Chương 5

PHƯƠNG PHÁP THỐNG KÊ XÁC SUẤT DÙNG TRONG TÍNH TOÁN THỦY VĂN

--- oOo ---

5.1 KHÁI NIỆM CHUNG

Các hiện tượng xảy ra trong thế giới tự nhiên có thể chia làm 2 loại: *hiện tượng tất nhiên và hiện tượng ngẫu nhiên.*

- **Hiện tượng tất nhiên:** là những hiện tượng phát sinh và diễn biến theo một qui luật nào đó theo những điều kiện nhất định. Khi các điều kiện hoặc trạng thái thay đổi, ta có thể biết trước được quá trình và tính chất của hiện tượng.

Ví dụ 5.1: Dưới áp suất không khí, khi đun nước đến 100°C, hiện tượng nước sôi và bốc hơi, còn khi hạ nhiệt độ xuống 0°C, hiện tượng nước đóng băng sẽ xảy ra.

- **Hiện tượng ngẫu nhiên:** là những hiện tượng mà ta không thể khẳng định trước sự phát sinh, phát triển của chúng. Trong một điều kiện nào đó, chúng có thể xảy ra như thế này hoặc thế khác và thậm chí không thể xảy ra.

Ví dụ 5.2: Thảy một con xúc sắc, ta hoàn toàn không biết trước mặt nào sẽ xuất hiện, có thể là 1 hoặc 2, hoặc 4 hoặc 6, ...Trong vài lần tung, ta hoàn toàn không xác định khả năng xuất hiện của một giá trị nào của nó.

Khi quan sát hiện tượng ngẫu nhiên trong một số ít lần ta không xác định được qui luật cụ thể, nhưng nếu quá trình lập đi lập lại nhiều lần và thống kê tất cả các xuất hiện đã xảy ra, ta có thể tìm thấy một qui luật chung nào đó, ta gọi đó là qui luật đám đông.

Việc phân loại như trên chỉ có tính chất tương đối. Thực tế có những hiện tượng vừa mang tính tất nhiên vừa mang tính ngẫu nhiên. Ví dụ: tại một địa điểm nhất định và với một thời điểm nào đó trong tương lai, ta không thể biết trước mưa sẽ rơi hay không? Khả năng mưa rơi này mang tính ngẫu nhiên. Tuy nhiên, với các thống kê lâu dài ta có thể phán đón khả năng có mưa hay không có vùng này vào thời điểm đó. Sự xác định này mang tính tất nhiên. Sở dĩ như vậy vì nguyên nhân phát sinh ra chúng rất phức tạp, tác động lẫn nhau, trong đó có các nguyên nhân bên trong của hiện tượng thúc đẩy nó phát sinh và diễn biến theo qui luật, lại có các nguyên nhân bên ngoài làm cho hiện tượng có tính ngẫu nhiên. Khi các nguyên nhân này biến động, tính chất của nó cũng biến động theo. ***Kết quả là cùng một hiện tượng, trong điều kiện này tính tất nhiên chiếm ưu thế, trong điều kiện khác tính ngẫu nhiên tăng lên có khi chiếm địa vị ưu thế.***

Xuất phát từ tính ngẫu nhiên của các hiện tượng thủy văn, ta có thể nghiên cứu áp dụng một số khái niệm và phương pháp trong lý thuyết thống kê xác xuất để tìm ra qui luật thống kê của các đặc trưng thủy văn, xác định trị số thiết kế công trình.

5.2 BIẾN CỐ NGẪU NHIÊN VÀ XÁC SUẤT

5.2.1. Biến cố

Do tính chất đám đông của hiện tượng ngẫu nhiên,muốn nghiên cứu qui luật của một hiện tượng ngẫu nhiên nào đó, ta phải tiến hành lập lại hoặc quan trắc rất nhiều một thực nghiệm. Tập hợp các hiện tượng có thể xảy ra trong quan trắc được gọi là biến cố.

- Biến cố nhỏ nhất hoặc đơn giản nhất trong quan trắc gọi là biến cố sơ cấp.
- Kết hợp các biến cố sơ cấp theo một tổng hợp nào đó tạo thành một biến cố ngẫu nhiên.
- Không gian chứa các biến cố sơ cấp được ký hiệu là E. Biến cố ngẫu nhiên được ký hiệu bằng các chữ hoa : A, B, C, ...

Ví dụ 5.3:

Tung một con súc sắc, ta nhận được 6 trường hợp xuất hiện của mặt số: 1, 2, 3, 4, 5 và 6. Mỗi mặt số được gọi là một biến số sơ cấp E. Tổ hợp các mặt số chẵn là 2, 4 và 6 được gọi là biến cố ngẫu nhiên A.

$$E = \{ 1, 2, 3, 4, 5, 6 \}$$
$$A = \{ 2, 4, 6 \}$$

Phân loại biến cố:

🕮 **Biến cố chắc chắn**

Là các biến cố chắc chắn xảy ra trong mọi trường hợp thực nghiệm.

Ví dụ 5.4:

Trong phép gieo xúc sắc, biến cố xuất hiện các mặt số nguyên nhỏ hơn hoặc bằng 6 là một biến cố chắc chắn.

🕮 **Biến cố không**

Là biến cố không bao giờ xuất hiện trong mọi lần thực nghiệm.

Ví dụ 5.5:

Biến cố xuất hiện mặt số 7 trong phép gieo xúc sắc là một biến cố không (ϕ).

📖 **Biến cố tổng**

Biến cố C được gọi là biến cố tổng của hai biến cố A và B khi C chứa ít nhất có 1 trong 2 biến cố A hoặc B xuất hiện. Nghĩa là, hoặc A xuất hiện (B không xuất hiện), hoặc B xuất hiện (A không xuất hiện) hoặc cả A và B đều xuất hiện.

Ký hiệu: **C = A + B** hoặc **C = A ∪ B** (đọc là A hội B) (5-1)

Ví dụ 5.6:

Tung một con xúc sắc: A = {e_1, e_2, e_3 }, B = {e_1, e_3, e_4}

Biến cố tổng C = A + B = { e_1, e_2, e_3, e_4 }

📖 **Biến cố tích**

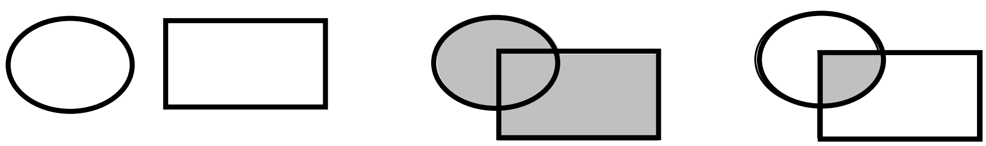
D được gọi là biến cố tích của hai biến cố A và B khi D chứa đồng thời các biến cố của A và của B .

Ký hiệu: **D = A . B** hoặc **D = A ∩ B** (đọc là A giao B) (5-2)

Ví dụ 5.7:

giống ví dụ trên, biến cố tích D = {e_1, e_3}

Hình vẽ 5.1 sau minh họa hai định nghĩa trên, A là tập hợp các biến cố điểm rơi vào hình tròn như hình (a), B là tập hợp các biến cố điểm rơi vào hình chữ nhật như hình (b). C là biến cố tổng của A và B là tập hợp các biến cố điểm như hình (c) và D là biến cố tích của A và B là tập hợp điểm như hình (d).



A hình (a) B hình (b) C = A + B hình (c) D = A.B hình (d)

Hình 5.1 Minh họa khái niệm tổng và tich 2 biến cố

📖 **Biến cố xung khắc**

Là biến cố không thể xuất hiện đồng thời đồng thời xuất hiện trong mỗi lần quan trắc. Các biến cố sơ cấp xung khắc với nhau từng đôi một.

Ví dụ 5.8:

Gieo một đồng xu, trong mỗi lần quan sát, không thể nào có sự xuất hiện cả mặt sấp lẫn mặt ngữa. Gieo một súc sắc cũng không thể có mặt số vừa chẵn và vừa lẻ.

🕮 Biến cố đối

$A^{/}$ là biến cố đối của biến cố A nếu $\mathbf{A \cup A^{/} = E}$ và $\mathbf{A \cap A^{/} = \phi}$. Biến cố đối cũng đồng thời là biến cố xung khắc, nhưng hai biến cố xung khắc không nhất thiết là hai biến cố đối nhau, bởi vì biến cố đối yêu cầu $A \cup A^{/} = E$, còn biến cố xung khắc không có yêu cầu điều kiện đó.

Ví dụ 5.9:

Tung một con xúc sắc, nếu : $A = \{e_2, e_4, e_6\}$ thì $A^{/} = \{e_1, e_3, e_5\}$

🕮 Biến cố độc lập

A và B là hai biền cố độc lập khi $A \cap B = \phi$. Nghiã là A và B không có chung một phần tử nào trong không gian biến cố sơ cấp, sự xuất hiện của A không ảnh hưởng gì đến sự xuất hiện của B và ngược lại .

Ví dụ 5.10:

A là tập hợp các biến cố trong các lần tung con xúc sắc, B là tập hợp các biến cố trong phép gieo đồng tiền. A và B là hai biến cố độc lập.

5.2.2 Xác suất

Các biến cố khác nhau có khả năng xuất hiện khác nhau, bằng cách dùng trị số số học để biểu thị cụ thể số đo số lần xuất hiện của 1 biến cố nào đó trong một loạt các quan trắc, trị số số học đó gọi là *xác suất* (probability). *Xác suất là số đo khả năng xuất hiện của biến số*. Xác suất càng lớn thì khả năng xuất hiện biến số càng lớn.

Xác suất thường ký hiệu là P, xác suất xuất hiện của biến cố A được ghi là P (A), bao giờ ta cũng có:

$$\mathbf{0 \le P(A) \le 1}$$

Trong một lần thực nghiệm có n biến cố sơ cấp, trong đó có m là biến cố sơ cấp thuận lợi cho biến cố A xuất hiện, thì xác xuất xuất hiện của A sẽ là :

$$\mathrm{P}(A) = \frac{m}{n} \qquad (5\text{-}3)$$

Ví dụ 5.11:

Một hộp có 5 viên bi, trong đó có 3 bi trắng và 2 bi đen. Bốc ngẫu nhiên 1 viên, Gọi A là khả năng bốc trúng 1 bi đen, xác suất của A sẽ là :

$$P(A) = \frac{2}{5}$$

5.2.3 Tần suất

Công thức tính xác suất chỉ phù hợp khi các biến cố xuất hiện là đồng khả năng, ví dụ con xúc sắc là cân đối, các viên là đồng chất, đồng kích thước . . . Tuy nhiên, trong tự nhiên có những hiện tượng xảy ra không đều, chẳng hạn như khả năng xuất hiện mực nước lớn trong từng năm, nên ta không thể đơn giản suy ra các qui luật xuất hiện của chúng. Qua thống kê nhiều năm, ghi nhận và phân cấp, ta có khái niệm *tần suất* (frequency) xuất hiện. ***Tần suất xuất hiện của biến cố A là tỷ số giữa số lần xuất hiện của biến cố đó, hay tần số m, với số lần quan trắc***.

Ví dụ 5.12: Mực nước lớn nhất tại một trạm thủy văn trong 15 năm cho ở bảng sau:

Năm	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94
H(m)	7,8	4,5	6,2	5,5	8,8	7,3	3,6	4,9	5,1	6,2	6,3	8,4	4,6	5,0	7,2
P	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$

Ở đây, tần suất xuất hiện của biến cố tổng là :

$$P(C) = \sum_{1}^{15} \frac{1}{15} = 1 = 100\%$$

Nói cách khác, ***xác suất xuất hiện của biến cố A là tần suất xuất hiện của biến cố đó khi số quan trắc tăng lên vô hạn, n $\rightarrow \infty$.***

$$P(A) = \lim_{n \to \infty} \frac{m}{n} \qquad (5\text{-}4)$$

Định lý cộng xác suất

Xác suất của tổng 2 biến cố xung khắc nhau bằng tổng xác suất của hai biến cố đó.

$$\mathbf{P(A + B) = P(A) + P(B)} \qquad (5\text{-}5)$$

Theo ví dụ 5.12 ở trên:

Tần suất xuất hiện trị số lớn hơn hoặc bằng H = 7,5 m sẽ là:

$$P(H \geq 7{,}8 \text{ m}) = P(7{,}8) + P(8{,}8) + P(8{,}4) = \frac{1}{15} + \frac{1}{15} + \frac{1}{15} = \frac{3}{15} = 0{,}20 \text{ hay } 20\%$$

Định lý nhân xác suất

Xác suất của tích hai biến cố bằng xác suất của biến cố thứ nhất nhân với xác suất có điều kiện của biến cố thứ hai.

$$\mathbf{P(A\,.B) = P(A)\,.\,P(B/A) = P(B)\,.\,P(A/B)} \qquad (5\text{-}6)$$

Trong đó P(B/A) là xác suất có điều kiện, là xác suất xuất hiện của biến cố B khi biến cố A đã xảy ra, còn P(A/B) là xác suất có điều kiện của biến cố A khi biến cố B đã xảy ra.

Ví dụ 5.13:

Bước đầu nghiên cứu quan hệ mưa rào và dòng chảy lũ của lưu vực X, người ta chia dòng chảy lũ thành 3 cấp lũ: lũ lớn A1, lũ trung bình A2 và lũ nhỏ A3 về mùa mưa cũng có 3 cấp tương ứng: mưa lớn B1, mưa trung bình B2 và mưa nhỏ B3. Thống kê 100 dòng chảy lũ ta có số các con lũ lớn, nhỏ xuất hiện tương ứng với các lượng mưa lớn, nhỏ như sau :

Biến cố	Lũ lớn A1	Lũ TB A2	Lũ nhỏ A3	Tổng
Mưa lớn B1	15	8	0	23
Mưa trung bình B2	4	49	4	57
Mưa nhỏ B3	0	6	14	20
Tổng	19	63	18	100

Từ bảng trên ta thấy :

Xác suất xuất hiện mưa lớn dưới điều kiện lũ lớn :

$$P\,(B1/A1) = \frac{15}{19} = 0{,}7895$$

Xác suất xuất hiện lũ lớn trong 100 con lũ:

$$P\,(A1\,) = \frac{19}{100} = 0{,}19$$

Xác suất xuất hiện lũ lớn cùng với mưa lớn trong tổng 100 con lũ

$$P\,(A1 \cap B1) = \frac{15}{100} = 0{,}15$$

Rõ ràng: $P(A1 \cap B1) = P(A1.B1) = P(A1)\,.\,P(B1/A1) = \frac{19}{100} \times \frac{15}{19} = 0.15$

Tính chất của biến cố điều kiện

Nếu A và B độc lập, thì

$$\mathbf{P(A/B) = P(A) \quad hay \quad P(A.B) = P(A)\,.\,P(B)} \qquad (5\text{-}7)$$

Ví dụ 5.14:

Một trạm bơm có 2 máy bơm hoạt động độc lập với nhau. Xác suất hư hỏng của mỗi máy là 10%. Vậy xác suất hư hỏng của cả 2 máy đồng thời sẽ là:

$$P = 10\%\ .\ 10\% = 1\%$$

5.3 PHÂN BỐ XÁC SUẤT CỦA BIẾN CỐ NGẪU NHIÊN

5.3.1 Đại lượng ngẫu nhiên

- *Đại lượng ngẫu nhiên rời rạc*: Lấy 1 đoạn [a, b] nào đó, nếu các đại lượng ngẫu nhiên trong [a, b] là một số đếm các trị số thì ta gọi đó là đại lượng ngẫu nhiên rời rạc.

Ví dụ 5.15:
Rút các con bài trong cổ bài 52 lá, tung một con súc sắc, gieo một cập đồng tiền ... là các đại lượng ngẫu nhiên rời rạc.

- *Đại lượng ngẫu nhiên liên tục*: Nếu trong đoạn [a, b], đại lượng ngẫu nhiên có vô cùng trị số, có thể lấy bất kỳ giá trị nào trong khoảng [a, b] thì ta gọi đó là đại lượng ngẫu nhiên liên tục.

Ví dụ 5.16:
Mực nước lũ biến thiên được xem như là một đại lượng ngẫu nhiên liên tục. Tuy nhiên, nếu lấy mực nước bình quân thời đoạn thì nó trở thành đại lượng ngẫu nhiên rời rạc.

Trong thủy văn, các trị số và lớn nhất như lưu lượng max... có ý nghĩa đặc biệt quan trọng. Cho nên khi tính toán cần xác định xác suất của X (mưa, dòng chảy,...) rơi vào khoảng trị số x_i nào đó đến trị số X max, nhưng thường Xmax không xác định chắc chắn trước nên ta thường tính xác suất để cho $X \geq x_i$ là: $P(X \geq x_i)$ với hàm ý là xác suất hay tần suất để x nằm trong khoảng $[x_i$, Xmax].

Từ khái niệm trên ta có thể hiểu rằng ứng với mỗi giá trị x của đại lượng ngẫu nhiên có một xác suất tương ứng, trong thủy văn còn gọi lá xác suất vượt, vì nó biểu thị xác suất của các giá trị đại lượng ngẫu nhiên lớn hơn hoặc bằng một giá trị nào đó, nó có thể mô tả một cách trực quan, ví dụ: xác suất vỡ đê = xác suất mực nước lớn hơn hoặc bằng cao trình đê ... Ta có thể *áp dụng công thức cộng xác suất để lũy tích (hay tích phân) xác suất (hay tần suất) của các khoảng nhỏ nằm trong đó, kết quả tìm được gọi là xác suất (hay tần suất) lũy tích*.

5.3.2 Mẫu và tổng thể

Muốn nghiên cứu biến ngẫu hiên nói chung và các đặc trưng thủy văn nói riêng ta cần có tài liệu quan trắc. Trong thống kê học, ta gọi n trị số riêng biệt x_1, x_2, x_3,...., x_n quan trắc được của một biến cố ngẫu nhiên nào đó được gọi là dung lượng mẫu và toàn thể trị số của biến cố ngẫu nhiên là tổng thể.

Thực tế trong thủy văn, ta không thể nào lấy được các tổng thể các trị số vì thời gian bắt đầu thu thập dữ liệu bao giờ cũng ngắn hơn thời gian tồn tại của hiện tượng tự nhiên.

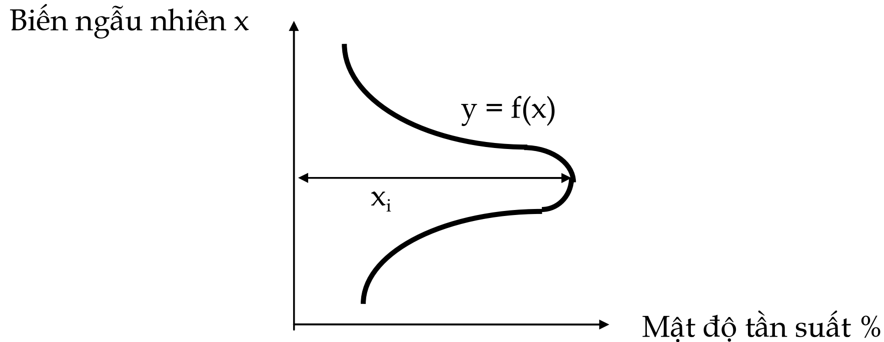
5.3.3 Phân bố xác suất của biến cố ngẫu nhiên

a) Đường phân bố mật độ tần suất

Giả thiết một cách lý tưởng là tổng thể các trị số của biến cố ngẫu nhiên trong đặc trưng thủy văn đều biết trong khoảng [a, b]. Chia khoảng [a, b] thành n khoảng nhỏ [a, x_1], [x_1, x_2], ..., [x_{n-1}, b]. Độ lớn của khoảng thứ i nào đó là $\Delta x_i = x_i - x_{i-1}$ và có tần suất tương ứng là P_i. Mật độ tần suất trung bình của khoảng này sẽ là $\frac{P_i}{\Delta x_i}$ khi $\Delta x_i \to 0$

Ta có: $\lim\limits_{\Delta x_i \to 0} \frac{p_i}{\Delta x_i} = f(x_i)$ được gọi là mật độ tần suất phân bố tại điểm x_i.

Do Δx_i có thể lấy nhỏ bao nhiêu tùy ý nên ứng với mỗi trị số của x đều có f(x) tương ứng nên f(x) gọi là hàm mật độ tần suất của biến cố ngẫu nhiên liên tục, như hình 5.2.

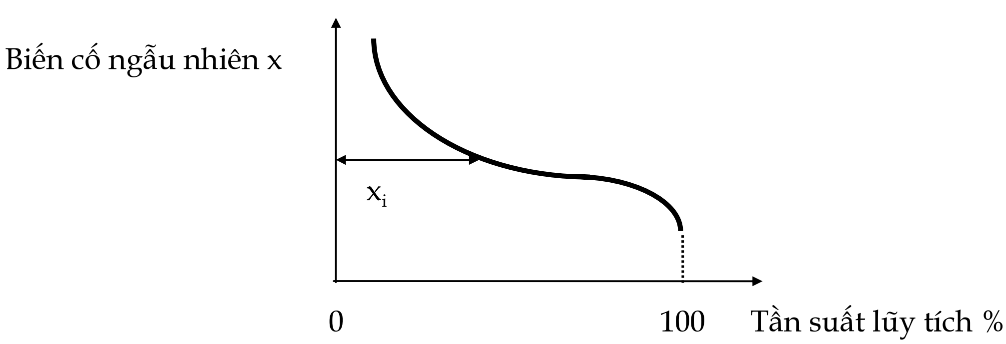


Hình 5.2

Đồ thị y = f (x) là đường phân bố mật độ tần suất của biến ngẫu nhiên liên tục

b) Đường tần suất lũy tích

Tích phân $\int_{x_i}^{b} f(x)d(x) = P(X \geq x_i)$ biểu thị tần xuất lũy tích của biến trong khoảng [x_i ,b]. Ứng với mỗi x_i ta có P (X ≥ x_i) được gọi là *hàm tần suất lũy tích*, còn đồ thị gọi là *đường tần suất*. Trong tính toán thủy văn, tần số lũy tích được gọi tắt là tần suất.



Hình 5.3: Đường tần suất của ngẫu nhiên x

Ví dụ 5.17:

Lấy 50 trị số đo đạt về lưu lượng của một trạm thủy văn từ năm 1920 đến 1969, trị số lớn nhất là 2 650 m^3/s, nhỏ nhất là 860 m^3/s, trung bình là 1 450 m^3/s. Vì biến cố ngẫu nhiên liên tục nên ta tiến hành phân cấp, mỗi cấp 300 m^3/s, sắp thứ tự từ lớn đến nhỏ rồi thống kê số lần xuất hiện các trị số rơi vào các cấp lưu lượng. Số lần xuất hiện này là tần số và ký hiệu là f.

- Tính tần suất mỗi cấp lưu lượng theo công thức: $P\ \% = \frac{f}{n} \times 100\ \%$ (n = 50)
- Lấy giá trị P chia cho độ lớn của cấp (300 m^3/s) ta được mật độ tần suất bình quân của cấp lưu lượng tương ứng.
- Lũy tích tần suất từ trên xuống dưới ta được tần suất lũy tích ứng với các trị số giới hạn dưới của mỗi cấp lưu lượng P ($Qm \geq Qm_i$).
- Lập bảng kết quả tính toán như sau:

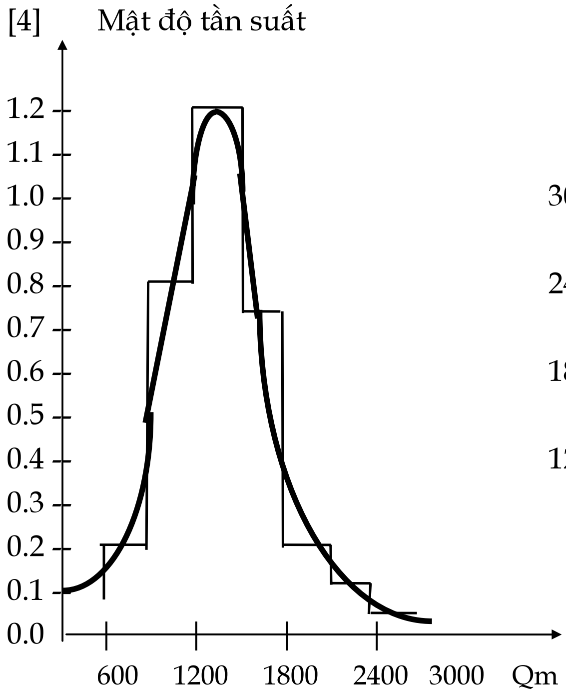
Bảng 5.1 : Tính tần suất Qmax

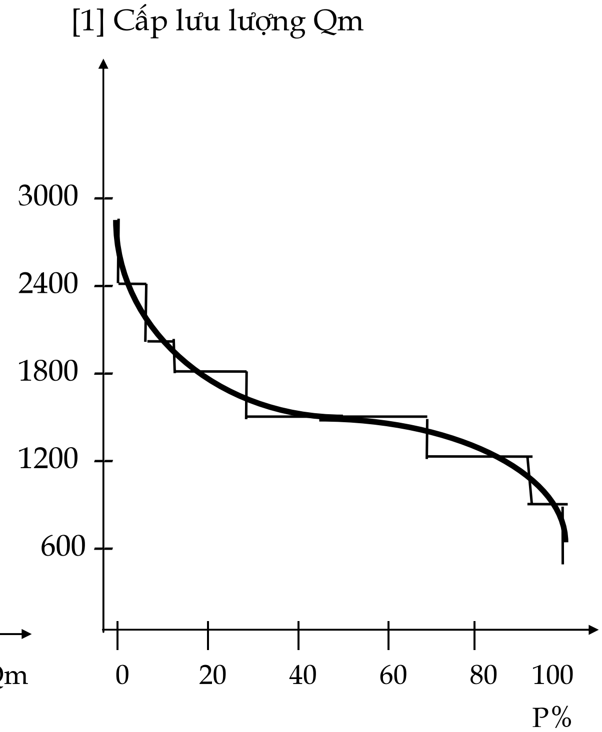
Cấp lưu lượng Qm (m^3/s)	Tần số f lần	Tần suất $P(\%) = \frac{f}{n} \times 100\ \%$	Mật độ tần suất $\frac{P}{300}\frac{(\%)}{(m^3/s)}$	Tần suất lũy tích $\Sigma P\% = P(Qm \geq Qm_i)$
[1]	[2]	[3]	[4]	[5]
2700 - 2400	1	2.0	0.06	2.0
2399 - 2100	2	4.0	0.12	6.0
2099 - 1800	3	6.0	0.20	12.0
1799 - 1500	11	22.0	0.73	34.0
1499 - 1200	18	36.0	1.20	70.0
1199 - 900	12	24.0	0.80	94.0
899 - 600	3	6.0	0.20	100.0
Tổng	50	100 %		

- Lấy cột [4] làm tung độ, cột [1] làm hoành độ, ta được đồ thị phân bố mật độ lưu lượng đỉnh lũ.

- Lấy cột [1] làm tung độ, cột [5] làm hoành độ, ta có đường tần suất lũy tích lưu lượng đỉnh lũ.

Đường phân bố mật độ tần suất lưu lượng đỉnh lũ và đường tần suất lũy tích lưu lượng được vẽ như hình 5.4 và 5.5





Hình 5.4
Đường phân bố mật độ tần suất
lưu lượng đỉnh lũ

Hình 5.5
Đường tần suất lũy tích
lưu lượng đỉnh lũ

5.4 ĐƯỜNG TẦN SUẤT KINH NGHIỆM

Đường tần suất kinh nghiệm trong thuỷ văn là đường tần suất được xây dựng từ mẫu tài liệu thực đo về một đặt trưng thủy văn nào đó của một trạm thủy văn nhất định, nó chỉ phản ánh tình hình đặc trưng của trạm đó mà không đúng với trạm khác.

5.4.1 Phương pháp vẽ đường tần suất kinh nghiệm

Muốn có đường tần suất, ta cần thu thập các chuỗi số liệu quan trắc nhiều năm, nếu số liệu dài trên 50 năm, ta có thể dùng cách phân cách thống kê như ở phần trước. Tuy nhiên, các trạm đo thủy văn hiện nay ở Việt Nam thường ngắn và bị gián đoạn.Trong trường hợp này ta không cần phân cấp thống kê mà có thể theo phương pháp vẽ đường tần suất kinh nghiệm sau:

- Sắp xếp theo thứ tự giảm dần từ lớn đến nhỏ.
- Tìm số lần suất hiện trị số “ bằng và lớn hơn “ một trị số nào đó rồi tính ra tần suất lũy tích :

$$P(x \geq x_i) = \frac{m}{n} 100\% \qquad (5\text{-}8)$$

với m là số trị thuận lợi, n là số năm quan trắc.

- Lấy X [Q, H, ...] làm tung độ và $P(x \geq x_i)$ làm hoành độ. Vẽ đường cong trơn đi qua trung tâm các điểm.

Ví dụ 5.18:
Một trạm thủy văn cho chuỗi số liệu lưu lượng đỉnh lũ hàng năm con sông A từ 1923 đến 1942 như cột [2] và [3] trong bảng 5.2.

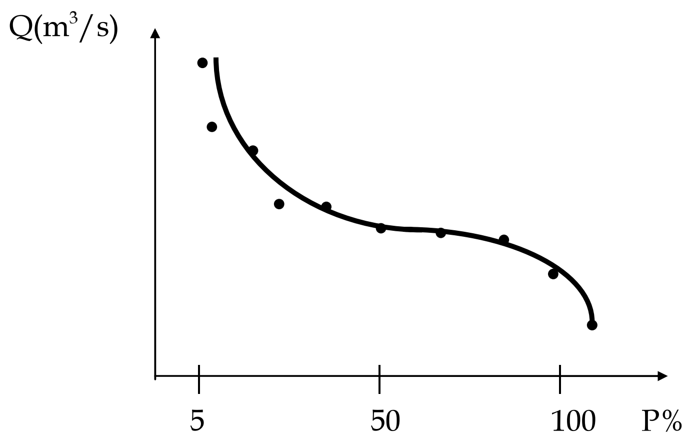
Bảng 5.2: Lưu lượng đỉnh lũ sông A

STT	Năm	Qmax (m^3/s)	STT	Năm	Qmax (m^3/s)
[1]	[2]	[3]	[1]	[2]	[3]
1	1923	176	11	1933	284
2	1924	212	12	1934	264
3	1925	234	13	1935	275
4	1926	147	14	1936	213
5	1927	288	15	1937	188
6	1928	215	16	1938	221
7	1929	262	17	1939	242
8	1930	250	18	1940	189
9	1931	192	19	1941	245
10	1932	167	20	1942	196

Yêu cầu vẽ đường tần suất dòng chảy lũ con sông A.

Giải:

- Tiến hành sắp xếp lưu lượng từ lớn đến nhỏ và tính tần suất xuất hiện trị số "lớn hơn hoặc bằng" trị tính toán như bảng tính ở trang kế.
- Sau đó vẽ đường tần suất kinh nghiệm Qmax sông A.



Hình 5.6 Đường tần suất kinh nghiệm Qmax sông A

Bảng 5.3: Tính tần suất lũy tích P ($x \ge x_i$)

STT	Năm	Qmax (m^3/s)	Qi sắp thứ tự lớn → nhỏ	Tần suất $P(Q \ge Qi) = \frac{m}{n} 100\ \%$
[1]	[2]	[3]	[4]	[5]
1	1923	176	288	(1/20) x 100 % = 5 %
2	1924	212	284	(2/20) x 100 % = 10 %
3	1925	234	275	(3/20) x 100 % = 15 %
4	1926	147	264	(4/20) x 100 % = 20 %
5	1927	288	262	(5/20) x 100 % = 25 %
6	1928	215	250	(6/20) x 100 % = 30 %
7	1929	262	245	(7/20) x 100 % = 35 %
8	1930	250	242	(8/20) x 100 % = 40 %
9	1931	192	234	(9/20) x 100 % = 45 %
10	1932	167	221	(10/20) x 100 % = 50 %
11	1933	284	215	(11/20) x 100 % = 55 %
12	1934	264	213	(12/20) x 100 % = 60 %
13	1935	275	212	(13/20) x 100 % = 65 %
14	1936	213	196	(14/20) x 100 % = 70 %
15	1937	188	192	(15/20) x 100 % = 75 %
16	1938	221	189	(16/20) x 100 % = 80 %
17	1939	242	188	(17/20) x 100 % = 85 %
18	1940	189	176	(18/20) x 100 % = 90 %
19	1941	245	167	(19/20) x 100 % = 95 %
20	1942	196	147	(20/20) x 100 % = 100 %

5.4.2 Công thức tính tần suất kinh nghiệm

Trường hợp trên tính $P(x \ge x_i) = \frac{m}{n} 100\ \%$ không hợp lý ở chỗ số hạng nhỏ nhất của mẫu có tần suất là 100 %, nghĩa là không có trị số nào sau này nhỏ hơn nữa, nên chỉ áp dụng khi n vô cùng lớn.

Trong thực tế tính toán hiện nay, người ta dùng một số công thức tính toán kinh nghiệm, các công thức này không được chứng minh bằng các thuật toán, nhưng được biến đổi từ thực tế kinh nghiệm để tăng khả năng an toàn khi sử dụng để khắc phục nhược điểm nói trên.

Các công thức sau được biến đổi từ $P(x \ge x_i) = \frac{m}{n} 100\ \%$ bằng cách làm tăng trị mẫu số hoặc/và giảm tử số.

- Công thức trung bình : $P_1 = \dfrac{m-0,5}{n} \times 100\%$ (5-9)

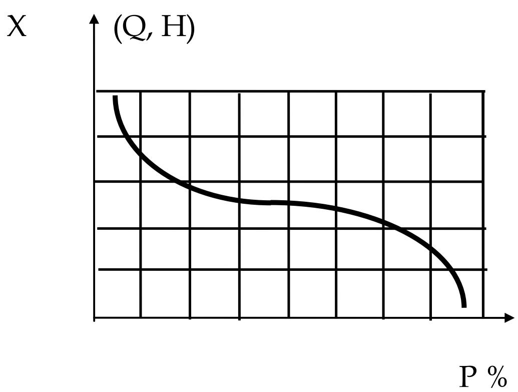
- Công thức vọng số : $P_2 = \dfrac{m}{n+1} \times 100\%$ (5-10)

- Công thức số giữa : $P_3 = \dfrac{m-0,3}{n+0,4} \times 100\%$ (5-11)

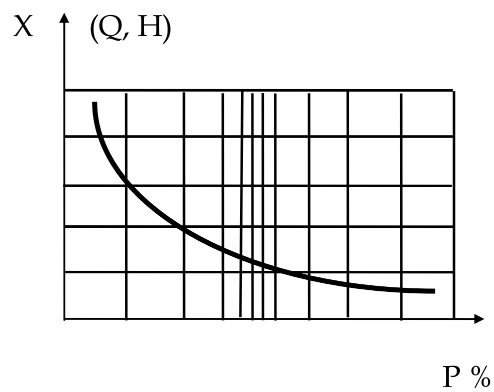
Qua 3 công thức trên, ta thấy trị P_2 là an toàn nhất nên được dùng phổ biến nhất.

5.4.3 Ngoại suy đường tần suất kinh nghiệm

Đường tần suất kẻ trên giấy ô vuông thường hai đầu rất dốc, đoạn này có ý nghĩa nhất, vì trị số đặt biệt ở hai đầu xuất hiện ít, do đó dễ dẫn đến sai lầm chủ quan. Thực tế, người ta thường dùng giấy tần suất để vẽ. Giấy tần suất là giấy vẽ sẵn hai trục, trục hoành hai đầu thưa ra, ở giữa tập trung theo phân bố logarit, còn trục tung thì chia đều. *(Xem mẫu giấy tần suất ở phần phụ lục, trang 137)*. Vẽ đường tần suất trên giấy này sẽ khắc phục được tình trạng dốc ở hai đầu.



Hình 5.7
Đường tần suất trên giấy ô vuông



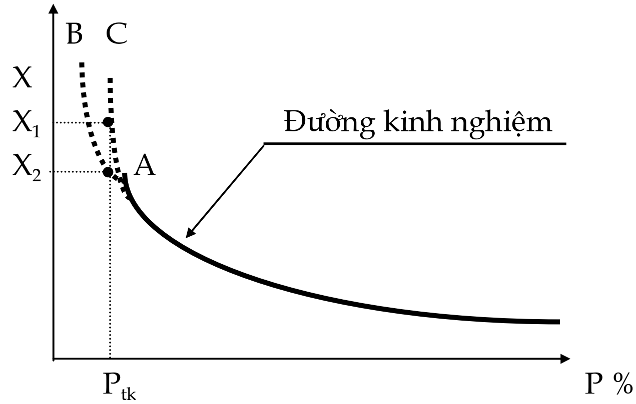
Hình 5.8
Đường tần suất trên giấy tần suất

Khi tính toán thủy văn cho các công trình chống lũ thường yêu cầu có các tần suất thiết kế rất nhỏ, ví dụ 1 % thậm chí 0,01 %. Trong khi đó, tài liệu đo đạt thủy văn thường rất ngắn, ví dụ n = 30 năm, thì P % tính theo

$$P\% = \frac{m}{n+1} \times 100\% = \frac{1}{30+1} \times 100\% = 3,22\%$$

Giá trị này vẫn lớn hơn trị cần thiết kế, giả sử là 0,1 % chẳng hạn. Người ta tìm cách kéo dài đường tần suất kinh nghiệm, nhưng rất khó phán đoán mức độ chính xác của đường kéo dài.

Ví dụ như hình 5.9



Hình 5.9: Kéo dài đường kinh nghiệm

Từ điểm A của đường tần suất kinh nghiệm, ta không thể khẳng định đường kéo dài nào AB hay AC là chính xác hơn ? Hai đường này sẽ dẫn đến 2 trị X_1 và X_2 rất sai biệt ứng với P_{tk}.

Do đó, người ta nghiên cứu một phương trình toán học nào đó để khái quát hóa đường phân bố mật độ tần suất và gọi đó là ***đường tần suất lý luận***. Việc xác định đường này có liên quan chặt chẽ đến một số trị đặt trưng cần xem xét sau đây.

5.5 CÁC TRỊ SỐ ĐẶT TRƯNG THỐNG KÊ

5.5.1 Các trị số đặt trưng biểu thị xu thế tập trung

a/ Số bình quân $\overline{x}$

Giả sử ta có một chuỗi các liệt quan trắc được x_1, x_2, x_3 ,..., x_n. Số bình quân của liệt số đó là:

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (5\text{-}12)$$

Nếu tần số của x_i là f_i (kể cả trường hợp chia nhóm) thì:

$$\overline{x} = \frac{x_1 f_1 + x_2 f_2 + \cdots + x_n f_n}{f_1 + f_2 + \cdots + f_n} = \frac{1}{n}\sum_{i=1}^{n} x_i f_i = \sum_{i=1}^{n} x_i \frac{f_i}{n} = \sum_{i=1}^{n} x_i P_i \qquad (5\text{-}13)$$

với $n = f_1 + f_2 + ... + f_n$ và p_i là tần suất của x_i .

Trị số bình quân rất quan trọng trong tính toán thủy văn, nó phản ánh tình hình chung của một liệt số, nhưng nếu mẫu không dài, trị bình quân dễ bị cực trị ảnh hưởng.

b/ Kỳ vọng toán học M(x)

Nếu X là đại lượng ngẫu nhiên liên tục và tồn tại kỳ vọng thì :

$$M(x) = \int_{-\infty}^{+\infty} x\, f(x)\, dx \qquad (5\text{-}14)$$

Nếu X là đại lượng ngẫu nhiên rời rạc và tồn tại kỳ vọng thì :

$$M(x) = \sum_{i=1}^{n} x_i\, p_i \qquad (5\text{-}15)$$

trong đó $p_i = P(x = x_i)$

Ví dụ 5.19:

Tính kỳ vọng của đại lượng ngẫu nhiên x có luật phân phối xác suất sau :

X	1	2	3
P_i	1/4	1/2	1/4

Theo công thức đại lượng ngẫu nhiên rời rạc, ta có

$$M(x) = 1\frac{1}{4} + 2\frac{1}{2} + 3\frac{1}{4} = 2$$

Nếu trong công thức kỳ vọng của đại lượng ngẫu nhiên rời rạc các giá trị x_i xuất hiện đều nhau (đồng khả năng) $p_i = 1/n$, ta có:

$$M(x) = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x} \qquad (5\text{-}16)$$

$\bar{x}$ chính là số bình quân của chuỗi x_i. Từ đó ta thấy kỳ vọng toán chính là số bình quân gia quyền mà lấy quyền số là xác suất xuất hiện của x_i. Vì vậy kỳ vọng toán biểu thị trung tâm của đại lượng ngẫu nhiên, vị trí trung tâm của hình mật độ xác suất.

c/ Số đông $X_đ$

Trên đường phân bố mật độ tần suất, trị số ứng với mật độ tần suất lớn nhất gọi là *số đông*, so với các số khác của biến cố ngẫu nhiên thì số đông có khả năng xuất hiện nhiều nhất. Đối với mẫu tài liệu thủy văn, muốn tìm số đông, cần phải tiến hành phân cấp thống kê.

Trong thí dụ số liệu trên sông A, cấp Qmax từ 1499 - 1200 m^3/s là cấp số đông. Số đông không bị các cực trị ảnh hưởng, nhưng cần phải phân cấp thống kê để xác định cụ thể trị số nào là số đông nên chỉ dùng trong lý luận, ít dùng trong tính toán.

5.5.2 Các trị số biểu thị xu thế phân tán

a/ Khoảng lệch lớn nhất

Khoảng lệch lớn nhất ∆m là sai biệt về trị số giữa trị số lớn nhất Xmax và trị số nhỏ nhất Xmin

$$\Delta m = Xmax - Xmin \qquad (5\text{-}17)$$

∆m dễ tìm nhưng nó không nói lên được mức độ phân tán của các biến số trong liệt số và nó dễ bị các cực trị ảnh hưởng nên ít được dùng.

b/ Khoảng lệch quân phương σ

Khoảng lệch quân phương σ là căn bậc 2 của bình quân các khoảng lệch bình phương.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n}} \qquad (5\text{-}18)$$

So với ∆m, σ có ưu điểm phản ánh mức độ phân tán toàn liệt, σ nhỏ thì liệt phân bố tập trung và ngược lại. Nhưng nó cũng có khuyết điểm: do khoảng lệch quân phương có thứ nguyên nên không thể dùng để so sánh mức độ phân tán giữa các liệt có thứ nguyên khác nhau. Hai liệt số có bình quân khác nhau xa cũng không thể dùng σ để so sanh được.

c/ Hệ số biến động Cv

Cv là tỉ số giữa σ và trị số bình quân của liệt số.

$$Cv = \frac{\sigma}{\bar{x}} = \frac{\sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n}}}{\bar{x}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i}{\bar{x}} - 1\right)^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(K_i - 1\right)^2} \qquad (5\text{-}19)$$

với $K_i = \frac{x_i}{\bar{x}}$ là hệ số module.

Cv là số không âm và không thứ nguyên, nó đã khắc phục được nhược điểm của σ, nên Cv là hệ số biểu thị mức độ phân tán tốt nhất.

Ví dụ 5.20:

Đại lượng ngẫu nhiên X [995, 1000, 1005] và Y = [5, 10, 15] đều có khoảng lệch quân phương là $\sigma_x = \sigma_y = 5$, nhưng trị số bình quân Mx = 1000 và My = 10 có độ chênh lệch nhau quá xa, vì vậy không thể dùng khoảng lệch quân phương để so sánh mức độ phân tán của chúng, lúc đó nếu ta tính Cv :

$$Cv_x = \frac{5}{1000}; \qquad Cv_y = \frac{5}{10}$$

ta thấy chuỗi số Y có độ phân tán lớn hơn chuỗi số X.

Tuy nhiên Cv cũng chưa phản ánh được hình dạng của đường phân bố mật độ tần suất nên người ta đưa vào hệ số thiên lệch Cs .

d/ Hệ số thiên lệch Cs :

Hệ số thiên lệch Cs của đại lượng ngẫu nhiên được tính như sau :

$$Cs = \frac{\int_{-\infty}^{+\infty}\left(x - M_x\right)^3 dx}{\sigma_x^{\,3}}$$ đối với đại lượng ngẫu nhiên liên tục (5-20)

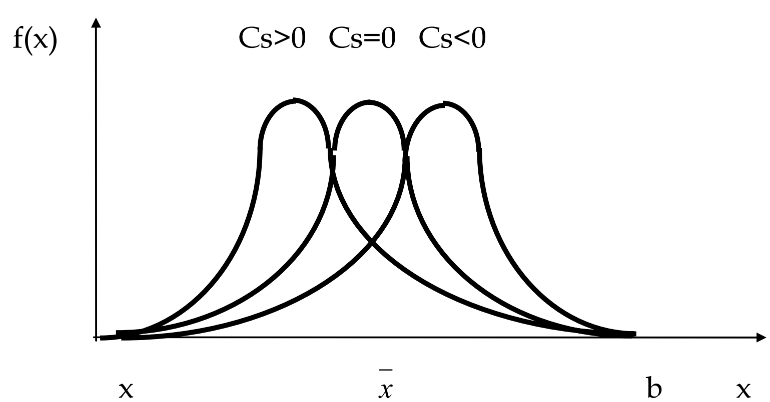
$$Cs = \frac{\sum_{i=1}^{n}(x_i - M_x)^3 . p}{\sigma_x^{\,3}}$$ đối với đại lượng ngẫu nhiên rời rạc (5-21)

$$Cs = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^3}{n\sigma_x^{\,3}}$$ khi $p_I = 1/n$ (5-22)

Công thức tính Cs phổ biến trong phân tích số liệu thủy văn hiện nay là:

$$Cs = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{n.Cv^3.\bar{x}^3} = \frac{\sum_{i=1}^{n}(K_i - 1)^3}{n.Cv^3} \quad \text{với} \quad K_i = \frac{x_i}{\bar{x}} \qquad (5\text{-}23)$$

Cs là một số không thứ nguyên. Trong công thức trên, vì Cv > 0 nên mẫu số dương, khi $\Sigma(Ki - 1)^3 > 0$ thì Cs > 0 làm đường phân bố sẽ lệch về phía trái của trị bình quân, ta gọi là *phân bố dương*, nó nói lên trị số nhỏ hơn trị bình quân là chiếm đa số. Ngược lại, khiCs < 0, ta có *phân bố âm*. Khi Cs = 0, dạng phân bố có dạng đối xứng qua trục quanh trị số bình quân



Hình 5.10 Hệ số thiên lệch biểu thị độ lệch của hình mật độ tần suất

Chú ý :

Thông thường trị số n không dài, dung lượng mẫu có giới hạn nên trong thực tế người ta dùng công thức sữa chữa sau :

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} = \bar{x}\sqrt{\frac{\sum_{i=1}^{n}\left(Ki-1\right)^2}{n-1}} \qquad (5\text{-}24)$$

$$Cv = \frac{1}{\bar{x}}\sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n}\left(Ki-1\right)^2}{n-1}} \qquad (5\text{-}25)$$

$$Cs = \frac{\sum_{i=1}^{n}\left(Ki-1\right)^3}{(n-3)Cv^3} \qquad (5\text{-}26)$$

5.6 TỔNG THỂ, MẪU VÀ SAI SỐ LẤY MẪU

5.6.1 Tổng thể

Tổng thể là tập hợp tất cả các gía trị mà đại lượng ngẫu nhiên X có thể nhận được. Số lượng tất cả các giá trị đó, ta gọi là *dung lượng của tổng thể*, ký hiệu là N.

5.6.2 Mẫu

Mẫu là một phần của tổng thể, mẫu được chọn ra theo đặt trưng riêng nào đó. Số lượng các giá trị của mẫu gọi là dung lượng mẫu, ký hiệu là n.

Ví dụ 5.21:

Tất cả các số liệu đo đạt liên quan đến dòng chảy của một con sông như lưu lượng, mực nước, chất lượng nước, bùn cát lơ lửng, ... được xem như tổng thể. Lấy riêng ra chuỗi số liệu về mực nước là một mẫu nghiên cứu.

5.6.3 Yêu cầu lấy mẫu thống kê thủy văn

Mục đích của phân tích thủy văn là tìm ra hàm phân phối xác suất F(x) của tổng thể, nhưng với một chuỗi nhỏ các mẫu x_1 , x_2 , ...,x_n thì ta có thể xác định hàm phân phối xác suất của mẫu F(x). Điều mong muốn của chúng ta là làm sao cho hàm phân phối xác suất của mẫu F(x) gần với hàm phân phối xác suất của tổng thể F(x). Do đó có một số yêu cầu nhất định đối với mẫu chọn ra :

- **Tính đồng nhất**: Mẫu phải đảm bảo tính đồng nhất các số liệu được lấy trong cùng một tổng thể.

- **Tính ngẫu nhiên độc lập**: Số liệu trong mẫu phải được lấy ngẫu nhiên và hoàn toàn độc lập với nhau.

- **Tính đại biểu**: Mẫu phải có dung lượng đủ lớn đặt trưng cho tất cả các trường hợp giá trị nhỏ, lớn và trung bình.

5.6.4 Công thức đánh giá sai số lấy mẫu

Sai số lấy mẫu thường tập trung chủ yếu vào sự sai biệt:

$$\Delta i = x_i - \overline{x} \tag{5-27}$$

Sai số do 3 nguyên nhân:

1. Tài liệu thu được là một mẫu của tổng thể nên khoảng lệch lớn nhất Δm của mẫu thường nhỏ hơn của tổng thể.

2. Sai số ngẫu nhiên trong đo đạc các trị số trong liệt tài liệu làm ảnh hưởng đến các giá trị Δi.

3. Dạng đường phân bố tần suất của tổng thể là liên tục còn của mẫu thực tế là không liên tục. Dựa vào mẫu để tính Δi và các tham số thống kê sẽ sai số so với tham số của tổng thể .

Trong 3 nguyên nhân này, nguyên nhân 1 là quan trọng hơn cả. Trong 3 tham số thống kê thường dùng thì sai số lấy mẫu của x nhỏ nhất sau đó đến Cv và Cs do có số mũ của khoảng lệch theo lũy thừa bậc 2 và 3. Do đó Cs sẽ xác định theo cách khác trừ trường hợp mẫu dài hơn 100 năm mới dùng công thức.

Hiện nay, để so sánh các tham số thống kê để tìm sai số lấy mẫu, thường dùng khoảng lệch quân phương của liệt các tham số thống kê để biểu thị gọi là ***sai số tiêu chuẩn.***

Giả sử ta chia tổng thể thành n mẫu có dung lượng bằng nhau, với mỗi mẫu ta tính x , Cv , Cs. Như vậy sẽ có n trị x, n trị Cv và trị Cs hợp thành 3 liệt tham số thống kê :

$\overline{x}_1, \overline{x}_2, \overline{x}_3, \ldots, \overline{x}_n$	: Liệt các trị bình quân
Cv_1 , Cv_2 , Cv_3 , ... , Cv_n	: Liệt các hệ số biến động
Cs_1 , Cs_2 ,Cs_3 , ... , Cs_n	: Liệt các hệ số thiên lệch

Với mối liệt tìm ra khoảng lệch quân phương, ta có σ là sai số tiêu chuẩn với mỗi liệt trên. Giá trị σ là sai số tuyệt đối còn σ % là sai số tương đối.

Sai số tiêu chuẩn của $\overline{x}$:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$
$$\sigma'_{\bar{x}} = \frac{100.Cv}{\sqrt{n}} 100\% \qquad (5\text{-}28)$$

Sai số tiêu chuẩn của Cv :

$$\sigma_{Cv} = \frac{Cv}{\sqrt{2n}}\sqrt{1+Cv^2}$$
$$\sigma'_{Cv} = \frac{100}{\sqrt{2n}}\sqrt{1+Cv^2}\,\% \qquad (5\text{-}29)$$

Sai số tiêu chuẩn của Cs :

$$\sigma_{Cs} = \sqrt{\frac{6}{n}(1+6Cv^2+5Cv^4)}$$
$$\sigma'_{Cs} = \frac{100}{Cs}\sqrt{\frac{6}{n}(1+6Cv^2+5Cv^4)}\,\% \qquad (5\text{-}30)$$

trong các công thức trên các giá trị n, σ, Cv, Cs lần lượt là dung lượng, khoảng lệch quân phương, hệ số biến động và hệ số thiên lệch của tập hợp mẫu đang tính toán.

Các sai số tiêu chuẩn tỷ lệ nghịch với căn bậc 2 của n, nên n càng nhỏ thì sai số càng lớn và ngược lại. Ứng dụng các công thức này giúp ta xác định tính đại biểu của mẫu tốt hay xấu. Ví dụ: với $\sigma_x \leq 5$ % thì mẫu đáng tin cậy có thể dùng tính toán cho công trình.

Ngược lại, nhiều khi qui phạm cho phép sai số không vượt quá bao nhiêu phần trăm nào đó, cần xem phải dùng bao nhiêu năm tài liệu (n) mới thoả mãn yêu cầu thống kê, khi được biết sai số tiêu chuẩn σ_{Cv} và Cv ta tính được n.

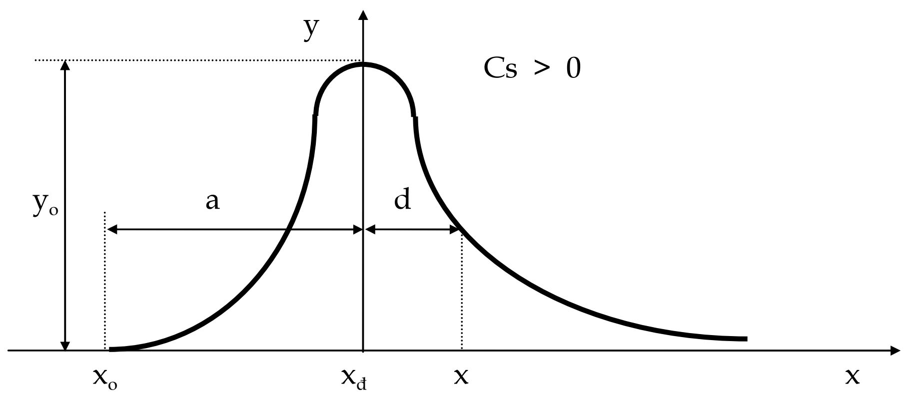
5.6.5 Đường tần suất "lý luận"

Như đã trình bày ở trước, do tài liệu quan trắc thường ngắn không đáp ứng được yêu cầu tính toán thủy văn với tần suất nhỏ. Để đáp ứng yêu cầu này người ta tập trung nghiên cứu đường phân bố mật độ của tổng thể dạng công thức toán học y = f(x). Tích phân đường này ta sẽ được đường tần suất tương ứng, gọi là ***đường tần suất "lý luận"***, để kéo dài bổ xung cho đường tần suất kinh nghiệm.

5.7 ĐƯỜNG PHÂN BỐ MẬT ĐỘ TẦN SUẤT PEARSON III (P.III)

5.7.1 Nguồn gốc của đường P.III

Năm 1795, Kan Pearson, một nhà thống kê sinh vật học người Anh căn cứ vào kết quả thống kê rất nhiều tài liệu, phát hiện thấy đường biểu diễn mật độ tần suất thường là quả hình chuông, chỉ có một số đông và tần suất 2 đầu giảm nhỏ tiến đến tiệm cận với hoành độ. Do đó, ông ra 2 điều kiện thành lập đường cong mật độ tần suất như sau:



Hình 5.11 Đường cong mật độ tần suất

Tại vị trí số đông, hệ số gốc của tiếp tuyến bằng 0. Nếu điểm gốc tọa độ đặt tại trị số bình quân thì khi x = - d, đạo hàm $\frac{dy}{dx} = 0$.

Ở đây, d là khoảng cách giữa số bình quân với số đông, gọi là bán kính lệch.

Hai đầu hoặc 1 đầu cong nhận trục hoành làm đường tiệm cận, nghiã là khi y = 0 thì $\frac{dy}{dx} = 0$

Với 2 điều kiện này ông đưa ra phương trình vi phân của họ đường cong phân bố mật độ tần suất như sau :

$$\frac{dy}{dx} = \frac{(x+d)y}{b_0 + b_1x + b_2x^2} \tag{5-31}$$

Giải phương trình bậc 2: $b_o + b_1x + b_2x^2 = 0$, ta được các loại nghiệm khác nhau, Pearson chia làm 13 loại đường cong. Đường Pearson III là 1 trong số các loại đường nói trên với $b_2 = 0$, tức là :

$$\frac{dy}{dx} = \frac{(x+d)y)}{b_0 + b_1x} \tag{5-32}$$

Di động trục tọa độ từ trị số bình quân về số đông và tích phân phương trình, ta được hàm số :

$$y = y_o\left(1+\frac{x}{a}\right)^{\frac{a}{d}}.e^{-(\frac{x}{d})} \qquad (5\text{-}33)$$

trong đó

a - khoảng cách từ khởi điểm của đường cong, tức là trị nhỏ nhất x_o, đến số đông

y_o - xác suất xuất hiện số đông (tung độ lớn nhất của đường cong)

e - cơ số logarit tự nhiên, e= 2.71828

Tính chất cơ bản của đường cong P.III là đầu trái có giới hạn (tồn tại trị số nhỏ nhất x_o) còn đầu phải vô hạn (không có trị số Xmax).

5.7.2 Ứng dụng của đường P.III

Qua dạng hàm số trên cho ta thấy đường P.III có 3 tham số: y_o, a, và d. Khi 3 tham số đã xác định thì ứng với x ta được y tương ứng. Đường cong được xác định hoàn toàn.

Qua phân tích, trong toán thống kê người ta xác định được quan hệ giữa các tham số đặc trưng thống kê như sau :

$$d = \left(\frac{Cv.Cs}{2}\right).\overline{X} \qquad (5\text{-}32)$$

$$a = \frac{2.C_V.\overline{X}}{C_s} - d \qquad (5\text{-}33)$$

$$y_o = \frac{2.Cs.\left(\frac{4}{Cs^2}-1\right)^{\left(\frac{4}{Cs^2}\right)}}{Cv.e^{\left(\frac{-4}{Cs^2}\right)}.\Gamma\left(\frac{4}{Cs^2}\right)} \qquad (5\text{-}34)$$

trong đó : $\Gamma\left(\frac{4}{C^2S}\right)$ là hàm Gamma, có có bảng tra sẵn (xem phụ lục số 2, sách Sổ tay kỹ thuật thủy lợi, tập II, bảng photocopy ở cuối chương). Do đó, nếu biết x, Cv, và Cs của mẫu thống kê thì đường mật độ tần suất P. III hoàn toàn xác định.

Để cho dễ tính, 2 ông Foxter và Rupkin đã dựa vào một số đặt tính của đường P.III tiến hành tích phân và lập bảng tra.

Do đường P. III có có đặt tính: Φ= f (Cs, P) = $\frac{Kp-1}{Cv}$, với Φ là khoảng lệch tung độ phụ thuộc vào Cv, dựa vào đặt tính này hai ông lấy Cv = 1 và các trị số Cs khác nhau, tiến hành tích phân tìm ra các Φ_p ứng với các tần suất khác nhau ở bảng Φ (Cs , P).

Khi áp dụng nếu Cv ≠ 1, thì dựa vào hệ thức trên suy ra :

$$Kp = \Phi\, Cv + 1 \tag{5-35}$$

Ví dụ 5.19:
Có 1 liệt thủy văn tính ra Cv = 0,5 và Cs = 1 thì tra ra $\Phi_{1\%} = 3{,}02$ và $\Phi_{75\%} = -0{,}75$. Vậy Kp ứng với 2 tần suất đó là:

$K_{1\%} = 3{,}02 \times 0{,}5 + 1 = 2{,}51$

$K_{75\%} = -0{,}75 \times 0{,}5 + 1 = 0{,}635$

Có Kp ta sẽ tìm được Xp:

$$Xp = Kp\,.\,\bar{x} = \bar{x}\,(\Phi\, Cv + 1) = \bar{x} + \Phi.\,\sigma \tag{5-36}$$

Đem các cặp trị số Kp và P (hoặc Xp và P) vẽ lên giấy tần suất ta được đường tần suất lý luận P. III của liệt cho.

5.7.3 Vài điểm cần chú ý khi ứng dụng đường P.III

a) Khi Cs< 0, vẫn dùng bảng Foxter - Rupkin nhưng sẽ biến đổi như sau :

$$\Phi_{p\,(Cs<0)} = -\,\Phi_{100-p\ \ Cs>0)} \tag{5-37}$$

Ví dụ 5.20:
Tìm Φ ứng với P = 1 % khi Cv = 0,5 và Cs = - 1.

Theo công thức trên, ta thấy :

$$\Phi_{1\%\,(Cs=-1)} = -\,\Phi_{100-p\,(Cs=1)} = -\,\Phi_{99\%\,(Cs=1)}$$

Tra bảng

$\Phi_{99\%} = -\,1{,}59$ (với P = 99 %, Cs = 1)

Do đó:

$\Phi_{1\%\,(Cs=-1)} = +\,1{,}59$

b) Khi dùng đường P. III, cần chú ý đến giới hạn thay đổi của Cs như sau :

$$2\,Cv \le Cs \le \frac{2Cv}{1-K\min} \tag{5-38}$$

Khi Cs < 2Cv , dường P.III xuất hiện trị số âm, không phù họp với hiện tượng thủy văn.

c) Căn cứ vào đặc tích phương trình đường P. III, khi Cs > 2, đường phân bố có dạng lưỡi liềm, không có số đông, không phù hợp với hiện tượng thủy văn cho nên chỉ khi *Cs < 2 thì mới được dùng đường P. III.*

5.8 ĐƯỜNG PHÂN BỐ MẬT ĐỘ TẦN SUẤT KRITXKI - MENKEN

Hai ông Kritxki - Menken, sau khi phân tích rất nhiều tài liệu trên các sông ngòi Liên Xô (cũ), đi tới kết luận:

Nhìn chung đường P. III tương đối phù hợp với tài liệu được đo. Về dòng chảy năm mà nói, đa số dòng sông đều có Cs giao động quanh 2.Cv, có một số sông có Cs < 2.Cv, lúc này P. III không hợp. Nhằm khắc phục nhược điểm của P.III trên, hai ông đề ra điều kiện để xây dựng đường phân bố mật độ của mình như sau:

1/ Có thể dùng 3 tham số $\overline{x}$,Cv, Cs để tính.

2/ Chỉ có một số đông.

3/ Trị của biến ngẫu nhiên cò thể thay đổi trong phạm vi : $-\infty \leq x \leq +\infty$.

Theo Kritxki - Menken, các điều kiện trên đây không những phù hợp với đặt tính dòng chảy mà còn sử dụng dễ dàng. Với mọi quan hệ Cs / Cv. Khởi điểm của đường cong đều qua điểm 0, không xuất hiện trị số âm.

Vì khi Cs = 2 Cv thì đường P. III hoàn toàn phù hợp với 3 diều kiện trên, cho nên hai ông lấy dạng hàm số của đường P. III với Cs = 2Cv làm cơ sở, tiến hành biến đổi hàm số tìm ra dạng đường của mính, gọi là đường Kritxki - Menken (tắt là đường K - M) sau đây :

$$y = f(x) = \frac{\alpha^{\alpha}}{\alpha^{\frac{\alpha}{b}}.b.\Gamma(\alpha)}.x^{\frac{\alpha}{b}-1}.e^{-\alpha.\left(\frac{x}{a}\right)^{\frac{1}{b}}} \quad \text{với } 0 \leq x \leq \infty \qquad (5\text{-}39)$$

Trong đó $\propto = \frac{1}{Cv^2}$, a và b đều là hằng số. Tích phân hàm mật độ tần suất trên ta sẽ được đường tần suất lý luận K -M.

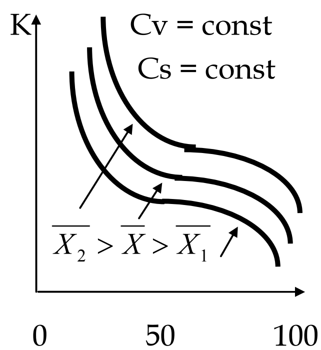
Tương tự như Foxter - Rupkin, hai ông lập bảng tính Kp với 7 trường hợp:
Cs = (1 ÷ 6) Cv. Dùng Kp tính ra Xp = Kp.$\overline{X}$. Vẽ lên giấy tần suất ta được đường tần suất lý luận K- M.

5.9 ẢNH HƯỞNG CỦA THAM SỐ THỐNG KÊ VỚI ĐƯỜNG TẦN SUẤT

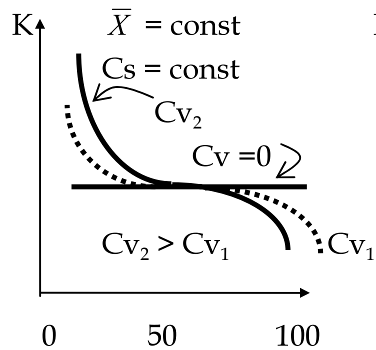
Để dễ dàng sử lý trong khi vẽ đường tần suất lý luận, ta cần tìm ảnh hưởng của chúng như sau :

- Nếu Cv, Cs không đổi khi ta thay đổi trị số bình quân thì các đường tần suất song song với nhau (Hình 5.12).
- Nếu cố định $\overline{X}$ và Cs thì Cv càng lớn thì phân bố càng phân tán, đường tần suất càng dốc. Khi Cv = 0, đường tần suất thành đường nằm ngang (Hình 5.13)
- Nếu Cv, $\overline{X}$ cố định, khi Cs càng lớn thì nửa trên đường tần suất càng dốc, nửa dưới càng bằng và ngược lại khi Cs giảm, nửa trên càng bằng và nửa dưới càng dốc (Hình 5.14).

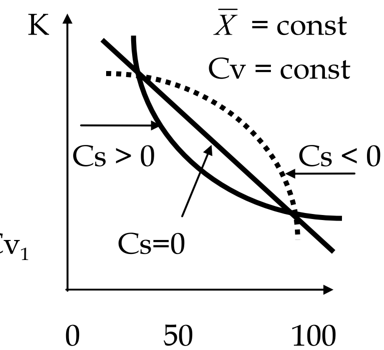
Cs > 0 : Đường tần suất lõm xuống phía dưới
Cs < 0 : Đường tần suất lồi lên
Cs = 0 : Đường tần suất là đường xiên góc.



Hình 5.12



Hình 5.13



Hình 5.14

5.10 PHƯƠNG PHÁP VẼ ĐƯỜNG TẦN SUẤT THƯỜNG DÙNG TRONG TÍNH TOÁN THỦY VĂN

5.10.1 Phương pháp đường thích hợp

Thực tế cho thấy đường tần suất lý luận P.III và Kritxki -Menken thường nằm cách xa điểm kinh nghiệm nên không thể dùng làm công cụ kéo dài đường tần suất kinh nghiệm, nguyên nhân chính là do sai số tính các tham số. Do đó không nhất thiết phải quá lệ thuộc vào công thức tính các tham số theo lý luận, mà có thể xác định các tham số thống kê sau cho đường tần suất lý luận phù hợp với các điểm kinh nghiệm là được. Cách này gọi là "PHƯƠNG PHÁP ĐƯỜNG THÍCH HỢP ".

Các bước tính như sau:

1/ Sắp xếp liệt số theo thứ tự từ lớn đến nhỏ, tính tần suất rời vẽ các điểm kinh nghiệm lên giấy tần suất .

2/ Tính $\overline{X}$ và Cv, còn Cs tính theo Cs = m Cv . Trong đó m là số cần chọn sau cho thích hợp, m có thể nhỏ hơn 2

3/ Giả định 1 trị số m, cùng Cv tra ra Kp. Tính $Xp = Kp.\overline{X}$. Vẽ đường tần suất lý luận Xp - P.

4/ So sánh các điểm kinh nghiệm nếu chưa phù hợp thì dựa vào ảnh hưởng của tham số thống kê đối với đường tần suất, điều chỉnh m, tức Cs, để cho phù hợp với điểm kinh nghiệm một cách tốt nhất.

Ưu điểm của phương pháp này là 2 lấy thực tiễn làm tiêu chuẩn kiểm nghiệm lý luận nên được dùng nhiều. Nhược điểm của phương pháp là có thể bị sai biệt tính toán do chủ quan. Có thể hai người tính cho 2 kết quả khác nhau nhất là phần kéo dài.

Ví dụ 5.21:

Tính tần suất theo phương pháp đường thích hợp cho tài liệu lưu lượng bình quân năm từ 1950 - 1969 tại một trạm thủy văn ghi ở cột [3] trên bảng 5.4. Tìm lưu lượng bình quân năm ứng với tần suất 10 % , 50% và 90%

Giải:

Các bước tính toán :

1/ Sắp xếp tài liệu lưu lượng theo thứ tự từ lớn đến nhỏ như cột [4] trong bảng 5.2.

2/ Tính trị số bình quân của liệt tài liệu:

$$\overline{Q} = \frac{\sum_{i=1}^{n} Q}{n} = \frac{7935}{20} = 396.75.m^3 / s$$

3/ Tính hệ số module:

$$K_i = \frac{Q_I}{\overline{Q}}, (Ki-1), (Ki-1)^2, (Ki-1)^3 \qquad \text{như ở cột [5] đến [9]}$$

4/ Tính hệ số biến động Cv:

$$Cv = \sqrt{\frac{\sum_{i=t}^{n}(K_i-1)^2}{n-1}} = \sqrt{\frac{1.366}{20-1}} = 0.268$$

Tính sai số của hệ biến động :

$$\sigma_{Cv} \frac{Cv}{\sqrt{2n}}\sqrt{1+Cv^2} = \frac{0.268}{\sqrt{2\times 20}}\sqrt{1+(0.268)^2} = 0.043 \qquad \text{[sai số tuyệt đối]}$$

$$\sigma'_{Cv} \frac{100}{\sqrt{2n}}\sqrt{1+Cv^2} = \frac{100}{\sqrt{2\times 20}}\sqrt{1+(0.268)^2} = 16.36\% \qquad \text{[sai số tương đối]}$$

Do đó: Cv = 0.268 ± 0.043

Bảng 5.4: Tính tần suất lưu lượng bình quâm tại một trạm thủy văn

TT	Năm	Qi (m^3/s)	Qi sắp thứ tự	Ki	(Ki - 1)	$(Ki-1)^2$	$(Ki-1)^3$	$P = \frac{m}{n+1}100\%$
[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
1	1950	570	592	1.492	0.492	0.242	0.119	4.8
2	1951	503	570	1.436	0.436	0.190	0.083	9.5
3	1952	313	503	1.267	0.267	0.071	0.019	14.3
4	1953	485	496	1.250	0.250	0.062	0.015	19.0
5	1954	460	485	1.222	0.222	0.049	0.010	23.8
6	1955	592	463	1.166	0.166	0.027	0.004	28.6
7	1956	215	460	1.159	0.159	0.025	0.004	33.3
8	1957	346	446	1.124	0.124	0.015	0.001	38.1
9	1958	333	445	1.121	0.121	0.014	0.001	42.9
10	1959	411	411	1.035	0.035	0.001	0.000	47.6
11	1960	263	399	0.005	0.005	0.000	0.000	52.4
12	1961	446	346	0.872	-0.127	0.016	-0.002	57.1
13	1962	445	342	0.862	-0.137	0.019	-0.002	61.9
14	1963	342	333	0.839	-0.160	0.025	-0.004	66.7
15	1964	274	313	0.788	-0.211	0.044	-0.009	71.4
16	1965	496	306	0.771	-0.228	0.0.52	-0.011	76.2
17	1966	399	274	0.690	-0.309	0.095	-0.029	80.0
18	1967	463	273	0.688	-0.311	0.097	-0.030	85.7
19	1968	273	263	0.662	-0.337	0.113	-0.038	90.5
20	1969	306	215	0.541	-0.458	0.209	-0.100	95.2
	Tổng	7935				1.366	0.033	

Khoảng lệch quân phương của trị bình quân:

$$\sigma = Cv \times \overline{Q} = 0.268 \times 396.75 = 106.329 m^2 / s$$

4/ Tính sai số của trị bình quân:

$$\sigma_\ell = \frac{\sigma}{\sqrt{n}} = \frac{106.329}{\sqrt{20}} = 23.775 m^3 / s \qquad \text{[Sai số tuyệt đối]}$$

$$\sigma'_\ell = \frac{Cv}{\sqrt{n}} 100\% = \frac{0.268}{\sqrt{20}} 100\% = 5.992\% \qquad \text{[Sai số tương đối]}$$

Do đó : $\overline{Q} = 396.75 \pm 23.775$ m^2/s

5/ Tính hệ số thiên lệch Cs:

$$Cs = \frac{\sum_{i=1}^{n}(Ki-1)^3}{(n-3)Cv^3} = \frac{0.033}{(20-3)(0.268)^3} = 0.100$$

Sai số tương đối của hệ số thiên lệch :

$$\sigma_{Cs} = \frac{100}{Cs}\sqrt{\frac{6}{n}\left(1+6Cv^2+5Cv^4\right)}\% = \frac{100}{0.100}\sqrt{\frac{6}{20}\left(1+6\times 0.268^2+5\times 0.268^4\right)} = 660.68\%$$

Do sai số của Cs quá lớn nên tính Cs = m Cv,ở đây chọn m = 2 ta được :

Cs = 2Cv = 2 x 0.27 = 0.54.

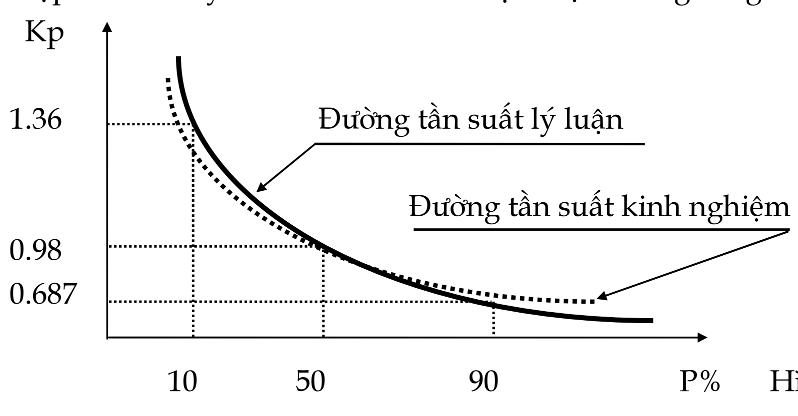
Căn cứ vào Cv và Cs = 2 Cv tra bảng ở phụ lục và nội suy giữa 2 dòng Cv= 0.25 và Cv = 0.30 ta được kết quả ở bảng 5.4.

Bảng 5.5 : Tọa độ đường tần suất lý luận

P %	0.1	1.0	3.0	5.0	10.0	20.0	50.0	75.0	90.0	95.0	99.0
Kp	2.05	1.74	1.52	1.49	1.36	1.21	0.98	0.81	0.867	0.60	0.48

6/ Vẽ các điểm tần suất kinh nghiệm Ki - P

Theo số liệu của cột [5] và [9] lên giấy tần suất, vẽ đường tần suất kinh nghiệm bằng một đường cong trơn. Đem kết quả bảng 2 vẽ lên giấy tần suất ta được tần suất lý luận (hình 5.15). So sánh 2 đường cong xem có phù hợp tốt, nếu như không phù hợp có thể thay đổi m cho đến khi đạt được đường cong thỏa mãn.



Hình 5.15

7/ Từ đường tần suất lý luận ta tra được lưu lượng bình quân năm ứng với tần suất thiết kế 10 % , 50 % ,và 90 % như ở bảng 5.6 dưới đây:

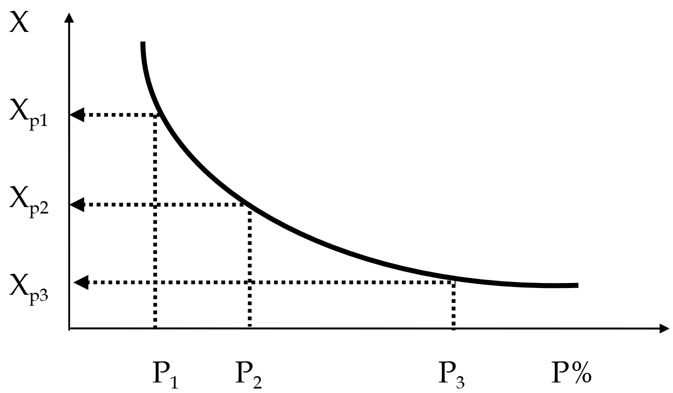
Bảng 5.6

P %	10 %	50 %	90 %
Kp	1.36	0.98	0.687
Qp = Kp . Q	540	288	265

5.10.2 Phương pháp 3 điểm

Để tránh mất thời giờ tính $\overline{X}$, Cv và việc thử dần m, Alecxâyev đưa ra phương pháp đồ giải 3 điểm để tính các thông số thống kê nhanh hơn và phù hợp hơn. Trên đường tần suất kinh nghiệm lấy 3 điểm:

(Xp_1, P_1), (Xp_2, P_2) và (Xp_3, P_3).



Hình 5.16: Xác định 3 điểm trên đường kinh nghiệm

Theo lý luận 3 điểm này sẽ nằm trên đường tần suất cần vẽ, ta có các hệ thống phương trình:

$$Xp_1 = \overline{X} + \sigma.\Phi(P_1, Cs) \quad [1] \tag{5-40}$$

$$Xp_2 = \overline{X} + \sigma.\Phi(P_2, Cs) \quad [2] \tag{5-41}$$

$$Xp_3 = \overline{X} + \sigma.\Phi(P_3, Cs) \quad [3] \tag{5-42}$$

Khử X và σ theo $\frac{[1]+[3]-2[2]}{[1]-[3]}$, ta có :

$$\frac{Xp_1 + Xp_3 - 2Xp_2}{Xp_1 - Xp_3} = \frac{\Phi(p_1, Cs) + \Phi(p_3, Cs) - 2\Phi(p_2, Cs)}{\Phi(p_1, Cs) - \Phi(p_3, Cs)} = S \tag{5-43}$$

S là tham số mới, nó là hàm của P và Cs.

Khi P đã chọn trước thì S = Φ (Cs), với Cs là hệ số thiên lệch.

Ba trị số P của 3 điểm lấy theo một trong 4 cách sau:

- P % = 1 - 50 - 99 %
- P % = 3 - 50 - 97 %
- P % = 5 - 50 - 95 %
- P % = 10 - 50 - 90 %

Dựa vào bảng tính sẵn Foxto - Rupkin, ta tính ra hệ S = Φ (Cs), có phụ luc tra. Như vậy từ Xp_1 , Xp_2 , Xp_3 đã biết thay vào phương trình S rồi tra phụ lục ta sẽ được Cs cần tìm.

Từ phương trình [1] và [3] ta rút ra :

$$\sigma = \frac{Xp_1 - Xp_3}{\Phi(P_1, Cs) - \Phi(P_3, Cs)} = \frac{Xp_1 - Xp_3}{\Phi(p_1) - \Phi(p_3)} \qquad (5\text{-}44)$$

Và từ [2] rút ra :

$$\overline{X} = Xp_2 - \sigma.\Phi(P_2, Cs) = Xp_2 - \sigma.\Phi(P_2) \qquad (5\text{-}45)$$

Thường P_2 = 50%, còn P_1 và P_3 là 5% và 90 % hoặc 10% và 90%, đã có bảng tra sẵn : Từ Cs tra ra Φ_{p1}, Φ_{p3} và Φ_{p2} ta sẽ tính được σ và X. Còn Cv = σ/ $\overline{X}$.

Có X, Cv, Cs, ta sẽ được tần suất lý luận Xp - P như trước.

Phương pháp này có các đặt điểm sau:

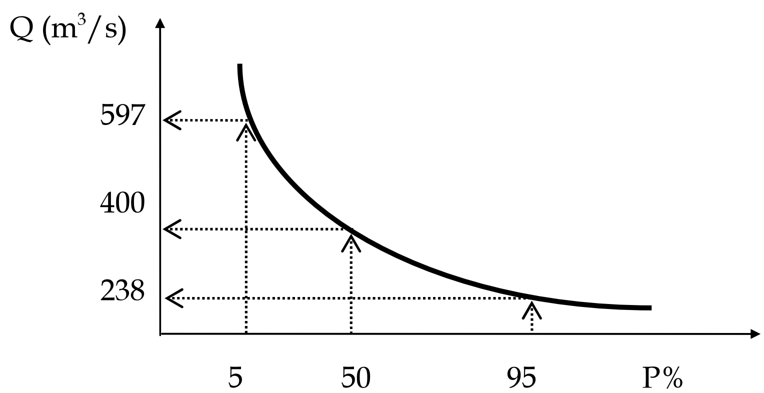
- Tính toán các tham số nhanh.
- Cv < 0.5 thì phương pháp cho kết quả khá phù hợp với điểm kinh nghiệm.
- Do phải vẽ đường kinh nghiệm để chọn điểm nên kết quả dễ bị chủ quan của người vẽ.

Ví dụ 5.22:

Vẽ đường tần suất theo phương pháp 3 điểm. Dùng lưu lượng bình quân năm ghi trong bảng 5.4 làm thí dụ.

Giải:

1/ Vẽ các điểm tần suất kinh nghiệm như tài liệu ở cột [4] và [9] trong bảng 5.4 rồi dùng thước cong vẽ đường tần suất kinh nghiệm. (như thí dụ trước).



Hình 5.17: Lấy 3 điểm theo ví dụ 5.22

Lấy 3 điểm trên đường tần suất kinh nghiệm:

$Q_{5\,\%}$ = 597 m^3/s $Q_{5\,\%}$ = 400 m^3/s $Q_{95\,\%}$ = 238 m^3/s

2/ Tính tham số thống kê

Hệ số lệch S

$$S = \frac{Q_{5\%} + Q_{95\%} - 2Q_{50\%}}{Q_{5\%} - Q_{95\%}} = \frac{597 + 238 - 2 \times 400}{597 - 238} = \frac{55}{359} = 0.15$$

Tra phụ lục 2 với S = 0.15 được Cs = 0.55

Tính σ và tra phụ lục 2 giá trị: $\phi_{5\%} - \phi_{95\%}$: (Trang 393)

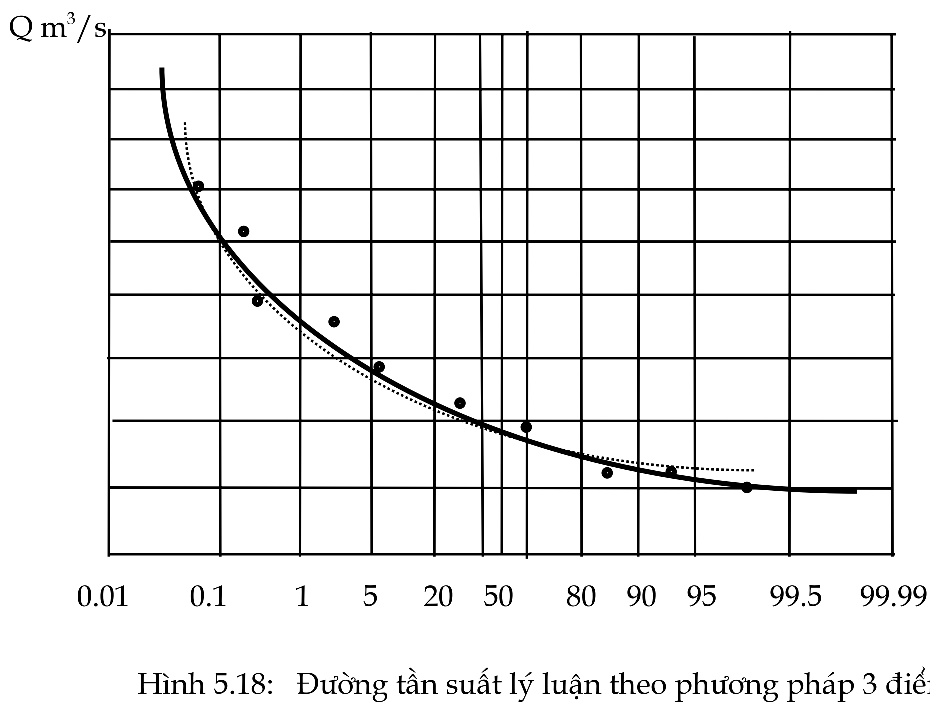
$$\sigma = \frac{Q_{5\%} - Q_{95\%}}{\phi_{5\%} - \phi_{95\%}} = \frac{597 - 238}{3.2625} = 109 m^3/s$$

Tính lưu lượng trung bình Q và hệ số biến động Cv :

$$\overline{Q} = Q_{50\%} - \sigma.\phi_{50\%} = 400 - 109\ (-0.092) = 393\ m^3/s$$

$$Cv = \frac{\sigma}{\overline{Q}} = \frac{109}{393} = 0.27$$

3/ Vẽ đường tần suất lý luận theo Q 393 m^3/ s , Cv = 0.27 và Cs = 0.55 như thường lệ theo phụ lục. Ta có đường tần suất lý luận khá phù hợp như hình vẽ, vậy không cần hiệu chỉnh thống kê.



Hình 5.18: Đường tần suất lý luận theo phương pháp 3 điểm
n = 20, Q 393 m^3/ s, Cv=0.27, Cs = 0.55

Ghi chú :

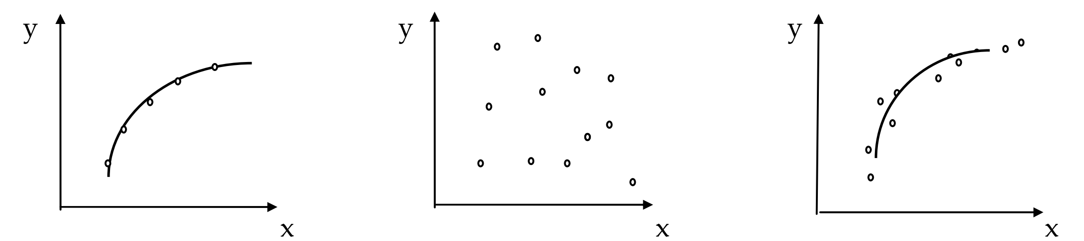
Đường tần suất lý luận được coi là phù hợp với đường tần suất kinh nghiệm khi :

$$\left[\overline{X}_{KN} - \overline{X}_{LL}\right] \leq \left[0.02.\overline{X}_{KN}\right] \qquad (5\text{-}46)$$

5.11 PHÂN TÍCH TƯƠNG QUAN

5.11.1 Khái niệm chung

Trong lý thuyết thống kê, khi phân tích quan hệ về số lượng của biến số này với biến số khác gọi là *phân tích tương quan* (correlation analysis). Quan hệ này có thể ở 3 dạng:



Quan hệ chặt chẽ (Hình 5-19-a) | Quan hệ rời rạc (Hình 5-19-b) | Quan hệ tương quan (Hình 5-19-c)

Hình 5.19: Các hình thức quan hệ

- ***Quan hệ chặt chẽ*** : quan hệ dạng hàm số x ~ y hay *tương quan hoàn toàn*. (Hình 5-19-a)
- ***Quan hệ rời rạc*** : còn gọi là *không tương quan*. (Hình 5-19-b)
- ***Quan hệ tương quan*** : đó là trường hợp ứng với mỗi trị số biến x có thể lấy trị số này hoặc trị số khác của y mà không khẳng định trước được. Nhưng chúng vẫn có thể có 1 xu thế rõ rệt nào đó gọi là *tương quan thống kê*. Nó là trường hợp trung gian giữa hai trường hợp cực đoan trên và là đối tượng nghiên cứu lý thuyết của thống kê. (Hình 5-19-c)

Tùy theo biến số xét đến nhiều hay ít mà phân hai loại: tương quan đơn và tương quan kép. Trong tính thủy văn chủ yếu dùng loại tương quan đơn, gồm có tương quan đường cong và phổ biến là tương quan đường thẳng .

5.11.2 Tương quan đường thẳng

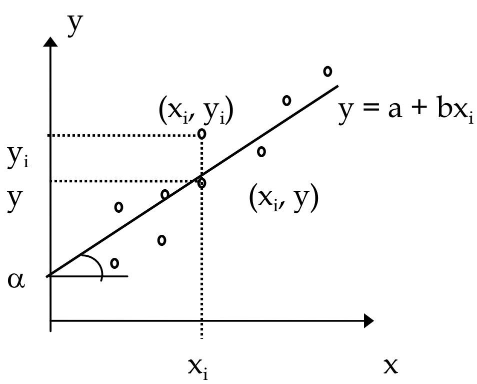
Có 2 phương pháp xác định tương quan được dùng hiện nay là : Phương pháp giải tích và phương pháp đồ giải.

5.11.2.1 Phương pháp giải tích

a/Phương pháp hồi qui:

Có dạng là $y = a + bx_1$ (5-47)

với x_1 là biến số độc lập và y là biến số phụ thuộc, a và b là 2 hệ số.



Hình 5.20: Tương quan đường thẳng

Hệ số góc a = tgα

Qua hình vẽ 5.20 cho thấy khỏang lệch giữa điểm thực đo đối với trị số bình quân có điều kiện nằm trên đường hồi qui là:

$$y_i - y = y_i - (a + bx_i) \quad (5\text{-}48)$$

Tiêu chuẩn để đường hồi qui tốt nhất là tổng bình phương các khoảng lệch phải tốt nhất :

$$\sum (y_i - y)^2 = \sum (y_i - a - bx_i)^2 = \min \quad (5\text{-}49)$$

Trong đó a và b là 2 biến số. Muốn cho biểu thức trên nhỏ nhất ta tìm đạo hàm riêng với a và b rồi cho nó bằng 0 :

$$\frac{\partial \sum (y_i - y)^2}{\partial a} = 0 \quad ; \quad \frac{\partial \sum (y_i - y)^2}{\partial a} = 0 \quad (5\text{-}50)$$

Giải hệ phương trình vi phân trên ta được :

$$a = \bar{y} - \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}\bar{x} \quad ; \quad b = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}\bar{x} \quad (5\text{-}51)$$

Thế a và b vào phương trình $y = (a + bx_i)$ ta được phương trình hồi qui y theo x

$$y - \bar{y} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}(x_i - \bar{x}) \quad (5\text{-}52)$$

trong đó : x_i, y_i - các trị số thực đo tương ứng với nhau

x, y - số bình quân của x_i và y_i

y - số bình quân có điều kiện ứng với x_i nào đó

Tương tự, nếu y là biến số độc lập, x là biến số phụ thuộc. Ta có phương trình hồi qui sau :

$$x = a_1 + b_1 y_I \tag{5-53}$$

và

$$x - \bar{x} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (y_i - \bar{y})^2}(y_i - \bar{y}) \tag{5-54}$$

trong đó : x là trị số bình quân có điện cần tìm ứng với y_I nào đó .

5.11.2.2 Hệ số tương quan

Phương trình hồi qui có thể biểu thị quan hệ tương quan giữa 2 biến số nhưng không đánh giá mức độ chặt chẽ của tương quan. Để đánh giá mức độ tương quan ta dùng 1 đặt trưng khác gọi là *hệ số tương quan* .

Xét 2 phương trình: $y = a + bx_i$ (5-55)

và $x = a_i + b_1 y_i$ (5-56)

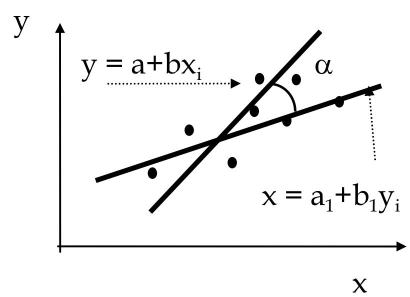
Hai phương trình này cắt nhau tại điểm $(\bar{x}, \bar{y})$ và góc kẹp là α (hình 5.21)

Khi α = 0 thì 2 đường trùng nhau và hệ số góc của chúng bằng nhau:

Tức là

$$b = \frac{1}{b_1} \Rightarrow bb_1 = 1$$

$$\Rightarrow \sqrt{b.b_1} = \mp 1$$



Hình 5.21 :
Góc kẹp α giữa 2 đường thẳng

Vậy nếu

- $\sqrt{b.b_1} = \mp 1 \Rightarrow$ x và y có quan hệ đường thẳng
- $\sqrt{b.b_1} \neq \pm 1 \Rightarrow$ x và y có tương quan thống kê
- $\gamma = \left|\sqrt{b.b_1}\right|$ càng nhỏ thì tương quan càng kém

Ta có sự quan hệ giữa sai số hồi qui với độ lệch quân phương và hệ số tương quan như sau :

$$S_y = \sigma_y \sqrt{1 - \gamma^2} \tag{5-57}$$

$$S_x = \sigma_x \sqrt{1 - \gamma^2} \tag{5-58}$$

- Khi $\gamma = \pm 1 \Rightarrow S_y = S_x = 0$: Các điểm đều nằm trên một đường thẳng và quan hệ là quan hệ hàm số.

- Khi $\gamma = 0 \Rightarrow S_y = \sigma_x$; $S_y = \sigma_y$ tức là $\overline{x} = x$ và $\overline{y} = y \Rightarrow$ Hai đường hồi qui song song với trục hoành và không có sự tương quan.

- Khi $0 < \gamma < 1$ là tương quan thống kê với S_y và S_x.

Sai số của hệ số tương quan:

$$\sigma_y = \frac{1-\gamma^2}{\sqrt{n}} \qquad (5\text{-}59)$$

Qua phân tích với nhiều tài liệu thủy văn, hầu như hiện nay đi đến điều kiện ứng dụng phương pháp tương đương đường thẳng là :

☒ Hệ số tương quan $\gamma \geq 0.8$

☒ Có 10 năm quan trắc đồng bộ $n \geq 10$ trở lên

Đặc điểm phương pháp

- Thông qua phương trình hồi qui để tính toán nên ít mắc sai số chủ quan.
- Có mức độ tương quan nên tránh được việc tùy tiện
- Hai đường hồi qui theo tính toán thí kết quả phải giống nhau
- Không loại trừ được điểm quá tản mạn do những nguyên nhân không xác đáng gây ra.

Để khắc phục nhược điểm này, trong thực tế thường phối hợp với tương quan đồ giải để bổ xung kéo dài tài liệu.

5.11.2.3 Phương pháp tương quan đồ giải

Trước khi áp dụng phải phân tích hệ số tương quan của biến số xem được hay không, nghĩa là tìm giá trị y. Sau đó chấm các điểm tương quan (x_i , y_i) lên trục tọa độ, nếu các điểm tương đối tập trung ($\gamma \geq 8$), ta vẽ đường thẳng đi qua trung tâm các điểm để kéo dài tài liệu. (Xem ví dụ ở cuối chương).

Do đó có thể dùng $b.b_1$ làm biểu thị cho mức độ tương quan thẳng giữa x và y gọi là *hệ số tương quan*. Vậy *hệ số tương quan là trung bình nhân của các hệ số góc của đường hồi qui* và ký hiệu:

$$\gamma = \pm\sqrt{b.b_1} \qquad (5\text{-}60)$$

Thế giá trị $b.b_1$ vào trên, ta có :

$$\gamma = \pm\sqrt{\frac{\left[\sum(x_i-\bar{x})(y_i-\bar{y})\right]\left[\sum(x_i-\bar{x})(y_i-\bar{y})\right]}{\sum(x_i-\bar{x})^2\sum(y_i-\bar{y})^2}} = \pm\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2\sum(y_i-\bar{y})^2}}$$

$$\gamma = \pm\frac{\sum(K_x-1).(K_y-1)}{\sqrt{\sum(K_x-1)^2.\sum(K_y-1)^2}}$$ trong đó : $K_x = \frac{x_i}{\bar{x}}, K_y = \frac{y_i}{\bar{y}}$ (5-61)

- γ là dấu cộng (+) khi quan hệ đồng biến, thí dụ: mưa và dòng chảy, mưa nhiều thì dòng chảy lớn.
- γ mang dấu trừ (-) khi quan hệ nghịch biến, thí dụ: bốc hơi nhiều thì dòng chảy nhỏ.

Hệ số hồi qui y theo x : $b = \frac{\sigma_y}{\sigma_x}\gamma$ (5-62)

Hệ số hồi qui x theo y : $b_1 = \frac{\sigma_x}{\sigma y}\gamma$ (5-63)

trong đó: σ_x và σ_y là khoảng lệch quân phương của 2 liệt số.

Phương trình hồi qui y theo x : $y - \bar{y} = \gamma\frac{\sigma_y}{\sigma_x}(x_i - \bar{x})$ (5-64)

Phương trình hồi qui x theo y : $x - \bar{x} = \gamma\frac{\sigma_x}{\sigma_y}(y_i - \bar{y})$ (5-65)

5.11.2.4 Sai số tương quan

Sai số của đường hồi qui :

$S_y = \sqrt{\frac{\sum(y_i-\bar{y})^2}{n-1}}$ khi tính y theo x (5-66)

$S_x = \sqrt{\frac{\sum(x_i-\bar{x})^2}{n-1}}$ khi tính x theo y (5-67)

5.11.3. Tương quan đường cong

Có một số trường hợp các điểm tương quan phân bố theo dạng parabol hay hyperbol đơn giản. Với giấy log, ta chuyển chúng thành đường thẳng.

5.11.3.1 Dạng parabol

Đường cong có dạng **$y = ax^m$** (5-68)

Lấy logarit hai vế : log y = log a + m log x (5-69)

Đây là dạng đường thẳng vẽ trên giấy pull - log với hệ số góc là m và cắt y ở điểm cách gốc là a vì khi x = 1 thì y = a.

5.11.3.2 Dạng hyperbol

Đường cong có dạng $y = \frac{a}{x^m}$ (5-70)

Lấy logarit hai vế: log y = log a - m log x (5-71)

Tương tự như trên , ta cũng vẽ được quan hệ này ở dạng đường thẳng trên giấy pull - log, hệ số góc lúc này là - m và cũng cách gốc 1 khoảng là a.

Những điểm cần chú ý khi áp dụng phương pháp thống kê và phân tích trongtính toán thủy văn:

1/ Liệt số thủy văn phải đồng nhất
2/ Các trị trong liệt số thủy văn phải độc lập với nhau.
3/ Liệt số thủy văn phải có tính đại biểu.

Ví dụ 5.23: Trạm thủy văn A có tài liệu thực đo 12 năm từ 1958 - 1969, tính ra module dòng chảy như cột [3] trong bảng 5.7. Trạm thủy văn B ở gần trạm A có tài liệu dài hơn, đo từ 1954 - 1969 như cột [4] trong bảng.

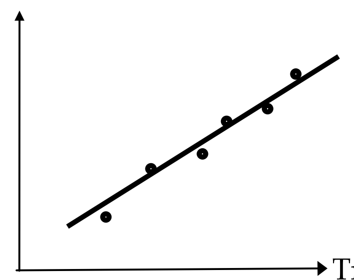
Bảng 5.7 : Module dòng chảy của trạm A và B

STT	NĂM	Trạm A ($l/s.km^2$)	Trạm B ($l/s.km^2$)	STT	NĂM	Trạm A ($l/s.km^2$)	Trạm B ($l/s.km^2$)
[1]	[2]	[3]	[4]	[1]	[2]	[3]	[4]
1	1954	-	6.9	9	1962	6.3	8.7
2	1955	-	7.2	10	1963	6.0	7.8
3	1956	-	5.9	11	1964	6.3	8.5
4	1957	-	6.4	12	1965	3.3	5.6
5	1958	4.8	6.2	13	1966	6.2	8.9
6	1959	4.1	5.3	14	1967	4.8	6.5
7	1960	5.3	6.8	15	1968	7.1	9.5
8	1961	5.0	8.0	16	1969	5.5	7.0

Yêu cầu dùng phương pháp tương quan để kéo dài tài liệu cho trạm A từ trạm B.

Giải :

Hai trạm A và B gần nhau, có điều kiện hình thành dòng chảy tương tự nhau, có trên 10 năm quan trắc đồng bộ (ứng với cùng thời gian) nên có thể xét dùng phương pháp tương quan.



Hình 5.22
Tương quan 2 trạm A và B

1/. Phương pháp tương quan giải tích

- Lấy tài liệu quan trắc đồng bộ trong 12 năm vẽ lên giấy kẻ ô vuông, thấy dải điểm phân bố theo xu thế đường thẳng nên có thể dùng phương pháp tương quan đường thẳng.

- Lập bảng số hạng trong phương pháp hồi qui như bảng 5.8 ở dưới:

Bảng 5.8: *Tính toán các giá trị tính tương quan*

TT	Năm	A [x_i]	B[y_i]	Kx	Ky	[Kx - 1]^2	[Ky - 1]^2	[Kx-1][Ky-1]
[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]
1	1958	4.8	6.2	0.89	0.84	0.0121	0.0256	0.0176
2	1959	4.1	5.3	0.76	0.72	0.0576	0.0784	0.0672
3	1960	5.3	6.8	0.98	0.92	0.0004	0.0064	0.0016
4	1961	5.0	8.0	0.93	1.08	0.0049	0.0064	- 0.0056
5	1962	6.3	8.7	1.17	1.17	0.0289	0.0289	0.0289
6	1963	6.0	7.8	1.11	1.05	0.0121	0.0025	0.0055
7	1964	6.3	8.5	1.17	1.15	0.0289	0.0225	0.0255
8	1965	3.3	5.6	0.61	0.76	0.1521	0.0576	0.0961
9	1966	6.2	8.9	1.15	1.20	0.0225	0.0400	0.0300
10	1967	4.8	6.5	0.89	0.88	0.0121	0.0144	0.0132
11	1968	7.1	9.5	1.32	1.28	0.1024	0.0784	0.0896
12	1969	5.5	7.0	1.02	0.95	0.0004	0.0025	- 0.0010
	Tổng	64.7	81.8			0.4344	0.3636	0.3641
	Bq =	5.4	7.4					

Hệ số tương quan :

$$\gamma = \pm\frac{\sum(K_x-1).(K_y-1)}{\sqrt{\sum(K_x-1)^2.\sum(K_y-1)^2}} = \frac{0.3641}{\sqrt{(0.4344)(0.3636)}} = 0.92$$

$\gamma = 0.92 > 0.8$ nên có thể dùng phương pháp tương quan để kéo dài tài liệu trạm A

Khoảng lệch quân phương:

$$\sigma_x = \bar{x}.\sqrt{\frac{\sum_{i=1}^{n}(Kx_i-1)^2}{n-1}} = 5.4\sqrt{\frac{0.4344}{11}} = 1.07$$

$$\sigma_y = \bar{y}.\sqrt{\frac{\sum_{i=1}^{n}(Ky_i-1)^2}{n-1}} = 7.4\sqrt{\frac{0.3636}{11}} = 1.35$$

Phương trình hồi qui y theo x :

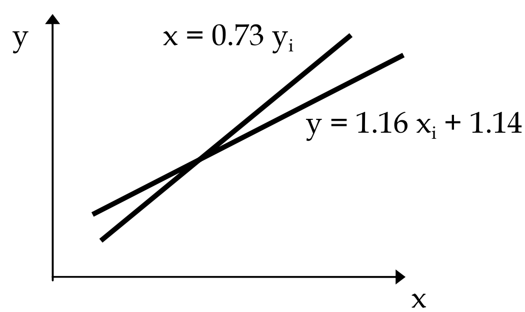
$$y-\bar{y} = \gamma\frac{\sigma_y}{\sigma_x}(x_i-\bar{x}) = y-7.4 = 0.92\frac{1.35}{1.07}(x_i-5.4)$$

$y = 1.16.x_I + 1.14$

Phương trình hồi qui x theo y :

$x = 0.73.y_i$

Dựa vào 2 phương trình hồi qui y theo x và x theo y, ta có thể vẽ đường hồi qui như sau :



Hình 5.23: Đường thẳng phương trình hồi qui x theo y và y theo x

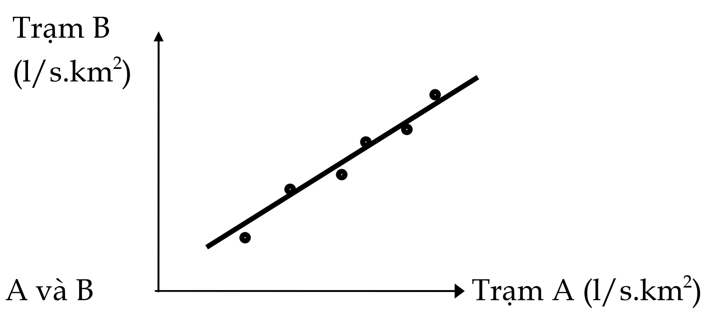
Căn cứ vào tài liệu 4 năm ở B (y_i đã biết), dùng phương trình của x theo y, ta tính được module dòng chảy trạm A từ 1954 - 1957 như ở bảng 5.9 sau :

Bảng 5.9: Kết quả tính toán kéo dài bổ sung số liệu trạm A

Năm	1954	1955	1956	1957
y_i (trị trạm B)	6.9	7.2	5.9	6.4
$x_i = 0.73\ y_i$ (trị trạm A)	5.04	5.26	4.31	4.67

2/. Dùng phương pháp tương quan đồ giải

Sau khi chấm các điểm tương quan (x_i, y_i), rồi định đường thẳng đi qua trung tâm các nhóm điểm để có hình tương quan sau :



Hình 5.23
Tương quan 2 trạm A và B

Căn cứ vào đường tương quan, dùng tài liệu đo đạt từ năm 1954 - 1957 của trạm B để suy ra trạm A :

Bảng 5.10: Kết quả kéo dài số liệu trạm A từ số liệu trạm B

Năm	1954	1955	1956	1957
M trạm A	5.0	5.3	4.2	4.6

===

Bài đọc thêm

MÔ HÌNH MƯA TIÊU

--- oOo ---

Các trận mưa lớn và dài ngày thường gây ngập úng cho các khu vực đồng bằng, việc xác định lượng mưa tiêu thiết kế gọi là ***tính toán mô hình mưa tiêu***.

Việc chọn lựa thời điểm tính toán rất quan trọng vì nó ảnh hưởng lớn đến qui mô và hiệu quả của công trình về sau. Thông thường, người ta chọn thời điểm tính toán mưa tiêu tương ứng với thời đoạn ngập do lũ lụt từ thượng nguồn dòng sông đổ về. Tần suất tính toán mưa cũng phù hợp với tần suất lũ. Thực tế, hai yếu tố này không phải lúc nào cũng xuất hiện đồng thời, nhất là các khu vực như Đồng bằng sông Cửu long. Các trận mưa lớn nhiều lúc xảy ra không cùng thời điểm của lũ lớn (lũ lớn đôi khi xảy ra trong thời kỳ mực nước lớn ứng với giai đoạn thủy triều tháng lớn nhất). Có những trận mưa lớn nhưng không gây úng cho cây trồng. Do vậy, cần cân nhắc khi chọn thời điểm tính toán mưa, mục đích tính mưa tiêu cho loại cây trồng nào, ở thời đoạn sinh trưởng nào.

Việc tính toán mô hình mưa phục vụ canh tác nông nghiệp, đặc biệt cho lúa và một số hoa màu có khả năng chịu ngập kém. Trong thời vụ tính toán, ta có thể chia các thời đoạn nhỏ hơn và thống kê lượng mưa ứng với thời đoạn nhỏ đó, ví dụ tính lượng mưa lớn nhất trong 1 ngày, 3 ngày, 5 ngày, 7 ngày, ... Theo nhiều thống kê khác nhau, người ta có kết quả tham khảo như sau:

- Đối với các vùng mưa bão (miền Bắc, miền Trung Việt Nam)

$$\overline{X}_{1\max} \leq \overline{X}_{3\max} \leq \overline{X}_{5\max} \leq \overline{X}_{7\max} \leq \ldots$$
$$Cv_{1\max} \geq Cv_{3\max} \geq Cv_{5\max} \geq Cv_{7\max} \geq \ldots$$

- Đối với các vùng mưa còn lại (miền Nam Việt Nam)

$$\overline{X}_{1\max} \leq \overline{X}_{3\max} \leq \overline{X}_{5\max} \leq \overline{X}_{7\max} \leq \ldots$$
$$Cv_{1\max} \leq Cv_{3\max} \leq Cv_{5\max} \leq Cv_{7\max} \leq \ldots$$

Đa số ngày xuất hiện lượng mưa max liên tiếp thường theo qui luật:

$$1 \text{ ngày max} \supset 3 \text{ ngày max} \supset 5 \text{ ngày max} \supset 7 \text{ ngày max} \supset \ldots$$

Sau khi có chuỗi mưa max trong nhiều năm, ta tiến hành tính toán tần suất xuất hiện theo các phương pháp đã biết (phương pháp đường thích hợp, phương pháp 3 điểm, ...). Sau đó ta xác định lượng mưa tính toán theo tần suất công trình.

Việc lựa chọn mô hình mưa đại biểu phải thỏa một số yêu cầu sau:

- Trận mưa lớn đã xảy ra gây úng lụt trong thực tế đại biểu cho một nguyên nhân gây mưa úng nhất định trong khu vực tiêu.
- Có số ngày mưa hiệu quả (mưa vượt quá tổn thất) thường xảy ra trong nhóm ngày mưa tính toán.
- Có lượng mưa toàn trận mưa xấp xỉ với lượng mưa trong khoảng thời khoảng khống chế T ứng với tần suất thiết kế.

Thu phóng mô hình mưa thiết kế, ví dụ mưa 5 ngày max liên tiếp, có 2 cách:

- ***Dùng phương pháp cùng tần suất***

  Theo phóng theo 3 tỉ số (chỉ số dưới p là lượng mưa theo tần suất, db là lượng mưa đại biểu, hay lượng mưa thực tế đã xảy ra):

  + $K_1 = \dfrac{X_{1\max p}}{X_{1\max db}}$ (thu phóng 1 ngày mưa lớn nhất)

  + $K_3 = \dfrac{X_{3\max p} - X_{1\max p}}{X_{3\max db} - X_{1\max db}}$ (thu phóng 2 ngày còn lại trong nhóm 3 ngày)

  + $K_5 = \dfrac{X_{5\max p} - X_{3\max p}}{X_{5\max db} - X_{3\max db}}$ (thu phóng 2 ngày còn lại trong nhóm 5 ngày)

- ***Phương pháp mô hình đại biểu***

  Thu phóng theo cùng tỉ số K_5' cho tất cả các ngày trong trận mưa được chọn

  + $K_1' = \dfrac{X_{5\max p}}{X_{5\max db}}$

Phương pháp này cũng có thể áp dụng cho việc thiết lập ***mô hình triều tiêu***.
(Phần thủy triều xem ở chương kế tiếp).

==