trong các giá trị $0,1,2, \ldots, n$ khi đó X có phân phối nhị thức.

Ta kí hiệu phân phối nhị thức là $X \sim B(n,p)$ với 2 tham số là n: số lần thực hiện phép thử và p là xác suất biến cố A xảy ra. Phân phối nhị thức được xem là phân phối đếm vì X là biến ngẫu nhiên rời rạc nhận giá trị x là số nguyên, không âm.

• **Hàm xác suất** của phân phối nhị thức là xác suất để biến cố A xảy ra x lần ($x = 0,1,2, \ldots, n$) trong dãy n phép thử và

$$p_x = C_n^x . p^x . (1-p)^{n-x} \quad (2.1)$$

trong đó n là tổng số lần thử; k là số lần biến cố A xảy ra; p là xác suất biến cố A xảy ra và $(1-p)$ là xác suất biến cố A không xảy ra của mỗi lần thử.

• **Hàm sinh xác suất** của biến ngẫu nhiên có phân phối nhị thức [3]

$$P(t) = [1 + p(t-1)]^n \quad (2.2)$$

• **Kỳ vọng và phương sai** của phân phối nhị thức là [2]

$$E(X) = np \text{ và } Var(X) = npq.$$

Phân phối nhị thức có phương sai bé hơn kỳ vọng. Nó phù hợp với các tập dữ liệu có phương sai quan sát mẫu nhỏ hơn kỳ vọng mẫu [4].

2.2. Phân phối Poisson

Trong lý thuyết xác suất và thống kê, phân phối Poisson là một phân phối xác suất rời rạc. Nhưng nó khác với các phân phối xác suất rời rạc khác ở chỗ thông tin cho biết không phải là xác suất để một biến cố (event) xảy ra (thành công) trong một lần thử như số lần mà biến cố đó xảy ra trong n lần thử của phân phối nhị thức, mà chính là trung bình số lần xảy ra thành công của một biến cố trong một khoảng thời gian nhất định. Giá trị trung bình này có giá trị thực và được gọi là λ.

Khi ta xem xét một biến ngẫu nhiên X nào đó và đếm số lần xuất hiện của nó trong một khoảng thời gian cho trước. Nếu số lần trung bình mà biến ngẫu nhiên đó xảy ra trong một khoảng thời gian là λ, thì X có phân phối Poisson, kí hiệu là $X \sim \wp(\lambda)$. Phân phối Poisson cũng được xem là phân phối đếm vì X là biến ngẫu nhiên rời rạc nhận giá trị x là số nguyên, không âm.

• **Hàm xác suất** của biến ngẫu nhiên có phân phối Poisson là xác suất của số lần xuất hiện của biến cố A xảy ra ($x = 0,1,2,\dots,n$) sẽ được tính như sau [5]

$$p_x = P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} \qquad (2.3)$$

• **Hàm sinh xác suất** của biến ngẫu nhiên có phân phối Poisson từ (1.3) là [3]

$$P(t) = E(t^X) = e^{\lambda(t-1)} \qquad (2.4)$$

• **Kỳ vọng và phương sai** được tính từ hàm sinh xác suất

$$E(X) = \lambda, Var(X) = \lambda.$$

Vì vậy phân phối Poisson có kỳ vọng bằng phương sai. Điều này khác với phân phối nhị thức có phương sai bé hơn kỳ vọng. Cho nên, phân phối Poisson là biểu diễn tốt hơn (gần với dữ liệu hơn) phân phối nhị thức đối với tập dữ liệu quan sát có phương sai bé hơn kỳ vọng.

Phân phối này có hai tính chất quan trọng.

Tính chất 1: Nếu $X_1, X_2, \dots, X_n$ là các biến ngẫu nhiên độc lập có phân phối Poisson với các tham số $\lambda_1, \lambda_2, \dots, \lambda_n$. Khi đó $X = X_1 + X_2 + \dots + X_n$ cũng có phân phối Poisson với tham số là $\lambda_1 + \lambda_2 + \dots + \lambda_n$ [4].

Tính chất 2: Giả sử rằng số biến cố X là biến ngẫu nhiên Poisson có kỳ vọng λ. Hơn nữa giả sử, mỗi biến cố có thể được phân lớp vào một trong n lớp với xác suất $p_1, p_2, \dots, p_n$ và nó độc lập với các biến cố khác. Khi đó, số biến cố $X_1, X_2, \dots, X_n$ tương ứng với $1,2,\dots,n$ loại biến cố là các biến ngẫu nhiên độc lập với kỳ vọng $\lambda p_1, \lambda p_2, \dots, \lambda p_n$ tương ứng. [4]

2.3. Phân phối nhị thức âm

Phân phối nhị thức âm (Negative Binomial Distribution): Giả sử trong một dãy n phép thử Bernoulli ta quan sát biến cố A. Mỗi phép thử có hai kết quả tiềm ẩn được gọi là "thành công" khi biến cố A xảy ra và "thất bại" khi biến cố A không xảy ra. Trong mỗi phép thử, xác suất A xảy ra là p và A không xảy ra là $(1-p)$. Chúng tôi quan sát số lần biến cố A không xảy ra trong dãy này cho đến khi biến cố A xảy ra ở lần thứ r. Gọi X là số lần biến cố A không xảy ra cho đến khi nó xảy ra ở lần thứ r (lần A xảy ra đầu

ISSN 2525-2615

tiên) được xác định trước, khi đó X nhận 1 trong các giá trị $0,1,2,\dots$ và X có phân phối nhị thức âm.

Ta kí hiệu phân phối nhị thức âm là $X \sim NB(r,p)$ với hai tham số: r là số lần xuất hiện của A cùng với p là xác suất xuất hiện của A trong mỗi phép thử.

Phân phối nhị thức âm cũng được xem là phân phối đếm vì X là biến ngẫu nhiên rời rạc nhận giá trị x là số nguyên, không âm. Dưới đây là các yếu tố chính cần lưu ý về thí nghiệm nhị thức âm:

- **Hàm xác suất** của phân phối nhị thức âm là xác suất biến ngẫu nhiên rời rạc X chỉ số lần thất bại trước khi thành công (lần thứ r) được tính theo công thức [1]

$$p_x = C_{x+r-1}^{r-1} p^r (1-p)^x \qquad (2.5)$$

trong đó n là tổng số lần thử; r là số lần thành công; x là số lần thất bại, $x = 0,1,2,\dots$; p là xác suất thành công của mỗi lần thử; và $(1-p)$ là xác suất thất bại của mỗi lần thử.

- **Hàm sinh xác suất** của phân phối nhị thức âm là [3]

$$P(t) = E(t^X) = \left[\frac{p}{1-(1-p)t}\right]^r \quad (2.6)$$

- **Kỳ vọng và phương sai** của phân phối nhị thức âm tương ứng là

$$E(X) = \frac{r(1-p)}{p} \text{ và } Var(X) = \frac{r(1-p)}{p^2}.$$

Vì p dương nên phương sai của phân phối nhị thức âm lớn hơn kỳ vọng. Điều này khác với phân phối Poisson có kỳ vọng bằng phương sai và phân phối nhị thức có phương sai bé hơn kỳ vọng. Cho nên, phân phối nhị thức âm là biểu diễn tốt hơn (gần với dữ liệu hơn) phân phối Poisson và phân phối nhị thức đối với tập dữ liệu quan sát có phương sai lớn hơn kỳ vọng.

2.4. Phân phối hình học

Phân phối hình học (Geometric Distribution) được xem là dạng đặc biệt của phân phối nhị thức âm. Nó liên quan tới số lần thử cần thiết (n) cho một lần thành công (biến cố A xảy ra) duy nhất. Vì vậy phân phối hình học là một phân phối nhị thức âm với số lần thành công $(r = 1)$. Gọi X là số lần biến cố A không xảy ra cho đến khi nó xảy ra ở lần đầu tiên, khi đó X nhận 1 trong các giá trị $0,1,2,\dots,n$ và X có phân hình học.

Ta kí hiệu phân phối hình học là $X \sim \mathcal{G}(p)$ với tham số p là xác suất xuất hiện của A trong mỗi phép thử. Phân phối hình học cũng được xem là phân phối đếm vì X là biến ngẫu nhiên rời rạc nhận giá trị x là số nguyên, không âm.

- **Hàm xác suất** của phân phối hình học là xác xuất để biến ngẫu nhiên rời rạc X chỉ số lần "thất bại" (biến cố A không xảy ra) đến khi "thành công" ở lần đầu tiên và

$$p_x = P(X = x) = p(1-p)^x \qquad (2.7)$$

trong đó n là tổng số lần thử; x là số lần thất bại, $x = 0,1,2, \ldots, n$; p là xác suất thành công của mỗi lần thử; và $(1-p)$ là xác suất thất bại của mỗi lần thử.

• **Hàm sinh xác suất** của phân phối của phân phối hình học là

$$P(t) = E(t^X) = \frac{p}{1-(1-p)t} \quad (2.8)$$

• **Do đó, kỳ vọng và phương sai** của phân phối của phân phối hình học tương ứng là

$$E(X) = \frac{1-p}{p} \text{ và } Var(X) = \frac{1-p}{p^2}.$$

Các hàm xác suất và các đặc trưng của 4 dạng phân phối đếm nêu trên được tổng hợp trong Bảng 1 ở cuối bài báo này.

3. Lớp phân phối xác suất $(a, b, 0)$

Đặt p_k là hàm xác suất của biến ngẫu nhiên rời rạc. Nó là phần tử của lớp phân phối $(a, b, 0)$ nếu tồn tại hai hằng số a và b sao cho [1]

$$\frac{p_x}{p_{x-1}} = a + \frac{b}{x}, \text{ với } x = 1,2, \ldots, n \quad (3.1)$$

trong công thức đệ qui này p_0 là xác suất tại 0 và nó có thể xác định khi đặt tổng các xác suất bằng 1. Lớp phân phối xác suất $(a, b, 0)$ là lớp có hai tham số là a và b.

Sau đây, bằng cách thế hàm xác suất của mỗi phân phối Poisson, nhị thức, nhị thức âm và hình học vào phần bên trái của công thức đệ qui ta sẽ nhận được giá trị của a và b cho ở Bảng 1.2 với $p_0 = P(X = 0)$ là giá trị khởi đầu của công thức đệ qui.

3.1. Xác định tham số a, b đối với phân phối nhị thức

Thế hàm xác suất của phân phối Poisson (2.1) vào vế trái của công thức đệ qui (3.1) ta được

$$\frac{C_n^x p^x (1-p)^{n-x}}{C_n^{x-1} p^{x-1} (1-p)^{n-x+1}}$$
$$= \frac{(n-x+1)}{x} \cdot \frac{p}{(1-p)}$$
$$= -\frac{p}{(1-p)} + \frac{(n+1)p}{(1-p)x}.$$

Do đó

$$a = -\frac{p}{1-p} \text{ và } b = \frac{(n+1)p}{1-p}.$$

3.2. Xác định tham số a, b đối với phân phối Poisson

Thế hàm xác suất của phân phối Poisson (2.3) vào vế trái của công thức đệ qui (3.1) ta được

$$\frac{\frac{\lambda^x e^{-\lambda}}{x!}}{\frac{\lambda^{x-1} e^{-\lambda}}{(x-1)!}} = \frac{\lambda^x (x-1)!}{x!\,\lambda^{x-1}} = \frac{\lambda}{x}.$$

Do đó: $\quad a = 0$ và $b = \lambda$.

3.3. Xác định tham số a, b đối với phân phối nhị thức âm

Thế hàm xác suất của phân phối Poisson (2.5) vào vế trái của công thức đệ qui (3.1) ta được

ISSN 2525-2615

$$\frac{C_{x+r-1}^{r-1}p^r(1-p)^x}{C_{x+r-2}^{r-1}p^r(1-p)^{x-1}}$$
$$=\frac{(x+r-1)(1-p)}{x}$$
$$=(1-p)+\frac{(r-1)(1-p)}{x}.$$

Do đó

$a=(1-p)$ và $b=(r-1)(1-p)$.

3.4. Xác định tham số a, b đối với phân phối hình học

Thế hàm xác suất của phân phối Poisson (2.7) vào vế trái của công thức đệ qui (3.1) ta được

$$\frac{p(1-p)^x}{p(1-p)^{x-1}}=(1-p).$$

Do đó

$a=(1-p)$ và $b=0$

Như vậy, chúng ta đã xác định được các tham số a và b của lớp phân phối $(a,b,0)$ từ công thức đệ qui (3.1) và các phần tử của lớp này được tổng hợp ở Bảng 2.

4. Kết luận

– Bài báo đã hệ thống các đặc tính quan trọng của 4 dạng phân phối thường gặp của biến ngẫu nhiên rời rạc X nhận giá trị nguyên, không âm. Các phân phối này thường được vận dụng trong các mô hình thống kê trong lĩnh vực bảo hiểm với số yêu cầu phát sinh một cách ngẫu nhiên và đếm được.

– Bên cạnh đó, dựa vào công thức đệ qui (3.1) và hai tham số a, b của lớp phân phối $(a,b,0)$ ta có thể xác định phân phối nào vừa khít với mô hình thực tế. Từ đó ta sẽ đưa ra các phương pháp ước lượng tham số phù hợp và cho kết quả chính xác hơn đối với mẫu số liệu thực tế. Việc chọn phương pháp ước lượng phù hợp độc giả có thể nghiên cứu thêm ở phần tài liệu tham khảo.

Bảng 1. *Tổng hợp hàm xác suất và các đặc trưng của biến ngẫu nhiên rời rạc.*

Phân phối	Hàm xác suất p_x	Hàm sinh xác suất $P(t)$	Kỳ vọng $E(X)$	Phương sai $Var(X)$
Nhị thức: $X\sim B(n,p)$	$C_n^x p^x(1-p)^{n-x}$	$[1+p(t-1)]^n$	np	npq
Poisson: $X\sim\wp(\lambda)$	$\frac{\lambda^x e^{-\lambda}}{x!}$	$e^{\lambda(t-1)}$	λ	λ
Nhị thức âm: $X\sim NB(r,p)$	$C_{x+r-1}^{r-1}p^r(1-p)^x$	$\left[\frac{p}{1-(1-p)t}\right]^r$	$\frac{r(1-p)}{p}$	$\frac{r(1-p)}{p^2}$.
Hình học $X\sim\mathcal{G}(p)$	$p(1-p)^x$	$\frac{p}{1-(1-p)t}$	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$

(Trong đó x có giá trị nguyên, không âm)

***Bảng 2.** Phần tử của lớp xác suất* $(a, b, 0)$

Phân phối	a	b	p_0
Nhị thức: $X \sim B(n, p)$	$-\dfrac{p}{1-p}$	$\dfrac{p(n+1)}{1-p}$	$(1-p)^n$
Poisson: $X \sim \wp(\lambda)$	0	λ	$e^{-\lambda}$
Nhị thức âm: $X \sim NB(r, p)$	$1-p$	$(r-1)(1-p)$	p^x
Hình học $X \sim \mathcal{G}(p)$	$1-p$	0	p

Tài liệu tham khảo

[1] [Stuart A. Klugman, Harry H. Panjer, and Gordon E. Willmot, Loss Models: From Data to Decisions (Wiley Series in Probability and Statistics), 5 ed., Wiley, 2019.

[2] Nguyễn Đức Khiêm, Toán 3, Vĩnh Long: Trường ĐHXD Miền Tây, 2008.

[3] Tô Anh Dũng, Lý thuyết xác suất và thống kê toán, HCM: Nhà xuất bản ĐHQG Tp Hồ Chí Minh, 2007.

[4] Panjer H. H., Wilmot G. E., Insurance Risk Models, Society of Actuaries, Chicago: Schaumburg, 1992.

[5] Nguyễn Duy Tiến và Vũ Viết Yên, Lý thuyết xác suất, Hà Nội: Nhà xuất bản Giáo dục, 2003.

[6] Phạm Xuân Kiều, Lý thuyết xác suất và Thống kê, Hà Nội: Nhà xuất bản Giáo dục, 2005.

[7] Nguyễn Duy Tiến, Lý thuyết xác suất, Hà Nội: Nhà xuất bản Giáo dục, 2000.

[8] Ronald E. Walpole et al, Probability & Statistics for Engineers & Sciences, 9 ed., Prentice Hall, 2012.

ISSN 2525-2615