XÁC ĐỊNH HÀM PHÂN BỐ LƯU LƯỢNG LỚN NHẤT TẠI TRẠM SƠN TÂY BẰNG PHƯƠNG PHÁP DÒ TÌM NGẪU NHIÊN VỚI CÁC QUÁ TRÌNH THÍCH NGHI

Nguyễn Văn Nhai, PTS. Nguyễn Hữu Bảo
Đại học Thủy Lợi

1. Đặt vấn đề

Việc khảo sát hàm mật độ cho phân phối lưu lượng lớn nhất Q_{max} tại trạm Sơn Tây là một nhu cầu rất quan trọng trong việc khảo sát và khống chế lũ ở hạ lưu sông Hồng, các phương pháp phổ biến trong thủy văn trước đây (chủ yếu là phương pháp moment) cho sai số khả dĩ chấp nhận được. Tuy nhiên, với sự phát triển của các phương pháp giải tích số và đặc biệt là phương pháp Monte-Carlo sai số có khả năng giảm hơn nữa.

2. Phương pháp dò tìm ngẫu nhiên với các quá trình tự thích nghi [2].

Bằng quy tắc kiểm định χ^2 để chấp nhận (với mức ý nghĩa 95%) giả thiết phân bố của Q_{max} tại trạm Sơn Tây tuân theo luật loga chuẩn tức là hàm mật độ có dạng:

$$f(x) = \frac{1}{x\sigma_{lnx}\sqrt{2\pi}} \exp\left(\frac{-(lnx-\mu)^2}{2\sigma_{lnx}^2}\right) \tag{1}$$

Ở đây σ_{lnx} và μ là hai tham số cần ước lượng sao cho tỷ sai các quan sát thực đo tại trạm Sơn Tây và quan sát tính toán là nhỏ nhất (với 89 quan sát đo được).

Quá trình kiểm định giả thuyết được tiến hành như sau:

1. Lấy lôga (cơ số a) các quan sát: $y_t = lnx_i$

2. Nhóm các kết quả y_i theo các khoảng sao cho các khoảng đó phủ toàn trục $(-\infty, +\infty)$ và sao cho số kết quả trong mỗi khoảng đủ lớn (trong mọi trường hợp không bé hơn 5, nhưng tốt hơn là không bé hơn 10).

3. Tính số m_i các kết quả quan trắc y_i rơi vào mỗi khoảng $(y_{i-1}; y_i)$.

4. Gọi $p_i = \Phi\left(\frac{y_i-\bar{y}}{S}\right) - \Phi\left(\frac{y_{i-1}-\bar{y}}{S}\right)$ (*)

Trong đó: $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})}$

(p_i tính theo công thức (*) bằng cách tra bảng phân phối chuẩn $\Phi(x)$).

5. Tính tổng $x^2 = \sum_{i=1}^{l} \frac{(m_i - np_i)^2}{np_i}$ (**)

với l là số các khoảng $(-\infty, y_1)$, $(y_1; y_2)$, ..., $(y_{l-1}; +\infty)$, n là các số quan sát $n = m_1 + m_2 + ... + m_l$.

6. Cho trước một mức ý nghĩa P (thường là 95%). Tra bảng x^2 với l–3 bậc tự do và mức ý nghĩa P ta được số α. Nếu $\alpha < x^2$ (x^2 tính được theo công thức (**)) thì giả thiết các quan sát tuân theo luật phân phối loga chuẩn với mức tin cậy 95% bị bác bỏ. Trong trường hợp ngược lại thì chấp nhận giả thiết đó đúng với độ tin cậy 95%.

Trong bài toán kiểm định giả thiết lưu lượng Q_{max} tại trạm Sơn Tây tuân theo luật phân phối loga chuẩn của chúng ta 89 quan sát được chia làm 8 khoảng. Tính x^2 theo công thức (**) ta được $x^2 = 5{,}4075$ còn cho trước mức ý nghĩa P = 95%, tra bảng giá trị tới hạn x^2 với số bậc tự do là 5 ta được số $\alpha = 11{,}07$. Như vậy, giả thiết các quan sát Q_{max} tuân theo luật phân phối loga chuẩn đã được chấp nhận với độ tin cậy 95%.

Phương pháp dò tìm tự thích nghi được sử dụng ở đây theo 2 giai đoạn:

Giai đoạn 1: Xác định miền biến thiên của σ_{lnx} và bằng quy tắc 3–σ, ta có:

$$\begin{aligned} \mu_1 &\le \mu \le \mu_2 \\ \sigma_1 &\le \sigma_{lnx} \le \sigma_2 \end{aligned} \tag{2}$$

$$\mu_1 = \overline{\ln x} - 3S; \qquad \mu_2 = \overline{\ln x} + 3S$$

$$\sigma_1 = S - 3\sigma_s; \qquad \sigma_2 = S + 3\sigma_s$$

$$S = \sqrt{\frac{(\ln x_i - \overline{\ln x})^2}{n-1}} \qquad \overline{\ln x} = \frac{1}{89}\sum_{i=1}^{89} \ln x_i$$

$$\sigma_s = \sqrt{\frac{M_4 - M_2^2}{4n \times M_2}} \quad (n = 89)$$

$$M_4 = \frac{1}{89}\sum_{i=1}^{89}(\ln x_i - \overline{\ln x})^4 \qquad M_2 = \frac{1}{89}\sum_{i=1}^{89}(\ln x_i - \overline{\ln x})^2$$

Chia nhỏ miền (2) theo lưới đều sao cho miền (2) thành 400 miền nhỏ. Lấy giá trị ở các điểm lưới thế vào hàm mật độ (1) để tìm điểm (μ_0, σ_0) có tỷ sai nhỏ nhất.

Giai đoạn 2: Coi (μ_0, σ_0) vừa tìm được là ước lượng khâu ban đầu và tinh chỉnh bằng phương pháp dò tìm ngẫu nhiên. Xây dựng một toán tử tự thích nghi sao cho theo toán tử này từ ước lượng thô (μ_0, σ_0) ban đầu có thể tìm được những ước lượng (μ_1, σ_1), (μ_2, σ_2)... các ước lượng tiếp theo sẽ có tỷ sai nhỏ hơn ước lượng (μ_0, σ_0) và tới khi dừng máy ta tìm được ước lượng tối ưu (μ^*, σ^*) có tỷ sai nhỏ nhất. Vì dãy các ước lượng (μ_0, σ_0), (μ_1, σ_1)... là một quá trình tự thích nghi, sự hội tụ nghiệm đã được chỉ ra trong lý thuyết sau khi lựa chọn toán tử tự thích nghi ([2]):

$$X_0 = (\mu_0, \sigma_0) \qquad X_i = (\mu_i, \sigma_i)$$

$$X_{i+1} - X_i = \Delta X_{i+1}$$

$$X_{i+1} = \begin{cases} g\xi \text{ nếu } Q(X_i) \leq Q(X_{i-1}) \\ -X_i \text{ nếu } Q(X_i) > Q(X_{i-1}) \end{cases}$$

Ở đây $Q(X_i)$ là tỷ sai tại bước thứ i, g: bước lưới, ξ: đại lượng ngẫu nhiên có phân phối đều trên đoạn [0,1]. Việc lựa chọn này thỏa mãn các điều kiện hội tụ của định lý hội tụ Martingale. ([1]).

3. Kết luận:

Việc áp dụng phương pháp này để dò tìm (μ, σ) trong mật độ loga chuẩn của phân phối Q_{max} tại trạm Sơn Tây theo các quan sát thu được từ năm 1900 tới năm 1990 (89 quan sát) cho thấy sai số giảm hẳn. Bằng phương pháp moment thường dùng trong thống kê thủy văn, sai số của bài toán này từ 0,069 đã giảm xuống còn 0,051. Tuy nhiên, đây là kết quả bước đầu, nếu cải tiến thuật toán thêm một bước nữa thì sai số tiếp tục còn nhỏ hơn nữa.

TÀI LIỆU THAM KHẢO

1. Л. А. Растригин. Случайный поиск в процесах адаптации. Издателство Мир. Рига, 1973.

2. Nguyễn Hữu Bảo. Phương pháp dò tìm ngẫu nhiên với các quá trình tự thích nghi và ứng dụng thủy văn của nó. Tập san Khí tượng thủy văn 12-1994.