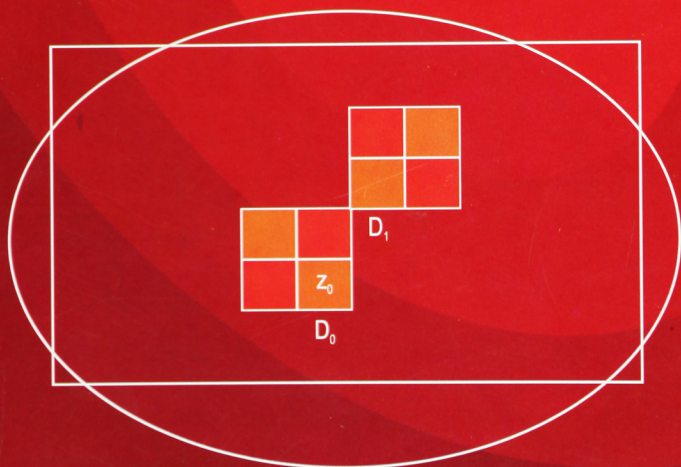




CK.0000068509

PGS. TS. NGUYỄN DOÃN Ý

QUY HOẠCH VÀ XỬ LÝ SỐ LIỆU THỰC NGHIỆM



NGUYỄN
HỌC LIỆU

5

NHÀ XUẤT BẢN XÂY DỰNG



PGS. TS. NGUYỄN DOÃN Ý

QUY HOẠCH VÀ XỬ LÝ SỐ LIỆU THỰC NGHIỆM

(Tái bản)



NHÀ XUẤT BẢN XÂY DỰNG
HÀ NỘI - 2010

LỜI NÓI ĐẦU

Trong khi giải quyết các vấn đề kỹ thuật, vấn đề đặt ra trước hết là cần phải có kết quả mô tả mối quan hệ phụ thuộc lẫn nhau của các yếu tố liên quan đảm bảo độ tin cậy và tính chính xác của các kết quả thực nghiệm với sai số cho phép.

Trong điều kiện đủ thông tin, có thể sử dụng các phương pháp toán học khác nhau để mô tả hệ thống, từ đo khảo sát và tìm cực trị đối với bài toán. Tuy nhiên khi thiếu thông tin, cần phải làm thực nghiệm để xây dựng mô hình toán học cho qua trình đo, sau đó tiến hành các bước tối ưu hóa.

Quy hoạch và xử lý số liệu thực nghiệm là một phương pháp toán học được sử dụng rộng rãi trong nghiên cứu, học tập và cả trong sản xuất thực tế.

Lý thuyết cơ bản được phát triển dựa vào các lý thuyết sau:

- Lý thuyết quy hoạch và phương pháp thực nghiệm.
- Lý thuyết hệ thống.
- Lý thuyết thống kê.
- Lý thuyết tối ưu hóa và ứng dụng.

Nhằm chọn một chiến lược tối ưu, trong điều kiện chưa hiểu một cách toàn diện một quá trình nào đó tác động vào quá trình tiến hành thực nghiệm, đồng thời phải thu được:

- Các số liệu cần thiết nhất.
- Số lượng thí nghiệm ít nhất.
- Độ tin cậy đặt ra trước.
- Với công thức toán học đơn giản nhất.

Và cuối cùng phải đạt được kết quả với hiệu quả kinh tế, kỹ thuật cao nhất.

Việc xử lý đánh giá, phân tích các kết quả thực nghiệm ngày càng cần thiết đối với học tập, nghiên cứu của sinh viên năm cuối, học viên cao học, nghiên cứu sinh và các nhà nghiên cứu.

Nội dung trình bày trong sách này không những nhằm thỏa mãn một phần yêu cầu của các nhà kỹ thuật nói chung, mà còn rất thuận lợi cho các nhà nghiên cứu trên các lĩnh vực khác nhau như: xã hội, giáo dục, kinh tế... khi cần tiến hành và xử lý các kết quả thực nghiệm.

Chương I

MỞ ĐẦU

1.1. MỞ ĐẦU

Quy hoạch và xử lý số liệu thực nghiệm, là một khoa học được sử dụng rộng rãi trong hầu hết các ngành kỹ thuật, kinh tế và xã hội. Trong thực tế, các kết quả thí nghiệm, thực nghiệm được đo đạc, thống kê luôn được thực hiện hàng ngày; Vấn đề đặt ra là phải đo đạc, thu nhận, thống kê các số liệu này như thế nào đạt được mức tối ưu nhất?

Vì vậy quy hoạch và xử lý số liệu thực nghiệm là một chiến lược, chiến thuật nhằm tác động toàn diện, tối ưu, đơn giản nhất vào các qua trình có tương quan với nhau, nhờ đó tiến hành thực nghiệm thu thập số liệu, đánh giá, xử lý số liệu đó, mô tả chúng bằng một mô hình toán học. Và có thể đề ra các phương án kỹ thuật, công nghệ, hoặc phát triển nội dung nghiên cứu khoa học, các giải pháp về kinh tế, xã hội một cách tối nhất.

Tùy theo mức độ hiểu biết các mối tương quan của quá trình; các cơ sở lý thuyết chỉ có giá trị định hướng ban đầu, hỗ trợ rút ngắn thời gian trong nghiên cứu thực nghiệm. Ngược lại, thực nghiệm làm sáng tỏ, bổ sung cho nghiên cứu lý thuyết.

Một quá trình càng phức tạp, càng có nhiều yếu tố phụ thuộc vào nhau, thì hiệu quả của quy hoạch và xử lý số liệu thực nghiệm càng cao. Vì vậy có thể nói quy hoạch và xử lý số liệu thực nghiệm là một ngành khoa học kết hợp một cách hợp lý nhất giữa nghiên cứu lý thuyết và nghiên cứu thực nghiệm.

Quy hoạch và xử lý số liệu thực nghiệm là môn học không thể thiếu đối với các sinh viên các ngành kỹ thuật và ngay cả với các ngành xã hội, đặc biệt đối với các học viên cao cấp, các nghiên cứu sinh các nhà nghiên cứu, những người mà thường xuyên phải giải quyết mối tương quan của mọi quá trình tự nhiên, xã hội, công nghệ...

Quy hoạch và xử lý số liệu thực nghiệm ngày càng phát triển không ngừng và thường xuyên được các nhà khoa học hoàn thiện. Theo quan điểm kinh tế, người ta mong muốn càng ít thực nghiệm càng tốt, trong khi đó lại phải đảm bảo đủ độ tin cậy chất lượng của mô hình tương quan và có thể tối ưu được qua trình đó.

Quy hoạch và xử lý số liệu thực nghiệm nhằm đem lại những ưu điểm cụ thể sau:

- Giảm đáng kể số lượng thí nghiệm đến mức nhỏ nhất.
- Mô hình toán học mô tả quá trình là đơn giản nhất
- Giảm thời gian, chi phí vật chất thấp nhất
- Lượng thông tin đầy đủ để đánh giá, tìm cực trị và với độ tin cậy cần thiết đặt ra trước.

1.2. NỘI DUNG CỦA QUY HOẠCH VÀ XỬ LÝ SỐ LIỆU THỰC NGHIỆM

Khi nghiên cứu mối quan hệ phụ thuộc các yếu tố với nhau, thường chưa biết, chưa rõ quy luật hoạt động của các mối quan hệ bên trong giữa các yếu tố. Thí dụ: Mòn chi tiết máy do nhiều yếu tố tác động lên bề mặt chi tiết, nhưng không rõ các yếu tố đó đã biến đổi cơ, lý, hóa trên bề mặt kim loại như thế nào? Ta chỉ cần biết với các yếu tố làm việc nhất định, và quan trắc, đo đạc được lượng mòn của chi tiết.

Như vậy ta coi những biến đổi đó như một hộp đen và mô tả bởi sơ đồ sau (hình 1).



Hình 1:

Trong đó:

X - các thông số vào, thông số điều khiển được hay còn gọi là thông số điều khiển;

Y - các thông số ra, là biến bị điều khiển;

Trường hợp có nhiều thông số ra, có thể xét véc tơ biến ra $y = (y_1, y_2, \dots, y_p)$. Nhưng cũng có thể xét riêng sự phụ thuộc từng thành phần, sau đó tổng hợp lại.

ξ - biến ngẫu nhiên (do tác động ngẫu nhiên, biến không điều khiển được). Thường giả thiết chúng có phương sai:

$$D(\xi) = \sigma^2 \text{ và kỳ vọng } E(\xi) = 0.$$

tức là chúng có phân phối: $\xi = N(0, \sigma^2)$.

Với một bộ n thí nghiệm và có k thông số vào thì ma trận các thông số vào X có thể viết như sau:

$$X = (X_{ij})_{n \times k} = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1k} \\ X_{21} & X_{22} & \dots & X_{2k} \\ \dots & \dots & \dots & \dots \\ X_{n1} & X_{n2} & \dots & X_{nk} \end{bmatrix} \quad (1.1)$$

Ta phải mô tả mối quan hệ giữa biến vào và biến ra dưới ảnh hưởng của biến ngẫu nhiên, có thể dự đoán hàm số mô tả mối quan hệ đó là một trong các hàm số hay gặp hoặc tổ hợp một vài hàm số có dạng đã biết:

$$y = \varphi(x_1, x_2, \dots, x_k) + \xi$$

$$y = \varphi(x) + \xi \quad (1.2)$$

Mỗi lần thực hiện một thí nghiệm, tức là tác động một lần các thông số vào x_{ik} có kết quả đầu ra sao cho:

$$y_i = \varphi_i(x_{i1}, x_{i2}, \dots, x_{ik}) + \xi \quad (1.3)$$

Với giả thiết $\xi = N(0, \sigma^2)$; sau khi loại trừ ảnh hưởng nhiễu, ta nhận được mô hình thống kê của hệ.

- Nếu mô hình thống kê chứa yếu tố thời gian là mô hình động, không chứa thời gian là mô hình tĩnh.
- Nếu mô hình chỉ áp dụng cho giới hạn nhất định của biến vào x_i gọi là mô hình địa phương.

Với mô hình địa phương, có thể khai triển Taylor đối với hàm $\varphi(x_1, \dots, x_k)$:

$$\varphi(x_1, \dots, x_k) = a_0 + \sum_{i=1}^k a_i x_i + \sum_{i=1, j=1}^{i, j, k} a_{ij} x_i x_j + \dots \quad (1.4)$$

Dựa vào các kết quả thực nghiệm, xác định các giá trị a_0, a_1, \dots được coi là nhận dạng mô hình thống kê đó. Phương trình nhận được, gọi là phương trình hồi quy thực nghiệm của hệ thống tương ứng với bộ thí nghiệm đã cho, ký hiệu là:

$$\varphi(x_1, \dots, x_k) = \sum_{i=1}^n \hat{a}_i f_i(x_1, \dots, x_k) \quad (1.5)$$

Phương trình hồi quy thực nghiệm phụ thuộc vào bộ n thí nghiệm, phương pháp nhận dạng mô hình thống kê.

Như vậy cần phải có chiến lược tác động vào các yếu tố vào, xây dựng bộ n thí nghiệm, sao cho mô hình thu được đạt: độ tin cậy đạt ra, đủ thông tin cần thiết, thuận tiện xử lý thông tin, tìm cực trị, để dùng sử dụng các công cụ tính toán hiện đại.

Chương 2

ĐẠI LƯỢNG NGẪU NHIÊN VÀ CÁC PHÂN PHỐI CỦA CHÚNG

Trong thực tế đời sống, hay trong kỹ thuật, thường xuyên chúng ta gặp những yếu tố ngẫu nhiên như sự biến động của giá cả, tỷ suất, nhiệt độ... Chúng là những đại lượng nhận được nhiều giá trị khác nhau với những điều kiện thí nghiệm không đổi.

Xét không gian Ω các biến cố ngẫu nhiên $A \in \Omega$. Biến ngẫu nhiên X biểu thị bằng số kết quả đo lường một đặc trưng, một chỉ tiêu nào đó có quan hệ với biến cố ngẫu nhiên. Do vậy biến cố ngẫu nhiên X sẽ nhận giá trị x với xác suất nào đó $P(X = x)$.

Đại lượng ngẫu nhiên X được gọi là rời rạc khi nó nhận một số giá trị hữu hạn hoặc vô hạn các giá trị đếm được khác nhau x_1, x_2, \dots, x_n .

Thí dụ: Số lần dừng máy trong ca sản xuất, có thể lấy giá trị: 0, 1, 2.

Nếu đại lượng X có thể nhận giá trị bất kỳ trong một khoảng của trục số, nó được gọi là ngẫu nhiên liên tục.

Thí dụ: Thời gian làm việc cho đến lúc máy hỏng, độ sai lệch đường kính danh nghĩa của một trục...

2.1. HÀM PHÂN PHỐI

Ta cần xác định không chỉ miền giá trị của đại lượng ngẫu nhiên, mà còn phải biết giá trị của đại lượng ngẫu nhiên rơi vào khoảng đã cho với một xác suất như thế nào? Đặc trưng là hàm phân phối xác suất của nó:

$$F(x) = P(X < x); \quad -\infty < x < +\infty \quad (2.1)$$

Biểu thức đó có nghĩa là: giá trị của hàm phân phối xác suất của đại lượng X tại điểm x bằng xác suất để đại lượng đó nhận một giá trị nhỏ hơn x .

2.1.1. Hàm phân phối của biến ngẫu nhiên rời rạc

Đại lượng ngẫu nhiên X lấy các giá trị rời rạc x_i ; $i = 1, 2, 3, \dots$ với xác suất p_i tương ứng:

X	x_1	x_2	x_3	...	x_i
$P(X = x_i)$	p_1	p_2	p_3	...	p_i

Hay viết dưới dạng biểu thức: