Chương 1

CÁC THÔNG TIN BAN ĐẦU TỪ LÝ THUYẾT XÁC SUẤT VÀ THỐNG KÊ TOÁN HỌC

1.1. CÁC LUẬN ĐIỂM XUẤT PHÁT TRONG CƠ SỞ SỬ DỤNG PHƯƠNG PHÁP LÝ THUYẾT XÁC SUẤT VÀ THỐNG KÊ TOÁN HỌC TRONG THUỶ VĂN

Phương pháp lý thuyết xác suất và thống kê toán học sử dụng trong các lĩnh vực khác nhau của thuỷ văn học. Tuy nhiên sử dụng rộng rãi nhất các phương pháp này trong tính toán và dự báo các đặc trưng của dòng chảy sông ngòi.

Khi thiết kế các dự án điều tiết dòng chảy , khi thi công và vận hành các thuỷ công trình, hệ thống tưới tiêu, cầu cống và khi thực hiện các biện pháp thuỷ công khác gắn liền với việc sử dụng tài nguyên nước đòi hỏi phải đánh giá định lượng các tham số dòng chảy sông ngòi thay đổi theo thời gian và không gian. Có nghĩa là nhất thiết xác định các đại lượng lưu lượng nước trung bình, cực đại và cực tiểu năm, phân phối dòng chảy trong năm, đại lượng dòng chảy phù sa v.v..

Các đại lượng sử dụng để thiết kế cần phải đặc trưng cho chế độ thuỷ văn của đối tượng nước nghiên cứu trong tương lai - thời kỳ vận hành trạm thuỷ lợi , tính toán cho hàng chục và hàng trăm năm sau.

Rõ ràng, bàn về các giá trị khả năng trong tương lai của tham số này hay tham số kia của chế độ thuỷ văn có thể nhận được chỉ khi dựa trên các tài liệu đo đạc thuỷ văn đã được tiến hành cho thời kỳ nhiều năm. Khi đó về nguyên tắc có thể sử dụng ba hướng.

Hướng thứ nhất là hướng tất định - xác định đại lượng ta quan tâm rồi sử dụng nó để liên kết dòng chảy với các nhân tố chi phối nó.

Hướng thứ hai dựa trên việc sử dụng đồng thời các qui luật nhân quả và thống kê đặc trưng cho dòng chảy sông ngòi và và các nhân tố xác định nó.

Con đường thứ ba gắn liền với sử dụng trực tiếp các qui luật thống kê thể hiện trong chuỗi các đại lượng thuỷ văn.

Sơ đồ dự báo dựa trên việc sử dụng các qui luật nhân quả (hướng thứ nhất), với sự phát triển của thuỷ văn hiện đại cho phép xác định đại lượng các đặc trưng thuỷ văn với thời hạn không vượt quá vài tháng. Hơn nữa, độ chính xác các ước lượng như vậy giảm nhanh khi tăng thời gian dự kiến.

Một vài nhà nghiên cứu cố gắng xác định khả năng của các sơ đồ dự báo khi xem xét chuỗi các đại lượng thuỷ văn như là một hàm tuần hoàn theo thời gian. Cơ sở của ý tưởng đó về dòng chảy thư nhất là chu kỳ thay đổi nước trong năm, có nghĩa là lần lượt các pha dòng chảy lặp lại mỗi năm theo một tuần tự giống nhau và khoảng lệch thời gian xuất hiện bé. Tuy nhiên, với sự hiện diện chu kỳ năm hiện tượng biến động dòng chảy không biến mất, cho nên coi nguyên nhân thứ hai của sự thay đổi đó kéo theo các ý tưởng về dao động tuần hoàn của bức xạ mặt trời và các quá trình địa vật lý khác. Các kết quả nhận được trong lĩnh vực này tới nay chứng tỏ rằng còn chưa giải quyết vấn đề ở chỗ mức độ nào đại lượng thuỷ văn là hàm xác định của thời gian và bằng cách nào dạng của nó có thể xác định trên cơ sở tài liệu quan trắc .

Như vậy, cần xét tới tinhg huống là việc xác định biến trình thời gian một đặc trưng thuỷ văn nào đó cho thời đoạn tính toán hàng chục năm hiện còn là vấn đề nan giải. Tuy vậy, dự báo các đặc trưng thuỷ văn với hạn nagứn là rất quan trọng, vì quá trình vận hành các trạm thuỷ lợi chúng cho phép gắn với các điều kiện cụ thể nào đó của chế độ nước. Các qui phạm dự báo cũng được sử dụng rộng rãi khi qui hoạch nhiều biện pháp thuỷ lợi.

Việc sử dụng đồng thời các quan hệ nhân quả gắn liền các đại lượng dòng chảy sông ngòi và các nhân tố xác định nó với ước lượng thống kê các tham số và biến của các quan hệ đó là cách giải quyết xét hiệu quả hơn có tính nguyên tắc bài toán đang xét. Tuy nhiên do độ tin cậy thấp của các phương trình quan hệ nhân quả và độ xử lý thấp của các phương pháp nhóm thống kê , việc sử dụng hướng này rất hạn chế. Thường chúng hay được áp dụng để tính toán các đặc trưng dòng chảy (cụ thể là lưu lượng nước với các suất đảm bảo khác nhau) sông ngòi chưa được nghiên cứu về phương diện thuỷ văn. Khi đó suất đảm bảo biến chính của quan hệ và đại lượng cần tìm là giống nhau, tức là sử dụng dạng đơn giản nhất xác định suất đảm bảo của hàm.

Ngày nay, các giá trị biến đổi mang các thủ thuật ước lượng giá trị tính toán của các đại lượng thuỷ văn mô tả theo qui luật thống kê đặc trưng cho chuỗi các đại lượng thuỷ văn. Khả năng sử dụng cách đó để thu được các giá trị tính toán tham số của chế độ thuỷ văn dựa trên giả thuyết rằng chuỗi các đại lượng đang xét được hình thành như là một tập ngẫu nhiên.

Sự tiếp nhận giả thuyết về sự phụ thuộc của dao động các đại lượng thuỷ văn theo các qui luật dao động đặc trưng bởi các số ngẫu nhiên có nghĩa là gắn thời gian xuất hiện đại lượng này hay đại lượng kia (ví dụ trong liệt quan trắc) là không lớn, ngẫu nhiên. Để mô tả tính chất tập các đại lượng như vậy trong sự lặp các giá trị khác nhau của đại lượng đó ở giới hạn tập đủ lớn.

Luận điểm này về tính chất ngẫu nhiên của sự hình thành chuỗi thuỷ văn không thể chứng minh trọn vẹn bằng lý thuyết, tuy nhiên áp dụng tới chuỗi dòng chảy sông ngòi (năm, cực đại) nó không chỉ một lần được khẳng định bằng kiểm chứng bởi việc đánh giá sự tương ứng của các đường cong đảm bảo dòng chảy thực nghiệm và sơ đồ lý thuyết.

Các kết quả phân tích như vậy được xét ở chương 4 chứng tỏ tính đúng đắn của việc sử dụng hướng thống kê làm cơ sở cho nhiều thủ thuật tính toán thuỷ văn.

Bằng các luận điểm lý thuyết dùng để làm cơ sở cho khả năng xem chuỗi các đại lượng ngẫu nhiên khác nhau như là một tập các biến cố ngẫu nhiên được gọi là các định luật tới hạn của lý thuyết xác suất.

Một trong những luận điểm cơ bản nhất của các định luật này dẫn tới qui luật số lớn, theo nó với một số lượng lớn các hiện tượng đồng nhất ngẫu nhiên, kết quả trung bình của chúng hầu như mất ngẫu nhiên và có thể dự đoán được với mức xác định cao.

Tính chất nêu trên của các hiện tượng ngẫu nhiên thể hiện khá rõ ràng ở trong chuỗi các đại lượng thuỷ văn ở chỗ theo độ tăng của số thành viên của tập, đường cong đảm bảo có dạng bền vững. Các ví dụ cụ thể theo vấn đề này sẽ dẫn trong chương 5.

Luận điểm thứ hai dẫn tới định luật tới hạn trung tâm, mà theo nó hiện tượng (biến cố) xuất hiện dưới tác động của tổng hoặc tích số lớn các nhân tố ngẫu nhiên độc lập (ít phụ thuộc) tạo nên tập ngẫu nhiên tuân theo các qui luật thống kê xác định.

Rõ ràng, nhiều hiện tượng thuỷ văn có thể xem xét thoả mãn sơ đồ này.

Thực vậy, khi xét điều kiện hình thành một đặc trưng thuỷ văn nào đó, thí dụ như lưu lượng nước cực đại của lũ xuân có thể dễ dàng xác định rằng sự xuất hiện hiện tượng này diễn ra hàng năm theo dạng của một qui luật tất định. Tuy nhiên đại lượng lưu lượng nước cực đại trong một năm cụ thể nào đó được hình thành dưới ảnh hưởng của rất nhiều nhân tố, xác định trữ lượng ẩm trong tuyết, mức độ ẩm ướt của lưu vực , lượng mưa trong quá trình hình thành lưu lượng nước cực đại, quá trình cường độ tan v.v.. Dưới tác động của nguyên nhân này hay nguyên nhân khác xuất hiện ở trong năm cụ thể đó một qui luật đã biết mang dạng của biến cố ngẫu nhiên.

Dựa trên các luận điểm thống kê đã nêu , cần thấy rằng chúng giả thiết thiếu sự biến đổi một chiều trong điều kiện hình thành hiện tượng thuỷ văn của đại lượng đang xét trong giới hạn của một tập đủ lớn (cụ thể cho n năm đủ dài). Nếu điều kiện hình thành thay đổi đơn trị như dướt tác động của hoạt động kinh tế trong đại lượng của các đặc trưng thuỷ văn thu được sau tác động đã nêu cần phải được hiệu chỉnh. Rõ ràng các hiệu chỉnh này thực hiện có ý nghĩa nếu như sự thay đổi một chiều trong điều kiện hình thành lớn đến mức ảnh hưởng tới các đại lượng của tham số thuỷ văn đang xét về qui mô nằm ngoài giới hạn sai số tính toán.

Coi luận điểm mực trung bình không thay đổi, xuất phát từ kết luận rằng các đại lượng thu được trên cơ sở phân tích chuỗi đại lượng thuỷ văn đang có với thời đại địa chất và khí hậu ngày nay vì sự thay đổi gắn với lịch sử địa chất trái đất (và các thay đổi khí hậu tương ứng) thực hiện trong thời kỳ lớn hơn nhiều lần thời đoạn tính toán. Cho nên chuỗi đo đạc thực tế đang có được coi như là một mẫu nào đó từ tập tổng thể lý thuyết gồm các số đại lượng vô hạn các đặc trưng thuỷ văn ta quan tâm. Khi đó tập mẫu đang có cần thực tế có nghĩa là mang tính dại diện đối với toàn tập. Nói cách khác, nó cần chứa đủ các năm nhiều, ít và nước trung bình nếu như xét đặc trưng dòng chảy sông ngòi. Tương tự, cần phải hình thành ngay cả chuỗi các đại lượng thuỷ văn khác để đánh gía chúng sử dụng các phương pháp lý thuyết xác suất.

Điều kiện quan trọng tiếp theo của tính ứng dụng các phương pháp lý thuyết xác suất trong thuỷ văn là yêu cầu nguyên tắc đồng nhất các đại lượng thuỷ văn trong một tập. Cụ thể (áp dụng cho các đặc trưng dòng chảy sông ngòi) điều này biểu hiện trước hết ở chỗ chọn các đại lượng dòng chảy đồng nhất căn nguyên (lưu lượng) vào trong một tập. Chỉ tiêu đồng nhất có thể là tính phụ thuộc của đại lượng dòng chảy đang xét trong một pha chu kỳ thay đổi trong năm của nước. ĐIều này thường được sử dụng khái niệm về lưu lượng nước đồng pha. Do dao động theo năm ngày bắt đầu và kết thúc các pha khác nhau của chu kỳ trong năm chọn đại lượng dòng chảy đồng pha đã nêu không thể gắn chặt với ngày này mà được thực hiện xuất phát từ việc đánh giá tính đồng nhất tổng thể của chúng theo bản chất.

Vậy, đồng nhất tổng thể là lưu lượng nước lũ xuân cực đại, lũ mưa, thể tích dòng chảy cho một pha đồng nhất trong năm (xuân, kiệt), giá trị năm của dòng chảy tổng cộng và dòng chảy thành phần của nó (mặt, ngầm) v.v..

2.2 CÁC PHƯƠNG PHÁP KHÁI QUÁT SỐ LIỆU THỐNG KÊ ĐƠN GIẢN NHẤT.

Kết quả quan trắc và đo đạc thuỷ văn có thể thể hiện bằng bảng, đồ thị hoặc dạng giải tích.

Nhiều thủ thuật có thể tìm thấy trong các Niên giám thuỷ văn , trong các chuyên khảo " Tài nguyên nước mặt Liên bang Xô viết" và các ấn phẩm khác trong đó trình bày các số liệu về thành phần của chế độ nước sông ngòi với các mức khái quát chúng khác nhau. Xử lý các số liệu này bản thân đã là một ví dụ của thống kê mô tả. Thường thông tin cơ bản ban đầu nằm trong các bảng rất cồng kềnh. Trong trường hợp này sử dụng nó ở dạng nguyên thuỷ khi nghiên cứu nhiều vấn đề thuỷ văn là khó khăn. Cho nên từ lâu đã sử dụng nhiều phương pháp khác nhau để khái quát bằng số thông tin ban đầu (tính giá trị trung bình, dẫn các giá trị tới hạn theo thành phần này hoặc kia của chế độ thuỷ văn và v.v..).

Một cách thể hiện đầy đủ hơn và đồng thời tiện lợi hơn có thể thực hiện được trên cơ sở sử dụng các phương pháp thống kê toán học, là khoa học phân tích định lượng các hiện tượng đại chúng đồng thời tính đến cả tính chất đặc thù của chúng.

Xét một vài ví dụ minh hoạ các thủ thuật sử dụng khi xử lý thống kê số liệu thuỷ văn và đồng thời dẫn một số khái niệm và định nghĩa thống kê .

Trong bảng 1.1 là kết quả quan trắc dòng chảy trung bình năm sông Dnhepr tại tuyến đo Loxmanskaia Kamenka từ năm 1918 đến 1962. để tăng độ trực quan của kết qủa quan trắc và để thể hiện chúng tiện lợi hơnkhi xử lí tiếp theo số liệu ban đầu thường được dẫn theo bảng nhóm. Với mục đích đó từ bảng 1.1 chọn các giá trị cực đại (Q_{max}) và cực tiểu (Q_{min}) của lưu lượng nước và tính hiệu giữa chúng đ, gọi là biên độ hay là khoảng dao động,

$$R = Q_{max} - Q_{min} = 3040 - 717 = 2323 \text{ m}^3/\text{s}.$$

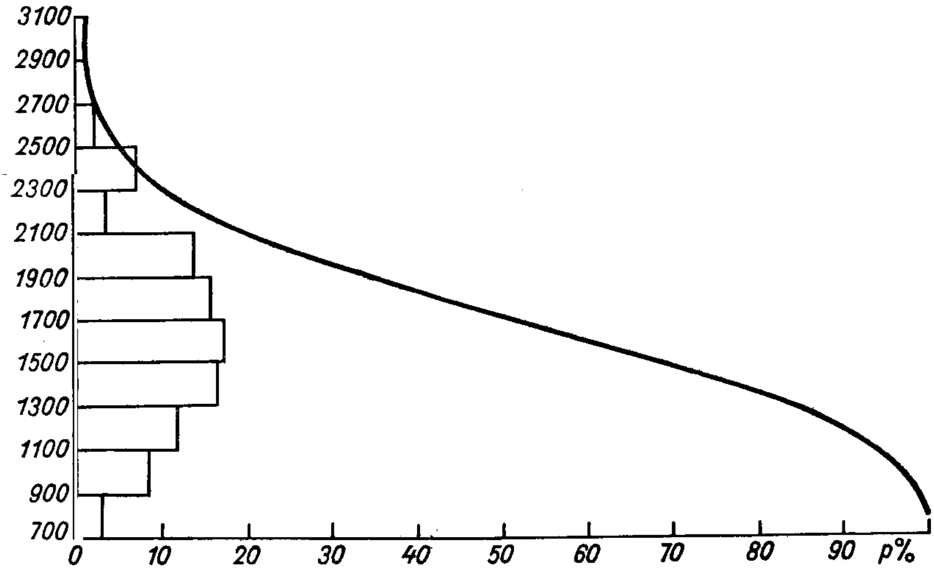
Bảng 1.1 Lưu lượng nước trung bình sông Dnhepr tại Loxmanskaia Kamenka

Năm	$\overline{Q}$	Năm	$\overline{Q}$	Năm	$\overline{Q}$	Năm	$\overline{Q}$

Biên độ chung của dao động đại lượng ngẫu nhiên có thể được chia ra các phần riêng biệt mà ngoài ranh giới giữa chúng nhận một số điểm (đại lượng) đặc trưng nào đó của chuỗi. Vậy, khi chia chuỗi các đại lượng ngẫu nhiên sắp xếp theo thứ tự giảm dần ra 4 phần , chia ra 4 đoạn: trên cùng hay là đoạn thứ nhất là những giá trị của biến mà dưới nó là ấ số hạng của tập, đoạn thứ hai là chiếm vị trí giữa chuỗi, và đoạn dưới hay là đoạn thứ ba mà dưới nó là ẩ số hạng của tập. Không hiếm khi người ta chia biên độ theo phần trăm. Các giá trị đại lượng ngẫu nhiên phân bố trên các ranh giới đó được gọi là ....

Trong trường hợp chung khi nhận một tập thống kê phân bố liên tục trong giới hạn cả biên độ thì có khả năng xét thành phần bất kỳ nào của tệp nằm giữa hai ranh giới chỉ định bất kỳ. Trong trường hợp này mọi giá trị (kể cả các giá trị nói trên) của biến nhận được các điểm đặc trưng nhất định gọi là điểm đoạn phần tư.

Biên độ nhận được có thể chia ra các khoảng, hoặc phân cấp và tính số lần đạt của dấu hiệu đang thử (lưu lượng nước) cho mỗi phân cấp. Các khoảng này có thể bằng và không bằng nhau theo giá trị. Trong thuỷ văn thường sử dụng các phân cấp bằng nhau theo giá trị. Số lượng các phân cấp thường được lựa chọn phụ thuộc vào dung lượng tài liệu đang xét đẻ nó có thể phản ánh các nét cơ bản nhất của chuỗi quan trắc đang xét. Khi đó với sự tăng độ dài của khoảng số lần đạt của biến nghiên cứu vào trong mỗi khoảng sẽ tăng lên và tăng độ tin cậy thống kê của tài liệu đang thể hiện. Nhưng với dung lượng quan trắc lớn và độ dài của khoảng lớn số phân cấp sẽ không lớn, và khi đó sẽ san bằng các nét đặc thù của chuỗi quan trắc này hoặc kia. Khi giảm độ dài của khoảng số lần đạt trong khoảng sẽ giảm và khả năng nguy hiểm xuất hiện qui luật không đặc trưng cho chuỗi thống kê đã cho. Đẩi với đánh giá sâu sắc số khoảng thường sử dụng các công thức kinh nghiệm, như $n_x \leq 5\lg N$, với n_x - số khoảng; N- dung lượng chung của quan trắc.



Hình 1.1 Biểu đồ phân bố và đường cong tích luỹ tần số dòng chảy năm sông Dnhepr tại Loxmanskaia Kamenka.

Nhận thấy rằng, các công thức như vậy không thể hiện qui tắc chung và vì thế có thể xét chỉ trong chính lần xấp xỉ đầu tiên, khi không có một thông tin bổ sung nào ngoài chuỗi quan trắc đang nghiên cứu và khi dung lượng số liệu ban đầu không lớn và không nhỏ lắm. Với dung lượng nhỏ số liệu thống kê sự nhóm theo khoảng hầu như trở thành một bài toán không thể thực hiện. Và với số quan trắc lớn, áp dụng công thức trên có thể đưa đến số phân cấp lớn và làm tăng khối lượng tính toán. Các phân cấp được chọn không cần phải phủ nhau để một và chỉ một giá trị chuỗi quan trắc không có thể rơi vào hai phân cấp.Nếu giá trị quan trắc rơi vào ranh giới phân cấp thì nó được coi là thuộc phân cấp lớn hơn.

Áp dụng cho ví dụ đang xét chỉ định tương ứng với những lập luận trên 12 phân cấp bằng nhau và tính số trường hợp đạt lưu lượng nước trong mỗi phân cấp.

Kết quả tính toán đưa vào bảng 1.2, trên đầu cột ghi tên phân cấp và dòng 1 - số trường hợp rơi lưu lượng nước vào mỗi phân cấp. Rõ ràng, tổng các trường hợp theo mọi phân cấp bằng số năm quan trắc . bảng được lập như vậy gọi là bảng phân bố thực nghiệm , hoặc là bảng tần số tuyệt đối. Khi biểu diễn tần số tuyệt đối bằng phần trăm so với tổng các trường hợp ta có phân bố tần số tương đối (dòng 2, bảng 1.2), tổng lần lượt nó cho ta các tần số luỹ tích tuyệt đối và tương đối (dòng 3 và 4 bảng 1.2).

Tổng các tần số tương đối bằng 100%, và có thể sử dụng khi kiểm tra tính đúng đắn của tính toán . Số liệu bảng 1.2 chỉ ra rằng thường xuyên nhất lưu lượng nước trung bình năm s. Dnhepr tại Loxmanskaia Kamenka được quan trắc trong khoảng 1500-1700 m^3/s; với sự tăng hoặc giảm lưu lượng nước số trường hợp giảm một cách có qui luật không tính đến những chênh lệch riêng lẻ với qui tắc này, nó có thể coi là các dao động ngẫu nhiên.

Bảng 1.2 Nhóm số liệu dòng chảy năm s. Dnhepr tại Loxmanskaia Kamenka.

Số liệu bảng 1.2 có thể thể hiện dưới dạng đồ thị (h. 1.1), trên đó theo trục tung đặt các phân cấp lưu lượng nước đã nhận, còn theo trục hoành ở dạng các hình chữ nhật - tần số tương đối. Cũng ở đây trong dạng một đường cong mềm mại chỉ rõ sự tăng trưởng tần số tương đối. Tổng tăng trưởng tần số gắn với giá trị lớn hơn của mỗi khoảng.

Đồ thị thu được của tần số tương đối gọi là ...., còn đồ thị tần số tương đối luỹ tích - .... hay là đường cong tích luỹ. Sự trình diễn bằng bảng hoặc đồ thị các tần số gọi là phân bố thực nghiệm , trong trường hợp này là dòng chảy năm s. Dnhepr tại Loxmanskaia Kamenka.

Diện tích mỗi phần riêng của ... bằng tích của kích thước phân cấpvà tần số tương đối, còn tổng diện tích .... - tổng của các tích đó.

Đường cong luỹ tích là đồ thị thể hiện độ lặp của lưu lượng nước lớn hơn giá trị cho trước.

Giả sử chúng ta quan tâm lưu lượng nước trung bình năm lớn hơn 1900 m^3/s thường quan trắc được không? Với lưu lượng này lấy từ đường cong luỹ tích giá trị độ lặp bằng 44,5%. Điều này có nghĩa là đại lượng lưu lượng nước 1900 m^3/s và lớn hơn được quan trắc trong 44,5 % mọi trường hợp. Nếu như chúng ta quan tâm vấn để độ lặp lại nào không vượt quá lưu lượng nước đã cho thì đáp số sẽ là 100% - 44,5% = 55,5%.

Trong thuỷ văn đường cong tần số luỹ tích tương đối được gọi là đường cong đảm bảo thực nghiệm . Và vì thế người ta nói rằng đại lượng lưu lượng nước bằng hoặc lớn hơn 1900 m^3/s được đảm bảo 44,5%, còn đại lượng lưu lượng nước 1900 m^3/s và nhỏ hơn đảm bảo 55,5%.

Chia tần số tương đối (hoặc tuyệt đối) lưu lượng nước cho độ dài của khoảng ta thu được tương ứng mật độ phân bố tương đối (hoặc tuyệt đối) (hàng 5 và 6 bảng 1.2). Mật độ phân bố sử dụng đặc biệt hợp lý khi cần nhận các phân cấp không đều theo nguyên nhân này hoặc kia. Diện tích bao bởi trục hoành và đường thẳng đặc trưng cho mật độ phân bố tương đối bằng 1 nếu tần suất tương đối được xác định bằng thập phân của đơn vị , hoặc bằng 100%, nếu như tần suất tương đối biểu diễn bằng phần trăm của tổng các trường hợp.

Ta xét thêm một ví dụ. Đối với việc mô tả thống kê bề mặt của một vi cảnh quan đầm lầy thông - cây bụi - rêu nước chỉ định một mặt cắt, trên đó cứ 10 cm xác định cao độ của bề mặt đầm lầy so với mực nước giả định.

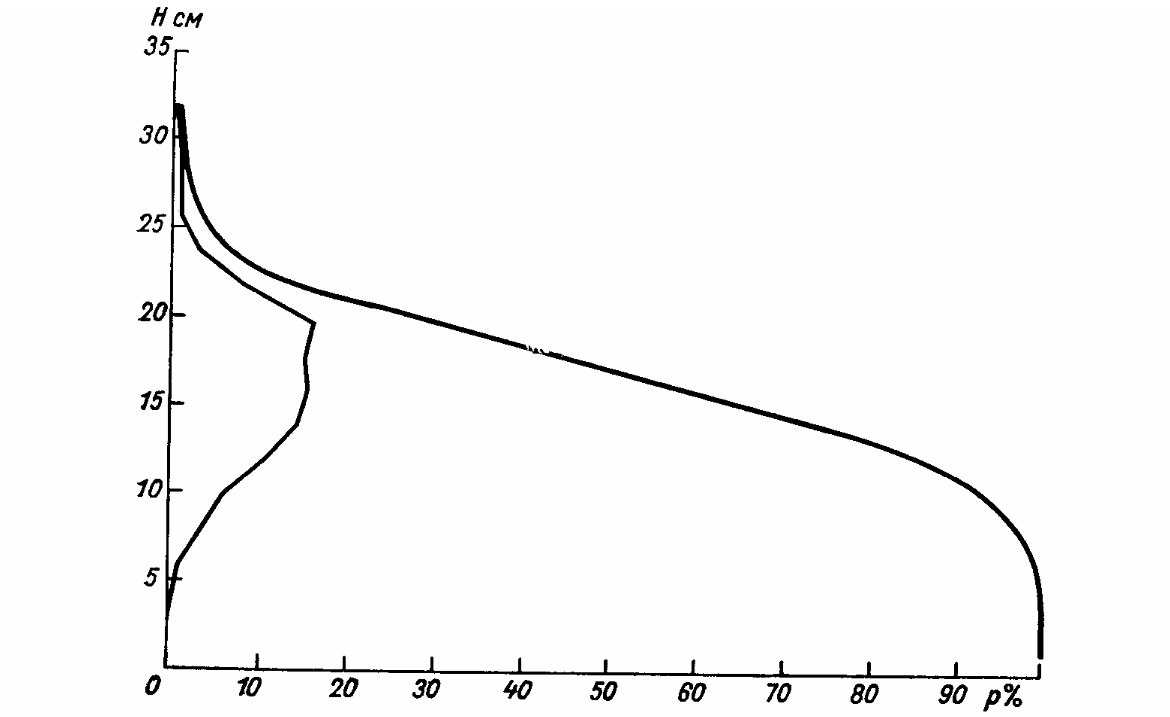
Kết quả quan trắc này (theo số liệu P. K. Varobiev) được cho vào bảng 1.3, trong đó cũng dẫn các đường cong phân bố thực nghiệm tính toán.

Tần suất tương đối trên h.1.2 đặt vào giữa khoảng, các điểm thu được được nối bằng các đường thẳng. Sự thể hiện tương tự các số liệu thống kê được gọi là đa giác phân bố (tần suất). Hay lặp nhất là cao độ bề mặt đầm lầy so với mực nước ngầm chiếm từ 15-20 cm. Đường cong đảm bảo được xây dựng cũng giống như ví dụ trước đây.

Từ đa giác tần số dẫn trên h. 1.2 suy ra rằng về cả hai phía của giá trị này tần suất tương đối giảm.

Bảng 1.3 Nhóm số liệu cao độ bề mặt vi cảnh quan đầm lầy

Đồ thị đã dẫn chứng tỏ rằng riêng các khái quát thành phần cơ bản cho phép thể hiện số liệu thống kê ban đầu ở dạng trực quan và tiện lợi hơn. Đồng thời có thể nhận thấy rằng các dạng khái quát tài liệu thống kê đang xét ứng với các đặc trưng thuỷ văn rất khác nhau cho phép phát hiện một vài qui luật thống kê chung. Cùng với nó phân bố lưu lượng nước năm và độ cao mặt đầm lầy có các đặc thù riêng có thể mô tả được nhờ sử dụng một vài khái niệm bổ sung mà chúng ta sẽ xem xét.



H.1.2 Đa giác phân bố và đường cong tích luỹ tần số cao độ vi cảnh quan (H) đầm lầy Lammin - Suo.

1. 3. KHÁI NIỆM XÁC SUẤT

A. N. Kolmogorov cho khái niệm xác suất đầy đủ nhất và kèm với nó là trừu tượng nhất; nó dựa trên 5 tiên đề dựa trên lý thuyết số đông. Không dừng lại ở các tiên đề của Kolmogorov vì điều đó đòi hỏi phải trình bày bổ sung một vài khái niệm của lý thuyết số đông, ta chuyển sang khám phá tư tưởng của khái niệm xác suất theo sơ đồ của Kolmogorov.

1. Giả sử rằng có tập các điều kiện S có thể lặp vô hạn lần. Dưới điều kiện S có thể hiểu như các nhân tố hình thành lưu lượng nước cực đại trong năm mà nó trôi cùng thời gian đồng nhất, có nghĩa là không quan sát thấy sự đổi hướng theo thời gian.

2. Dưới tác động của điều kiện S hình thành trong trường hợp này tập các lưu lượng nước cực đại (Q_{max}) cho thời đoạn đủ dài.

3. Khi tuân thủ vài điều kiện cho mỗi lưu lượng nước có thể quan trắc đọc hoặc không cho n năm có thể liên tưởng một số thực xác định $P(Q_{max})$ gọi là xác suất xuất hiện của đại lượng đang xét. Số $P(Q_{max})$ có các tính chất sau:

1) khi lặp điều kiện S một số lần đủ lớn tần suất tương đối $\frac{m}{n}$ của lưu lượng Q_{max} trong các khoảng đã cho sẽ không khác mấy xác suất $P(Q_{max})$. Ở đây m ký hiệu số trường hợp xuất hiện Q_{max} trong n lần lặp điều kiện S;

2) nếu giá trị xác suất $P(Q_{max})$ rất bé thì có thể không liều mà khẳng định rằng với sự thực hiện một lần điều kiện S giá trị lưu lượng đã cho Q_{max} không xuất hiện.

Định nghĩa cổ điển xác suất dựa trên nguyên tắc khả năng đồng đều. Khi đó thường dẫn các ví dụ đã trở thành kinh điển như tung đồng tiền (rơi mặt hình hay số) và con xúc xắc (rơi mặt nào đó trong sáu khả năng). Trong trường hợp thứ nhất xác suất xuất hiện hình hay số bằng ẫ, còn trường hợp thứ hai xác suất xuất hiện mặt nào đó của con xúc xắc là 1/6. Tất nhiên ở đây bàn đến các đồng tiền và xúc xắc đồng nhất.

Nguyên tắc khả năng đồng đều trong quan trắc thuỷ văn nếu được thực hiện thì cũng rất hiếm. Trong các trường hợp như vậy tiên nghiệm , trước khi thử, xác định xác suất xuất hiện biến cố nào đó là không thể. Nó chỉ được đánh giá trên cơ sở xác định xác suất thực nghiệm hoặc thường xuyên, là chung nhất và bao gồm cả xác định xác suất kinh điển như là trường hợp riêng.

Xác suất thực nghiệm của một biến cố A nào đó được gọi là thương số mà tử là số trường hợp xuất hiện biến cố A, còn mẫu là tổng các trường hợp thuộc một cấp xác định nào đó của thực nghiệm ngẫu nhiên.

Khi tăng số thực nghiệm đến vô cùng thì xác suất thực nghiệm tiến đến giới hạn của mình - xác suất lý thuyết. Thực vậy, nếu chúng ta tung đồng tiền giả ử là 10 lần thì hoàn toàn không nhất thiết có 5 lần hình và 5 lần số. Trong trường hợp đó xác suất thực nghiệm không bằng ẫ. Nếu số lần tung đồng tiền tăng lên thì rõ ràng xác suất thực nghiệm ngày càng gần với giá trị ẫ, tứclà gần với giới hạn lý thuyết .

Khi nghiên cứu các tập thống kê các đại lượng thuỷ văn không biết trước xác suất lý thuyết . Cho nên để ước lượng xác suất lý thuyết thường sử dụng xác suất thực nghiệm, càng gần nhất với lý thuyết khi dung lượng quan trắc (tệp) càng lớn.

Xác suất thực nghiệm biến cố A, ký hiệu qua P(A), bằng m/n, tức là:

$$P(A) = \frac{m}{n},$$

với m - số trường hợp thuận cho biến cố A, n tổng các trường hợp đang xét (dung lượng tệp). Xác suất thực nghiệm biến cố đối A, ký hiệu là $P(\overline{A})$ bằng:

$$P(\overline{A}) = \frac{n-m}{n} = 1 - P(A).$$

Rõ ràng P(A) + $P(\overline{A})$ =1. Xác suất xuất hiện biến cố thay đổi từ 0 tới 1, tức là 0≤ P(A)≤1. Đôi khi xác suất xuất hiện biến cố đang xét được biểu diễn bằng phần trăm. Trong trường hợp này thì giới hạn dao động của nó từ 0 đến 100%. Xác suất biến cố xuất hiện chắc chắn bằng 1, còn xác suất của biến cố không thể xuất hiện bằng 0.

Thể hiện trên h. 1.1 tổ chức đồ phân bố lưu lượng nước trung bình năm s. Dnhepr tại Loxmanskaia Kamenka có thể xem như phân bố xác suất thực nghiệm vì khái niệm tần suất tương đối của lưu lượng nước trong giới hạn phân cấp trong trường hợp đã cho là đồng nghĩa với khái niệm xác suất thực nghiệm.

Khi tăng dung lượng tệp, có ngiã là trong trường hợp đã cho tăng số năm quan trắc dòng chảy năm s. Dnhepr tại Loxmanskaia Kamenka có thể giảm kích thước phân cấp . Nếu số thành viên chuỗi tiến đến vô hạn, còn kích thước phân cấp tiến đến 0 ta nhận được dạng giới hạn của tổ chức đồ phân bố tương ứng với đường cong phân bố xác suất lý thuyết . Khi chuyển qua giới hạn diện tích chặn bởi đường cong phân bố xác suất và trục hoành tiến đến 1 thì diện tích này bằng xác suất cái gọi là đại lượng ngẫu nhiên đã cho nhận bất kỳ giá trị nào, tức là xác suất của biến cố chắc chắn.

Trong mối phụ thuộc vào đặc điểm hình thành của các tập thống kê hình dạng đồ thị của tổ chức đồ và các đường cong phân bố xác suất tương ứng có thể rất đa dạng. Trong số các đồ thị một đỉnh có thể chí ra các dạng cơ bản sau: 1) đối xứng, 2) bất đối xứng vừa phải, 3) rất bất đối xứng và 4) dạng chữ U (h.1.3).

Phân bố đối xứng là các dạng có tần số (xác suất) hai giá trị bất kỳ của biến nằm trên khoảng cách đồng đều về hai phía của giá trị trung bình nào đó bằng nhau.

Phân bố không đối xứng hoặc bất đối xứng là các dạng có tần số của đối số cách giá trị trung bình một khoảng nào đó về một phía luôn lớn hơn tần số phía ngược lại.

Phân bố rất không đối xứng là phạm trù ứng với các dạng khi mà tần số lớn nhất ứng với giá trị cực đại hoặc cực tiểu tưcs là mọi tần số phân bố theo một chiều nhất định so với tần số cực đại.

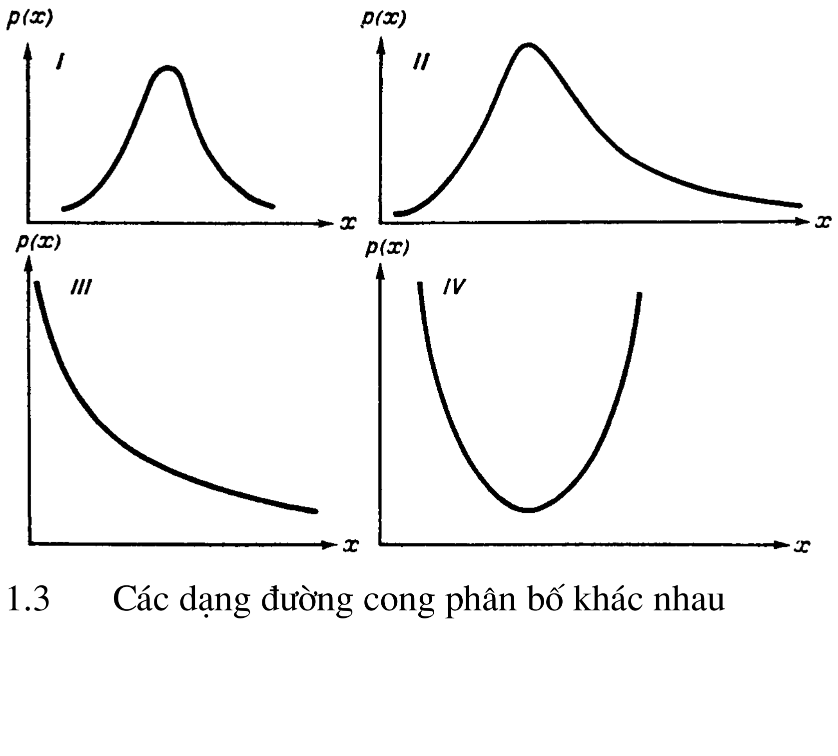
Phân bố dạng chữ U đặc trưng cho sự hiện diện trong đó một khoảng giữa có tần số nhỏ hơn phần còn lại và đột ngột tăng ở các giá trị biên của phân bố.

Trong thuỷ văn thường gặp phân bố bất đối xứng vừa phải và phân bố đối xứng, còn phân bố rất không đối xứng hiếm khi gặp.

Với các tính toán thuỷ văn thường nảy sinh việc cần mô tả đường cong phân bố xác suất thực nghiệm bằng giải tích, khi đó người ta thường sử dụng các qui luật phân bố đại lượng ngẫu nhiên khác nhau sẽ xét trong chương 2. Nhận làm các tham số mô tả các qui luật thống kê của chuỗi các đặc trưng thuỷ văn và các đường cong phân bố giải tích tương ứng người ta sử dụng giá trị trung bình (trung bình số học, trung vị và mod), mức độ phân tán (độ lệch quân phương hoặc độ lệch tuyệt đối), các chỉ số bất đối xứng, độ nhọn và v.v.. Các tham số này của các tập thống kê được trình bày trong các bài tiếp theo.

1. 4 TRUNG BÌNH SỐ HỌC VÀ CÁC TÍNH CHẤT CỦA NÓ. KỲ VỌNG TOÁN HỌC.

Một trong những tham số cơ bản nhất của chuỗi thống kê là giá trị trung bình của đại lượng mẫu, hay là trung tâm mà các thành viên của tập được phân bố.



Hình 1.3 Các dạng đường cong phân bố khác nhau

Tham số này hoặc tự mình, học kết hợp với các đặc trưng đang xét khác sau đây của chuỗi thống kê thường được sử dụng để mô tả qui luật thống kê của các tập riêng biệt.

Bên cạnh trung bình số học, nhận làm đặc trưng của trung tâm còn có trung vị, trung bình điều hoà và trung bình nhân sẽ xét trong các bài khác.

Trung bình số học chuỗi các đại lượng x được xác định theo công thức:

$$\overline{x} = \frac{1}{n}(x_1 + x_2 + \ldots + x_n = \frac{1}{n}\sum_{i=1}^{n} x_i . \qquad (1.1)$$

Tính trung bình khi nhóm các số liệu đo đạc thường thực hiện theo biểu thức:

$$\overline{x} = \frac{\sum_{i=1}^{k} n_i x_i}{\sum_{i=1}^{k} n_i} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i \qquad (1.2)$$

với k - số phân cấp; n_i - tần số tuyệt đối của phân cấp ; x_i - điểm giữa của khoảng.

Tính toán trung bình số học theo công thức (1.2) đơn giản hơn nhiều và giảm khối lượng tính toán , đặc biệt với n lớn. Khi đó ta coi phân bố đều dấu hiệu dao động trong phân cấp và đại lượng phân cấp càng bé thì càng đúng. Với dung lượng mẫu ít như thường gặp ở các chuỗi quan trắc thuỷ văn sử dụng công thức (1.1) tốt hơnvà chỉ khi dung lượng số liệu nhiều mới thực hiện nhóm theo (1.2).

Khi xét tới đẳng thức $n_i/n = P_i$ và $n = \sum_{i=1}^{k} n_i$, công thức (1.2) có thể dễ dàng chuyển về dạng:

$$\overline{x} = \sum_{i=1}^{k} P_i x_i \qquad (1.3)$$

với P_i - tần số tương đối hoặc là xác suất thực nghiệm .

Trung bình số học thường xuyên mang thứ nguyên của đại lượng đo đạc mà ta tính toán.

Xét các đặc trưng cơ bản của trung bình số học .

1. Tổng độ lệch mọi số liệu quan trắc với trung bình số học bằng 0.

$$\sum_{i=1}^{n} (x_i - \overline{x}) = 0 \qquad (1.4)$$

Tính chất này của trung bình số học thường được sử dụng để kiểm tra tính đúng đắn của tính toán độ lệch của số liệu quan trắc so với trung bình số học.

2. Tổng bình phương độ lệch các thành của chuỗi với trung tâm biểu diễn dưới dạng trung bình số học đạt cực tiểu so với tổng tương tự so với một số a bất kỳ $a \neq \overline{x}$

$$S = \sum_{i=1}^{n} (x_i - \overline{x})^2 = \min. \qquad (1.5)$$

3. Trung bình số học của chuỗi nhận được bằng cách trộn các nhóm thống kê đồng nhất tạo ra giá trị trung bình trọng lượng của các trung bình đưa vào trong tính toán với trọng số bằng giá trị tính theo dung lượng cả tập trộn.

$$\overline{x} = \frac{\sum_{k=1}^{m} n_k x_k}{\sum_{k=1}^{m} n_{ki}} \qquad (1.6)$$

Tính chất này của trung bình số học thường được sử dụng khi tính toán các đại lượng trung bình năm của các đặc trưng thuỷ văn theo các giá trị tháng của chúng. Xét đến tính không đều số ngày trong các tháng giá trị trung bình năm cần được xác định như trung bình trọng lượng theo số ngày trong mỗi tháng. Tuy nhiên khi chú ý tới sự thay đổi không lớn số ngày trong tháng (từ 28 đến 31) ta có trong trường hợp này trung bình số học tính từ các giá trị trung bình tháng không khác mấy so với giá trị trung bình trọng lượng theo số ngày trong các tháng. Trong trường hợp khác nhau nhiều dung lượng của các tập trộn việc xác định trung bình nhất định phải tuân thủ theo trung bình trọng lượng của mỗi tập riêng.

Trung bình số học áp dụng cho mọi chuỗi biến đổi bất kỳ nào đều bảo tồn ý nghĩa của tham số thống kê . tuy nhiên nếu tương quan với tập có đại lượng biến đổi, theo bản chất nó không có một giá trị hằng số nào, khi đó vai trò của trung bình số học cũng hạn chế, thì trong trường hợp đó, khi chuỗi thống kê được tạo thành do sự thay đổi một vài đại lượng có giá trị không đổi về nguyên tắc, trung bình số hcọ được coi như là giá trị gần đúng của đại lượng đó. Thí dụ như trong quan hệ của tập trung bình năm, cực đại và cực tiểu và các lưu lượng nước đặc trưng khác đại lượng trung bình số học có thể coi như một tham số thống kê bởi vì trong trường hợp này chỉ xét các đại lượng về nguyên tắc không có một giá trị không đổi nào hết.

Tương tự, giá trị trung bình của lưu lượng nước thay đổi trong thời đoạn chế độ không dừng như trong nhánh lên của lũ xuân không thể coi như là giá trị gần đúng với giá trị thực.

Một ví dụ kiểu khác có thể xét trường hợp đo lưu lượng nước trong sông ngòi trong giai đoạn kiệt ổn định khi giá trị lưu lượng không thay đổi trong thời gian đo.

Thường xuyên lặp lại đo đạc ta được tập các đại lượng mà giá trị trung bình sẽ được coi là tham số thống kê của chuỗi trong dạng gần đúng nhất với giá trị thực của lưu lượng nước trong thời đoạn đang xét.

Tính chất nêu trên của trung bình số học được sử dụng như là đánh giá gia nhập khu giữa trên đoạn dông như là hiệu lưu lượng nước đo ở hai trạm thuỷ văn . Với khoảng cách không lớn giữa hai tuyến đo hiệu này được coi là không đáng kể so với sai số đo ddạc lưu lượng nưcớ và do vậy không tin cậy. Để tăng độ tin cậy của các đánh giá tương tự thường đo một số lần lưu lượng nước trong mỗi tuyến đo trong khoảng một thời gian tương đối ngắn, trong giới hạn đó sự thay đổi thực của lưu lượng nước có thể coi là không đáng kể. Khi đó trung bình số học trên mỗi tuyến đo là giá trị xác suất nhất của giá trị chân lý lưu lượng nước còn hiệu giữa chúng như là đại lượng đủ tin cậy của gia nhập khu giữa.

Rõ ràng ở mức độ mà điều kiện hình thành lưu lượng nước không phải là dừng, các kết luận nêu trên không có ý nghĩa. Khi đánh giá khả năng của thủ thuật đã nêu tất nhiên phải hiểu là độ chính xác của trung bình nhận được không thể cao hơn độ chính xác của các đo đạc đưcợ ứng dụng và độ chính bxác của dụng cụ đo.

Tính trung bình số học theo công thức (1.1) hoặc (1.2) thường không gặp khó khăn và cho nên chỉ dẫn ra các kết quả tính toán cuối cùng. Vậy trung bình số học từ chuỗi các lưu lượng nước trung bình năm s. Dnhepr tại Loxmanskaia Kamenka tính theo công thức (1.1) là 1642 m^3/s, còn theo công thức (1.2) :1651 m^3/s; Như đã thấy tính toán trung bình số học theocác công thức trên hầu như trùng nhau vì gắn với dung lượng quan trắc lớn 145 năm.

Trung bình độ cao của bề mặt đầm lầy so với mực nước ngầm tính theo công thức (1.2) là 16,06 cm.

Với dung lượng tính toán lớn hiện nay trung bình số học cũng như các tham số thống kê khác của chuỗi thường tính trên máy tính điện tử. Do độ dài hạn chế của chuỗi quan trắc thuỷ văn không thể tăng theo ý muốn của nhà thuỷ văn bằng cách tiến hành thực nghiệm bổ sung, trong tính toán thuỷ văn thường thực hiện việc dẫn trung bình số học nhận được theo mẫu quan trắc hạn chế về thời đoạn dài. các phương pháp như vậy được trình bày trong chương 6.

Dẫn về thời đoạn dài các giá trị trung bình số học theo chuỗi quan trắc nhiều năm của một đặc trưng thuỷ văn này hoặc kia được gọi là chuẩn.

Nếu trong quá trình hình thành dòng chảy sông ngòi bắt đầu tác động của một nhân tố nào đó chưa được tính đến như hoạt động kinh tế trên lưu vực thì nó cần được tính và được hiệu chỉnh tương ứng trung bình số học của giai đoạn tiếp theo - giai đoạn vận hành công trình.

Các tính chất bổ sung giá trị trung bình số học của mẫu nhận được từ một tập chung nào đó sữ được xét trong chương 5. Ở đây chỉ nhận xét rằng trung bình của chuỗi quan trắc thống kê với tính không thay đổi của các điều kiện hình thành nó và khi tăng số thành viên của mẫu tới trung bình chung của tập, hoặc tiến đến kỳ vọng toán học .

Như vậy, giá trị trung bình số học của chuỗi quan trắc thống kê là tham số mà xung quanh nó thực hiện dao động của chuỗi thống kê đã cho, hoặc như thường nói là tham số trung tâm nhóm số liệu thống kê.

Nói chung, khái niệm kỳ vọng toán học áp dụng trong các giả thiết thuỷ văn là trừu tượng toán học vì chuỗi quan trắc thuỷ văn có độ dài vô hạn không tồn tại. Ngoài ra, xuất phát từ các hình ảnh vật lý hoặc hình ảnh chung của sự hình thành dòng chảy sông ngòi cũng không nên xác định kỳ vọng toán học . Tính điều kiện của thuật ngữ kỳ vọng toán học càng sâu sắc còn bởi trong thiên nhiên nói chung, và trong dao động dòng chảy sông ngòi nói riêng, ta biết được hướng của sự thay đổi. Cho nên nói về kỳ vọng toán học như của dòng chảy năm trong các tính toán thiết kế thường được hiểu là trung bình số học không phai cho thời đoạn vô hạn mà chỉ có hàng chục hoặc hàng trăm năm. Trong trường hợp như thế nói một cách nghiêm túc không nên sử dụng thuật ngữ " kỳ vọng toán học ".

1.5 TRUNG VỊ (SỐ GIỮA)

Sau trung bình số học trung vị là đặc trưng trung tâm nhóm quan trọng tiếp theo, nó bằng giá trị của thành viên chuỗi biến đổi nằm ở vị trí giữa trong trường hợp chuỗi được sắp xếp theo trật tự tăng hoặc giảm dần.

Nếu số thành viên của chuỗi x_i là lẻ và bằng 2m+1 thì trung vị chuỗi này là thành viên x_{m+1} trong chuỗi đã sắp xếp (tăng hoặc giảm dần) của tài liệu quan trắc, tức là:

$$Me = X_{m+1} \tag{1.7}$$

Nếu số thành viên của chuỗi x_i là chẵn, tác là 2m thì trung vị được coi là nhận giá trị trung bình giữa các giá trị trung tâm của đại lượng chuỗi đã sắp xếp:

$$Me = \frac{1}{2}(x_m + x_{m+1}). \tag{1.8}$$

Xác định trung vị theo số liệu thực nghiệm thường không khó, đặc biệt khi chuỗi quan trắc không dài. Thực vậy, để làm điều này cần sắp xếp chuỗi theo trật tự giảm (hoặc tăng) dần và chọn thành viên chính giưã trong trường hợp số thành viên chuỗi lẻ và hai thành viên giưa (tính giá trị trung bình của chúng) nếu chuỗi có số thành viên là chẵn. Thế nên trung vị của chuỗi lưu lượng nước trung bình năm s. Dnhepr tại Loxmanskaia Kamenka sau khi sắp xếp chuỗi này giảm dần là 1620 m^3/s.

Trong trường hợp số liệu quan trắc được nhóm trung vị tính theo công thức gần đúng sau:

$$Me = N_1 - \frac{h\left(\frac{n}{2} - S\right)}{m}, \tag{1.9}$$

với N_1 - điểm cuối của khoản giữa, h - kích thước khoảng, n- số thành viên của chuỗi, S - tần số tích luỹ đến giá trị N_1; m- số trường hợp trong khoảng.

Tính toán trung vị theo công thức đã cho càng chính xác khi phân bố số liệu quan trắc trong khoảng giữa càng đều. Trung vị đối với chuỗi dòng chảy năm s. Dnhepr, tính theo công thức (1.9) sẽ bằng:

$$Me = 1699 - \frac{200\left(\frac{145}{2} - 63\right)}{25} = 1623 m^3/s.$$

Tiến hành tính toán trung vị số liệu nhóm về địa hình đầm lầy đã dẫn trên: N_1=16; h=2; n=903; S=422; m=144. Thế các giá trị này vào công thức (1.9), ta có:

$$Me = 16 - \frac{2\left(\frac{903}{2} - 422\right)}{144} = 15{,}6 cm.$$

Từ các tính toán đưa ra thấy rằng giá trị trung vị, khác với trung bình số học, là được xác định chỉ bởi đại lượng giữa hay là hai giá trị nằm giữa chuỗi phân bố theo trật tự giảm dần mà không phụ thuộc vào các thành viên còn lại của chuỗi.

Nói cách khác là trung vị không thay đổi nếu như giá trị bất kỳ nào của biên nhỏ hơn trung vị thay đổi thoải mái nếu vẫn bảo tồn tính chất nhỏ hơn Me, và mọi thành viên lớn hơn Me cũng thay đổi tuỳ thích trong khoảng lớn hơn Me. Sự thay đổi như vậy chỉ tác động lên giá trị trung bình số học. Tính chất này của trung vị được sử dụng một cách hợp lý hơn so với trung bình số học trong trường hợp khi các thành viên cuối của chuỗi không chính xác và kém tin cậy. Nhưng trung vị so với trung bình số học cũng có nhược điểm là không dễ cho bằng các phép giải tích, chẳng hạn như đối với nó không thể sử dụng định lý cộng.

Vậy khi trộn hai chuỗi với nhau không thể nói gì về trung vị của chuỗi tổng cộng mặc dù đã biết trung vị chuỗi thành phần vì không thể tính được từ chúng.

Đừng vuông góc tại điểm tương ứng với giá trị trung vị với trục của đại lượng biến đổi chia biểu đồ ra hai phàan bằng nhau.

Nhận xét không chứng minh tính chất cơ bản của trung vị là tổng các giá trị độ lệch tuyệt đối các thành viên của chuỗi thống kê với trung vị là cực tiểu so với tổng tương tự được cấu thành từ bất kỳ giá trị nào của chuỗi khác Me.

1.6 SỐ ĐÔNG (MOD)

Số đông được gọi là đại lượng xác suất nhất (thường hay gặp nhất) trong chuỗi thống kê đã cho. Nói cách khác số đông là tung độ lớn nhất của đường cong phân bố trong trường hợp phân bố một đỉnh. Trong trường hợp tổng quát đường cong phân bố có thể có một số đỉnh và tương ứng nó có một và số đông.

Xác định số đông qua giá trị cực đại của đường cong phân bố - là bài toán khá phức tạp, còn với chuỗi quan trắc không lớn thì hầu như không thực hiện được. Giá trị gần đúng của số đông có thể coi là điểm giữa của khoảng có tần số cực đại với số liệu quan trắc được nhóm. Thế nên đối với lưu lượng nước trung bình s. Dnhepr tại Loxmanskaia Kamenka Mo = 1600 m^3/s, còn đối với mặt vi cảnh quan Mo = 19,5 cm.

Đối với phân bố một đỉnh và không quá bất đối xứng số đông có thể tính theo đẳng thức gần đúng của K. Piecson :

$$Mo = \bar{x} + 3(Me - \bar{x}). \qquad (1.10)$$

Giá trị số đông tính theo công thức này đối với lưu lượng nước trung bình năm bằng 1576 m^3/s, đối với bề mặt vi cảnh quanđầm lầy là 26,3 cm.

Sử dụng số đông cúng như trung vị hợp lý khi phân tích các phân bố cực bất đối xứng, khi mà giá trị trung bình tham số đại diện đầy đủ của phân bố và cần được bổ sung bởi trung vị và số đông.

1.7 TRUNG BÌNH NHÂN VÀ TRUNG BÌNH ĐIỀU HOÀ.

Đôi khi vì mục đích nhận được sự phù hợp nhất qui luật phân bố chuỗi thực nghiệm và một vài sơ đồ thống kê (lý thuyết) người ta thực hiệnviệc biến đổi đại lượng tập thực nghiệm. Người ta thường sử dụng logarit hoá đại lượng chuỗi ban đầu cho biến đổi như vậy. Trong trường hợp đó thế vào chỗ các đại lượng ban đầu x_1, x_2, x_3, ..., x_n cấu tạo chuỗi gồm $\lg x_1$, $\lg x_2$, $\lg x_3$, ..., $\lg x_n$. Trung bình số học của chuỗi mới bằng:

$$\lg G = \frac{1}{n}(\lg x_1 + \lg x_2 + \lg x_3 + \ldots + \lg x_n). \qquad (1.11)$$

Hệ thức (1.11) được nhận bằng cách logarit hoá biểu thức ban đầu có dạng:

$$G = \sqrt[n]{x_1 x_2 x_3 \ldots x_n}\,, \tag{1.12}$$

thể hiện là giá trị trung bình nhân của biến x nhận các giá trị dương x_1, x_2, x_3, ..., x_n.

Từ các hệ thức (1.11) và (1.12) suy ra rằng logarit của trung bình nhân (G) bằng trung bình số học của logarit giá trị các đại lượng chuỗi thống kê đang xét.

Như đã chứng minh trong toán học thống kê [111]. trung bình nhân luôn nhỏ hơn trung bình số học.

Một trong các hình thức biến dạng chuỗi ban đầu của tập thống kê người ta biến tập các đại lượng dương x_1, x_2, x_3, ..., x_n thành chuỗi dạng $\frac{1}{x_1}, \frac{1}{x_2}, \frac{1}{x_3}, \ldots, \frac{1}{x_n}$.

Giá trị trung bình số học của chuỗi biến hình trên bằng:

$$\frac{1}{H} = \frac{1}{n}\left(\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_2} + \ldots, \frac{1}{x_n}\right)$$

từ đó:

$$H = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \ldots + \frac{1}{x_n}}\,. \tag{1.13}$$

Đại lượng H với giá trị nghịch đảo của nó bàng trung bình số học các giá trị nghịch đảo của biến x được gọi là trung bình điều hoà của đại lượng x.

Một số ví dụ về biến đổi logarit chuỗi ban đầu và sử dụng trung bình điều hoà trong tính toán thuỷ văn được xét ở chương 2.

Như vậy, giá trị trung bình (giá trị trung bình số học, trung vị, số đông và v.v..) là các đặc trưng mà xung quanh chúng thực hiện việc nhóm các chuỗi biến đổi hoặc thường gọi là các trung tâm nhóm. Giá trị trung bình mô tả một số tính chất quan trọng của tập thống kê nhưng chưa phải là các đặc trưng toàn diện của chúng. Thực vậy có thể tưởng tượng rằng có hai chuỗi đại lượng giá trị trung bình của chúng bằng nhau, còn tính chất phân tán tương đối của chúng khác nhau. Chẳng hạn như giá trị trung bình của mực nước trong một thuỷ vực kín với sóng mạnh và sóng yếu có cùng một trị số. Tuy nhiên tính chất dao động mực nước với nhiều lần đo sẽ khác nhau. Do vậy để mô tả các tập tương tự cần phải xét các đặc trưng mức độ phân tán mà ta sẽ bàn tới ở bài sau.

1.8 CÁC MỨC ĐỘ PHÂN TÁN ĐƠN GIẢN NHẤT.

Thước đo độ phân tán (độ biến đổi) đơn giản nhất của chuỗi thống kê là biên độ hay là hiệu biến đổi trước khí nhóm của số liệu quan trắc :

$$A = x_{max} - x_{min} \tag{1.14}$$

Từ công thức này suy ra đối với việc tính biên độ nhất định phải biết giá trị cực đại và giá trị cực tiểu của chuỗi quan trắc . Vậy biên độ dao động dòng chảy năm s. Dnhepr tại Loxmanskaia Kamenka là A = 3040 - 717 = 2323 m^3/s. còn biên độ dao động bề mặt đầm lầy đang xét A = 32 - 1 = 31 cm.

Với sự tăng độ dài của chuỗi quan trắc biên độ dao động chỉ có thể tăng và mang tính không xác định trong việc sử dụng biên độ như là một đặc trưng phân tán. Đưa hiệu chỉnh trên số thành viên của chuỗi khi tính toán biên độ trong nhiều trường hợp hoàn toàn không làm mất tồn tại đó. Ngoài ra biên độ có dao động ngẫu nhiên lớn từ mẫu này sang mẫu khác và càng làm khó khăn khi sử dụng nó. Tuy có nhược điểm như vậy biên độ cũng được sử dụng trong một số trường hợp tuyến tính thuỷ văn. Chẳng hạn như khi đánh giá tính chuẩn xác của các dự báo thuỷ văn bên cạnh các phương pháp hoàn thiện hơn đôi khi còn sử dụng 20% biên độ. Nếu hiệu của đại lượng dự báo và thực tế nhỏ hơn 1/5 A, thì dự báo coi như thoả mãn, còn ngược lại - không thoả mãn.

Người ta sử dụng độ lệch trung bình tuyệt đối làm đặc trưng mức độ phân tán khác, nó được tính theo công thức :

$$d = \frac{\sum_{i=1}^{n} \left|(x_i - \overline{x})\right|}{n}, \tag{1.15}$$

với n- số thành viên của chuỗi x_i và i từ 1 đến n; $\overline{x}$ - giá trị trung bình số học.

Nhược điểm lớn nhất của độ lệch trung bình tuyệt đối là ở chỗ khi tính toán nó không xét dấu (x_i - $\overline{x}$), gây khó khăn cho sự hoàn thiện sơ đồ tính nó. Sự đóng góp của độ lệch lớn và nhỏ của x_i so với $\overline{x}$ được coi là bằng nhau, điều đó làm giảm giá trị của tham số này như là thước đo mức biến đổi.

Đại lượng độ lệch tuyệt đối trung bình đối với dòng chảy năm s. Dnhepr tại Loxmanskaia Kamenka bằng 360 m^3/s, còn đối vơí bề mặt vi cảnh quan đầm lầy - 3,6cm.

1.9 ĐỘ LỆCH QUÂN PHƯƠNG . PHƯƠNG SAI. HỆ SỐ BIẾN ĐỔI.

Thước đo phân tán của chuỗi thống kê hay sử dụng nhất tương ứng với đại lượng trung bình số học của nó là độ lệch quân phương σ_{x_i} , hay là chuẩn:

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}} \qquad (1.16)$$

Độ lệch quân phương bảo lưu thứ nguyên của chuỗi quan trắc gốc.

Trong trường hợp sử dụng bảng số liệu nhóm độ lệch quân phương có thể tính theo công thức :

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^{k} n_i (x_i - \overline{x})^2}{n}} \qquad (1.17)$$

với n_i - tần số tuyệt đối chuỗi thống kê trong đoạn thứ i.

Bình phương của độ lệch quân phương gọi là phương sai.

Trong nhiều trường hợp rất có ích khi tính độ lệch quân phương trên cơ sở phương pháp sai phân:

$$\sigma_*^2 = \frac{\sum_{i=1}^{n-1}(x_i - x_{i+1})^2}{2(n-1)}. \qquad (1.18)$$

Tính toán theo công thức này không đòi hỏi phải tính trước giá trị trung bình số học. Cần nhận thấy rằng độ mạnh của độ lệch quân phương tính theo công thức (1.18) chỉ bằng khoảng 2/3 độ lệch quân phương tính theo (1.16). Cho nên công thức (1.18) thường ít được sử dụng . Tuy nhiên nó có thể hữu ích khi mà trong chuỗi gốc xuất hiện dao động tuần hoàn hoặc dao động có hướng. Trong các trường hợp như vậy, trong giá trị độ lệch quân phương tính theo (1.18) loại trừ được dao động của giá trị trung bình , trong khi đó nếu tính theo (1.16) các dao động của giá trị trung bình còn nằm trong độ lệch quân phương.

Xét các tính chất chung của phương sai

1. Nếu tổng bình phương độ lệch của chuỗi quan trắc x_i cạnh giá trị a gọi là x_i gần a

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - a)^2 = \sigma_a^2 ,$$

thì phương sai σ_a^2 đạt cực tiểu khi $a = \overline{x}$. Tính chất này của phương sai được xét khi mô tả tính chất của trung bình số học.

2. Nếu một đại lượng y_i nào đó liên quan với x_i bởi phương trình dạng $y_i = ax_i + b$, với a và b - các đại lượng hằng số thì $\sigma_y^2 = a^2\sigma_x^2$, hay nói cách khác:

$$\sigma_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - b)^2 - (\overline{x} - b)^2$$

Đẳng thức này thường được sử dụng khi tính phương sai. Khi đó các hằng số được chọn sao cho hiệu $(x_i - b)$ là cực tiểu , tức là để thuận lợi cho việc tính.

3. Xét một tính chất rất quan trọng cộng phương sai thường hay sử dụng trong tính toán thuỷ văn và khi xử lý các vấn đề lý thuyết khác nhau của thống kê.

Phương sai chung của tổng k chuỗi quan trắc của đại lượng ngẫu nhiên x bằng trung bình số học của các phương sai thành phần, cấu tạo bởi phương sai của trung bình thành phần cạnh trung bình tổng cộng.

$$\sigma^2 = \frac{\sum_{h=1}^{k}\sigma_h^2}{k} + \frac{\sum_{h=1}^{k}(\overline{x}_h - \overline{x})^2}{k} , \qquad (1.19)$$

với:

$$
\begin{matrix}
x_{11} & x_{12} & \dots & x_{1i} & \dots & x_{1n_1} \\
x_{21} & x_{22} & \dots & x_{2i} & \dots & x_{2n_2} \\
. & . & . & . & . & . \\
. & . & . & . & . & . \\
. & . & . & . & . & . \\
x_{h1} & x_{h2} & \dots & x_{hi} & \dots & x_{hn_h} \\
. & . & . & . & . & . \\
. & . & . & . & . & . \\
. & . & . & . & . & . \\
x_{k1} & x_{k2} & \dots & x_{ki} & \dots & x_{kn_k}
\end{matrix}
.
$$

- các chuỗi quan trắc ban đầu;

$$\overline{x}_1 = \frac{\sum_{i=1}^{n_1} x_{1i}}{n_1};\ \overline{x}_2 = \frac{\sum_{i=1}^{n_2} x_{2i}}{n_2};\ \dots;\ \overline{x}_k = \frac{\sum_{i=1}^{n_k} x_{ki}}{n_k};$$

- trung bình số học của các tập thành phần:

$$\sigma_1^2 = \frac{\sum_{i=1}^{n_1} (x_{1i} - \overline{x}_1)^2}{n_1};\ \sigma_2^2 = \frac{\sum_{i=1}^{n_2} (x_{2i} - \overline{x}_2)^2}{n_2};\ \dots;\ \sigma_k^2 = \frac{\sum_{i=1}^{n_k} (x_{ki} - \overline{x}_k)^2}{n_k}$$

- độ lệch quân phương của các tập thành phần; n_1, n_2, ..., n_k - dung lượng các tập thành phần; $\overline{x}$ - trung bình chung cho tất cả quan trắc ; $\sigma = \frac{\sum_{j=1}^{n} (x_j - \overline{x})^2}{n}$ - phương sai chung cho tất cả quan trắc, với j bao cả i=1 đến $i=n_h$ và từ h=1 đến h=k; dung lượng chung tất cả số liệu bằng $n = \sum_{h=1}^{k} n_h$.

Tính chất cộng phương sai có ứng dụng thực tiễn khi một số phân bố được trộn vào một hoặc ngược alị khi một phân bố được phân ra nhiều phần, đối với chúng tính riêng $\overline{x}_h$ và σ_h^2, chẳng hạn như trong tính toán thuỷ văn đôi khi người ta gộp vào một phân bố đặc trưng thuỷ văn này hoặc kia nhận được theo các sông riêng biệt (tham số dòng chảy cực đại hoặc dòng chảy năm). Trong trường hợp như vậy, phương sai chung của chuỗi tổng cộng có thể được tính theo các phương sai đã biết và trung bình của chuỗi nhân tạo với việc sử dụng tính chất đã nêu của phép cộng phương sai.

Một số tính chất bổ sung của phương sai mẫu sẽ được xét ở chương 5. Ở đây chỉ ra mà không chứng minh rằng phương sai mẫu khác biệt càng ít càng tốt với phương sai tập tổng nếu số thành viên quan trắc đủ lớn. Trong tính toán thực tế phương sai theo chuỗi quan trắc cần sử dụng công thức :

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}, \qquad (1.20)$$

công thức này cúng đưa ra mà không chứng minh và nó cho xấp xỉ tốt nhất của phương sai mẫu với phương sai tập tổng, không có sai số hệ thống. Nhận thấy rằng, với n > 30 hiệu đính hằng số dịch chuyển σ^2 (thế vào n ở mẫu số là n-1) không lớn lắm. tuy nhiên, khi xét đến độ phức tạp của hiệu chỉnh do dịch chuyển σ^2 sử dụng công thức (1.20) chắc chắn hơn (1.16)-(1.18).

Trong tính toán thuỷ văn thường xuất hiện nhu cầu so sánh tính biến đổi của các chuỗi cấu tạo từ các đặc trưng thuỷ văn rất khác nhau theo giá trị . Trong trường hợp này độ lệch quân phương của các chuỗi đang xét là không so sánh được. Chẳng hạn như độ lệch quân phương của chuỗi lưu lượng cực đại trong năm của s. Dnhepr tại Kiev và s. Sor ở Slavgorod sẽ rất khác nhau dù rằng vì đại lượng tuyệt đối của các chuỗi đó rất khác nhau. Việc so sánh độ biến đổi của các chuỗi như vậy thực hiện được nhờ hệ số biến đổi (C_v) là tỷ số của độ lệch quân phương và giá trị trung bình của chuỗi:

$$C_{vx} = \frac{\sigma_x}{\bar{x}} \qquad (1.21)$$

Hệ số biến đổi là đặc trưng không thứ nguyên của độ biến đổi chuỗi thống kê . Đôi khi hệ số biến đổi được biểu diễn bằng phần trăm với đại lượng trung bình $C_{vx} = \frac{\sigma_x}{\bar{x}} 100\%$.

Trong tính toán thuỷ văn hệ số biến đổi hay được xác định theo công thức:

$$C_{vx} = \sqrt{\frac{\sum_{i=1}^{n}(k_i - 1)^2}{n-1}} \qquad (1.22)$$

với $k = x_i/\overline{x}$ - hệ số mô đun.

Biểu thức (1.22) dễ dàng nhận được từ các biến đổi đơn giản sau:

$$C_{vx} = \frac{\sigma_x}{\overline{x}} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{(n-1)\overline{x}^2}} = \sqrt{\frac{\sum_{i=1}^{n}\left(\frac{x_i - \overline{x}}{\overline{x}}\right)^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n}\left(\frac{x_i}{\overline{x}} - 1\right)^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n}(k_i - 1)^2}{n-1}}$$

Trong trường hợp số liệu quan trắc nhóm biểu thức (1.22) có dạng:

$$C_{vx} = \sqrt{\frac{\sum_{i=1}^{m} n_i (k_i - 1)^2}{n-1}}$$

với n_i - tần số tuyệt đối của phân cấp thứ i.

Giữa độ lệch trung bình tuyệt đối (d) và độ lệch quân phương (σ) trong trường hợp nếu như chuỗi thống kê tuân theo qui luật phân bố chuẩn, đẳng thức sau được sử dụng :

$$d = \sqrt{\frac{2}{\pi}}\sigma = 0{,}8\sigma \qquad (1.24)$$

Hệ thức này có thể sử dụng để ước lượng gần đúng σ khi biết độ lệch tuyệt đối. Tuy nhiên sai số ước lượng này có thể lớn nếu chuỗi thống kê đang xét lệch nhiều so với phân bố chuẩn.

Một đặc trưng quan trọng tuy ít dùng hơn (so với phương sai và chuẩn) của độ phân tán chuỗi thống kê tương ứng trị trung bình của nó là độ lệch xác suất (Δ), xác định theo đẳng thức:

$$\Delta = \frac{x_{25} - x_{75}}{2} \qquad (1.25)$$

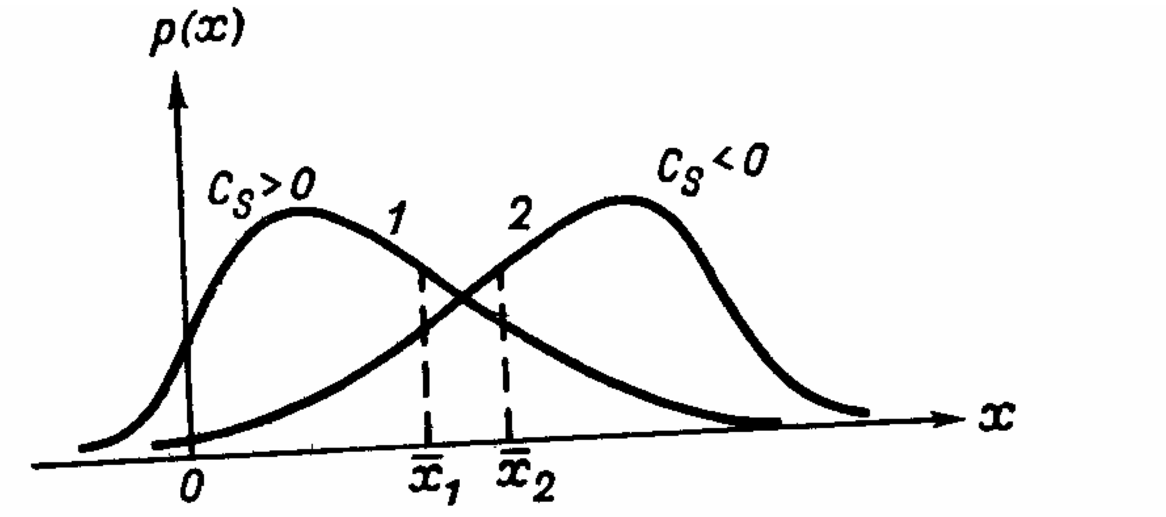
với x_{25} và x_{75} - đoạn tứ phân thứ nhất và thứ ba.

Rõ ràng, giữa x_{25} và x_{75} gồm một nửa thành viên tập thống kê . Nói cách khác, giá trị Δ, phân bố về hai phía của trung vị phân bố, xác định biên mà trong giới hạn của nó gồm một nửa thành viên chuỗi thống kê. Cho nên đối với một thành viên lấy ngẫu nhiên của tập thống kê xác suất có mặt trong hoặc ngoài đoạn đó. Rõ ràng rằng biên đó bố trí càng gần trung vị thì độ phân tán chuỗi thống kê càng ít hơn. Cho nên Δ được nhận làm thước đo độ phân tán biến của phân bố. Đối với điều kiện phân bố chuẩn Δ gắn với σ bằng hệ thức không đổi:

$$\Delta = 0{,}674\sigma.$$

1.10 BẤT ĐỐI XỨNG VÀ ĐỘ NHỌN

Giá trị trung bình xác định các trung tâm nhóm của các tập thống kê cũng như các thước đo phân tán khác nhau xét trong các phần trên vẫn chưa mô tả đầy đủ các tính chất cơ bản nhất của chuỗi thống kê.



H.1.4 Đường cong phân bố xác suất với bất đối xứng dương (1) và âm (2).

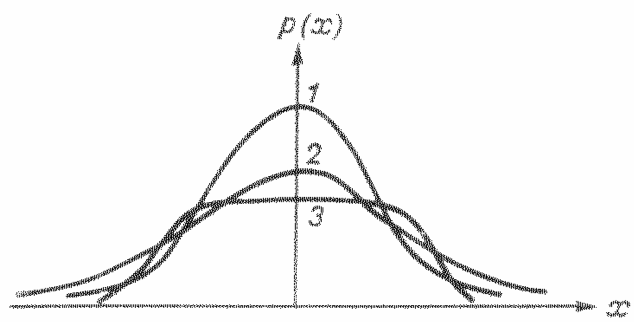
Cụ thể là một nét đặc trưng quan trọng của phân bố là tính bất đối xứng của nó, tức là sự phân tán khác nhau các thành viên của chuỗi so với trị trung bình hoặc số đông. Nguyên do là giữa các tập thống kê có thể quan sát thấy các tập có giá trị trung bình số học và phương sai bằng nhau, tuy nhiên tuy nhiên nhóm tương ứng với trung tâm phân bố của đại lượng cấu tạo nên tập đó có thể hoặc đối xứng hoặc bất đối xứng.

Coi đặc trưng đối xứng (bất đối xứng) của chuỗi thống kê là giá trị trung bình của lập phương độ lệch các thành viên chuỗi với giá trị trung bình số học:

$$\mu_3 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3. \tag{1.26}$$

Khi các thành viên chuỗi phân bố đối xứng so với trị trung bình, các giá trị về phía dương và âm cách đều trị trung bình bằng nhau. Trong các trường hợp khác có thể độ lệch dương (năm nhiều nước) coá thể lặp nhiều hơn và giá trị thường lặp nhất (số đông) thường nhỏ hơn giá trị trung bình . Đây là trường hợp bất đối xứng dương. (h.1.4, đường cong 1). Trong trường hợp ngược lại, ta có bất đối xứng âm. (h.1.4. đường cong 2).

Đối với đường cong (1) đặc trưng bởi đỉnh lớn hơn giá trị trung bình , đối với



H.1.5 Sơ đồ minh hoạ ảnh hưởng của độ nhọn đến dạng của đường cong phân bố.

(2) thì có hệ thức ngược lại.

Nếu chuỗi là đối xứng, thì độ lạch lập phương sẽ có dấu khác nhau và tổng của chúng sẽ bằng 0.

Với bất đối xứng dương, dãy gồm một loạt các thành viên dương có giá trị lớn và một loạt thành viên khác có giá trị bé, khi lập phương độ lệch dương đó lại càng tăng nhiều hơn so với độ lệch âm và tổng lập phương độ lệch sẽ dương.

Với bất đối xứng âm, kết quả như vậy sẽ cho dấu âm.

Để nhận được biểu thức vô thứ nguyên đối với đặc trưng bất đối xứng của chuỗi lấy giá trị trung bình của lập phương chia cho lập phương của độ lệch quân phương. Thương số này gọi là hệ số bất đối xứng.

$$C_s = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^3}{n\sigma^3}, \qquad (1.27)$$

Thay $\sigma = C_v \, \overline{x}$, ta nhận được:

$$C_s = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^3}{nC_v^3\overline{x}^3} = \frac{\sum_{i=1}^{n}\left(\frac{x_i - \overline{x}}{\overline{x}}\right)^3}{nC_v^3} = \frac{\sum_{i=1}^{n}(k_i - 1)^3}{nC_v^3}. \tag{1.28}$$

Biểu tức (1.28) thường được sử dụng trong thực tiễn tính toán thuỷ văn khi xác định các tham số của chuỗi thống kê.

Nhận thấy từ công thức đã dẫn, hệ số bất đối xứng là đại lượng vô thứ nguyên.

Thỉnh thoảng người ta coi đặc trưng bất đối xứng của đường cong phân bố và hệ số nén (S), biểu diễn qua tứ đoạn phân của nó bởi hệ thức:

$$S = \frac{x_p + x_{100-p} - 2x_{50}}{x_p - x_{100-p}}, \tag{1.29}$$

với x_p, x_{100-p}, - tung độ đường cong đảm bảo (tứ đoạn phân) phân bố trên các khoảng cách bằng nhau (theo suất đảm bảo) từ trung tâm (trung vị) của phân bố (x_{50}).

Khi nghiên cứu các tính chất chung của chuỗi thống kê đôi khi xuất hiện việc cần thiết phải so sánh mức độ là trơn hay độ nhọn của đồ thị phân bố xác suất của tập đang xét với đường cong phân bố chuẩn. Người ta nhận tham số độ nhọn (độ góc) làm đặc trưng như vậy cho tính chất nêu trên của đường cong phân bố.

$$\varepsilon = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^4}{\left[\sum_{i=1}^{n}(x_i - \overline{x})^2\right]^2} - 3. \tag{1.30}$$

Ứng dụng với qui luật phân bố chuẩn, thương số nằm trong biểu thức độ nhọn abừng 3 và do vậy độ nhọn sẽ bằng 0. Độ lệch của thương số này và 3 chứng tỏ sự khác biệt về mức độ nhọn (hay làm trơn) của đường cong đang xét cuả chuỗi tjk với qui luật phân bố chuẩn (h.1.5)

1 - $\varepsilon > 0$; 2 - $\varepsilon = 0$; 3 - $\varepsilon < 0$;

Trong thực tiễn tính toán đại lượng độ nhọn khi xử lý các đại lượng thuỷ văn thường không được sử dụng vì tham số này xác định theo chuỗi chỉ có vài chục thành viên hoàn toàn không chính xác.

1. 11 MÔMEN CÁC TẬP (CHUỖI) THỐNG KÊ.

Trong các bài trước đã xét các tính chất cơ bản nhất của chuỗi thống kê - trị trung bình số học (kỳ vọng toán học) , thước đo độ phân tán (phương sai , độ lệch quân phương và v.v..), mức độ đối xứng (độ bất đối xứng) và độ nhọn được thể hiện độc lập và đẳng hướng với nhau. Giữa chúng còn có mối tương quan liên hệ cấu tạo nên một hệ thống nhất dựa trên khái niệm về mômen các chuỗi thống kê. Khái niệm mômen được đưa vào thống kê trên cơ sở từ cơ học, nơi sử dụng nó để đặc trưng cho phân bố khối lượng (mômen tĩnh, mômen quán tính v.v..).

Khi mô tả các tính chất của tập thống kê thường hay sử dụng mômen dưới hai dạng: mômen gốc và mômen trung tâm.

Mômen gốc bậc k của đại lượng ngẫu nhiên rời rạc x là biểu thức dạng:

$$\alpha_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k \qquad (1.31)$$

Như vậy, mômen gốc bậc k của đại lượng ngẫu nhiên x là kỳ vọng toán học 52m_x) hay nói cách khác là giá trị trung bình ($\overline{x}$) bậc k của đại lượng ngẫu nhiên này.

Cụ thể, với k = 1 có mômen gốc bậc 1hay là kỳ vọng toán học (giá trị trung bình số học) của biến đang xét .

Mô men trung tâm (μ) bậc k của đại lượng ngẫu nhiên x là giá trị trung bình của độ lệch đại lượng ngẫu nhiên x với trung bình của nó (kỳ vọng toán học m_x) bậc k, tức là:

$$\mu_k = \frac{1}{n}\sum_{i=1}^{n} (x - \overline{x})^k \qquad (1.32)$$

Rõ ràng, với mômen trung tâm bậc 0, cũng như mômen gốc bậc 0 bằng 1, còn mômen trung tâm bậc nhất bằng 0.

Hiệu giữa các giá trị đại lượng ngẫu nhiên x và kỳ vọng toán học (giá trị trung bình) trong thống kê gọi là đại lượng ngẫu nhiên trung tâm.

$$\overset{o}{x} = x - \overline{x}$$

Sử dụng khái niệm này cho phép xác định mômen trung tâm bậc k của đại lượng ngẫu nhiên x ngắn gọn hơn, như là kỳ vọng toán học (trị trung bình) của đại lượng trung tâm luỹ thừa bậc k.

$$\mu_k = \frac{1}{n}\sum_{i=1}^{n} \overset{o}{x}^{k} \qquad (1.33)$$

Hệ thức giữa các tham số thống kê cơ bản đã xét trước với các mômen tập thống kê được dẫn ở bảng 1.4.

Bảng 1.4 Biểu thức các tham số $\bar{x}$, C_v , C_s qua các mômen .

Tham số của tập thống kê	Biểu thức tham số qua mômen
Giá trị trung bình số học $\bar{x}$ Kỳ vọng toán học m_x	Mômen gốc bậc nhất α_1
Phương sai σ^2	Mômen trung tâm bậc hai μ_2
Độ lệch quân phương (chuẩn) σ	Căn bậc hai của mômen trung tâm $\sqrt{\mu_2}$
Hệ số biến đổi C_v	Căn bậc hai của mômen trung tâm chia cho mômen gốc bậc nhất $\frac{\sqrt{\mu_2}}{\alpha_1}$
Hệ số bất đối xứng C_s	Mômen trung tâm bậc ba chia cho lập phương độ lệch quân phương $\frac{\mu_3}{\sigma^3} = \frac{\mu_3}{\mu_2^{3/2}}$
Độ nhọn ε	Thương số (Mômen trung tâm bậc bốn chia cho độ lệch quân phương luỹ thừa bốn) trừ đi ba $\frac{\mu_4}{\sigma^4} - 3 = \frac{\mu_4}{\mu_2^2} - 3$

Ở dạng chung nhất mômen có thể được xem không chỉ tương ứng với gốc toạ độ (mômen gốc) hay hay kỳ vọng toán học (mômen trung tâm) mà có thể là một điểm a bất kỳ:

$$\gamma_k = \frac{1}{n}\sum_{i=1}^{n}(x_i - a)^k \qquad (1.34)$$

Rõ ràng, với a = 0 biểu thức (1.34) trùng với mômen gốc , còn $a = \bar{x}$ trùng với khái niệm mômen trung tâm. Đôi khi với các tính toán thực nghiệm nhưng chủ yếu là với phân tích lý thuyết qui luật phân bố các tập thống kê mômen trung tâm cần được biểu diễn qua mômen gốc hoặc mômen so với điểm a bất kỳ. Biểu thức gắn mômen trung tâm bậc k (μ_k) với mômen k tương ứng gốc bất kỳ a (γ_k), để dàng nhận được từ biểu thức sau:

$$\mu_k = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^k = \frac{1}{n}\sum_{i=1}^{n}\left[(x_i - a) - (\bar{x} - a)\right]^k = \frac{1}{n}\sum_{i=1}^{n}[(x_i - a)^k - C_k^1 (x_i - a)^{k-1}(\bar{x} - a)$$
$$+ C_k^2 (x_i - a)^{k-2}(\bar{x} - a)^2 - C_k^3 (x_i - a)^{k-3}(\bar{x} - a)^3 + \ldots + (-1)^S C_k^S (x_i - a)^{k-S}(\bar{x} - a)^S + \ldots +$$
$$+ (-1)^{k-1} C_k^{k-1}(x_i - a)(\bar{x} - a)^{k-1} + (-1)^k (\bar{x} - a)^k] \qquad (1.35)$$

với C_k^S - số kết hợp từ k theo S (S = 1, 2, 3, ... , k).

Khai triển ngoặc vuông từ phương trình (1.35) và biểu thức đối với mômen bậc nhất tương ứng điểm a:

$$\gamma_1 = \frac{1}{n}\sum_{i=1}^{n}(x_i - a) = \frac{1}{n}\sum_{i=1}^{n}x_i - \frac{1}{n}\sum_{i=1}^{n}a = \bar{x} - a,$$

Cuối cùng ta thu được:

$$\mu_k = \gamma_k - C_k^1\gamma_{k-1}\gamma_1 + C_k^2\gamma_{k-2}\gamma_1^2 - C_k^3\gamma_{k-3}\gamma_1^3 + \ldots + (-1)^{k-1}C_k^{k-1}\gamma_{k-2}\gamma_1^{k-1} + (-1)^k\gamma_1^k. \qquad (1.36)$$

Khi nghiên cứu tập thống kê các đại lượng thuỷ văn mômen cao hơn bậc bốn, và thực hiện các tính toán thực nghiệm mômen cao hơn bậc ba thường không được sử dụng bởi sai số tính toán các mômen này rất lớn. Sự xuất hiện sai số đáng kể khi xác định mômen từ bậc ba trở đi là hệ quả của chuỗi quan trắc về các đặc trưng thuỷ văn không dài.

Sử dụng biểu thức chung (1.36) để nhận công thức gắn bốn mômen trung tâm đầu tiên (μ) với mômen tương ứng điểm bất kỳ a - (γ). Khi đó cần tính tới:

$$\mu_0 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^0 = \frac{n}{n} = 1,$$

$$\mu_1 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^1 = 0.$$

Khi đó:

$$\mu_2 = \gamma_2 - C_2^1\gamma_1\gamma_1 + C_2^2\gamma_0\gamma_1^2 = \gamma_2 - 2\gamma_1^2 + \gamma_1^2 = \gamma_2 - \gamma_1^2,$$

$$\mu_3 = \gamma_3 - C_3^1\gamma_2\gamma_1 + C_3^2\gamma_1\gamma_1^2 - C_3^3\gamma_0\gamma_1^3 = \gamma_3 - 3\gamma_2\gamma_1 + 3\gamma_1^3 - \gamma_{13}^3 = \gamma_3 - 3\gamma_2\gamma_1 + 2\gamma_1^3,$$

$$\mu_4 = \gamma_4 - C_4^1\gamma_3\gamma_1 + C_4^2\gamma_2\gamma_1^2 - C_4^3\gamma_1\gamma_1^3 + C_4^4\gamma_0\gamma_1^4 = \gamma_4 - 4\gamma_3\gamma_1 + 6\gamma_2\gamma_1^2 - 3\gamma_1^4.$$

Tương tự , có thể xây dựng biểu thức chung để gắn các mômen trung tâm (μ) và gốc (α), nó có dạng:

$$\mu_k = \alpha_k - C_k^1\alpha_{k-1}\alpha_1 + C_k^2\alpha_{k-2}\alpha_1^2 - C_k^3\alpha_{k-3}\alpha_1^3 + \ldots + (-1)^{k-1}C_k^{k-1}\alpha_1^{k-1} + (-1)^k\alpha_1^k \quad (1.37)$$

Biểu thức chung này đối với bốn mômen trung tâm đầu tiên viết dưới dạng:

$\mu_0 = 1$

$\mu_1 = 0$

$\mu_2 = \alpha_2 - \alpha_1^2$

$\mu_3 = \alpha_3 - 3\alpha_2\alpha_1 + 2\alpha_1^3$

$\mu_4 = \alpha_4 - 4\alpha_3\alpha_1 + 6\alpha_2\alpha_1^2 - 3\alpha_1^4$ (1.38)

Tiếp theo, đôi khi để hoà đồng chỉ số sử dụng trong thuỷ văn có thể ký hiệu mômen bằng chữ cái m.

Khi xác định lưu lượng nước trung bình năm nhận được với các đoạn đường quá trình dòng chảy khác nhau , các tham số phân bố (mômen mẫu) có thay đổi chút ít. Vấn đề này được G.G. Svanhide và A.N. Kilasonia [127, 128] nghiên cứu khá chi tiết.