NGHIÊN CỨU CÁC ĐẶC TRƯNG SỐ CỦA TỔNG CÁC BIẾN NGẪU NHIÊN CÓ PHÂN PHỐI GAMMA PHỤ THUỘC

INVESTIGATION OF THE NUMERICAL CHARACTERISTICS OF THE SUM OF DEPENDENT GAMMA DISTRIBUTION RANDOM VARIABLES

Trần Quốc Chiến
Trường Đại học Sư phạm,
Đại học Đà Nẵng

Nguyễn Văn Hưng
Trường Cao đẳng Nghề Đà Nẵng

TÓM TẮT

Nghiên cứu này chú trọng vào các đặc trưng số cơ bản của tổng các biến ngẫu nhiên mà trong đó các biến của quá trình ngẫu nhiên là phụ thuộc và có phân phối không chuẩn.

Nội dung nghiên cứu trình bày các vấn đề cơ sở bao gồm hàm mật độ xác suất của các đại lượng ngẫu nhiên có phân phối gamma, các đặc trưng số cơ bản của phân phối gamma, mô hình tự hồi quy gamma bậc 1 (GAR(1), các biến ngẫu nhiên GAR(1) và tổng của các biến ngẫu nhiên GAR(1). Bằng phương pháp phân tích lý thuyết chúng tôi đã đưa ra được kết quả là các biểu thức giải tích biểu diễn kỳ vọng toán học và phương sai của tổng các biến ngẫu nhiên GAR(1).

Các kết quả đạt được trong nghiên cứu này sẽ rất hữu ích để nghiên cứu tìm lời giải cho lớp các bài toán trong thực tế liên quan đến các quá trình ngẫu nhiên và các ứng dụng của nó. Chẳng hạn bài toán tính dung lượng trung bình hồ chứa, bài toán tính lưu lượng dòng chảy trong lãnh vực thuỷ văn, các bài toán tính khoảng thời gian đến và phục vụ trong lý thuyết xếp hàng, điều khiển dự trữ và điều phối các yêu cầu phục vụ trong lãnh vực viễn thông, vv….

ABSTRACT

The main purpose of the study is to center on the basic numerical charateristics of the sum of random variables in which the variables of the stochastic processes are generally dependent and not normally distributed.

The content of the study comprises an investigation of random variables that have gamma distribution, in which the gamma distribuition function and its basic numerical charateristics, the first order gamma autoregressive (GAR(1)) model, the sequences of gamma variables and the sum of GAR(1) variables are concerned. By a theoretical analysis we have obtained the analytical expressions of the expected value and variance of the sum of GAR(1) varriables.

These results will be useful for further study to solve some problems of stochastic processes and their application in practice such as the mean range of reservoir storage, streamflow in stochastic hydrology, interarrival and service times in queuing theory, lead times and demands in inventory control and telecommunication, etc.

Phân phối gamma có vai trò rất quan trọng trong việc nghiên cứu các quá trình ngẫu nhiên và các ứng dụng của nó. Đối với các quá trình ngẫu nhiên mà trong đó chuỗi các biến ngẫu nhiên có độ lệch và phụ thuộc thì áp dụng mô hình hồi quy gamma bậc 1 (GAR(1)) là rất hiệu quả.

Trong thực tế có nhiều bài toán liên quan đến tổng các biến ngẫu nhiên có phân phối gamma phụ thuộc theo mô hình GAR(1), chẳng hạn bài toán tính dung lượng trung bình hồ chứa, bài toán tính thời gian chờ đợi trung bình được phục vụ với các dòng vào là các quá trình ngẫu nhiên có phân phối gamma phụ thuộc. Để giải các bài toán này, các đặc trưng số cơ bản của tổng các biến ngẫu nhiên theo mô hình GAR(1) cần phải được xem xét. Trong phạm vi nghiên cứu ở đây chúng tôi phân tích và đưa ra kết quả các đặc trưng số cơ bản gồm kỳ vọng toán học và phương sai của tổng các biến ngẫu nhiên GAR(1).

1. Phân phối Gamma

Một biến ngẫu nhiên liên tục X được gọi là có phân phối gamma 3 tham số nếu hàm mật độ xác suất của nó là:

$$f(x) = \frac{(x-c)^{\mathrm{a}-1} e^{-(x-c)/b}}{b^{\mathrm{a}}\Gamma(\mathrm{a})} \tag{1}$$

trong đó a>0, b>0, c>0, x ≥ c; a, b, c tương ứng là các tham số độ nhọn, tỉ lệ và vị trí.

Hàm $\Gamma(\mathrm{a})$ được xác định bởi

$$\Gamma(\mathrm{a}) = \int_0^\infty t^{\mathrm{a}-1} e^{-t} dt \qquad ,\mathrm{a} > 0$$

Đây là hàm đệ quy :

$$\Gamma(\mathrm{a}+1) = \mathrm{a}\Gamma(\mathrm{a})$$

Khi a = k (k là một số nguyên dương) ta có :

$$\Gamma(\mathrm{k}) = (\mathrm{k}-1)! = 1*2*.....(\mathrm{k}-1)$$

Khi c = 0 ta có phân phối gamma 2 tham số ;

Khi c = 0 và b = 1 ta có phân phối gamma 1 tham số

2. Các đặc trưng số của phân phối gamma

Các đặc trưng số của phân phối gamma 3 tham số được tính như sau:

- Kỳ vọng: $\mathrm{E(X)} = \mathrm{ab} + \mathrm{c}$
- Phương sai: $Var(X) = \mathrm{ab}^2$
- Hệ số lệch: $2/\sqrt{\mathrm{a}}$

Trường hợp với 2 tham số ta có:

- Kỳ vọng: $\mathrm{E(X)} = \mathrm{ab}$

- Phương sai: $Var(X) = ab^2$

- Hệ số lệch: $2/\sqrt{a}$

Trường hợp với 1 tham số ta có:

- Kỳ vọng: $E(X) = a$

- Phương sai: $Var(X) = a$

- Hệ số lệch: $2/\sqrt{a}$

Hàm phân phối gamma:

$$F(X) = P(X \le x) = (\Gamma(a))^{-1} \int_0^y t^{a-1} e^{-t} dt$$

Với $y = (x - c)/b$

Khi a = k (một số nguyên dương) ta có

$$F(X) = 1 - \left(1 + y + y^2/2! + ... + y^{k-1}/(k-1)!\right) e^{-y}$$

Với $y = (x - c)/b$

3. Mô hình tự hồi quy gamma bậc 1 (GAR(1))

Để giải những bài toán trong thực tế mà trong đó các quá trình ngẫu nhiên là phụ thuộc và có phân phối không chuẩn khi đó chuỗi các biến ngẫu nhiên gamma được nghiên cứu và áp dụng rất hiệu quả.

Đã có nhiều công trình nghiên cứu đề xuất các mô hình sinh ra chuỗi các biến ngẫu nhiên gamma phụ thuộc, trong đó mô hình được đề xuất bởi LAWRANCE và LEWIS(1981) tỏ ra rất hiệu quả và được ứng dụng phổ biến.

Mô hình tự hồi quy gamma bậc 1 (GAR(1)) được đề xuất bởi LAWRANCE và LEWIS(1981) như sau:

$$X_i = \Phi X_{i-1} + e_i \quad (2)$$

Trong đó: X_i là biến ngẫu nhiên biểu diễn quá trình phụ thuộc ở thời điểm i

Φ là hệ số hồi quy

e_i là biến ngẫu nhiên độc lập cần được xác định.

X_i có phân phối gamma 3 tham số có hàm mật độ xác suất như ở phương trình 1 Quá trình được xác định bởi phương trình 2 được gọi là mô hình GAR(1).

Trong thực tế có nhiều bài toán liên quan đến tổng các biến ngẫu nhiên có phân phối gamma phụ thuộc theo mô hình GAR(1), Để giải các bài toán này ; các đặc trưng số của tổng các biến ngẫu nhiên theo mô hình GAR(1) cần phải được xem xét. Trong phạm vi nghiên cứu ở đây chúng tôi sẽ xem xét các đặc trưng cơ bản là kỳ vọng và phương sai của tổng các biến ngẫu nhiên GAR(1).

4. Các đặc trưng số cơ bản của tổng các biến ngẫu nhiên GAR(1)

Trong mô hình GAR(1) các biến ngẫu nhiên được xác định theo mô hình ở phương trình (2), khi đó tổng của n biến ngẫu nhiên GAR(1) đầu tiên được ký hiệu là S_n và

$$S_n = X_1 + X_2 + ... + X_n \tag{3}$$

4.1. Kỳ vọng của tổng của các biến ngẫu nhiên GAR(1)

Kỳ vọng của tổng của n biến ngẫu nhiên GAR(1) gọi là $E(S_n)$ được tính như sau:

Tổng S_n có dạng:

$$S_n = \sum_{i=1}^{n} X_i$$

trong đó: X_i, i = 1, 2, …, n là các biến GAR(1).

Theo định lý cộng của kỳ vọng toán học, với các trường hợp X_i, i = 1, 2, …, n là phụ thuộc hoặc độc lập thì kỳ vọng của tổng các biến ngẫu nhiên X_i là:

$$E(S_n) = E(\sum_{i=1}^{n} X_i)$$

$$= \sum_{i=1}^{n} E(X_i)$$

Vì vậy kỳ vọng của tổng của n biến ngẫu nhiên GAR(1) là:

- Trường hợp các biến ngẫu nhiên có phân phối gamma 3 tham số:

$$E(S_n) = n(ab + c) \tag{4}$$

- Trường hợp các biến ngẫu nhiên có phân phối gamma 2 tham số:

$$E(S_n) = nab \tag{5}$$

- Trường hợp các biến ngẫu nhiên có phân phối gamma 1 tham số:

$$E(S_n) = na \tag{6}$$

4.2. Phương sai của tổng của các biến ngẫu nhiên GAR(1)

Phương sai của tổng của n biến ngẫu nhiên GAR(1) gọi là $Var(S_n)$ được biểu diễn như sau:

$$Var(S_n) = Var\sum_{i=1}^{n} X_i$$

Do các biến ngẫu nhiên X_i, i = 1, 2, …, n là phụ thuộc vì vậy ta có:

$$Var(\sum_{i=1}^{n} X_i) = \sum_{i,j=1}^{n} Cov(X_i, X_j)$$

$$= \sum_{i=1}^{n} Var(X_i) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} Cov(X_i, X_j)$$

$Cov(X_i, X_j)$ gọi là momen tương quan của X_i và X_j

Đặt j =i+k ta có :

$$Cov(X_i, X_j) = Cov(X_i, X_{i+k}) = r_k \sigma(X_i)\sigma(X_{i+k}) \tag{7}$$

Ở đây r_k là hệ số tương quan bậc k của chuỗi $\{X_i\}$ và $\sigma(X_i)$ là độ lệch tiêu chuẩn của X_i.

Ta có $\sigma(X_i) = [Var(X_i)]^{1/2}$

Do $Var(X_i) = Var(X_j)$ với mọi i,j :

$$\sigma(X_i)\sigma(X_j) = [\sigma(X_i)]^2 = Var(X_i) \tag{8}$$

Từ phương trình (7) và phương trình (8) ta có :

$$2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} Cov(X_i, X_j) = 2Var(X_i)\sum_{i=1}^{n-1}\sum_{k=1}^{n-i} r_k$$

$$= 2Var(X_i)\sum_{k=1}^{n-1}(n-k)r_k$$

Vì vậy ta có :

$$Var(S_n) = Var(X_i)\left[n + \sum_{k=1}^{n-1}(n-k)r_k\right] \tag{9}$$

Với quá trình ngẫu nhiên theo mô hình GAR(1) ta chứng minh được rằng

$$r_k = \Phi^k \tag{10}$$

trong đó Φ là hệ số hồi quy của mô hình GAR(1).

Để đơn giản, ta xem xét trường hợp các biến ngẫu nhiên có phân phối gamma 1 tham số:

* Chứng minh với trường hợp hệ số hồi quy bậc 1:

$$Cov(X_i, X_{i+1}) = E[(X_i - a)(X_{i+1} - a)] \tag{11}$$

Thay thế phương trình (2) vào phương trình (11) ta có:

$$Cov(X_i, X_{i+1}) = E[(X_i - a)(\Phi X_i + e_{i+1} - a)]$$
$$= E(\Phi X_i^2 + X_i e_{i+1} - a\ X_i - \Phi a X_i - a e_{i+1} + a^2)$$

Vì X_i và e_{i+1} là độc lập lẫn nhau do đó:

$$Cov(X_i, X_{i+1}) = \Phi E(X_i^2) + E(X_i)E(e_{i+1})$$
$$- aE(X_i) - \Phi a E(X_i) - aE(e_{i+1}) + E(a^2)$$

Trong mô hình GAR(1) tất cả các biến ngẫu nhiên X_i , i=0,1,…,n có kỳ vọng và phương sai:

$$E(X_i) = a \text{ và } Var(X_i) = a$$

Và $$E(X_i^2) = E(X_i X_i) = E(X_i)E(X_i) + Var(X_i) = a^2 + a$$

Từ phương trình (11) ta có

$$Cov\left(X_i, X_{i+1}\right) = \Phi(a^2 + a) - \Phi a^2 = \Phi a$$

Theo lý thuyết xác suất, hệ số tương quan:

$$r_1 = Cov(X_i, X_{i+1}) / \sigma(X_i)\sigma(X_{i+1})$$

Suy ra $r_1 = \Phi$

* Chứng minh với trường hợp hệ số hồi quy bậc 2:

$$\begin{aligned}
Cov(X_i, X_{i+2}) &= E[(X_i - a)(X_{i+2} - a)] \\
&= E[(X_i - a)(\Phi X_{i+1} + e_{i+2} - a)] \\
&= E[(X_i - a)(\Phi(\Phi X_i + e_{i+1}) + e_{i+2} - a)] \\
&= E[(X_i - a)(\Phi^2 X_i + \Phi e_{i+1} + e_{i+2} - a)] \\
&= E(\Phi^2 X_i^2 + \Phi X_i e_{i+1} + X_i e_{i+2} - aX_i - \Phi^2 aX_i - \Phi ae_{i+1} - ae_{i+2} + a^2 \\
&= \Phi^2 E(X_i^2) + \Phi E(X_i)E(e_{i+1}) + E(X_i)E(e_{i+2}) - aE(X_i) \\
&\quad - \Phi^2 aE(X_i) - \Phi aE(e_{i+1}) - aE(e_{i+2}) + E(a^2) \\
&= \Phi^2(a^2 + a) + \Phi aE(e_{i+1}) + aE(e_{i+2}) - a^2 \\
&\quad - \Phi^2 a^2 - \Phi aE(e_{i+1}) - aE(e_{i+2}) + a^2 \\
&= \Phi^2 a
\end{aligned}$$

Suy ra $r_2 = \Phi^2$

* Chứng minh với trường hợp hệ số hồi quy bậc k:

Bằng phương pháp quy nạp, Giả sử phương trình (10) đúng với hệ số tương quan bậc k-1, ta có:

$$r_{k-1} = \Phi^{k-1}$$

Hay $$\begin{aligned} Cov(X_i, X_{i+k-1}) &= E[(X_i - a)(X_{i+k-1} - a)] \\ &= \Phi^{k-1} a \end{aligned} \tag{12}$$

Trường hợp với hệ số tương quan bậc k ta có:

$$\begin{aligned}
Cov(X_i, X_{i+k}) &= E[(X_i - a)(X_{i+k} - a)] \\
&= E[(X_i - a)(\Phi X_{i+k-1} + e_{i+k} - a)] \\
&= E[(X_i - a)(\Phi X_{i+k-1} - \Phi a + \Phi a + e_{i+k} - a)] \\
&= \Phi E[(X_i - a)(X_{i+k-1} - a)] + E[(X_i - a)(\Phi a + e_{i+k} - a)]
\end{aligned} \tag{13}$$

Số hạng thứ hai của phương trình (13) bằng 0 vì vậy, từ phương trình (12) ta có:

$$\begin{aligned} Cov(X_i, X_{i+k}) &= \Phi\Phi^{k-1}a \\ &= \Phi^k a \end{aligned}$$

Suy ra $r_k = \Phi^k$ (phương trình (10) đã được chứng minh).

Phương trình (10) cũng đúng với các trường hợp các biến ngẫu nhiên GAR(1) có phân phối gamma có 2 hoặc 3 tham số, hoặc chuỗi các biến ngẫu nhiên có phân phối chuẩn.

Từ mối quan hệ giữa hệ số tương quan bậc 1 và bậc k đã được khẳng định ở trên: ta suy ra phương sai của tổng của các biến ngẫu nhiên GAR(1) như sau:

$$\begin{aligned} Var(S_n) &= Var(X_i)\left[n + 2\sum_{k=1}^{n-1}(n-k)r_k\right] \\ &= Var(X_i)\left[n + 2\sum_{k=1}^{n-1}(n-k)\Phi^k\right] \\ &= nVar(X_i) + 2Var(X_i)[(n-1)\Phi + (n-2)\Phi^2 + \ldots + \Phi^{n-1}] \end{aligned} \tag{14}$$

Từ phương trình (14) ta có phương sai của tổng của các biến ngẫu nhiên GAR(1) như sau:

- Trường hợp các biến ngẫu nhiên có phân phối gamma 3 tham số:

$$Var(S_n) = nab^2 + 2ab^2[(n-1)\Phi + (n-2)\Phi^2 + \ldots + \Phi^{n-1}] \tag{15}$$

- Trường hợp các biến ngẫu nhiên có phân phối gamma 2 tham số:

$$Var(S_n) = nab^2 + 2ab^2[(n-1)\Phi + (n-2)\Phi^2 + \ldots + \Phi^{n-1}] \tag{16}$$

- Trường hợp các biến ngẫu nhiên có phân phối gamma 1 tham số:

$$Var(S_n) = na + 2a[(n-1)\Phi + (n-2)\Phi^2 + \ldots + \Phi^{n-1}] \tag{17}$$

Kỳ vọng và phương sai của tổng các biến ngẫu nhiên GAR(1) đã được xác định bởi các phương trình (4)-(6) và (15)-(17) sẽ là cơ sở để tìm lời giải cho lớp các bài toán với các tham số là tổng của các đại lượng ngẫu nhiên có phân phối gamma phụ thuộc./.

TÀI LIỆU THAM KHẢO

[1] AHRENS, J.H. and DIETER, U. (1982), Generating Gamma Variates by a Modified Rejection Technique, ACM Transactions on Mathematical Software, Vol. 25, No. 1, pp. 47-54.

[2] D.I. KAZAKEVITS, (2005), Cơ sở Lý thuyết Hàm ngẫu nhiên và ứng dụng trong Khí tượng Thủy văn. (Người dịch: Phạm Văn Huấn, Nguyễn Thanh Sơn, Phan Văn Tân). NXB Đại học Quốc gia Hà Nội

[3] DO, L.M. (1988), Generating Gamma Variates, ACM Transactions on Mathematical Software, Vol. 14, No. 3, pp. 261-266.

[4] FERNANDEZ, B and SALAS, J.D. (1990), Gamma – Autoregressive Models for Streamflow Simulation, J. of Hydraulic Engineering, Vol. 116, No. 11, pp. 1403-1414.

[5] GAVE, D.P. and LEWIS, P.A.W. (1980), First Order Autoregressive Gamma Sequences and Point Process, Adv. Appl. Prob., Vol. 12, No.3, pp. 727-745.

[6] LAWRANCE, A.J. and LEWIS, P.A.W. (1981), A New Autoregressive Time Series Model in Exponential Variables (NEAR(1)), Adv. Appl. Prob. Vol. 13, No. 4, pp. 826-845.

[7] OBEYSEKERA, J.T.B. and YEVJEVICH, V. (1985), A Note on Simulation of Samples of Gamma – Autoregressive Variables, Water Resources - 41 -Research, Vol. 21, No. 10. pp. 1569-1572.