Chương 5

ƯỚC LƯỢNG THỐNG KÊ CÁC THAM SỐ PHÂN PHỐI CỦA ĐẠI LƯỢNG NGẪU NHIÊN

5.1. KHÁI NIỆM CHUNG

Một vấn đề rất quan trọng của phân tích thống kê là ước lượng các tham số phân phối được dùng để mô tả chuỗi đại lượng ngẫu nhiên được đem ra nghiên cứu.

Ở chương II khi mô tả các luật phân phối đã chỉ rõ các tham số đó là kỳ vọng toán (trị bình quân), khoảng lệch trung bình bình phương (hoặc hệ số biến đổi) và hệ số không đối xứng. Để mô tả các luật phân phối (thí dụ phân phối gama ba tham số và phân phối nhị thức dưới dạng tổng quát) phải sử dụng ba tham số; trong trường hợp sử dụng luật phân phối chuẩn biết kỳ vọng toán và khoảng lệch trung bình bình phương là đủ; luật Poatxông được dùng chỉ cần một tham số là kỳ vọng toán v .v... Trong thuỷ văn, khi tính dao động nhiều năm của dòng chảy, người ta thường dùng các đường phân phối có tham số độc lập không lớn hơn ba; trị bình quân, hệ số biến đổi và hệ số không đối xứng.

Trong tất cả các trường hợp, khi luật phân phối xác suất (dạng đường) được chọn ra phải xuất phát từ những nhận thức chung bao gồm những khoảng biến thiên của đại lượng ngẫu nhiên, tính không đối xứng của phân phối v.v... Một bài toán đặt ra là ước lượng các tham số nối với những điều kiện của chuỗi cụ thể cho trước. Rõ ràng là bài toán này chỉ có thể dựa vào lượng thông tin chứa trong tài liệu quan trắc thực tế yếu tố nghiên cứu của chế độ thuỷ văn. Cố nhiên là khi có chuỗi dài như thế nào đó, nhưng với những thời đoạn khác nhau có thể nhận được những giá trị tham số của luật phân phối đặc trưng cho yếu tố nghiên cứu của chế độ thuỷ văn là khác nhau. Nghĩa là bất kỳ giá trị nào của tham số cần tìm được tính theo mẫu hữu hạn luôn luôn có chứa yêú tố ngẫu nhiên. Giá trị ngẫu nhiên gần đúng đó được gọi là ước lượng của tham số. Thí dụ như tính ước lượng của kỳ vọng toán sẽ dùng là trị bình quân số học của những giá trị quan trắc được trong n lần thực hiện phép thử. Khi số lần xuất hiện phép thử rất lớn thì việc thay kỳ vọng toán bằng trị bình quân sẽ đưa đến ít sai số. Thông thường sai số này càng lớn khi số lần thực hiện phép thử n không lớn thì việc thay kỳ vọng toán bằng trị bình quân sẽ đưa đến những sai số.

Thường thường sai số này càng lớn khi số lần thực hiện phép thử càng nhỏ và hệ số biến đổi càng lớn. Vấn đề ước lượng các tham số chưa biết khác tương tự như vậy.

Tóm lại, bài toán đặt ra là sử dụng những mẫu tài liệu quan trắc các đặc trưng thuỷ văn tương đối ngắn để ước lượng các tham số của phân phối xác suất làm sao tốt nhất. Giải bài toán này phải dựa vào những luận chứng cơ bản của phương pháp chọn mẫu.

Lý thuyết mẫu có một ý nghĩa rất đặc biệt khi chỉnh lý tài liệu thuỷ văn quan trắc được, vì các tham số của tổng thể thường chưa biết trước mà xác định chúng phải dựa vào các mẫu thông thường có dung lượng không lớn.

Cần phải chú ý đến tình trạng là trong nhiều trường hợp, thường thường trong tính toán dòng chảy sông ngòi dung lượng mẫu không thể tăng được nữa do nó bị quyết định bởi thời gian quan trắc. Trong một số nghiên cứu chẳng hạn như sự biến động tốc độ của dòng chảy, khả năng đó về nguyên tắc có thể thực hiện được.

Khi sử dụng các phương pháp phân tích phân tích mẫu phải chú ý là các chuỗi thống kê đem nghiên cứu biểu diễn các giá trị đồng nhất về mặt định tính, thuộc cùng một tổng thể với các điều kiện đó có ý nghĩa đặc biệt trong tính toán thuỷ văn và phân tích tính đại biểu của chuỗi đại lượng thuỷ văn, nghĩa là đánh giá tính đại biểu của tài liệu mẫu so với tổng thể. Phép kéo dài các chuỗi thuỷ văn ngắn có một ý nghĩa quan trọng: bên cạnh sự phân tích tài liệu gốc một cách tổng quát, nó còn phải tiến hành trước những tính toán thống kê. Cấp chỉ tiêu đồng nhất đã được xét ở chương IV đều phải dựa vào lý thuyết mẫu.

5.2. NHỮNG YÊU CẦU CƠ BẢN ĐỐI VỚI ƯỚC LƯỢNG CÁC THAM SỐ CỦA PHÂN PHỐI.

Rõ ràng là ước lượng của tham số được coi là tốt nhất nếu nó gần đúng nhất với giá trị thực của tham số đó. Giả sử, ta xét hàm mật độ phân phối của xác suất của một dạng đã biết là $P(x, a_1, a_2, ... a_k)$ có chứa các tham số chưa biết là: $P(a_1, a_2, ... a_k)$ có chứa các tham só chưa biết là $p(a_1, a_2, ... a_k)$ và mẫu ngẫu nhiên có n số phối hợp với hàm phân phối $P(x, a_1, a_2, ... a_k)$ trên. khi đó $k \ll n$. Yêu cầu ước lượng những giá trị của các tham số chưa biết theo tài liệu mẫu. Ta sẽ làm rõ hơn định nghĩa tổng quát đó.

Ước lượng các tham số của tổng thể ta ký hiệu là $\tilde{a}_1, \tilde{a}_2 ... \tilde{a}_k$, như trên đã chỉ rõ, chúng đều phụ thuộc (làm hàm số) vào những tài liệu mẫu $\tilde{a}_1$ (x_1, x_2, ... x_n), $\tilde{a}_2$ (x_1, x_2, ... x_n) ... $\tilde{a}_k$ (x_1, x_2, ... x_n) vì vậy chính chúng là những giá trị ngẫu nhiên biến thiên từ mẫu này sang mẫu khác. Luật phân phối của $\tilde{a}_i$ phụ thuộc, thứ nhất là luật phân phối của x (thông htường vào chính tham số chưa biết a) thứ hai là vào số lần thực hiện phép thử n.

Hàm dạng $\tilde{a}_i$ (x_1, x_2, ... x_n) được dùng để xác định các giá trị mẫu của tham số nói chung có thể đưa ra rất nhiều. Thí dụ để xác định kỳ vọng toán ta có thể dùng trị bình quân số học hay trị bình quân hình học, số giữa các định vị khác nhau, nửa tổng của các số hạng cực đoạn của mẫu. Song các đặc trưng này tiện cận với giá trị chân thực phải tìm của thám số ở mức độ khác nhau, số lượng hàm có thể dùng để xác định những giá trị mẫu của tham số được lược bớt nếu đối với ước lượng các tham số có một số yêu cầu xác định được đặt ra

Điều kiện thứ nhất mà ước lượng [(1)] $\tilde{a}$ (x1, x2, ... xn) được xây dựng một cách hợp lý đối với tham số a cần phải thoả mãn, là sự hội tụ theo xác suất của ước lượng đến tham số càn tìm dùng lượng mẫu của tài liệu n tăng lên vô hạn, nghĩa là:

$$\tilde{a}\ (x1, x2, \dots xn) \underset{n \to \infty}{\longrightarrow} a \qquad (5\text{-}1)$$

Ta biết rằng đối với một đại lượng ngẫu nhiên (trong trường hợp này là $\tilde{a}$) được gọi là hội tụ theo xác suất về giá trị a nếu khi n tăng lên vô hạn phải thoả mãn với đẳng thức sau:

$$\lim P_{n \to \infty} \{ |\tilde{a} - a| < \varepsilon \} = 1 \qquad (5\text{-}2)$$

trong đó: ε - giá trị nhỏ tuỳ ý.

Những ước lượng thoả mãn với yếu cầu (5.1) và (5.2) được gọi là ước lượng đúng. Tính đúng đắn của ước lượng thống kê sẽ đảm bảo nó gần đúng thực tế (bất luận khi nào với n lớn hơn) với tham số phải tìm.

[(1)] Để đơn giản trong phân tích, ta hạn chế đối với trường hợp các mẫu được rút ra từ một chuỗi được mô tả bằng hàm mật độ xác suất P (Px,a) chỉ chứa mọt tham số chưa biết là a.

Thí dụ, ước lượng mẫu của trị bình quân và phương sai đều là ước lượng đúng. Từ tính chất đúng đắn của ước lượng ta vẫn chưa thể đi đến kết luận đầy đủ về sự thích hợp của nó đối với việc xác định gần đúng tham số khi n nhỏ, vì trong các điều kiện đó ước lượng gần đúng có thể chệch theo hướng này hay hướng khác so với giá trị chân thực của tham số.

Vì vậy, điều kiện thứ hai đặt ra đối với ước lượng thống kê là điều kiện không chệch nghĩa là với mọi n ước lượng không có sai số hệ thống. Đối với ước lượng không kỳ vọng của nó phải trùng với tham số của tổng thể.

$$M(\tilde{a}) = a. \qquad (5.3)$$

Ở đây đáng chú ý là kỳ vọng toán của ước lượng $\tilde{a}$ được xác định là tập hợp các mẫu của biến ngẫu nhiên x có dung lượng là N với bất kỳ số hạng n trong mẫu là hữu hạn (n < N). Ước lượng được gọi là chệch dương, nếu $M(\tilde{a}) > a$ và là lệch âm, nếu $M(\tilde{a}) < a$.

Yêu cầu không chệch của ước lượng là đặc biệt quan trọng, khi dung lượng mẫu (chuỗi tài liệu quan trắc) nhỏ đối với người làm công tác thuỷ văn thường gặp phải. Trị bình quân số học của mẫu không chỉ là ước lượng đúng mà còn là ước lượng không chệch của kỳ vọng toán.

$$M(\bar{x}) = \mu \qquad (5.4)$$

Biểu thức này được rút ra từ những quan hệ rõ ràng sau đây:

$M(x_i) = \mu$

và nghĩa là :

$$M(\bar{x}) = M\left(\frac{\sum_{i=1}^{n} xi}{n}\right) = \frac{M\left(\frac{n}{1} xi\right)}{n} = \frac{\sum_{i}^{n}[M(x_i)]}{n} = \frac{n\mu}{n} = \mu$$

Khi n nhỏ, giá trị mẫu của phương sai:

$$S^2 = \frac{\sum_{1}^{n}(x_i - \bar{x})^2}{n}$$ là ước lượng chệch âm của phương sai tổng thể.

$$M(S^2) = \frac{n-1}{n}\delta^2 \qquad (5.5)$$

Giá trị S^2 được hiệu chỉnh đối với giá trị chệch, vì vậy, nó là ước lượng không chệch, và có thể được xác định qua công thức (5.5) theo quan hệ:

$$S^2 = \frac{\sum_{i=1}^{n}(xi - \bar{x})^2}{n - i} \qquad (5.6)$$

biểu thức (5.6) được nhận từ các phép biến đổi sau đây:

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n}[(x_i - \mu) - (\bar{x} - \mu)]^2 =$$

$$\frac{1}{n}\sum_{i=1}^{n}[(x_i - \mu)^2 - 2(\bar{x} - \mu)(x_i - \mu) + (\bar{x} - \mu)^2] =$$

$$= \frac{1}{n}\left[\sum_{i=1}^{n}(x_i - \mu)^2 - 2(\bar{x} - \mu)\sum_{i=1}^{n}(x_i - \mu) + n(\bar{x} - \mu)^2\right] =$$

$$= \frac{1}{n}\left[\sum_{i=1}^{n}(x_i - \mu)^2 - 2n(\bar{x} - \mu)\frac{\sum_{i=1}^{n}(x_i - \mu)}{n} + n(\bar{x} - \mu)^2\right] =$$

$$= \frac{1}{n}\left[\sum_{i=1}^{n}(x_i - \mu)^2 - n(\bar{x} - \mu)^2\right] = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 - \left[\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)\right]^2$$

Trong trường hợp không có nối quan hệ giữa các số hạng của chuỗi gốc có thể viết:

$$M\{(xi - \mu)(x_{i+k} - \mu)\} = M(xi - \mu).M(x_{i+k} - \mu) = 0$$

Vậy khi chuyển phương sai về kỳ vọng toán ta được:

$$M(S^2) = \frac{1}{n}\sum_{i=1}^{n}M\left\{(x_i - \mu)^2 - \frac{1}{n^2}\right\}\sum_{i=1}^{n}M\{(x_i - \mu)^2\} = \sigma^2 - \frac{1}{n}\sigma^2 = \frac{n-1}{n}\sigma^2$$

Mối quan hệ vừa nhận được này là cơ sở của mối quan hệ (5.6) thường được dùng trong các " tài liệu hướng dẫn thuỷ văn " để ước lượng giá trị mẫu của phương sai khi n < 20. Khi n > 20 số hiệu chỉnh sự chênh lệch không lớn lắm cho nên nó không được xét.

Phương pháp trên đã khử được tính chệch của các giá trị mẫu của tham số bằng cách đưa vào giá trị hiệu chỉnh trong nhiều trường hợp nó là hợp lý hơn việc xây dựng những hàm phức tạp để nhận các giá trị mẫu đã bị tước bỏ các yếu tố chênh lệch. Phương pháp khử tính chệch bằng cách đưa vào những giá trị hiệu chỉnh về sau được sử dụng khi phân tích phương pháp thử thống kê để ước lượng giá trị mẫu của tham số.

Theo định nghĩa, hệ số biến đổi là tỷ số giữa khoảng lệch trung bình bình phương với trị bình quân $C_V = \sigma_x/x$. Ta biết rằng trị bình quân là ước lượng không chệch của kỳ vọng toán mà tính chệch của phép ước lượng khoảng lệch trung bình bình phương có thể khử theo công thức (5.6) ta có thể hy vọng là hệ só biến đổi được xác định không có sai số hệ thống. Song E.G.Blôkhinôv [18] bằng phân tích lý luận đã thấy rằng đối với trường hợpC_S=2C_V hệ số biến đổi không có sai số hệ thống có thể được xác định từ công thức gần đúng:

$$M(C_v) = C_v - \frac{C_v}{C_n}(1 + 3C_V^2) \qquad (5.7)$$

trong đó n - số số hạng của chuỗi; C_V - ước lượng không chệch của hệ số biến đổi; $M(C_V)$ - kỳ vọng toán của ước lượng chệch của hệ số biến đổi.

Kết luận trên, theo ý kiến của Blikhônôv, được rút ra từ điều kiện là tính không chệch của các biến số (x và σ_x), không quy định tính không chệch của hàm (C_V). Giá trị hiểu chỉnh tính chệch của ước lượng mẫu C_V, rút ra từ công thức (5.7) khi n > 20 chiếm 2 - 5 %, vì vậy thường không được xét đến. Tương tự như vậy, giá trị hiệu chỉnh Blôkhinôv nhận được để khử tính chệch âm của ước lượng hệ số không đối xứng có dạng:

$$\frac{n^2}{(n-1)(n-2)} \qquad (5.8)$$

Song giá trị hiệu chỉnh này không có độ chính xác cần thiết và được làm sáng tỏ trên cơ sở phương pháp thử thống kê ta sẽ phân tích sau đây:

Những ước lượng đúng, không chệch nhận được bằng những phương pháp khác nhau có thể có sự phân tán khác nhau. Thí dụ, đối với luật phân phối chuẩn của xác suất, (để ước lượng tâm phân phối (kỳ vọng toán) có thể sử dụng trị bình quân số học cũng như là số giữa thực nghiệm. Cả hai ước lượng đó đều đúng với và không chệch, song giá trị phân tán của chúng không bằng nhau. Vì vậy, cần phải đánh giá ước lượng mẫu của tham số, theo phương sai của chúng. Lúc này để làm chỉ tiêu hiệu của hai ước lượng $\tilde{a}_1$ và $\tilde{a}_2$ đối với cùng một tham số $\tilde{a}$ người ta lấy tỷ số:

$$e = \frac{D\tilde{a}_1}{D\tilde{a}_2} \tag{5.9}$$

Nếu e < 1, thì ước lượng $\tilde{a}_1$ hiệu quả hơn $\tilde{a}_2$ (và ngược lại) sự phân tích nhỏ hơn. Ước lượng đó $\tilde{a}_1$ mà puhơng sai của nó không vượt qúa giá trị lý luận nhỏ nhất có thể có được gọi là ước lượng hiệu quả có phương sai nhỏ nhất có thể có:

Thí dụ, ta đánh giá tính hiệu quả của việc ứng dụng chỉ tiêu để ước lượng kỳ vọng toán của tổng thể tuận theo luật chuẩn là trị bình quân số học và số giữa. Phương sai của những dao động ngẫu nhiên của trị bình quân số học đối với đại lượng ngẫu nhiên có n số hạng, được xác định bằng biểu thức:

$$D(\bar{x}) = \frac{D(x)}{n} \tag{5.10}$$

trong đó D(x) - phương sai tổng thể. Thật vậy, sau khi xác định phương sai và trị bình quân số học ta có:

$$D(\tilde{x}) = M\left[(\bar{x} - x_0)^2\right]$$
$$\bar{x} = \frac{1}{n}\sum_1^n x_i$$

Ở đây M - đàu kỳ vọng toán; x_0 - kỳ vọng của tổng thể.

Do x_i độc lập với nhau có thể viết:

$$D(\bar{x}) = M\left[\frac{1}{n_2}(\sum_1^n (x_i - x_0)^2\right] = \frac{1}{n^2}\sum_1^n M\left[(x_i - x_0)^2\right] = \frac{nD(x)}{n^2} = \frac{D(x)}{n}$$

Vì $D(x) = M\left[(x_i - x_0)^2\right]$

Như vậy, khoảng lệch trung bình bình phương (sai số ngẫu nhiên) của bình quân số học bằng khoảng lệch trung bình bình phương cả tổng chia cho $\sqrt{n}$

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} \tag{5.11}$$

Thay vào (5.11) $\sigma_x = C_{vx}\overline{X}$ ta nhận thấy được biểu thức thường dùng trong thuỷ văn:

$$\sigma_{\bar{x}} = \frac{C_{v_x}\,\overline{x}}{\sqrt{n}} \tag{5.12}$$

Sử dụng quan hệ (5.12) ta có thể xác định được số năm cần thiết n ta có thể xác định được số năm cần thiết n, để tính giá trị chuẩn của đại lượng nghiên cứu (thí dụ chuẩn dòng chảy) ứng với sai số ngẫu nhiên và những giá trị khác nhau của hệ số biến đổi C_V cho trước. Đối với tổng thể tuân theo luật phân phối chuẩn, phương sai của số giữa tính từ các mẫu có n số hạng được xác định theo công thức sau:

$$D(Me) = \sigma^2_{Me} = \frac{\pi}{2}.\frac{\sigma^2_x}{n} \tag{5.13}$$

Sử dụng các quan hệ (5.11) và (5.13) ta tính được tính hiệu quả của số giữa so với bình quân số học:

$$e(Me) = \frac{\sigma^2_{\bar{x}}}{\sigma^2_{Me}} = \frac{\sigma^2_{\bar{x}}}{n} \bigg/ \frac{\pi}{2}.\frac{\sigma^2_{\bar{x}}}{n} = \frac{2}{\pi} = 0{,}64$$

Hệ thức trên chỉ ra rằng đối với việc đánh giá tâm trung bình số học chỉ cần 64% lượng quan trắc ddể giải quyết bài toán qua trung vị.

Ở phần này sẽ được trình bày những thông báo tóm tắt về các phương pháp ước lượng dộ chính xác của các tham số mẫu, thường tuhòng về các phương pháp khử sự chệch nhận được bằng phân tích lý luận có giá trị minh chứng hơn hẳn vì chúng thuộc về các chuỗi tuân theo luật phân phối chuẩn hoặc phân phối gama với điều kiện $C_s = 2C_V$. Nghiên cứu đầy đủ hơn về độ chính xác của ước lượng tham số mẫu được thực hiện dựa vào phương pháp thử thống kê sẽ được trình bày ở mục 4 chương này.

5.3. CÁC PHƯƠNG PHÁP ƯỚC LƯỢNG THAM SỐ CỦA HÀM PHÂN PHỐI.

Khi giải bài toán thuỷ văn người ta thường sử dụng các phương pháp ước lượng tham số của phân phối sau đây:

1. Phương pháp mômen.
2. Phương pháp định vị
3. Phương pháp thích hợp tối đa

Phương pháp mômen được ứng dụng nhiều nhất trong thực tế tính toán thuỷ văn.Thực chất phương pháp này là các tham số cần tìm của phân phối được biểu diễn qua các mômen, mà ước lượng của chúng lấy bằng giá trị mômen của phân phối thực nghiệm thường thường có số hiệu chỉnh loại trừ sự chệch. Số mômen thực nghiệm được xác định theo tài liệu quan trắc được bằng tham số số của luật phân phối nghiên cứu. Như chúng ta đã biết độ tin cậy của ước lượng mômen thực nghiệm với dung lượng mẫu cho trước nếu số bậc của nó lớn thì nó cũng giảm. Vì vậy, trong thuỷ văn người ta thường không sử dụng các phân phối được xác định nhiều tham số (lớn hơn ba).

Các tham số của luật phân phối được biểu diễn qua các mômen thống kê bằng các công thức (1.10), (1.22) và (1.27). Những ước lượng tham số nhận được bằng phương pháp mômen đối với luật phân phối dùng trong thuỷ văn đều là những ước lượng đúng đối với các chuỗi không có mối quan hệ nội tại. Sự chệch của ước lượng không lớn có thể khử được bằng những giá trị hiệu chỉnh đơn giản.

Phương pháp định vị được dùng vào việc sử dụng mối quan hệ của các tham số mẫu với những giá trị của các định vị tương ứng. Trong tính toán thuỷ văn người ta thường ứng dụng loại bản đồ giải các phương pháp định vị do G,A,Alekxxev nghiên cứu [3,5]. Các phương pháp kỹ thuật sử dụng phương pháp này được nghiên cứu ở chương III.

Thực chất của **phương pháp thích hợp tối đa** như sau: đối với phép thử độc lập xác suất xuất hiện đồng thời giá trị x_1, x_2, x_n trong mẫu với các biên:

Đối với số hạng thứ nhất từ x_1 đến $x_1 + \Delta x_1$

Đối với số hạng thứ nhất từ x_2 đến $x_2 + \Delta x_2$

....

Đối với số hạng thứ n từ x_a đến $x_1 + \Delta x_2$

(đối với tham số có một tham số a)

$P = P_1 (x_1, a) P_2 (x_2, a) ... P(x_n, a) \Delta x_1 \Delta x_2 ... \Delta x_n$

Ở đây Pi $(x_i, a) \Delta x_i$ - xác suất của giá trị xi lấy bất kỳ trong mẫu mà rơi vào trong khoảng (xi - xi+1). Khi khoảng chia khá nhỏ, các xác suất này có thể xem gần như tỷ lệ với tung độ đường phân phối mật độ xác suất (P) được dùng để mô tả chuỗi nghiên cứu.

Bởi vì những giá trị mẫu $x_1, x_2, ... x_n$ của đại lượng ngẫu nhiên thực tế đã quan trắc được trong quá trình thí nghiệm, nếu giá trị đó của tham số a hay một tham số có thể lớn nhất, nếu xác suất này là lớn nhất, nghĩa là trong đó tính xác suất (hay tổng lôgarit của xác suất cũng như vậy thôi) của các giá trị quan trắc được đạt giá trị cực đại.

$$L ((x_1, x_2, ... x_n, a) = P_1 (x_1, a)B_2 (x_2, a) ... P_n (x_n,a) \qquad (5.14)$$

Biểu thức (5.14) được gọi là hàm thích hợp.

Như vậy, để làm ước lượng cho tham số chưa biết a người ta lấy giá trị nhận được từ phương trình

$$\frac{dL}{da} = 0 \qquad (5.15)$$

Chú ý là LnL và L đạt giá trị cực đại với cùng một giá trị a, phương trình (4.15) có thể thay bằng một phương trình thích hợp đơn giản hơn.

$$\frac{1}{L}\frac{dL}{da} = \frac{dLnL}{da} = \sum_1^n \frac{d.\ln P(xi, a)}{da} = 0 \qquad (5.16)$$

Phương pháp thích hợp tối đa được suy ra trực tiếp từ biểu thức (5.16) có uy thế lớn đối với những giá trị mẫu của đại lượng ngẫu nhiên, mà tương ứng với chúng là những xác suất lớn. Tính chất này của phương pháp thích hợp tối đa biểu hiện nhiều trong các phân phối đối xứng. Trong phương pháp mômen thống kê thì ngược lại, có uy thế đối với các số hạng cực đoan của mẫu và nhỏ đối với các số hạng giữa. Vì vậy phương pháp thích hợp tối đa sẽ cho những ước lượng tham số ổn định hơn so với những phương pháp mômen. Phương pháp thích hợp tối đa sẽ cho những ước lượng đúng, được phân phối tiệm cận với luật chuẩn có phương sai có thể có là nhỏ nhất so với các phân phối khác, cũng tiệm cậm với ước lượng chuẩn, và tốt nhất vì các ước lượng sử dụng lượng thông tin về các tham số chưa biết có trong mẫu. Những ước lượng tính theo phương pháp thích hợp tối đa có thể la bị chệch. Sự chệch này nhìn chung không lớn, có thể khử bằng cách đưa vào những giá trị hiệu chỉnh tương ứng.

Đối với luật phân phối chuẩn các ước lượng nhận được bằng phương pháp thích hợp tối đa trùng với các ước lượng mômen. Đối với luật phân phối các phương pháp này cho ta những kết quả thường là khác nhau.

Phương pháp thích hợp tối đa đã được X.N.Kriski - M.F.Menkel [68] dựa vào thực tế tính toán thuỷ văn. Trước khi sử dụng thực tế, phương pháp này đã được trình bày trong các công trình [20, 22] của E.G.Blôkhinôv.

Để làm thí dụ chúng ta sẽ xét việc ước lượng tham số của phân phối chuẩn bằng phương pháp thích hợp tối đa.

Như đã biết, luật phân phối chuẩn đối với mẫu có dung lượng bằng n có thể viết dưới dạng.

$$P(x_1, x_0, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x_i - x_0)^2} \qquad (5.17)$$

Chúng ta sẽ xây dựng hàm thch hợp (5.11) để có như vậy ta sơ bộ tìm lôgarit các giá trị P (x_i, x_0, σ)

$$\ln P(x_i, x_0, \sigma) = -\ln\sigma - \frac{1}{2}\ln 2\pi - \frac{1}{2\sigma^2}(x_i - x_0)^2 \qquad (5.18)$$

Cộng những giá trị ln P đối với tất cả những giá trị x_i quan trắc được, ta nhận được hàm thích hợp:

$$L = \sum_1^n \ln P(x_i, x_0, \sigma) = -\sum_1^n \ln\sigma - \frac{1}{2}\sum_1^n \ln 2\pi - \frac{1}{2\sigma^2}\sum_1^n (x_i - x_0)^2$$

Sau đó ta tìm phương trình thích hợp $\frac{dL}{dx_0} = 0$ đối với ước lượng tham số x_0

$$\frac{dL}{dx_0} = -\frac{1}{\sigma^2}\sum_1^n (x_i, x_0) = 0$$

từ đó đối với $\sum_1^n x_0 = nx_0$ ta nhận được

$$x_0 = \frac{\sum_1^n x_i}{n}$$

Như vậy, ước lượng thống kê của tham số x_0 (kỳ vọng toán) trong trường hợp này là trị bình quân số học của chuỗi những giá trị x_i

Để nhận được biểu thức ước lượng thống kê của tham số σ bằng phương pháp thích hợp tối đa chúng ta biến đổi phương trình thích hợp $\frac{dL}{d\sigma} = 0$

hay là

$$\frac{dL}{d\sigma} = \sum_1^n \frac{1}{\sigma} + \frac{1}{\sigma^3}\sum_1^n (x_i - x_0)^2$$

hay $$\frac{dL}{d\sigma} = \frac{1}{\sigma^3}\sum_1^n (x_i - x_0)^2 - \frac{n}{\sigma} = 0$$

từ đó ta nhận được

$$\sigma = \sqrt{\frac{\sum (x_i - x_0)^2}{n}} \qquad (5.19)$$

nghĩa là biểu thức của khoảng lệch trung bình bình phương.

Từ phân tích trên rút ra đối với luật phân phối chuẩn, ước lượng của tham số nhận được bằng phương pháp thích hợp tối đa và phương pháp mômen trùng nhau. Đối với các đường phân phối khác chưa chắc có sự trùng hợp hoàn toàn như vậy.

Để làm thí dụ thứ hai, ta xét việc ứng dụng phương pháp thích hợp tối đa về ước lượng tham số của đường phân phối nhị thức với $C_S = 20_V$ theo sự phân tích của X.N.Kriski và M.F.Menkel [68]. Như đã rõ ở chương I phương trình đường phân phối nhị thức với $C_S = 20_V$ có dạng:

$$P(xi, c; \gamma) = \frac{1}{a^\gamma}\frac{\gamma^\gamma}{\Gamma(\gamma)} e^{-\frac{\gamma}{a} x_{i x_i} . j-1} \qquad (5.20)$$

Theo trình tự phân tích như trên ta có:

$$P(x_i, a, \gamma) = -\gamma \ln a + \gamma \ln \gamma - \ln\Gamma(\gamma) - \frac{\gamma}{a} x_i + \gamma \ln x_i - \ln x_i \quad (5.21)$$

Trong trường hợp này hàm thích hợp (5.18) sẽ có dạng:

$$L = \sum_1^n \ln(xi; a; \gamma) = -\sum_1^n \gamma \ln a + \sum_1^n \gamma \ln \gamma - \sum_1^n \ln a\Gamma(\gamma) - \sum_1^n \frac{\gamma}{a} x_i + \sum_1^n \gamma \ln x_i - \sum_1^n \gamma \ln x_i$$

Phương trình thích hợp tương ứng đối với ước lượng tham số a ta viết được dưới dạng:

$$\frac{dL}{da} = -\gamma \sum_1^n \frac{1}{a} - \sum_1^n \frac{\gamma}{a^2} x_i = 0$$

từ đó rút ra

$$\sum_1^n \frac{\gamma}{a^2} x_i = \frac{\gamma n}{a}$$
$$\frac{1}{a} \sum_1^n x_i = n \qquad (5.22)$$

hay là $a = \frac{\sum_1^n x_i}{n}$ (5.23)

Từ phương trình (5.23) suy ra là ước hiệu quả nhất của tham số a trong phương trình đường phân phối nhị thức với $C_S = 2C_V$ là trị bình quân số học của chuỗi thống kê.

Để ước lượng tham số γ phương trình thích hợp ta sẽ có dạng:

$$\frac{dL}{d\gamma} = -\sum_1^n \ln e + \sum_1^n (\ln \gamma + 1) - \sum_1^n \frac{a\Gamma\gamma}{d\gamma} - \sum_1^n \frac{xi}{a} - \sum_1^n \ln x_i = 0$$

Từ đó:

$$-n \ln a + n(\ln \gamma + 1) - n \frac{d\Gamma(\gamma)}{d\gamma} = \sum_1^n \frac{x_i}{a} - \sum_1^n \ln x_i$$

Theo đẳng thức (5.22) ta nhận được:

$$\frac{a\Gamma(\gamma)}{d\gamma} - \ln\gamma = \frac{\sum_{1}^{n}\ln x_i}{n} - \ln a \qquad (5.24)$$

Biểu thức vế phải của đẳng thức (5.24) có thể viết dưới dạng:

Như vậy, cuối cùng ta có:

$$\frac{1}{n}\sum_{1}^{n}\ln\frac{x_i}{x} = \frac{a}{a}\ln\Gamma(\gamma) - \ln\gamma \qquad (5.25)$$

Từ biểu thức (5.25) suy ra ước lượng hiệu quả nhất của tham số γ trong phương trình đường phân phối nhị thức khi $C_S=2C_V$ là trị bình quân mẫu của chuỗi logarit hệ số môđul đại lượng nghiên cứu.

Giữa giá trị $\lambda = \frac{1}{n}\sum_{1}^{n}\ln\frac{x_i}{\overline{x}}$ và hệ số biến đổi C_V có mối quan hệ hàm số quy định bằng phương trình (5.25) vì rằng tham số $\gamma = \frac{1}{C_V^2}$

Các giá trị của hệ số biến đổi phụ thuộc vào:

$\lambda = \frac{1}{n}\sum_{1}^{n}\lg\frac{x_i}{\overline{x}}$ theo phương trình (5.25) đã được trình bày dưới dạng bảng có trong công trình [22] với điều kiện $C_S = 2C_V$

Theo đúng với lược đồ xác định ước lượng các tham số bằng phương pháp thích hợp tối đa đối với các luật phân phối chuẩn và nhị thức (với $C_S=2C_V$) ta sẽ xét (theo sự nghiên cứu của Blôkhinôv) [22] Phương pháp xác định những ước lượng đó đối với luật phân phối gama ba tham số. Mật độ phân phối xác suất của luật phân phối này được viết như sau:

$$P(x,\overline{x}_0,\gamma,b) = \left[\frac{\Gamma(\gamma+b)}{\Gamma(\gamma)}\right]^{\gamma/b} \cdot \frac{1}{|b|\Gamma(\gamma)x_0^{\gamma/b}} x^{\gamma/b}$$
$$x.\exp\left\{\left[\frac{\Gamma(\gamma+b)}{\Gamma(\gamma)}\cdot\frac{x}{x_0}\right]^{1/b}\right\} \qquad (5.26)$$

Trong phương trình này x_0 - trị bình quân; γ và b - các tham số có quan hệ với hệ số biến đổi C_V và hệ số không đối xứng C_S bằng phương trình sau đây:

$$C_V = \left[\frac{\Gamma(\gamma)\Gamma(\gamma+2b)}{\Gamma^2(\gamma+b)} - 1 \right]^{1/2} \tag{5.27}$$

$$C_S \frac{\dfrac{\Gamma^2(\gamma)\Gamma(\gamma+3b)}{\Gamma^3(\gamma+b)} - 3\dfrac{\Gamma(\gamma)\Gamma(\gamma+2b)}{\Gamma^2(\gamma+b)} + 2}{\left[\dfrac{\Gamma(\gamma)\Gamma(\gamma+2b)}{\Gamma^2(\gamma+b)} - 1 \right]^{3/2}} \tag{5.28}$$

trong đó $\Gamma(\gamma)$ và $\Gamma(\gamma + b)$ v . v... là ký hiệu hàm gama của các đối số tương ứng.

Từ phương trình (5.26) ta đều rõ luật phân phối nghiên cứu được quy định bởi ba tham số: trị bình quân x_0, hệ số biến đổi C_V và hệ số không đối xứng C_S. Yêu cầu xác định những ước lượng của các tham số đó bằng phương pháp thích hợp tối đa.

Ta biến đổi hàm thích hợp (5.14) bằng cách tìm logarit của các giá trị $Pi(x_i, x_0, \gamma, b)$

$$\ln P(x_i, x_0, \gamma, b) = (\frac{\gamma}{b} - 1)\ln\frac{x_i}{x_0} - \left[\frac{\Gamma(\gamma+b)x_i}{\Gamma(\gamma)x_0} \right]^{1/b} + \\ \frac{\gamma}{b}\ln\frac{\Gamma(\gamma+b)}{\Gamma(\gamma)} - \mathrm{Ln}|b| - \mathrm{Ln}(\gamma) - \mathrm{Ln}(x_0) \tag{5.29}$$

Trong trường hợp này hàm thích hợp (dưới dạng logarit) có dạng:

$$\ln L = \sum_1^n \ln P(x_i) = (\frac{\gamma}{b} - 1)\sum_1^n \ln\frac{x_i}{x_0} - \left[\frac{\Gamma(\gamma+b)}{\Gamma(\gamma)} \right]^{1/b} \sum_1^n (\frac{x_i}{x_0})^{1/b} \\ + \frac{\gamma}{b}\ln\frac{\Gamma(\gamma+b)}{\Gamma(\gamma)} - \ln|b| - \mathrm{Ln}\Gamma(\gamma) - \mathrm{Ln}x_0$$

Để tiếp tục phân tích được tiện lợi hơn ta sử dụng trị bình quân hình học của mật độ xác suất mẫu đem nghiên cứu nghiã là giá trị L = (L).1/n, trong đó n - dung lượng mẫu. Phép biến đổi đó về mặt nguyên tắc không làm thay đổi ước lượng cần tìm của các tham số. Với chú thích trên, phương trình (5.29) có thể viết dưới dạng:

$$\mathrm{LnL}' = (\frac{\gamma}{b} - 1)\frac{\sum_1^n \ln\frac{x_i}{x_0}}{n} - \left[\frac{\Gamma(\gamma+b)}{\Gamma(\gamma)} \right]^{1/b} \frac{\sum_1^n (\frac{x_i}{x_0})^{1/b}}{n} + \\ + \frac{\gamma}{b}\ln\frac{\Gamma(\gamma+b)}{\Gamma(\gamma)} - \mathrm{Ln}|b| - \mathrm{Ln}\Gamma(\gamma) - \mathrm{Ln}x_0 \tag{5.30}$$

Để nhận được các quan hệ cho phép ta xác định các tham số mà ta quan tâm (x_0, γ, b) bằng phương pháp thích hợp tối đa, ta vi phân hoá phương trình (5.30).

Để nhận được các quan hệ cho phép ta xác định các tham số mà ta quan tâm (x_0, γ, b) bằng phương pháp thích hợp tối đa ta vi phân hoá phương trình (5.30) lần lượt theo x_0, γ và b, nghĩa là tìm đạo hàm lôgarit các hàm thích hợp $\frac{\partial LnL'}{\partial x_0}$; $\frac{\partial LnL'}{\partial \gamma}$; $\frac{\partial LnL'}{\partial_b}$. Khi cho các đạo hàm lôgarit đó bằng không (0) ta có thể nhận được hệ phương trình sau đây:

$$\left.\begin{aligned} &\left[\frac{\Gamma(\gamma+b)}{\Gamma(\gamma)}\right]^{1/b}\lambda 1-\gamma=0 \\ \gamma_2+\mathrm{Ln}\,&\frac{\Gamma(\gamma+b)}{\Gamma(\gamma)}-b\frac{\partial}{\partial\tau}\mathrm{Ln}\,\Gamma(\gamma)=0 \\ &\gamma\lambda_2\left[\frac{\Gamma(\gamma+b)}{\Gamma(\gamma)}\right]^{1/b}\lambda_3+b=0 \end{aligned}\right\} \quad (5.31)$$

$$\lambda_1=\frac{\sum_1^n\left(\frac{x_i}{x_0}\right)^{1/b}}{n}$$

$$\lambda_2=\frac{\sum_1^n\frac{x_i}{x_0}}{n}$$

$$\lambda_3=\frac{\sum_1^n\left(\frac{x_i}{x_0}\right)^{1/b}\ln\frac{x_i}{x_0}}{n}$$

Giải hệ (5.31) đối với x_0, γ, b đối với mẫu cho trước (x_1, x_2, ... x_n) mà các giá trị λ_1, λ_2 và λ_3 được tính theo mẫu này ta sẽ nhận được ước lượng tốt nhất của các tham số đó.

Thay vào các biểu thức (5.27) và (5.28) ước lượng của tham số γ và b được tính theo hệ các phương trình (5.31) ta nhận được những ước lượng tốt nhất của hệ số biến đổi C_V và hệ số không đối xứng C_S. Rõ ràng là phương pháp ước lượng các tham số x_0, C_V, C_S này mặc dù về mặt nguyên tắc rất chặt chẽ, song có khó khăn khi giải phương trình siêu việt (5.31) mà ở đây chúng ta chỉ hạn chế là trình bày theo lược đồ tổng quát chứ không đưa ra các phép biến đổi toán học một cách chi tiết. Điều đó rất cần thiết như trong công trình [22] E.G.Blôkhinôv dựa vào nguyên tắc thích hợp tối đa đã nghiên cứu các phương pháp đơn giản ước lượng tham số của phân phối gama ba tham số. Ước lượng của các tham số nhận được theo phương sai giản đơn không mấy khác só với các ước lượng bằng lược đồ đầy đủ.

Phương pháp đơn giản đã được Blôkhinôv nghiên cứu theo 2 phương án:

a/ Cho việc xác định đồng thời và độc lập tất cả ba tham số (x_0, C_V, C_S) theo các giá trị thống kê được tính theo tài liệu quan trắc được.

b/ Cho ước lượng hệ số biến đổi C_V với những quan hệ C_S/C_V được quy định tuỳ ý.

Với mục đích đơn giản hơn, lược đồ đơn giản cơ bản Blôkhinôv đã thay thế giá trị thống kê λ_1, λ_2 và λ_3 bằng các giá trị thống kê sau:

$$\lambda'_1 = \frac{\sum_1^n x_i}{n}$$

$$\lambda'_2 = \frac{\sum_1^n \ln \frac{x_i}{x}}{n}$$

$$\lambda'_3 = \frac{\sum_1^n \frac{x_i}{\overline{x}} \ln \frac{x_i}{\overline{x}}}{n}$$

Sử dụng giá trị đầu tiên của các tham số thống kê này nghĩa là lấy trị bình quân số học $\overline{x}$ làm ước lượng của tham số x_0 người ta sử dụng trị bình quân mũ. Ở mục bài 2 đã chỉ rõ ước lượng trị bình quân như vậy là rất hiệu quả đối với nhiều luật phân phối kể cả phân phối gama ba tham số. Giá trị thống kê thứ hai $x\lambda_2$ như đã thấy phù hợp hoàn toàn với ước lượng của nó được rút ra từ lược đồ đầy đủ ứng dụng phương pháp thích hợp tối đa đó xác định các tham số của phân phối gama ba tham số. Giá trị thống kê thứ ba λ_2 trong phương án nghiên cứu của phương pháp đơn giản hợp với giá trị thứ ba λ_3 của lược đồ đầy đủ với b = 1 hay chính $C_S = 2C_V$

Do tính chất trên của tham số thống kê thứ hai, mối quan hệ của nó với hệ số biến đổi người ta có thể biểu diễn được bằng phương trình của hệ (5.31) xây dựng cho lược đồ đầy đủ.

$$\lambda_2 = b\frac{\partial}{\partial\gamma}.\ln\Gamma(\gamma) - \ln\frac{\Gamma(\gamma+b)}{\Gamma(\gamma)} \qquad (5.32)$$

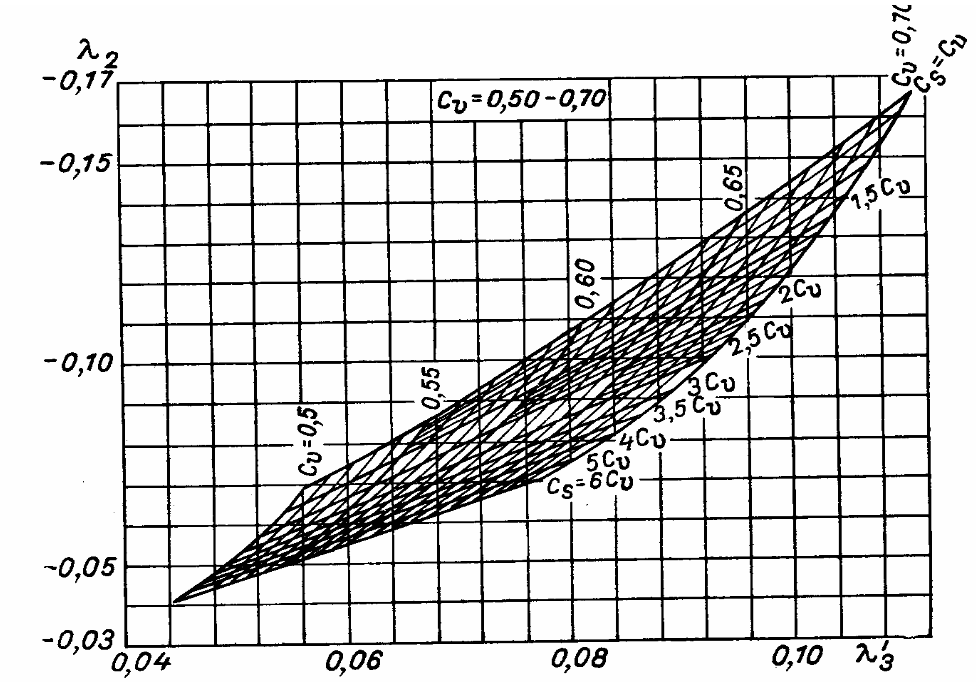
Đối với quan hệ mới của giá trị thống kê thứ ba λ_3 với hệ số không đối xứng Blôkhinôv đã nhận được phương trình:

$$\lambda_3 = b\frac{\partial}{\partial\gamma}.\ln\Gamma(\gamma+b) - \ln\frac{\Gamma(\gamma+b)}{\Gamma(\gamma)} \qquad (5.33)$$

Nghiên cứu đồng thời các phương trình (5.32) và (5.33) cho phép tạo được một hệ

$$\left.\begin{aligned} & b\frac{\partial}{\partial\gamma}\ln\Gamma(\gamma) - \ln\frac{\Gamma(\gamma+b)}{\Gamma(\gamma)} - \lambda_2 = 0 \\ & b\frac{\partial}{\partial\gamma}\ln\Gamma(\gamma+b) - \ln\frac{\Gamma(\gamma+b)}{\Gamma(\gamma)} - \lambda'_2 = 0 \end{aligned}\right\} \qquad (5.34)$$

Có khả năng với các tham số thống kê λ_2 và λ_3 tính được theo chuỗi tài liệu



Hình 5.1 Toán đồ xác định tham số C_V và C_S của phân phối gama ba tham số bằng phương pháp thích hợp tối đa.

quan trắc, tìm được các giá trị của tham số γ và b. Khi đã biết được giá trị γ và b, theo phương trình (5.27) và (5.28) có thể ước lượng được các tham số C_V và C_S. trong khi đơn giản hoá lược đồ sử dụng phương pháp thích hợp tối đa như trên lời giải của hệ phương trình (5.34) đối với từng mẫu chỉ có thể nhận được bằng phép lựa chọn, cho nên khối lượng tính toán rất lớn. Vì để có khả năng sử dụng phương pháp này trong thực tế Blôkhinôv đã xây dựng các toán đồ. Trục hoành của các toán đồ này là trục của các giá trị λ_3 còn trục tung là của các giá trị λ_3. Điểm cắt nhau của các giá trị λ_2 và λ_3 xác định giá trị của các tham số cần tìm C_V và C_S.

Các toán đồ đã xây dựng cho C_V từ 0,25 đén 1,5. Song vì ước lượng bằng phương pháp thích hợp tối đa và bằng phương pháp mômen C_V <<0,5 thực tế không khác nhau cho nên trên hình 5.1 chỉ trình bày các toán đồ với C_V =0,5 + 1,5; và C_S/C_V = 1+6.

Ý nghĩa sử dụng phương pháp đơn giản ước lượng các tham số của phân phối gama ba tham số rút ra như sau. Như đã rõ m trong hệ các phương trình (5.31) các giá trị thông kê:

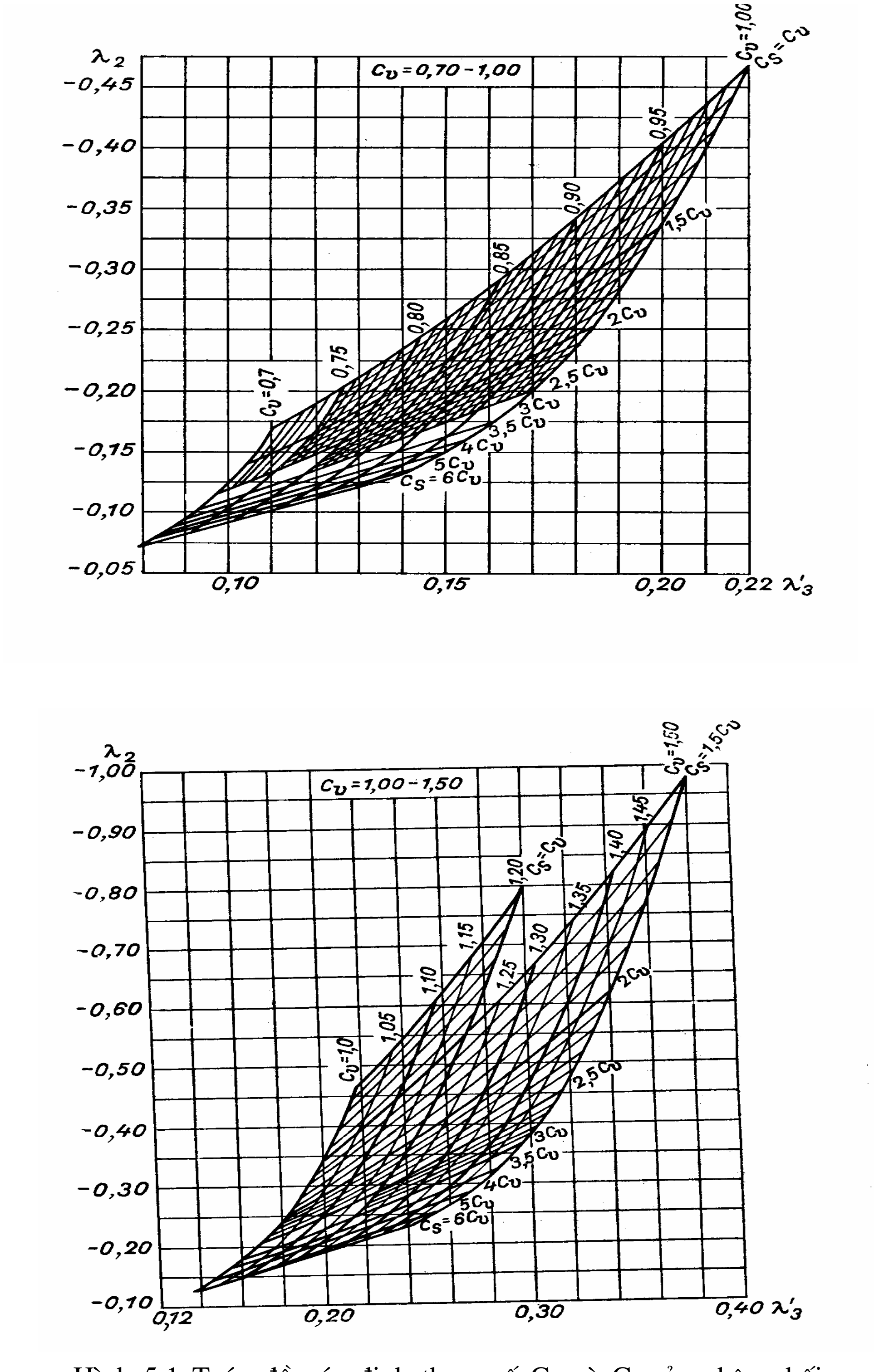
$$\lambda_1 = \frac{\sum_1^n \left(\frac{-x_i}{x_0}\right)^{1/b}}{n} \qquad \text{và} \qquad \lambda_3 = \frac{\sum_1^n \left(\frac{x_i}{x_0}\right)^{1/b} \ln\frac{x_i}{x_0}}{n}$$

có quan hệ với nhau và có quan hệ với các tham số cần tìm (x_0, C_V, C_S) thông quan tham số trung gian b. Điều đó làm rất phức tạp cho việc giải hệ phương trình nghiên cứu và làm cho nó ít thích ứng với việc sử dụng vào thực tế ngay cả khi dùng MTĐT.

Việc sử dụng các tham số thống kê về mặt cấu trúc của mình gần giống như các tham số thống kê dùng trong phương pháp thích hợp tối đa đầy đủ, nhưng đã khử đi mối quan hệ bên trong như trên để cho phép Blôkhinôv nhận được một hệ phương trình (5.34) đơn giản hơn, được mô tả bằng các toán đồ 5.1.

Phương án thứ hai của phương pháp đơn giản khong dựa vào các phương trình (5.27) liên kết giữa hệ số biến đổi C_V với tham số thống kê λ_2 mà dựa vào phương trình (5.32). Thực chất của việc thay thế là ở chỗ phương trình (5.27) liên kết các giá trị đó đối với trường hợp riêng C_S=2C_V còn phương trình (5.32) cố định mối quan hệ đó, cho quan hệ tuỳ ý giữa các tham số C_V và C_S cho nó được suy ra từ lời giải tổng quát trở lên ta có:

$$b\frac{\lambda\partial}{\partial\gamma}\text{Ln}\Gamma(\gamma) - \text{en}\frac{\Gamma(\gamma+b)}{\Gamma(\gamma)} - \lambda_2 = 0 \qquad (5.35)$$



Hình 5.1 Toán đồ xác định tham số C_V và C_S của phân phối gama ba tham số bằng phương pháp thích hợp tối đa.

trong đó $\lambda_2 = \dfrac{\sum_{1}^{n} \ln \dfrac{x_i}{x}}{n}$ giá trị thống kê được tính toán mẫu

Bảng 5.1 Tính các tham số $\overline{Q}, C_v, C_s$ *bằng phương pháp thích hợp tối đa chuỗi lưu lượng nước cực đại sông Tixư trạm Delovôi.*

Số thứ tự	Năm	Q_{max} (m^3/s)	Qmax trong trật tự giảm dần	$ki \frac{Q_i}{\overline{Q}}$	lgk		k.lgk
1	2	3	4	5	6	7	8
1	1933	490	745	2,35	0,37107		0,872
2	1934	293	700	2,21	0,34439		0,760
3	1935	405	518	1,63	0,21219		0,346
4	1936	240	504	1,59	0,10140		0,320
5	1937	215	490	1,55	0,19033		0,295
6	1938	240	460	1,46	0,16435		0240
7	1939	146	462	1,46	0,10435		0,240
8	1940	282	408	1,29	0,11059		0,143
9	1941	504	405	1,28	0,10721		0,137
10	1945	338	338	1,07	0,02938		-0,031
11	1946	162	310	0,978	1,99034	-0,00966	-0,0094
12	1947	462	310	0,978	1,99034	-0,00966	-0,0094
13	1948	700	293	0,924	1,96567	-0,03433	-0,0317
14	1949	462	293	0,294	1,96567	-0,03433	-0,0317
15	1950	225	282	0,890	1,94939	-0,05061	-0,0650
16	1951	235	265	0,836	1,92201	-0,07779	-0,0650
17	1952	518	252	0,795	1,90037	-0,09963	-0,0792
18	1953	171	240	0,757	1,87910	-0,02090	-0,0915
19	1954	220	240	0,757	1,87910	-0,18090	-0,0915
20	1955	293	240	0,757	1,87910	-0,12090	-0,0915
21	1956	265	235	0,741	1,86982	-0,13018	-0,0965
22	1957	228	228	0,719	1,85673	-0,14327	-0,103
23	1958	310	225	0,710	1,85126	-0,14847	-0,106
24	1959	408	220	0,694	1,84136	-0,15864	-0,110
25	1960	173	215	0,678	1,83123	-0,16877	-0,114
26	1961	159	173	0,546	1,73719	-0,26281	-0,143
27	1962	252	171	0,539	1,73159	0,26841	-0,145
28	1963	146	162	0,511	1,70842	-0,29158	-0.149
29	1964	310	159	0,502	1,70070	-0,29930	-0,150
30	1965	240	146	0,461	1,66370	-0,33638	-0,155
31	1966	745	146	0,461	1,66370	-0,33630	-0,155

n= 31 năm, Q = 317 m^3/s Σlgk = - 1,328 Σklgk = 1,412

$$\lambda_2 = \frac{\sum \lg k}{n-1} = \frac{-1,328}{30} = -0,044; \lambda_3 = \frac{\sum k \lg k}{n-1} = \frac{1,412}{30} = 0,047$$

$$\gamma = \frac{1}{C_V^2}$$

Quan hệ (5.32) không thể giải trực tiếp với tham số C_V mà có thể biểu diễn dưới dạng [22] tính bằng phương pháp lựa chọn theo các điều kiện được viết dưới dạng phương trình (5.35)

Ta sẽ nghiên cứu trình tự sử dụng phương pháp giản đơn ước lượng các tham số của phân phối gama ba tham số. Việc tính toán các tham số của chuỗi lưu lượng lớn nhất nước sông Tixa - trạm Đêlôvôi (r = 1190km^2) bằng phương pháp thích hợp tối đa và được trình bày trong bảng 5.1.

Dựa vào các tham số thống kê λ_2 và λ_3 theo toán đồ (hình 5.1) ta xác định được $C_V = 0{,}51$; $C_S = 5C_V = 2{,}55$. Theo tham số thống kê λ_2 với các giá trị bảng 5.1 ta có thể xác định được C_V nếu như chấp nhận quan hệ cố định nào đó giữa C_S và C_V (C_S/C_V). Vì trong trường hợp này ta tính toán lưu lượng lớn nhất của lũ do mưa lấy $C_S = 4C_V$ với $\lambda_2 =0{,}044$ ta có $C_V = 0{,}49$.

5.4. ỨNG DỤNG PHƯƠNG PHÁP THỐNG KÊ ĐỂ ƯỚC LƯỢNG CÁC THAM SỐ CỦA PHÂN PHỐI.

Phương pháp thử thống kê (phương pháp Mônte - Carlô) được ứng dụng nhiều nhất trong tính toán thuỷ văn. Việc tính toán theo phương pháp này rất thích hpựp trên máy tính điện tử (MTĐT). Trong tài liệu thuỷ văn Xô Viết X.N.Kriski và MF.Menkel là những người đầu tiên nghiên cứu những lý luận cơ bản có tính nguyên tắc của việc mô hình.

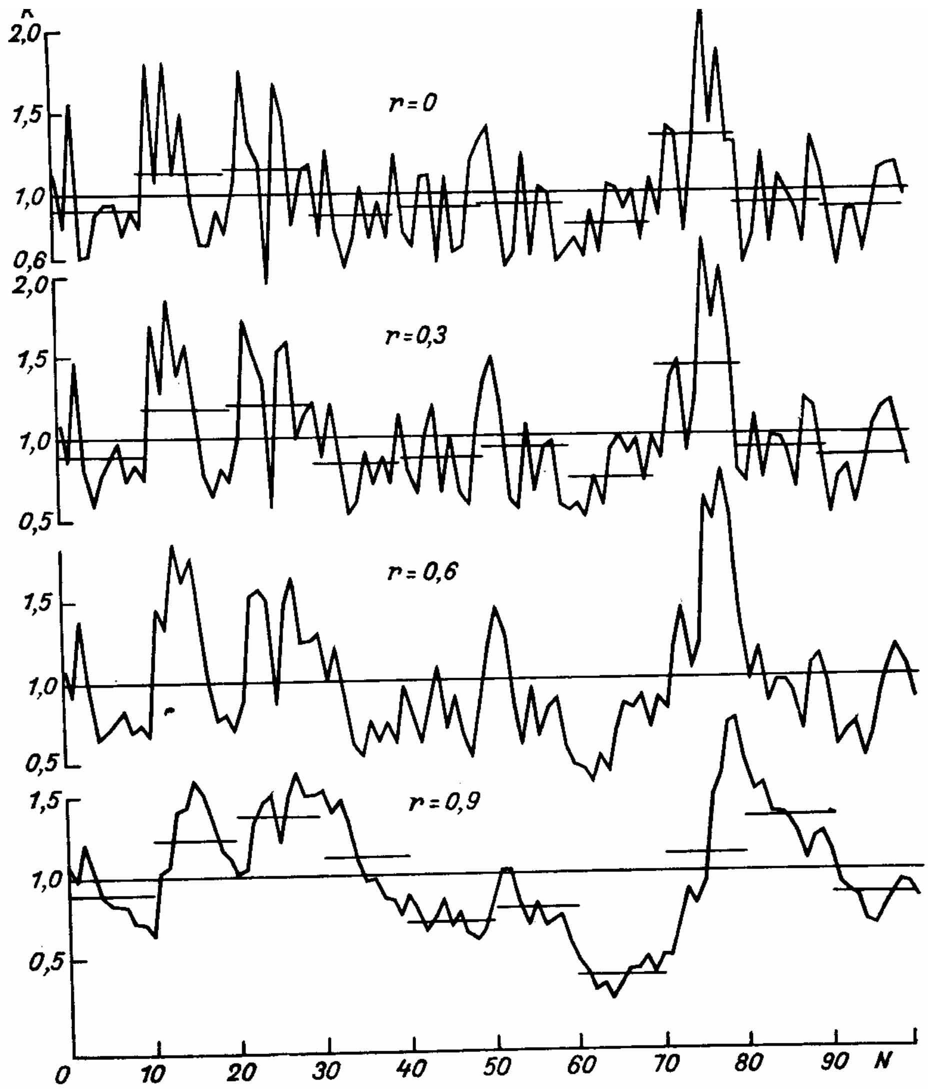
Tính các tham số Q ==, C_V và C_S của chuỗi dòng chảy lớn nhất nước sông Tixa trạm bằng phương pháp thích hợp tối đa.

hoá thống kê, song việc ứng dụng rộng rãi phương pháp mà trong tuyến tính dòng chảy sông ngòi được bắt đầu từ những công trình nghiên cứu của G.G.Xynidze [122]. Nghiên cứu những năm ít ưnớc và nhiều nước bằng phương pháp thống kê đã được V.G.Andreianôv và các tác giả khác[1] dùng đối với các chuỗi thuỷ văn. Về sau hướng này đã được Đ.Ta.Ratkôvich [95 - 97] phát triển E.G Blôkhinôv [18, 20, 22] và A.S.Rênhicôvski [37] dã sử dụng phương pháp thử thống kê đó xác định phân phối của ước lượng mẫu của các tham số phân phối (trị bình quân, hệ số biến đổi và hệ số không đối xứng).

Trong các công trình nghiên cứu của Blôkhinôv đặc biệt chí ý đến việc so sánh phân phối của các ước lượng mẫu tĩnh theo phương pháp mômen và phương pháp thích hợp tối đa, còn trong công trình của Rernicôvski nghiên cứu ảnh hưởng của các mức quan hệ của các chuỗi thuỷ văn đến các ước lượng mẫu của các tham số phân phối. Ngoài ra Blôkhinôv [20] đã nghiên cứu những đặc điểm phân phối của các tung độ đường tần suất gama ba tham số với ước lượng mẫu của hai tham số (x, C_V và C_S). Khi đó như trên đã nói đối với ước lượng mẫu của tham số Blôkhinôv được tiến hành so sánh những phân phối của ước lượng mẫu tung độ đường tần suất nhận được bằng phương pháp mômen và phương pháp thích hợp tối đa. Để đánh giá so sánh như vậy với 100 mẫu là hoàn toàn đủ. Song để giải thích những sai số ngẫu nhiên của tung độ đường tần suất không thể thực hiện dưới dạng so sánh hai phương pháp tính chúng mà ở dạng các giá trị tuyệt đối; số lượng như vậy vẫn chưa đủ. Để nhận được sai số ngẫu nhiên của đại lượng nghiên cứu, kết luận như vậy phải dựa vào số lượng mẫu rất lớn. Vì vậy, Z.F.Vôlkôva [39] đã tiến hành đánh giá sai số ngẫu nhiên của tung độ đường phân phối gama với số lượng mẫu lên đến 1000. Bà đã sử dụng các mẫu có dung lượng khác nhau (n=25,5 và 100) để tiến hành nghiên cứu với quan hệ $C_S = 2C_V$. Ước lượng của các tham số được tính toán theo phương pháp mômen và phương pháp thích hợp tối đa.

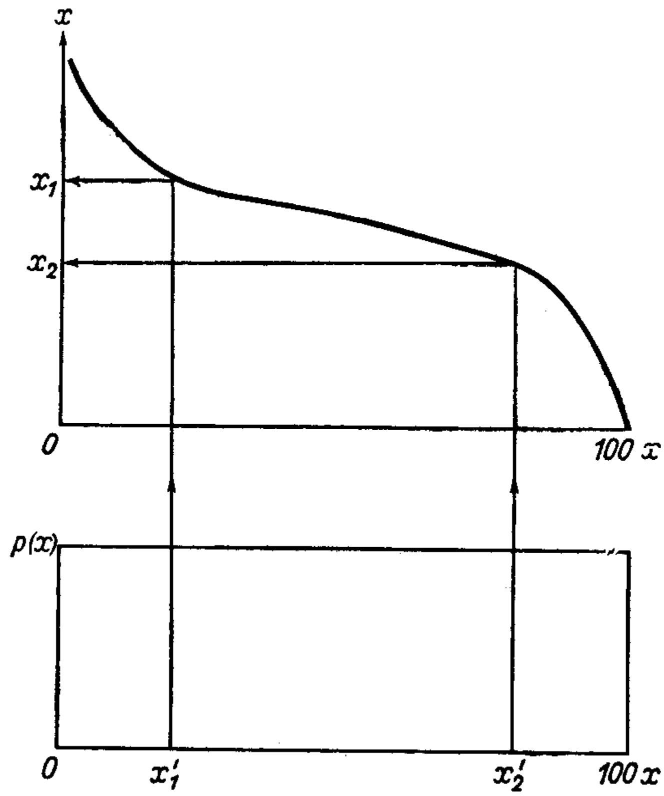
Những công trình nghiên cứu của Blôkhinôv, Rêmicô và các tác giả khác đều dựa vào những phương pháp thử thống kê có ý nghĩa rất quan trọng về thực tế và phương pháp luận, có ý nghĩa rất quan trọng, nhưng bài toán làm sáng tỏ độ chính xác ước lượng thống kê các tham số của luật phân phối đại lượng ngẫu nhiên chỉ giải được đối với những trường hợp riêng. Thí dụ như Blôkhinôv đã nghiên cứu các chuỗi không có mối quan hệ nối tại giữa các số hạngRênicôvski có xét đến mối tương quan nội tại của chuỗi nhưng đối với trường hợp riêng r = -,3 và 0,5 và quan hệ $C_S = 2C_V$. Khi nghiên ưúu ta thấy rằng mối quan hệ nội tại của chuỗi có ảnh hưởng rất lớn đến độ chệch của ước lượng tham số, và đặc biệt đối với những dao động ngẫu nhiên của chúng, cho nên cần phải tiến hành nghiên cứu toàn bộ những ảnh hưởng của tham số này ($r_{i, i+1}$) đối với các luật phân phối của các đặc trưng thuỷ văn.

Hậu quả quan trọng của mối quan hệ nội tại chuỗi là làm giảm đi có hệ thống (chệch âm) những ước lượng mẫu của các tham số trong chuỗi thống kê. Sự ảnh hưởng của mối quan hệ nội tại chuỗi đến cấu trúc của các chuỗi hình hoá được minh hoạ trên hình 5.2. Khi mô hình hoá để làm tham số gốc người ta lấy x = 1,0; C_V=0,3; C_S = 0,6 $r_{i,i+1}$=0,0; 0,3; 0,6; 0,9. Dung lượng chung của chuỗi N = 100, và các mẫu n = 10. sự có mặt của mối tương quan nội tại giữa các số hạng kề nhau của chuỗi đã hạ thấp tính đại biểu của mẫu số trị cho trước và sẽ nhận được những giá trị tham số chệch so với các tham số tổng thể.



Hình 5.2 Các chuỗi mô hình hoá bằng phương pháp thử thống kê với các hệ số tương quan giữa các số hạng kề nhau của chuỗi là khác nhau.

Nguyên nhân trực tiếp của những sai số khác đó là mối quan hệ giữa các số hạng của chuỗi tăng lên sẽ dẫn đến sự san bằng những dao động trong mẫu tương đối ngắn (khoảng thời gian quan trắc ngắn). Kết quả đó là sự tích luỹ các khoảng lệch cùng một dấu trong mẫu ngẫu nhiên hữu hạn. Trong kết quả luỹ tích các khoảng lệch đó việc xác định trị bình quân các chuỗi có mối quan hệ nội tại có thể thực hiện được với độ chính xác hơn so với mẫu các giá trị độc lập có cùng số hạng. Ngoài ra đối với phân phối của ước lượng của tham số nào đó (hệ số biến đổi và đặc biệt là hệ số không đối xứng) chịu ảnh hưởng rất lớn của hệ số không đối xứng hay quan hệ C_S/C_V. Khi phân tích đầy đủ độ chính xác ước lượng thống kê các tham số phân phối của các đại lượng thuỷ văn nghiên cứu cần phải xét đến tính hình đó. Do lời giải lý thuyết của bài toán nghiên cứu chỉ có thể thực hiện được trong các trường hợp riêng (thí dụ như đối với các điều kiện của luật phân phối nhị thức khi $C_S=2C_V$), phương pháp thử thống kê được dùng làm phương pháp thực tế chủ yếu để giải bài toán dưới dạng tổng quát.



Hình 5.3. Lược đồ minh hoạ phép biến đổi luật phân phối đều sang một luật bất kỳ cho trước.

Việc mô hình hoá thống kê chuỗi ngẫu nhiên độc lập tuân theo một luật phân phối nhất định, theo lược đồ dùng trong tuyến tính thuỷ văn và tính toán thuỷ lợi được thực hiện rất đơn giản. Để làm việc này chuỗi các số ngẫu nhiên phân phối đều được dùng là các tần suất, từ đó theo đường tần suất cho trước các giá trị của chuỗi ngẫu nhiên được xác định với luật phân phối cho trước (hình 5.3). Khi sử dụng MTĐT số, thường xuyên người ta dùng các chương trình mẫu cho phép nhận được các số ngẫu nhiên phân phối đều. Khi tiến hành các phép toán mô hình hoá, luật phân phối thường được cho dưới dạng giải tích hoặc dạng bảng.

Để mô hình hoá chuỗi đại lượng ngẫu nhiên có mối tương quan giữa các số hạng kề nhau (xích Markoov đơn giản) ngày nay trong các công trình nghiên cứu thuỷ văn được sử dụng rất nhiều giả thiết, trong số đó có các công trình nghiên cứu G.G.Xvanidze[120], N.A.Kartvêlisvili [61], Đ.Ta.Ratkôvich [95-97], E.G.Blôkhinôv [25], G,A.Alêkxeev [9]. Trong số đó ta chú ý những điều kiện sau đây:

1. Chuỗi thống kê gốc của các đại lượng ngẫu nhiên tuân theo luật phân phối chuẩn có sử dụng tương quan chuẩn được biến đổi thành chuỗi có mối tương quan nội tại. Sử dụng tần suất của các số hạng trọng chuỗi sẽ được biến đổi đó làm (acgumen) đối số thông qua luật phân phối gama ta phục hồi chuỗi đại lượng ngẫu nhiên tuân theo luật đó và có mối tương quan nội tại. Giả thiết này đã được N.A.Kartvêlisvili [61] và G.A Alêkxêev [9] nghiên cứu lược đồ đó và đã được G.G.Xvanidze [129] nghiên cứu rất cụ thể. Nó được ứng dụng ngày một nhiều hơn đặc biệt là đối với mô hình hoá từng nhóm của các chuỗi thuỷ văn. Svanidze gọi nó là phương pháp luận thống nhất của mô hình hoá thống kê.

2. Chuỗi đại lượng ngẫu nhiên phân phối đều (các tần suất) đem dùng bằng phương pháp tương quan đặc biệt do I.O.Xarmanôv [118] nghiên cứu, được biến đổi sang chuỗi có mối quan hệ nội tại. Chuỗi giá trị được biến đổi bằng phương pháp trên, theo luật phân phối gama về sau được chuyển sang chuỗi cũng tuân theo luật phân phối đó. Khi mô hình đã được Đ.Ta.Ratcôvich [95 - 97] sử dụng khi nghiên cứu các nhóm năm của pha nước khác nhau.

3. Chuỗi thống kê của đại lượng ngẫu nhiên tuân theo luật phân phối gama được biến đổi chuỗi có mối tương quan nội tại. Phép biến đổi đó được thực hiện nhờ tương quan gama. Mô hình này là do E.G.Blôkhinôv và I.O.Xarmanôv [25,118] nghiên cứu các luật phân phối của ước lượng mẫu các thám số .

4. Chuỗi thống kê gốc tuân theo luật phân phối gama được biến đổi dang chuỗi có mối quan hệ nội tại. Phép biến đổi này được tiến hành nhờ tương quan chuẩn (giả thiết của Kriski - Menkel). Mô hình này được gọi là mô hình hồi quy và được A.S.Reznikôvski và các tác giả khác [37] sử dụng rộng rãi khi ước lượng các tham số phân phối và tính toán thuỷ văn.

Nói chung, việc lựa chọn phương pháp mô hình hoá thống kê trước hết dựa vào các chuỗi thực nghiệm ví dụ như dòng chảy sông ngòi để xây dựng các mô hình thống kê. Không dựa vào tài liệu quan trắc thực tế mà chỉ từ những kết luận viển vông sẽ khó tìm được phương pháp mô hình hoá thích hợp. Khi vấn đề đó không đưa vào phân tích đặc biệt, ta sẽ trình bày các kết quả ứng dụng mô hình hồi quy do Viện nghiên cứu thuỷ văn Liên Xô (G.G.I) dùng để tiến hành một luật những tính toán đánh giá các luật phạn phối của các tham số mẫu và tung độ đường tần suất.

Mô hình hoá thống kê theo lược đồ này được thực hiện theo phương trình:

$$K_{i+1} = 1 + r_{i,i+1}(k_i - 1) + \phi_{i+1}.Cv\sqrt{(1-r^2)} \qquad (5.36)$$

Ý nghĩa của phương trình (5.36) được làm rõ ở những lập luận sau:

Do có mối tương quan giữa các sô hạng của chuỗi thống kê các chuỗi x_i và x_{i+1} có thể xem như là hai chuỗi được liên hệ với nhau bằng phương trình hồi quy tuyến tính:

$$\bar{x}_{i+1,dk} - \bar{x}_{i+1} = r_{i,i+1}\frac{\delta xi+1}{\delta xi}(\bar{x}_i - \bar{x}_i) \qquad (5.37)$$

Ở đây $\bar{x}_{i+1,dk}$ - kỳ vọng toán học (trị bình quân) có điều kiện của chuỗi các giá trị x_{i+1}, ứng với giá trị x_i cho trước δ_{xi+1}, δ_{xi} khoảng lệch trung bình bình phương không có điều kiện của các chuỗi x_{i+1} và x_i; x_{i+1}; x_i - kỳ vọng toán học (trị bình quân) không có điêu kiện của các chuỗi x_{i+1} và x_i; $r_{i+1,\ i}$ - hệ số tương quan giữa các số hạng của các chuỗi x_i và x_{i+1}.

Do mối tướng quan giữa các số hạng của chuỗi nào. Có được xét là chệch nhau một số thứ tự (x_i và x_{i+1}) cho nên

$$\delta\ xi+1 = \delta\ xi = \delta x$$

$$x\ \ i+1 = xi = x \qquad (5.38)$$

Khi các chuỗi được viết dưới dạng hệ số môdul thì $\bar{x}_{i+1} = \bar{x}_i = 1$. Trong trường hợp này (5.37) có dạng:

$$\bar{k}_{i+1,dk} = 1 + r_{i,i+1}(k_{i-1}) \qquad (5.39)$$

Như vậy hai số hạng đầu tiên trong phương trình (5.36) là kỳ vọng toán có điều kiện của hai số hạng k_{i+1} phụ thuộc vào k_i

Khi mô hình hoá các chuỗi có mối tương quan nội tại ta cần phải xét đến sự phân tán ngẫu nhiên quanh kỳ vọng toán có điều kiện kê trên ($\overline{K}_{i+1,\,đk}$). Sự phân tán đó được xác định bằng đường tần suất có điều kiện, các tham số của nó là:

a/ Trị bình quan có điều kiện $k_{i+1,\,đk} = 1\text{-}r_{i,\,i+1}$ (ki-1)

b/ Hệ số biến đổ có điều kiện:

$$C_{v,dk} = C_v\sqrt{1 - r_{i,i+1}^2} \qquad (5.40)$$

Ở đây C_V - hệ số biến đổi có điều kiện chuỗi gốc

c/ Hệ số không đối xứng có điều kiện $C_{S,\,dk} = kC_{V,dk}$ giá trị lấy cũng như đối với đường phân phối không điều kiện. Khi tính toán như trên các giá trị $C_{V,\,đk}$ và $C_{S,đk}$ giá trị k lấy cũng như đối với đường phân phối không điều kiện. Khi tính toán như trên các giá trị $C_{V,\,đk}$ và $C_{S,đk}$ lấy như nhau đối với mọi k_i.

Số hạng thứ ba trong phương trình (5.36) phản ánh sự phân tán của biến ngẫu nhiên quanh kỳ vọng có điều kiện. Với (5.39) và (5.40), phương trình (5.36) có thể biểu diễn dưới dạng:

$$k_{i+1} = k_{i+1} + C_{V,đk}\,\phi_{i+1}$$

Trong tài liệu thuỷ văn lược đồ nghiên cứu này đôi khi được gọi là mô hình hoá các chuỗi ngẫu nhiên có mối quan hệ nội tại dùng tương quan chuẩn. Tên gọi đó là do trong phương pháp mô hình hoá này, phương sai có điều kiện được lấy bằng hằng số theo đẳng thức (5.40), thực ra điều đó chỉ đúng với các chuỗi phân phối theo luật chuẩn.

Trong trường hợp riêng, khi các chuỗi đem nghiên cứu có mối tương quan nội tại ($r_{i,i+1}=0$), phương trình (5.37) lấy dạng thông thường.

$$x_{i+1} = 1 + \phi_{i+1}C_V \qquad (5.41)$$

đối với đường tần suất không có điều kiện, để mô hình hoá các số ngẫu nhiên như trên.

Kỹ thuật mô hình hoá chuỗi ngẫu nhiên có mối tương quan nội tại tuân theo luật phân phối cho trước (Kriski Menkel hay nhị thức) gồm các phép như sau:

1/ Lấy số ngẫu nhiên là tần suất.

2/ Xác định những giá trị ki theo số ngẫu nhiên và $C_{V,}$ C_S/C_V cho trước.

3/ Tình toán số hạng X_{i+1} của chuỗi theo phương trình (5.36)

Tổng quát hoá các số ngẫu nhiên phân phối đều trong khoảng từ 0 đến 1 theo phương trình của V.X.Gubenkô và A.I.Trekuôv trình bày trong cuốn sách của V.Đ.Liasenkô (86). Độ dài của khoảng khong tuần hoàn bằng 2^{23} - 1 = 8.589.934.591, nghĩa là biến thiên thực tế có chu kỳ vô hạn, vì vậy sử dụng phương trình này có thể nhận chuỗi các số ngẫu nhiên phân phối đều thực tế có độ dài vô hạn. Vì các bảng phân phối gama ba tham số và luật nhị thức được thiết lập đối với các tần suất từ 0,001 đến 99,9% khoảng từ 0 đến 1, trong phạm vi đó các số ngẫu nhiên phải phối hợp được khôi phục và chuyển sang khoảng mới 0,001 đến 99,9% nhờ phép biến đổi tuyến tính.

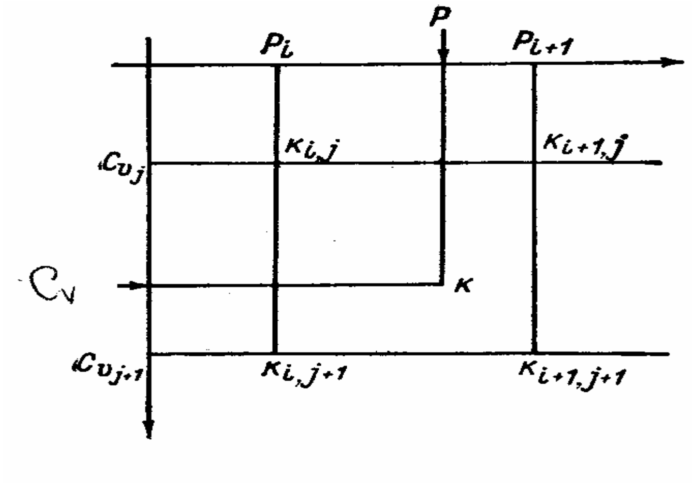
$$P = \frac{d - C}{b - a} x + C - a = 99{,}899 + 0{,}001 \qquad (5.42)$$

trong đó (b-a) khoảng gốc (0-1); (a-c) - khoảng biến đổi 0,001 - 99,9%.

Phép biến đổi trên được thực hiện, bởi vì khi N đủ lớn trong các miền tần suất 0,0 - 0,001% và 99,9% - 100% có thể có những số ngẫu nhiên tuỳ ý, trong khi đó các bảng sử dụng cac phân phối gama ba tham và luật nhị thức các khoảng này của tần suất bị bỏ mất.

Lấy những số ngẫu nhiên cho phân phối đều vừa được trong khoảng (0,001 - 99,9%) là tần suất biểu diễn dưới dạng phân trăm, đưa chúng vào bảng phân phối gama ba tham số (hay luật nhị thức) với C_V và C_S/C_V cho trước ta xác định các hệ số modul. Sau đó theo phương trình (5.36) ta tính được chuỗi mô hình hoá ứng với các tham số gốc (x, C^*_V, C_S/C_V, $r^*_{i,i+1}$, N).

Khi xác định các hệ số môdul theo $k(P,C_V)$ cần phải tiến hành nội suy các giá trị bảng đối với C^*_V gốc và số ngẫu nhiên P* nhận được (hình 5.4). Để thực hiện phép toán đó ta sử dụng công thức nội suy của Newton tới các bước không đề u.



Hình 5.4. Lược đồ minh hoạ

$$k = k_{i,j} + \frac{k_{i+1} - k_{i,j}}{P_{i+1} - P_i}(P^* - Pi) + \frac{k_{i,j+1} - k_{i,j}}{C_{vj+1} - C_{vi}}(C_v^* - C_{vj}) +$$
$$\frac{k_{j,i} + k_{i+1,j+1} - k_{i,j+1} - k_{i+1,j}}{(P_{i+1} - P_i)(C_{vj+1} - C_{vi})}(P^* - Pi)(C_v^* - C_{vj}) \qquad (543)$$

trong đó $k_{i,j}$; $k_{i+1,\, j}$, $k_{i,\, j+1}$ -những giá trị bảng của hệ số môdul.

P* - số ngẫu nhiên lấy là tần suất.

Pi - Giá trị bảng của tần suất nhỏ hơn và gần bằng P*; Pi+1 giá trị bảng của tần suất lớn hơn và gần bằng P*.

C^*_V - giá trị gốc của hệ số biến đổi.

C_{VJ} - giá trị bảng của C_V nhỏ hơn và gần bằng C^*_V;

C_{VJ+1} - giá trị bảng của C_V lớn hơn và gần bằng C^*_V.

Trong trường hợp luật phân phối nhị thức bảng số được biểu diễn dưới dạng khoảng lệch tung độ so với trị bình quân $\phi = \frac{k-1}{C_v}$ có dựa vào sử dụng công thức tương tự theo P và C_S.

Việc sử dụng 4 số hạng của công thức Newton đã đem đến việc nội suy có độ chính xác cao. Những tính toán theo lược đồ này đã được thực hiện trên MTĐT. Thuật toán của chương trình tính toán gồm 4 giai đoạn.

Ở giai đoạn thứ nhất tiến hành mô hình hoá thống kê chuỗi ngẫu nhiên (x_1, x_2, ... x_n) dưới dạng xích Marcôv đơn giản ứng với đường tần suất gama ba tham số hay nhị thức. Để làm việc này trên MTĐT người ta dựa vào:

a/ Các tham số tiêu chuẩn của phân phối (trị bình quân số học x*, hệ số biến đổi $C^*_{S\ \text{hoặc}}$ tỷ số $k^* = C^*_S/C^*_V$).

b/ Hệ số tương quan giữa những số hạng kề nhau của chuỗi $r_{i,i+1}$;

c/ Phân phối gama ba tham số được biểu diễn dưới dạng bảng Blôkhinôv - Nikôlskaia, hay bảng phân phối nhị thức.

d/ Tổng số hạng trong chuỗi mô hình hoá (N)

e/ Số số hạng mẫu (n)

Ở giai đoạn thứ hai của tính toán , thực chất là được thực hiện đồng thời với giai đoạn 1, các chuỗi mô hình hoá với dung lượng bằng N được chia ra làm các mẫu có dung lượng n nhỏ hơn (n < N) với các mẫu này tiến hành tính toán các tham số tiêu chuẩn của phân phối (x, C_V, C_S) và $r_{i,i+1}$.

<table>
<tr><td>Giai đoạn I</td><td>Giai đoạn II</td><td>Giai đoạn IV</td></tr>
<tr>
<td>$\begin{array}{ccccc} x_{11} & x_{12} \cdots & x_{1i} & \cdots & x_{1n} \\ x_{21} & x_{22} \cdots & x_{2i} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ x_{j1} & x_{j2} \cdots & x_{ji} & \cdots & x_{jn} \\ \vdots & \vdots & \vdots & & \vdots \\ x_{L1} & x_{L2} \cdots & x_{Li} & \cdots & x_{Ln} \end{array}$</td>
<td>$\begin{array}{cccc} \bar{x}_1 & C_{v_1} & C_{s_1} & r_1 \\ \bar{x}_2 & C_{v_2} & C_{s_2} & r_2 \\ \vdots & \vdots & \vdots & \vdots \\ \bar{x}_j & C_{v_j} & C_{s_j} & r_j \\ \vdots & \vdots & \vdots & \vdots \\ x_L & C_{v_L} & C_{s_L} & r_L \end{array}$</td>
<td>$\begin{array}{ccccc} x_{1;0,001} & x_{1;0,01} \cdots & x_{1;p} & \cdots & x_{1;99,9} \\ x_{2;0,001} & x_{2;0,01} \cdots & x_{2;p} & \cdots & x_{2;99,9} \\ \vdots & \vdots & \vdots & & \vdots \\ x_{j;0,001} & x_{j;0,01} \cdots & x_{j;p} & \cdots & x_{j;99,9} \\ \vdots & \vdots & \vdots & & \vdots \\ x_{L;0,001} & x_{L;0,01} \cdots & x_{L;p} & \cdots & x_{L;99,9} \end{array}$</td>
</tr>
<tr>
<td></td>
<td>$\begin{array}{cccc} \bar{x}_{\bar{x}} & \bar{C}_v & \bar{C}_s & \bar{r} \\ C_{v_{\bar{x}}} & C_{v_{C_v}} & C_{v_{C_s}} & C_{v_r} \\ C_{s_{\bar{x}}} & C_{s_{C_v}} & C_{s_{C_s}} & C_{s_r} \end{array}$</td>
<td>$\begin{array}{ccccc} \bar{x}_{p=0,001} & \bar{x}_{p=0,01} \cdots & \bar{x}_p & \cdots & \bar{x}_{p=99,9} \\ C_{v_{x_{p=0,001}}} & C_{v_{x_{p=0,01}}} \cdots & C_{v_{x_p}} & \cdots & C_{v_{x_{p=99,9}}} \\ C_{s_{x_{p=0,001}}} & C_{s_{x_{p=0,01}}} \cdots & C_{s_{x_p}} & \cdots & C_{s_{x_{p=99,9}}} \end{array}$</td>
</tr>
<tr><td></td><td>Giai đoạn III</td><td></td></tr>
</table>

Hình 5.5 Sơ đồ tính

Ta đưa vào lược đồ tổng quát của tính toán chỉ số j (từ j = 1 đến j = L) để ký hiệu số thứ tự của mỗi số hạng trong chuỗi với dung lượng bằng n (hình 5.5).

Việc tính toán các tham số tiêu chuẩn của phân phối và hệ số tương quan giữa các số hạng kề nhau của chuỗi đối với từng mẫu j (từ 1 đến L) được tiến hành theo các công thức đã biết.

$$\overline{X}_j = \frac{\sum_{i=1}^{n} x_{i,j}}{n} \tag{5.44}$$

$$C_{vj} = \frac{\delta_j}{\overline{X}_j}\sqrt{\frac{\frac{1}{n-1}\sum_{i=1}^{n}(\overline{X}_{j,i} - \overline{X}_j)^2}{\overline{X}_j^2}} \tag{5.45}$$

$$C_{S,j}\, \frac{n}{(n-1)(n-2)} \cdot \frac{\sum_{i=1}^{n}(X_{j,i} - \overline{X}_j)^3}{\delta_j^{\delta}} \tag{5.46}$$

$$r_{(i,i+1)j} = \frac{1}{n-1\delta_j^2}\sum_{i=1}^{n}(X_{j,i} - \overline{X})(X_{j,i+1} - \overline{X}) \tag{5.47}$$

trong tính toán trên máy tính điện tử chúng được biểu diễn qua mômen gốc.

Vì vậy, trong kết quả tính toán ở giai đoạn này ta nhận được X_j, C_{vj}, C_{si}, $r_{(i,i+1)j}$; trong đó i biến thiên từ 1 đến L. Những tính toán tiếp theo được chia ra làm 2 bài toán độc lập (giai đoạn 3 và 4).

Ở giai đoạn ba tiến hành tính toán các tham số tiêu chuẩn của phân phối (x, C_V,C_S) theo các chuỗi x_j, C_{vj}, C_{sj} r $(_{i,\ i+1})$,j các công thức (5.44) - (5.47), trong đó j biến thiên từ 1 đến L. Cuối cùng ở giai đoạn này ta nhận được trị bình quân (x), hệ số biến đổi (C_{VX}, hệ số không đối xứng (C_{SX}) của trị bình quan mẫu x_j, trị bình quân (C_V), hệ số biến đổi (Cv_{CS}) và hệ số không đối xứng (Cs_{CV}) của hệ số biến đổi mẫu C_V; trị bình quân $(\bar{r}_{i,i+1})$ hế số biến đổi (Cv_{CS}) và hệ số không đối xứng (Cs_{CS}) của hệ số không đối xứng mẫu, và cuối cùng trị bình quân $(r_{i,i+1})$ hệ số biến đổi ($C_{vri,i+1}$) và hệ số không đối xứng($Cs_{ri,i+1}$) của hế số tương quan mẫu giữa các số hạng kề nhau của chuỗi r $_{(i,i+1)j}$ ... Các tham số được tính đưa ra in.

Ở giai đoạn thứ tư (cuối cùng) tiến hành tính toán các tham số tiêu chuẩn (X_{pj}, $\delta_{x,py}$, Cs_{xpj}) của các tung độ mẫu phân phối gama ba tham số hay phân phối nhị thức:

Để làm việc đó ta sử dụng các tham số x_j, C_{vj} (j = 1 đến L) nhận được giữa giai đoạn thứ hai. Giá trị hệ số không đối xứng đối với phân phối Kriski - Menkel là theo quan hệ chuẩn hoá với hệ số biến đổi ($C_S = kC_V$) vì vậy ta chỉ sử dụng hai tham số tự do (X_j và C_{vj}) để xác định tung độ đường tần suất của phân phối gama ba tham số. Kết quả ta nhận được X_{pj} (j = 1, 2, ... L) đối với các tần suất P = 0,001; 0,01; 0,03; 0,05; 0,1; 0,3; 0,5; 1; 3; 5; 10; 20; 25; 30; 30; 40; 50; 60; 70; 75; 80; 90; 95; 97; 99; 99,5; 99,7; 99,9%. Đối vơi luật phân phối nhị thức cũng như trên ta xét tính toán với hai tham số (X, C_V và $C_S = kC_V$) và ba tham số (x, C_V và C_S).

Cuối cùng các công thức (5.44) - (5.46) ta xác định được các tham số của tung độ đường phân phối gama ba tham số (x_p, Cv_{xp}, Cs_{xp}) hay những phân phối nhị thức với những giá trị tần suất trên. Trong tính toán đó, biến ngẫu nhiên là những giá trị X_{pj} (j = 1 đến L) còn P = 0,001; 0,01 ... 99,9%.

Như vậy, trong chương trình thực hiện thuật toán trình bày trên đây tính toán được xét trước tiên có sử dụng lượng thông tin gốc theo sáu tham số: n (n<N); $\overline{x}*;\overline{C}_V^* k = C_s / C_v$ quyết định bảng phân phó gama ba tham số hay bảng phân phối nhị thức gốc, cũng được thực hiện trên MTĐT $r*_{i,i+1}$ và cuối cùng là N quyết định số lượng mẫu L = N/n.

Trong kết quả tính toán được in ra những đặc trưng sau đây:

1/ Các tham số phân phối $\overline{x}_N, C_{VN}, C_{SN}$ và hệ số tương quan giữa các số hạng kề nhau của chuỗi ($r_{(i, i+1)}$, N) theo các chuỗi đã được mô hình hoá với dung lượng cho trước N. Căn cứ vào lượng thông tin này tiến hành đánh giá chất lượng của mô hình hoá bằng cách so sánh ccs tham số mới nhận được với các tham số cho trước. Lúc này các đặc trưng cho trước $\overline{x}*$ $C*_V$, $C*_S$, $r*_{i,i+1}$ là các tham số của tổng thể còn $\overline{X}_N$ C_{VN}; $C_{SôNG\ NGòI}$; r_N- các tham số mẫu tính theo chuỗi n có dung lượng hữu hạn. Khi đủ lớn các tham số cho trước và tính được sẽ sai khác rất ít, lúc đó chuỗi mô hình hoá có dung lượng là N có thể lấy làm tổng thể.

2/ Các tham số tiêu chuẩn của các tham số phân phối mẫu có dung lượng cho trước $n(\overline{x}; C_{V\overline{V}}; C_{SX}; \overline{r}; C_{Vr}; C_{Sr}$. Khi so sánh các tham số cho trước với các giá trị bình quân của chúng mà nhận được theo các mẫu có dung lượng bằng $n(\overline{x}n), \overline{C}_{Vn}; \overline{C}_{Sn}; \overline{r}n$ có thể đánh giá được độ chênh lệch của các tham số mẫu, Khoảng lệch trung bình bình phương ngẫu nhiên của các tham số mẫu ứng với dung lượng mẫu cho trước n δx (Cvx) ; Cv_{cv}; Cv_{cs} $\delta r(C_{vr})$ khi tổ hợp với $C_{s\overline{x}}; C_{Scv}; C_{sr}$ quyết định luật phân phối của các tham số mẫu.

3/ Các tham số phân phối của 27 tung độ với p = 0,001; 0,01; 99,9%. Độ chệch của tung độ đường tần suất có thể xác định được bằng cách so sánh giá trị (thực) bảng x^*_p với giá trị tính toán được $\overline{x}_p$. Biết các tham số phân phối của tung độ đường tần suất và lấy luật phân phối gama làm luật mô tả các dao động của tung độ mẫu ta dễ dàng xác định được giới hạn tin cậy cho tất cả các tần suất đem dùng, hay trong tính toán cuối cùng cho các phân phối Kriski - Menkel hay phân phối nhị thức dùng để mô tả các tài liệu mẫu có dung lượng hữu hạn. Ta nhận thấy rằng các giới hạn tin cậy tương tự, theo phương trình cho này được thực hiện với hai tham số tự do đem dùng ($\overline{x}, C_v$) và quan hệ cố định C_S/C_V đối với luật phân phối Kriski - Menkel và nhị thức, cũng như đối với ba tham số tự do được dùng cho luật phân phối nhị thức.

Sự thực hiện phương trình nghiên cứu này đối với các tham số gốc khác nhau: $\overline{x}^*$, C^*_V, C^*_S; $r^*_{i,\ i+1}$, n và N có thể tìm ra được các quy luật thông kê, việc ước lượng chúng một cách đúng đắn không thể nhận được khi dựa vào phân tích lý luận hay cũng rất phức tạp, cho nên con đường này không thể sử dụng vào thực tế để giải các bài toán thống kê được. Cũng cần chú ý là độ chính xác của các lời giải dựa vào mô hình hoá nói chung có thể là rất cao khi N đủ. Song khi xây dựng các thí nghiệm số trị tương tự thường thường phải tìm hiểu điều kiện tối ưu mà trong đó sai số của lời giải không lớn lắm và có thể thực hiện tốt chương trình trên máy tính số hiện đại có tốc độ lớn. Ta nhận thấy rằng chương trình tính toán được xét đến như khả năng biến đổi của C_V và C_S được hạn chế bằng giá trị của tham số đó đã trình bày trong bảng của Blôkhinôv - Nikôlskaia [24] của phân phối nhị thức (C_S); giá trị $r_{i,\ i+1}$ có thể biến đổi trong toàn bộ phạm vi của mình nghĩa là ±1,n<N còn $N \le 2^{33}$-1.

5.5. KẾT QUẢ ƯỚC LƯỢNG THAM SỐ MẪU CỦA PHÂN PHỐI.

Như chúng ta đã biết ở bài 3, việc xác định ước lượng thống kê của các tham số thống kê của các tham số của luật phân phối với dung lượng đầy đủ (nghĩa là phụ thuộc vào dung lượng mẫu, độ biến đổi và không đối xứng cử các chuỗi với các giá trị khác nhau của hệ số tương quan nội tại) dựa vào phân tích lý luận không thể được. Những ước ulợng đó có khả năng xác định được dựa vào việc sử dụng phương pháp phép thử thống kê. Những kết luận cơ bản sẽ trình bày sau đây dựa theo hướng này đều dựa vào mô hình hoá thống kê được tiến hành theo chương trình ở bài 3. Khi thực hiện chương trình này các chuỗi phải được biểu diễn dưới dạng hệ số môdul trong trường hợp này như ta đã biết trị bình quân của chuỗi bằng 1.

Dung lượng chung của chuỗi đem mô hình hoá (N) lấy bằng 50000, còn trong các trường hợp riêng lấy đến 100000, như vậy số lượng mẫu (L) với n = 10 là 5000; đối với n = 25; L = 2000; đối với n = 50; L = 1000; đối với n = 100; L = 500 và cuối cùng với n = 200; l = 250. Các hệ số biến đổi khi mô hình hoá các chuỗi lấy bằng 0,1; 0,3; 0,5; 0,7 ; 1,0 hệ số không đối xứng $C_S = 0$; $C_S = C_V$; $C_S = 2C_V$; $C_S = 3C_V$; $C_S = 4C_V$ hệ số tương quan giữa các số hạng kề nhau của chuỗi r = 0,0; 0,3; 0,5; 0,7; 0,9. Các chuỗi bao gồm số số hạng như nhau 10; 25; 50; 100 và 200. Tổng số phương án tính toán đối với mỗi phan tích thử nghiệm (Kriski - Menkel và nhị thức) là 500 và đối với phân phối chuẩn là 125 phương án.

Những tài liệu nhận đượcđề cập rất đầy đủ đến vấn đề nghiên cứu đã được các nhà xuất bản khoa học chuyên ngành công bố. Ở đây ta chỉ trình bày những kết luận cơ bản.

5.5.1 Ước lượng trị bình quân mẫu.

Ở bài 2 của chương này ta đã xác nhận trị bình quân của số học ước lượng đúng, không chệch, đủ và có hiệu quả của kỳ vọng toán (x_0) hay các giá trị chuẩn như thuật ngữ thường dùng trong thuỷ văn. Các trị bình quân số học $\overline{x}$ nhận được theo nhiều mẫu của tổng thể các giá trị độc lập tuân theo luật phân phối chuẩn sẽ tạo nên một chuỗi cũng tuân theo luật phân phối chuẩn với các tham số.

$$\overline{x} = x_0 \text{ và } \sigma_{\overline{x}} = \frac{\sigma_x}{\sqrt{n}} \qquad (5.48)$$

trong đó σ_x - khoảng lệch trung bình bình phương của chuỗi các giá trị x; σ_x - khoảng lệch trung bình bình phương của tổng thể; n - số số hạng trong mẫu.

Luật phân phối của trị bình quân được bảo tồn và đối với các mẫu bị chệch so với phân phối chuẩn nếu như dung lượng mẫu không đủ lớn.

Đối với các chuỗi thống kê có mối tương quan nội tại giữa các số hạng kề nhau (Markôv đơn giản) Kriski và Menkel [78] đã nhận được quan hệ.

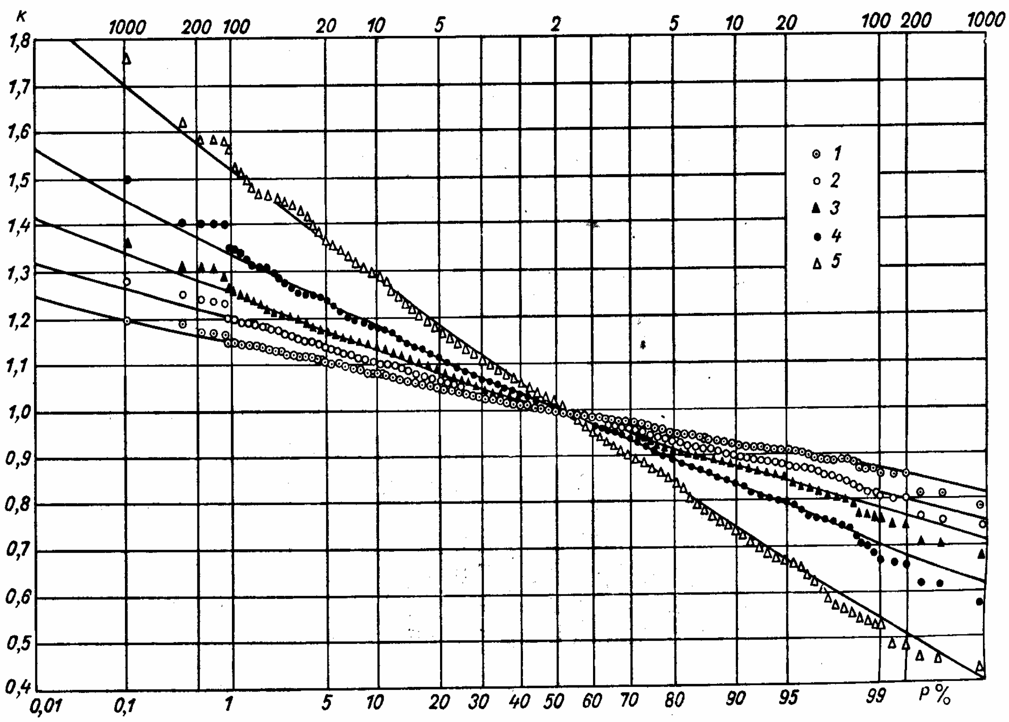
$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}\sqrt{\frac{1+\frac{2n}{n(1-r)}(n-\frac{1-r^n}{1-r})}{1-\frac{2r}{n(n-1)(1-r)}(n-\frac{1-r^n}{1-r})}} \tag{5.49}$$

trong đó r - hệ số tương quan giữa các số hạng kề nhau của chuỗi. Các ký hiệu khác cũng như trên. Khi r = 0 công thức (5.48) và (5.49) rõ ràng là trùng nhau. Khi r nhỏ công tưức (5 - 48) và (5.49) rõ ràng là trùng nhau. Khi r nhỏ công thức (5.49) có thể biểu diễn dưới dạng quan hệ gần đúng:

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}\sqrt{\frac{\pi-r}{1-r}} \tag{5.50}$$

được đưa ra dùng trong thực tế.

Sự minh hoạ các luật tích phân phân phối của trị bình quân mẫu đã được trình



Hình 5.6 Các đường cong đảm bảo trị trung bình của mẫu nhận được từ phương pháp mô hình hoá Monte - Karlo với các hệ số tương quan giữa các trị thành viên
1- r = 0.0 2- r =0.3 3- r = 0.5 4- r = 0.7 5- r = 0.9

bày trên hình (5.6).

Các đường tần suất trình bày trên hình (5.6) xây dựng trên cơ sở các chuỗi đã được mô hình hoá bằng phương pháp thử thống kê với các tham số gốc: x = 1; C_V 0,3; C_S = 0,6; n = 25; N = 15000; L = 6000; $r_{i, i+1}$ = 0,0; 0,3; 0,5; 0,9.

Nhũng kết quả của mô hình hoá cho thấy rằng các trị bình quân số học mẫu với tất cả các tham số gốc thực tế là không chệch (Tất cả các đường tần suất đàu cắt nhau ở điểm x = 1 với P = 50%). Phương sai của trị bình quân mẫu tăng lên khi hệ số tương quan giữa các số hạng kề nhau của chuoĩ tăng. Phân tích của trị bình quân mẫu có hệ só không đối xứng dương và lớn lên theo hệ số r tăng.

Hình 5.6 trình bày đồng thời các hàm phân phối nhị thức với các đường phân phối thực nghiệm. Hàm lý luận được xây dựng trên cơ sở các tham số của chuỗi trị bình quân mẫu ((X1, x2, ... xn); Cv_X; Cs_X). Sự phù hợp tốt giữa các đường tần suất lý luận và thực nghiệm đã cho ta cơ sở tất cả các phân tích của chuỗi thống kê mô hình hoá thực hiện dựa vào việc sử dụng các tham số thống kê ứng với các mẫu đem mô hình hoá. Cách này sử dụng lượng thông tin trong các chuỗi đêm mô hình hoá có ưu điểm là chuyển sang hàm lý luận mô tả các chuỗi thống kê sẽ được mô hình hoá các tham số. đảm bảo việc khái quát hoá tài liệu mẫu san bằng những nhiễu động ngẫu nhiên riêng biệt của chúng. Ngoài ra, hàm phân phối lý luận các ước lượng mẫu (thí dụ như trị bình quân) cho phép ta biểu diễn lượng thông tin dưới dạng gon (compac). Điều đó đã mở ra khả năng ký thuật chỉnh lý các chuỗi ngẫu nhiên có độ dài thực tế là vô hạn.

Thật vậy, khi sử dụng các giải thích mô tả các chuỗi thống kê trên MTĐT trong toàn bộ bộ nhớ sẽ chỉ chứa thông tin về tham số của các phân phối này. Để khôi phục các hàm phân phối thực nghiệm. MTĐT cần phải chứa đầy đủ lượng thông tin về các số hạng của chuỗi nghiên cứu. Trong trường hợp này các chuỗi thống kê là các chuỗi gồm các trị bình quân mẫu. Bên cạnh đó cần chú ý rằng sự phân phối các số hạng của mẫu trong chuỗi san bằng chiếm rất nhiều thời gian của máy. Cuối cùng tính toán trên MTĐT các hàm phân phối thực nghiệm theo 600 mẫu chiếm khoảng 40 phút thời gian của máy, còn tính các tham số thống kê của hàm phân phối lý luận theo 2000 mẫu chiếm khoảng 5 phút. Cũng cần phải chú ý là MTĐT có thể thực hiện đồng thời (theo dung lượng bộ nhớ) chỉnh lý khoảng 700 mẫu đối với bài toán xây dựng hàm phân phối thực nghiệm, còn khi sử dụng MTĐT chỉ để ước lượng các tham số thống kê của mẫu. số lượng các mẫu nghiên cứu thực tế là vô hạn.

Như vậy, dạng được dùng đưa ra từ MTĐT lượng thông tin dưới dạng các tham số thống kê tiêu chuẩn đã mở ra thêm khả năng sử dụng phương pháp thử thống kê. Những hiểu biết trên hoàn toàn phụ thuộc về việc nghiên cứu các luật phân phối của các hệ só biến đổi, không đối xứng mẫu và hệ số tương quan giữa các số hạng kề nhau của chuỗi cũng như của các tung độ đường tần suất.

Sử dụng phương pháp thử thống kê để xác định khoảng lệch trung bình bình phương của trị bình quân mẫu cho phép ta đi đến những kết luận sau đây:

1/ Việc ứng dụng những hàm phân phối khác nhau (Kriski - Menkel nhị thức) để mô hình hoá không được thể hiện ở giá trị của sai số trung bình bình phương ngẫu nhiên của trị bình quân mẫu trong toàn miền các tham số phân phối gốc cho trước (C_V, $C_{S/CV}$) với r và n khác nhau.

2/ Trong các giá trị của sai số trung bình bình phương của trị bình quân mẫu không được biểu hiện quan hệ $C_{S/}$ C_v gốc cho trước..

3/ Sai số trung bình bình phương của trị bình quân mẫu càng tăng, khi hệ số biến đổi của phân phối gốc tăng, khi dung ulọng mẫu giảm và khi hệ số tương quan gưữa các số hạng kề nhau của chuỗi tăng. Khi giá trị của hệ số tương quan bằng 0,3 - 0,5 khoảng lệch trung bình bình phương của trị bình quân mẫu tăng lên khoảng 30 - 40%.

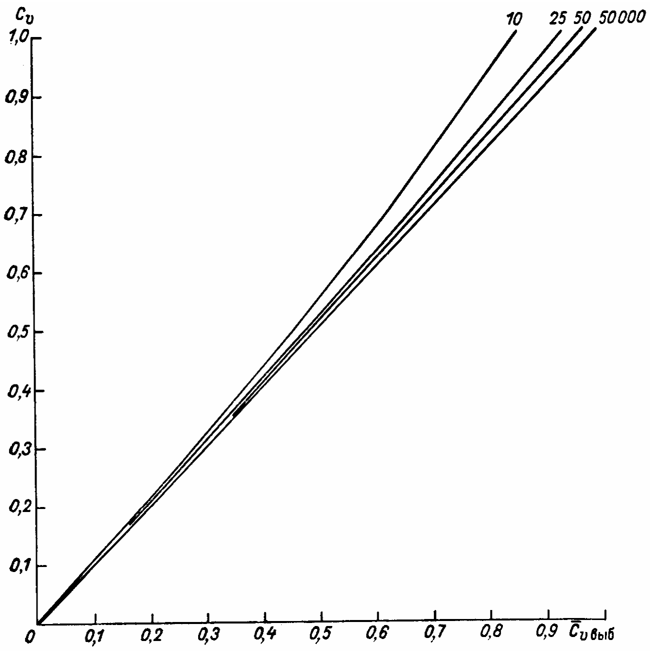
4/ Sự sai lệch khác nhau giữa các giá trị của khoảng lệch trung bình bình phương theo công thức (5.49) và tính được theo phương pháp thống kê thường không lớn lắm. Chúng tăng khi hệ số tương quan giữa các số hạng kề nhau của chuỗi và hệ số biến đổi của phân phối gốc tăng. Khi sử dụng mẫu không lớn(n=10), giá trị của hệ số biến đổi lớn (C_V = 1,0) và hệ số tương quan r = 0,5 thì phương sai khác đó chiếm khoảng 0,05, còn khi C_V = 1,0; r = 0,7; n = 10 sự sai khác đó đạt tới 1,0 + 0,2. Giá trị của khoảng lệch trung bình bình phương của trị bình quân mẫu tính theo công thức (5.49) lớn hơn sơ với tính bằng phương pháp thử thống kê. Như vậy, hệ số tương quan giữa các số hạng kề nhau của chuỗi không lớn hơn 0,5 ($r \le 0,5$) và số số hạng của chuỗi n >10, sự khác nhau đó không có, vì vậy công thức (5.49) dùng trong thực tế có đủ độ chính xác cần thiết.

Ngoài ra, có thể nhận thấy rằng sự khác nhau không lớn lắm ở ước ượng khoảng lệch trung bình bình phương của trị bình quân mẫu nhận được là độc lập, nghĩa là nhận được bằng phương pháp thử thống kê và dựa vào công thức (5.49) đều là sự khẳng định thêm tính của việc tiến hành mô hình hoá thống kê.

5.5.2. Ước lượng hệ số biến đổi màu.

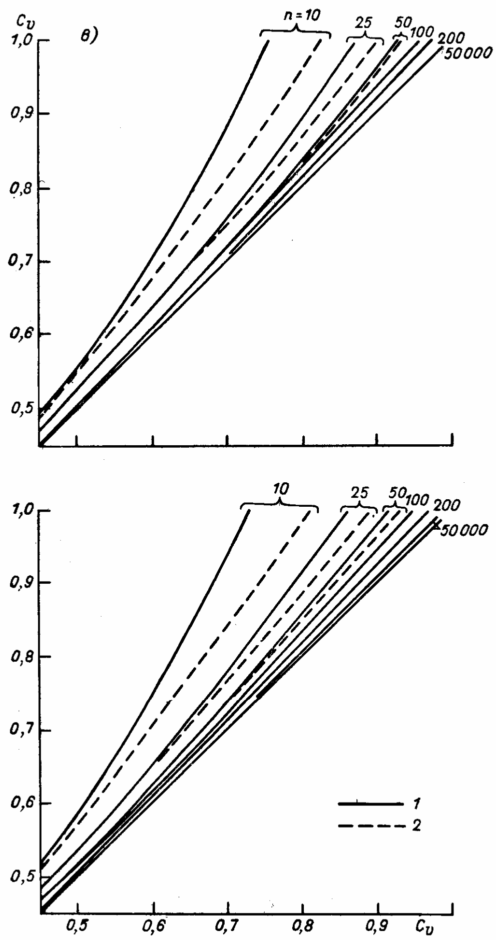
Khoảng lệch trung bình bình phương mẫu và hệ số biến đổi ngay cả đối với các chuỗi có mối tương quan nội tại là ước lượng tham số của tổng thể chệch âm. Sự chệch này được khử riêng bằng cách đưa vào một số nhân $\frac{n}{n-1}$, trong đó n - dung lượng mẫu.

Trong các công trình của E.G. Blôkhinôv [18] và A.S. Rêznikôsla [37] đã chú ý rằng, sự chệch âm của hệ số biến đổi mẫu (khoảng lệch trung bình bình phương) thường không vượt quá 3 -4 %, song điều đó chỉ đúng nếu quan hệ gốc $C_S/Cv \leq 2$.



Hình 5.7 Đồ thị quan hệ Cv = f(Cv mẫu) với các dung lượng mẫu khác nhau (n=10; n=25; n=50) dựng theo công thức (5.53) với Cs = 2Cv và $r_{i,\,i+1}$ =

Khi tỷ số C_S/Cv tăng sự chệch tăng, chỉ khi hệ số biến đổi nhỏ hơn 0,5, sự chệch mới ít phụ thuộc vào tỷ số C_S/Cv và không lớn lắm. Với những giá trị khác của tham số của chuỗi, điều đó không như vậy. Thí dụ, theo tài liệu mô hình hoá thống kê đối với các chuỗi có mối tương quan nội tại khi khi $C_S/Cv = 1$ và $C_V = 1{,}0$ dung ượng mẫu n = 10 giá trị chệch âm chiếm 25%, cũng với các tham số thống kê như vậy nhưng n = 25 giá trị chệch nhận được bằng 10%. Những giá trị chệch đó ứng với phân phối nhị thức được dùng làm phân phối gốc. Mối quan hệ giữa kỳ vọng toán (trị bình quân) của các giá trị mẫu của hệ số biến đổi (C_V mẫu) và ước lượng không chệch của hệ số biến đổi (C_V) đối với các chuỗi không có mối tương quan nội tại được biểu diễn bằng công thức của E.G. Blôkhinôv [18].



Hình 5.8. Đồ thị quan hệ Cv = f(Cv mẫu) với việc sử dụng phương pháp thử nghiệm thống kê với các quan hệ Cs/Cv =3,4 với $r_{i,i+1}$ = 0.0; 0.3 và n=10,25, 100, 200,50000 đối với phân phối nhị thức (1) và K-M (2)

$$\overline{C}_{v\text{mẫu}} = Cv - \frac{C_v}{4n}(1 + 3C_v^2) \qquad (5.51)$$

Mối quan hệ nội tại đã được làm tăng độ chệch âm của hệ số biến đổi mẫu. Tình trạng này đã được xét trong công thức Kriski - Menkel [78].

$$\overline{C}_{v\text{mẫu}} = C_{v\text{mẫu}}\sqrt{1 - \frac{2n}{n(n-1)(n-r)}(n - \frac{1-r^n}{1-r})} \qquad (5.52)$$

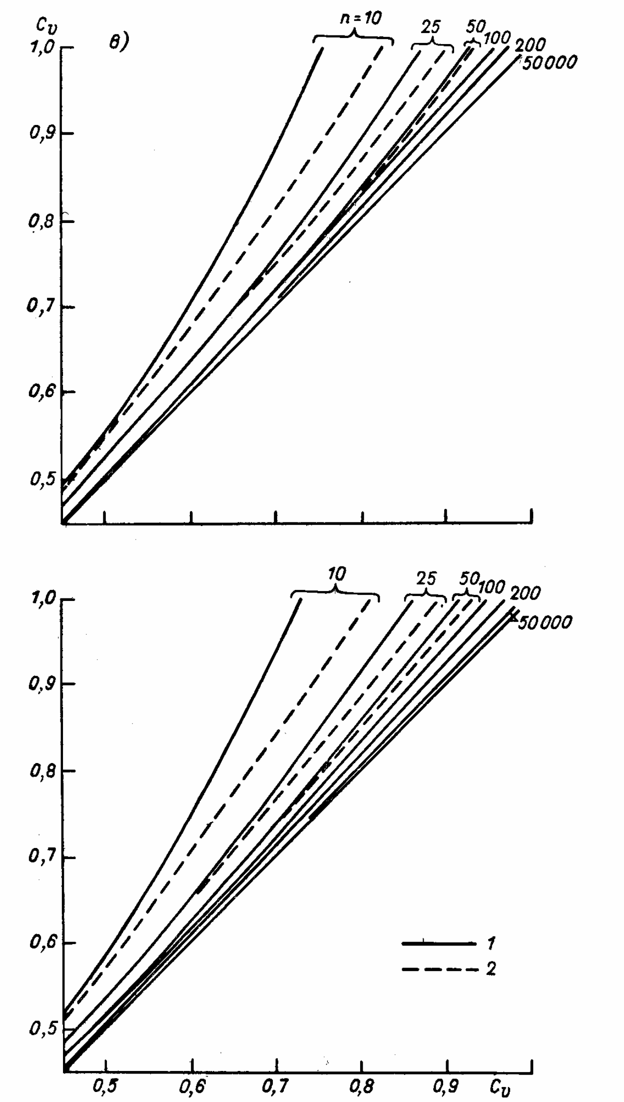
Ở đây Cv mẫu - trị bình quân của hệ số biến đổi với các chuỗi có dung lượng bằng n, và có hệ số tương quan nội tại r; Cv mẫu - trị bình quân của hệ số biến đổi mẫu đối với các chuỗi không có mối quan hệ nội tại.

Thay vào quan hệ (5.52) đẳng thức (5.51) ta nhận được:

$$C_{v\text{mẫu}} = \left[C_V - \frac{C_V}{4n}(1 + 3C_V^2)\right]\sqrt{1 - \frac{2r}{n(n-1)(n-r)}(n - \frac{1-r^n}{1-r})} \qquad (5.53)$$

Mối quan hệ (5.53) phù hợp với đẳng thức Cs = 2Cv trong chuỗi của gái trị gốc. Từ phương trình (5.53) có thể nhận được biểu thức cho ước lượng không chệch pha của hệ số biến đổi (C_V) phụ thuộc vào trị bình quân của các ước lượng hệ số biến đổi (C_V mẫu) vào số hàng của chuỗi n và là số tương quan nội tại r. Song, lời giải này rất tốn công và không thuận lợi. Vì vậy, tốt hơn hết là nên sử dụng quan hệ đồ thị C_V = f(C_V mẫu; n; r) Thí dụ quan hệ này với r = 0,3 đã được trình bày ở hình 5.7.

Việc sử dụng phương pháp thử thống kê đã cho phép ta đánh giá được khả năng ứng dụng công thức (5.53) với C_S khác $2C_V$. Hình như khi $C_V \le 0,5$ các ước lượng không chệch của hệ số biến đổi nhận được theo các quan hệ rút ra từ phương trình (5.53) khá phù hợp với ước lượng nhận được bằng phương pháp thử thống kê với tất cả quan hệ Cs/Cv được dùng trong tính toán thuỷ văn. Khi các hệ số biến đổi gần bằng 1,0 và tỷ số Cs/Cv > 2 giá trị C_V không chệch nhận được theo các quan hệ (hình 5.7) rút ra từ phương trình (5.53) là nhỏ hơn so với nhận được theo phương pháp thử thống kê. Do đó, cho nên giá trị chệch của hệ số biến đổi mẫu với những giá trị khác nhau của các tham số phân phối gốc được tính bằng phương trình (5.53) là không đầy đủ. Việc tính toán sự chệch của ước lượng dẫn đến đối mẫu một cách chính xác hơn có thể phải dựa vào ứng dụng các quan hệ Cv = f (có mẫu x,h,Cs/Cv) nhận được bằng phép thử thống kê.



Hình 5.8 Đồ thị quan hệ Cv = f(Cv mẫu) với việc sử dụng phương pháp thử nghiệm thống kê với các quan hệ Cs/Cv =3,4 với $r_{i,i+1}$ = 0.0; 0.3 và n=10,25, 100, 200,50000 đối với phân phối nhị thức (1) và K-M (2)

Những thí dụ của các quan hệ như vậy ứng với phân phối nhị thức và phân phối Kriski-Menkal, khi Cs/Cv = 3và 4, và r=0, r=0,3 đã được trình bày trên hình 5.8. Các thí dụ này cho thấy rằng các ước lượng hệ số biến đổi mâu thuẫn được sử dụng đường Kriski - Menkel có độ chệch âm nhỏ hơn ước lượng ứng với phân phối nhị thức. Song cần phải chú ý rằng sự khác nhau đó là rất cơ bản chỉ đối với dung lượng mẫu nhỏ và tỷ số C_S/C_V lớn; sự khác nhau này cũng tăng khi mối tương quan nội tại của chuỗi tăng. Với các giá trị thực $C_V = 1,0$; $C_S = 4C_V$; r = 0,3 và n = 10 đối với phân phối nhị thức độ chênh lệch tới 27%, còn đối với phân phối Kriski - Menkel là 19%. Khi giảm C_V, C_S/C_V và tăng r, n độ chệch âm của hệ số biến đổi mẫu cũng giảm.

Chúng ta sẽ xét sự phân tán ngẫu nhiên của hệ số biến đổi mẫu (σ_{cV}). Khái niệm chung về sự dao động ngẫu nhiên của hệ số biến đổi mẫu xác định đối với tổng thể có các tham số x = 1,0; Cv ≠ 00,3; Cs = 0,6 đã nhận được từ các đường tần suất trình bày trên hình 5.9. để xây dựng các đường tần suất người ta sử dụng 600 mẫu và mỗi mẫu có 25 số hạng. Các giá trị của hệ số biến đổi mẫu được xác định bằng phương pháp mômen ứng với các hệ số tương quan giữa các số hạng của chuỗi r = 0,0; 0,5; 0,7; 0,9. Trên hình 5.9 ta có thể tháy rằng sự phân tán của các hệ số biến đổi mẫu tăng khi hệ số tương quan giữa các số hạng kề nhau của chuỗi cũng tăng. Trên hình 5.9 bên cạnh các đường tần suất lý luận là đường tần suất của phân phối nhị thức. Đường phân phối lý luận được xây dựng theo các tham số (Cv; Cv_{Cv}, Cs_{Cv}) của các chuỗi được mô hình hoá. Sự trùng nhau giữa các đường tần suất lý luận và thực nghiệm sẽ tạo ra cơ sở trong chỉnh lý các chuỗi nhận được bằng phương pháp thử thống kê, sử dụng không phải là tất cả chuỗi các giá trị tạo nên chuỗi thống kê của hệ số biến đổi, mà chỉ sử dụng các tham số thống kê của các chuỗi đó. Kết luận này đã được rút ra khi nghiên cứu ước lượng trị bình quân mẫu.

Đối với các chuỗi không có mối tương quan nội tại, khoảng lệch trung bình bình phương của hệ số biến đổi mẫu (Cv) tính bằng phương pháp mômen thống kê có thể xác định theo công thức [18].

$$\sigma_{C_v} = \frac{Cv}{\sqrt{2n}}\sqrt{1+C_V^2} \qquad (5.54)$$

đã được giới thiệu trong cuốn. Hướng dẫn xác định các đặc trưng thuỷ văn tuyến tính "XN435 - 72.

Về sau E.G Blôkhinôv (18) đã đưa vào số hiệu chỉnh được xác định bằng con đường thực nghiệm.

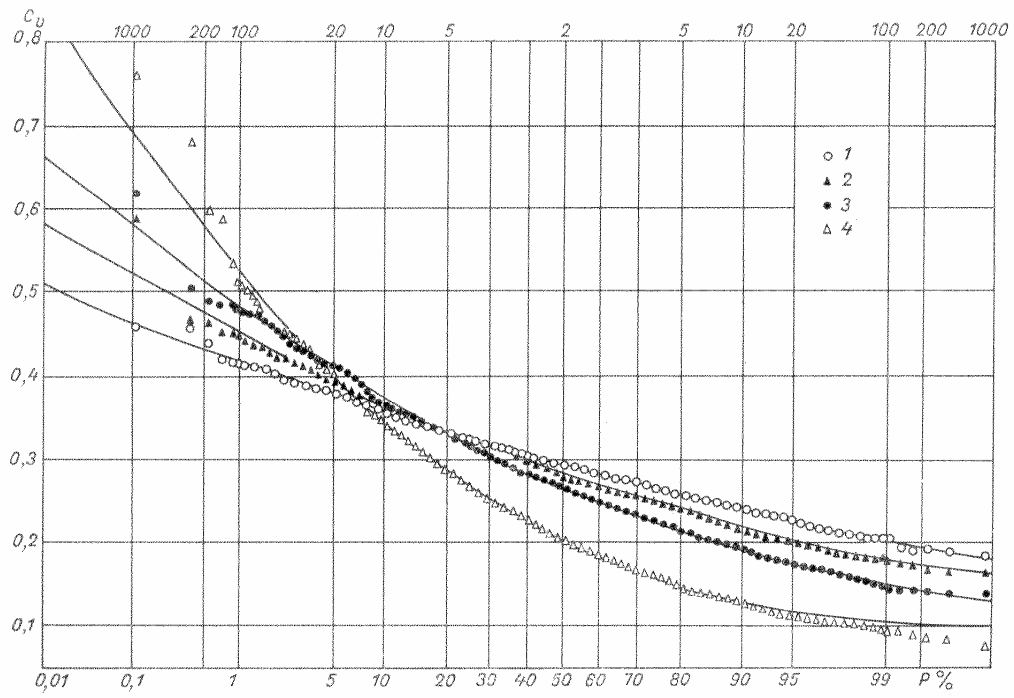
$$\sigma_{C_v} = \frac{C_v}{n + 4C_V^2}\sqrt{\frac{n(1 + C_v^2)}{2}} \tag{5.55}$$

trong đó n - dung lượng mẫu; Cv - hệ số biến đổi của chuỗi gốc.

Công thức sai số tiêu chuẩn của hệ số biến đổi mẫu nhận được bằng phương pháp thích hợp tối đa có dạng:

$$\sigma_{C_v} = \frac{Cv}{\sqrt{2n}}\sqrt{\frac{3}{3 + C_v^2}} \tag{5.56}$$

Mối quan hệ do Blôkhinôv tìm được và được trình bày trong cuốn "Hướng dẫn xác định các đặc trưng thuỷ văn tính toán "XN435 - 72. Các công thức (5.54) - (5.56) ứng với điều kiện là không có mối tương quan nội tại của chuỗi và nhận được đối với tỷ số Cs/Cv = 2.



Hình 5.9 Các đường cong đảm bảo hệ số biến đổi của mẫu theo mô hình hoá M-K (x=1, Cv=0,3; Cs=0,6) với các hệ số tương quan chập khác nhau 1- r=0; 2- r=0,5; 3- r=0,7; 4- r=0,9

A.S.Rêznikôvski (37) đối với Cv vẫn nhận được bằng phương pháp mômen, dựa vào phương pháp thử thống kê đã đánh giá sự ảnh hưởng tương quan nội tại của chuỗi cho trường hợp Cs = 2Cv đối voi giá trị Cv. Khi dựa vào quan hệ (5.55) số hiệu chỉnh thực nghiệm Rêznikôvshi đã biểu diễn nó dưới dạng:

$$\sigma_{Cv} = \frac{Cv}{N+4C_V^2}\sqrt{\frac{n(1+C_V^2)}{2}(1+\frac{3C_V r^2}{1+r})} \qquad (5.57)$$

Dựa vào rất nhiều tài liệu ở trên nhận được bằng phương pháp thử thống kê G.G.I đã tiến hành đánh gía công thức (5.57) theo khả năng sử dụng nó với các tỷ số Cs/Cv ≠2 và luật phân phối Kriski - Menkel và luật nhị thức làm phân phối gốc. Sự so sánh trên đã được thực hiện với hệ số biến đổi mẫu vì công thức (5.57) được viết dưới dạng:

$$Cv_{Cv} = \frac{1}{n+4C_V^2}\sqrt{\frac{n(1+C_V^2)}{2}(1+\frac{3Cvr^2}{1+r})} \qquad (5.58)$$

Những kết quả so sánh đã chứng minh rằng khi không có mối tương quan nội tại trong chuỗi gốc (r = 0) và Cs = 2Cv công thức (5.58) cho thấy sự phù hợp tốt với tài liệu của mô hình hoá thống kê +. Khi các tỷ số gốc khác Cs = 2Cv, các giá trị ước lượng nhận được theo công thức (5.58) nhỏ hơn ước lượng nhận được bằng phương pháp thử thống kê. Sự khác biệt đó đặc biệt quan trọng khi n nhỏ và hệ số biến đổi của chuỗi gốc (Cv), tỷ số Cs / Cv và hệ số tương quan nội tại của r đều lớn. Đối với các điều kiện đó, số hiệu chỉnh thực nghiệm của Rêznikôvski được dựa vào biểu thức (5.55) cần phải làm chính xác. Trước khi kết thúc sự nghiên cứu này cần phải tiến hành đánh giá sự phân tán ngẫu nhiên của hệ số biến đổi mẫu với tỷ số Cs ≠Cv và có mối tương quan nội tại của chuỗi, bằng cách sử dụng tài liệu nhận được bằng phương pháp thử thống kê.

Các hệ số biến đổi của hệ số biến đổi mẫu (Cv_{Cv}) nhận được theo các mẫu mô hình hoá với sự sử dụng luật phân phối nhị thức và phân phối Kriski - Menkel khi Cs≠ 2Cv cơ bản là khác nhau. Các chuỗi hệ số biến đổi mẫu đều có thể không đối xứng dương. Hệ số không đối xứng này tăng khi tỷ số Cs/Cv, Cv và r tăng; và giảm khi dung lượng mẫu tăng Tỷ số giữa hệ số không đổi đối xứng với hệ số biến đổi của hệ số biến đổi mẫu (Cs_{Cv}/ Cv_{Cv}) tăng khi hệ số biến đổi tăng.

5.5.3. Ước lượng hệ số thống kê không đối xứng của mẫu.

Để minh hoạ sự phân tán ngẫu nhiên của hệ số không đối xứng trên hình 5.10 đã trình bày các đường tần suất của tham số này. Khi xây dựng các đường tần suất trên người ta sử dụng 600 giá trị Cs nhận được từ các mẫu, mỗi mẫu có 25 số hạng. Dung lượng chung của chuỗi đem mô hình hoá là N = 15000; của tham số gốc được dùng để mô hình hoá là x = 1,0; Cv = 0,3; Cs = 0,6; hệ số tương quan giữa các số hạng kề nhau của chuỗi được lấy bằng r = 0,0; 0,5; 0,7; 0,9.

Cũng như khi nghiên cứu các chuỗi thống kê mô hình hoá của các tham số khác (trị bình quân, hệ số biến đổi ta phải tính các tham số tiêu chuẩn (Cs; Cv_{Cs}; Cs_{Cs}) của chuỗi nghiên cứu. dựa vào các tham số này để tiến hành xây dựng đường tần suất nhị thức (hình 5.10)) chúng rất phù hợp với phân phối của hệ số không đối xứng mẫu (Cs).

Trên hình 5.10 ta có thể thấy rằng phân phối của hệ số không đối xứng mẫu có sự phân tán ngẫu nhiên lớn hơn so với các chuỗi x và Cv. Phân phối này có chệch âm tăng khi hệ số tương quan giữa các số hạng kề nhau của chuỗi tăng. Sự chệch âm của hệ số không đối xứng mẩu, khi không có mối tương quan nội tại của chuỗi có thể khử bằng cách tính ước lượng không chệch ($C_{sk,\ chệch}$) theo một số các công thức sau đây:

$$C_{s\ k\ .\ chệch} = (1+8.5/n)\ C_{s\ mẫu} \qquad (5.59)$$

(công thức các nhà thuỷ văn Mỹ dùng).

$$Cs\ k.chệch = [n+5+2C_V(1+3C_V^2)]/n\ C_S\ mẫu \qquad (5.60)$$

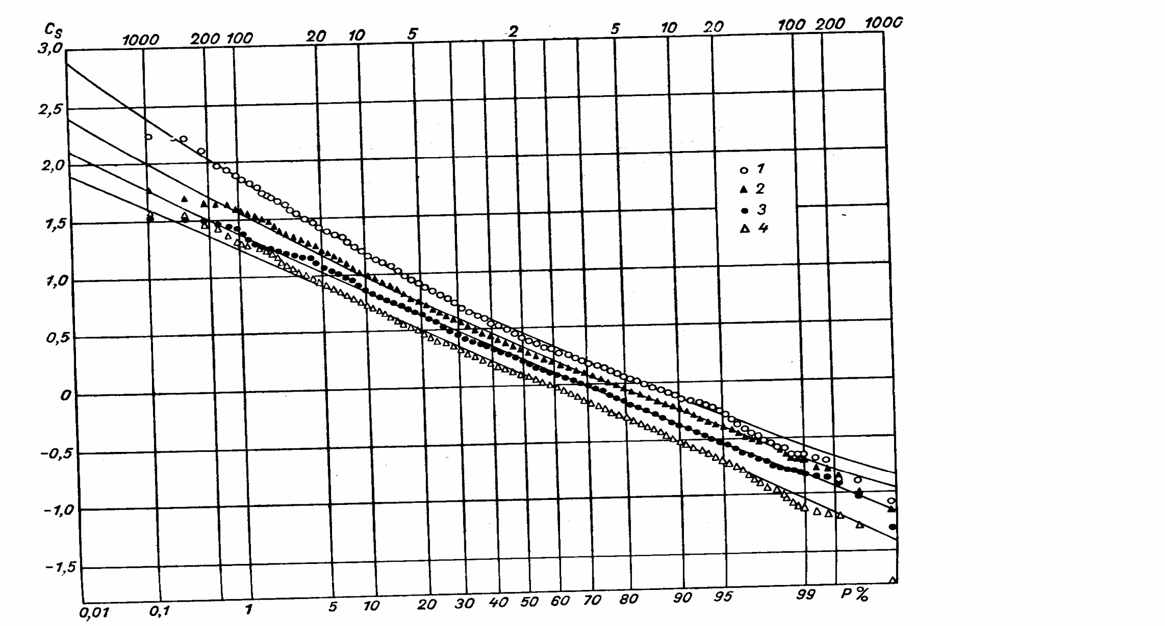
(công thức của E,G. Blôkhinôv)

$$Cs\ k\ .\ chệch = [n+4+4Cv)\ (a+2r^2)]/n\ Cs\ mẫu. \qquad (5.61)$$

(công thức của A.S. Rêznikôki)

có xét ảnh hưởng của tướng quan giữa các chuỗi đối với độ chệch của hệ số không đối xứng mẫu.

Các công thức trên đều đúng đối với trường hợp khi phân phối gốc Cs = 2Cv. Để đánh giá độ chính xác của công thức ta đem so sánh những giá trị $C_{smẫu}$ nhận được theo các biểu thức đó với kết quả ước lượng bằng phương pháp thử thống kê với Cs = 2Cv.



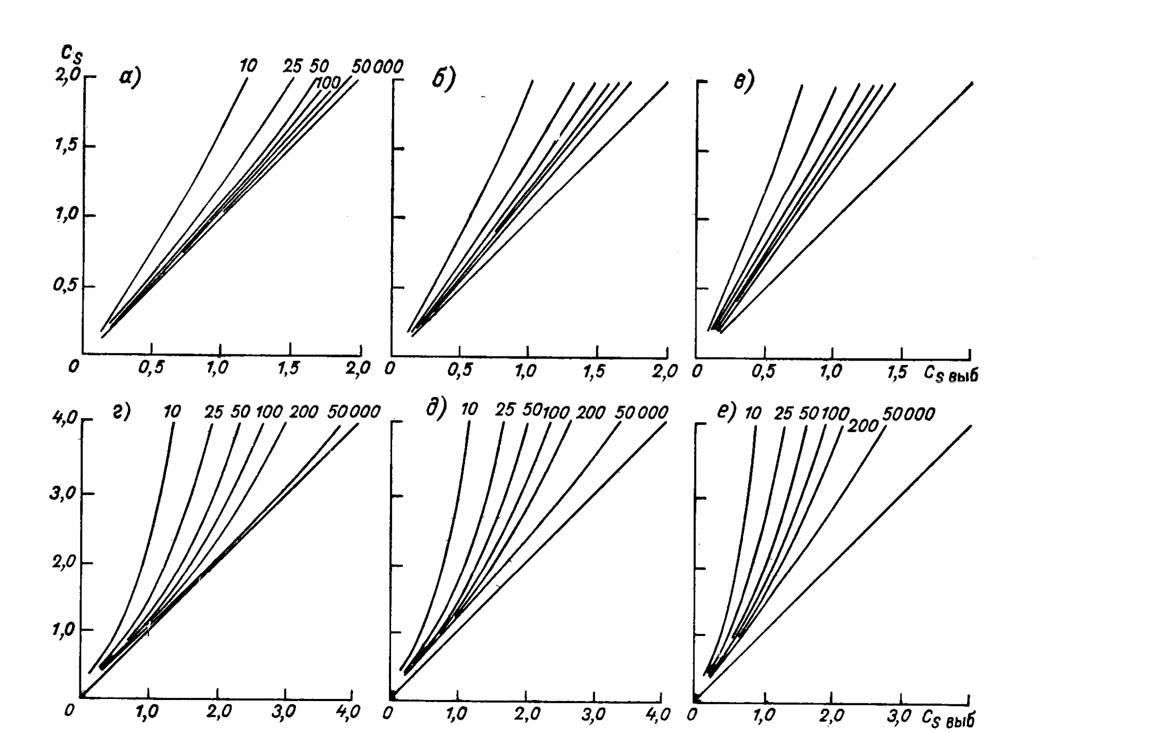
Hình 5.10 Đường cong đảm bảo hệ số bất đối xứng của mẫu theo số liệu phương pháp M-K với x=1; Cv=0,3 và Cs=0,6 theo các hệ số tương quan 1- r=0,3; 2- r=0,5; 3- r=0,7; 4- r=0,9.

Những so sánh này đã chứng minh rằng công thức (5.61) rất phù hợp với tài liệu thử thống kê khi Cs = 2Cv. Một lưòi giải tổng quát hơn với việc đánh giá sự chệch của hệ số không đối xứng với các tỷ số Cs/Cv khác nhau dung lượng mẫu khác nhau và hệ số tương quan nội tại của chuỗi cung khác nhau và các đồ thị của mối quan hệ các ước lượng Cs chệch và không chệch được xây dựng cho Cs/Cv = 1, 2, 3, 4; n = 10, 25, 50, 100, 200, 5000; r = 0,0; 0,3; 0,5. Ngoài ra đối với các mẫu được mô hình hoá còn sử dụng luật phân phối nhị thức và phân phối Kriski - Menkel. Các tài liệu mô hình hoá thống kê đã cho thấy rằng độ chệch của hệ số không đối xứng mẫu tăng khi tăng các tham số của chuỗi gốc Cs, Cs/Cv, r, và giảm dung lượng mẫu n.

Những xây dựng trên hình 5.11 đã cho phép ta khử những chệch âm của hệ số không đối xứng mẫu.

Sự phân tán ngẫu nhiên của hệ số không đối xứng mẫu được xác định bằng khoảng lệch trung bình bình phương hay hệ số biến đổi của các hệ số không đối xứng của chuỗi các hệ số không đối xứng mâu (Cs_{Cs}).

Để đánh giá khoảng lệch tiêu chuẩn của hệ số không đối xứng mẫu trong thuỷ văn thường được dùng công thức lý luận của kriski - Menkel (78) ứng với các chuỗi có Cs = 2Cv và không có mối tương quan nội tại (r = 0)



Hình 5.11 Đồ thị quan hệ Cs = f(Csmẫu) theo phương pháp thử xác suất với tỷ lệ Cs/Cv =2; 4; và các giá trị $r_{i,i+1}$=0,0; 0,3; 0,5 với các dung lượng mẫu n=10, 25, 50, 100, 200, 50000 đối với phân bố nhị thức

a. Cs=2Cv, r=0; b.Cs=2Cv, r=0,3; c. Cs=2Cv, r=0,5

d. Cs=4Cv, r=0; e.Cs=4Cv, r=0,3; f. Cs=4Cv, r=0,5

$$C_S = \sqrt{\frac{6}{n}(1+6^2+5^2)} \qquad (5.62)$$

Trên cơ sở tài liệu mô hình hoá thống kê đối với các chuỗi có tỷ số Cs/Cv = 2 và r = 0, ReZmkôvski (37) biểu diễn biểu thức dưới dạng:

$$\delta C_S = \sqrt{\frac{6}{n}(1+Cv^2)} \qquad (5.63)$$

Trong trường hợp đó người ta xét các chuỗi của hệ số môdul Cv = và với $Cv_{Cs} = \dfrac{\delta C_S}{C_S}$, các quan hệ (5.62) và (5.63) có thể biểu diễn qua Cv.

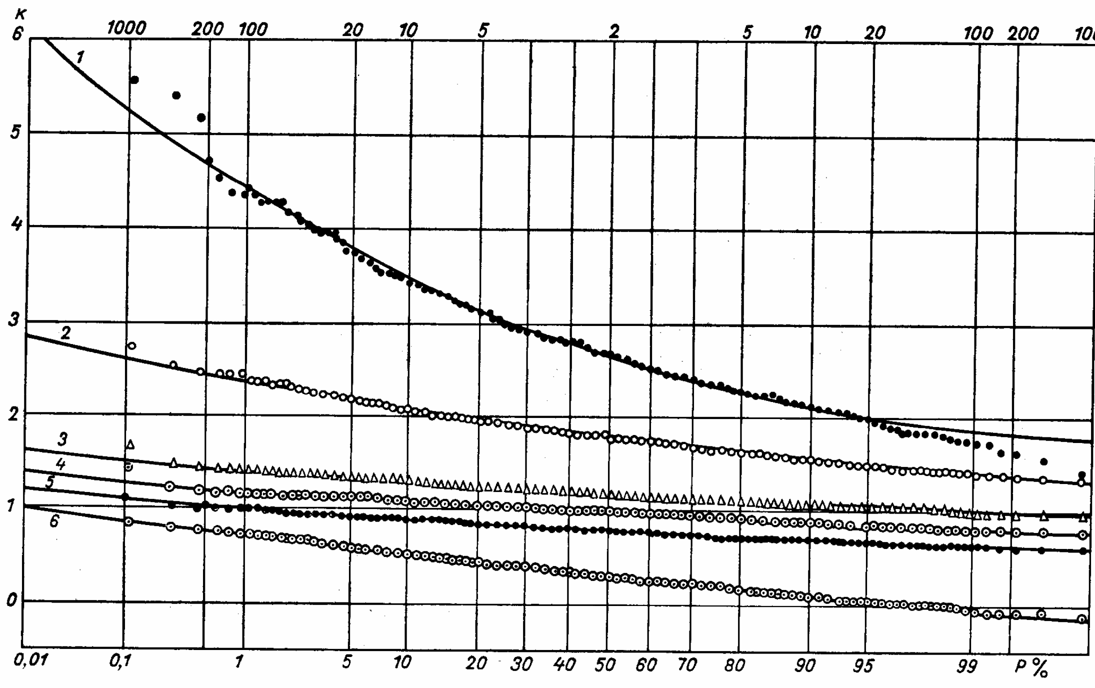
Sự so sánh các giá trị Cv_{Cs} tính theo các công thức (5.62) và (5.63) và bằng phương pháp thử thống kê đối với trường hợp Cs = 2Cv và r = 0 đã chứng minh rằng công thức (5.63) đối với trường hợp Cs= 2Cv rất phù hợp với tài liệu mô hình hoá thống kê được sử dụng để xây dựng nó. Các quan hệ khác của các tham số trên công thức (5.63) không đề cập đến bởi vì đánh giá sai số ngẫu nhiên của hệ số không đối xứng mẫu trong các trường hợp đó cần dựa vào các kết quả nhận được bằng phương pháp thử thống kê với Cs = 2Cv.

Dựa vào ba tham số tự do đó tiến hành xây dựng 600 đường tần suất mẫu, được dùng để xác định các giá trị xp ứng với những tần suất khác nhau. Rõ ràng là đối với mỗi giá trị xác định P ta nhận được 600 giá trị xp. Theo các số liệu này ta xây dựng được các đường tần suất vẽ trên hình 5.12 hình dạng của chúng được xác định bởi sự phân bố của các điểm tương ứng trên đồ thị (các đường tần suất thực nghiệm. Với mỗi giá trị của P có 600 giá trị xp đã cho phép ta xác định được các tham số của đường tần suất xp (xp, Cv_{XP}, Cs_{XP})) và căn cứ vào đó xây dựng đường tần suất lý luận (đường nhị thức) phù hợp tốt nhất với đường tần suất thực nghiệm. Sự so sánh này đã cho thấy rằng, để xây dựng các đường tần suất nghiên cứu khi khái quát hoá tài liệu bằng phương pháp thử thống kê chỉ có thể dựa vào các đường tần suất lý luận quy định bởi ba tham số, không tiến hành xây dựng các đường tần suất thực nghiệm. Sự cần thiết của phép toán đó đã được xét ở trên.

5.5. 6. Ước lượng tung độ đường cong tần suất mẫu.

Ở mục trước ta đã xét ước lượng của các tham số tính theo mẫu thống kê hay còn gọi là các tham số mẫu của phân phối. Nhưng tài liệu đó đã cho phép ta giải thích đầy đủ nhất sự chệch, sự phân tán ngẫu nhiên của các tham số mẫu khi luật phân phối gốc khác nhau, dung lượng mẫu khác nhau và hệ số tương quan giữa các số hạng kề của chuỗi cung khác nhau. Rõ ràng sự chệch của ước lượng tham số (Cv, Cs) của các mẫu thống kê phải được thể hiện ở sự chệch của tung độ đường tần suất mẫu, còn sự phân tán ngẫu nhiên của các tham số mẫu cung phải được thể hiện ở sự phân tán của tung độ đường tần suất mẫu. Việc ước lượng các chệch và phân tán ngẫu nhiên của chúng là một cần thiết đặc biệt trong tính toán thuỷ văn, và đặc trưng tính toán cuối cùng trong thuỷ văn thường là tung độ đường tần suất. Việc giới thiệu tổng quát về sự chệch, phân tán ngẫu nhiên và không đối xứng của tung độ đường tần suất mẫu với P = 0,001; 1,0; 25; 50; 75; 99,9% đã được thể hiện trên hình 5.12.

Các đường tần suất biểu diễn trên hình 5.12 được xây dựng như sau. Đầu tiên theo phương pháp thử thống kê tạo mẫu một tổng thể có 1500 số hạng, với các tham số : x = 1,0; Cv = 0,3; Cs = 0,6 và hệ số tương quan giữa các số hạng kề nhau của chuỗi r = 0,3; Tổng thể này được chia ra làm 600 mẫu, mối mẫu có 25 số hạng. Đối với từng mẫu bằng phương pháp mômen xác định các tham số x, Cv, Cs.



Hình 5.12 Đường cong đảm bảo của tung độ các đường cong tần suất mẫu theo phương pháp thử ngẫu nhiên (x=1; Cv=0,3; Cs=o,6; r=0,3) với các tần suất p khác nhau

1- p=0,001%; 2- p=1%; 3-p=25%; 4-p=50%; 5-p=75%; 6-p=99,9%

Trên cơ sở ba tham số chỉ định đã nêu đã xây dựng hơn 600 đường cong tần suất mẫu, dùng để xác định giá trị x_p với các suất đảm bảo khác nhau. Rõ ràng đối với một giá trị P xác định có 600 x_p. Hình 5.12 xây dựng theo các số liệu này. Sự hiện diện của 600 giá trị x_p đối với mỗi P cho phép xác định tham số đường cong đảm bảo (x_p, Cv_{xp}, Cs_{xp}) trên cơ sở xây dựng đường cong giải tích (nhị thức) đảm bảo phù hợ với đường cong thực nghiệm. Sự trùng hợp này chứng tỏ rằng việc xây dựng các đường cong đã nêu khi khái quát tài liệu bằng phương pháp thử ngẫu nhiên có thể dựa trên các đường giải tích xác định theo ba tham số. Vấn đề này trước đây đã tường minh.

5.5.7. Ước lượng độ lệch của tung độ đường tần suất mẫu.

Trong tài liệu thuỷ văn người đầu tiên chú ý đến sự chệch của tung độ đường tần suất lẫm là E.G. Blôkhinôv (20). Kết luận này được dựa vào sự phân tích các chuỗi được mô hình hoá bằng phương pháp thử thống kê. Khi mô hình hoá các giá trị của tham số gốc như hệ số biến đổi lấy bằng cv = 0,5 và 1,0, tỷ số Cs/Cv = 2, các mẫu gồm 25 hoặc 50 số hạng, số lượng mẫu bằng 100. Các chuỗi đem nghiên cứu không có mối tương quan nội tại.

Những kết quả đánh giá sự chệch của tung độ đường tần suất mẫu dựa vào lượng thông tin đầy đủ nhất, thí dụ nhưu có xét mối tương quan nội tại của chuỗi cho thấy độ chệch của tung độ đường tần suất mẫu thường không lớn lắm. Ở miền tần suất nhỏ chệch âm, coi ở miền tần suất lớn chệch dương. Độ chệch tăng khi hệ số biến đổi gốc và tỷ số Cs/Cv tăng, độ chệch cũng tăng như vậy khi hệ số tương quan giữa các số hạng kề nhau của chuỗi tăng. Tăng dung lượng mẫu sẽ làm giảm độ chệch của tung độ đường tần suất mẫu. Độ chệch của tung độ đường tần suất có thể nhận biết được trong tính toán thực tế đối với các chuỗi nhỏ (n = 10 + 20). Hệ số biến đổi của chuỗi gốc lớn (Cv = 1,0) và hệ số tương quan nội tại của chuỗi lớn (r = 0,3 + 0,5).

Trong tính toán khi sử dụng các hệ số không đối xứng mẫu (nghĩa là trong trường hợp ba tham tự do) độ chệch của tung độ tần suất tăng lên so với trường hợp dùng 2 tham số tự do (x và Cv) là do độ chệch của hệ số không đối xứng.

Độ chệch của tung độ đường tần suất mã có thể khử bằng cách đưa vào một số hiệu chỉnh. Việc khả sự chệch của tung độ đường tần suất có thể tiến hành bằng cách đưa vào những hiệu chỉnh tương ứng để khử sự chệch của hệ số không đối xứng và hệ số biến đổi mẫu mà được dùng để tính trung độ đường tần suất.

5.5.8. Ước lượng dao động ngẫu nhiên của tung độ đường tần suất mẫu

Trong thống kê tính toán [111] đã chứng minh rằng phương sai của ước lượng tham số thống kê γ, chẳng hạn như những đại lượng mà ta quan tâm là hệ số biến đổi, hệ số không đối xứng và tung độ đường tần suất đều là các hàm số có tâm phân phối x và mômen trung tâm μ_2, μ_3, μ_4 dưới dạng tổng quát biểu diễn bằng quan hệ:

$$\sigma_\gamma^2 = \lambda_1^2\sigma_x^2 + \lambda_2^2\sigma_{\mu_2}^2 + \lambda_3^2\sigma_{\mu_3}^2 + \lambda_4^2\sigma_{\mu_4}^2 + 2\lambda_1\lambda_2\sigma_x - \sigma_{\mu_2} 2r_{x\mu_2} +$$
$$+ 2\lambda_1\lambda_3\sigma_x - \sigma_{\mu_3} 2r_{x\mu_3} + 2\lambda_1\lambda_4\sigma_x - \sigma_{\mu 4} 2r_{x\mu 4} + 2\lambda_2\lambda_3\sigma_{\mu_2}\sigma_{\mu 3} r_{\mu_2 x\mu_3} +$$
$$+ 2\lambda_2\lambda_4\sigma_{\mu_2}\sigma_{\mu 4} r_{\mu_2\mu_4} + 2\lambda_3\lambda_4\sigma_{\mu 3}\sigma_{\mu 4} r_{\mu_3\mu_4}$$

(5.64)

trong đó $\lambda_1 = \frac{\partial\varphi}{\partial x}; \lambda_2 = \frac{\partial\varphi}{\partial\mu_2}; \lambda_3 = \frac{\partial\varphi}{\partial\mu_3}; \lambda_4 = \frac{\partial\varphi}{\partial\mu_4}$ là các hàm riêng quy định sự phụ thuộc cảu tham số nghiên cứu () của đường phân phối các mômen thống kê tương ứng.

Khoảng lệch trung bình bình phương và hệ số biến đổi của các đại lượng được ký hiệu bằng chỉ số.

Dựa vào biểu thức đó để đánh giá phương sai ta có thể nhận được các công thức viết trên (5.48), (5.62). Ta sẽ xét lược đồ tổng quát sử dụng phương trình (5.64) ở thí dụ chứng minh công thức sai số tiêu chuẩn của tung độ đường tần suất đối với phân phối nhị thức khi Cs = 2Cv.

Trong trường hợp nghiên cứu hàm liên kết giá trị x_p với các giá trị xác định của nó, như ta đã biết có dạng:

$$x_p = \overline{x} + \sigma_x\Phi = \overline{x} + \sqrt{\mu_2\Phi}$$

Trong đó giá trị bảng của luật nhị thức phụ thuộc vào hệ số không đối xứng (Cs) và tần suất (P) ứng với nó là Xp nghĩa là = ϕ(Cs, P) - khoảng lệch trung bình bình phương của các chuỗi có giá trị X.

Nếu Cs = 2Cv giá trị là hàm số ước lượng mômen trung tâm mẫu bậc hai vì Cs không tham gia với tính chất là một tham số độc lập. Đối với trường hợp đang nghiên cứu $\lambda_3=\lambda_4=0$ phương trình (5.64) sẽ đơn giản:

$$\sigma_{xp}^2 = \sigma_x^2 + \left(\frac{\partial x_p}{\partial\mu_2}\right)^2 \sigma_{\mu_2} + \frac{2dx_p}{d\mu_2}\sigma_x - \sigma_{\mu_2} r_{x\mu_2} \qquad (5.65)$$

Trong đó

$$\frac{\partial x_1}{\partial \mu_2} = \frac{4}{2\sqrt{\mu_2}} = 2\sqrt{\mu_2}\frac{\partial \mu}{\partial \mu_2} \qquad (5.66)$$

$$\sigma_\chi^2 = \frac{\mu_2}{n}$$

Với $\sigma_{\mu_2}^2 = 2\mu_2(1+3\mu_2)$

$$\sigma_{\bar{x}} - \sigma_{\mu 2} r_{\bar{x}\mu} = \frac{2\mu_2^2}{n}$$

ta nhận được:

$$\sigma_{x_p}^2 = \frac{\mu_2}{n} + \left(\frac{\Phi}{2\sqrt{\mu_2}} + \sqrt{\mu_2}\Phi'\right)^2 \frac{2\mu_2^2}{n}(1+3\mu_2) + 2\left(\frac{\Phi}{2\sqrt{\mu_2}} + \mu_2\Phi'\right)\frac{2\mu_2^2}{n}$$

trong đó $\Phi' = \dfrac{\partial \Phi}{\partial \mu_2}$

Với $\mu_2 = \sigma^2$ và $\Phi' = \dfrac{\partial \Phi}{\partial \mu_2} = \dfrac{1}{\sigma}.\dfrac{\partial \Phi}{\sigma Cs}$

ta có thể nhận được biểu thức cuối cùng:

$$\sigma_{\bar{x}p} = \frac{\sigma}{\sqrt{n}}\sqrt{1+2(\frac{\Phi}{2}+\sigma\Phi_s)^2(1+3\sigma^2)+4\sigma(\frac{\Phi}{2}+\sigma\Phi_s)} = \frac{\sigma}{\sqrt{n}}\sqrt{A},$$

trong đó $A = 1 + 2(\dfrac{\Phi}{2} + \Phi_s\sigma)^2(1+3\sigma^2) + 4\sigma(\dfrac{\Phi}{2} + \Phi_s\sigma)$

$\Phi = \dfrac{x_p - \bar{x}}{\sigma} = \dfrac{k-1}{Cv}$ khoảng lệch chuẩn hoá của tung độ đường tần suất so với trị bình quân. Các giá trị ϕ đã được trình bày trong bảng luật phân phối nhị thức; $\Phi_s = \dfrac{\partial \Phi}{\partial Cs}$ đạo hàm ϕ theo Cs được xác định như là tỷ số những số gia $\dfrac{\Delta\Phi}{\Delta C_S}$ đối với các tần suất p cần tính giá trị σ_{xp}.

Trong trường hợp nếu chỉ xây dựng đường tần suất giá trị Cs được xác định không phải theo giá trị của hệ số biến đổi mà trực tiếp từ ước lượng mẫu của mômen trung tâm bậc hai và ba ($C_S = \dfrac{\mu_3}{\mu_1\mu_2}$) giá trị δxp được biểu diễn bằng phương trình:

$$\sigma^*_{x_p} = \frac{\sigma}{\sqrt{n}}\sqrt{A + (6\Phi_S^2 + 4\Phi_S\sigma + 34\Phi_S^2\sigma^2 + 24\Phi_s\sigma^4 - 4\Phi_s\sigma^2}$$

Số hạng thứ hai trong dấu ngoặc dưới dấu căn bậc hai của biểu thức (5.69) biểu diễn phần phân tán của xp, phụ thuộc vào những dao động của ước ượng hệ số Cs mẫu được đưa ra mô tả luật phân phối dưới dạng một tham số độc lập. Về sau kriski - Menkel (79) đã sử dụng công thức chính xác của sai số trung bình bình phương của cá tung độ đường tần suất mẫu có hai tham số tự do (x, Cv) với tỷ lệ cố định Cs / Cv = 2. Sinh Gu Anh làm chính xác thêm với việc xét cả hệ số tương quan giữa các ước lượng mẫu. Đối với các mẫu rút ra từ tổng thể tuân theo luật phân phối chuẩn thì mối quan hẹ này không có. Đối với phân phối nhị thức không đối xứng tồn tại mối tương quan này nó sẽ ảnh hưởng đến việc đánh giá số tiêu chuẩn của hệ số biến đổi mẫu và tung độ đường tần suất (định vi) mẫu. Với các điều kiện trên công thức (5.68) có thể được biểu diễn bằng biểu thức:

$$\sigma_{xp} = \frac{\delta}{\sqrt{n}}\sqrt{1 + 2\left(\frac{\Phi}{2} + \Phi sCv\right)^2 (1 + 3C_v^2) + 4Cv\left(\frac{\Phi}{2} + \Phi sCv\right) - 4\Phi sC_v^2(1 + \Phi Cv + \Phi' C_v^2)}$$

(5.70)

Ở đây biểu thức dưới dấu căn là giá trị đã được chính xác hoá của tham số A trong công thức (5.68)

$$\sigma^*_{x_p} = \frac{\sigma}{\sqrt{n}}\sqrt{A + (6\Phi_S^2 + 4\Phi_S\sigma + 34\Phi_S^2\sigma^2 + 24\Phi_s\sigma^4 - 4\Phi_s\sigma^2}$$

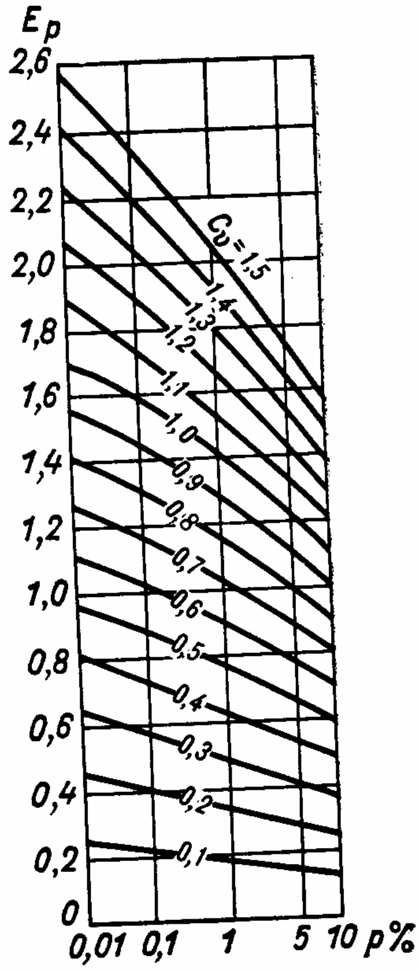
(5.71)

Như vậy, biểu thức (5.68) có thể viết dưới dạng:

$$\sigma_{xp} = \frac{\sigma}{\sqrt{n}}\sqrt{A'}$$

Đưa giá trị của tham số A' (vị trí A) vào phương trình (5.69) ta nhận được biểu thức chính xác hoá đối với δ*xp ứng với trường hợp sử dụng ba tham số tự do (x, Cv, Cs). Biểu thức (5.70) đã được sử dụng để xây dựng toán đồ hiệu chỉnh an toàn (trong "Hướng dẫn xác định các đặc trưng thuỷ văn tính toán " XN 435 - 72. Toán đề này đã được trình bày trên hình 5.13). Khi xây dựng toán đồ này đã sử dụng phương trình (5.70) dưới dạng:

$$Ep\frac{\sigma_{xp}\sqrt{n}}{\overline{x}} = \frac{\sigma}{\overline{x}}\sqrt{A'} = f(Cv, P)$$



Hình 5.13 Toán đồ để xác định hệ số định vị

Trong văn bản chuẩn hoá trước đây, để tính toán lưu lượng lớn nhất (GOST 3999/48) toàn thể được sử dụng đã dựa vào công thức (5.68) E,G.Blôkhinôv đã dưa ra biểu thức cơ cấu trúc đơn giản để đánh giá sai số trung bình bình phương tung độ đường phân phối nhị thức khi $C_S = 2C_V$.

$$\sigma_{xp} = \frac{\sigma}{\sqrt{n}}\sqrt{k_p^2 - \frac{1+C_v^2}{2}\left(\frac{\partial k_p}{\partial C_v}\right)^2} \qquad (5.72)$$

trong đó giá trị của hệ số môdul σ_{xp} lấy trong bảng.

Việc chứng minh công thức (5.72) đã được trình bày trong công trinh trình (79) của Kriski - Menkel.

Bảng 5.2 Độ lệch quân phương của tung độ đường cong đảm bảo mẫu σ_{xp} với các hệ số Cs khác nhau theo n=52; r=0, Cs/Cv=2; bằng phương pháp thử ngẫu nhiên bằng các công thức (5.68-5.71)

P%	C_V	λ	Hai tham số tự do (x, C_V)						
			Công thức		Phương pháp thử thống kê		Công thức		Phương pháp thử thống kê Phương pháp mômen
			(5.68)	(5.69)	Phương pháp mômen	Phương pháp tổng hợp tối đa	(5.70)	(5.71)	
1	2	3	4	5	6	7	8	9	10
0,01	0,1	1,42	0,07	0,07	0,07	0,08	0,14	0,14	0,12
	0,3	2,52	0,34	0,32	0,31	0,30	0,56	0,55	0,42
	0,5	3,98	0,84	0,77	0,73	0,66	1,30	1,26	0,96
	0,7	5,81	1,65	1,48	1,35	1,20	2,48	2,36	1,65
	1,0	9,21	3,72	3,13	2,74	2,28	5,36	4,97	3,00
0,1	0,1	1,34	0,06	0,06	0,06	0,07	0,107	0,107	0,08
	0,3	2,19	0,26	0,25	0,25	0,24	0,39	0,39	0,32
	0,5	3,27	0,62	0,58	0,54	0,51	0,90	0,87	0,68
	0,7	4,56	1,16	1,07	0,99	0,89	1,70	1,66	1,15
	1.0	6,91	2,62	2,20	1,93	1,72	3,63	3,36	2,07
1,0	0,1	1,25	0,04	0,04	0,04	0,05	0,056	0,056	0,05
	0,3	1,83	0,18	0,18	0,17	0,17	0,23	0,22	0,20
	0,5	2,51	0,40	0,39	0,36	0,35	0,50	0,48	0,41
	0,7	3,29	0,72	0,68	0,63	0,59	0,91	0,86	0,68
	1,0	4,61	1,44	1,30	1,16	1,08	1,85	1,71	1,20
10,0	0,1	1,13	0,03	0,03	0,03	0,03	0,029	0,029	0,03
	0,3	1,14	0,10	0,11	0,10	0,10	0,10	0,10	0,10
	0,5	1,67	0,18	0,20	0,19	0,19	0,20	0,20	0,19
	0,7	1,94	0,29	0,30	0,30	0,29	0,32	0,32	0,30
	1,0	2,30	0,49	0,51	0,48	0,47	0,53	0,53	0,48

Do $\sigma_{xp} = Cv_{XP}k_p$, từ công thức (5.72) có thể nhận được biểu thức đối với hệ số biến đổi của tung độ đường tần suất mâu.

$$Cv_{XP} = \frac{\sigma_{xp}}{k_p} = \frac{Cv}{\sqrt{n}}\sqrt{1 + \frac{1+Cv^2}{2k_p^2}\left(\frac{\partial k_p}{\partial C_v}\right)^2} \qquad (5.73)$$

Sai số trung bình bình phương của tung độ đường tần suất (σ_{xp}) với Cv khác nhau và n cố định n = 25, r = 0 và Cs/Cv = 2 nhận được theo phương pháp thử thống kê và theo các công thức (5.68) và (5.71)

Việc đánh giá độ chính xác tính toán sai số trung bình bình phương tung độ đường tần suất mẫu cần phải dựa vào tài liệu mô hình hoá thống kê. Những kết quả đánh giá như vậy đã được trình bày trong bảng 5.2 cho thấy rằng sai số trung bình bình phương của tung độ đường tần suất mẫu có hai tham số tự do tính theo công thức (5.68) và (5.69) lớn hơn so với sai số nhận được bằng phương pháp thử thống kê. Về phàn mình ước lượng δxp nhận được bằng phương pháp thử thống kê đối với các tham số được tính theo phương pháp mômen là lớn hơn ước lượng khi sử dụng các tham số được tính theo phương pháp thích hợp tối đa (tài liệu của Volcova (39). Sự sai khác đó giảm khi tần suất p tăng và rát nhỏ khi P > 5. Các giá trị δxp, được tính bằng phương pháp thử thống kê sử dụng ba tham số tự do.

Những giá trị nhận được theo công thức (5.68) và (5.69) lớn hơn nhiều so với những giá trị tính toán theo phương pháp thử thống kê. Từ đó suy ra việc chính xác hóa tham số A của công thức (5.70) được tiến hành có xét đến mối tương quan giữa các giá trị mẫu x và δ là vẫn chữa đủ.

Những biểu thức lý luận trình bày ở trên cũng để tính như ra đã rõ chỉ đối với các chuỗi không có mối tương quan nội tại và thoả mãn với biểu thức Cs = 2Cv. Những kết luận tổng quát hơn và các giá trị σ_{xp} (hay Cv_{Xp}) phụ thuộc vào dung lượng chuỗi, các hệ số tương quan nội tại và tỷ số Cs/Cv đã nhận được bằng phương pháp thử thống kê. Tài liệu này đã cho thấy rằng hệ số biến đổi của tung độ đường tần suất mẫu tăng khi một số biến đổi gốc, tỷ số Cs/Cv và hệ số tương quan giữa các số hạng kề nhau của chuỗi tăng, giá trị Cv_{xp} giảm khi dung lượng mẫu tăng.

Những hệ số biến đổi của tung độ đường tần suất mẫu (Cs_{Xp}) tăng khi tăng các tham số Cv, Cs/Cv và giảm khi tăng dung lượng mẫu. hệ số tương quan giữa hai số hạng kề nhau của chuỗi sẽ không ảnh hưởng đến tham số này.

Giá trị Cs_{Xp}được xác định đối với các đường tần suất tính theo 2 tham số tự do đứng ra nhỏ hơn các giá trị đó được xác định đối với các đường tần suất tính theo ba tham số tự do khi Cs = 2Cv. Với các quan hệ khác giữa các tham số đó trong tổng thể dạng hàm phân phối sẽ không có ảnh hưởng thêm đến giá trị Cs_{Xp}.

Từ những tài liệu tóm tắt đã được trình bày trên về việc ước lượng các tham số của luật phân phối trên cơ sở phương pháp mô hình hoá thống kê có thể rút ra là phương pháp này cho phép ta giải bài toán nghiên cứu dưới dạng đầy đủ nhất. Khi đó dung lượng thông tin đầy đủ đối với luật phân phối của ước lượng tung độ đường tần suất mẫu là có thể chuyển sang hệ thống rất chặt chẽ có lôgíc của ước lượng tần suất lưu lượng tính toán. Thật vậy thực tế ứng dụng ước lượng, chẳng hạn như lưu lượng tính toán lớn nhất đã sử dụng sự chuẩn hoá xét việc đưa vào tính toán những giá trị có xác suất vượt hàng năm 1 lần trong 1000 năm, diều đó chưa đáng lắm về mặt lôgic. Đối với các giá trị này của lưu lượng nước ứng với xác suất vượt hàng năm còn được dựa vào số hiệu chỉnh an toàn tính theo công thức:

$$\Delta Q = \frac{aE_p}{\sqrt{n}} Qp = \frac{a\sigma_{x_p}}{\overline{Q}} Qp \qquad (5.74)$$

trong đó Qp - lưu lượng nước lớn nhất với tần suất p; $E_p = \frac{\sigma_p \sqrt{n}}{\overline{Q}}$ hàm bổ trợ được xác định theo toán đồ (hình 5.13) σ_{xp}- khoảng lệch trung bình bình phương của tung độ đường tần suất mẫu; n - số số dâng của chuỗi; Qp - trị bình quân của chuỗi; Qp - lưu lượng tính toán ứng với tần suất P; a - hệ số chuẩn hoá đặc trưng cho sự biến động thuỷ văn của sông; a = 0,7 đối với các sông nằm trong các vùng thuỷ văn được nghiên cứu; a = 1,5 đối với các sông nằm ở những vùng thuỷ văn nghiên cứu ít.

Cơ sở của số hiệu chỉnh an toàn là sai số trung bình bình phương của tung độ đường tần suất do sự hạn chế của thời gian quan trắc thuỷ văn (n năm). Cần phải chú ý rằng toán đồ được dùng để tính toán số hiệu chỉnh an toàn đã được xây dựng đối với phân phối xác suất cố định tỷ số Cs/Cv = 2 và đối với các chuỗi không có mối tương quan nội tại.

Như vậy, hệ này rất có ý nghĩa thực tế nhưng chưa phải là đủ liên kết tất cả các khâu của nó.

Có lượng thông tin đầy đủ (nhận được trên cơ sở mô hình hoá thống kê) về đường phân phối tung độ đường tần suất tính toán mẫu đã cho phép xây dựng một hệ rất chặt chẽ các đường tần suất tính toán đem dùng. Trong trường hợp này tần suất tính toán gốc, cơ bản có thể lấy bằng thời đoạn tính toán (thiết kế) khai thác công trình. Thí dụ nếu một công trình thuỷ lợi dự kiến khai thác trong 100 năm thì giá trị tính toán gốc được lấy là đặc trưng thuỷ văn (lưu lượng nước) ứng với tần suất 1%, nếu thời đoạn tính toán công trình làm việc 50 năm thì phải xác định lưu lượng ứng với tần suất là 2% v.v ... Như vậy đó lập lại trung bình của lưu lượng nước được lấy theo thời gian khai thác công trình.

Tiếp theo, phụ thuộc vào loại đầu tư vốn (hay mức độ quan trọng) của công trình mà lấy mức tin cậy chuẩn hoá của tung độ đường tần suất dùng (q%) hay mức sử dụng (1 - q%). Do đó mức độ tin cậy của tính toán thuỷ văn lấy theo loại vồn đầu tư của công trình. Thí dụ, đối với các công trình loại 1 và loại 2 được tính với thời gian khai thác như nhau (dường như 100 năm) lưu lượng nước lớn nhất tính toán cần phải lấy cùng tần suất (1%). Song nếu lấy xác suất lưu lượng đó (giới hạn trên của độ tin cậy) đối với công trình loại 1 sẽ lấy chẳng hạn là 99,9% thì đối với công trình loại 2 nó có thể lấy bằng 99% v.v...

Một hệ thống sử dụng tương tự của các tần suất tính toán không thể thực hiện được khi không có phương pháp mô hình hoá thống kê, vì lời giải lý luận về sai số trung bình bình phương của các chuỗi có độ dài khác nhau, có mối tương quan nội tại và đặc trưng bằng các tỷ số khác nhau của Cs và Cv, không nhận được. Ưu điểm đặc biệt đó là đường tần suất x_p, hệ số không đối xứng chỉ có thể nghiên cứu được bằng phương pháp thử thống kê.