

**TCVN**

**TIÊU CHUẨN QUỐC GIA**

**TCVN 8244-1:2010**

**ISO 3534-1:2006**

Xuất bản lần 1

**THỐNG KÊ HỌC – TỪ VỰNG –  
PHẦN 1: THUẬT NGỮ CHUNG VỀ THỐNG KÊ VÀ  
THUẬT NGỮ DÙNG TRONG XÁC SUẤT**

*Statistics - Vocabulary and symbols –*

*Part 1: General statistical terms and terms used in probability*

**HÀ NỘI - 2010**

**Mục lục**

	Trang
Lời nói đầu .....	5
Lời giới thiệu .....	6
Phạm vi áp dụng .....	9
1 Thuật ngữ chung về thống kê .....	9
2 Thuật ngữ dùng trong xác suất .....	46
Phụ lục A (tham khảo) Các ký hiệu .....	92
Phụ lục B (tham khảo) Sơ đồ khái niệm về thống kê .....	94
Phụ lục C (tham khảo) Sơ đồ khái niệm về xác suất .....	100
Phụ lục D (tham khảo) Phương pháp luận sử dụng trong việc xây dựng từ vựng .....	104
Thư mục tài liệu tham khảo .....	110
Chỉ mục theo bảng chữ cái .....	111

**Contents**

	Page
Foreword .....	5
Introduction .....	6
Scope .....	9
1 General statistical terms .....	9
2 Terms used in probability .....	46
Annex A (informative) Symbols .....	92
Annex B (informative) Statistical concept diagrams .....	94
Annex C (informative) Probability concept diagrams .....	100
Annex D (informative) Methodology used in development of the vocabulary .....	104
Bibliography .....	110
Alphabetical index .....	113

## Lời nói đầu

TCVN 8244-1:2010 hoàn toàn tương đương với ISO 3534-1:2006;

TCVN 8244-1:2010 do Ban kỹ thuật Tiêu chuẩn Quốc gia TCVN/TC 69 *Ứng dụng các phương pháp thống kê* biên soạn, Tổng cục Tiêu chuẩn Đo lường Chất lượng đề nghị, Bộ Khoa học và Công nghệ công bố.

Bộ TCVN 8244 gồm các phần dưới đây có tên chung "Thống kê học – Từ vựng và ký hiệu":

- TCVN 8244-1:2010 (ISO 3534-1:2006), Phần 1: Thuật ngữ chung về thống kê và thuật ngữ dùng trong xác suất
- TCVN 8244-2:2010 (ISO 3534-2:2006), Phần 2: Thống kê ứng dụng

Bộ ISO 3534 còn có phần dưới đây có tên chung "Statistics – Vocabulary and symbols":

- Part 3: Design of experiments

## Lời giới thiệu

Tiêu chuẩn này và TCVN 8244-2 tương thích nhau. Hai tiêu chuẩn có chung mục đích là giới hạn mức độ toán học tương ứng đến mức tối thiểu cần thiết để có được những định nghĩa cơ động và chính xác. Các thuật ngữ trong Phần 1 sử dụng trong xác suất và thống kê là các thuật ngữ cơ bản, do vậy được trình bày ở mức độ toán học tương đối phức tạp. Thực tế là người sử dụng TCVN 8244-2 (ISO 3534-2) hoặc các tiêu chuẩn khác về thống kê ứng dụng của TC 69 thường có thể tham khảo tiêu chuẩn này đối với định nghĩa của các thuật ngữ nhất định, nhiều thuật ngữ được mô tả theo cách ít thiên về kỹ thuật hơn trong nội dung chú thích và được minh họa bằng các ví dụ. Mặc dù cách mô tả này không thay thế cho các định nghĩa chính thức nhưng chúng có thể cung cấp định nghĩa cho các khái niệm thực tế, cho người không có chuyên môn, do đó phục vụ cho nhu cầu của nhiều người sử dụng các tiêu chuẩn thuật ngữ này. Nhằm hỗ trợ thêm cho người sử dụng các tiêu chuẩn như TCVN 8244-2 (ISO 3534-2) hoặc TCVN 6910 (ISO 5725), các chú thích và ví dụ được đưa ra làm cho tiêu chuẩn này dễ tiếp cận hơn.

Tập hợp các thuật ngữ về xác suất và thống kê được định nghĩa đầy đủ và hoàn chỉnh rất cần thiết cho việc xây dựng và sử dụng hiệu quả các tiêu chuẩn về thống kê. Các định nghĩa được nêu ở đây phải đủ chính xác và phức tạp về mặt toán học nhằm giúp người xây dựng các tiêu chuẩn về thống kê không bị nhầm lẫn. Việc giải thích chi tiết hơn các khái niệm, ngữ cảnh và lĩnh vực ứng dụng có thể tìm trong các sách hướng dẫn về xác suất, thống kê.

Các sơ đồ khái niệm được đưa ra trong phụ lục tham khảo cho từng nhóm thuật ngữ: 1) thuật ngữ chung về thống kê (trong Phụ lục B) và 2) thuật ngữ dùng trong xác suất (trong Phụ lục C). Có sáu sơ đồ

## Introduction

The current versions of ISO 3534-1 and ISO 3534-2 are intended to be compatible. They share the common goal of restricting their respective mathematical levels to the minimum levels necessary to attain coherent, correct and concise definitions. Part 1 on terms used in probability and statistics is fundamental, so by necessity is presented at a somewhat sophisticated mathematical level. Recognizing that users of ISO 3534-2 or other TC 69 standards on applied statistics may occasionally consult this part of ISO 3534 for the definition of certain terms, several of the terms are described in a less technical manner within the notes and are illustrated with examples. Although these informal descriptions are not a substitute for formal definitions, they may provide a working, layman definition of concepts, thus serving the needs of multiple users of these terminology standards. To accommodate further the applied user who normally would be involved with standards such as ISO 3534-2 or ISO 5725, for example, notes and examples are offered to make this part of ISO 3534 more accessible.

A well-defined and reasonably complete set of probability and statistical terms is essential to the development of and effective use of statistical standards. The definitions provided here must be sufficiently accurate and mathematically sophisticated to enable statistical standards developers to avoid ambiguities. Of course, more detailed explanations of concepts, their contexts and their realms of application can be found in introductory probability and statistics textbooks.

Concept diagrams are provided in an informative annex for each group of terms: 1) general statistical terms (in Annex B) and 2) terms used

khái niệm đối với các thuật ngữ chung về thống kê và bốn sơ đồ cho các thuật ngữ liên quan đến xác suất. Một số thuật ngữ xuất hiện trong nhiều sơ đồ để cho thấy mối liên kết giữa tập hợp khái niệm này với tập hợp khái niệm khác. Phụ lục D đưa ra giới thiệu tóm tắt về các sơ đồ khái niệm và giải thích chúng.

Các sơ đồ này là công cụ hướng dẫn trong tiêu chuẩn vì chúng hỗ trợ việc mô tả các mối quan hệ giữa các thuật ngữ khác nhau. Các sơ đồ này cũng rất hữu ích cho việc chuyển dịch tiêu chuẩn sang các ngôn ngữ khác.

Như một chú thích chung cho phần lớn nội dung tiêu chuẩn, nếu không có quy định nào khác, các định nghĩa đều liên quan đến trường hợp một chiều (một biến). Điều này được nêu ở đây nhằm tránh việc đề cập lại nhiều lần về phạm vi một chiều cho hầu hết các định nghĩa.

in probability (in Annex C). There are six concept diagrams for general statistical terms and four concept diagrams for terms related to probability. Some terms appear in multiple diagrams to provide a link from one set of concepts to another. Annex D provides a brief introduction to Concept Diagrams and their interpretation.

These diagrams were instrumental in constructing this revision as they assist in delineating the interrelationships of the various terms. These diagrams are also likely to be useful in translating the standard into other languages.

As a general comment with respect to much of the standard, unless otherwise indicated, the definitions relate to the one-dimensional (univariate) case. This provision is admitted here to eliminate the need to mention repetitively the one-dimensional scope for most of the definitions.

## Thống kê học – Từ vựng và ký hiệu – Phần 1: Thuật ngữ chung về thống kê và thuật ngữ dùng trong xác suất

### Statistics – Vocabulary and symbols – Part 1: General statistical terms and terms used in probability

#### Phạm vi áp dụng

Tiêu chuẩn này quy định các thuật ngữ chung về thống kê và thuật ngữ dùng trong xác suất được sử dụng để soạn thảo các tiêu chuẩn khác. Ngoài ra, tiêu chuẩn này còn quy định các ký hiệu cho một số thuật ngữ.

Các thuật ngữ được phân loại thành:

- a) thuật ngữ chung về thống kê (điều 1);
- b) thuật ngữ dùng trong xác suất (điều 2).

Phụ lục A đưa ra danh mục các ký hiệu và chữ viết tắt khuyến nghị sử dụng trong tiêu chuẩn này. Các điều mục trong tiêu chuẩn này được sắp xếp phù hợp với các sơ đồ khái niệm trong Phụ lục B và C.

#### 1 Thuật ngữ chung về thống kê

##### 1.1

##### Tổng thể

Toàn bộ cá thể được xem xét.

CHÚ THÍCH 1: Một tổng thể có thể là thực và hữu hạn, thực và vô hạn hoặc hoàn toàn là giả thuyết. Đôi khi, thuật ngữ "tổng thể hữu hạn" được sử dụng, đặc biệt là trong lấy mẫu điều tra. Tương tự, thuật ngữ

#### Scope

This part of ISO 3534 defines general statistical terms and terms used in probability which may be used in the drafting of other International Standards. In addition, it defines symbols for a limited number of these terms.

The terms are classified as:

- a) general statistical terms (Clause 1);
- b) terms used in probability (Clause 2).

Annex A gives a list of symbols and abbreviations recommended to be used for this part of ISO 3534. The entries in this part of ISO 3534 are arranged in association with concept diagrams provided as Annexes B and C.

#### 1 General statistical terms

##### 1.1

##### population

totality of items under consideration

NOTE 1 A population may be real and finite, real and infinite or completely hypothetical. Sometimes the term "finite population" is used, especially in survey sampling. Likewise the term "infinite

"tổng thể vô hạn" được sử dụng trong trường hợp lấy mẫu từ một tập không đếm được. Trong điều 2, tổng thể được xem xét như không gian mẫu theo ngữ cảnh xác suất (2.1).

CHÚ THÍCH 2: Tổng thể giả thuyết cho phép tương tự tính chất của dữ liệu theo nhiều giả định. Do đó, tổng thể giả thuyết rất hữu ích ở giai đoạn thiết kế trong các điều tra thống kê, đặc biệt để xác định cỡ mẫu phù hợp. Tổng thể giả thuyết có thể có số lượng hữu hạn hoặc vô hạn. Khái niệm này đặc biệt hữu ích trong thống kê suy luận để hỗ trợ việc đánh giá độ mạnh của bằng chứng trong điều tra thống kê.

CHÚ THÍCH 3: Hoán cảnh nghiên cứu có thể chỉ ra bản chất của tổng thể. Ví dụ, nếu ba ngôi làng được chọn để điều tra nhân khẩu học hoặc sức khỏe, khi đó tổng thể bao gồm các cư dân của ba ngôi làng cụ thể này. Một cách khác, nếu ba ngôi làng này được chọn ngẫu nhiên trong số tất cả các làng trong một khu vực cụ thể, khi đó tổng thể sẽ gồm tất cả các cư dân trong khu vực đó.

## 1.2

### Đơn vị mẫu

một trong các phần riêng lẻ hợp thành **tổng thể** (1.1).

CHÚ THÍCH: Tùy từng trường hợp, phần quan tâm nhỏ nhất có thể là một cá thể, một hộ gia đình, một khu trường học, một đơn vị hành chính, v.v...

## 1.3

### Mẫu

Tập hợp con của **tổng thể** (1.1) gồm một hoặc nhiều **đơn vị mẫu** (1.2)

CHÚ THÍCH 1: Đơn vị mẫu có thể là các cá thể, các trị số hoặc thậm chí là các thực thể trừu tượng tùy thuộc vào tổng thể nghiên cứu.

CHÚ THÍCH 2: Định nghĩa về mẫu trong TCVN 8244-2 (ISO 3534-2) có ví dụ về khuôn khổ lấy mẫu thiết yếu trong việc lấy mẫu ngẫu nhiên từ một tổng thể hữu hạn.

"population" is used in the context of sampling from a continuum. In Clause 2, population will be viewed in a probabilistic context as the **sample space** (2.1).

NOTE 2 A hypothetical population allows one to imagine the nature of further data under various assumptions. Hence, hypothetical populations are useful at the design stage of statistical investigations, particularly for determining appropriate sample sizes. A hypothetical population could be finite or infinite in number. It is a particularly useful concept in inferential statistics to assist in evaluating the strength of evidence in a statistical investigation.

NOTE 3 The context of an investigation can dictate the nature of the population. For example, if three villages are selected for a demographic or health study, then the population consists of the residents of these particular villages. Alternatively, if the three villages were selected at random from among all of the villages in a specific region, then the population would consist of all residents of the region.

## 1.2

### sampling unit

one of the individual parts into which a **population** (1.1) is divided

NOTE Depending on the circumstances the smallest part of interest may be an individual, a household, a school district, an administrative unit and so forth.

## 1.3

### sample

subset of a **population** (1.1) made up of one or more **sampling units** (1.2)

NOTE 1 The sampling units could be items, numerical values or even abstract entities depending on the population of interest.

NOTE 2 The definition of sample in ISO 3534-2 includes an example of sampling frame which is essential in drawing a random sample from a finite population.



## 1.4

**Giá trị quan trắc**

Giá trị thu được của một tính chất gắn với một đơn vị của mẫu (1.3).

CHÚ THÍCH 1: Các từ đồng nghĩa là "thể hiện"; "dữ liệu".

CHÚ THÍCH 2: Định nghĩa này không quy định nguồn gốc hay cách thức thu được giá trị này. Giá trị có thể biểu diễn một thể hiện của **biến ngẫu nhiên** (2.10). Ngoài ra giá trị có thể là một trong số nhiều giá trị sẽ được phân tích thống kê sau đó. Mặc dù kết luận đúng đòi hỏi một số cơ sở thống kê vững chắc nhưng không loại trừ việc dùng các tính toán bằng số hoặc đồ thị bằng hình các giá trị quan trắc. Chỉ khi có các vấn đề như xác định xác suất quan trắc một tập hợp giá trị cụ thể thì bộ máy thống kê trở nên thích đáng và thiết yếu. Giai đoạn phân tích sơ bộ các giá trị quan trắc thường được coi là một phần của phân tích dữ liệu.

## 1.5

**Thống kê mô tả**

Việc mô tả các **giá trị quan trắc** (1.4) bằng đồ thị, trị số hoặc mô tả tóm tắt khác.

VÍ DỤ 1: Tóm tắt bằng số bao gồm **trung bình** (1.15), **độ rộng** (1.10), **độ lệch chuẩn mẫu** (1.17), v.v...

VÍ DỤ 2: Ví dụ về tóm tắt bằng đồ thị bao gồm biểu đồ hộp, biểu đồ, đồ thị Q-Q, đồ thị phân vị chuẩn, đám mây điểm, đám mây điểm nhiều chiều và biểu đồ phân bố.

## 1.6

**Mẫu ngẫu nhiên**

**Mẫu** (1.3) được chọn bằng phương pháp chọn ngẫu nhiên.

CHÚ THÍCH 1: Định nghĩa này rộng hơn định nghĩa trong TCVN 8244-2 (ISO 3534-2) để áp dụng được với tổng thể vô hạn.

## 1.4

**observed value**

obtained value of a property associated with one member of a **sample** (1.3)

NOTE 1 Common synonyms are "realization" and "datum".

NOTE 2 The definition does not specify the genesis or how this value has been obtained. The value may represent one realization of a **random variable** (2.10), but not exclusively so. It may be one of several such values that will be subsequently subjected to statistical analysis. Although proper inferences require some statistical underpinnings, there is nothing to preclude computing summaries or graphical depictions of observed values. Only when attendant issues such as determining the probability of observing a specific set of realizations does the statistical machinery become both relevant and essential. The preliminary stage of an analysis of observed values is commonly referred to as data analysis.

## 1.5

**descriptive statistics**

graphical, numerical or other summary depiction of **observed values** (1.4)

EXAMPLE 1 Numerical summaries include **average** (1.15), **range** (1.10), **sample standard deviation** (1.17), and so forth.

EXAMPLE 2 Examples of graphical summaries include boxplots, diagrams, Q-Q plots, normal quantile plots, scatterplots, multiple scatterplots and histograms.

## 1.6

**random sample**

**sample** (1.3) which has been selected by a method of random selection

NOTE 1 This definition is less restrictive than that given in ISO 3534-2 to allow for infinite populations.

CHÚ THÍCH 2: Khi một mẫu gồm  $n$  đơn vị mẫu được chọn từ không gian mẫu (2.1) hữu hạn, mỗi tổ hợp của  $n$  đơn vị mẫu có xác suất (2.5) lấy mẫu cụ thể. Đối với phương án lấy mẫu điều tra, xác suất riêng cho từng tổ hợp có thể được tính trước.

CHÚ THÍCH 3: Đối với việc lấy mẫu điều tra từ một không gian mẫu hữu hạn, một mẫu ngẫu nhiên có thể được chọn bằng các phương án lấy mẫu khác nhau như lấy mẫu ngẫu nhiên phân tầng, lấy mẫu ngẫu nhiên hệ thống, lấy mẫu theo nhóm, lấy mẫu với xác suất lấy mẫu tỷ lệ với cỡ biến phụ trợ và nhiều khả năng khác.

CHÚ THÍCH 4: Định nghĩa này nói chung đề cập đến giá trị quan trắc (1.4) thực tế. Các giá trị quan trắc này được coi là thể hiện của các biến ngẫu nhiên (2.10), trong đó mỗi giá trị quan trắc tương ứng với một biến ngẫu nhiên. Khi hàm ước lượng (1.12), thống kê kiểm nghiệm cho các kiểm nghiệm thống kê (1.48) hoặc khoảng tin cậy (1.28) được suy ra từ mẫu ngẫu nhiên, thì định nghĩa này nói đến các biến ngẫu nhiên phát sinh từ các thực thể trừu tượng chứ không phải giá trị quan trắc thực tế của các biến ngẫu nhiên này.

CHÚ THÍCH 5: Mẫu ngẫu nhiên lấy từ tổng thể vô hạn thường được tạo ra bởi việc lấy lặp lại từ không gian mẫu, dẫn đến một mẫu có các biến ngẫu nhiên độc lập cùng phân bố khi áp dụng giải thích về định nghĩa này đề cập trong Chú thích 4.

## 1.7

### Mẫu ngẫu nhiên đơn giản

mẫu ngẫu nhiên (1.6) của <tổng thể hữu hạn> sao cho mỗi tập hợp con với cỡ mẫu đã cho xác suất lấy như nhau.

CHÚ THÍCH: Định nghĩa này phù hợp với định nghĩa nêu trong TCVN 8244-2 (ISO 3534-2), mặc dù cách diễn đạt hơi khác nhau.

NOTE 2 When the sample of  $n$  sampling units is selected from a finite sample space (2.1), each of the possible combinations of  $n$  sampling units will have a particular probability (2.5) of being taken. For survey sampling plans, the particular probability for each possible combination may be calculated in advance.

NOTE 3 For survey sampling from a finite sample space, a random sample can be selected by different sampling plans such as stratified random sampling, systematic random sampling, cluster sampling, sampling with probability of sampling proportional to the size of an auxiliary variable and many other possibilities.

NOTE 4 The definition generally refers to actual observed values (1.4). These observed values are considered as realizations of random variables (2.10), where each observed value corresponds to one random variable. When estimators (1.12), test statistics for statistical tests (1.48) or confidence intervals (1.28) are derived from a random sample, the definition accommodates reference to the random variables arising from abstract entities in the sample rather than the actual observed values of these random variables.

NOTE 5 Random samples from infinite populations are often generated by repeated draws from the sample space, leading to a sample consisting of independent, identically distributed random variables using the interpretation of this definition mentioned in Note 4.

## 1.7

### simple random sample

<finite population> random sample (1.6) such that each subset of a given size has the same probability of selection

NOTE This definition is in harmony with the definition given in ISO 3534-2, although the wording here is slightly different.

## 1.8

**Thống kê**

Hàm số hoàn toàn xác định của các **biến ngẫu nhiên** (2.10)

CHÚ THÍCH 1: Thống kê là hàm số của các biến ngẫu nhiên trong một **mẫu ngẫu nhiên** (1.6) theo nghĩa nêu trong Chú thích 4 của 1.6.

CHÚ THÍCH 2: Theo Chú thích 1, nếu  $\{X_1, X_2, \dots, X_n\}$  là mẫu ngẫu nhiên lấy từ **phân bố chuẩn** (2.50) với **trung bình** (2.35)  $\mu$  chưa biết và **độ lệch chuẩn** (2.37)  $\sigma$  chưa biết, khi đó  $(X_1 + X_2 + \dots + X_n)/n$  là một **thống kê**, **trung bình mẫu** (1.15), trong đó  $[(X_1 + X_2 + \dots + X_n)/n] - \mu$  không phải là **thống kê** vì nó có giá trị của **tham số** (2.9)  $\mu$  chưa biết.

CHÚ THÍCH 3: Định nghĩa được nêu ở đây là định nghĩa kỹ thuật, tương ứng với cách xử lý trong thống kê toán học.

## 1.9

**Thống kê thứ tự**

**Thống kê** (1.8) xác định bởi thứ tự của nó trong một sắp xếp không giảm của các **biến ngẫu nhiên** (2.10).

VÍ DỤ: Cho các giá trị quan trắc của một mẫu là 9, 13, 7, 6, 13, 7, 19, 6, 10, và 7. Giá trị quan trắc của các **thống kê** thứ tự là 6, 6, 7, 7, 7, 9, 10, 13, 13, 19. Các giá trị này tạo thành thể hiện của  $X_{(1)}$  đến  $X_{(10)}$ .

CHÚ THÍCH 1: Cho **giá trị quan trắc** (1.4) của một **mẫu ngẫu nhiên** (1.6) là  $\{x_1, x_2, \dots, x_n\}$  và khi sắp xếp theo thứ tự không giảm được ấn định là  $x_{(1)} \leq \dots \leq x_{(k)} \leq \dots \leq x_{(n)}$ . Khi đó,  $(x_{(1)}, \dots, x_{(k)}, \dots, x_{(n)})$  là giá trị quan trắc của **thống kê** thứ tự  $(X_{(1)}, \dots, X_{(k)}, \dots, X_{(n)})$  và  $x_{(k)}$  là giá trị quan trắc của **thống kê** thứ tự thứ  $k$ .

CHÚ THÍCH 2: Thực tế là việc có được các **thống kê** thứ tự của một tập hợp dữ liệu cũng chính là sắp xếp các dữ liệu như mô tả trong Chú thích 1 ở trên. Khi đó, kiểu sắp xếp của tập hợp dữ liệu cũng có thể dùng để thu được **thống kê** tổng hợp hữu ích như

## 1.8

**statistic**

completely specified function of **random variables** (2.10)

NOTE 1 A statistic is a function of random variables in a **random sample** (1.6) in the sense given in Note 4 of 1.6.

NOTE 2 Referring to Note 1, if  $\{X_1, X_2, \dots, X_n\}$  is a random sample from a **normal distribution** (2.50) with unknown **mean** (2.35)  $\mu$  and unknown **standard deviation** (2.37)  $\sigma$ , then the expression  $(X_1 + X_2 + \dots + X_n)/n$  is a statistic, the **sample mean** (1.15), whereas  $[(X_1 + X_2 + \dots + X_n)/n] - \mu$  is not a statistic as it involves the unknown value of the **parameter** (2.9)  $\mu$ .

NOTE 3 The definition given here is a technical one, corresponding to the treatment found in mathematical statistics.

## 1.9

**order statistic**

**statistic** (1.8) determined by its ranking in a non-decreasing arrangement of **random variables** (2.10)

EXAMPLE Let the observed values of a sample be 9, 13, 7, 6, 13, 7, 19, 6, 10, and 7. The observed values of the order statistics are 6, 6, 7, 7, 7, 9, 10, 13, 13, 19. These values constitute realizations of  $X_{(1)}$  through  $X_{(10)}$ .

NOTE 1 Let the **observed values** (1.4) of a **random sample** (1.6) be  $\{x_1, x_2, \dots, x_n\}$  and once sorted in non-decreasing order designated as  $x_{(1)} \leq \dots \leq x_{(k)} \leq \dots \leq x_{(n)}$ . Then  $(x_{(1)}, \dots, x_{(k)}, \dots, x_{(n)})$  is the observed value of the order statistic  $(X_{(1)}, \dots, X_{(k)}, \dots, X_{(n)})$  and  $x_{(k)}$  is the observed value of the  $k$ th order statistic.

NOTE 2 In practical terms, obtaining the order statistics for a data set amounts to sorting the data as formally described in Note 1. The sorted form of the

## TCVN 8244-1:2010

nêu trong một số định nghĩa tiếp theo.

CHÚ THÍCH 3: Thống kê thứ tự liên quan đến các giá trị mẫu xác định bởi vị trí của chúng sau khi sắp xếp theo thứ tự không giảm. Như trong ví dụ, có thể dễ dàng hiểu việc sắp xếp các giá trị mẫu (thể hiện của biến ngẫu nhiên) hơn là việc sắp xếp của các biến ngẫu nhiên không được quan tâm. Tuy nhiên, ta có thể hiểu về biến ngẫu nhiên từ **mẫu ngẫu nhiên** (1.6) được sắp xếp theo thứ tự không giảm. Ví dụ, có thể nghiên cứu giá trị lớn nhất trong số  $n$  biến ngẫu nhiên trước khi nhận giá trị của nó.

CHÚ THÍCH 4: Mỗi thống kê thứ tự riêng là một thống kê, là hàm số xác định hoàn toàn của một biến ngẫu nhiên. Hàm số này đơn giản là hàm phân định vị trí hoặc thứ tự trong tập hợp các biến ngẫu nhiên đã được sắp xếp.

CHÚ THÍCH 5: Các giá trị trùng lặp gây ra một vấn đề tiềm ẩn, đặc biệt đối với các biến ngẫu nhiên rời rạc và các thể hiện có khả năng tách biệt thấp. Từ "không giảm" được sử dụng tốt hơn từ "tăng" như một cách tiếp cận tế nhị vấn đề này. Cần nhấn mạnh rằng các giá trị trùng lặp được duy trì và không được gộp vào như một giá trị duy nhất. Trong ví dụ ở trên, hai thể hiện của 6 và 6 là giá trị trùng lặp.

CHÚ THÍCH 6: Thứ tự được xác định theo trục của đường thẳng thực và không theo giá trị tuyệt đối của các biến ngẫu nhiên.

CHÚ THÍCH 7: Tập hợp đầy đủ của thống kê thứ tự gồm một biến ngẫu nhiên  $n$  chiều, trong đó  $n$  là số quan trắc trong mẫu đó.

CHÚ THÍCH 8: Các thành phần của thống kê thứ tự cũng được coi như thống kê thứ tự nhưng với một chỉ số cho biết vị trí của nó trong chuỗi giá trị mẫu đã được sắp xếp.

CHÚ THÍCH 9: Giá trị nhỏ nhất, lớn nhất, và đối với cỡ mẫu số lẻ, **trung vị mẫu** (1.13), là trường hợp đặc biệt của thống kê thứ tự. Ví dụ, đối với cỡ mẫu 11,

data set then lends itself to obtaining useful summary statistics as given in the next few definitions.

NOTE 3 Order statistics involve sample values identified by their position after ranking in non-decreasing order. As in the example, it is easier to understand the sorting of sample values (realizations of random variables) rather than the sorting of unobserved random variables. Nevertheless, one can conceive of random variables from a **random sample** (1.6) being arranged in a non-decreasing order. For example, the maximum of  $n$  random variables can be studied in advance of its realized value.

NOTE 4 An individual order statistic is a statistic which is a completely specified function of a random variable. This function is simply the identity function with the further identification of position or rank in the sorted set of random variables.

NOTE 5 Tied values pose a potential problem especially for discrete random variables and for realizations that are reported to low resolution. The word "non-decreasing" is used rather than "ascending" as a subtle approach to the problem. It should be emphasized that tied values are retained and not collapsed into the single tied value. In the example above, the two realizations of 6 and 6 are tied values.

NOTE 6 Ordering takes place with reference to the real line and not to the absolute values of the random variables.

NOTE 7 The complete set of order statistics consist of an  $n$  dimensional random variable, where  $n$  is the number of observations in the sample.

NOTE 8 The components of the order statistic are also referred to as order statistics but with a qualifier that gives the number in the sequence of ordered values of the sample.

NOTE 9 The minimum, the maximum, and for odd-numbered sample sizes, the **sample median**

$X_{(1)}$  là nhỏ nhất,  $X_{(11)}$  là lớn nhất và  $X_{(6)}$  là trung vị mẫu.

### 1.10

#### Độ rộng mẫu

Thống kê thứ tự lớn nhất (1.9) trừ đi thống kê thứ tự nhỏ nhất.

VÍ DỤ: Tiếp tục với ví dụ ở 1.9, độ rộng mẫu quan trắc là  $19 - 6 = 13$ .

CHÚ THÍCH: Trong kiểm soát thống kê quá trình, độ rộng mẫu thường được dùng để theo dõi độ phân tán theo thời gian của quá trình, đặc biệt khi cỡ mẫu tương đối nhỏ.

### 1.11

#### Tâm độ rộng

trung bình (1.15) của thống kê thứ tự (1.9) nhỏ nhất và lớn nhất

VÍ DỤ: Tâm độ rộng quan trắc của các giá trị trong ví dụ ở 1.9 là  $(6+19)/2 = 12,5$ .

CHÚ THÍCH: Tâm độ rộng cho phép đánh giá nhanh và đơn giản điểm giữa của các tập hợp dữ liệu nhỏ.

### 1.12

#### Hàm ước lượng

$\hat{\theta}$

thống kê (1.8) dùng trong phép ước lượng (1.36) tham số  $\theta$ .

CHÚ THÍCH 1: Hàm ước lượng có thể là trung bình mẫu (1.15) để ước lượng trung bình của tổng thể (2.35), biểu thị bằng  $\mu$ . Đối với phân bố (2.11) như phân bố chuẩn (2.50), hàm ước lượng "tự nhiên" của trung bình tổng thể  $\mu$  là trung bình mẫu.

CHÚ THÍCH 2: Để ước lượng tính chất của tổng thể [ví dụ mode (2.27) đối với một phân bố đơn biến (2.16)], một ước lượng thích hợp có thể là hàm của (các) ước lượng của (các) tham số của phân bố hoặc có thể là hàm phức hợp của mẫu ngẫu nhiên (1.6).

CHÚ THÍCH 3: Thuật ngữ "hàm ước lượng" được

(1.13), are special cases of order statistics. For example, for sample size 11,  $X_{(1)}$  is the minimum,  $X_{(11)}$  is the maximum and  $X_{(6)}$  is the sample median.

### 1.10

#### sample range

largest order statistic (1.9) minus the smallest order statistic

EXAMPLE Continuing with the example from 1.9, the observed sample range is  $19 - 6 = 13$ .

NOTE In statistical process control, the sample range is often used to monitor the dispersion over time of a process, particularly when the sample sizes are relatively small.

### 1.11

#### mid-range

average (1.15) of smallest and largest order statistics (1.9)

EXAMPLE The observed mid-range for the example values in 1.9 is  $(6+19)/2 = 12,5$ .

NOTE The mid-range provides a quick and simple assessment of the middle of small data sets.

### 1.12

#### estimator

$\hat{\theta}$

statistic (1.8) used in estimation (1.36) of the parameter  $\theta$

NOTE 1 An estimator could be the sample mean (1.15) intended to estimate the population mean (2.35), which could be denoted by  $\mu$ . For a distribution (2.11) such as the normal distribution (2.50), the "natural" estimator of the population mean  $\mu$  is the sample mean.

NOTE 2 For estimating a population property [e.g. the mode (2.27) for a univariate distribution (2.16)], an appropriate estimator could be a function of the estimator(s) of the parameter(s) of a distribution or could be a complex function of a

dùng ở đây với nghĩa rộng. Nó bao gồm các ước lượng điểm đối với tham số, cũng như ước lượng khoảng có thể dùng để dự đoán (đôi khi được gọi là hàm dự đoán). Hàm ước lượng cũng có thể bao gồm các hàm như ước lượng nhân và thống kê với mục đích đặc biệt khác. Thông tin thêm được nêu trong chú thích của 1.36.

1.13

**Trung vị mẫu**

**Thống kê thứ tự** thứ  $[(n+1)/2]$  (1.9), nếu cỡ mẫu (xem TCVN 8244-2 (ISO 3534-2), 1.2.26)  $n$  là số lẻ; tổng của thống kê thứ tự thứ  $(n/2)$  và  $(n/2) + 1$  chia cho 2, nếu cỡ mẫu  $n$  là chẵn.

VÍ DỤ: Tiếp theo ví dụ của 1.9, giá trị của 8 là thể hiện trung vị mẫu. Trong trường hợp này (ngay cả cỡ mẫu bằng 10), giá trị thứ 5 và thứ 6 là 7 và 9, trung bình là 8. Trên thực tế, điều này được ghi là "trung vị mẫu là 8", mặc dù nói một cách chặt chẽ thì trung vị mẫu được định nghĩa là biến ngẫu nhiên.

CHÚ THÍCH 1: Đối với mẫu ngẫu nhiên (1.6) có cỡ mẫu  $n$ , trong đó biến ngẫu nhiên (2.10) được sắp xếp theo thứ tự không giảm từ 1 đến  $n$ , trung vị mẫu là biến ngẫu nhiên thứ  $(n+1)/2$  nếu cỡ mẫu là lẻ. Nếu cỡ mẫu  $n$  là chẵn thì trung vị mẫu là trung bình của biến ngẫu nhiên thứ  $(n/2)$  và  $(n+1)/2$ .

CHÚ THÍCH 2: Về lý thuyết thì dường như không thể sắp xếp các biến ngẫu nhiên chưa được quan trắc. Tuy nhiên, có thể thiết lập cấu trúc của các thống kê thứ tự biết rõ sao cho có thể tiến hành phân tích dựa trên quan trắc. Trong thực tế, khi thu được các giá trị quan trắc và thông qua việc sắp xếp các giá trị, ta sẽ có được thể hiện của các thống kê thứ tự. Sau đó, các thể hiện này có thể được giải thích từ cấu trúc của thống kê thứ tự từ mẫu ngẫu nhiên.

random sample (1.6).

NOTE 3 The term "estimator" is used here in a broad sense. It includes the point estimator for a parameter, as well as the interval estimator which is possibly used for prediction (sometimes referred to as a predictor). Estimator also can include functions such as kernel estimators and other special purpose statistics. Additional discussion is provided in the notes to 1.36.

1.13

**sample median**

$[(n+1)/2]$ th order statistic (1.9), if the sample size (see ISO 3534-2:2006, 1.2.26)  $n$  is odd; sum of the  $(n/2)$ th and  $[(n/2) + 1]$ th order statistics divided by 2, if the sample size  $n$  is even

EXAMPLE Continuing with the example of 1.9, the value of 8 is a realization of the sample median. In this case (even sample size of 10), the 5th and 6th values were 7 and 9, whose average equals 8. In practice, this would be reported as "the sample median is 8", although strictly speaking, the sample median is defined as a random variable.

NOTE 1 For a random sample (1.6) of sample size  $n$  whose random variables (2.10) are arranged in non-decreasing order from 1 to  $n$ , the sample median is the  $(n+1)/2$ th random variable if the sample size is odd. If the sample size  $n$  is even, then the sample median is the average of the  $(n/2)$ th and  $(n+1)/2$ th random variables.

NOTE 2 Conceptually, it may seem impossible to conduct an ordering of random variables which have not yet been observed. Nevertheless, the structure for understanding order statistics can be established so that upon observation, the analysis may proceed. In practice, one obtains observed values and through sorting the values, one obtains realizations of the order statistics. These realizations can then be interpreted from the structure of order statistics from a random sample.

CHÚ THÍCH 3: Trung vị mẫu cung cấp hàm ước lượng điểm giữa của phân bố, mỗi phía chứa một nửa mẫu.

CHÚ THÍCH 4: Trên thực tế, trung vị mẫu có ích trong việc đưa ra ước lượng không nhạy với các giá trị cực trị trong tập dữ liệu. Ví dụ, trung vị thu nhập và trung vị giá nhà ở thường được báo cáo là giá trị tóm tắt.

#### 1.14

##### Mômen mẫu bậc $k$

$$E(X^k)$$

Tổng các biến ngẫu nhiên (2.10) lũy thừa  $k$  trong mẫu ngẫu nhiên (1.6) chia cho số quan trắc trong mẫu (1.3).

CHÚ THÍCH 1: Đối với mẫu ngẫu nhiên cỡ  $n$ , nghĩa là  $\{X_1, X_2, \dots, X_n\}$ , mômen mẫu bậc  $k$ ,  $E(X^k)$ , là

$$\frac{1}{n} \sum_{i=1}^n X_i^k$$

CHÚ THÍCH 2: Ngoài ra, khái niệm này có thể được mô tả như là mômen mẫu bậc  $k$  so với điểm không.

CHÚ THÍCH 3: Mômen mẫu bậc 1 được đề cập trong định nghĩa tiếp theo là trung bình mẫu (1.15).

CHÚ THÍCH 4: Mặc dù định nghĩa này được đưa ra cho  $k$  bất kỳ, thường dùng trong các trường hợp thực tế  $k = 1$  [trung bình mẫu (1.15)],  $k = 2$  [kèm theo phương sai mẫu (1.16) và độ lệch chuẩn mẫu (1.17)],  $k = 3$  [liên quan đến hệ số bất đối xứng của mẫu (1.20)] và  $k = 4$  [liên quan đến hệ số nhọn của mẫu (1.21)].

CHÚ THÍCH 5: "E" trong  $E(X^k)$  lấy từ "giá trị kỳ vọng" hoặc "kỳ vọng" của biến ngẫu nhiên  $X$ .

#### 1.15

##### trung bình mẫu

##### trung bình

##### trung bình số học

tổng các biến ngẫu nhiên (2.10) trong mẫu ngẫu nhiên

NOTE 3 The sample median provides an estimator of the middle of a distribution, with half of the sample to each side of it.

NOTE 4 In practice, the sample median is useful in providing an estimator that is insensitive to very extreme values in a data set. For example, median incomes and median housing prices are frequently reported as summary values.

#### 1.14

##### sample moment of order $k$

$$E(X^k)$$

sum of  $k^{\text{th}}$  power of random variables (2.10) in a random sample (1.6) divided by the number of observations in the sample (1.3)

NOTE 1 For a random sample of sample size  $n$ , i.e.  $\{X_1, X_2, \dots, X_n\}$ , the sample moment of order  $k$ ,  $E(X^k)$ , is

$$\frac{1}{n} \sum_{i=1}^n X_i^k$$

NOTE 2 Furthermore, this concept can be described as the sample moment of order  $k$  about zero.

NOTE 3 The sample moment of order 1 will be seen in the next definition to be the sample mean (1.15).

NOTE 4 Although the definition is given for arbitrary  $k$ , commonly used instances in practice involve  $k = 1$  [sample mean (1.15)],  $k = 2$  [associated with the sample variance (1.16) and sample standard deviation (1.17)],  $k = 3$  [related to sample coefficient of skewness (1.20)] and  $k = 4$  [related to sample coefficient of kurtosis (1.21)].

NOTE 5 The "E" in  $E(X^k)$  comes from the "expected value" or "expectation" of the random variable  $X$ .

#### 1.15

##### sample mean

##### average

##### arithmetic mean

sum of random variables (2.10) in a random

nhiên (1.6) chia cho số các số hạng trong tổng đó.

VÍ DỤ: Tiếp theo ví dụ ở 1.9, thể hiện của trung bình mẫu là 9,7 vì tổng các giá trị quan trắc là 97 và cỡ mẫu là 10.

CHÚ THÍCH 1: Coi như một thống kê, trung bình mẫu là hàm của các biến ngẫu nhiên từ mẫu ngẫu nhiên theo nghĩa nêu trong Chú thích 3 của 1.8. Ta phải phân biệt hàm ước lượng này với trị số của trung bình mẫu tính được từ các giá trị quan trắc (1.4) trong mẫu ngẫu nhiên đó.

CHÚ THÍCH 2: Trung bình mẫu được coi là thống kê thường được sử dụng như ước lượng cho trung bình của tổng thể (2.35). Từ đồng nghĩa thường dùng là trung bình số học.

CHÚ THÍCH 3: Đối với mẫu ngẫu nhiên cỡ  $n$ , nghĩa là  $\{X_1, X_2, \dots, X_n\}$ , trung bình mẫu là:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

CHÚ THÍCH 4: Trung bình mẫu có thể coi là mômen mẫu bậc 1.

CHÚ THÍCH 5: Đối với cỡ mẫu 2, trung bình mẫu, trung vị mẫu (1.13) và tâm độ rộng (1.11) là như nhau.

### 1.16

#### Phương sai mẫu

$s^2$

Tổng bình phương các độ lệch của các biến ngẫu nhiên (2.10) trong mẫu ngẫu nhiên (1.6) so với trung bình mẫu (1.15) của chúng chia cho số các số hạng trong tổng đó trừ đi một.

VÍ DỤ: Tiếp theo ví dụ số ở 1.9, phương sai mẫu được tính là 17,57. Tổng các bình phương của trung bình mẫu quan trắc là 158,10 và cỡ mẫu là 10 trừ đi 1 còn 9, cho mẫu số thích hợp.

CHÚ THÍCH 1: Coi như một thống kê (1.8), phương

sample (1.6) divided by the number of terms in the sum

EXAMPLE Continuing with the example from 1.9, the realization of the sample mean is 9,7 as the sum of the observed values is 97 and the sample size is 10.

NOTE 1 Considered as a statistic, the sample mean is a function of random variables from a random sample in the sense given in Note 3 of 1.8. One must distinguish this estimator from the numerical value of the sample mean calculated from the **observed values** (1.4) in the random sample.

NOTE 2 The sample mean considered as a statistic is often used as an estimator for the population mean (2.35). A common synonym is arithmetic mean.

NOTE 3 For a random sample of sample size  $n$ , i.e.  $\{X_1, X_2, \dots, X_n\}$ , the sample mean is:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

NOTE 4 The sample mean can be recognized as the sample moment of order 1.

NOTE 5 For sample size 2, the sample mean, the **sample median** (1.13) and **mid-range** (1.11) are the same.

### 1.16

#### sample variance

$s^2$

sum of squared deviations of **random variables** (2.10) in a **random sample** (1.6) from their **sample mean** (1.15) divided by the number of terms in the sum minus one

EXAMPLE Continuing with the numerical example of 1.9, the sample variance can be computed to be 17,57. The sum of squares about the observed sample mean is 158,10 and the sample size 10 minus 1 is 9, giving the appropriate denominator.

NOTE 1 Considered as a **statistic** (1.8), the sample



sai mẫu  $S^2$  là hàm số của các biến ngẫu nhiên từ một mẫu ngẫu nhiên. Ta phải phân biệt hàm ước lượng (1.12) với trị số của phương sai mẫu tính được từ các giá trị quan trắc (1.4) trong mẫu ngẫu nhiên. Trị số này được gọi là phương sai mẫu thực nghiệm hoặc phương sai mẫu quan trắc và thường được ký hiệu là  $s^2$ .

CHÚ THÍCH 2: Đối với mẫu ngẫu nhiên cỡ  $n$ , nghĩa là  $\{X_1, X_2, \dots, X_n\}$  có trung bình mẫu  $\bar{X}$  thì phương sai mẫu là:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

CHÚ THÍCH 3: Phương sai mẫu là một thống kê "gần như" trung bình bình phương độ lệch của các biến ngẫu nhiên (2.10) so với trung bình mẫu của chúng (chỉ "gần như" vì trong mẫu số  $n-1$  được sử dụng thay cho  $n$ ). Việc sử dụng  $n-1$  cung cấp ước lượng không chệch (1.34) của phương sai (2.36) tổng thể.

CHÚ THÍCH 4: Đại lượng  $n-1$  được gọi là bậc tự do (2.54).

CHÚ THÍCH 5: Phương sai mẫu có thể coi là mômen mẫu bậc hai của biến ngẫu nhiên của mẫu chuẩn hóa (1.19).

## 1.17

### độ lệch chuẩn mẫu

$S$

Căn bậc hai không âm của phương sai mẫu (1.16).

VÍ DỤ: Tiếp theo ví dụ số ở 1.9, độ lệch chuẩn mẫu quan trắc là 4,192 vì phương sai mẫu quan trắc là 17,57.

CHÚ THÍCH 1: Trên thực tế, độ lệch chuẩn mẫu được dùng để ước lượng độ lệch chuẩn (2.37). Ở đây cũng cần nhấn mạnh rằng  $S$  cũng là biến ngẫu nhiên (2.10) chứ không phải là thể hiện của mẫu ngẫu nhiên (1.6).

variance  $S^2$  is a function of random variables from a random sample. One has to distinguish this estimator (1.12) from the numerical value of the sample variance calculated from the observed values (1.4) in the random sample. This numerical value is called the empirical sample variance or the observed sample variance and is usually denoted by  $s^2$ .

NOTE 2 For a random sample of sample size  $n$ , i.e.  $\{X_1, X_2, \dots, X_n\}$  with sample mean  $\bar{X}$  the sample variance is:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

NOTE 3 The sample variance is a statistic that is "almost" the average of the squared deviations of the random variables (2.10) from their sample mean (only "almost" since  $n-1$  is used rather than  $n$  in the denominator). Using  $n-1$  provides an unbiased estimator (1.34) of the population variance (2.36).

NOTE 4 The quantity  $n-1$  is known as the degrees of freedom (2.54).

NOTE 5 The sample variance can be recognized to be the 2<sup>nd</sup> sample moment of the standardized sample random variables (1.19).

## 1.17

### sample standard deviation

$S$

non-negative square root of the sample variance (1.16)

EXAMPLE Continuing with the numerical example of 1.9, the observed sample standard deviation is 4,192 since the observed sample variance is 17,57.

NOTE 1 In practice, the sample standard deviation is used to estimate the standard deviation (2.37). Here again, it should be emphasized that  $S$  is also a random variable (2.10) and not a realization from a random sample (1.6).

CHÚ THÍCH 2: Độ lệch chuẩn mẫu là thước đo độ phân tán của phân bố (2.11).

1.18

**Hệ số biến động mẫu**

Độ lệch chuẩn mẫu (1.17) chia cho trung bình mẫu (1.15).

CHÚ THÍCH: Như với hệ số biến động (2.38), việc sử dụng thống kê này được giới hạn ở các tổng thể có giá trị dương. Hệ số biến động thường được lấy theo phần trăm.

1.19

**Biến ngẫu nhiên của mẫu chuẩn hóa**

Biến ngẫu nhiên (2.10) trừ đi trung bình mẫu (1.15) chia cho độ lệch chuẩn mẫu (1.17).

VÍ DỤ: Đối với ví dụ ở 1.9, trung bình mẫu quan trắc là 9,7 và độ lệch chuẩn mẫu quan trắc là 4,192. Do đó, biến ngẫu nhiên chuẩn hóa quan trắc (lấy đến hai chữ số thập phân) là:

-0,17; 0,79; -0,64; -0,88; 0,79; -0,64; 2,22; -0,88; 0,07; -0,62.

CHÚ THÍCH 1: Biến ngẫu nhiên của mẫu chuẩn hóa được phân biệt với biến ngẫu nhiên chuẩn hóa (2.33) lý thuyết tương ứng. Mục đích của việc chuẩn hóa là để chuyển đổi thành các biến ngẫu nhiên có trung bình "không" và độ lệch chuẩn đơn vị, để dễ dàng giải thích và so sánh.

CHÚ THÍCH 2: Giá trị quan trắc chuẩn hóa có trung bình quan trắc bằng không và độ lệch chuẩn quan trắc bằng 1.

1.20

**Hệ số bất đối xứng của mẫu**

Trung bình số học của lũy thừa bậc ba của biến ngẫu nhiên của mẫu chuẩn hóa (1.19) từ mẫu ngẫu nhiên (1.6).

VÍ DỤ: Tiếp theo ví dụ ở 1.9, hệ số bất đối xứng của

NOTE 2 The sample standard deviation is a measure of the dispersion of a distribution (2.11).

1.18

**sample coefficient of variation**

sample standard deviation (1.17) divided by the sample mean (1.15)

NOTE As with the coefficient of variation (2.38), the utility of this statistic is limited to populations that are positive valued. The coefficient of variation is commonly reported as a percentage.

1.19

**standardized sample random variable**

random variable (2.10) minus its sample mean (1.15) divided by the sample standard deviation (1.17)

EXAMPLE For the example of 1.9, the observed sample mean is 9,7 and the observed sample standard deviation is 4,192. Hence, the observed standardized random variables (to two decimal places) are:

-0,17; 0,79; -0,64; -0,88; 0,79; -0,64; 2,22; -0,88; 0,07; -0,62.

NOTE 1 The standardized sample random variable is distinguished from its theoretical counterpart standardized random variable (2.33). The intent of standardizing is to transform random variables to have zero means and unit standard deviations, for ease in interpretation and comparison.

NOTE 2 Standardized observed values have an observed mean of zero and an observed standard deviation of 1.

1.20

**sample coefficient of skewness**

arithmetic mean of the third power of the standardized sample random variables (1.19) from a random sample (1.6)

EXAMPLE Continuing with the example from

mẫu quan trắc được tính bằng 0,971 88. Với cỡ mẫu 10 trong ví dụ này, hệ số bất đối xứng của mẫu biến động lớn, do đó phải sử dụng một cách thận trọng. Sử dụng công thức thay thế trong Chú thích 1, tính được giá trị là 1,349 83.

CHÚ THÍCH 1: Công thức tương ứng với định nghĩa là

$$\frac{1}{n} \sum_{i=1}^n \left( \frac{X_i - \bar{X}}{S} \right)^3$$

Một số bộ chương trình thống kê sử dụng công thức dưới đây cho hệ số bất đối xứng của mẫu để hiệu chỉnh độ chệch (1.33):

$$\frac{n}{(n-1)(n-2)} \sum_{i=1}^n Z_i^3$$

trong đó

$$Z_i = \frac{X_i - \bar{X}}{S}$$

Đối với cỡ mẫu lớn, sự khác biệt giữa hai ước lượng này là không đáng kể. Tỷ số giữa ước lượng không chệch và ước lượng chệch là 1,389 đối với  $n = 10$ , 1,031 đối với  $n = 100$  and 1,003 đối với  $n = 1\,000$ .

CHÚ THÍCH 2: Hệ số bất đối xứng đề cập tới sự không đối xứng. Giá trị của thống kê này gần với "không" gọi ý là phân bố đang xét là gần đối xứng, trong khi các giá trị khác không tương ứng với phân bố khi có các giá trị cực trị quá xa về một phía của tâm phân bố. Dữ liệu bất đối xứng cũng được phản ánh trong các giá trị của **trung bình mẫu** (1.15) và **trung vị mẫu** (1.13) là không giống nhau. Dữ liệu bất đối xứng dương (bất đối xứng phải) chỉ ra khả năng có một vài quan trắc cực trị, lớn. Tương tự, dữ liệu bất đối xứng âm (bất đối xứng trái) chỉ ra khả năng có một vài quan trắc cực trị, nhỏ.

CHÚ THÍCH 3: Hệ số bất đối xứng của mẫu có thể coi là mômen mẫu bậc ba của biến ngẫu nhiên của mẫu chuẩn hóa (1.19).

1.9, the observed sample coefficient of skewness can be computed to be 0,971 88. For a sample size such as 10 in this example, the sample coefficient of skewness is highly variable, so it must be used with caution. Using the alternative formula in Note 1, the computed value is 1,349 83.

NOTE 1 The formula corresponding to the definition is

$$\frac{1}{n} \sum_{i=1}^n \left( \frac{X_i - \bar{X}}{S} \right)^3$$

Some statistical packages use the following formula for the sample coefficient of skewness to correct for bias (1.33):

$$\frac{n}{(n-1)(n-2)} \sum_{i=1}^n Z_i^3$$

where

$$Z_i = \frac{X_i - \bar{X}}{S}$$

For a large sample size, the distinction between the two estimates is negligible. The ratio of the unbiased to the biased estimate is 1,389 for  $n = 10$ , 1,031 for  $n = 100$  and 1,003 for  $n = 1\,000$ .

NOTE 2 Skewness refers to lack of symmetry. Values of this statistic close to zero suggest that the underlying distribution is approximately symmetric, whereas non-zero values would likely correspond to a distribution having occasional extreme values on one side of the centre of the distribution. Skewed data would also be reflected in values of the **sample mean** (1.15) and **sample median** (1.13) that are dissimilar. Positively skewed (right-skewed) data indicate the possible presence of a few extreme, large observations. Similarly, negatively skewed (left-skewed) data indicate the possible presence of a few extreme, small observations.

NOTE 3 The sample coefficient of skewness can be recognized to be the 3<sup>rd</sup> sample moment of the **standardized sample random variables** (1.19).

1.21

**hệ số nhọn của mẫu**

trung bình số học của lũy thừa bậc bốn của biến ngẫu nhiên của mẫu chuẩn hóa (1.19) từ mẫu ngẫu nhiên (1.6).

VÍ DỤ: Tiếp theo ví dụ ở 1.9, hệ số nhọn của mẫu quan trắc được tính bằng 2,674 19. Với cỡ mẫu 10 trong ví dụ này, hệ số nhọn của mẫu biến động lớn, do đó phải sử dụng một cách thận trọng. Sử dụng công thức thay thế trong Chú thích 1, tính được giá trị là 1,349 83. Các bộ chương trình thống kê sử dụng nhiều cách điều chỉnh khác nhau khi tính toán hệ số nhọn của mẫu (xem Chú thích 2 của 2.40). Sử dụng công thức thay thế cho trong Chú thích 1, giá trị tính được là 0,436 05. Hai giá trị 2,674 19 và 0,436 05 không thể so sánh trực tiếp. Muốn so sánh, lấy 2,674 19 – 3 (liên hệ với hệ số nhọn của phân bố chuẩn là 3) được –0,325 81, lúc này có thể so sánh thích hợp với 0,436 05.

CHÚ THÍCH 1: Công thức tương ứng với định nghĩa là

$$\frac{1}{n} \sum_{i=1}^n \left( \frac{X_i - \bar{X}}{S} \right)^4$$

Một số bộ chương trình thống kê sử dụng công thức dưới đây cho hệ số nhọn của mẫu để hiệu chỉnh độ chệch (1.33) và để chỉ thị độ lệch so với hệ số nhọn của phân bố chuẩn (bằng 3):

$$\frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^n Z_i^4 - \frac{3(n-1)^2}{(n-2)(n-3)}$$

trong đó

$$Z_i = \frac{X_i - \bar{X}}{S}$$

Số hạng thứ hai trong biểu thức xấp xỉ 3 đối với cỡ mẫu  $n$  lớn. Đôi khi hệ số nhọn được lấy theo giá trị xác định ở 2.40 trừ đi 3 để nhấn mạnh việc so sánh với phân bố chuẩn. Rõ ràng là người thực hiện cần nhận thức được các điều chỉnh, nếu có, khi tính toán

1.21

**sample coefficient of kurtosis**

arithmetic mean of the fourth power of the standardized sample random variables (1.19) from a random sample (1.6)

EXAMPLE Continuing with the example from 1.9, the observed sample coefficient of kurtosis can be computed to be 2,674 19. For such a sample size as 10 in this example, the sample coefficient of kurtosis is highly variable, so it must be used with caution. Statistical packages use various adjustments in computing the sample coefficient of kurtosis (see Note 3 of 2.40). Using the alternate formula given in Note 1, the computed value is 0,436 05. The two values 2,674 19 and 0,436 05 are not comparable directly. To do so, take 2,674 19 – 3 (to relate to the kurtosis of the normal distribution which is 3) which equals –0,325 81 which now can be appropriately compared to 0,436 05.

NOTE 1 The formula corresponding to the definition is:

$$\frac{1}{n} \sum_{i=1}^n \left( \frac{X_i - \bar{X}}{S} \right)^4$$

Some statistical packages use the following formula for the sample coefficient of kurtosis to correct for bias (1.33) and to indicate the deviation from the kurtosis of the normal distribution (which equals 3):

$$\frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^n Z_i^4 - \frac{3(n-1)^2}{(n-2)(n-3)}$$

where

$$Z_i = \frac{X_i - \bar{X}}{S}$$

The second term in the expression is approximately 3 for large  $n$ . Sometimes the kurtosis is reported as a value as defined in 2.40 minus 3 to emphasize comparisons to the normal distribution. Obviously, a practitioner needs to be aware of the adjustments, if

bộ chương trình thống kê.

**CHÚ THÍCH 2:** Hệ số nhọn đề cập đến tính chất nặng đuôi của một phân bố (một mốt). Đối với phân bố chuẩn (2.50), hệ số nhọn mẫu xấp xỉ bằng 3, tùy thuộc vào độ biến động mẫu. Trên thực tế, hệ số nhọn của phân bố chuẩn cung cấp một mốc hoặc giá trị cơ sở. Những phân bố (2.11) có hệ số nhọn nhỏ hơn 3 có đuôi nhẹ hơn so với phân bố chuẩn, những phân bố có hệ số nhọn lớn hơn 3 có đuôi nặng hơn phân bố chuẩn.

**CHÚ THÍCH 3:** Đối với các giá trị quan trắc được của hệ số nhọn lớn hơn 3 rất nhiều, có khả năng phân bố đang xét có đuôi nặng hơn nhiều so với phân bố chuẩn. Một khả năng khác có thể tìm ra sự có mặt của những giá trị bất thường có thể có.

**CHÚ THÍCH 4:** Hệ số nhọn của mẫu có thể coi là mômen mẫu thứ tư của biến ngẫu nhiên của mẫu chuẩn.

## 1.22

### Hiệp phương sai mẫu

$S_{XY}$

Tổng các tích độ lệch của cặp biến ngẫu nhiên (2.10) trong một mẫu ngẫu nhiên (1.6) so với các trung bình mẫu (1.15) chia cho số các số hạng trong tổng đó trừ đi một.

**VÍ DỤ 1:** Xét minh họa bằng số dưới đây, sử dụng 10 bộ ba giá trị quan trắc. Đối với ví dụ này, chỉ xét  $x$  và  $y$ .

**Bảng 1 – Kết quả dùng cho ví dụ 1**

$i$	1	2	3	4	5	6	7	8	9	10
$x$	38	41	24	60	41	51	58	50	65	33
$y$	73	74	43	107	65	73	99	72	100	48
$z$	34	31	40	28	35	28	32	27	27	31

Trung bình mẫu quan trắc đối với  $X$  là 46,1 và đối với  $Y$  là 75,4. Hiệp phương sai mẫu bằng

$$[(38 - 46,1) \times (73 - 75,4) + (41 - 46,1) \times (74 - 75,4) + \dots + (33 - 46,1) \times (48 - 75,4)]/9 = 257,178$$

any, in statistical package computations.

**NOTE 2** Kurtosis refers to the heaviness of the tails of a (unimodal) distribution. For the normal distribution (2.50), the sample coefficient of kurtosis is approximately 3, subject to sampling variability. In practice, the kurtosis of the normal distribution provides a benchmark or baseline value. Distributions (2.11) with values smaller than 3 have lighter tails than the normal distribution; distributions with values larger than 3 have heavier tails than the normal distribution.

**NOTE 3** For observed values of kurtosis much larger than 3, the possibility exists that the underlying distribution has genuinely heavier tails than the normal distribution. Another possibility to be investigated is the presence of potential outliers.

**NOTE 4** The sample coefficient of kurtosis can be recognized to be the 4th sample moment of the standardized sample random variables.

## 1.22

### sample covariance

$S_{XY}$

sum of products of deviations of pairs of random variables (2.10) in a random sample (1.6) from their sample means (1.15) divided by the number of terms in the sum minus one

**EXAMPLE 1** Consider the following numerical illustration using 10 observed 3-tuples (triplets) of values. For this example, consider only  $x$  and  $y$ .

**Table 1 – Results for Example 1**

$i$	1	2	3	4	5	6	7	8	9	10
$x$	38	41	24	60	41	51	58	50	65	33
$y$	73	74	43	107	65	73	99	72	100	48
$z$	34	31	40	28	35	28	32	27	27	31

The observed sample mean for  $X$  is 46,1 and for  $Y$  is 75,4. The sample covariance is equal to

$$[(38 - 46,1) \times (73 - 75,4) + (41 - 46,1) \times (74 - 75,4) + \dots + (33 - 46,1) \times (48 - 75,4)]/9 = 257,178$$

VÍ DỤ 2: Trong bảng của ví dụ trên, chỉ xét  $y$  và  $z$ . Trung bình mẫu quan trắc đối với  $Z$  là 31,3. Hiệp phương sai mẫu bằng

$$[(73 - 75,4) \times (34 - 31,3) + (74 - 75,4) \times (74 - 31,3) + \dots + (48 - 75,4) \times (31 - 31,3)]/9 = -54,356$$

CHÚ THÍCH 1: Coi như một thống kê (1.8), hiệp phương sai mẫu là hàm số của các cặp biến ngẫu nhiên  $[(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)]$  từ mẫu ngẫu nhiên cỡ  $n$  theo nghĩa nêu trong Chú thích 3 của 1.6. Ước lượng (1.12) này cần được phân biệt với trị số của hiệp phương sai mẫu tính được từ các cặp giá trị đơn vị mẫu (1.2) quan trắc  $[(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)]$  trong mẫu ngẫu nhiên. Trị số này được gọi là hiệp phương sai mẫu thực nghiệm hoặc hiệp phương sai mẫu quan trắc.

CHÚ THÍCH 2: Hiệp phương sai mẫu  $S_{XY}$  được cho bằng:

$$\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$$

CHÚ THÍCH 3: Việc sử dụng mẫu số bằng  $n - 1$  cung cấp ước lượng không chệch (1.34) của hiệp phương sai tổng thể (2.43).

CHÚ THÍCH 4: Ví dụ trong Bảng 1 gồm có ba biến trong khi định nghĩa đề cập đến các cặp biến. Trên thực tế, thường gặp phải những tình huống có nhiều biến.

### 1.23

#### Hệ số tương quan mẫu

$r_{xy}$

Hiệp phương sai mẫu (1.22) chia cho tích các độ lệch chuẩn mẫu (1.17) tương ứng.

VÍ DỤ 1: Tiếp theo Ví dụ 1 của 1.22, độ lệch chuẩn quan trắc là 12,945 đối với  $X$  và 21,329 đối với  $Y$ . Do đó, hệ số tương quan của mẫu quan trắc (với  $X$  và  $Y$ ) được cho bằng:

$$257,118/(12,948 \times 21,329) = 0,9312$$

EXAMPLE 2 In the table of the previous example, consider only  $y$  and  $z$ . The observed sample mean for  $Z$  is 31,3. The sample covariance is equal to

$$[(73 - 75,4) \times (34 - 31,3) + (74 - 75,4) \times (74 - 31,3) + \dots + (48 - 75,4) \times (31 - 31,3)]/9 = -54,356$$

NOTE 1 Considered as a statistic (1.8) the sample covariance is a function of pairs of random variables  $[(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)]$  from a random sample of size  $n$  in the sense given in Note 3 of 1.6. This estimator (1.12) needs to be distinguished from the numerical value of the sample covariance calculated from the observed pairs of values of the sampling units (1.2)  $[(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)]$  in the random sample. This numerical value is called the empirical sample covariance or the observed sample covariance.

NOTE 2 The sample covariance  $S_{XY}$  is given as:

$$\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$$

NOTE 3 Using  $n - 1$  provides an unbiased estimator (1.34) of the population covariance (2.43).

NOTE 4 The example in Table 1 consists of three variables whereas the definition refers to a pair of variables. In practice, it is common to encounter situations with multiple variables.

### 1.23

#### sample correlation coefficient

$r_{xy}$

sample covariance (1.22) divided by the product of the corresponding sample standard deviations (1.17)

EXAMPLE 1 Continuing with Example 1 of 1.22, the observed standard deviation is 12,945 for  $X$  and 21,329 for  $Y$ . Hence, the observed sample correlation coefficient (for  $X$  and  $Y$ ) is given by:

VÍ DỤ 2: Tiếp theo Ví dụ 2 của 1.22, độ lệch chuẩn quan trắc là 21,329 đối với  $Y$  và 4,165 đối với  $Z$ . Do đó, hệ số tương quan của mẫu quan trắc (với  $Y$  và  $Z$ ) được cho bằng:

$$-54,356/(21,329 \times 4,165) = -0,612$$

CHÚ THÍCH 1: Hệ số tương quan mẫu được tính theo công thức:

$$\frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

Biểu thức này tương đương với tỷ số giữa hiệp phương sai mẫu với căn bậc hai của tích các độ lệch chuẩn. Đôi khi, kí hiệu  $r_{xy}$  được dùng để chỉ hệ số tương quan mẫu. Hệ số tương quan mẫu quan trắc dựa trên  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ .

CHÚ THÍCH 2: Hệ số tương quan mẫu quan trắc có thể lấy giá trị trong phạm vi  $[-1, 1]$ , với các giá trị gần 1 chỉ ra mối tương quan dương mạnh và các giá trị gần -1 chỉ ra mối tương quan âm mạnh. Các giá trị gần 1 hoặc -1 chỉ ra các điểm gần như nằm trên một đường thẳng.

## 1.24

### Sai số tiêu chuẩn

$$\sigma_{\hat{\theta}}$$

độ lệch chuẩn (2.37) của hàm ước lượng (1.12)

$$\hat{\theta}$$

VÍ DỤ: Nếu trung bình mẫu (1.15) là ước lượng của trung bình (2.35) tổng thể và độ lệch chuẩn của một biến ngẫu nhiên (2.10) là  $\sigma$ , thì sai số tiêu chuẩn của trung bình mẫu là  $\sigma/\sqrt{n}$  trong đó  $n$  là số quan trắc trong mẫu. Ước lượng của sai số tiêu chuẩn là  $S/\sqrt{n}$  trong đó  $S$  là độ lệch chuẩn mẫu (1.17).

$$257,118/(12,948 \times 21,329) = 0,9312$$

EXAMPLE 2 Continuing with Example 2 of 1.22, the observed standard deviation is 21,329 for  $Y$  and 4,165 for  $Z$ . Hence, the observed sample correlation coefficient (for  $Y$  and  $Z$ ) is given by:

$$-54,356/(21,329 \times 4,165) = -0,612$$

NOTE 1 Notationally, the sample correlation coefficient is computed as:

$$\frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

This expression is equivalent to the ratio of the sample covariance to the square root of the product of the standard deviations. Sometimes the symbol  $r_{xy}$  is used to denote the sample correlation coefficient. The observed sample correlation coefficient is based on realizations  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ .

NOTE 2 The observed sample correlation coefficient can take on values in  $[-1, 1]$ , with values near 1 indicating strong positive correlation and values near -1 indicating strong negative correlation. Values near 1 or -1 indicate that the points are nearly on a straight line.

## 1.24

### standard error

$$\sigma_{\hat{\theta}}$$

standard deviation (2.37) of an estimator (1.12)

$$\hat{\theta}$$

EXAMPLE If the sample mean (1.15) is the estimator of the population mean (2.35) and the standard deviation of a single random variable (2.10) is  $\sigma$ , then the standard error of the sample means is  $\sigma/\sqrt{n}$  where  $n$  is the number of observations in the sample. An estimator of the standard error is  $S/\sqrt{n}$  where  $S$  is the sample

CHÚ THÍCH 1: Trong thực tế, sai số tiêu chuẩn cung cấp ước lượng tự nhiên của độ lệch chuẩn của một hàm ước lượng.

CHÚ THÍCH 2: Không có thuật ngữ phụ (nhận biết được) về sai số "phi tiêu chuẩn". Sai số tiêu chuẩn có thể coi như từ viết tắt của "độ lệch chuẩn của ước lượng". Trên thực tế, sai số tiêu chuẩn thường có ý đề cập đến độ lệch chuẩn của trung bình mẫu. Ký hiệu của sai số tiêu chuẩn của trung bình mẫu là  $\sigma_{\bar{x}}$ .

1.25

**Ước lượng khoảng**

Khoảng giới hạn bởi **thống kê** (1.8) giới hạn trên và **thống kê** giới hạn dưới.

CHÚ THÍCH 1: Một trong các điểm cuối có thể là  $+\infty$ ,  $-\infty$  hoặc giới hạn tự nhiên của giá trị **tham số**. Ví dụ, 0 là giới hạn dưới của **ước lượng khoảng** của **phương sai** (2.36) tổng thể. Trong trường hợp như vậy, khoảng thường được đề cập là khoảng một phía.

CHÚ THÍCH 2: Ước lượng khoảng có thể được cho cùng với **ước lượng** (1.36) **tham số** (2.9). Ước lượng khoảng được giả định là chứa **tham số** theo tỷ lệ cơ hội đã định, trong điều kiện lấy mẫu lặp lại hoặc theo nghĩa xác suất nhất định khác.

CHÚ THÍCH 3: Ba loại ước lượng khoảng thông thường là **khoảng tin cậy** (1.28) đối với (các) **tham số**, **khoảng dự đoán** (1.30) đối với các quan trắc tương lai và **khoảng dung sai thống kê** (1.26) về tỷ lệ của **phân bố** (2.11) chứa trong đó.

1.26

**Khoảng dung sai thống kê**

Khoảng xác định từ **mẫu ngẫu nhiên** (1.6) sao cho với mức tin cậy quy định rằng khoảng này có thể phủ ít nhất một tỷ lệ quy định của **tổng thể**

**standard deviation** (1.17).

NOTE 1 In practice, the standard error provides a natural estimate of the standard deviation of an estimator.

NOTE 2 There is no (sensible) complementary term "non-standard" error. Standard error can be viewed as an abbreviation for the expression "standard deviation of an estimator". Commonly, in practice, standard error is implicitly referring to the standard deviation of the sample mean. The notation for the standard error of the sample mean is  $\sigma_{\bar{x}}$ .

1.25

**Interval estimator**

interval, bounded by an upper limit **statistic** (1.8) and a lower limit **statistic**

NOTE 1 One of the end points could be  $+\infty$ ,  $-\infty$  or a natural limit of the value of a **parameter**. For example, 0 is a natural lower limit for an **interval estimator** of the population **variance** (2.36). In such cases, the intervals are commonly referred to as one-sided intervals.

NOTE 2 An interval estimator can be given in conjunction with **parameter** (2.9) **estimation** (1.36). The interval estimator is presumed to contain a parameter on a stated proportion of occasions, under conditions of repeated sampling, or in some other probabilistic sense.

NOTE 3 Three common types of interval estimators include **confidence intervals** (1.28) for parameter(s), **prediction intervals** (1.30) for future observations, and **statistical tolerance intervals** (1.26) on the proportion of a **distribution** (2.11) contained.

1.26

**statistical tolerance interval**

interval determined from a **random sample** (1.6) in such a way that one may have a specified level of confidence that the interval covers at



(1.1) được lấy mẫu.

**CHÚ THÍCH:** Mức tin cậy trong trường hợp này là tỷ lệ của khoảng thiết lập theo cách này trong một thời gian dài ít nhất sẽ chứa tỷ lệ quy định của tổng thể được lấy mẫu.

### 1.27

#### Giới hạn dung sai thống kê

**Thống kê (1.8)** biểu diễn đầu mút của **khoảng dung sai thống kê (1.26)**

**CHÚ THÍCH:** Khoảng dung sai thống kê có thể là  
– một phía (có một trong hai giới hạn cố định ở giới hạn tự nhiên của biến ngẫu nhiên), trong đó có thể có giới hạn dung sai thống kê trên hoặc dưới, hoặc  
– hai phía, trong đó có cả hai giới hạn dung sai thống kê.

Giới hạn tự nhiên của biến ngẫu nhiên có thể cung cấp giới hạn cho giới hạn một phía.

### 1.28

#### Khoảng tin cậy

**Ước lượng khoảng (1.25)** ( $T_0$ ,  $T_1$ ) đối với tham số (2.9)  $\theta$  có các **thống kê (1.8)**  $T_0$  và  $T_1$  là giới hạn khoảng và đảm bảo là  $P[T_0 < \theta < T_1] \geq 1 - \alpha$

**CHÚ THÍCH 1:** Độ tin cậy phản ánh tỷ lệ các trường hợp khoảng tin cậy chứa giá trị tham số thực trong một loạt dài các **mẫu ngẫu nhiên (1.6)** lặp lại trong các điều kiện giống nhau. Khoảng tin cậy không phản ánh **xác suất (2.5)** để khoảng quan trắc chứa giá trị thực của tham số (hoặc chứa hoặc không chứa).

**CHÚ THÍCH 2:** Kèm theo khoảng tin cậy là đặc trưng hiệu năng  $100(1 - \alpha)$  %, trong đó  $\alpha$  thường là một số rất nhỏ. Đặc trưng hiệu năng, còn gọi là hệ số tin cậy hoặc mức tin cậy, thường bằng 95 % hoặc 99 %. Bất đẳng thức  $P[T_0 < \theta < T_1] \geq 1 - \alpha$  đúng với giá trị  $\theta$  cụ thể bất kỳ nhưng chưa biết của tổng thể.

least a specified proportion of the sampled **population (1.1)**

**NOTE** The confidence in this context is the long-run proportion of intervals constructed in this manner that will include at least the specified proportion of the sampled population.

### 1.27

#### statistical tolerance limit

**statistic (1.8)** representing an end point of a **statistical tolerance interval (1.26)**

**NOTE** Statistical tolerance intervals may be either  
– one-sided (with one of its limits fixed at the natural boundary of the random variable), in which case they have either an upper or a lower statistical tolerance limit, or

– two-sided, in which case they have both.

A natural boundary of the random variable may provide a limit for a one-sided limit.

### 1.28

#### confidence interval

**interval estimator (1.25)** ( $T_0$ ,  $T_1$ ) for the **parameter (2.9)**  $\theta$  with the **statistics (1.8)**  $T_0$  and  $T_1$  as interval limits and for which it holds that  $P[T_0 < \theta < T_1] \geq 1 - \alpha$

**NOTE 1** The confidence reflects the proportion of cases that the confidence interval would contain the true parameter value in a long series of repeated **random samples (1.6)** under identical conditions. A confidence interval does not reflect the **probability (2.5)** that the observed interval contains the true value of the parameter (it either does or does not contain it).

**NOTE 2** Associated with this confidence interval is the attendant performance characteristic  $100(1 - \alpha)$  %, where  $\alpha$  is generally a small number. The performance characteristic, which is called the confidence coefficient or confidence level, is often

## 1.29

**Khoảng tin cậy một phía**

**Khoảng tin cậy (1.28)** có một trong hai đầu mút cố định ở  $+\infty$ ,  $-\infty$ , hoặc giới hạn cố định tự nhiên.

CHÚ THÍCH 1: Định nghĩa 1.28 áp dụng với  $T_0$  đặt ở  $-\infty$  hoặc  $T_1$  đặt ở  $+\infty$ . Khoảng tin cậy một phía phát sinh trong trường hợp quan tâm tập trung vào một phía. Ví dụ, trong kiểm tra âm lượng về an toàn liên quan đến điện thoại di động, giới hạn tin cậy trên cần được quan tâm cho biết giới hạn trên đối với âm lượng tạo ra trong các điều kiện an toàn giả định. Đối với phép thử kết cấu cơ khí, giới hạn tin cậy dưới về lực tại đó thiết bị hỏng cần quan tâm.

CHÚ THÍCH 2: Một ví dụ khác về khoảng tin cậy một phía là trường hợp tham số có một giới hạn tự nhiên là "không". Đối với phân bố Poisson (2.47) liên quan đến xử lý khiếu nại của khách hàng thì "không" là giới hạn dưới. Trong một ví dụ khác, khoảng tin cậy đối với độ tin cậy của linh kiện điện tử là (0,98, 1), trong đó 1 là giới hạn trên tự nhiên.

## 1.30

**Khoảng dự đoán**

Dãy giá trị của một biến, tạo ra từ một mẫu ngẫu nhiên (1.6) các giá trị từ một tổng thể liên tục, sao cho có thể đảm bảo với mức tin cậy cho trước rằng, trong một mẫu ngẫu nhiên khác lấy từ chính tổng thể (1.1) đó, không có ít hơn số lượng giá trị cho trước rơi vào khoảng đó.

CHÚ THÍCH: Thông thường, quan tâm tập trung vào một quan trắc đơn phát sinh từ tình huống tương tự như các quan trắc là cơ sở của khoảng dự đoán. Một hướng thực tế khác là phân tích hồi quy, trong đó khoảng dự đoán được thiết lập cho một dãy các giá trị độc lập.

95 % or 99 %. The inequality  $P[T_0 < \theta < T_1] \geq 1 - \alpha$  holds for any specific but unknown population value of  $\theta$ .

## 1.29

**one-sided confidence interval**

**confidence interval (1.28)** with one of its end points fixed at  $+\infty$ ,  $-\infty$ , or a natural fixed boundary

NOTE 1 Definition 1.28 applies with either  $T_0$  set at  $-\infty$  or  $T_1$  set at  $+\infty$ . One-sided confidence intervals arise in situations where interest focuses strictly on one direction. For example, in audio volume testing for safety concerns in cellular telephones, an upper confidence limit would be of interest indicating an upper bound for the volume produced under presumed safe conditions. For structural mechanical testing, a lower confidence limit on the force at which a device fails would be of interest.

NOTE 2 Another instance of one-sided confidence intervals occurs in situations where a parameter has a natural boundary such as zero. For a Poisson distribution (2.47) involved in modelling customer complaints, zero is a lower bound. As another example, a confidence interval for the reliability of an electronic component could be (0,98, 1), where 1 is the natural upper boundary limit.

## 1.30

**prediction interval**

range of values of a variable, derived from a random sample (1.6) of values from a continuous population, within which it can be asserted with a given confidence that no fewer than a given number of values in a further random sample from the same population (1.1) will fall

NOTE Commonly, interest focuses on a single further observation arising from the same situation as the observations which are the basis of the prediction interval. Another practical context is regression analysis in which a prediction interval is constructed for a spectrum of independent values.

## 1.31

**Ước lượng**

**Giá trị quan trắc (1.4) của một hàm ước lượng (1.12).**

**CHÚ THÍCH:** Ước lượng muốn nói đến trị số thu được từ các giá trị quan trắc. Đối với **phép ước lượng (1.36)** của một **tham số (2.9)** từ **phân bố xác suất (2.11)** giả thuyết, hàm ước lượng đề cập đến **thống kê (1.8)** dùng để ước lượng tham số và ước lượng đề cập đến kết quả sử dụng các giá trị quan trắc. Đôi khi, tính từ "điểm" được đặt sau từ ước lượng để nhấn mạnh rằng một giá trị đơn được tạo ra hơn là một khoảng các giá trị. Tương tự, tính từ khoảng được đặt sau từ ước lượng trong trường hợp thực hiện phép ước lượng khoảng.

## 1.32

**Sai số của phép ước lượng**

**Ước lượng (1.31) trừ đi tham số (2.9) hoặc tính chất của tổng thể dự định để ước lượng.**

**CHÚ THÍCH 1:** Tính chất của tổng thể có thể là hàm số của tham số hoặc các tham số hoặc đại lượng khác liên quan đến phân bố xác suất (2.11).

**CHÚ THÍCH 2:** Sai số ước lượng có thể do lấy mẫu, độ không đảm bảo đo, làm tròn số hoặc các nguyên nhân khác. Trong thực tế, sai số ước lượng thể hiện sự quan tâm ở mức thấp nhất của người thực hiện. Việc xác định thành phần chính gây sai số ước lượng mới là yếu tố quan trọng trong nỗ lực cải tiến chất lượng.

## 1.33

**Độ chệch**

**Kỳ vọng (2.12) của sai số của phép ước lượng (1.32)**

**CHÚ THÍCH 1:** Định nghĩa này khác với TCVN 8244-2 (ISO 3534-2) (3.3.2) và TCVN 6165 (VIM) (5.25 và 5.28). Ở đây độ chệch được sử dụng theo nghĩa tổng quát như nêu trong Chú thích 1 của 1.34.

## 1.31

**estimate**

**observed value (1.4) of an estimator (1.12)**

**NOTE** Estimate refers to a numerical value obtained from observed values. With respect to **estimation (1.36)** of a **parameter (2.9)** from an hypothesized **probability distribution (2.11)**, estimator refers to the **statistic (1.8)** intended to estimate the parameter and estimate refers to the result using observed values. Sometimes the adjective "point" is inserted before estimate to emphasize that a single value is being produced rather than an interval of values. Similarly, the adjective "interval" is inserted before estimate in cases where interval estimation is taking place.

## 1.32

**error of estimation**

**estimate (1.31) minus the parameter (2.9) or population property that it is intended to estimate**

**NOTE 1** Population property may be a function of the parameter or parameters or another quantity related to the probability distribution (2.11).

**NOTE 2** Estimator error could involve contributions due to sampling, measurement uncertainty, rounding, or other sources. In effect, estimator error represents the bottom line performance of interest to practitioners. Determining the primary contributors to estimator error is a critical element in quality improvement efforts.

## 1.33

**bias**

**expectation (2.12) of error of estimation (1.32)**

**NOTE 1** This definition differs from TCVN 8244-2 (ISO 3534-2) (3.3.2) and TCVN 6165 (VIM) (5.25 and 5.28). Here bias is used in a generic sense as indicated in Note 1 in 1.34.

**NOTE 2** The existence of bias can lead to

CHÚ THÍCH 2: Sự có mặt của độ chệch có thể dẫn đến những hậu quả đáng tiếc trong thực tế. Ví dụ, việc ước lượng sai độ bền của vật liệu có thể dẫn đến hỏng thiết bị. Trong lấy mẫu điều tra, độ chệch có thể dẫn đến những quyết định sai của cuộc thăm dò chính trị.

### 1.34

#### Ước lượng không chệch

Ước lượng (1.12) có độ chệch (1.33) bằng không

VÍ DỤ 1: Đối với mẫu ngẫu nhiên (1.6) có  $n$  biến ngẫu nhiên (2.10) độc lập, mỗi biến có cùng phân bố chuẩn (2.50) có trung bình (2.35)  $\mu$  và độ lệch chuẩn (2.37)  $\sigma$ , thì trung bình mẫu  $\bar{X}$  (1.15) và phương sai mẫu (1.16)  $S^2$  là các ước lượng không chệch đối với trung bình  $\mu$  và phương sai (2.36)  $\sigma^2$ , tương ứng.

VÍ DỤ 2: Như đề cập trong Chú thích 1 của 1.37 ước lượng hợp lý cực đại (1.35) của phương sai  $\sigma^2$  sử dụng mẫu số  $n$  thay cho  $n - 1$  và do đó là ước lượng chệch. Trong các ứng dụng, độ lệch chuẩn mẫu (1.17) được sử dụng nhiều nhưng điều quan trọng là lưu ý rằng căn bậc hai của phương sai mẫu sử dụng  $n - 1$  là một ước lượng chệch của độ lệch chuẩn (2.37) của tổng thể.

VÍ DỤ 3: Đối với mẫu ngẫu nhiên có  $n$  cặp biến ngẫu nhiên độc lập, mỗi cặp có cùng phân bố chuẩn hai chiều (2.65) có hiệp phương sai (2.43) bằng  $\rho\sigma_X\sigma_Y$ , thì hiệp phương sai mẫu (1.22) là một ước lượng không chệch đối với hiệp phương sai của tổng thể. Ước lượng hợp lý cực đại dùng  $n$  thay cho  $n - 1$  trong mẫu số và do đó là ước lượng chệch.

CHÚ THÍCH: Ước lượng không chệch được mong muốn theo nghĩa về trung bình, chúng cho giá trị chính xác. Chắc chắn là ước lượng không chệch cung cấp điểm bắt đầu hữu ích trong việc tìm kiếm ước lượng "tối ưu" các tham số của tổng thể. Định nghĩa nêu ở đây là định nghĩa có tính chất thống kê.

Trong ứng dụng hàng ngày, người thực hiện cố gắng

unfortunate consequences in practice. For example, underestimation of the strength of materials due to bias could lead to unexpected failures of a device. In survey sampling, bias could lead to incorrect decisions from a political poll.

### 1.34

#### unbiased estimator

estimator (1.12) having bias (1.33) equal to zero

EXAMPLE 1 For a random sample (1.6) of  $n$  independent random variables (2.10), each with the same normal distribution (2.50) with mean (2.35)  $\mu$  and standard deviation (2.37)  $\sigma$ , the sample mean  $\bar{X}$  (1.15) and the sample variance (1.16)  $S^2$  are unbiased estimators for the mean  $\mu$  and the variance (2.36)  $\sigma^2$ , respectively.

EXAMPLE 2 As is mentioned in Note 1 to 1.37 the maximum likelihood estimator (1.35) of the variance  $\sigma^2$  uses the denominator  $n$  instead of  $n - 1$  and thus is a biased estimator. In applications, the sample standard deviation (1.17) receives considerable use but it is important to note that the square root of the sample variance using  $n - 1$  is a biased estimator of the population standard deviation (2.37).

EXAMPLE 3 For a random sample of  $n$  independent pairs of random variables, each pair with the same bivariate normal distribution (2.65) with covariance (2.43) equal to  $\rho\sigma_X\sigma_Y$ , the sample covariance (1.22) is an unbiased estimator for population covariance. The maximum likelihood estimator uses  $n$  instead of  $n - 1$  in the denominator and thus is biased.

NOTE Estimators that are unbiased are desirable in that on average, they give the correct value. Certainly, unbiased estimators provide a useful starting point in the search for "optimal" estimators of population parameters. The definition given here is of a statistical nature.

tránh tạo ra độ chệch cho nghiên cứu bằng cách, ví dụ, đảm bảo rằng mẫu ngẫu nhiên là đại diện của tổng thể quan tâm.

### 1.35

#### Ước lượng hợp lý cực đại

Ước lượng (1.12) ấn định giá trị của tham số (2.9) tại đó hàm hợp lý (1.38) đạt được hoặc tới gần giá trị lớn nhất của nó.

CHÚ THÍCH 1: Ước lượng hợp lý cực đại là phương pháp tốt để thu được tham số ước lượng khi phân bố (2.11) đã được quy định [ví dụ, phân bố chuẩn (2.50), phân bố gamma (2.56), phân bố Weibull (2.63), v.v...]. Các hàm ước lượng này có tính chất thống kê mong muốn (ví dụ, phép biến đổi đơn điệu) và trong nhiều trường hợp cung cấp phương pháp chọn phép ước lượng. Trong các trường hợp hàm ước lượng hợp lý cực đại là chệch, đôi khi có sự hiệu chỉnh đơn giản độ chệch (1.33). Như đề cập trong ví dụ 2 của 1.34, ước lượng hợp lý cực đại đối với phương sai (2.36) của phân bố chuẩn là ước lượng chệch nhưng có thể hiệu chỉnh bằng cách sử dụng  $n - 1$  thay cho  $n$ . Phạm vi độ chệch trong những trường hợp như vậy giảm khi cỡ mẫu tăng.

CHÚ THÍCH 2: Từ viết tắt MLE thường được sử dụng cho cả hàm ước lượng hợp lý cực đại và phép ước lượng hợp lý cực đại với ngữ cảnh chỉ ra sự lựa chọn thích hợp.

### 1.36

#### Phép ước lượng

Quy trình thu được đại diện thống kê của tổng thể (1.1) từ mẫu ngẫu nhiên (1.6) lấy từ tổng thể này.

CHÚ THÍCH 1: Đặc biệt, quy trình này liên quan đến tiến trình từ một hàm ước lượng (1.12) đến một ước lượng cụ thể (1.31) cấu thành phép ước lượng.

CHÚ THÍCH 2: Phép ước lượng được hiểu theo nghĩa rộng hơn để bao gồm cả ước lượng điểm, ước lượng khoảng hoặc ước lượng tính chất của tổng thể.

In everyday usage, practitioners try to avoid introducing bias into a study by ensuring, for example, that the random sample is representative of the population of interest.

### 1.35

#### maximum likelihood estimator

estimator (1.12) assigning the value of the parameter (2.9) where the likelihood function (1.38) attains or approaches its highest value

NOTE 1 Maximum likelihood estimation is a well-established approach for obtaining parameter estimates where a distribution (2.11) has been specified [for example, normal (2.50), gamma (2.56), Weibull (2.63), and so forth]. These estimators have desirable statistical properties (for example, invariance under monotone transformation) and in many situations provide the estimation method of choice. In cases in which the maximum likelihood estimator is biased, a simple bias (1.33) correction sometimes takes place. As mentioned in Example 2 of 1.34 the maximum likelihood estimator for the variance (2.36) of the normal distribution is biased but it can be corrected by using  $n - 1$  rather than  $n$ . The extent of the bias in such cases decreases with increasing sample size.

NOTE 2 The abbreviation MLE is commonly used both for maximum likelihood estimator and maximum likelihood estimation with the context indicating the appropriate choice.

### 1.36

#### estimation

procedure that obtains a statistical representation of a population (1.1) from a random sample (1.6) drawn from this population

NOTE 1 In particular, the procedure involved in progressing from an estimator (1.12) to a specific estimate (1.31) constitutes estimation.

NOTE 2 Estimation is understood in a rather broad context to include point estimation, interval

CHÚ THÍCH 3: Thông thường, đại diện thống kê đề cập đến phép ước lượng một tham số (2.9) hoặc các tham số hoặc hàm của tham số từ một mô hình giả định. Tổng quát hơn, đại diện của tổng thể có thể ít cụ thể hơn, ví dụ như các thống kê liên quan đến tác động của các thảm họa tự nhiên (chết, bị thương, mất tài sản và tổn thất về nông nghiệp – tất cả những điều mà nhà quản lý tình trạng khẩn cấp mong muốn ước lượng).

CHÚ THÍCH 4: Việc xem xét thống kê mô tả (1.5) có thể gợi ý rằng mô hình giả định cung cấp đại diện không đầy đủ các dữ liệu, như được chỉ ra nhờ thước đo sự phù hợp của mô hình đó đối với các dữ liệu. Trong trường hợp như vậy, cần xem xét các mô hình khác và quá trình ước lượng được tiếp tục.

### 1.37

#### Phép ước lượng hợp lý cực đại

Phép ước lượng (1.36) dựa trên hàm ước lượng hợp lý cực đại (1.35)

CHÚ THÍCH 1: Đối với phân bố chuẩn (2.50), trung bình mẫu (1.15) là hàm ước lượng hợp lý cực đại (1.35) của tham số (2.9)  $\mu$  trong khi phương sai mẫu (1.16), sử dụng mẫu số  $n$  thay cho là  $n - 1$ , cung cấp hàm ước lượng hợp lý cực đại  $\sigma^2$ . Mẫu số  $n - 1$  thường được dùng vì giá trị này cung cấp ước lượng không chệch (1.34).

CHÚ THÍCH 2: Phép ước lượng hợp lý cực đại đôi khi được dùng để mô tả dẫn xuất của hàm ước lượng (1.12) từ hàm hợp lý.

CHÚ THÍCH 3: Mặc dù trong một số trường hợp, cho ra biểu thức dạng hiển việc sử dụng phép ước lượng hợp lý cực đại rõ ràng minh, nhưng vẫn có các trường hợp khác, trong đó hàm ước lượng hợp lý cực đại đòi hỏi việc giải lặp đối với một tập hợp các phương trình.

CHÚ THÍCH 4: Từ viết tắt MLE thường được sử dụng cho cả hàm ước lượng hợp lý cực đại và phép ước lượng hợp lý cực đại theo ngữ cảnh chỉ ra sự

estimation or estimation of properties of populations.

NOTE 3 Frequently, a statistical representation refers to the estimation of a parameter (2.9) or parameters or a function of parameters from an assumed model. More generally, the representation of the population could be less specific, such as statistics related to impacts from natural disasters (casualties, injuries, property losses and agricultural losses – all of which an emergency manager might wish to estimate).

NOTE 4 Consideration of descriptive statistics (1.5) could suggest that an assumed model provides an inadequate representation of the data, such as indicated by a measure of the goodness of fit of the model to the data. In such cases, other models could be considered and the estimation process continued.

### 1.37

#### maximum likelihood estimation

estimation (1.36) based upon the maximum likelihood estimator (1.35)

NOTE 1 For the normal distribution (2.50), the sample mean (1.15) is the maximum likelihood estimator (1.35) of the parameter (2.9)  $\mu$  while the sample variance (1.16), using the denominator  $n$  rather than  $n - 1$ , provides the maximum likelihood estimator of  $\sigma^2$ . The denominator  $n - 1$  is typically used since this value provides an unbiased estimator (1.34).

NOTE 2 Maximum likelihood estimation is sometimes used to describe the derivation of an estimator (1.12) from the likelihood function.

NOTE 3 Although in some cases, a closed-form expression emerges using maximum likelihood estimation, there are other situations in which the maximum likelihood estimator requires an iterative solution to a set of equations.

NOTE 4 The abbreviation MLE is commonly used both for maximum likelihood estimator and maximum likelihood estimation with the context indicating

lựa chọn thích hợp.

### 1.38

#### Hàm hợp lý

Hàm mật độ xác suất (2.26) đánh giá tại các giá trị quan trắc (1.4) và được coi là hàm số của các tham số (2.9) của họ phân bố (2.8) đó.

VÍ DỤ 1: Xét trường hợp mười cá thể được chọn ngẫu nhiên từ một tổng thể (1.1) rất lớn và 3 trong số đó có đặc trưng riêng. Từ mẫu này, ước lượng (1.31) trực giác của tỷ lệ tổng thể có đặc trưng là 0,3 (3 trong số 10). Trong mô hình phân bố nhị phân (2.46), hàm xác suất (hàm khối lượng xác suất như hàm của  $p$  với  $n$  cố định tại 10 và  $x$  tại 3) đạt giá trị lớn nhất tại  $p = 0,3$ , vì vậy thống nhất với trực giác.

[Điều này có thể xác nhận thêm bằng cách vẽ hàm khối lượng xác suất của phân bố nhị phân (2.46)  $120 p^3 (1 - p)^7$  theo  $p$ .]

VÍ DỤ 2: Đối với phân bố chuẩn (2.50) đã biết độ lệch chuẩn (2.37), nói chung có thể chứng tỏ rằng hàm xác suất có giá trị lớn nhất tại  $\mu$  bằng trung bình mẫu.

### 1.39

#### Hàm hợp lý biên duyên

Hàm xác suất (1.38) như hàm của một tham số (2.9) duy nhất với tất cả các tham số khác được đặt để hàm đó đạt cực đại

### 1.40

#### Giả thuyết

$H$

Nhận định về tổng thể (1.1).

CHÚ THÍCH: Thông thường, nhận định về tổng thể liên quan đến một hoặc nhiều tham số (2.9) trong một họ phân bố (2.8) hoặc về họ phân bố đó.

the appropriate choice.

### 1.38

#### likelihood function

probability density function (2.26) evaluated at the observed values (1.4) and considered as a function of the parameters (2.9) of the family of distributions (2.8)

EXAMPLE 1 Consider a situation in which ten items are selected at random from a very large population (1.1) and 3 of the items are found to have a specific characteristic. From this sample, an intuitive estimate (1.31) of the population proportion having the characteristic is 0,3 (3 out of 10). Under a binomial distribution (2.46) model, the likelihood function (probability mass function as a function of  $p$  with  $n$  fixed at 10 and  $x$  at 3) achieves its maximum at  $p = 0,3$ , thus agreeing with intuition.

[This can be further verified by plotting the probability mass function of the binomial distribution (2.46)  $120 p^3 (1 - p)^7$  versus  $p$ .]

EXAMPLE 2 For the normal distribution (2.50) with known standard deviation (2.37), it can be shown in general that the likelihood function takes its maximum at  $\mu$  equal to the sample mean.

### 1.39

#### profile likelihood function

likelihood function (1.38) as a function of a single parameter (2.9) with all other parameters set to maximize it

### 1.40

#### hypothesis

$H$

statement about a population (1.1)

NOTE Commonly the statement about the population concerns one or more parameters (2.9) in a family of distributions (2.8) or about the family of distributions.

## 1.41

**Giả thuyết không** $H_0$ 

**Giả thuyết** (1.40) cần được kiểm nghiệm bằng các **phép kiểm nghiệm thống kê** (1.48).

VÍ DỤ 1: Trong một **mẫu ngẫu nhiên** (1.6) các biến **ngẫu nhiên** (2.10) độc lập có cùng **phân bố chuẩn** (2.50) với **trung bình** (2.35) và **độ lệch chuẩn** (2.37) chưa biết, giả thuyết không đối với trung bình  $\mu$  có thể là trung bình này nhỏ hơn hoặc bằng giá trị  $\mu_0$  đã cho và điều này thường được viết theo cách sau:

$H_0: \mu \leq \mu_0$ .

VÍ DỤ 2: Giả thuyết không có thể là mô hình thống kê đối với một **tổng thể** (1.1) là **phân bố chuẩn**. Đối với loại giả thuyết không này, **trung bình** và **độ lệch chuẩn** không được cho trước.

VÍ DỤ 3: Giả thuyết không có thể là mô hình thống kê đối với **tổng thể** gồm **phân bố đối xứng**. Đối với loại giả thuyết không này, **dạng phân bố** không được cho trước.

**CHÚ THÍCH 1:** Rõ ràng là giả thuyết không có thể bao gồm một tập con từ một tập hợp các phân bố xác suất có thể.

**CHÚ THÍCH 2:** Định nghĩa này không nên coi là tách biệt với **đối giả thuyết** (1.42) và **phép kiểm nghiệm thống kê** (1.48), vì việc áp dụng đúng kiểm nghiệm giả thuyết đòi hỏi tất cả các thành phần này.

**CHÚ THÍCH 3:** Trong thực tế, ta không bao giờ chứng minh giả thuyết không nhưng việc đánh giá trong tình huống đặt ra có thể không thỏa đáng để bác bỏ giả thuyết không. Động cơ ban đầu của việc tiến hành kiểm nghiệm giả thuyết bắt nguồn từ mong muốn rằng kết luận sẽ nghiêng về đối giả thuyết cụ thể liên quan đến vấn đề được xét.

**CHÚ THÍCH 4:** Việc không bác bỏ giả thuyết không không phải là "chứng minh" về hiệu lực của nó mà có thể chỉ ra rằng không có đủ bằng chứng để phân bác. Hoặc giả thuyết không (hoặc gần giống) thực tế là

## 1.41

**null hypothesis** $H_0$ 

**hypothesis** (1.40) to be tested by means of a **statistical test** (1.48)

**EXAMPLE 1** In a **random sample** (1.6) of independent **random variables** (2.10) with the same **normal distribution** (2.50) with unknown **mean** (2.35) and unknown **standard deviation** (2.37), a null hypothesis for the mean  $\mu$  may be that the mean is less than or equal to a given value  $\mu_0$  and this is usually written in the following way:  $H_0: \mu \leq \mu_0$ .

**EXAMPLE 2** A null hypothesis may be that the statistical model for a **population** (1.1) is a normal distribution. For this type of null hypothesis, the mean and standard deviation are not specified.

**EXAMPLE 3** A null hypothesis may be that the statistical model for a population consists of a symmetric distribution. For this type of null hypothesis, the form of the distribution is not specified.

**NOTE 1** Explicitly, the null hypothesis can consist of a subset from a set of possible probability distributions.

**NOTE 2** This definition should not be considered in isolation from **alternative hypothesis** (1.42) and **statistical test** (1.48), as proper application of hypothesis testing requires all of these components.

**NOTE 3** In practice, one never proves the null hypothesis, but rather the assessment in a given situation may be inadequate to reject the null hypothesis. The original motivation for conducting the hypothesis test would likely have been an expectation that the outcome would favour a specific alternative hypothesis relevant to the problem at hand.

**NOTE 4** Failure to reject the null hypothesis is not "proof" of its validity but may rather be an



đúng hoặc cỡ mẫu không đủ để phát hiện sự khác biệt so với giả thuyết.

CHÚ THÍCH 5: Trong nhiều tình huống, quan tâm ban đầu tập trung vào giả thuyết không nhưng khả năng sai khác cũng có thể được quan tâm. Sự xem xét thích hợp cỡ mẫu và hiệu lực trong việc phát hiện sai lệch hoặc thay đổi cụ thể có thể dẫn đến việc thiết lập một quy trình kiểm nghiệm để đánh giá một cách thích hợp giả thuyết không.

CHÚ THÍCH 6: Việc chấp nhận đối giả thuyết trái ngược với việc không bác bỏ giả thuyết không là một kết quả tích cực ở chỗ nó hỗ trợ mỗi quan tâm phỏng đoán. Việc bác bỏ giả thuyết không, ủng hộ đối giả thuyết là kết quả rõ ràng hơn kết quả như "không bác bỏ giả thuyết không tại thời điểm này".

CHÚ THÍCH 7: Giả thuyết không là cơ sở để thiết lập **thống kê kiểm nghiệm** (1.52) tương ứng dùng để đánh giá giả thuyết không.

CHÚ THÍCH 8: Giả thuyết không thường được biểu thị là  $H_0$ .

CHÚ THÍCH 9: Nếu có thể, tập con nhận biết giả thuyết không cần được chọn sao cho nhận định không phù hợp với phỏng đoán cần nghiên cứu. Xem chú thích 2 của 1.48 và ví dụ trong 1.49.

indication that there is insufficient evidence to dispute it. Either the null hypothesis (or a close proximity to it) is in fact true, or the sample size is insufficient to detect a difference from it.

NOTE 5 In some situations, initial interest is focused on the null hypothesis, but the possibility of a departure may be of interest. Proper consideration of sample size and power in detecting a specific departure or alternative can lead to the construction of a test procedure for appropriately assessing the null hypothesis.

NOTE 6 The acceptance of the alternative hypothesis in contrast to failing to reject the null hypothesis is a positive result in that it supports the conjecture of interest. Rejection of the null hypothesis in favour of the alternative is an outcome with less ambiguity than an outcome such as "failure to reject the null hypothesis at this time."

NOTE 7 The null hypothesis is the basis for constructing the corresponding **test statistic** (1.52) used to assess the null hypothesis.

NOTE 8 The null hypothesis is often denoted  $H_0$  ( $H$  having a subscript of zero although the zero is sometimes pronounced "oh" or "nought").

NOTE 9 The subset identifying the null hypothesis should, if possible, be selected in such a way that the statement is incompatible with the conjecture to be studied. See Note 2 to 1.48 and the example in 1.49.

## 1.42

### Đối giả thuyết

$H_A, H_1$

Nhận định chọn một tập hợp hoặc một tập con tất cả các **phân bố xác suất** (2.11) có thể chấp nhận được không nằm trong **giả thuyết không** (1.41).

VÍ DỤ 1: Đối giả thuyết của giả thuyết không nêu trong ví dụ 1 của 1.41 là **trung bình** (2.35) lớn hơn giá trị quy định, được viết theo cách sau:  $H_A: \mu > \mu_0$ .

VÍ DỤ 2: Đối giả thuyết của giả thuyết không nêu

## 1.42

### alternative hypothesis

$H_A, H_1$

statement which selects a set or a subset of all possible admissible **probability distributions** (2.11) which do not belong to the **null hypothesis** (1.41)

EXAMPLE 1 The alternative hypothesis to the null hypothesis given in Example 1 of 1.41 is that the **mean** (2.35) is larger than the specified value, which

## TCVN 8244-1:2010

trong ví dụ 2 của 1.41 là mô hình thống kê của tổng thể không phải là phân bố chuẩn (2.50).

VÍ DỤ 3: Đối giả thuyết của giả thuyết không nêu trong ví dụ 3 của 1.41 là mô hình thống kê của tổng thể gồm một phân bố không đối xứng. Đối với đối giả thuyết này, dạng không đối xứng cụ thể không được quy định.

CHÚ THÍCH 1: Đối giả thuyết là phần bù của giả thuyết không.

CHÚ THÍCH 2: Đối giả thuyết có thể được biểu thị bằng  $H_1$  hoặc  $H_A$  mà không có sự ưu tiên rõ ràng miễn là cách kí hiệu tương đương với kí hiệu của giả thuyết không.

CHÚ THÍCH 3: Đối giả thuyết là nhận định trái ngược với giả thuyết không. Thống kê kiểm nghiệm (1.52) tương ứng được dùng để quyết định giữa giả thuyết không và đối giả thuyết.

CHÚ THÍCH 4: Không nên tách biệt đối giả thuyết với giả thuyết không cũng như phép kiểm nghiệm thống kê (1.48).

CHÚ THÍCH 5: Việc chấp nhận đối giả thuyết trái với việc không bác bỏ giả thuyết không là một kết quả tích cực ở chỗ nó hỗ trợ mỗi quan tâm phỏng đoán.

### 1.43

#### Giả thuyết đơn

Giả thuyết (1.40) quy định một phân bố duy nhất trong một họ phân bố (2.8).

CHÚ THÍCH 1: Giả thuyết đơn là giả thuyết không (1.41) hoặc đối giả thuyết (1.42) trong đó tập con được chọn chỉ gồm một phân bố xác suất (2.11) duy nhất.

CHÚ THÍCH 2: Trong một mẫu ngẫu nhiên (1.6) các biến ngẫu nhiên (2.10) độc lập có cùng phân bố chuẩn (2.50) với trung bình (2.35) chưa biết và độ

is written in the following way:  $H_A: \mu > \mu_0$ .

EXAMPLE 2 The alternative hypothesis to the null hypothesis given in Example 2 of 1.41 is that the statistical model of the population is not a normal distribution (2.50).

EXAMPLE 3 The alternative hypothesis to the null hypothesis given in Example 3 of 1.41 is that the statistical model of the population consists of an asymmetric distribution. For this alternative hypothesis, the specific form of asymmetry is not specified.

NOTE 1 The alternative hypothesis is the complement of the null hypothesis.

NOTE 2 The alternative hypothesis can also be denoted  $H_1$  or  $H_A$  with no clear preference as long as the symbolism parallels the null hypothesis notation.

NOTE 3 The alternative hypothesis is a statement which contradicts the null hypothesis. The corresponding test statistic (1.52) is used to decide between the null and alternative hypotheses.

NOTE 4 The alternative hypothesis should not be considered in isolation from the null hypothesis nor statistical test (1.48).

NOTE 5 The acceptance of the alternative hypothesis in contrast to failing to reject the null hypothesis is a positive result in that it supports the conjecture of interest.

### 1.43

#### simple hypothesis

hypothesis (1.40) that specifies a single distribution in a family of distributions (2.8)

NOTE 1 A simple hypothesis is a null hypothesis (1.41) or alternative hypothesis (1.42) for which the selected subset consists of only a single probability distribution (2.11).

NOTE 2 In a random sample (1.6) of independent random variables (2.10) with the same normal distribution (2.50) with unknown mean (2.35) and

lệch chuẩn (2.37)  $\sigma$  đã biết, giả thuyết đơn đối với trung bình  $\mu$  là trung bình bằng giá trị  $\mu_0$  đã cho và điều này thường được viết như sau:  $H_0: \mu = \mu_0$ .

CHÚ THÍCH 3: Giả thuyết đơn quy định được hoàn toàn phân bố xác suất (2.11).

#### 1.44

##### Giả thuyết hợp

Giả thuyết (1.40) quy định nhiều hơn một phân bố (2.11) trong một họ phân bố (2.8).

VÍ DỤ 1: Giả thuyết không (1.41) và đối giả thuyết (1.42) nêu trong ví dụ ở 1.41 và 1.42 đều là ví dụ của giả thuyết hợp.

VÍ DỤ 2: Trong 1.48, giả thuyết không trong Trường hợp 3 của Ví dụ 3 là một giả thuyết đơn. Giả thuyết không trong Ví dụ 4 cũng là một giả thuyết đơn. Các giả thuyết khác trong 1.48 là giả thuyết hợp.

CHÚ THÍCH: Giả thuyết hợp là một giả thuyết không hoặc đối giả thuyết trong đó tập con được chọn gồm nhiều phân bố xác suất.

#### 1.45

##### Mức ý nghĩa

$\alpha$

xác suất (2.5) lớn nhất <của phép kiểm nghiệm thống kê> bác bỏ giả thuyết không (1.41) trong khi trên thực tế giả thuyết không là đúng.

CHÚ THÍCH: Nếu giả thuyết không là một giả thuyết đơn (1.43), thì xác suất bác bỏ giả thuyết không khi nó đúng sẽ là một giá trị duy nhất.

#### 1.46

##### Sai lầm loại I

Bác bỏ giả thuyết không (1.41) trong khi trên thực tế giả thuyết không là đúng.

CHÚ THÍCH 1: Trong thực tế, sai lầm loại I là một

known standard deviation (2.37)  $\sigma$ , a simple hypothesis for the mean  $\mu$  is that the mean is equal to a given value  $\mu_0$  and this is usually written in the following way:  $H_0: \mu = \mu_0$ .

NOTE 3 A simple hypothesis specifies the probability distribution (2.11) completely.

#### 1.44

##### composite hypothesis

hypothesis (1.40) that specifies more than one distribution (2.11) in a family of distributions (2.8)

EXAMPLE 1 The null hypotheses (1.41) and the alternative hypotheses (1.42) given in the examples in 1.41 and 1.42 are all examples of composite hypotheses.

EXAMPLE 2 In 1.48, the null hypothesis in Case 3 of Example 3 is a simple hypothesis. The null hypothesis in Example 4 is also a simple hypothesis. The other hypotheses in 1.48 are composite.

NOTE A composite hypothesis is a null hypothesis or alternative hypothesis for which the selected subset consists of more than a single probability distribution.

#### 1.45

##### significance level

$\alpha$

<statistical test> maximum probability (2.5) of rejecting the null hypothesis (1.41) when in fact it is true

NOTE If the null hypothesis is a simple hypothesis (1.43), then the probability of rejecting the null hypothesis if it were true becomes a single value.

#### 1.46

##### Type I error

rejection of the null hypothesis (1.41) when in fact it is true

NOTE 1 In fact, a Type I error is an incorrect

quyết định sai. Do đó, ta mong muốn duy trì **xác suất** (2.5) đưa ra quyết định sai như vậy càng nhỏ càng tốt. Để đạt được xác suất bằng không sai lầm loại I, ta có thể không bao giờ bác bỏ giả thuyết không. Nói cách khác là không quan tâm đến bằng chứng khi đưa ra quyết định.

CHÚ THÍCH 2: Có khả năng trong một số tình huống (ví dụ, phép kiểm nghiệm tham số nhị phân  $p$ ), mức ý nghĩa quy định trước 0,05 là không thể đạt được do sự rời rạc của các kết quả.

1.47

**Sai lầm loại II**

Việc không bác bỏ **giả thuyết không** (1.41) trong khi trên thực tế **giả thuyết không** không đúng.

CHÚ THÍCH: Trong thực tế, sai lầm loại II là một quyết định sai. Do đó, ta mong muốn duy trì **xác suất** (2.5) đưa ra quyết định sai như vậy càng nhỏ càng tốt. Sai lầm loại II thường xảy ra trong trường hợp cỡ mẫu không đủ để phát hiện ra sự sai lệch so với giả thuyết không.

1.48

**Phép kiểm nghiệm thống kê**

Kiểm nghiệm mức ý nghĩa

Quy trình quyết định xem liệu có bác bỏ giả thuyết không (1.41) và chấp nhận đối giả thuyết (1.42) hay không.

VÍ DỤ 1: Nếu một biến ngẫu nhiên liên tục (2.29) thực tế có thể lấy giá trị từ  $-\infty$  đến  $+\infty$  và có nghi ngờ rằng phân bố xác suất thực không phải là phân bố chuẩn (2.50), thì các giả thuyết sẽ được lập như dưới đây.

- Phạm vi của trường hợp này là tất cả các phân bố xác suất liên tục (2.23), có thể lấy giá trị từ  $-\infty$  đến  $+\infty$ .
- Phỏng đoán là phân bố xác suất thực không phải là phân bố chuẩn.
- Giả thuyết không là phân bố xác suất là phân bố chuẩn.
- Đối giả thuyết là phân bố xác suất không phải là phân bố chuẩn.

decision. Hence, it is desired to keep the **probability** (2.5) of making such an incorrect decision as small as possible. To obtain a zero probability of a Type I error, one would never reject the null hypothesis. In other words, regardless of the evidence, the same decision is made.

NOTE 2 It is possible that in some situations (for example, testing the binomial parameter  $p$ ) that a prespecified significance level such as 0,05 is not attainable due to discreteness of outcomes.

1.47

**Type II error**

failure to reject the **null hypothesis** (1.41) when in fact the **null hypothesis** is not true

NOTE In fact, a Type II error is an incorrect decision. Hence, it is desired to keep the **probability** (2.5) of making such an incorrect decision as small as possible. Type II errors commonly occur in situations where the sample sizes are insufficient to reveal a departure from the null hypothesis.

1.48

**statistical test**

significance test

procedure to decide if a null hypothesis (1.41) is to be rejected in favour of an alternative hypothesis (1.42)

EXAMPLE 1 As an example, if an actual, continuous random variable (2.29) can take values between  $-\infty$  and  $+\infty$  and one has a suspicion that the true probability distribution is not a normal distribution (2.50), then the hypotheses will be formulated, as follows.

- The scope of the situation is all continuous probability distributions (2.23), which can take values between  $-\infty$  and  $+\infty$ .
- The conjecture is that the true probability distribution is not a normal distribution.
- The null hypothesis is that the probability distribution is a normal distribution.

VÍ DỤ 2: Nếu biến ngẫu nhiên tuân thủ phân bố chuẩn với độ lệch chuẩn (2.37) đã biết và nghi ngờ rằng giá trị kỳ vọng  $\mu$  sai lệch so với giá trị  $\mu_0$  đã cho, khi đó các giả thuyết sẽ được lập theo Trường hợp 3 trong ví dụ tiếp theo.

VÍ DỤ 3: Ví dụ này xem xét ba khả năng trong phép kiểm nghiệm thống kê.

Trường hợp 1. Phòng đoán rằng trung bình quá trình cao hơn trung bình mục tiêu của  $\mu_0$ . Phòng đoán này dẫn đến các giả thuyết sau:

Giả thuyết không:  $H_0: \mu \leq \mu_0$

Đối giả thuyết:  $H_1: \mu > \mu_0$

Trường hợp 2. Phòng đoán rằng trung bình quá trình thấp hơn trung bình mục tiêu của  $\mu_0$ . Phòng đoán này dẫn đến các giả thuyết sau:

Giả thuyết không:  $H_0: \mu \geq \mu_0$

Đối giả thuyết:  $H_1: \mu < \mu_0$

Trường hợp 3. Phòng đoán rằng trung bình quá trình không bằng trung bình quá trình nhưng không quy định hướng. Phòng đoán này dẫn đến các giả thuyết sau:

Giả thuyết không:  $H_0: \mu = \mu_0$

Đối giả thuyết:  $H_1: \mu \neq \mu_0$

Trong cả ba trường hợp, việc thiết lập các giả thuyết đều bắt nguồn từ phỏng đoán liên quan đến đối giả thuyết và sự sai lệch của đối giả thuyết so với điều kiện cơ sở.

VÍ DỤ 4: Ví dụ này xét trong phạm vi tất cả các tỷ lệ khuyết tật trong hai lô 1 và 2,  $p_1$  và  $p_2$  nhận các giá trị từ không đến một. Ta có thể nghi ngờ rằng hai lô là khác nhau và do đó phỏng đoán rằng tỷ lệ khuyết tật trong hai lô là khác nhau. Phòng đoán này dẫn đến các giả thuyết sau:

– The alternative hypothesis is that the probability distribution is not a normal distribution.

EXAMPLE 2 If the random variable follows a normal distribution with known standard deviation (2.37) and one suspects that its expectation value  $\mu$  deviates from a given value  $\mu_0$ , then the hypotheses will be formulated according to Case 3 in the next example.

EXAMPLE 3 This example considers three possibilities in statistical testing.

Case 1. It is conjectured that the process mean is higher than the target mean of  $\mu_0$ . This conjecture leads to the following hypotheses:

Null hypothesis:  $H_0: \mu \leq \mu_0$

Alternative hypothesis:  $H_1: \mu > \mu_0$

Case 2. It is conjectured that the process mean is lower than the target mean of  $\mu_0$ . This conjecture leads to the following hypotheses:

Null hypothesis:  $H_0: \mu \geq \mu_0$

Alternative hypothesis:  $H_1: \mu < \mu_0$

Case 3. It is conjectured that the process mean is not compatible with the process mean but the direction is not specified. This conjecture leads to the following hypotheses:

Null hypothesis:  $H_0: \mu = \mu_0$

Alternative hypothesis:  $H_1: \mu \neq \mu_0$

In all three cases, the formulation of the hypotheses was driven by a conjecture regarding the alternative hypothesis and its departure from a baseline condition.

EXAMPLE 4 This example considers as its scope all proportions  $p_1$  and  $p_2$  between zero and one of defectives in two lots 1 and 2. One might suspect that the two lots are different and therefore conjecture that the proportions of defects in the two lots are different. This conjecture leads to the

Giả thuyết không:  $H_0: p_1 = p_2$

Đối giả thuyết:  $H_1: p_1 \neq p_2$

**CHÚ THÍCH 1:** Phép kiểm nghiệm thống kê là một quy trình, đúng trong những điều kiện quy định, để quyết định việc phân bố xác suất thực tuân theo giả thuyết không hay đối giả thuyết, dựa vào các quan trắc trên mẫu.

**CHÚ THÍCH 2:** Trước khi tiến hành phép kiểm nghiệm thống kê, tập hợp các phân bố xác suất được xác định trước trên cơ sở thông tin có sẵn. Tiếp đó, các phân bố xác suất, có thể đúng trên cơ sở phỏng đoán cần nghiên cứu, được nhận biết để thiết lập đối giả thuyết. Sau cùng, lập nên giả thuyết không để bổ sung cho đối giả thuyết. Trong nhiều trường hợp, tập hợp có thể của các phân bố xác suất và từ đó giả thuyết không và đối giả thuyết cũng có thể được xác định bằng cách tham chiếu các tập hợp giá trị của các tham số liên quan.

**CHÚ THÍCH 3:** Vì quyết định dựa trên cơ sở các quan trắc mẫu nên có thể dẫn đến **sai lầm loại I** (1.46), bác bỏ giả thuyết không trong khi giả thuyết không là đúng, hoặc **sai lầm loại II** (1.47), không bác bỏ giả thuyết không ủng hộ đối giả thuyết trong khi đối giả thuyết là đúng.

**CHÚ THÍCH 4:** Trường hợp 1 và 2 trong ví dụ 3 ở trên là ví dụ về **phép kiểm nghiệm một phía**. Trường hợp 3 là ví dụ về **phép kiểm nghiệm hai phía**. Trong cả ba trường hợp, lựa chọn một phía hay hai phía được xác định bằng việc xem xét vùng tham số  $\mu$  tương ứng với đối giả thuyết. Tổng quát hơn, phép kiểm nghiệm một phía và hai phía có thể bị chi phối bởi vùng bác bỏ giả thuyết không ủng hộ thống kê kiểm nghiệm đã chọn. Nghĩa là, thống kê kiểm nghiệm có một vùng tới hạn thuận cho đối giả thuyết, nhưng nó có thể không liên quan trực tiếp đến mô tả đơn giản không gian tham số như trong các trường hợp 1, 2 và 3.

**CHÚ THÍCH 5:** Cần thận trọng đối với việc đưa ra các

following hypotheses:

Null hypothesis:  $H_0: p_1 = p_2$

Alternative hypothesis:  $H_1: p_1 \neq p_2$

**NOTE 1** A statistical test is a procedure, which is valid under specified conditions, to decide, by means of observations from a sample, whether the true probability distribution belongs to the null hypothesis or the alternative hypothesis.

**NOTE 2** Before a statistical test is carried out the possible set of probability distributions is at first determined on the basis of the available information. Next the probability distributions, which could be true on the basis of the conjecture to be studied, are identified to constitute the alternative hypothesis. Finally, the null hypothesis is formulated as the complement to the alternative hypothesis. In many cases, the possible set of probability distributions and hence also the null hypothesis and the alternative hypothesis can be determined by reference to sets of values of relevant parameters.

**NOTE 3** As the decision is made on the basis of observations from a sample, it may be erroneous leading to either a **Type I error** (1.46), rejecting the null hypothesis when in fact it is correct, or a **Type II error** (1.47), failure to reject the null hypothesis in favour of the alternative hypothesis when the alternative hypothesis is true.

**NOTE 4** Case 1 and Case 2 of Example 3 above are instances of **one-sided tests**. Case 3 is an example of a **two-sided test**. In all three of these cases, the one-sided versus two-sided qualifier is determined by consideration of the region of the parameter  $\mu$  corresponding to the alternative hypothesis. More generally, one-sided and two-sided tests can be governed by the region for rejection of the null hypothesis corresponding to the chosen test statistic. That is, the test statistic has an associated critical region favouring the alternative hypothesis, but it may not relate directly to a simple description of the parameter space as in Cases 1, 2 and 3.

giả định cơ bản hoặc việc ứng dụng sai phép kiểm nghiệm thống kê. Phép kiểm nghiệm thống kê dẫn đến những kết luận ổn định ngay cả trong trường hợp quy định sai của các giả định cơ bản được gọi là **ổn định**. Phép kiểm nghiệm  $t$  một mẫu đối với trung bình là một ví dụ về phép kiểm nghiệm được coi là rất ổn định ở các phân bố không chuẩn. Phép kiểm nghiệm Bartlett đối với tính thuần nhất của các phương sai là một ví dụ về quy trình không ổn định, có khả năng dẫn đến việc bác bỏ quá mức sự bằng nhau của các phương sai trong các trường hợp phân bố trong đó các phương sai trên thực tế là như nhau.

#### 1.49

##### $p$ -giá trị

**xác suất (2.5) quan trắc thấy giá trị thống kê kiểm nghiệm (1.52) được quan trắc hoặc giá trị khác bất kỳ ít nhất là không thuận cho giả thuyết không (1.41).**

VÍ DỤ: Xét ví dụ số ban đầu nêu trong 1.9. Giả sử để minh họa rằng các giá trị này là giá trị quan trắc từ quá trình mà thông thường mong muốn có trung bình là 12,5, từ kinh nghiệm trước đó, người kỹ sư cùng với quá trình cảm thấy rằng quá trình này thấp hơn giá trị mong muốn trên. Một nghiên cứu được thực hiện và một mẫu ngẫu nhiên cỡ mẫu 10 được chọn với các kết quả số ở 1.9. Các giả thuyết thích hợp là:

Giả thuyết không:  $H_0: \mu \geq 12,5$

Đối giả thuyết:  $H_0: \mu < 12,5$

Trung bình mẫu là 9,7 theo hướng của phỏng đoán nhưng có đủ cách xa 12,5 để hỗ trợ phỏng đoán hay không? Đối với ví dụ này **thống kê kiểm nghiệm (1.52)** là  $-1,976\ 4$  với  $p$ -giá trị tương ứng 0,040. Điều này có nghĩa là có ít hơn bốn cơ hội trong một trăm quan trắc giá trị thống kê kiểm nghiệm là  $-1,976\ 4$  hoặc thấp hơn, nếu trong thực tế trung bình thực của quá trình là 12,5. Nếu mức ý nghĩa quy định trước ban đầu là 0,05, thì thông thường ta sẽ bác bỏ giả thuyết không và chấp nhận đối giả thuyết.

Giả sử khác là vấn đề được trình bày hơi khác đối

NOTE 5 Careful attention to the underlying assumptions must be made or the application of statistical testing may be flawed. Statistical tests that lead to stable inferences even under possible misspecification of the underlying assumptions are referred to as **robust**. The one-sample  $t$  test for the mean is an example of a test considered very robust under non-normal distributions. Bartlett's test for homogeneity of variances is an example of a non-robust procedure, possibly leading to the excessive rejection of equality of variances in distributional cases for which the variances were in fact identical.

#### 1.49

##### $p$ -value

**probability (2.5) of observing the observed test statistic (1.52) value or any other value at least as unfavourable to the null hypothesis (1.41)**

EXAMPLE Consider the numerical example originally introduced in 1.9. Suppose for illustration that these values are observations from a process that is nominally expected to have a mean of 12,5, and from previous experience the engineer associated with the process felt that the process was consistently lower than the nominal value. A study was undertaken and a random sample of size 10 was collected with the numerical results from 1.9. The appropriate hypotheses are:

Null hypothesis:  $H_0: \mu \geq 12,5$

Alternative hypothesis:  $H_0: \mu < 12,5$

The sample mean is 9,7 which is in the direction of the conjecture, but is it sufficiently far from 12,5 to support the conjecture? For this example the **test statistic (1.52)** is  $-1,976\ 4$  with corresponding  $p$ -value 0,040. This means that there are less than four chances in one hundred of observing a test statistic value of  $-1,976\ 4$  or lower, if in fact the true process mean is at 12,5. If the original pre-specified significance level had been 0,05, then typically one would reject the null hypothesis in favour of the

chút. Hình dung là vấn đề liên quan là quá trình chệch khỏi mục tiêu 12,5 nhưng hướng không được xác định. Điều này dẫn đến các giả thuyết sau:

Giả thuyết không:  $H_0: \mu = 12,5$

Đối giả thuyết:  $H_1: \mu \neq 12,5$

Cho dữ liệu tương tự thu thập từ mẫu ngẫu nhiên, thống kê kiểm nghiệm cũng bằng  $-1,9764$ . Đối với đối giả thuyết này, vấn đề quan tâm là "xác suất gặp phải giá trị cực trị như vậy hoặc vượt quá giá trị đó là bao nhiêu?". Trong trường hợp này, có hai vùng liên quan, các giá trị nhỏ hơn hoặc bằng  $-1,9764$  hoặc các giá trị lớn hơn hoặc bằng  $1,9764$ . Xác suất của thống kê kiểm nghiệm  $t$  xuất hiện trong một trong hai vùng này là  $0,080$  (bằng hai lần giá trị một phía). Có tám cơ hội trong một trăm quan trắc giá trị thống kê kiểm nghiệm đạt cực trị này hoặc lớn hơn. Vì vậy, giả thuyết không không bị bác bỏ ở mức ý nghĩa  $0,05$ .

CHÚ THÍCH 1: Nếu  $p$ -giá trị, ví dụ, trở thành  $0,029$ , thì theo giả thuyết không, sẽ có ít hơn ba cơ hội trên một trăm lần xuất hiện cực trị của thống kê kiểm nghiệm hoặc vượt quá cực trị. Dựa vào thông tin này, ta có thể buộc phải bác bỏ giả thuyết không vì đó là một  $p$ -giá trị tương đối nhỏ. Chính thức hơn thì nếu mức ý nghĩa được đặt là  $0,05$ , thì rõ ràng  $p$ -giá trị là  $0,029$  nhỏ hơn  $0,05$  dẫn đến việc bác bỏ giả thuyết không.

CHÚ THÍCH 2: Thuật ngữ  $p$ -giá trị đôi khi được gọi là xác suất ý nghĩa mà không nên nhầm với mức ý nghĩa (1.45) là một hằng số quy định trong một ứng dụng.

### 1.50

#### Hiệu lực của phép kiểm nghiệm

Một trừ đi xác suất (2.5) mắc sai lầm loại II (1.47).

CHÚ THÍCH 1: Hiệu lực của phép kiểm nghiệm đối

alternative hypothesis.

Suppose alternatively that the problem were formulated somewhat differently. Imagine that the concern was that the process was off the 12,5 target but the direction was unspecified. This leads the following hypotheses:

Null hypothesis:  $H_0: \mu = 12,5$

Alternative hypothesis:  $H_1: \mu \neq 12,5$

Given the same data collected from a random sample, the test statistic is the same,  $-1,9764$ . For this alternative hypothesis, a question of interest is "what is the probability of seeing such an extreme value or more extreme?". In this case, there are two relevant regions, values less than or equal to  $-1,9764$  or values greater than or equal to  $1,9764$ . The probability of a  $t$  test statistic occurring in one of these regions is  $0,080$  (twice the one-sided value). There are eight chances in one hundred of observing a test statistic value this extreme or more so. Thus, the null hypothesis is not rejected at the significance level  $0,05$ .

NOTE 1 If the  $p$ -value, for example, turns out to be  $0,029$ , then there are less than three chances in one hundred that such an extreme value of the test statistic or a more extreme one, would occur under the null hypothesis. On the basis of this information, one might feel compelled to reject the null hypothesis, as this is a fairly small  $p$ -value. More formally, if the significance level had been established as  $0,05$ , then definitely the  $p$ -value of  $0,029$  being less than  $0,05$  would lead to the rejection of the null hypothesis.

NOTE 2 The term  $p$ -value is sometimes referred to as the significance probability which should not be confused with **significance level** (1.45) which is a specified constant in an application.

### 1.50

#### power of a test

one minus the **probability** (2.5) of the **Type II error** (1.47)



với một giá trị quy định của một **tham số** (2.9) chưa biết trong một **họ phân bố** (2.8) bằng **xác suất bác bỏ giả thuyết không** (1.41) đối với giá trị tham số đó.

**CHÚ THÍCH 2:** Trong hầu hết các trường hợp thực tế, việc tăng cỡ mẫu sẽ làm tăng hiệu lực của phép kiểm nghiệm. Nói cách khác là xác suất bác bỏ giả thuyết không khi **đối giả thuyết** (1.42) là đúng sẽ tăng khi cỡ mẫu tăng, từ đó làm giảm xác suất sai lầm loại II.

**CHÚ THÍCH 3:** Các trường hợp kiểm nghiệm thường mong muốn khi cỡ mẫu trở nên cực lớn, thậm chí sai lệch nhỏ so với giả thuyết không phải được phát hiện, dẫn tới bác bỏ giả thuyết không. Nói cách khác, hiệu lực của phép kiểm nghiệm cần đạt đến 1 đối với từng đối giả thuyết của giả thuyết không khi cỡ mẫu lớn đến vô cùng. Phép kiểm nghiệm như vậy gọi là **nhất quán**. Khi so sánh hai phép kiểm nghiệm về hiệu lực, phép kiểm nghiệm có hiệu lực cao hơn có vẻ có hiệu quả hơn với điều kiện các mức ý nghĩa giống nhau cũng như các giả thuyết không và đối giả thuyết. Có các mô tả toán học chính thức hơn cho cả tính nhất quán và hiệu quả không thuộc phạm vi của tiêu chuẩn này. (Tham khảo các sách thống kê hoặc sách toán thống kê.)

### 1.51

#### Đường hiệu lực

Tập hợp các giá trị **hiệu lực** của **phép kiểm nghiệm** (1.50) là hàm số của **tham số** (2.9) tổng thể từ một **họ phân bố** (2.8).

**CHÚ THÍCH:** Hàm hiệu lực bằng một trừ đi đường hiệu quả.

### 1.52

#### Thống kê kiểm nghiệm

**Thống kê** (1.8) sử dụng cùng với **phép kiểm nghiệm thống kê** (1.48)

**NOTE 1** The power of the test for a specified value of an unknown **parameter** (2.9) in a **family of distributions** (2.8) equals the **probability** of rejecting the **null hypothesis** (1.41) for that parameter value.

**NOTE 2** In most cases of practical interest, increasing the sample size will increase the power of a test. In other words, the probability of rejecting the null hypothesis, when the **alternative hypothesis** (1.42) is true increases with increasing sample size, thereby reducing the probability of a Type II error.

**NOTE 3** It is desirable in testing situations that as the sample size becomes extremely large, even small departures from the null hypothesis ought to be detected, leading to the rejection of the null hypothesis. In other words, the power of the test should approach 1 for every alternative to the null hypothesis as the sample size becomes infinitely large. Such tests are referred to as **consistent**. In comparing two tests with respect to power, the test with the higher power is deemed the more **efficient** provided the significance levels are identical as well as the particular null and alternative hypotheses. There are more formal, mathematical descriptions of both consistency and efficiency that are beyond the scope of this part of ISO 3534. (Consult the various encyclopaedia in statistics or mathematical statistics textbooks.)

### 1.51

#### power curve

collection of values of the **power of a test** (1.50) as a function of the population **parameter** (2.9) from a **family of distributions** (2.8)

**NOTE** The power function is equal to one minus the operating characteristic curve.

### 1.52

#### test statistic

**statistic** (1.8) used in conjunction with a **statistical test** (1.48)

CHÚ THÍCH: Thống kê kiểm nghiệm được dùng để đánh giá **phân bố xác suất** (2.11) được xét có phù hợp với **giả thuyết không** (1.41) hoặc **đổi giả thuyết** (1.42) hay không.

1.53

**Thống kê mô tả đồ thị**

**Thống kê mô tả** (1.5) dưới dạng hình ảnh.

CHÚ THÍCH: Mục đích của thống kê mô tả thường là để giảm số lượng lớn các giá trị xuống còn một số ít để sử dụng hoặc biểu diễn giá trị theo cách dễ hình dung. Ví dụ về tổng hợp đồ họa bao gồm biểu đồ hộp, biểu đồ xác suất, biểu đồ Q-Q, biểu đồ phân vị chuẩn, đám mây điểm, đám mây điểm nhiều chiều và biểu đồ phân bố (1.61).

1.54

**Thống kê mô tả dạng số**

**Thống kê mô tả** (1.5) dưới dạng số.

CHÚ THÍCH: Thống kê mô tả bằng số bao gồm **trung bình** (1.15), **độ rộng mẫu** (1.10), **độ lệch chuẩn mẫu** (1.17), **độ rộng giữa các tứ phân vị**, ...

1.55

**Các lớp**

CHÚ THÍCH: Các lớp được giả định là đầy đủ và loại trừ lẫn nhau. Đường thẳng thực là tất cả các số thực nằm trong khoảng từ  $-\infty$  đến  $+\infty$ .

1.55.1

**Lớp**

<đặc trưng định tính> Tập hợp con các cá thể lấy từ **mẫu** (1.3).

1.55.2

**Lớp**

<đặc trưng thứ tự> Tập hợp một hoặc nhiều loại được sắp xếp theo một thang thứ tự.

1.55.3

**Lớp**

<đặc trưng định lượng> Khoảng của đường thẳng thực.

NOTE The test statistic is used to assess whether the **probability distribution** (2.11) at hand is consistent with the **null hypothesis** (1.41) or the **alternative hypothesis** (1.42).

1.53

**graphical descriptive statistics**

**descriptive statistics** (1.5) in pictorial form

NOTE The intent of descriptive statistics is generally to reduce a large number of values to a manageable few or to present the values in a way to facilitate visualization. Examples of graphical summaries include boxplots, probability plots, Q-Q plots, normal quantile plots, scatterplots, multiple scatterplots and histograms (1.61).

1.54

**numerical descriptive statistics**

**descriptive statistics** (1.5) in numerical form

NOTE Numerical descriptive statistics include **average** (1.15), **sample range** (1.10), **sample standard deviation** (1.17), **interquartile range**, and so forth.

1.55

**classes**

NOTE The classes are assumed to be mutually exclusive and exhaustive. The real line is all the real numbers between  $-\infty$  and  $+\infty$ .

1.55.1

**class**

<qualitative characteristic> subset of items from a **sample** (1.3)

1.55.2

**class**

<ordinal characteristic> set of one or more adjacent categories on an ordinal scale

1.55.3

**class**

<quantitative characteristic> interval of the real line

## 1.56

**Giới hạn lớp****Biên giới lớp**

<đặc trưng định lượng> Giá trị xác định cận trên và cận dưới của lớp (1.55).

CHÚ THÍCH: Định nghĩa này đề cập đến giới hạn lớp kèm theo đặc trưng định lượng.

## 1.57

**Điểm giữa lớp**

<đặc trưng định lượng> Trung bình (1.15) giữa giới hạn lớp (1.56) trên và dưới.

## 1.58

**Độ rộng lớp**

<đặc trưng định lượng> Giới hạn trên của lớp trừ đi giới hạn dưới của lớp (1.55).

## 1.59

**Tần số**

Số lần xuất hiện hoặc giá trị quan trắc (1.4) trong một lớp (1.55) quy định

## 1.60

**Phân bố tần số**

Mối quan hệ theo thực nghiệm giữa các lớp (1.55) và số lần xuất hiện của chúng hoặc giá trị quan trắc (1.4).

## 1.61

**Biểu đồ phân bố**

Cách biểu diễn phân bố tần số (1.60) bằng đồ thị gồm các hình chữ nhật liền nhau, mỗi hình có độ rộng bằng với độ rộng lớp (1.58) và diện tích tỷ lệ với tần số lớp.

CHÚ THÍCH: Cần chú ý đối với trường hợp dữ liệu trong các lớp có độ rộng lớp không bằng nhau.

## 1.62

**Biểu đồ cột**

Cách biểu diễn phân bố tần số (1.60) của một tính

## 1.56

**class limits****class boundaries**

<quantitative characteristic> values defining the upper and lower bounds of a class (1.55)

NOTE This definition refers to class limits associated with quantitative characteristics.

## 1.57

**mid-point of class**

<quantitative characteristic> average (1.15) of upper and lower class limits (1.56)

## 1.58

**class width**

<quantitative characteristic> upper limit of a class minus the lower limit of a class (1.55)

## 1.59

**frequency**

number of occurrences or observed values (1.4) in a specified class (1.55)

## 1.60

**frequency distribution**

empirical relationship between classes (1.55) and their number of occurrences or observed values (1.4)

## 1.61

**histogram**

graphical representation of a frequency distribution (1.60) consisting of contiguous rectangles, each with base width equal to the class width (1.58) and area proportional to the class frequency

NOTE Care needs to be taken for situations in which the data arises in classes having unequal class widths.

## 1.62

**bar chart**

graphical representation of a frequency

chất định danh bằng đồ thị gồm một tập hợp các hình chữ nhật có độ rộng bằng nhau và chiều cao tỷ lệ với tần số (1.59).

CHÚ THÍCH 1: Các hình chữ nhật đôi khi được vẽ bằng hình ảnh ba chiều để tạo tính thẩm mỹ, mặc dù việc này không cung cấp thêm thông tin và không phải là cách biểu diễn được khuyến nghị. Đối với biểu đồ cột, các hình chữ nhật không nhất thiết phải liền nhau.

CHÚ THÍCH 2: Sự phân biệt giữa biểu đồ phân bố và biểu đồ cột ngày càng trở nên mờ nhạt do phần mềm sẵn có không phải lúc nào cũng tuân thủ các định nghĩa nêu ở đây.

### 1.63

#### Tần số tích lũy

Tần số (1.59) cộng dồn đối với các lớp tính đến và bao gồm cả giới hạn quy định.

CHÚ THÍCH: Định nghĩa này chỉ áp dụng cho các giới hạn quy định tương ứng với các giới hạn lớp (1.56).

### 1.64

#### Tần số tương đối (tần suất)

Tần số (1.59) chia cho tổng số lần xuất hiện hoặc giá trị quan trắc (1.4).

### 1.65

#### Tần số tương đối tích lũy (tần suất tích lũy)

Tần số tích lũy (1.63) chia cho tổng số lần xuất hiện hoặc giá trị quan trắc (1.4).

## 2 Thuật ngữ dùng trong xác suất

### 2.1

#### Không gian mẫu

$\Omega$

Tập hợp tất cả các kết quả có thể có.

VÍ DỤ 1: Xét thời gian pin do người tiêu dùng mua sử dụng được. Nếu pin không có điện khi sử dụng lần đầu thì thời gian sử dụng bằng 0. Nếu pin hoạt động trong một khoảng thời gian thì thời gian sử dụng

distribution (1.60) of a nominal property consisting of a set of rectangles of uniform width with height proportional to frequency (1.59)

NOTE 1 The rectangles are sometimes depicted as three-dimensional images for apparently aesthetic purposes, although this adds no additional information and is not a recommended presentation. For a bar chart, the rectangles need not be contiguous.

NOTE 2 The distinction between histograms and bar charts has become increasingly blurred as available software does not always follow the definitions given here.

### 1.63

#### cumulative frequency

frequency (1.59) for classes up to and including a specified limit

NOTE This definition is only applicable for specified values that correspond to class limits (1.56).

### 1.64

#### relative frequency

frequency (1.59) divided by the total number of occurrences or observed values (1.4)

### 1.65

#### cumulative relative frequency

cumulative frequency (1.63) divided by the total number of occurrences or observed values (1.4)

## 2 Terms used in probability

### 2.1

#### sample space

$\Omega$

set of all possible outcomes

EXAMPLE 1 Consider the failure times of batteries purchased by a consumer. If the battery has no power upon initial use, its failure time is 0. If the battery does function for a while, it produces a failure

bằng một số giờ. Do đó, không gian mẫu gồm các kết quả {pin hỏng ngay lần đầu} và {pin hỏng sau  $x$  giờ, trong đó  $x$  lớn hơn 0}. Ví dụ này sẽ được sử dụng trong toàn bộ điều này. Cụ thể, thảo luận mở rộng của ví dụ này được nêu trong 2.68.

VÍ DỤ 2: Một hộp gồm 10 điện trở được ghi nhãn 1, 2, 3, 4, 5, 6, 7, 8, 9, 10. Nếu hai điện trở được lấy mẫu ngẫu nhiên không hoàn lại từ bộ điện trở này, thì không gian mẫu gồm 45 kết quả sau: (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (1, 7), (1, 8), (1, 9), (1, 10), (2, 3), (2, 4), (2, 5), (2, 6), (2, 7), (2, 8), (2, 9), (2, 10), (3, 4), (3, 5), (3, 6), (3, 7), (3, 8), (3, 9), (3, 10), (4, 5), (4, 6), (4, 7), (4, 8), (4, 9), (4, 10), (5, 6), (5, 7), (5, 8), (5, 9), (5, 10), (6, 7), (6, 8), (6, 9), (6, 10), (7, 8), (7, 9), (7, 10), (8, 9), (8, 10), (9, 10). Biến cố (1, 2) được coi là giống với (2, 1), vậy thứ tự lấy mẫu điện trở không quan trọng. Nếu thứ tự là quan trọng thì (1, 2) được coi là khác với (2, 1), khi đó có tổng số 90 kết quả trong không gian mẫu.

VÍ DỤ 3: Nếu trong ví dụ trước, việc lấy mẫu được thực hiện có hoàn lại thì cần thêm vào các biến cố (1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6), (7, 7), (8, 8), (9, 9) và (10, 10). Trong trường hợp thứ tự không quan trọng, trong không gian mẫu sẽ có 55 kết quả. Trường hợp thứ tự là quan trọng, trong không gian mẫu có 100 kết quả.

CHÚ THÍCH 1: Các kết quả có thể phát sinh từ thực nghiệm thực tế hoặc thực nghiệm giả thuyết hoàn toàn. Tập hợp này có thể là danh sách rõ ràng, một tập hợp đếm được ví dụ như các số nguyên dương, {1, 2, 3, ...}, hoặc đường thẳng thực.

CHÚ THÍCH 2: Không gian mẫu là thành phần đầu tiên của không gian xác suất (2.68).

time of some number of hours. The sample space therefore consists of the outcomes {battery fails upon initial attempt} and {battery fails after  $x$  hours where  $x$  is greater than zero hours}. This example will be used throughout this clause. In particular, an extensive discussion of this example is given in 2.68.

EXAMPLE 2 A box contains 10 resistors that are labelled 1, 2, 3, 4, 5, 6, 7, 8, 9, 10. If two resistors were randomly sampled without replacement from this collection of resistors, the sample space consists of the following 45 outcomes: (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (1, 7), (1, 8), (1, 9), (1, 10), (2, 3), (2, 4), (2, 5), (2, 6), (2, 7), (2, 8), (2, 9), (2, 10), (3, 4), (3, 5), (3, 6), (3, 7), (3, 8), (3, 9), (3, 10), (4, 5), (4, 6), (4, 7), (4, 8), (4, 9), (4, 10), (5, 6), (5, 7), (5, 8), (5, 9), (5, 10), (6, 7), (6, 8), (6, 9), (6, 10), (7, 8), (7, 9), (7, 10), (8, 9), (8, 10), (9, 10). The event (1, 2) is deemed the same as (2, 1), so that the order in which resistors are sampled does not matter. If alternatively the order does matter, so (1, 2) is considered different from (2, 1), then there are a total of 90 outcomes in the sample space.

EXAMPLE 3 If in the preceding example, the sampling were performed with replacement, then the additional events (1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6), (7, 7), (8, 8), (9, 9), and (10, 10) would also need to be included. In the case where ordering does not matter, there would be 55 outcomes in the sample space. In the ordering matters situation, there would be 100 outcomes in the sample space.

NOTE 1 Outcomes could arise from an actual experiment or a completely hypothetical experiment. This set could be an explicit list, a countable set such as positive integers, {1, 2, 3, ...}, or the real line, for example.

NOTE 2 Sample space is the first component of a probability space (2.68).

## 2.2

## Biến cố

A

Tập con của không gian mẫu (2.1)

VÍ DỤ 1: Tiếp theo ví dụ 1 của 2.1, dưới đây là các ví dụ về biến cố  $\{0\}$ ,  $(0, 2)$ ,  $\{5,7\}$ ,  $[7, +\infty)$ , ứng với pin hỏng ngay từ đầu, pin ban đầu hoạt động nhưng chưa đến hai giờ làm việc thì hỏng, pin hỏng ở chính xác là 5,7 h, và sau 7 h pin vẫn không hỏng.  $\{0\}$  và  $\{5,7\}$  mỗi tập chứa một giá trị;  $(0, 2)$  là khoảng mở của đường thẳng thực;  $[7, +\infty)$  là khoảng vô hạn đóng bên trái của đường thẳng thực.

VÍ DỤ 2: Tiếp theo ví dụ 1 của 2.1, lưu ý đến việc chọn mà không hoàn lại và không ghi lại thứ tự chọn. Một biến cố có thể xảy ra là A định nghĩa bởi {ít nhất một trong các điện trở 1 hoặc 2 nằm trong mẫu}. Biến cố này chứa 17 kết quả (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (1, 7), (1, 8), (1, 9), (1, 10), (2, 3), (2, 4), (2, 5), (2, 6), (2, 7), (2, 8), (2, 9) và (2, 10). Một biến cố khác là B {không có điện trở nào trong số 8, 9 hoặc 10 nằm trong mẫu}. Biến cố này chứa 21 kết quả (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (1, 7), (2, 3), (2, 4), (2, 5), (2, 6), (2, 7), (3, 4), (3, 5), (3, 6), (3, 7), (4, 5), (4, 6), (4, 7), (5, 6), (5, 7), (6, 7).

VÍ DỤ 3: Tiếp tục với ví dụ 2, phần giao nhau giữa biến cố A và B (nghĩa là một trong các điện trở 1 và 2 nằm trong mẫu nhưng không có điện trở nào trong số 8, 9 và 10), chứa 11 kết quả sau: (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (1, 7), (2, 3), (2, 4), (2, 5), (2, 6), (2, 7).

Hợp hai biến cố A và B chứa 27 kết quả sau: (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (1, 7), (1, 8), (1, 9), (1, 10), (2, 3), (2, 4), (2, 5), (2, 6), (2, 7), (2, 8), (2, 9), (2, 10), (3, 4), (3, 5), (3, 6), (3, 7), (4, 5), (4, 6), (4, 7), (5, 6), (5, 7) và (6, 7).

Số kết quả trong hợp của hai biến cố A và B (nghĩa là ít nhất một trong hai điện trở 1 và 2 hoặc không có điện trở nào trong số 8, 9 và 10 nằm trong mẫu) là 27 đúng bằng  $17 + 21 - 11$ , là số kết quả trong A cộng số kết quả trong B trừ đi số kết quả trong phần giao

## 2.2

## event

A

subset of the sample space (2.1)

EXAMPLE 1 Continuing with Example 1 of 2.1, the following are examples of events  $\{0\}$ ,  $(0, 2)$ ,  $\{5,7\}$ ,  $[7, +\infty)$ , corresponding to an initially failed battery, a battery that works initially but fails before two hours, a battery that fails at exactly 5,7 h, and a battery that has not yet failed at 7 h. The  $\{0\}$  and  $\{5,7\}$  are each sets containing a single value;  $(0, 2)$  is an open interval of the real line;  $[7, +\infty)$  is a left closed infinite interval of the real line.

EXAMPLE 2 Continuing with Example 2 of 2.1, restrict attention to selection without replacement and without recording the selection order. One possible event is A defined by {at least one of the resistors 1 or 2 is included in the sample}. This event contains the 17 outcomes (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (1, 7), (1, 8), (1, 9), (1, 10), (2, 3), (2, 4), (2, 5), (2, 6), (2, 7), (2, 8), (2, 9), and (2, 10). Another possible event B is {none of the resistors 8, 9 or 10 is included in the sample}. This event contains the 21 outcomes (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (1, 7), (2, 3), (2, 4), (2, 5), (2, 6), (2, 7), (3, 4), (3, 5), (3, 6), (3, 7), (4, 5), (4, 6), (4, 7), (5, 6), (5, 7), (6, 7).

EXAMPLE 3 Continuing with Example 2, the intersection of events A and B (i.e. that at least one of the resistors 1 and 2 is included in the sample, but none of the resistors 8, 9 and 10), contains the following 11 outcomes (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (1, 7), (2, 3), (2, 4), (2, 5), (2, 6), (2, 7).

The union of the events A and B contains the following 27 outcomes: (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (1, 7), (1, 8), (1, 9), (1, 10), (2, 3), (2, 4), (2, 5), (2, 6), (2, 7), (2, 8), (2, 9), (2, 10), (3, 4), (3, 5), (3, 6), (3, 7), (4, 5), (4, 6), (4, 7), (5, 6), (5, 7), and (6, 7).

Incidentally, the number of outcomes in the union of the events A and B (i.e., that at least one of the

nhau của hai biến cố.

**CHÚ THÍCH:** Cho trước biến cố và kết quả của thực nghiệm, biến cố được gọi là xảy ra nếu kết quả thuộc về biến cố đó. Các biến cố thực tế sẽ thuộc về **sigma đại số của các biến cố** (2.69), thành phần thứ hai của **không gian xác suất** (2.68). Các biến cố tự nhiên xảy ra trong trường hợp trò chơi may rủi (bài poker, rulet, ...) xác định số kết quả thuộc về biến cố xác định các số chênh để cá cược.

### 2.3

#### Biến cố bù

$A^c$

**Không gian mẫu** (2.1) loại đi **biến cố** (2.2) đã cho.

**VÍ DỤ 1:** Tiếp tục với pin ở ví dụ 1 của 2.1, phần bù của biến cố  $\{0\}$  là biến cố  $(0, +\infty)$  tương đương với phần bù của biến cố ban đầu pin hoạt động. Tương tự, biến cố  $[0,3)$  ứng với các trường hợp pin ban đầu không hoạt động hoặc hoạt động trong khoảng ít hơn ba giờ. Phần bù của biến cố này là  $[3, \infty)$  ứng với trường hợp pin làm việc ở 3 h và thời gian làm việc lớn hơn giá trị này.

**VÍ DỤ 2:** Tiếp tục với ví dụ 2 của 2.2. Có thể dễ dàng tìm ra số kết quả trong biến cố  $B$  bằng cách xem xét biến cố bù cho  $B = \{\text{mẫu chứa ít nhất một trong các điện trở } 8, 9 \text{ hoặc } 10\}$ . Biến cố này chứa  $7 + 8 + 9 = 24$  kết quả  $(1, 8), (2, 8), (3, 8), (4, 8), (5, 8), (6, 8), (7, 8), (1, 9), (2, 9), (3, 9), (4, 9), (5, 9), (6, 9), (7, 9), (8, 9), (1, 10), (2, 10), (3, 10), (4, 10), (5, 10), (6, 10), (7, 10), (8, 10), (9, 10)$ . Vì toàn bộ không gian mẫu chứa 45 kết quả nên biến cố  $B$  chứa  $45 - 24 = 21$  kết quả [đó là:  $(1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (1, 7), (2, 3), (2, 4), (2, 5), (2, 6), (2, 7), (3, 4), (3, 5), (3, 6), (3, 7), (4, 5), (4, 6), (4, 7), (5, 6), (5, 7), (6, 7)$ ].

resistors 1 and 2 or none of the resistors 8, 9, and 10, is included in the sample) is 27 which also equals  $17 + 21 - 11$ , namely the number of outcomes in  $A$  plus the number of outcomes in  $B$  minus the number of outcomes in the intersection is equal to the number of outcomes in the union of the events.

**NOTE** Given an event and an outcome of an experiment, the event is said to have occurred, if the outcome belongs to the event. Events of practical interest will belong to the **sigma algebra of events** (2.69), the second component of the **probability space** (2.68). Events naturally occur in gambling contexts (poker, roulette, and so forth) where determining the number of outcomes that belong to an event determines the odds for betting.

### 2.3

#### complementary event

$A^c$

**sample space** (2.1) excluding the given event (2.2)

**EXAMPLE 1** Continuing with the battery Example 1 of 2.1, the complement of the event  $\{0\}$  is the event  $(0, +\infty)$  which is equivalent to the complement of the event that the battery did not function initially is the event that the battery did function initially. Similarly, the event  $[0,3)$  corresponds to the cases that either the battery was not functioning initially or it did function less than three hours. The complement of this event is  $[3, \infty)$  which corresponds to the case that a battery was working at 3 h and its failure time is greater than this value.

**EXAMPLE 2** Continuing with Example 2 of 2.2. The number of outcomes in  $B$  can be found easily by considering the complementary event to  $B = \{\text{the sample contains at least one of the resistors } 8, 9 \text{ or } 10\}$ . This event contains the  $7 + 8 + 9 = 24$  outcomes  $(1, 8), (2, 8), (3, 8), (4, 8), (5, 8), (6, 8), (7, 8), (1, 9), (2, 9), (3, 9), (4, 9), (5, 9), (6, 9), (7, 9), (8, 9), (1, 10), (2, 10), (3, 10), (4, 10), (5, 10), (6, 10), (7, 10), (8, 10), (9, 10)$ . As the entire sample space

CHÚ THÍCH 1: Biến cố bù là phần bù của biến cố trong không gian mẫu.

CHÚ THÍCH 2: Biến cố bù cũng là một biến cố.

CHÚ THÍCH 3: Đối với biến cố  $A$ , biến cố bù cho  $A$  thường được kí hiệu là  $A^c$ .

CHÚ THÍCH 4: Trong nhiều trường hợp, có thể dễ dàng tính toán xác suất của phần bù của một biến cố hơn là tính xác suất của biến cố đó. Ví dụ, biến cố xác định bởi "ít nhất một khuyết tật xuất hiện trong mẫu gồm 10 cá thể được chọn ngẫu nhiên từ tổng thể gồm 1 000 cá thể, với giả định một phần trăm khuyết tật" có một lượng lớn các kết quả được liệt kê. Phần bù của biến cố này (không phát hiện thấy khuyết tật) dễ xử lí hơn nhiều.

contains 45 outcomes in this case, the event  $B$  contains  $45 \cap 24 = 21$  outcomes [namely: (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (1, 7), (2, 3), (2, 4), (2, 5), (2, 6), (2, 7), (3, 4), (3, 5), (3, 6), (3, 7), (4, 5), (4, 6), (4, 7), (5, 6), (5, 7), (6, 7)].

NOTE 1 The complementary event is the complement of the event in the sample space.

NOTE 2 The complementary event is also an event.

NOTE 3 For an event  $A$ , the **complementary event** to  $A$  is usually designated by the symbol  $A^c$ .

NOTE 4 In many situations, it may be easier to compute the probability of the complement of an event than the probability of the event. For example, the event defined by "at least one defect occurs in a sample of 10 items chosen at random from a population of 1 000 items, having an assumed one percent defectives" has a huge number of outcomes to be listed. The complement of this event (no defects found) is much easier to deal with.

## 2.4

### Biến cố độc lập

Cặp biến cố (2.2) sao cho xác suất (2.5) của phần giao nhau giữa hai biến cố đó là tích của các xác suất riêng lẻ.

VÍ DỤ 1: Xét trường hợp trò tung hai xúc xắc, một đỏ và một trắng có được 36 kết quả, mỗi kết quả có xác suất  $1/36$ .  $D_i$  được định nghĩa là biến cố trong đó tổng số chấm trên hai xúc xắc là  $i$ .  $W$  được định nghĩa là xúc xắc trắng có một chấm. Biến cố  $D_7$  và  $W$  độc lập nhau, trong khi biến cố  $D_i$  và  $W$  lại không độc lập với  $i = 2, 3, 4, 5$  hoặc  $6$ . Các biến cố không độc lập được gọi là biến cố phụ thuộc.

VÍ DỤ 2: Các biến cố độc lập và phụ thuộc xảy ra tự nhiên trong các ứng dụng. Trong trường hợp các biến cố hoặc tình huống là phụ thuộc thì việc biết kết quả của biến cố liên quan rất có ích. Ví dụ, một người chuẩn bị thực hiện một cuộc phẫu thuật tim sẽ có khả năng thành công rất khác nhau, nếu đó là trường hợp người này có tiền sử hút thuốc hoặc yếu tố rủi ro

## 2.4

### independent events

pair of events (2.2) such that the **probability** (2.5) of the intersection of the two events is the product of the individual probabilities

EXAMPLE 1 Consider a two die tossing situation, with one red die and one white die so as to distinguish the 36 possible outcomes with probability  $1/36$  assigned to each.  $D_i$  is defined as the event where the sum of the dots on the red and white die is  $i$ .  $W$  is defined as the event that the white die shows one dot. The events  $D_7$  and  $W$  are independent, whereas the events  $D_i$  and  $W$  are not independent for  $i = 2, 3, 4, 5$  or  $6$ . Events that are not independent are referred to as dependent events.

EXAMPLE 2 Independent and dependent events arise naturally in applications. In cases where events or circumstances are dependent, it is quite useful to know of the outcome of a related event. For example, an individual about to undergo heart surgery could



khác. Vì vậy, hút thuốc và chết do tiến triển bệnh là phụ thuộc. Ngược lại, cái chết có thể độc lập với ngày của tuần mà người này sinh ra. Trong ngữ cảnh độ tin cậy thì các thành phần có nguyên nhân hỏng chung thì không có thời gian làm việc đến khi hỏng độc lập. Các thanh nhiên liệu trong lò phản ứng có xác suất thấp xảy ra vỡ nhưng đưa ra rằng một thanh nhiên liệu bị gãy thì xác suất gãy thanh liền kề về cơ bản có thể tăng lên.

VÍ DỤ 3: Tiếp theo ví dụ 2 của 2.2, giả định rằng việc lấy mẫu được thực hiện bằng cách lấy mẫu ngẫu nhiên đơn giản, sao cho tất cả các kết quả có cùng xác suất  $1/45$ . Khi đó  $P(A) = 17/45 = 0,3778$ ,  $P(B) = 21/45 = 0,4667$  và  $P(A \text{ và } B) = 11/45 = 0,2444$ . Tuy nhiên, tích  $P(A) \times P(B) = (17/45) \times (21/45) = 0,1763$ , khác với  $0,2444$ , do đó biến cố  $A$  và  $B$  là không độc lập.

CHÚ THÍCH: Định nghĩa này được nêu trong ngữ cảnh hai biến cố nhưng có thể mở rộng thêm. Đối với biến cố  $A$  và  $B$ , điều kiện độc lập  $P(A \cap B) = P(A)P(B)$ . Đối với ba biến cố  $A$ ,  $B$  và  $C$  độc lập, điều kiện là:

$$P(A \cap B \cap C) = P(A)P(B)P(C)$$

$$P(A \cap B) = P(A)P(B)$$

$$P(A \cap C) = P(A)P(C) \text{ và}$$

$$P(B \cap C) = P(B)P(C)$$

Nói chung, đối với nhiều hơn hai biến cố,  $A_1, A_2, \dots, A_n$ , là độc lập nếu xác suất phần giao nhau của tập con các biến cố cho trước bất kỳ bằng tích xác suất các biến cố riêng lẻ, điều kiện này duy trì cho từng và mọi tập con. Có thể đặt ra ví dụ trong đó mỗi cặp biến cố là độc lập nhưng ba biến cố lại không độc lập (nghĩa là theo cặp chứ không hoàn toàn độc lập).

## 2.5

### Xác suất của biến cố $A$

have very different prospects for success, if it is the case that this individual had a smoking history or other risk factors. Thus, smoking and death from invasive procedures could be dependent. In contrast, death would likely be independent of the day of the week that this person was born. In a reliability context, components having a common cause of failure do not have independent failure times. Fuel rods in a reactor have a presumably low probability of cracks occurring but given that a fuel rod cracks, the probability of an adjacent rod cracking may increase substantially.

EXAMPLE 3 Continuing Example 2 of 2.2, assume that the sampling has been done by simple random sampling, such that all outcomes have the same probability  $1/45$ . Then  $P(A) = 17/45 = 0,3778$ ,  $P(B) = 21/45 = 0,4667$  and  $P(A \text{ and } B) = 11/45 = 0,2444$ . However, the product  $P(A) \times P(B) = (17/45) \times (21/45) = 0,1763$ , which is different from  $0,2444$ , so the events  $A$  and  $B$  are not independent.

NOTE This definition is given in the context of two events but can be extended. For events  $A$  and  $B$ , the independence condition is  $P(A \cap B) = P(A)P(B)$ . For three events  $A$ ,  $B$  and  $C$  to be independent, it is required that:

$$P(A \cap B \cap C) = P(A)P(B)P(C)$$

$$P(A \cap B) = P(A)P(B)$$

$$P(A \cap C) = P(A)P(C) \text{ and}$$

$$P(B \cap C) = P(B)P(C)$$

In general, for more than two events,  $A_1, A_2, \dots, A_n$  are independent if the probability of the intersection of any given subset of the events equals the product of the individual events, this condition holding for each and every subset. It is possible to construct an example in which each pair of events is independent, but the three events are not independent (i.e. pairwise, but not complete independence).

## 2.5

### probability of an event $A$

$P(A)$ 

Số thực thuộc khoảng đóng  $[0, 1]$  được ấn định cho một biến cố (2.2).

VÍ DỤ: Tiếp theo ví dụ 2 của 2.1, có thể tìm được xác suất của một biến cố bằng cách cộng các xác suất của tất cả các kết quả cấu thành biến cố đó. Nếu tất cả 45 kết quả có cùng một xác suất thì mỗi kết quả sẽ có xác suất là  $1/45$ . Xác suất của biến cố có thể tìm được bằng cách đếm số kết quả rồi chia cho 45.

CHÚ THÍCH 1: Độ đo xác suất (2.70) gán các số thực cho mọi biến cố trong không gian mẫu. Việc gán bởi độ đo xác suất một biến cố đơn lẻ cung cấp xác suất của biến cố đó. Nói cách khác, độ đo xác suất lập ra tập hợp hoàn chỉnh các giá trị ấn định cho tất cả các biến cố, trong khi xác suất đại diện cho một ấn định cụ thể cho một biến cố riêng lẻ.

CHÚ THÍCH 2: Định nghĩa này đề cập đến xác suất như là xác suất của một biến cố cụ thể. Xác suất có thể liên quan đến tần suất xuất hiện trong một thời gian dài hoặc mức độ tin tưởng vào khả năng xuất hiện biến cố. Thông thường, xác suất của biến cố  $A$  được biểu thị bằng  $P(A)$ . Ký hiệu  $p(A)$  sử dụng chữ  $p$  được dùng trong trường hợp có nhu cầu xem xét rõ ràng sự chính thống của không gian xác suất (2.68).

## 2.6

### Xác suất có điều kiện

 $P(A|B)$ 

Xác suất (2.5) của phần giao nhau giữa biến cố  $A$  và  $B$  chia cho xác suất của  $B$ .

VÍ DỤ 1: Tiếp theo ví dụ 1 của 2.1, xét biến cố (2.2)  $A$  xác định là {pin hoạt động trong ít nhất là ba giờ}, là  $[3, \infty)$ . Biến cố  $B$  được xác định là {pin hoạt động ngay từ đầu}, là  $(0, \infty)$ . Xác suất có điều kiện của  $A$  cho trước  $B$  tính đến việc xử lý trường hợp pin hoạt

 $P(A)$ 

real number in the closed interval  $[0, 1]$  assigned to an event (2.2)

EXAMPLE Continuing with Example 2 of 2.1, the probability for an event can be found by adding the probabilities for all outcomes constituting the event. If all the 45 outcomes have the same probability, each of them will have the probability  $1/45$ . The probability of an event can be found by counting the number of outcomes and dividing this number by 45.

NOTE 1 Probability measure (2.70) provides assignment of real numbers for every event of interest in the sample space. Taking an individual event, the assignment by the probability measure gives the probability associated with the event. In other words, probability measure yields the complete set of assignments for all of the events, whereas probability represents one specific assignment for an individual event.

NOTE 2 This definition refers to probability as probability of a specific event. Probability can be related to a long-run relative frequency of occurrences or to a degree of belief in the likely occurrence of an event. Typically, the probability of an event  $A$  is denoted by  $P(A)$ . The notation  $p(A)$  using the script letter  $p$  is used in contexts where there is the need to explicitly consider the formality of a probability space (2.68).

## 2.6

### conditional probability

 $P(A|B)$ 

probability (2.5) of the intersection of  $A$  and  $B$  divided by the probability of  $B$

EXAMPLE 1 Continuing the battery Example 1 of 2.1, consider the event (2.2)  $A$  defined as {the battery survives at least three hours}, namely  $[3, \infty)$ . Let the event  $B$  be defined as {the battery functioned initially}, namely  $(0, \infty)$ . The conditional probability of

động từ ban đầu.

VÍ DỤ 2: Tiếp theo ví dụ 2 của 2.1, nếu là lựa chọn không hoàn lại, xác suất chọn điện trở 2 trong lần lấy mẫu thứ hai là bằng không nếu biết trước nó đã được chọn trong lần lấy mẫu thứ nhất. Nếu các xác suất đối với mọi điện trở được chọn là bằng nhau thì xác suất chọn điện trở 2 trong lần lấy mẫu thứ hai bằng 0,111 1 nếu biết trước nó không được chọn trong lần lấy mẫu thứ nhất.

VÍ DỤ 3: Tiếp tục với ví dụ 2 của 2.1, nếu là lựa chọn có hoàn lại và các xác suất bằng nhau đối với tất cả các điện trở được chọn trong mỗi lần lấy mẫu thì xác suất chọn điện trở 2 trong lần lấy mẫu thứ hai sẽ là 0,1 hoặc là điện trở 2 đã được chọn trong lần đầu hoặc là không được chọn trong lần đầu. Vì vậy các kết quả của lần lấy mẫu thứ nhất và thứ hai là các biến cố độc lập.

CHÚ THÍCH 1: Xác suất của biến cố  $B$  cần phải lớn hơn không.

CHÚ THÍCH 2: " $A$  cho trước  $B$ " có thể nêu đầy đủ hơn là "biến cố  $A$  cho trước biến cố  $B$  đã xảy ra". Vạch thẳng đứng trong ký hiệu đối với xác suất có điều kiện đọc là "cho trước".

CHÚ THÍCH 3: Nếu xác suất có điều kiện của biến cố  $A$  cho trước biến cố  $B$  xảy ra bằng xác suất xảy ra biến cố  $A$ , thì khi đó biến cố  $A$  và  $B$  là độc lập. Nói cách khác, việc biết được sự xuất hiện của  $B$  không ảnh hưởng đến xác suất của  $A$ .

## 2.7

### Hàm phân bố của biến ngẫu nhiên $X$

$F(x)$

Hàm số của  $x$  cho biết xác suất (2.5) của biến cố (2.2)  $(-\infty, x]$ .

CHÚ THÍCH 1: Khoảng  $(-\infty, x]$  là tập hợp tất cả các giá trị nhỏ hơn và bao gồm cả  $x$ .

CHÚ THÍCH 2: Hàm phân bố mô tả toàn bộ phân bố

$A$  given  $B$  takes into account that one is dealing with the initially functional batteries.

EXAMPLE 2 Continuing with Example 2 of 2.1, if the selection is without replacement, the probability of selecting resistor 2 in the second draw is equal to zero given that it has been selected in the first draw. If the probabilities are equal for all resistors to be selected, the probability for selecting resistor 2 in the second draw equals 0,111 1 given that it has not been selected in the first draw.

EXAMPLE 3 Continuing with Example 2 of 2.1, if the selection is done with replacement and the probabilities are the same for all resistors to be selected within each draw, then the probability of selecting resistor 2 in the second draw will be 0,1 either if resistor 2 has been selected in the first draw or if it is not selected in the first draw. Thus the outcomes of the first and the second draw are independent events.

NOTE 1 The probability of the event  $B$  is required to be greater than zero.

NOTE 2 " $A$  given  $B$ " can be stated more fully as "the event  $A$  given the event  $B$  has occurred". The vertical bar in the symbol for conditional probability is pronounced "given".

NOTE 3 If the conditional probability of the event  $A$  given that the event  $B$  occurred is equal to the probability of  $A$  occurring, the events  $A$  and  $B$  are independent. In other words, the knowledge of occurrence of  $B$  suggests no adjustment to the probability of  $A$ .

## 2.7

### distribution function of a random variable $X$

$F(x)$

function of  $x$  giving the probability (2.5) of the event (2.2)  $(-\infty, x]$

NOTE 1 The interval  $(-\infty, x]$  is the set of all values up to and including  $x$ .

NOTE 2 The distribution function completely

xác suất (2.11) của biến ngẫu nhiên (2.10). Phân loại phân bố cũng như phân loại biến ngẫu nhiên thành các lớp rời rạc hoặc liên tục đều dựa trên phân loại hàm phân bố.

CHÚ THÍCH 3: Vì biến ngẫu nhiên lấy giá trị là các số thực hoặc bộ có thứ tự của  $k$  số thực, hàm ý trong định nghĩa này là  $x$  cũng là một số thực hoặc bộ có thứ tự của  $k$  số thực. Hàm phân bố đối với phân bố nhiều chiều (2.17) cho biết xác suất (2.5) mà mỗi biến ngẫu nhiên đơn của phân bố nhiều chiều nhỏ hơn hoặc bằng giá trị quy định. Về ký hiệu, hàm phân bố nhiều chiều được cho bởi  $F(x_1, x_2, \dots, x_n) = P[X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n]$ . Hàm phân bố cũng là hàm không giảm. Trường hợp một chiều, hàm phân bố được cho bởi  $F(x) = P[X \leq x]$ , đưa ra xác suất của biến cố là biến ngẫu nhiên  $X$  lấy các giá trị nhỏ hơn hoặc bằng  $x$ .

CHÚ THÍCH 4: Thông thường, hàm phân bố được chia thành: hàm phân bố rời rạc (2.22) và hàm phân bố liên tục (2.23) nhưng cũng có các khả năng khác. Trờ lại ví dụ về pin ở 2.1, hàm phân bố có thể như sau:

$$F(x) = \begin{cases} 0 & \text{nếu } x < 0 \\ 0,1 & \text{nếu } x = 0 \\ 0,1 + 0,9[1 - \exp(-x)] & \text{nếu } x > 0 \end{cases}$$

Từ quy định về hàm phân bố này, tuổi thọ của pin là không âm. Có 10 % cơ hội pin không hoạt động ngay từ lần đầu. Nếu trên thực tế ban đầu pin không hoạt động thì tuổi thọ của pin có phân bố mũ (2.58) với tuổi thọ trung bình là 1 h.

CHÚ THÍCH 5: Chữ viết tắt cdf (hàm phân bố tích lũy) thường được dùng cho hàm phân bố.

describes the probability distribution (2.11) of the random variable (2.10). Classifications of distributions as well as classifications of random variables into discrete or continuous classes are based on classifications of distribution functions.

NOTE 3 Since random variables take values that are real numbers or ordered  $k$ -tuples of real numbers, it is implicit in the definition that  $x$  is also a real number or an ordered  $k$ -tuple of real numbers. The distribution function for a multivariate distribution (2.17) gives the probability (2.5) that each of the random variables of the multivariate distribution is less than or equal to a specified value. Notationally, a multivariate distribution function is given by  $F(x_1, x_2, \dots, x_n) = P[X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n]$ . Also, a distribution function is non-decreasing. In a univariate setting, the distribution function is given by  $F(x) = P[X \leq x]$ , which gives the probability of the event that the random variable  $X$  takes on a value less than or equal to  $x$ .

NOTE 4 Commonly, distribution functions are classified into discrete distribution (2.22) functions and continuous distribution (2.23) functions but there are other possibilities. Recalling the battery example of 2.1, one possible distribution function is, as follows:

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 0,1 & \text{if } x = 0 \\ 0,1 + 0,9[1 - \exp(-x)] & \text{if } x > 0 \end{cases}$$

From this specification of the distribution function, battery life is non-negative. There is a 10 % chance that the battery does not function on the initial attempt. If the battery does in fact function initially, then its battery life has an exponential distribution (2.58) with mean life of 1 h.

NOTE 5 Often the abbreviation cdf (cumulative distribution function) is given for distribution function.

**Họ phân bố**

Tập hợp các phân bố xác suất (2.11).

CHÚ THÍCH 1: Tập hợp các phân bố xác suất thường được xác định bởi tham số (2.9) của phân bố xác suất.

CHÚ THÍCH 2: Trung bình (2.35) và/hoặc phương sai (2.36) của phân bố xác suất thường được dùng như chỉ số của họ phân bố hoặc bộ phận của chỉ số trong trường hợp cần nhiều hơn hai tham số để xác định họ phân bố. Trong các trường hợp khác, trung bình và phương sai không nhất thiết phải là các tham số rõ ràng trong họ phân bố mà đúng hơn là hàm của các tham số.

**2.9****Tham số**

Chỉ số của họ phân bố (2.8).

CHÚ THÍCH 1: Tham số có thể là một chiều hoặc nhiều chiều.

CHÚ THÍCH 2: Tham số đôi khi được gọi là tham số định vị, đặc biệt nếu tham số tương ứng trực tiếp với trung bình của họ phân bố. Một số tham số được mô tả như tham số thang đo, đặc biệt nếu chúng chính là hoặc tỷ lệ với độ lệch chuẩn (2.37) của phân bố. Các tham số không phải là tham số định vị cũng không phải là tham số thang đo thường được gọi là tham số định dạng.

**2.10****Biến ngẫu nhiên**

Hàm xác định trên không gian mẫu (2.1) trong đó các giá trị của hàm này là bộ có thứ tự của  $k$  số thực.

VÍ DỤ: Tiếp theo ví dụ về pin trong 2.1, không gian mẫu gồm các biến cố được mô tả bằng lời (pin hỏng ngay lần đầu tiên, ban đầu pin hoạt động nhưng sau đó hỏng ở  $x$  giờ). Các biến cố như vậy khó giải quyết bằng toán học, do đó phải kèm theo mỗi biến cố thời gian tại đó pin hỏng (cho bằng số thực). Nếu biến

**family of distributions**

set of probability distributions (2.11)

NOTE 1 The set of probability distributions is often indexed by a parameter (2.9) of the probability distribution.

NOTE 2 Often the mean (2.35) and/or the variance (2.36) of the probability distribution is used as the index of the family of distributions or as part of the index in cases where more than two parameters are needed to index the family of distributions. On other occasions, the mean and variance are not necessarily explicit parameters in the family of distributions but rather a function of the parameters.

**2.9****parameter**

index of a family of distributions (2.8)

NOTE 1 The parameter may be one-dimensional or multi-dimensional.

NOTE 2 Parameters are sometimes referred to as location parameters, particularly if the parameter corresponds directly to the mean of the family of distributions. Some parameters are described as scale parameters, particularly if they are exactly or proportional to the standard deviation (2.37) of the distribution. Parameters that are neither location nor scale parameters are generally referred to as shape parameters.

**2.10****random variable**

function defined on a sample space (2.1) where the values of the function are ordered  $k$ -tuplets of real numbers

EXAMPLE Continuing the battery example introduced in 2.1, the sample space consists of events which are described in words (battery fails upon initial attempt, battery works initially but then fails at  $x$  hours). Such events are awkward to work with mathematically, so it is natural to associate with

ngẫu nhiên lấy giá trị 0, thì ta có thể thấy rằng kết quả này tương ứng với sự cố xảy ra ngay từ đầu. Đối với giá trị của biến ngẫu nhiên lớn hơn không, có thể hiểu là ban đầu pin hoạt động và sau đó hỏng ở giá trị cụ thể này. Biểu diễn của biến ngẫu nhiên cho phép trả lời câu hỏi: "xác suất pin có tuổi thọ vượt quá thời gian bảo hành của nó, nghĩa là 6 h, là bao nhiêu?".

CHÚ THÍCH 1: Ví dụ về bộ  $k$  có thứ tự là  $(x_1, x_2, \dots, x_k)$ . Nói cách khác, bộ có thứ tự  $k$  là vectơ  $k$  chiều (vectơ hàng hoặc cột).

CHÚ THÍCH 2: Thông thường, biến ngẫu nhiên có số chiều biểu thị bằng  $k$ . Nếu  $k = 1$  thì biến ngẫu nhiên được gọi là một chiều hoặc một biến. Đối với  $k > 1$ , biến ngẫu nhiên được gọi là nhiều chiều. Trên thực tế, khi số chiều là một số cho trước,  $k$ , biến ngẫu nhiên được gọi là  $k$  chiều.

CHÚ THÍCH 3: Biến ngẫu nhiên một chiều là hàm giá trị thực xác định theo không gian mẫu (2.1) là bộ phận của không gian xác suất (2.68).

CHÚ THÍCH 4: Biến ngẫu nhiên có giá trị thực được cho như các cặp thứ tự được gọi là hai chiều. Định nghĩa này mở rộng khái niệm cặp thứ tự sang bộ  $k$  có thứ tự.

CHÚ THÍCH 5: Thành phần thứ  $j$  của biến ngẫu nhiên  $k$ -chiều là biến ngẫu nhiên tương ứng với thành phần thứ  $j$  duy nhất của bộ  $k$ . Thành phần thứ  $j$  của biến ngẫu nhiên  $k$  chiều tương ứng với không gian xác suất trong đó biến cố (2.2) chỉ được xác định về giá trị của thành phần được xét.

2.11

**Phân bố xác suất**

**Phân bố**

**Độ đo xác suất** (2.70) được xác định bởi một biến ngẫu nhiên (2.10).

VÍ DỤ: Tiếp theo ví dụ về pin ở 2.1, phân bố tuổi thọ

each event, the time (given as a real number) at which the battery fails. If the random variable takes the value 0, then one would recognize that this outcome corresponds to an initial failure. For a value of the random variable greater than zero, it would be understood that the battery initially worked and then subsequently failed at this specific value. The random variable representation allows one to answer questions such as, "what is the probability that the battery exceeds its warranty life, i.e. 6 h?".

NOTE 1 An example of an ordered  $k$ -tuple is  $(x_1, x_2, \dots, x_k)$ . An ordered  $k$ -tuple is, in other words, a vector in  $k$  dimensions (either a row or column vector).

NOTE 2 Typically, the random variable has dimension denoted by  $k$ . If  $k = 1$ , the random variable is said to be one-dimensional or univariate. For  $k > 1$ , the random variable is said to be multi-dimensional. In practice, when the dimension is a given number,  $k$ , the random variable is said to be  $k$ -dimensional.

NOTE 3 A one-dimensional random variable is a real-valued function defined on the sample space (2.1) that is part of a probability space (2.68).

NOTE 4 A random variable with real values given as ordered pairs is said to be two-dimensional. The definition extends the ordered pair concept to ordered  $k$ -tuples.

NOTE 5 The  $j^{\text{th}}$  component of a  $k$ -dimensional random variable is the random variable corresponding to only the  $j^{\text{th}}$  component of the  $k$ -tuple. The  $j^{\text{th}}$  component of a  $k$ -dimensional random variable corresponds to a probability space where events (2.2) are determined only in terms of values of the component considered.

2.11

**probability distribution**

**distribution**

**probability measure** (2.70) induced by a random variable (2.10)

EXAMPLE Continuing with the battery example from

của pin mô tả hoàn toàn các xác suất xuất hiện các giá trị cụ thể. Không biết chắc thời gian hỏng của pin sẽ là bao nhiêu cũng như không biết (trước thử nghiệm) việc pin có hoạt động ngay từ ban đầu hay không. Phân bố xác suất mô tả hoàn toàn tính chất xác suất của một kết quả không chắc chắn. Trong Chú thích 2 của 2.7, đưa ra một khả năng biểu diễn phân bố xác suất, gọi là hàm phân bố.

**CHÚ THÍCH 1:** Có nhiều cách biểu diễn toán học tương đương cho phân bố bao gồm **hàm phân bố (2.7)**, **hàm mật độ xác suất (2.27)**, nếu có, và **hàm đặc trưng**. Với các mức độ khó khác nhau, các cách biểu diễn này cho phép xác định xác suất tại đó biến ngẫu nhiên lấy giá trị trong vùng đã cho.

**CHÚ THÍCH 2:** Vì biến ngẫu nhiên là hàm số từ các tập hợp con của không gian mẫu lên đường thẳng thực nên, chẳng hạn, xác suất của biến ngẫu nhiên nhận giá trị thực bất kỳ là 1. Đối với ví dụ về pin,  $P\{X \geq 0\} = 1$ . Trong nhiều trường hợp, nếu xử lý trực tiếp biến ngẫu nhiên và một trong các cách thể hiện của nó sẽ dễ dàng hơn nhiều so với việc liên hệ đến độ đo xác suất. Tuy nhiên, khi chuyển đổi từ cách thể hiện này sang cách thể hiện khác, độ đo xác suất đảm bảo được tính nhất quán.

**CHÚ THÍCH 3:** Biến ngẫu nhiên có một thành phần được gọi là phân bố xác suất một chiều hoặc đơn biến. Nếu biến ngẫu nhiên có hai thành phần, thì đó là phân bố xác suất hai chiều hoặc hai biến, còn nếu có nhiều hơn hai thành phần thì ta nói rằng biến ngẫu nhiên có phân bố xác suất nhiều chiều hoặc đa biến.

## 2.12

### Kỳ vọng

Tích phân của hàm của **biến ngẫu nhiên (2.10)** theo **độ đo xác suất (2.70)** lấy trên toàn bộ **không gian mẫu (2.1)**.

2.1, the distribution of battery life completely describes the probabilities with which specific values occur. It is not known with certainty what the failure time of a given battery will be nor is it known (prior to testing) if the battery will even function upon the initial attempt. The probability distribution completely describes the probabilistic nature of an uncertain outcome. In Note 4 of 2.7, one possible representation of the probability distribution was given, namely a distribution function.

**NOTE 1** There are numerous, equivalent mathematical representations of a distribution including **distribution function (2.7)**, **probability density function (2.27)**, if it exists, and **characteristic function**. With varying levels of difficulty, these representations allow for determining the probability with which a random variable takes values in a given region.

**NOTE 2** Since a random variable is a function on subsets of the sample space to the real line, it is the case, for example, that the probability that a random variable takes on any real value is 1. For the battery example,  $P\{X \geq 0\} = 1$ . In many situations, it is much easier to deal directly with the random variable and one of its representations than to be concerned with the underlying probability measure. However, in converting from one representation to another, the probability measure ensures the consistency.

**NOTE 3** A random variable with a single component is called a one-dimensional or univariate probability distribution. If a random variable has two components, one speaks about a two-dimensional or bivariate probability distribution, and with more than two components, the random variable has a multidimensional or multivariate probability distribution.

## 2.12

### expectation

integral of a function of a **random variable (2.10)** with respect to a **probability measure (2.70)** over the **sample space (2.1)**

CHÚ THÍCH 1: Kỳ vọng của hàm  $g$  của biến ngẫu nhiên  $X$  được biểu thị bằng  $E[g(X)]$  và được tính bằng:

$$E[g(X)] = \int_{\Omega} g(X)dP = \int_{R^k} g(x)dF(x)$$

trong đó  $F(x)$  là hàm phân bố tương ứng.

CHÚ THÍCH 2: “E” trong  $E[g(X)]$  lấy từ “giá trị kỳ vọng” hoặc “kỳ vọng” của biến ngẫu nhiên  $X$ . E có thể được coi như một toán tử hoặc hàm số chiếu biến ngẫu nhiên lên đường thẳng thực theo công thức trên.

CHÚ THÍCH 3: Có hai tích phân được cho đối với  $E[g(X)]$ . Tích phân đầu tiên lấy trên không gian mẫu chỉ có nghĩa về khái niệm chứ không dùng trong thực tiễn. Tích phân thứ hai mô tả việc tính toán không gian  $R^k$ , được quan tâm nhiều hơn trong thực tiễn.

CHÚ THÍCH 4: Trong nhiều trường hợp, tích phân nói trên rút gọn về dạng nhận biết được từ phép tính. Các ví dụ được cho trong chú thích của moment bậc  $r$  (2.34) trong đó  $g(x) = x^r$ , trung bình (2.35) trong đó  $g(x) = x$  và phương sai (2.36) trong đó  $g(x) = [x - E(X)]^2$ .

CHÚ THÍCH 5: Định nghĩa này không giới hạn ở tích phân một chiều như các ví dụ và chú thích trước. Đối với trường hợp số chiều lớn hơn, xem 2.43.

CHÚ THÍCH 6: Đối với biến ngẫu nhiên rời rạc (2.28), tích phân thứ hai trong chú thích 1 được thay bằng ký hiệu tổng. Các ví dụ được cho trong 2.35.

NOTE 1 The expectation of the function  $g$  of a random variable  $X$  is denoted by  $E[g(X)]$  and is computed as:

$$E[g(X)] = \int_{\Omega} g(X)dP = \int_{R^k} g(x)dF(x)$$

where  $F(x)$  is the corresponding distribution function.

NOTE 2 The “E” in  $E[g(X)]$  comes from the “expected value” or “expectation” of the random variable  $X$ . E can be viewed as an operator or function that maps a random variable to the real line according to the above calculation.

NOTE 3 Two integrals are given for  $E[g(X)]$ . The first treats the integration over the sample space which is conceptually appealing but not of practical use. The second integral depicts the calculation over the  $R^k$ , which is of greater practical interest.

NOTE 4 In many cases of practical interest, the above integral reduces to a form recognizable from calculus. Examples are given in the notes to moment of order  $r$  (2.34) where  $g(x) = x^r$ , mean (2.35) where  $g(x) = x$  and variance (2.36) where  $g(x) = [x - E(X)]^2$ .

NOTE 5 The definition is not restricted to one-dimensional integrals as the previous examples and notes might suggest. For higher dimensional situations, see 2.43.

NOTE 6 For a discrete random variable (2.28), the second integral in Note 1 is replaced by the summation symbol. Examples can be found in 2.35.

2.13

*p*-phân vị

*p*-fractile

$$X_p, x_p$$

Giá trị của  $x$  bằng cận dưới của tất cả các  $x$  sao cho hàm phân bố (2.7)  $F(x)$  lớn hơn hoặc bằng  $p$ , đối với  $0 < p < 1$ .

VÍ DỤ 1: Xét phân bố nhị thức (2.46) có hàm khối lượng xác suất cho trong Bảng 2. Tập hợp các giá trị này ứng với phân bố nhị thức có tham số  $n = 6$  và  $p =$

2.13

*p*-quantile

*p*-fractile

$$X_p, x_p$$

value of  $x$  equal to the infimum of all  $x$  such that the distribution function (2.7)  $F(x)$  is greater than or equal to  $p$ , for  $0 < p < 1$

EXAMPLE 1 Consider a binomial distribution (2.46) with probability mass function given in Table 2. This set of values corresponds to a binomial



0,3. Đối với trường hợp này, một số  $p$ -phân vị được chọn là:

$$\begin{aligned}x_{0,1} &= 0 \\x_{0,25} &= 1 \\x_{0,5} &= 2 \\x_{0,75} &= 3 \\x_{0,90} &= 3 \\x_{0,95} &= 4 \\x_{0,99} &= 5 \\x_{0,999} &= 5\end{aligned}$$

Tính rời rạc của phân bố nhị thức dẫn đến các giá trị tích phân của các  $p$ -phân vị.

**Bảng 2 – Ví dụ về phân bố nhị thức**

$X$	$P\{X=x\}$	$P\{X \leq x\}$	$P\{X > x\}$
0	0,117 649	0,117 649	0,882 351
1	0,302 526	0,420 175	0,579 825
2	0,324 135	0,744 310	0,255 690
3	0,185 220	0,929 530	0,070 470
4	0,059 535	0,989 065	0,010 935
5	0,010 206	0,999 271	0,000 729
6	0,000 729	1,000 000	0,000 000

VÍ DỤ 2: Xét một phân bố chuẩn chuẩn hóa (2.51) có các giá trị được chọn từ hàm phân bố cho trong Bảng 3. Một số  $p$ -phân vị được chọn là:

**Bảng 3 – Ví dụ về phân bố chuẩn chuẩn hóa**

$p$	$x$ sao cho $P\{X \leq x\} = p$
0,1	-1,282
0,25	-0,674
0,5	0,000
0,75	0,674
0,841 344 75	1,000
0,9	1,282
0,95	1,645
0,975	1,960
0,99	2,326
0,995	2,576
0,999	3,090

distribution with parameters  $n = 6$  and  $p = 0,3$ . For this case, some selected  $p$ -quantiles are:

$$\begin{aligned}x_{0,1} &= 0 \\x_{0,25} &= 1 \\x_{0,5} &= 2 \\x_{0,75} &= 3 \\x_{0,90} &= 3 \\x_{0,95} &= 4 \\x_{0,99} &= 5 \\x_{0,999} &= 5\end{aligned}$$

The discreteness of the binomial distribution leads to integral values of the  $p$ -quantiles.

**Table 2 – Binomial distribution example**

$X$	$P\{X=x\}$	$P\{X \leq x\}$	$P\{X > x\}$
0	0,117 649	0,117 649	0,882 351
1	0,302 526	0,420 175	0,579 825
2	0,324 135	0,744 310	0,255 690
3	0,185 220	0,929 530	0,070 470
4	0,059 535	0,989 065	0,010 935
5	0,010 206	0,999 271	0,000 729
6	0,000 729	1,000 000	0,000 000

EXAMPLE 2 Consider a standardized normal distribution (2.51) with selected values from its distribution function given in Table 3. Some selected  $p$ -quantiles are:

**Table 3 – Standardized normal distribution example**

$p$	$x$ such that $P\{X \leq x\} = p$
0,1	-1,282
0,25	-0,674
0,5	0,000
0,75	0,674
0,841 344 75	1,000
0,9	1,282
0,95	1,645
0,975	1,960
0,99	2,326
0,995	2,576
0,999	3,090

Vi phân bố của  $X$  là liên tục nên tiêu đề của cột thứ hai cũng có thể là:  $x$  sao cho  $P\{X < x\} = p$ .

**CHÚ THÍCH 1:** Đối với phân bố liên tục (2.23), nếu  $p$  là 0,5 thì 0,5-phân vị ứng với trung vị (2.14). Đối với  $p$  bằng 0,25, 0,25-phân vị được coi là tứ phân vị dưới. Đối với phân bố liên tục, 25 % của phân bố thấp hơn 0,25 phân vị trong khi 75 % là cao hơn 0,25 phân vị. Đối với  $p$  bằng 0,75, 0,75-phân vị được coi là tứ phân vị trên.

**CHÚ THÍCH 2:** Nói chung, 100  $p$  % phân bố là thấp hơn  $p$ -phân vị; 100(1 -  $p$ ) % phân bố cao hơn  $p$ -phân vị. Việc xác định trung vị gặp khó khăn đối với các phân bố rời rạc vì có thể phải thảo luận vì còn có nhiều giá trị thỏa mãn định nghĩa.

**CHÚ THÍCH 3:** Nếu  $F$  liên tục và tăng ngặt thì  $p$ -phân vị là lời giải cho  $F(x) = p$ . Trong trường hợp này, từ "cận dưới" trong định nghĩa có thể thay bằng "tối thiểu".

**CHÚ THÍCH 4:** Nếu hàm phân bố là hằng số và bằng  $p$  trong một khoảng thì tất cả các giá trị trong khoảng đó là  $p$ -phân vị đối với  $F$ .

**CHÚ THÍCH 5:**  $p$ -phân vị được xác định cho các phân bố một chiều (2.16).

## 2.14

### Trung vị

0,5-phân vị (2.13).

**VÍ DỤ:** Đối với ví dụ về pin ở chú thích 4 trong 2.7, trung vị là 0,587 8, là lời giải cho  $x$  trong  $0,1 + 0,9[1 - \exp(-x)] = 0,5$ .

**CHÚ THÍCH 1:** Trung vị là một trong các  $p$ -phân vị (2.13) được áp dụng phổ biến nhất trong ứng dụng thực tế. Trung vị của một phân bố một chiều (2.16) liên tục là giá trị sao cho một nửa của tổng thể (1.1) lớn hơn hoặc bằng trung vị và một nửa của tổng thể nhỏ hơn hoặc bằng trung vị.

**CHÚ THÍCH 2:** Trung vị được xác định cho các phân bố một chiều (2.16).

Since the distribution of  $X$  is continuous, the second column heading could also be:  $x$  such that  $P\{X < x\} = p$ .

**NOTE 1** For continuous distributions (2.23), if  $p$  is 0,5 then the 0,5-quantile corresponds to the median (2.14). For  $p$  equal to 0,25, the 0,25-quantile is known as the lower quartile. For continuous distributions, 25 % of the distribution is below the 0,25 quantile while 75 % is above the 0,25 quantile. For  $p$  equal to 0,75, the 0,75-quantile is known as the upper quartile.

**NOTE 2** In general, 100  $p$  % of a distribution is below the  $p$ -quantile; 100(1 -  $p$ ) % of a distribution is above the  $p$ -quantile. There is a difficulty in defining the median for discrete distributions since it could be argued to have multiple values satisfying the definition.

**NOTE 3** If  $F$  is continuous and strictly increasing, the  $p$ -quantile is the solution to  $F(x) = p$ . In this case, the word "infimum" in the definition could be replaced by "minimum".

**NOTE 4** If the distribution function is constant and equal to  $p$  in an interval, then all values in that interval are  $p$ -quantiles for  $F$ .

**NOTE 5**  $p$ -quantiles are defined for univariate distributions (2.16).

## 2.14

### median

0,5-quantile (2.13)

**EXAMPLE** For the battery example of Note 4 in 2.7, the median is 0,587 8, which is the solution for  $x$  in  $0,1 + 0,9[1 - \exp(-x)] = 0,5$

**NOTE 1** The median is one of the most commonly applied  $p$ -quantiles (2.13) in practical use. The median of a continuous univariate distribution (2.16) is such that half of the population (1.1) is greater than or equal to the median and half of the population is less than or equal to the median.

**NOTE 2** Medians are defined for univariate distributions (2.16).

**2.15****Tứ phân vị**

0,25-phân vị (2.13) hoặc 0,75-phân vị.

VÍ DỤ: Tiếp theo ví dụ về pin ở 2.14, có thể thấy rằng 0,25-phân vị là 0,182 3 và 0,75-phân vị là 1,280 9.

CHÚ THÍCH 1: 0,25 phân vị cũng được coi là **tứ phân vị dưới**, trong khi 0,75 phân vị cũng được gọi là **tứ phân vị trên**.

CHÚ THÍCH 2: Tứ phân vị được xác định cho các **phân bố một chiều** (2.16).

**2.16****Phân bố xác suất đơn biến****Phân bố một chiều**

**Phân bố xác suất** (2.11) của một **biến ngẫu nhiên** (2.10).

CHÚ THÍCH: Phân bố xác suất đơn biến là một chiều. **Phân bố nhị thức** (2.46), **Poisson** (2.47), **phân bố chuẩn** (2.50), **phân bố gamma** (2.56), **phân bố  $t$**  (2.53), **phân bố Weibull** (2.63) và **phân bố beta** (2.59) là ví dụ của phân bố xác suất một chiều.

**2.17****Phân bố xác suất đa biến****Phân bố nhiều chiều**

**Phân bố xác suất** (2.11) của hai hay nhiều **biến ngẫu nhiên** (2.10).

CHÚ THÍCH 1: Vì trường hợp phân bố một chiều phổ biến hơn nên khi không có quy định nào khác thì thường giả định là phân bố một chiều.

CHÚ THÍCH 2: Phân bố nhiều chiều đôi khi được gọi là phân bố đồng thời.

CHÚ THÍCH 3: **Phân bố đa thức** (2.45), **phân bố chuẩn hai chiều** (2.65) và **phân bố chuẩn đa biến** (2.64) là các ví dụ của phân bố xác suất đa biến được đề cập trong tiêu chuẩn này.

**2.15****quartile**

0,25-quantile (2.13) or 0,75-quantile

EXAMPLE Continuing with the battery example of 2.14, it can be shown that the 0,25-quantile is 0,182 3 and the 0,75-quantile is 1,280 9.

NOTE 1 The 0,25 quartile is also known as the **lower quartile**, while the 0,75 quartile is also known as the **upper quartile**.

NOTE 2 Quartiles are defined for **univariate distributions** (2.16).

**2.16****univariate probability distribution****univariate distribution**

**probability distribution** (2.11) of a single **random variable** (2.10)

NOTE Univariate probability distributions are one-dimensional. The **binomial** (2.46), **Poisson** (2.47), **normal** (2.50), **gamma** (2.56),  **$t$**  (2.53), **Weibull** (2.63) and **beta** (2.59) distributions are examples of univariate probability distributions.

**2.17****multivariate probability distribution****multivariate distribution**

**probability distribution** (2.11) of two or more **random variables** (2.10)

NOTE 1 Since univariate distribution is in preponderance, it is customary to presume a univariate situation unless otherwise stated.

NOTE 2 The multivariate distribution is sometimes referred to as the joint distribution.

NOTE 3 The **multinomial distribution** (2.45), **bivariate normal distribution** (2.65) and the **multivariate normal distribution** (2.64) are examples of multivariate probability distributions covered in this part of ISO 3534.

## 2.18

**Phân bố xác suất biên duyên****Phân bố biên duyên**

**Phân bố xác suất** (2.11) của một tập hợp con thực sự, không rỗng gồm các thành phần của một **biến ngẫu nhiên** (2.10).

VÍ DỤ 1: Đối với một phân bố có ba biến ngẫu nhiên  $X$ ,  $Y$  và  $Z$ , sẽ có ba phân bố biên duyên với hai biến ngẫu nhiên, đó là  $(X, Y)$ ,  $(X, Z)$  và  $(Y, Z)$  và ba phân bố biên duyên có một biến ngẫu nhiên là  $X$ ,  $Y$  và  $Z$ .

VÍ DỤ 2: Đối với **phân bố chuẩn hai chiều** (2.65) của cặp biến  $(X, Y)$ , phân bố của từng biến  $X$  và  $Y$  xét riêng rẽ là các phân bố biên duyên và đều là **phân bố chuẩn** (2.50).

VÍ DỤ 3: Đối với **phân bố đa thức** (2.45), phân bố của  $(X_1, X_2)$  là phân bố biên duyên  $k > 3$ . Phân bố của  $X_1, X_2, \dots, X_k$  riêng rẽ cũng là các phân bố biên duyên. Các phân bố biên duyên này đều là **phân bố nhị thức** (2.46).

CHÚ THÍCH 1: Đối với phân bố đồng thời  $k$  chiều, một ví dụ về phân bố biên duyên bao gồm phân bố xác suất của một tập con gồm  $k_1 < k$  biến ngẫu nhiên.

CHÚ THÍCH 2: Cho một phân bố xác suất nhiều chiều (2.17) liên tục (2.23) biểu diễn bởi hàm mật độ xác suất (2.26), hàm mật độ xác suất của phân bố xác suất biên duyên được xác định bằng tích phân hàm mật độ xác suất trong phạm vi các biến không được xét trong phân bố biên duyên.

CHÚ THÍCH 3: Cho một phân bố xác suất nhiều chiều rời rạc (2.22) biểu diễn bởi hàm khối lượng xác suất (2.24), hàm khối lượng xác suất của phân bố xác suất biên duyên được xác định bằng tổng hàm khối lượng xác suất trong phạm vi các biến không được xét trong phân bố biên duyên.

## 2.18

**marginal probability distribution****marginal distribution**

**probability distribution** (2.11) of a non-empty, strict subset of the components of a **random variable** (2.10)

EXAMPLE 1 For a distribution with three random variables  $X$ ,  $Y$  and  $Z$ , there are three marginal distributions with two random variables, namely for  $(X, Y)$ ,  $(X, Z)$  and  $(Y, Z)$  and three marginal distributions with a single random variable, namely for  $X$ ,  $Y$  and  $Z$ .

EXAMPLE 2 For the bivariate normal distribution (2.65) of the pair of variables  $(X, Y)$ , the distribution of each of the variables  $X$  and  $Y$  considered separately are marginal distributions, which are both normal distributions (2.50).

EXAMPLE 3 For the multinomial distribution (2.45), the distribution of  $(X_1, X_2)$  is a marginal distribution if  $k > 3$ . The distributions of  $X_1, X_2, \dots, X_k$ , separately are also marginal distributions. These marginal distributions are each binomial distributions (2.46).

NOTE 1 For a joint distribution in  $k$  dimensions, one example of a marginal distribution includes the probability distribution of a subset of  $k_1 < k$  random variables.

NOTE 2 Given a continuous (2.23) multivariate probability distribution (2.17) represented by its probability density function (2.26), the probability density function of its marginal probability distribution is determined by integrating the probability density function over the domain of the variables that are not considered in the marginal distribution.

NOTE 3 Given a discrete (2.22) multivariate probability distribution represented by its probability mass function (2.24), the probability mass function of its marginal probability distribution is determined by summing the probability mass function over the

## 2.19

**Phân bố xác suất có điều kiện****Phân bố có điều kiện**

**Phân bố xác suất** (2.11) giới hạn trong tập con không rỗng của **không gian mẫu** (2.1) và điều chỉnh để có xác suất của toàn bộ không gian mẫu giới hạn.

VÍ DỤ 1: Trong ví dụ pin ở 2.7, chú thích 4, phân bố có điều kiện của tuổi thọ pin dựa vào hàm pin ban đầu là hàm mũ (2.58).

VÍ DỤ 2: Đối với **phân bố chuẩn hai chiều** (2.65), **phân bố xác suất có điều kiện** của  $Y$  cho trước  $X = x$  phản ánh tác động đến  $Y$  khi biết  $X$ .

VÍ DỤ 3: Xét biến ngẫu nhiên  $X$  mô tả phân bố của phí tổn bảo hiểm hàng năm ở Florida do các biến cố bão được công bố. Phân bố này có xác suất khác không là phí tổn hàng năm bằng không do khả năng không có bão tác động đến Florida trong một năm cho trước. Phân bố có điều kiện của phí tổn trong những năm thực tế xảy ra sự việc có thể là mối quan tâm.

CHÚ THÍCH 1: Như một ví dụ về phân bố có hai biến ngẫu nhiên  $X$  và  $Y$ , có các phân bố có điều kiện đối với  $X$  và phân bố có điều kiện đối với  $Y$ . Phân bố của  $X$  lấy điều kiện thông qua  $Y = y$  được biểu thị là "phân bố có điều kiện của  $X$  cho trước  $Y = y$ ", trong khi phân bố của  $Y$  lấy điều kiện  $X = x$  được biểu thị "phân bố có điều kiện của  $Y$  cho trước  $X = x$ ".

CHÚ THÍCH 2: **Phân bố xác suất biên duyên** (2.18) có thể coi như phân bố không điều kiện.

CHÚ THÍCH 3: Ví dụ 1 ở trên minh họa trường hợp phân bố một chiều được điều chỉnh thông qua điều kiện để có được phân bố một chiều khác, mà trong trường hợp này là phân bố khác. Ngược lại, đối với phân bố hàm mũ, phân bố có điều kiện mà sự cố sẽ xảy ra trong giờ tiếp theo, biết rằng không có sự cố nào xảy ra trong vòng 10 h đầu tiên, là hàm mũ có cùng tham số.

domain of the variables that are not considered in the marginal distribution.

## 2.19

**conditional probability distribution****conditional distribution**

**probability distribution** (2.11) restricted to a non-empty subset of the **sample space** (2.1) and adjusted to have total probability one on the restricted sample space

EXAMPLE 1 In the battery example of 2.7, Note 4, the conditional distribution of battery life given that the battery functions initially is **exponential** (2.58).

EXAMPLE 2 For the **bivariate normal distribution** (2.65), the **conditional probability distribution** of  $Y$  given that  $X = x$  reflects the impact on  $Y$  from knowledge of  $X$ .

EXAMPLE 3 Consider a random variable  $X$  depicting the distribution of annual insured loss costs in Florida due to declared hurricane events. This distribution would have a non-zero probability of zero annual loss costs owing to the possibility that no hurricane impacts Florida in a given year. Of possible interest is the conditional distribution of loss costs for those years in which an event actually occurs.

NOTE 1 As an example for a distribution with two random variables  $X$  and  $Y$ , there are conditional distributions for  $X$  and conditional distributions for  $Y$ . A distribution of  $X$  conditioned through  $Y = y$  is denoted as "conditional distribution of  $X$  given  $Y = y$ ", while a distribution of  $Y$  conditioned by  $X = x$  is denoted "conditional distribution of  $Y$  given  $X = x$ ".

NOTE 2 **Marginal probability distributions** (2.18) can be viewed as unconditional distributions.

NOTE 3 Example 1 above illustrates the situation where a univariate distribution is adjusted through conditioning to yield another univariate distribution, which in this case is a different distribution. In contrast, for the exponential distribution, the conditional distribution that a failure will occur within

CHÚ THÍCH 4: Phân bố có điều kiện có thể phát sinh đối với các phân bố rời rạc nhất định trong đó không thể có các kết quả cụ thể. Ví dụ, phân bố Poisson có thể đúng như một mô hình đối với số bệnh nhân ung thư trong một tổng thể gồm các bệnh nhân nhiễm bệnh nếu lấy điều kiện là dương tính hoàn toàn (bệnh nhân u bướu không được định nghĩa là nhiễm bệnh).

CHÚ THÍCH 5: Phân bố có điều kiện phát sinh trong ngữ cảnh giới hạn không gian mẫu về một tập con cụ thể. Đối với  $(X, Y)$  có phân bố chuẩn hai chiều (2.65), có thể quan tâm xét phân bố có điều kiện của  $(X, Y)$  cho trước kết quả phải xuất hiện trong hình vuông đơn vị  $[0, 1] \times [0, 1]$ . Một khả năng khác là phân bố có điều kiện của  $(X, Y)$  cho trước  $X^2 + Y^2 \leq r$ .

Trường hợp này tương ứng với trường hợp trong đó ví dụ một bộ phận đáp ứng dung sai và ta có thể quan tâm thêm các tính chất dựa vào việc đạt được tính năng này.

2.20

**Đường hồi quy**

Tập hợp các giá trị của kỳ vọng (2.12) của xác suất phân bố có điều kiện (2.19) của một biến ngẫu nhiên (2.10)  $Y$  cho trước biến ngẫu nhiên  $X = x$ .

CHÚ THÍCH: Ở đây, đường hồi quy được xác định trong ngữ cảnh  $(X, Y)$  có phân bố hai chiều (xem chú thích 1 của 2.17). Do đó, đây là một khái niệm khác với phân tích hồi quy, trong đó  $Y$  liên quan đến tập hợp các giá trị độc lập xác định trước.

2.21

**Mặt hồi quy**

Tập hợp các giá trị của kỳ vọng (2.12) của phân bố xác suất có điều kiện (2.19) của một biến ngẫu nhiên (2.10)  $Y$  cho trước các biến ngẫu nhiên  $X_1 = x_1$  và  $X_2 = x_2$ .

CHÚ THÍCH: Ở đây, như trong 2.20, mặt hồi quy

the next hour, given that no failures have occurred during the first 10 h, is exponential with the same parameter.

NOTE 4 Conditional distributions can arise for certain discrete distributions where specific outcomes are impossible. For example, the Poisson distribution could serve as a model for number of cancer patients in a population of infected patients if conditioned on being strictly positive (a patient with no tumours is not by definition infected).

NOTE 5 Conditional distributions arise in the context of restricting the sample space to a particular subset. For  $(X, Y)$  having a bivariate normal distribution (2.65) it may be of interest to consider the conditional distribution of  $(X, Y)$  given that the outcome must occur in the unit square  $[0, 1] \times [0, 1]$ . Another possibility is the conditional distribution of  $(X, Y)$  given that  $X^2 + Y^2 \leq r$ . This case corresponds to a situation where for example a part meets a tolerance and one might be interested in further properties based on achieving this performance.

2.20

**regression curve**

collection of values of the expectation (2.12) of the conditional probability distribution (2.19) of a random variable (2.10)  $Y$  given a random variable  $X = x$

NOTE Here, regression curve is defined in the context of  $(X, Y)$  having a bivariate distribution (see Note 1 to 2.17). Hence, it is a different concept than those found in regression analysis in which  $Y$  is related to a deterministic set of independent values.

2.21

**regression surface**

collection of values of the expectation (2.12) of the conditional probability distribution (2.19) of a random variable (2.10)  $Y$  given the random variables  $X_1 = x_1$  and  $X_2 = x_2$

NOTE Here, as in 2.20, regression surface is

được xác định trong ngữ cảnh  $(Y, X_1, X_2)$  là một phân bố nhiều chiều (2.17). Như với đường hồi quy, mặt hồi quy liên quan đến khái niệm khác với trong phân tích hồi quy và phương pháp luận về mặt đáp ứng.

## 2.22

### Phân bố xác suất rời rạc

#### Phân bố rời rạc

**Phân bố xác suất** (2.11) trong đó không gian mẫu  $\Omega(2.1)$  là hữu hạn hoặc vô hạn đếm được.

VÍ DỤ: Ví dụ về phân bố rời rạc trong tiêu chuẩn này là phân bố đa thức (2.45), nhị thức (2.46), Poisson (2.47), siêu hình học (2.48) và nhị thức âm (2.49).

CHÚ THÍCH 1: "Rời rạc" có nghĩa là không gian mẫu có thể được cho trong một danh mục hữu hạn hoặc bắt đầu của danh mục vô hạn trong đó có thứ tự rõ ràng, ví dụ như số khuyết tật là 0, 1, 2, ... Ngoài ra, phân bố nhị thức tương ứng với không gian mẫu hữu hạn  $\{0, 1, 2, \dots, n\}$  trong khi phân bố Poisson tương ứng với không gian mẫu vô hạn đếm được  $\{0, 1, 2, \dots\}$ .

CHÚ THÍCH 2: Các trường hợp có dữ liệu định tính trong lấy mẫu chấp nhận thuộc về phân bố rời rạc.

CHÚ THÍCH 3: Hàm phân bố (2.7) của phân bố rời rạc có giá trị rời rạc.

## 2.23

### Phân bố xác suất liên tục

#### Phân bố liên tục

**Phân bố xác suất** (2.11) trong đó hàm phân bố (2.7) đánh giá tại  $x$  có thể biểu thị như tích phân của hàm không âm từ  $-\infty$  đến  $x$ .

VÍ DỤ: Trường hợp có phân bố liên tục xuất hiện trong hầu hết các trường hợp liên quan đến các biến kiểu dữ liệu trong các ứng dụng công nghiệp.

CHÚ THÍCH 1: Ví dụ về phân bố liên tục là phân bố chuẩn (2.50), phân bố chuẩn chuẩn hóa (2.51),  $t$  (2.53),  $F$  (2.55), gamma (2.56), khi bình phương

defined in the context of  $(Y, X_1, X_2)$  being a multivariate distribution (2.17). As with the regression curve, the regression surface involves a concept distinct from those found in regression analysis and response surface methodology.

## 2.22

### discrete probability distribution

#### discrete distribution

**probability distribution** (2.11) for which the **sample space**  $\Omega(2.1)$  is finite or countably infinite

EXAMPLE Examples of discrete distributions in this document are multinomial (2.45), binomial (2.46), Poisson (2.47), hypergeometric (2.48) and negative binomial (2.49).

NOTE 1 "Discrete" implies that the sample space can be given in a finite list or the beginnings of an infinite list in which the subsequent pattern is apparent, such as the number of defects being 0, 1, 2, ... Additionally, the binomial distribution corresponds to a finite sample space  $\{0, 1, 2, \dots, n\}$  whereas the Poisson distribution corresponds to a countably infinite sample space  $\{0, 1, 2, \dots\}$ .

NOTE 2 Situations with attribute data in acceptance sampling involve discrete distributions.

NOTE 3 The **distribution function** (2.7) of a discrete distribution is discrete valued.

## 2.23

### continuous probability distribution

#### continuous distribution

**probability distribution** (2.11) for which the **distribution function** (2.7) evaluated at  $x$  can be expressed as an integral of a non-negative function from  $-\infty$  to  $x$

EXAMPLE Situations where continuous distributions occur are virtually any of those involving variables type data found in industrial applications.

NOTE 1 Examples of continuous distributions are normal (2.50), standardized normal (2.51),  $t$  (2.53),

(2.57), hàm mũ (2.58), beta (2.59), đều (2.60), cực trị loại I (2.61), cực trị loại II (2.62), cực trị loại III (2.63) và loga chuẩn (2.52).

CHÚ THÍCH 2: Hàm không âm đề cập trong định nghĩa là hàm mật độ xác suất (2.26). Đây là hạn chế quá chặt chẽ để nhấn mạnh rằng hàm phân bố có thể khác biệt ở mọi nơi. Tuy nhiên, đối với các xem xét thực tế, nhiều phân bố liên tục được sử dụng phổ biến có tính chất là đạo hàm của hàm phân bố cung cấp hàm mật độ xác suất tương ứng.

CHÚ THÍCH 3: Trường hợp dữ liệu định lượng trong các ứng dụng lấy mẫu chấp nhận tương ứng với phân bố xác suất liên tục.

## 2.24

### Hàm khối lượng xác suất

hàm <phân bố rời rạc> cho biết xác suất (2.5) để biến ngẫu nhiên (2.10) bằng một giá trị cho trước.

VÍ DỤ 1: Hàm khối lượng xác suất mô tả biến ngẫu nhiên  $X$  bằng số lượng mặt ngửa xuất hiện khi tung ba đồng xu là:

$$P(X=0) = 1/8$$

$$P(X=1) = 3/8$$

$$P(X=2) = 3/8$$

$$P(X=3) = 1/8$$

VÍ DỤ 2: Nhiều hàm khối lượng xác suất được đưa ra trong việc xác định phân bố rời rạc (2.22) thường gặp trong các ứng dụng. Các ví dụ về phân bố rời rạc một chiều bao gồm: phân bố nhị thức (2.46), Poisson (2.47), siêu hình học (2.48) và phân bố nhị thức âm (2.49). Ví dụ về phân bố rời rạc nhiều chiều là phân bố đa thức (2.45).

CHÚ THÍCH 1: Hàm khối lượng xác suất có thể được cho là  $P(X=x_i) = p_i$ , trong đó  $X$  là biến ngẫu nhiên,  $x_i$  là giá trị cho trước và  $p_i$  là xác suất tương ứng.

$F$  (2.55), gamma (2.56), chi-squared (2.57), exponential (2.58), beta (2.59), uniform (2.60), Type I extreme value (2.61), Type II extreme value (2.62), Type III extreme value (2.63), and lognormal (2.52).

NOTE 2 The non-negative function referred to in the definition is the **probability density function** (2.26). It is unduly restrictive to insist that a distribution function be differentiable everywhere. However, for practical considerations, many commonly used continuous distributions enjoy the property that the derivative of the distribution function provides the corresponding probability density function.

NOTE 3 Situations with variables data in acceptance sampling applications correspond to continuous probability distributions.

## 2.24

### probability mass function

<discrete distribution> function giving the probability (2.5) that a random variable (2.10) equals a given value

EXAMPLE 1 The **probability mass function** describing the random variable  $X$  equal to the number of heads resulting from tossing three fair coins is:

$$P(X=0) = 1/8$$

$$P(X=1) = 3/8$$

$$P(X=2) = 3/8$$

$$P(X=3) = 1/8$$

EXAMPLE 2 Various probability mass functions are given in defining common **discrete distributions** (2.22) encountered in applications. Subsequent examples of univariate discrete distributions include the **binomial** (2.46), **Poisson** (2.47), **hypergeometric** (2.48) and **negative binomial** (2.49). An example of a multivariate discrete distribution is the **multinomial** (2.45).

NOTE 1 The probability mass function can be given as  $P(X=x_i) = p_i$ , where  $X$  is the random variable,  $x_i$  is



CHÚ THÍCH 2: Hàm khối lượng xác suất được đưa vào trong  $p$ -phân vị ví dụ 1 của 2.13 bằng cách sử dụng phân bố nhị thức (2.46).

## 2.25

### Mốt của hàm khối lượng xác suất

(Các) giá trị ở đó hàm khối lượng xác suất (2.24) đạt cực đại địa phương.

VÍ DỤ: Phân bố nhị thức (2.46) với  $n = 6$  và  $p = 1/3$  là phân bố một mốt có mốt tại 3.

CHÚ THÍCH: Phân bố rời rạc (2.22) là một mốt nếu hàm khối lượng xác suất của nó chỉ có một mốt, hai mốt nếu hàm khối lượng xác suất có hai mốt và nhiều mốt nếu hàm khối lượng xác suất có nhiều hơn hai mốt.

## 2.26

### Hàm mật độ xác suất

$f(x)$

Hàm không âm khi lấy tích phân từ  $-\infty$  đến  $x$  cho giá trị của hàm phân bố (2.7) tại  $x$  của phân bố liên tục (2.23).

VÍ DỤ 1: Nhiều hàm mật độ xác suất được cho trong việc xác định phân bố xác suất thường gặp trong thực tế. Các ví dụ bao gồm phân bố chuẩn (2.50), chuẩn chuẩn hóa (2.51),  $t$  (2.53),  $F$  (2.55), gamma (2.56), khi bình phương (2.57), phân bố mũ (2.58), beta (2.59), phân bố đều (2.60), phân bố chuẩn nhiều chiều (2.64) và phân bố chuẩn hai chiều (2.65).

VÍ DỤ 2: Đối với hàm phân bố xác định bởi  $F(x) = 3x^2 - 2x^3$  trong đó  $0 \leq x \leq 1$ , hàm mật độ xác suất tương ứng sẽ là  $f(x) = 6x(1 - x)$  trong đó  $0 \leq x \leq 1$ .

VÍ DỤ 3: Tiếp theo ví dụ về pin ở 2.1, không có hàm mật độ xác suất đi kèm với hàm phân bố quy định, do xác suất dương của kết quả không. Tuy nhiên, phân bố có điều kiện cho rằng pin hoạt động lúc đầu có  $f(x) = \exp(-x)$  đối với  $x > 0$  là hàm mật độ xác suất, tương

a given value, and  $p_i$  is the corresponding probability.

NOTE 2 A probability mass function was introduced in the  $p$ -quantile Example 1 of 2.13 using the binomial distribution (2.46).

## 2.25

### mode of probability mass function

value(s) where a probability mass function (2.24) attains a local maximum

EXAMPLE The binomial distribution (2.46) with  $n = 6$  and  $p = 1/3$  is unimodal with mode at 3.

NOTE A discrete distribution (2.22) is unimodal if its probability mass function has exactly one mode, bimodal if its probability mass function has exactly two modes and multi-modal if its probability mass function has more than two modes.

## 2.26

### probability density function

$f(x)$

non-negative function which when integrated from  $-\infty$  to  $x$  gives the distribution function (2.7) evaluated at  $x$  of a continuous distribution (2.23)

EXAMPLE 1 Various probability density functions are given in defining the common probability distributions encountered in practice. Subsequent examples include the normal (2.50), standardized normal (2.51),  $t$  (2.53),  $F$  (2.55), gamma (2.56), chi-squared (2.57), exponential (2.58), beta (2.59), uniform (2.60), multivariate normal (2.64) and bivariate normal distributions (2.65).

EXAMPLE 2 For the distribution function defined by  $F(x) = 3x^2 - 2x^3$  where  $0 \leq x \leq 1$ , the corresponding probability density function is  $f(x) = 6x(1 - x)$  where  $0 \leq x \leq 1$ .

EXAMPLE 3 Continuing with the battery example of 2.1, there does not exist a probability density function associated with the specified distribution function, owing to the positive probability of a zero outcome.

## TCVN 8244-1:2010

ứng với phân bố mũ.

CHÚ THÍCH 1: Nếu hàm phân bố  $F$  là khả vi liên tục thì hàm mật độ xác suất là

$$f(x) = dF(x)/dx$$

tại điểm  $x$  nơi có đạo hàm.

CHÚ THÍCH 2: Đồ thị của  $f(x)$  theo  $x$  đưa ra các mô tả như đối xứng, đỉnh, đuôi nặng, một mốt, hai mốt, v.v... Đồ thị thích hợp của  $f(x)$  nằm trên biểu đồ phân bố cung cấp đánh giá bằng mắt về sự phù hợp giữa phân bố thích hợp và dữ liệu.

CHÚ THÍCH 3: Chữ viết tắt thông dụng của hàm mật độ xác suất là pdf.

### 2.27

#### Mode của hàm mật độ xác suất

(Các) giá trị tại đó hàm mật độ xác suất (2.26) đạt cực đại địa phương.

CHÚ THÍCH: Phân bố liên tục (2.23) là một mode nếu hàm mật độ xác suất của nó chỉ có một mode, hai mode nếu hàm mật độ xác suất có hai mode và nhiều mode nếu hàm mật độ xác suất có nhiều hơn hai mode.

CHÚ THÍCH 2: Phân bố tại đó các mode tạo thành một tập hợp liên tục cũng được gọi là một mode.

### 2.28

#### Biến ngẫu nhiên rời rạc

Biến ngẫu nhiên (2.10) có phân bố rời rạc (2.22).

CHÚ THÍCH: Biến ngẫu nhiên rời rạc được xem xét trong tiêu chuẩn này bao gồm biến ngẫu nhiên phân bố nhị thức (2.46), Poisson (2.47), siêu hình học (2.48) và đa thức (2.45).

### 2.29

#### Biến ngẫu nhiên liên tục

However, the conditional distribution given that the battery is initially functioning has  $f(x) = \exp(-x)$  for  $x > 0$  as its probability density function, which corresponds to the exponential distribution.

NOTE 1 If the distribution function  $F$  is continuously differentiable, then the probability density function is

$$f(x) = dF(x)/dx$$

at the points  $x$  where the derivative exists.

NOTE 2 A graphical plot of  $f(x)$  versus  $x$  suggests descriptions such as symmetric, peaked, heavy-tailed, unimodal, bi-modal and so forth. A plot of a fitted  $f(x)$  overlaid on a histogram provides a visual assessment of the agreement between a fitted distribution and the data.

NOTE 3 A common abbreviation of probability density function is pdf.

### 2.27

#### mode of probability density function

value(s) where a probability density function (2.26) attains a local maximum

NOTE 1 A continuous distribution (2.23) is unimodal if its probability density function has exactly one mode, bimodal if its probability density function has exactly two modes and multi-modal if its probability density function has more than two modes.

NOTE 2 A distribution where the modes constitute a connected set is also said to be unimodal.

### 2.28

#### discrete random variable

random variable (2.10) having a discrete distribution (2.22)

NOTE Discrete random variables considered in this part of ISO 3534 include the binomial (2.46), Poisson (2.47), hypergeometric (2.48) and multinomial (2.45) random variables.

### 2.29

#### continuous random variable

**Biến ngẫu nhiên (2.10) có phân bố liên tục (2.23).**

**CHÚ THÍCH:** Biến ngẫu nhiên liên tục được đề cập trong tiêu chuẩn này bao gồm biến ngẫu nhiên phân bố chuẩn (2.50), phân bố chuẩn chuẩn hóa (2.51), phân bố  $t$  (2.53), phân bố  $F$  (2.55), gamma (2.56), khi bình phương (2.57), mũ (2.58), beta (2.59), phân bố đều (2.60), cực trị loại I (2.61), cực trị loại II (2.62), cực trị loại III (2.63), loga chuẩn (2.52), chuẩn nhiều chiều (2.64) và chuẩn hai chiều (2.65).

### 2.30

**Phân bố xác suất quy tâm**

**Phân bố xác suất (2.11) của biến ngẫu nhiên quy tâm (2.31).**

### 2.31

**Biến ngẫu nhiên quy tâm**

**Biến ngẫu nhiên (2.10) có được bằng cách lấy một biến ngẫu nhiên trừ đi giá trị trung bình (2.35) của nó.**

**CHÚ THÍCH 1:** Biến ngẫu nhiên quy tâm có trung bình bằng không.

**CHÚ THÍCH 2:** Thuật ngữ này chỉ áp dụng với các biến ngẫu nhiên có trung bình. Ví dụ, không có trung bình của phân bố  $t$  (2.53) một bậc tự do.

**CHÚ THÍCH 3:** Nếu biến ngẫu nhiên  $X$  có trung bình (2.35) bằng  $\mu$  thì biến ngẫu nhiên tương ứng sẽ là  $X - \mu$ , có trung bình bằng không.

### 2.32

**Phân bố xác suất chuẩn hóa**

**Phân bố xác suất (2.11) của biến ngẫu nhiên chuẩn hóa (2.33).**

### 2.33

**Biến ngẫu nhiên chuẩn hóa**

**Biến ngẫu nhiên quy tâm (2.31) có độ lệch chuẩn (2.37) bằng 1.**

**CHÚ THÍCH 1:** Biến ngẫu nhiên (2.10) được tự

**random variable (2.10) having a continuous distribution (2.23)**

**NOTE** Continuous random variables considered in this part of ISO 3534 include the normal (2.50), standardized normal (2.51),  $t$  distribution (2.53),  $F$  distribution (2.55), gamma (2.56), chi-squared (2.57), exponential (2.58), beta (2.59), uniform (2.60), Type I extreme value (2.61), Type II extreme value (2.62), Type III extreme value (2.63), lognormal (2.52), multivariate normal (2.64) and bivariate normal (2.65).

### 2.30

**centred probability distribution**

**probability distribution (2.11) of a centred random variable (2.31)**

### 2.31

**centred random variable**

**random variable (2.10) with its mean (2.35) subtracted**

**NOTE 1** A centred random variable has mean equal to zero.

**NOTE 2** This term only applies to random variables with a mean. For example, the mean of the  $t$  distribution (2.53) with one degree of freedom does not exist.

**NOTE 3** If a random variable  $X$  has a mean (2.35) equal to  $\mu$ , the corresponding centred random variable is  $X - \mu$ , having mean equal to zero.

### 2.32

**standardized probability distribution**

**probability distribution (2.11) of a standardized random variable (2.33)**

### 2.33

**standardized random variable**

**centred random variable (2.31) whose standard deviation (2.37) is equal to 1**

**NOTE 1** A random variable (2.10) is automatically

động chuẩn hóa nếu trung bình của nó bằng không và độ lệch chuẩn bằng 1. Phân bố đều (2.60) trong khoảng  $(-3^{0.5}, 3^{0.5})$  có trung bình là không và độ lệch chuẩn bằng 1. Phân bố chuẩn chuẩn hóa (2.51) đương nhiên là được chuẩn hóa.

CHÚ THÍCH 2: Nếu phân bố (2.11) của biến ngẫu nhiên  $X$  có trung bình (2.35)  $\mu$  và độ lệch chuẩn  $\sigma$ , thì biến ngẫu nhiên chuẩn hóa tương ứng sẽ là  $(X - \mu)/\sigma$ .

2.34

Momen bậc  $r$

Momen thứ  $r$

Kỳ vọng (2.12) của lũy thừa  $r$  của một biến ngẫu nhiên (2.10).

VÍ DỤ: Xét một biến ngẫu nhiên có hàm mật độ xác suất (2.26)  $f(x) = \exp(-x)$  đối với  $x > 0$ . Sử dụng công thức tích phân từng phần, có thể chứng tỏ rằng  $E(X) = 1$ ,  $E(X^2) = 2$ ,  $E(X^3) = 6$  và  $E(X^4) = 24$ , hoặc nói chung,  $E(X^r) = r!$  Đây là ví dụ của phân bố mũ (2.58).

CHÚ THÍCH 1: Trong trường hợp rời rạc một chiều, công thức tương ứng:

$$E(X^r) = \sum_{i=1}^n x_i^r p(x_i)$$

đối với số hữu hạn  $n$  kết quả và

$$E(X^r) = \sum_{i=1}^{\infty} x_i^r p(x_i)$$

đối với số kết quả vô hạn đếm được. Trong trường hợp liên tục một chiều, công thức tương ứng:

$$E(X^r) = \int_{-\infty}^{\infty} x^r f(x) dx$$

CHÚ THÍCH 2: Nếu biến ngẫu nhiên có  $k$  chiều thì lũy thừa  $r$  được hiểu là áp dụng cho thành phần.

CHÚ THÍCH 3: Momen được nêu ở đây sử dụng biến ngẫu nhiên  $X$  nâng lên một lũy thừa. Một cách tổng quát hơn, ta có thể xét momen bậc  $r$  của  $X - \mu$  hoặc  $(X - \mu)/\sigma$ .

standardized if its mean is zero and its standard deviation is 1. The uniform distribution (2.60) on the interval  $(-3^{0.5}, 3^{0.5})$  has mean zero and standard deviation equal to 1. The standardized normal distribution (2.51) is, of course, standardized.

NOTE 2 If the distribution (2.11) of the random variable  $X$  has mean (2.35)  $\mu$  and standard deviation  $\sigma$ , then the corresponding standardized random variable is  $(X - \mu)/\sigma$ .

2.34

moment of order  $r$

$r$ th moment

expectation (2.12) of the  $r^{\text{th}}$  power of a random variable (2.10)

EXAMPLE Consider a random variable having probability density function (2.26)  $f(x) = \exp(-x)$  for  $x > 0$ . Using integration by parts from elementary calculus, it can be shown that  $E(X) = 1$ ,  $E(X^2) = 2$ ,  $E(X^3) = 6$ , and  $E(X^4) = 24$ , or in general,  $E(X^r) = r!$  This is an example of the exponential distribution (2.58).

NOTE 1 In the univariate discrete case, the appropriate formula is:

$$E(X^r) = \sum_{i=1}^n x_i^r p(x_i)$$

for a finite number  $n$  of outcomes and

$$E(X^r) = \sum_{i=1}^{\infty} x_i^r p(x_i)$$

for a countably infinite number of outcomes. In the univariate continuous case, the appropriate formula is:

$$E(X^r) = \int_{-\infty}^{\infty} x^r f(x) dx$$

NOTE 2 If the random variable has dimension  $k$ , then the  $r^{\text{th}}$  power is understood to be applied componentwise.

NOTE 3 The moments given here use a random variable  $X$  raised to a power. More generally,

## 2.35 Trung bình

### 2.35.1

#### Trung bình

##### Mômen bậc $r = 1$

$\mu$

Mômen <phân bố liên tục> bậc  $r$  trong đó  $r$  bằng 1, tính như tích phân của tích giữa  $x$  và hàm mật độ xác suất (2.26),  $f(x)$ , lấy trên đường thẳng thực.

VÍ DỤ 1: Xét biến ngẫu nhiên liên tục (2.29)  $X$  có hàm mật độ xác suất  $f(x) = 6x(1-x)$ , trong đó  $0 \leq x \leq 1$ .

Trung bình của  $X$  là:

$$\int_0^1 6x^2(1-x)dx = 0,5$$

VÍ DỤ 2: Tiếp theo ví dụ pin từ 2.1 và 2.7, trung bình là 0,9 vì với xác suất 0,1 trung bình của phần rời rạc của phân bố là 0 và với xác suất 0,9 trung bình của phần liên tục của phân bố là 1. Phân bố này là phân bố hỗn hợp liên tục và rời rạc.

CHÚ THÍCH 1: Trung bình của phân bố liên tục (2.23) được biểu thị bằng  $E(X)$  và được tính là:

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

CHÚ THÍCH 2: Trung bình không tồn tại đối với tất cả các biến ngẫu nhiên (2.10). Ví dụ, nếu  $X$  được xác định bằng hàm mật độ xác suất  $f(x) = [\pi(1+x^2)]^{-1}$ , tích phân tương ứng với  $E(X)$  là phân kỳ.

### 2.35.2

#### Trung bình

$\mu$

tổng <phân bố rời rạc> tích của  $x_i$  và hàm khối lượng xác suất (2.24)  $p(x_i)$ .

one could consider moments of order  $r$  of  $X - \mu$  or  $(X - \mu)/\sigma$ .

## 2.35 Means

### 2.35.1

#### mean

##### moment of order $r = 1$

$\mu$

<continuous distribution> moment of order  $r$  where  $r$  equals 1, calculated as the integral of the product of  $x$  and the probability density function (2.26),  $f(x)$ , over the real line

EXAMPLE 1 Consider a continuous random variable (2.29)  $X$  having probability density function  $f(x) = 6x(1-x)$ , where  $0 \leq x \leq 1$ . The mean of  $X$  is:

$$\int_0^1 6x^2(1-x)dx = 0,5$$

EXAMPLE 2 Continuing with the battery example from 2.1 and 2.7, the mean is 0,9 since with probability 0,1 the mean of the discrete part of the distribution is 0 and with probability 0,9 the mean of the continuous part of the distribution is 1. This distribution is a mixture of continuous and discrete distributions.

NOTE 1 The mean of a continuous distribution (2.23) is denoted by  $E(X)$  and is computed as:

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

NOTE 2 The mean does not exist for all random variables (2.10). For example, if  $X$  is defined by its probability density function  $f(x) = [\pi(1+x^2)]^{-1}$ , the integral corresponding to  $E(X)$  is divergent.

### 2.35.2

#### mean

$\mu$

<discrete distribution> summation of the product of  $x_i$  and the probability mass function (2.24)  $p(x_i)$

EXAMPLE 1 Consider a discrete random variable

VÍ DỤ 1: Xét biến ngẫu nhiên rời rạc  $X$  (2.28) đại diện cho số mặt ngửa xuất hiện khi tung ba đồng xu. Hàm khối lượng xác suất là

$$P(X=0) = 1/8$$

$$P(X=1) = 3/8$$

$$P(X=2) = 3/8$$

$$P(X=3) = 1/8$$

Vì vậy, trung bình của  $X$  là

$$0(1/8) + 1(3/8) + 2(3/8) + 3(1/8) = 12/8 = 1,5$$

VÍ DỤ 2: Xem ví dụ 2 trong 2.35.1.

CHÚ THÍCH: Trung bình của phân bố rời rạc (2.22) biểu thị bằng  $E(X)$  và được tính là:

$$E(X) = \sum_{i=1}^n x_i p(x_i)$$

đối với số lượng hữu hạn các kết quả, và

$$E(X) = \sum_{i=1}^{\infty} x_i p(x_i)$$

đối với số lượng kết quả vô hạn đếm được.

### 2.36

#### Phương sai

$V$

Môment bậc  $r$  (2.34) trong đó  $r$  bằng 2 theo phân bố xác suất quy tâm (2.30) của biến ngẫu nhiên (2.10).

VÍ DỤ 1: Đối với biến ngẫu nhiên rời rạc (2.28) trong ví dụ của 2.24 phương sai là

$$\sum_{i=0}^3 (x_i - 1,5)^2 P(X = x_i) = 0,75$$

VÍ DỤ 2: Đối với biến ngẫu nhiên liên tục (2.29) trong ví dụ 1 của 2.26, phương sai là

$$\int_0^1 (x_i - 0,5)^2 6x(1-x) dx = 0,05$$

VÍ DỤ 3: Đối với ví dụ pin ở 2.1, có thể xác định phương sai bằng cách nhận thức rằng phương sai của  $X$  là  $E(X^2) - [E(X)]^2$ . Từ ví dụ 3 của 2.35,  $E(X) =$

$X$  (2.28) representing the number of heads resulting from the tossing of three fair coins. The probability mass function is

$$P(X=0) = 1/8$$

$$P(X=1) = 3/8$$

$$P(X=2) = 3/8$$

$$P(X=3) = 1/8$$

Hence, the mean of  $X$  is

$$0(1/8) + 1(3/8) + 2(3/8) + 3(1/8) = 12/8 = 1,5$$

EXAMPLE 2 See Example 2 in 2.35.1.

NOTE The mean of a discrete distribution (2.22) is denoted by  $E(X)$  and is computed as:

$$E(X) = \sum_{i=1}^n x_i p(x_i)$$

for a finite number of outcomes, and

$$E(X) = \sum_{i=1}^{\infty} x_i p(x_i)$$

for a countably infinite number of outcomes.

### 2.36

#### variance

$V$

moment of order  $r$  (2.34) where  $r$  equals 2 in the centred probability distribution (2.30) of the random variable (2.10)

EXAMPLE 1 For the discrete random variable (2.28) in the example of 2.24 the variance is

$$\sum_{i=0}^3 (x_i - 1,5)^2 P(X = x_i) = 0,75$$

EXAMPLE 2 For the continuous random variable (2.29) in Example 1 of 2.26, the variance is

$$\int_0^1 (x_i - 0,5)^2 6x(1-x) dx = 0,05$$

EXAMPLE 3 For the battery example of 2.1, the variance can be determined by recognizing that the variance of  $X$  is  $E(X^2) - [E(X)]^2$ . From Example 3 of

0,9. Lập luận tương tự, có thể chứng tỏ  $E(X^2)$  bằng 1,8. Do đó, phương sai của  $X$  là  $1,8 - (0,9)^2$  bằng 0,99.

**CHÚ THÍCH:** Phương sai có thể xác định tương đương như kỳ vọng (2.12) của bình phương biến ngẫu nhiên trừ đi trung bình (2.35) của nó. Phương sai của biến ngẫu nhiên  $X$  được ký hiệu bằng  $V(X) = E\{(X-E(X))^2\}$ .

### 2.37

#### Độ lệch chuẩn

$\sigma$

Căn bậc hai dương của phương sai (2.36)

VÍ DỤ: Đối với ví dụ về pin của 2.1 và 2.7, độ lệch chuẩn là 0,995.

### 2.38

#### Hệ số biến động

CV

<biến ngẫu nhiên dương> độ lệch chuẩn (2.37) chia cho trung bình (2.35).

VÍ DỤ: Đối với ví dụ về pin của 2.1 và 2.7, hệ số biến động là  $0,99/0,995$  bằng  $0,994\ 97$ .

**CHÚ THÍCH 1:** Hệ số biến động thường được tính bằng phần trăm.

**CHÚ THÍCH 2:** Thuật ngữ "độ lệch chuẩn tương đối" trước đây không giống với thuật ngữ hệ số biến động.

### 2.39

#### Hệ số bất đối xứng

$\gamma_1$

Mômen bậc 3 (2.34) trong phân bố xác suất chuẩn hóa (2.32) của một biến ngẫu nhiên (2.10).

VÍ DỤ: Tiếp theo ví dụ pin của 2.1 và 2.7 có phân bố hỗn hợp liên tục-rời rạc, sử dụng các kết quả từ ví dụ ở 2.34, ta có

$$E(X) = 0,1(0) + 0,9(1) = 0,9$$

$$E(X^2) = 0,1(0^2) + 0,9(2) = 1,8$$

2.35,  $E(X) = 0,9$ . Using the same type of conditioning argument,  $E(X^2)$  can be shown to be 1,8. Thus, the variance of  $X$  is  $1,8 - (0,9)^2$  which equals 0,99.

**NOTE** The variance can equivalently be defined as the expectation (2.12) of the square of the random variable minus its mean (2.35). The variance of a random variable  $X$  is denoted by  $V(X) = E\{(X-E(X))^2\}$ .

### 2.37

#### standard deviation

$\sigma$

positive square root of the variance (2.36)

**EXAMPLE** For the battery example of 2.1 and 2.7, the standard deviation is 0,995.

### 2.38

#### coefficient of variation

CV

<positive random variable> standard deviation (2.37) divided by the mean (2.35).

**EXAMPLE** For the battery example of 2.1 and 2.7, the coefficient of variation is  $0,99/0,995$  which equals  $0,994\ 97$ .

**NOTE 1** The coefficient of variation is commonly reported as a percentage.

**NOTE 2** The predecessor term "relative standard deviation" is deprecated by the term coefficient of variation.

### 2.39

#### coefficient of skewness

$\gamma_1$

moment of order 3 (2.34) in the standardized probability distribution (2.32) of a random variable (2.10)

**EXAMPLE** Continuing with the battery example of 2.1 and 2.7 having a mixed continuous-discrete distribution, one has, using results from the example in 2.34,

$$E(X^3) = 0,1(0) + 0,9(6) = 5,4$$

$$E(X^4) = 0,1(0) + 0,9(24) = 21,6$$

Để tính hệ số bất đối xứng, lưu ý là  $E\{[X - E(X)]^3\} = E(X^3) - 3 E(X) E(X^2) + 2 [E(X)]^3$  và từ 2.37 có độ lệch chuẩn là 0,995. Do đó, hệ số bất đối xứng là  $[5,4 - 3(0,9)(1,8) + 2(0,9)^3]/(0,995)^3$  hoặc 1,998.

CHÚ THÍCH 1: Một định nghĩa tương đương dựa trên kỳ vọng (2.12) bậc ba của  $(X-\mu)/\sigma$ , là  $E[(X-\mu)^3/\sigma^3]$ .

CHÚ THÍCH 2: Hệ số bất đối xứng là thước đo tính đối xứng của phân bố (2.11) và đôi khi được ký hiệu bằng  $\sqrt{\beta_1}$ . Đối với các phân bố đối xứng, hệ số bất đối xứng bằng 0 (với điều kiện tồn tại mômen thích hợp trong định nghĩa). Ví dụ về phân bố có độ bất đối xứng bằng không bao gồm phân bố chuẩn (2.50), phân bố beta (2.59) với điều kiện  $\alpha = \beta$  và phân bố  $t$  (2.53) với điều kiện có các mômen.

2.40

Hệ số nhọn

$$\beta_2$$

Mômen bậc 4 (2.34) trong phân bố xác suất chuẩn hóa (2.32) của một biến ngẫu nhiên (2.10).

VÍ DỤ: Tiếp theo ví dụ về pin của 2.1 và 2.7, để tính hệ số nhọn, lưu ý rằng

$$E\{[X-E(X)]^4\} = E(X^4) - 4 E(X) E(X^3) + 6[E(X)]^2 E(X^2) - 3 [E(X)]^4$$

Do đó, hệ số nhọn bằng

$$[21,6 - 4(0,9)(5,4) + 6(0,9)^2(2) - 3(0,9)^4] / (0,995)^4$$

hoặc 8,94.

CHÚ THÍCH 1: Một định nghĩa tương đương dựa trên

$$E(X) = 0,1(0) + 0,9(1) = 0,9$$

$$E(X^2) = 0,1(0^2) + 0,9(2) = 1,8$$

$$E(X^3) = 0,1(0) + 0,9(6) = 5,4$$

$$E(X^4) = 0,1(0) + 0,9(24) = 21,6$$

To compute the coefficient of skewness, note that  $E\{[X - E(X)]^3\} = E(X^3) - 3 E(X) E(X^2) + 2 [E(X)]^3$  and from 2.37 the standard deviation is 0,995. The coefficient of skewness is thus  $[5,4 - 3(0,9)(1,8) + 2(0,9)^3]/(0,995)^3$  or 1,998.

NOTE 1 An equivalent definition is based on the expectation (2.12) of the third power of  $(X-\mu)/\sigma$ , namely  $E[(X-\mu)^3/\sigma^3]$ .

NOTE 2 The coefficient of skewness is a measure of the symmetry of a distribution (2.11) and is sometimes denoted by  $\sqrt{\beta_1}$ . For symmetric distributions, the coefficient of skewness is equal to 0 (provided the appropriate moments in the definition exist). Examples of distributions with skewness equal to zero include the normal distribution (2.50), the beta distribution (2.59) provided  $\alpha = \beta$  and the  $t$  distribution (2.53) provided the moments exist.

2.40

coefficient of kurtosis

$$\beta_2$$

moment of order 4 (2.34) in the standardized probability distribution (2.32) of a random variable (2.10)

EXAMPLE Continuing with the battery example of 2.1 and 2.7, to compute the coefficient of kurtosis, note that

$$E\{[X-E(X)]^4\} = E(X^4) - 4 E(X) E(X^3) + 6[E(X)]^2 E(X^2) - 3 [E(X)]^4$$

The coefficient of kurtosis is thus

$$[21,6 - 4(0,9)(5,4) + 6(0,9)^2(2) - 3(0,9)^4] / (0,995)^4$$

or 8,94.



kỳ vọng (2.12) bậc bốn của  $(X - \mu)/\sigma$ , gọi là  $E[(X - \mu)^4/\sigma^4]$ .

CHÚ THÍCH 2: Hệ số nhọn là thước đo độ lớn phần đuôi của phân bố (2.11). Đối với phân bố đều (2.60), hệ số nhọn là 1,8; đối với phân bố chuẩn (2.50), hệ số nhọn là 3; đối với phân bố mũ (2.58), hệ số nhọn là 9.

CHÚ THÍCH 3: Cần chú ý khi xem xét các giá trị nhọn được công bố, vì một số người thực hiện trừ đi 3 (hệ số nhọn của phân bố chuẩn) từ giá trị được tính theo định nghĩa.

#### 2.41

**Mômen hỗn hợp bậc  $r$  và  $s$**

Trung bình (2.35) của tích lũy thừa bậc  $r$  của một biến ngẫu nhiên (2.10) và bậc  $s$  của một biến ngẫu nhiên khác lấy theo phân bố xác suất (2.11) đồng thời của chúng.

#### 2.42

**Mômen hỗn hợp quy tâm bậc  $r$  và  $s$**

Trung bình (2.35) tích lũy thừa bậc  $r$  của biến ngẫu nhiên quy tâm (2.31) và bậc  $s$  của một biến ngẫu nhiên quy tâm khác trong phân bố xác suất (2.11) đồng thời của chúng.

#### 2.43

**Hiệp phương sai**

$\sigma_{XY}$

Trung bình (2.35) của tích hai biến ngẫu nhiên quy tâm (2.31) lấy theo phân bố xác suất (2.11) đồng thời của chúng.

CHÚ THÍCH 1: Hiệp phương sai là mômen trung tâm hỗn hợp bậc 1 và 1 (2.42) đối với hai biến ngẫu nhiên.

CHÚ THÍCH 2: Về ký hiệu, hiệp phương sai là

$$E[(X - \mu_X)(Y - \mu_Y)],$$

NOTE 1 An equivalent definition is based on the expectation (2.12) of the fourth power of  $(X - \mu)/\sigma$ , namely  $E[(X - \mu)^4/\sigma^4]$ .

NOTE 2 The coefficient of kurtosis is a measure of the heaviness of the tails of a distribution (2.11). For the uniform distribution (2.60), the coefficient of kurtosis is 1,8; for the normal distribution (2.50), the coefficient of kurtosis is 3; for the exponential distribution (2.58), the coefficient of kurtosis is 9.

NOTE 3 Caution needs to be exercised in considering reported kurtosis values, as some practitioners subtract 3 (the kurtosis of the normal distribution) from the value that is computed from the definition.

#### 2.41

**joint moment of orders  $r$  and  $s$**

mean (2.35) of the product of the  $r^{\text{th}}$  power of a random variable (2.10) and the  $s^{\text{th}}$  power of another random variable in their joint probability distribution (2.11)

#### 2.42

**joint central moment of orders  $r$  and  $s$**

mean (2.35) of the product of the  $r^{\text{th}}$  power of a centred random variable (2.31) and the  $s^{\text{th}}$  power of another centred random variable in their joint probability distribution (2.11)

#### 2.43

**covariance**

$\sigma_{XY}$

mean (2.35) of the product of two centred random variables (2.31) in their joint probability distribution (2.11)

NOTE 1 The covariance is the joint central moment of orders 1 and 1 (2.42) for two random variables.

NOTE 2 Notationally, the covariance is

$$E[(X - \mu_X)(Y - \mu_Y)],$$

trong đó  $E(X) = \mu_x$  and  $E(Y) = \mu_y$ .

2.44

**Hệ số tương quan**

Trung bình (2.35) của tích hai biến ngẫu nhiên chuẩn hóa (2.33) lấy theo phân bố xác suất (2.11) đồng thời của chúng.

CHÚ THÍCH: Hệ số tương quan đôi khi được gọi tắt đơn giản là tương quan. Tuy nhiên, việc sử dụng này lại chồng chéo với cách giải thích về tương quan như sự liên kết giữa hai biến.

2.45

**Phân bố đa thức**

Phân bố rời rạc (2.22) có hàm khối lượng xác suất (2.24)

$$P(X_1 = x_1, X_2 = x_2, \dots, X_k = x_k) = \frac{n!}{x_1! x_2! \dots x_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}$$

trong đó

$x_1, x_2, \dots, x_k$  là các số nguyên dương sao cho  $x_1 + x_2 + \dots + x_k = n$  với các tham số  $p_i > 0$  đối với tất cả  $i = 1, 2, \dots, k$  với  $p_1 + p_2 + \dots + p_k = 1$   
 $k$  số nguyên lớn hơn hoặc bằng 2

CHÚ THÍCH: Phân bố đa thức cho biết xác suất để trong  $n$  phép thử độc lập có được một tổ hợp cho trước về số lần xuất hiện các biến cố trong đó  $k$  kiểu biến cố loại trừ nhau, giả định rằng xác suất xuất hiện của mỗi loại biến cố là như nhau trong tất cả các phép thử.

2.46

**Phân bố nhị thức**

Phân bố rời rạc (2.22) có hàm khối lượng xác suất (2.24)

$$P(X = x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

where  $E(X) = \mu_x$  and  $E(Y) = \mu_y$ .

2.44

**correlation coefficient**

mean (2.35) of the product of two standardized random variables (2.33) in their joint probability distribution (2.11)

NOTE Correlation coefficient is sometimes more briefly referred to as simply correlation. However, this usage overlaps with interpretations of correlation as an association between two variables.

2.45

**multinomial distribution**

discrete distribution (2.22) having the probability mass function (2.24)

$$P(X_1 = x_1, X_2 = x_2, \dots, X_k = x_k) = \frac{n!}{x_1! x_2! \dots x_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}$$

where

$x_1, x_2, \dots, x_k$  are non-negative integers such that  $x_1 + x_2 + \dots + x_k = n$  with parameters  $p_i > 0$  for all  $i = 1, 2, \dots, k$  with  $p_1 + p_2 + \dots + p_k = 1$   
 $k$  an integer greater than or equal to 2

NOTE The multinomial distribution gives the probability of the number of times each of  $k$  possible outcomes have occurred in  $n$  independent trials where each trial has the same  $k$  mutually exclusive events and the probabilities of the events are the same for all trials.

2.46

**binomial distribution**

discrete distribution (2.22) having the probability mass function (2.24)

$$P(X = x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

trong đó  $x = 0, 1, 2, \dots, n$  và với các tham số  $n = 1, 2, \dots$ , và  $0 < p < 1$ .

VÍ DỤ: Hàm khối lượng xác suất mô tả trong ví dụ 1 của 2.24 có thể thấy là tương ứng với phân bố nhị thức với các tham số xác định  $n = 3$  và  $p = 0,5$ .

CHÚ THÍCH 1: Phân bố nhị thức là một trường hợp đặc biệt của phân bố đa thức (2.45) với  $k = 2$ .

CHÚ THÍCH 2: Phân bố nhị thức cho xác suất về số lần một trong số hai kết quả xảy ra trong  $n$  phép thử độc lập, trong đó mỗi phép thử có cùng hai biến cố (2.2) loại trừ nhau và xác suất (2.5) của mỗi biến cố là như nhau đối với các phép thử.

CHÚ THÍCH 3: Trung bình (2.35) của phân bố nhị thức bằng  $np$ . Phương sai (2.36) của phân bố nhị thức bằng  $np(1 - p)$ .

CHÚ THÍCH 4: Hàm khối lượng xác suất nhị thức cũng có thể biểu thị bằng cách sử dụng hệ số nhị thức cho bởi

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

## 2.47

### Phân bố Poisson

Phân bố rời rạc (2.22) có hàm khối lượng xác suất (2.24)

$$P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

trong đó  $x = 0, 1, 2, \dots$  và tham số  $\lambda > 0$ .

CHÚ THÍCH 1: Giới hạn của phân bố nhị thức (2.46) khi  $n$  tiến đến  $\infty$  và  $p$  tiến đến không sao  $np$  tiến đến  $\lambda$  là phân bố Poisson với tham số  $\lambda$ .

CHÚ THÍCH 2: Trung bình (2.35) và phương sai (2.36) của phân bố Poisson đều bằng  $\lambda$ .

CHÚ THÍCH 3: Hàm khối lượng xác suất (2.24) của phân bố Poisson cho xác suất về số lần xuất hiện một tính chất của một quá trình trong một khoảng thời gian

where  $x = 0, 1, 2, \dots, n$  and with indexing parameters  $n = 1, 2, \dots$ , and  $0 < p < 1$ .

EXAMPLE The probability mass function described in Example 1 of 2.24 can be seen to correspond to the binomial distribution with index parameters  $n = 3$  and  $p = 0,5$ .

NOTE 1 The binomial distribution is a special case of the multinomial distribution (2.45) with  $k = 2$ .

NOTE 2 The binomial distribution gives the probability of the number of times each of two possible outcomes have occurred in  $n$  independent trials where each trial has the same two mutually exclusive events (2.2) and the probabilities (2.5) of the events are the same for all trials.

NOTE 3 The mean (2.35) of the binomial distribution equals  $np$ . The variance (2.36) of the binomial distribution equals  $np(1 - p)$ .

NOTE 4 The binomial probability mass function may be alternately expressed using the binomial coefficient given by

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

## 2.47

### Poisson distribution

discrete distribution (2.22) having the probability mass function (2.24)

$$P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

where  $x = 0, 1, 2, \dots$  and with parameter  $\lambda > 0$ .

NOTE 1 The limit of the binomial distribution (2.46) as  $n$  approaches  $\infty$  and  $p$  tends to zero in such a way that  $np$  tends to  $\lambda$  is the Poisson distribution with parameter  $\lambda$ .

NOTE 2 The mean (2.35) and the variance (2.36) of the Poisson distribution are both equal to  $\lambda$ .

NOTE 3 The probability mass function (2.24) of the Poisson distribution gives the probability for the

bằng độ dài đơn vị đáp ứng các điều kiện nhất định, vì dự cường độ xuất hiện phụ thuộc vào thời gian.

2.48

**Phân bố siêu hình học**

Phân bố rời rạc (2.22) có hàm khối lượng xác suất (2.24)

$$P(X = x) = \frac{\binom{M}{x} \binom{M-N}{n-x}}{\binom{M-N+n}{n}}$$

trong đó  $\max(0, M - N) \leq x \leq \min(M, n)$  có tham số là số nguyên

$$N = 1, 2, \dots$$

$$M = 0, 1, 2, \dots, N - 1$$

$$n = 1, 2, \dots, N$$

CHÚ THÍCH 1: Phân bố (2.11) siêu hình học xuất hiện khi số cá thể được đánh dấu trong một mẫu ngẫu nhiên đơn giản (1.7) cỡ  $n$ , được lấy mà không hoàn lại, từ một tổng thể (hoặc lô) cỡ  $N$  chứa chính xác  $M$  cá thể đánh dấu.

CHÚ THÍCH 2: Việc hiểu về phân bố siêu hình học sẽ dễ dàng hơn khi tham khảo Bảng 4.

**Bảng 4 – Ví dụ về phân bố siêu hình học**

Tập hợp tham chiếu	Cá thể đánh dấu hoặc không đánh dấu	Cá thể đánh dấu	Cá thể không đánh dấu
Tổng thể	$N$	$M$	$N - M$
Các cá thể trong mẫu	$n$	$x$	$N - x$
Các cá thể ngoài mẫu	$N - n$	$M - x$	$N - n - M + x$

number of occurrences of a property of a process in a time interval of unit length satisfying certain conditions, e.g. intensity of occurrence independent of time.

2.48

**hypergeometric distribution**

discrete distribution (2.22) having the probability mass function (2.24)

$$P(X = x) = \frac{\binom{M}{x} \binom{M-N}{n-x}}{\binom{M-N+n}{n}}$$

where  $\max(0, M - N) \leq x \leq \min(M, n)$  with integer parameters

$$N = 1, 2, \dots$$

$$M = 0, 1, 2, \dots, N - 1$$

$$n = 1, 2, \dots, N$$

NOTE 1 The hypergeometric distribution (2.11) arises as the number of marked items in a simple random sample (1.7) of size  $n$ , taken without replacement, from a population (or lot) of size  $N$  containing exactly  $M$  marked items.

NOTE 2 An understanding of the hypergeometric distribution may be facilitated with Table 4.

**Table 4 – Hypergeometric distribution example**

Reference set	Marked or unmarked items	Marked items	Unmarked items
Population	$N$	$M$	$N - M$
Items in sample	$n$	$x$	$N - x$
Items not in sample	$N - n$	$M - x$	$N - n - M + x$

CHÚ THÍCH 3: Trong những điều kiện nhất định (ví dụ  $n$  tương đối nhỏ so với  $N$ ), thì phân bố siêu hình học có thể xấp xỉ bằng phân bố nhị thức  $n$  và  $p = M/N$ .

CHÚ THÍCH 4: Trung bình (2.35) của phân bố siêu hình học bằng  $(nM)/N$ . Phương sai (2.36) của phân bố siêu hình học bằng

$$n \frac{M}{N} \left(1 - \frac{M}{N}\right) \frac{N-n}{N-1}$$

## 2.49

### Phân bố nhị thức âm

Phân bố rời rạc (2.22) có hàm khối lượng xác suất (2.24)

$$P(X = x) = \frac{(c+x-1)!}{x!(c-1)!} p^c (1-p)^x$$

trong đó  $x = 0, 1, 2, \dots, n$ , với tham số  $c > 0$  và tham số  $p$  thỏa mãn  $0 < p < 1$ .

CHÚ THÍCH 1: Nếu  $c = 1$ , phân bố nhị thức âm được coi là phân bố hình học và mô tả xác suất (2.5) biến cố (2.2) có xác suất là  $p$ , sẽ xuất hiện lần đầu tiên trong phép thử thứ  $(x+1)$ .

CHÚ THÍCH 2: Hàm khối lượng xác suất cũng có thể viết theo cách tương đương như sau:

$$P(X = x) = \binom{-c}{x} p^c (1-p)^x$$

Thuật ngữ "phân bố nhị thức âm" xuất phát từ cách viết hàm khối lượng xác suất này.

CHÚ THÍCH 3: Dạng hàm khối lượng xác suất nêu trong định nghĩa này thường được gọi là "phân bố Pascal" với điều kiện  $c$  là số nguyên lớn hơn hoặc bằng 1. Trong trường hợp đó, hàm khối lượng xác suất mô tả xác suất lần xuất hiện thứ  $c$  của biến cố (2.2), có xác suất (2.5) là  $p$ , xuất hiện trong lần thử  $(c+x)$ .

CHÚ THÍCH 4: Trung bình (2.35) của phân bố nhị thức âm là  $(cp)/(1-p)$ . Phương sai (2.36) của nhị

NOTE 3 Under certain conditions (for example,  $n$  is small relative to  $N$ ), then the hypergeometric distribution can be approximated by the binomial distribution with  $n$  and  $p = M/N$ .

NOTE 4 The mean (2.35) of the hypergeometric distribution equals  $(nM)/N$ . The variance (2.36) of the hypergeometric distribution equals

$$n \frac{M}{N} \left(1 - \frac{M}{N}\right) \frac{N-n}{N-1}$$

## 2.49

### negative binomial distribution

discrete distribution (2.22) having the probability

mass function (2.24)

$$P(X = x) = \frac{(c+x-1)!}{x!(c-1)!} p^c (1-p)^x$$

where  $x = 0, 1, 2, \dots, n$  with parameter  $c > 0$  and parameter  $p$  satisfying  $0 < p < 1$ .

NOTE 1 If  $c = 1$ , the negative binomial distribution is known as the geometric distribution and describes the probability (2.5) that the first incident of the event (2.2) whose probability is  $p$ , will occur in trial  $(x+1)$ .

NOTE 2 The probability mass function may also be written in the following, equivalent way:

$$P(X = x) = \binom{-c}{x} p^c (1-p)^x$$

The term "negative binomial distribution" emerges from this way of writing the probability mass function.

NOTE 3 The version of the probability mass function given in the definition is often called the "Pascal distribution" provided  $c$  is an integer greater than or equal to 1. In that case, the probability mass function describes the probability that the  $c^{\text{th}}$  incident of the event (2.2), whose probability (2.5) is  $p$ , occurs in trial  $(c+x)$ .

thức âm là  $(cp)/(1 - p)^2$ .

**2.50**

**Phân bố chuẩn**

**Phân bố Gaussian**

**Phân bố liên tục (2.23) có hàm mật độ xác suất (2.26)**

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

trong đó  $-\infty < x < \infty$  và có tham số  $-\infty < \mu < \infty$  và  $\sigma > 0$ .

**CHÚ THÍCH 1:** Phân bố chuẩn là một trong các phân bố xác suất (2.11) được sử dụng rộng rãi nhất trong thống kê ứng dụng. Do dạng của hàm mật độ, nó thường được gọi là đường "hình chuông". Bên cạnh việc sử dụng như một mô hình dùng cho các hiện tượng ngẫu nhiên, phân bố này còn như phân bố trung bình (1.15) giới hạn. Như một phân bố tham chiếu trong thống kê, phân bố này được sử dụng rộng rãi để đánh giá tính bất thường của các kết quả thực nghiệm.

**CHÚ THÍCH 2:** Tham số định vị  $\mu$  là trung bình (2.35) và tham số thang đo  $\sigma$  là độ lệch chuẩn (2.37) của phân bố chuẩn.

**2.51**

**Phân bố chuẩn chuẩn hóa**

**Phân bố Gaussian chuẩn hóa**

**Phân bố chuẩn (2.50) có  $\mu = 0$  và  $\sigma = 1$ .**

**CHÚ THÍCH:** Hàm mật độ xác suất (2.26) của phân bố chuẩn chuẩn hóa là

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

trong đó  $-\infty < x < \infty$ . Các bảng phân bố chuẩn có hàm mật độ xác suất này, đưa ra làm ví dụ, vùng phủ  $f$  đối với các giá trị trong khoảng  $(-\infty, \infty)$ .

**NOTE 4** The mean (2.35) of the negative binomial distribution is  $(cp)/(1 - p)$ . The variance (2.36) of the negative binomial is  $(cp)/(1 - p)^2$ .

**2.50**

**normal distribution**

**Gaussian distribution**

**continuous distribution (2.23) having the probability density function (2.26)**

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

where  $-\infty < x < \infty$  and with parameters  $-\infty < \mu < \infty$  and  $\sigma > 0$ .

**NOTE 1** The normal distribution is one of the most widely used probability distributions (2.11) in applied statistics. Owing to the shape of the density function, it is informally referred to as the "bell-shaped" curve. Aside from serving as a model for random phenomena, it arises as the limiting distribution of averages (1.15). As a reference distribution in statistics, it is widely used to assess the unusualness of experimental outcomes.

**NOTE 2** The location parameter  $\mu$  is the mean (2.35) and the scale parameter  $\sigma$  is the standard deviation (2.37) of the normal distribution.

**2.51**

**standardized normal distribution**

**standardized Gaussian distribution**

**normal distribution (2.50) with  $\mu = 0$  and  $\sigma = 1$**

**NOTE** The probability density function (2.26) of the standardized normal distribution is

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

where  $-\infty < x < \infty$ . Tables of the normal distribution involve this probability density function, giving for example, the area under  $f$  for values in  $(-\infty, \infty)$ .

## 2.52

**Phân bố lôga chuẩn**

**Phân bố liên tục** (2.23) có hàm mật độ xác suất (2.26).

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}$$

trong đó  $x > 0$  và với các tham số  $-\infty < \mu < \infty$  và  $\sigma > 0$ .

**CHÚ THÍCH 1:** Nếu  $Y$  có phân bố chuẩn (2.50) với trung bình (2.35)  $\mu$  và độ lệch chuẩn (2.37)  $\sigma$ , thì phép biến đổi cho bởi  $X = \exp(Y)$  có hàm mật độ xác suất nêu trong định nghĩa. Nếu  $X$  có phân bố lôga chuẩn có hàm mật độ xác suất như nêu trong định nghĩa thì  $\ln(X)$  có phân bố chuẩn với trung bình  $\mu$  và độ lệch chuẩn  $\sigma$ .

**CHÚ THÍCH 2:** Trung bình của phân bố lôga chuẩn là  $\exp[\mu + (\sigma^2)/2]$  và phương sai là  $\exp(2\mu + \sigma^2) \times [\exp(\sigma^2) - 1]$ . Điều này chỉ ra rằng trung bình và phương sai của phân bố lôga chuẩn là hàm số của tham số  $\mu$  và  $\sigma^2$ .

**CHÚ THÍCH 3:** Phân bố lôga chuẩn và phân bố Weibull (2.63) được sử dụng phổ biến trong các ứng dụng liên quan tới độ tin cậy.

## 2.53

**Phân bố  $t$** **Phân bố Student**

**Phân bố liên tục** (2.23) có hàm mật độ xác suất (2.26)

$$f(t) = \frac{\Gamma[(v+1)/2]}{\sqrt{\pi v} \Gamma(v/2)} \times \left(1 + \frac{t^2}{v}\right)^{-(v+1)/2}$$

trong đó  $-\infty < t < \infty$  và với tham số  $v$ , là số nguyên dương.

**CHÚ THÍCH 1:** Phân bố  $t$  được sử dụng rộng rãi trong thực tế để đánh giá trung bình mẫu (1.15) trong trường hợp chung khi độ lệch chuẩn của tổng thể được ước lượng từ dữ liệu. Thống kê mẫu  $t$  có thể so sánh với phân bố  $t$  có  $n - 1$  bậc tự do để đánh giá

## 2.52

**lognormal distribution**

**continuous distribution** (2.23) having the **probability density function** (2.26)

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}$$

where  $x > 0$  and with parameters  $-\infty < \mu < \infty$  and  $\sigma > 0$ .

**NOTE 1** If  $Y$  has a normal distribution (2.50) with mean (2.35)  $\mu$  and standard deviation (2.37)  $\sigma$ , then the transformation given by  $X = \exp(Y)$  has the probability density function given in the definition. If  $X$  has a lognormal distribution with density function as given in the definition, then  $\ln(X)$  has a normal distribution with mean  $\mu$  and standard deviation  $\sigma$ .

**NOTE 2** The mean of the lognormal distribution is  $\exp[\mu + (\sigma^2)/2]$  and the variance is  $\exp(2\mu + \sigma^2) \times [\exp(\sigma^2) - 1]$ . This indicates that the mean and variance of the lognormal distribution are functions of the parameters  $\mu$  and  $\sigma^2$ .

**NOTE 3** The lognormal distribution and Weibull distribution (2.63) are commonly used in reliability applications.

## 2.53

 **$t$  distribution****Student's distribution**

**continuous distribution** (2.23) having the **probability density function** (2.26)

$$f(t) = \frac{\Gamma[(v+1)/2]}{\sqrt{\pi v} \Gamma(v/2)} \times \left(1 + \frac{t^2}{v}\right)^{-(v+1)/2}$$

where  $-\infty < t < \infty$  and with parameter  $v$ , a positive integer.

**NOTE 1** The  $t$  distribution is widely used in practice to evaluate the sample mean (1.15) in the common case where the population standard deviation is estimated from the data. The sample  $t$  statistic can be compared to the  $t$  distribution with  $n$

trung bình quy định như mô tả của trung bình tổng thể thực.

CHÚ THÍCH 2: Phân bố  $t$  phát sinh như phân bố thương số của hai biến ngẫu nhiên (2.10) độc lập, trong đó tử số có phân bố chuẩn chuẩn hóa (2.51) và mẫu số được phân bố theo căn bậc hai dương của phân bố khi bình phương (2.57) sau khi chia cho số bậc tự do của nó. Tham số  $\nu$  được gọi là **bậc tự do** (2.54).

CHÚ THÍCH 3: Hàm gamma được định nghĩa trong 2.56.

**2.54**

**Bậc tự do**

$\nu$

Số các số hạng trong một tổng trừ đi số các ràng buộc đối với các số hạng của tổng đó.

CHÚ THÍCH: Khái niệm này đã gặp trước đây trong ngữ cảnh sử dụng  $n - 1$  trong mẫu số của **hàm ước lượng** (1.12) của **phương sai mẫu** (1.16). Số bậc tự do được dùng để hiệu chỉnh các tham số. Thuật ngữ bậc tự do cũng được sử dụng rộng rãi trong ISO 3534-3 trong đó bình phương trung bình được cho như tổng các bình phương chia cho số bậc tự do thích hợp.

**2.55**

**Phân bố  $F$**

**Phân bố liên tục** (2.23) có **hàm mật độ xác suất** (2.26)

$$f(x) = \frac{\Gamma[(\nu_1 + \nu_2)/2]}{\Gamma(\nu_1/2)\Gamma(\nu_2/2)} (\nu_1)^{\nu_1/2} (\nu_2)^{\nu_2/2} \frac{x^{(\nu_1/2)-1}}{(\nu_1 x + \nu_2)^{(\nu_1+\nu_2)/2}}$$

trong đó

$x > 0$

$\nu_1$  và  $\nu_2$  là số nguyên dương

$\Gamma$  là hàm gamma xác định trong 2.56.

CHÚ THÍCH 1: Phân bố  $F$  là phân bố tham chiếu hữu ích để đánh giá tỉ số của các **phương sai** (2.36) độc lập.

CHÚ THÍCH 2: Hàm  $F$  phát sinh khi phân bố là

$- 1$  degrees of freedom to assess a specified mean as a depiction of the true population mean.

NOTE 2 The  $t$  distribution arises as the distribution of the quotient of two independent random variables (2.10), the numerator of which has a **standardized normal distribution** (2.51) and the denominator is distributed as the positive square root of a **chi-squared distribution** (2.57) after dividing by its degrees of freedom. The parameter  $\nu$  is referred to as the **degrees of freedom** (2.54).

NOTE 3 The gamma function is defined in 2.56.

**2.54**

**degrees of freedom**

$\nu$

number of terms in a sum minus the number of constraints on the terms of the sum

NOTE This concept was previously encountered in the context of using  $n - 1$  in the denominator of the **estimator** (1.12) of the **sample variance** (1.16). The number of degrees of freedom is used to modify parameters. The term degrees of freedom is also widely used in ISO 3534-3 where mean squares are given as sums of squares divided by the appropriate degrees of freedom.

**2.55**

**$F$  distribution**

**continuous distribution** (2.23) having the **probability density function** (2.26)

where

$x > 0$

$\nu_1$  and  $\nu_2$  are positive integers

$\Gamma$  is the gamma function defined in 2.56.

NOTE 1 The  $F$  distribution is a useful reference distribution for assessing the ratio of independent **variances** (2.36).

NOTE 2 The  $F$  distribution arises as the distribution



thương số của hai biến ngẫu nhiên độc lập, mỗi biến có một phân bố khi bình phương (2.57), chia cho bậc tự do (2.54) của nó. Tham số  $\nu_1$  bậc tự do ở tử số còn  $\nu_2$  là bậc tự do mẫu số của phân bố  $F$ .

## 2.56

### Phân bố gamma

Phân bố liên tục (2.23) có hàm mật độ xác suất (2.26)

$$f(x) = \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)}$$

trong đó  $x > 0$  và các tham số  $\alpha > 0, \beta > 0$ .

CHÚ THÍCH 1: Phân bố gamma được sử dụng trong các ứng dụng liên quan tới độ tin cậy đối với mô hình thời gian tính đến khi hỏng. Phân bố này bao gồm phân bố mũ (2.58) như trường hợp đặc biệt cũng như các trường hợp khác với tỉ lệ hỏng tăng theo tuổi đời.

CHÚ THÍCH 2: Phân bố gamma được xác định bởi

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx.$$

Đối với các giá trị nguyên của  $\alpha$ ,  $\Gamma(\alpha) = (\alpha - 1)!$

CHÚ THÍCH 3: Trung bình (2.35) của phân bố gamma là  $\alpha\beta$ . Phương sai (2.36) của phân bố gamma là  $\alpha\beta^2$ .

## 2.57

### Phân bố khi-bình phương

#### Phân bố $\chi^2$

Phân bố liên tục (2.23) có hàm mật độ xác suất (2.26).

$$f(x) = \frac{x^{\frac{\nu}{2}-1} e^{-x/2}}{2^{\nu/2} \Gamma(\nu/2)}$$

trong đó  $x > 0$  và với  $\nu > 0$ .

CHÚ THÍCH 1: Đối với dữ liệu phát sinh từ phân bố chuẩn (2.50) có độ lệch chuẩn (2.37)  $\sigma$  đã biết, thống kê  $nS^2/\sigma^2$  có phân bố khi-bình phương với  $n - 1$

of the quotient of two independent random variables each having a chi-squared distribution (2.57), divided by its degrees of freedom (2.54). The parameter  $\nu_1$  is the numerator degrees of freedom and  $\nu_2$  is the denominator degrees of freedom of the  $F$  distribution.

## 2.56

### gamma distribution

continuous distribution (2.23) having the probability density function (2.26)

$$f(x) = \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)}$$

where  $x > 0$  and parameters  $\alpha > 0, \beta > 0$ .

NOTE 1 The gamma distribution is used in reliability applications for modelling time to failure. It includes the exponential distribution (2.58) as a special case as well as other cases with failure rates that increase with age.

NOTE 2 The gamma function is defined by

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx.$$

For integer values of  $\alpha$ ,  $\Gamma(\alpha) = (\alpha - 1)!$

NOTE 3 The mean (2.35) of the gamma distribution is  $\alpha\beta$ . The variance (2.36) of the gamma distribution is  $\alpha\beta^2$ .

## 2.57

### chi-squared distribution

#### $\chi^2$ distribution

continuous distribution (2.23) having the probability density function (2.26)

$$f(x) = \frac{x^{\frac{\nu}{2}-1} e^{-x/2}}{2^{\nu/2} \Gamma(\nu/2)}$$

where  $x > 0$  and with  $\nu > 0$ .

NOTE 1 For data arising from a normal distribution (2.50) with known standard deviation (2.37)  $\sigma$ , the statistic  $nS^2/\sigma^2$  has a chi-squared

bậc tự do. Kết quả này là cơ sở để thu được khoảng tin cậy đối với  $\sigma^2$ . Lĩnh vực ứng dụng khác của phân bố khi-bình phương là như phân bố tham chiếu đối với sự phù hợp của phép kiểm nghiệm.

CHÚ THÍCH 2: Phân bố này là một trường hợp đặc biệt của phân bố gamma (2.56) với các tham số  $\alpha = \nu/2$  và  $\beta = 2$ . Tham số  $\nu$  được gọi là bậc tự do (2.54).

CHÚ THÍCH 3: Trung bình (2.35) của phân bố khi bình phương là  $\nu$ . Phương sai (2.36) của phân bố khi bình phương là  $2\nu$ .

2.58

**Phân bố mũ**

Phân bố liên tục (2.23) có hàm mật độ xác suất (2.26).

$$f(x) = \beta^{-1} e^{-x/\beta}$$

trong đó  $x > 0$  và tham số  $\beta > 0$ .

CHÚ THÍCH 1: Phân bố mũ cung cấp cơ sở cho các ứng dụng liên quan đến độ tin cậy, tương ứng với trường hợp "không bị lão hóa" hoặc tính chất không có nhớ.

CHÚ THÍCH 2: Phân bố mũ là một trường hợp đặc biệt của phân bố gamma (2.56) với  $\alpha = 1$  hoặc tương đương phân bố khi bình phương (2.57) với  $\nu = 2$ .

CHÚ THÍCH 3: Trung bình (2.35) của phân bố mũ là  $\beta$ . Phương sai (2.36) của phân bố mũ là  $\beta^2$ .

2.59

**Phân bố beta**

Phân bố liên tục (2.23) có hàm mật độ xác suất (2.26).

$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}$$

trong đó  $0 \leq x \leq 1$  và với tham số  $\alpha, \beta > 0$ .

CHÚ THÍCH: Phân bố beta rất linh hoạt, có hàm mật độ xác suất có nhiều hình dạng khác nhau "đơn

distribution with  $n - 1$  degrees of freedom. This result is the basis for obtaining confidence intervals for  $\sigma^2$ . Another area of application for the chi-squared distribution is as the reference distribution for goodness of fit tests.

NOTE 2 This distribution is a special case of the gamma distribution (2.56) with parameters  $\alpha = \nu/2$  and  $\beta = 2$ . The parameter  $\nu$  is referred to as the degrees of freedom (2.54).

NOTE 3 The mean (2.35) of the chi-squared distribution is  $\nu$ . The variance (2.36) of the chi-squared distribution is  $2\nu$ .

2.58

**exponential distribution**

continuous distribution (2.23) having the probability density function (2.26)

$$f(x) = \beta^{-1} e^{-x/\beta}$$

where  $x > 0$  and with parameter  $\beta > 0$ .

NOTE 1 The exponential distribution provides a baseline in reliability applications, corresponding to the case of "lack of aging" or memory-less property.

NOTE 2 The exponential distribution is a special case of the gamma distribution (2.56) with  $\alpha = 1$  or equivalently, the chi-squared distribution (2.57) with  $\nu = 2$ .

NOTE 3 The mean (2.35) of the exponential distribution is  $\beta$ . The variance (2.36) of the exponential distribution is  $\beta^2$ .

2.59

**beta distribution**

continuous distribution (2.23) having the probability density function (2.26)

$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}$$

where  $0 \leq x \leq 1$  and with parameters  $\alpha, \beta > 0$ .

NOTE The beta distribution is highly flexible, having a probability density function that has a

thức, dạng chữ "j", dạng chữ "u"). Phân bố này có thể sử dụng như một mô hình của độ không đảm bảo đi kèm với một tỷ lệ. Ví dụ, trong ứng dụng bảo hiểm bão, tỷ lệ hư hỏng mong muốn đối với một loại kết cấu với tốc độ gió cho trước có thể là 0,40, mặc dù không phải gia đình nào gặp phải trường gió này cũng là do cùng loại hư hại. Phân bố beta có trung bình 0,40 có thể dùng cho mô hình thiệt hại không giống nhau theo loại kết cấu này.

## 2.60

### Phân bố đều

#### Phân bố hình chữ nhật

Phân bố liên tục (2.23) có hàm mật độ xác suất (2.26)

$$f(x) = \frac{1}{b-a}$$

trong đó  $a \leq x \leq b$ .

CHÚ THÍCH 1: Phân bố đều với  $a = 0$  và  $b = 1$  là phân bố cơ bản đối với bộ tạo số ngẫu nhiên điển hình.

CHÚ THÍCH 2: Trung bình (2.35) của phân bố đều là  $(a+b)/2$ . Phương sai (2.36) của phân bố đều là  $(b-a)^2/12$ .

CHÚ THÍCH 3: Phân bố đều là trường hợp đặc biệt của phân bố beta với  $\alpha = 1$  và  $\beta = 1$ .

## 2.61

### Phân bố cực trị loại I

#### Phân bố Gumbel

Phân bố liên tục (2.23) có hàm phân bố (2.7)

$$F(x) = e^{-e^{-x-a}}$$

trong đó  $-\infty < x < \infty$  với các tham số  $-\infty < a < \infty$ ,  $b > 0$ .

CHÚ THÍCH: Phân bố cực trị đưa ra phân bố tham chiếu thích hợp cho các thống kê thứ tự (1.9) cực trị  $X_{(1)}$  và  $X_{(n)}$ . Ba phân bố giới hạn có thể khi  $n$  có xu hướng tiến đến  $\infty$  được đưa ra bởi ba loại phân bố

variety of shapes (unimodal, "j"-shaped, "u"-shaped). The distribution can be used as a model of the uncertainty associated with a proportion. For example, in an insurance hurricane modelling application, the expected proportion of damage on a type of structure for a given wind speed might be 0,40, although not all houses experiencing this wind field will accrue the same damage. A beta distribution with mean 0,40 could serve to model the disparity in damage to this type of structure.

## 2.60

### uniform distribution

#### rectangular distribution

continuous distribution (2.23) having the probability density function (2.26)

$$f(x) = \frac{1}{b-a}$$

where  $a \leq x \leq b$ .

NOTE 1 The uniform distribution with  $a = 0$  and  $b = 1$  is the underlying distribution for typical random number generators.

NOTE 2 The mean (2.35) of the uniform distribution is  $(a+b)/2$ . The variance (2.36) of the uniform distribution is  $(b-a)^2/12$ .

NOTE 3 The uniform distribution is a special case of the beta distribution with  $\alpha = 1$  and  $\beta = 1$ .

## 2.61

### type I extreme value distribution

#### Gumbel distribution

continuous distribution (2.23) having the distribution function (2.7)

$$F(x) = e^{-e^{-x-a}}$$

where  $-\infty < x < \infty$  with parameters  $-\infty < a < \infty$ ,  $b > 0$ .

NOTE Extreme value distributions provide appropriate reference distributions for the extreme order statistics (1.9)  $X_{(1)}$  and  $X_{(n)}$ . The three possible

cực trị nêu trong 2.61, 2.62 và 2.63.

**2.62**

**Phân bố cực trị loại II**

**Phân bố Fréchet**

**Phân bố liên tục (2.23) có hàm phân bố (2.7).**

$$F(x) = e^{-\left(\frac{x-a}{b}\right)^k}$$

trong đó  $x > a$  và các tham số  $-\infty < a < \infty$ ,  $b > 0$ ,  $k > 0$ .

**2.63**

**Phân bố cực trị loại III**

**Phân bố Weibull**

**Phân bố liên tục (2.23) có hàm phân bố (2.7)**

$$F(x) = 1 - e^{-\left(\frac{x-a}{b}\right)^k}$$

trong đó  $x > a$  với các tham số  $-\infty < a < \infty$ ,  $b > 0$ ,  $k > 0$

**CHÚ THÍCH 1:** Ngoài việc dùng làm một trong ba phân bố giới hạn của thống kê thứ tự cực trị, phân bố Weibull chiếm vị trí quan trọng trong các ứng dụng khác nhau, đặc biệt trong nghiên cứu độ tin cậy và kỹ thuật. Phân bố Weibull đã chứng tỏ cung cấp tính phù hợp về thực nghiệm cho nhiều loại tập dữ liệu khác nhau.

**CHÚ THÍCH 2:** Tham số  $a$  là tham số định vị theo nghĩa là giá trị nhỏ nhất có thể có được trong phân bố Weibull. Tham số  $b$  là tham số thang đo [liên quan đến độ lệch chuẩn (2.37) của phân bố Weibull]. Tham số  $k$  là tham số định dạng.

**CHÚ THÍCH 3:** Đối với  $k = 1$ , phân bố Weibull có dạng phân bố hàm mũ. Nâng lên phân bố mũ với  $a = 0$  và tham số  $b$  lên lũy thừa  $1/k$  tạo thành phân bố Weibull theo định nghĩa. Một trường hợp đặc biệt khác của phân bố Weibull là phân bố Rayleigh (với  $a = 0$  và  $k = 2$ ).

limiting distributions as  $n$  tends to  $\infty$  are provided by the three types of extreme value distributions given in 2.61, 2.62 and 2.63.

**2.62**

**type II extreme value distribution**

**Fréchet distribution**

**continuous distribution (2.23) having the distribution function (2.7)**

$$F(x) = e^{-\left(\frac{x-a}{b}\right)^k}$$

where  $x > a$  and with parameters  $-\infty < a < \infty$ ,  $b > 0$ ,  $k > 0$ .

**2.63**

**type III extreme value distribution**

**Weibull distribution**

**continuous distribution (2.23) having distribution function (2.7)**

$$F(x) = 1 - e^{-\left(\frac{x-a}{b}\right)^k}$$

where  $x > a$  with parameters  $-\infty < a < \infty$ ,  $b > 0$ ,  $k > 0$

**NOTE 1** In addition to serving as one of the three possible limiting distributions of extreme order statistics, the Weibull distribution occupies a prominent place in diverse applications, particularly reliability and engineering. The Weibull distribution has been demonstrated to provide empirical fits to a variety of data sets.

**NOTE 2** The parameter  $a$  is a location parameter in the sense that it is the minimum value that the Weibull distribution can achieve. The parameter  $b$  is a scale parameter [related to the standard deviation (2.37) of the Weibull distribution]. The parameter  $k$  is a shape parameter.

**NOTE 3** For  $k = 1$ , the Weibull distribution is seen to include the exponential distribution. Raising an exponential distribution with  $a = 0$  and parameter  $b$  to the power  $1/k$  produces the Weibull distribution in the definition. Another special case of the Weibull

## 2.64

**Phân bố chuẩn nhiều chiều**

**Phân bố liên tục (2.23) có hàm mật độ xác suất (2.26)**

$$f(x) = (2\pi)^{-n/2} |\Sigma|^{-n/2} e^{-\frac{(x-\mu)^T \Sigma^{-1} (x-\mu)}{2}}$$

trong đó

$-\infty < x_i < \infty$  đối với mỗi  $i$ ;

$\mu$  là vector tham số  $n$  chiều;

$\Sigma$  là ma trận xác định dương, đối xứng  $n \times n$  tham số; và

chữ đậm chỉ thị vector  $n$  chiều.

**CHÚ THÍCH:** Mỗi phân bố biên duyên (2.18) của phân bố nhiều chiều trong điều này có phân bố chuẩn. Tuy nhiên, có nhiều phân bố nhiều chiều có phân bố biên duyên khác dạng phân bố nêu trong điều này.

## 2.65

**Phân bố chuẩn hai chiều**

**Phân bố liên tục (2.23) có hàm mật độ xác suất (2.26)**

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)} \left[ \left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2 \right]\right\}$$

trong đó

$-\infty < x < \infty$ ,

$-\infty < y < \infty$

$-\infty < \mu_x < \infty$ ,

$\infty < \mu_y < \infty$ ,

$\sigma_x > 0$

$\sigma_y > 0$

$|\rho| < 1$

**CHÚ THÍCH:** Theo ký hiệu gợi ý, đối với  $(X, Y)$  có hàm mật độ xác suất (2.26) nêu trên.  $E(X) = \mu_x$ ,  $E(Y) = \mu_y$ ,

distribution is the Rayleigh distribution (for  $a = 0$  and  $k = 2$ ).

## 2.64

**multivariate normal distribution**

**continuous distribution (2.23) having the probability density function (2.26)**

$$f(x) = (2\pi)^{-n/2} |\Sigma|^{-n/2} e^{-\frac{(x-\mu)^T \Sigma^{-1} (x-\mu)}{2}}$$

where

$-\infty < x_i < \infty$  for each  $i$ ;

$\mu$  is an  $n$ -dimensional parameter vector;

$\Sigma$  is an  $n \times n$  symmetric, positive definite matrix of parameters; and

the boldface indicates  $n$ -dimensional vectors.

**NOTE** Each of the marginal distributions (2.18) of the multivariate distribution in this clause have a normal distribution. However, there are many other multivariate distributions having normal marginal distributions besides the version of the distribution given in this clause.

## 2.65

**bivariate normal distribution**

**continuous distribution (2.23) having the probability density function (2.26)**

where

$-\infty < x < \infty$ ,

$-\infty < y < \infty$

$-\infty < \mu_x < \infty$ ,

$\infty < \mu_y < \infty$ ,

$\sigma_x > 0$

$\sigma_y > 0$

$|\rho| < 1$

**NOTE** As the notation suggests, for  $(X, Y)$  having the above probability density function (2.26),  $E(X)$

$V(X) = \sigma_x^2$ ,  $V(Y) = \sigma_y^2$ , và  $\rho$  là hệ số tương quan (2.44) giữa  $X$  và  $Y$ .

2.66

**Phân bố chuẩn chuẩn hóa hai chiều**

**Phân bố chuẩn hai chiều** (2.65) có các thành phần **phân bố chuẩn chuẩn hóa** (2.51).

2.67

**Phân bố mẫu**

Phân bố của một thống kê.

CHÚ THÍCH: Minh họa về phân bố mẫu cụ thể được nêu trong chú thích 1 của 2.53, chú thích 1 của 2.55 và chú thích 1 của 2.57.

2.68

**Không gian xác suất**

$(\Omega, \mathcal{N}, \rho)$

**Không gian mẫu** (2.1), và **sigma đại số của biến cố** (2.69), và **độ đo xác suất** (2.70).

VÍ DỤ 1: Một trường hợp đặc biệt, không gian mẫu có thể chứa tất cả 105 cá thể sản xuất trong một ngày quy định tại một nhà máy. Sigma đại số của biến cố gồm tất cả những tập con có thể. Các biến cố này bao gồm {không có cá thể nào}, {cá thể 1}, {cá thể 2}, ... {cá thể 105}, {cá thể 1 và cá thể 2}, ..., {tất cả 105 cá thể}. Một thước đo xác suất có thể được xác định là số cá thể trong một biến cố chia cho tổng số cá thể được sản xuất. Ví dụ, biến cố {cá thể 4, cá thể 27, cá thể 92} có độ đo xác suất là 3/105.

VÍ DỤ 2: Xét ví dụ thứ hai đối với tuổi thọ của pin. Nếu pin đến tay khách hàng và không hoạt động thì tuổi thọ là 0 h. Nếu pin hoạt động thì tuổi thọ tuân theo một **phân bố xác suất** (2.11) nhất định, ví dụ như **phân bố mũ** (2.58). Khi đó, thời gian tồn tại bị chi phối bởi xác suất pha trộn giữa rời rạc (tỷ lệ pin không hoạt động từ ban đầu) và liên tục (thời gian sống thực). Để đơn giản cho ví dụ này, giả định rằng tuổi thọ của pin tương đối ngắn so với thời gian nghiên cứu và thời gian tồn tại được đo liên tục. Tất nhiên, trên thực tế có thể bị ngắt theo dõi thời gian tồn tại bên phải hoặc

$= \mu_x$ ,  $E(Y) = \mu_y$ ,  $V(X) = \sigma_x^2$ ,  $V(Y) = \sigma_y^2$ , and  $\rho$  is the **correlation coefficient** (2.44) between  $X$  and  $Y$ .

2.66

**standardized bivariate normal distribution**

**bivariate normal distribution** (2.65) having **standardized normal distribution** (2.51) components

2.67

**sampling distribution**

distribution of a statistic

NOTE Illustrations of specific sampling distributions are given in Note 2 of 2.53, Note 1 of 2.55 and Note 1 of 2.57.

2.68

**probability space**

$(\Omega, \mathcal{N}, \rho)$

**sample space** (2.1), an associated **sigma algebra of events** (2.69), and a **probability measure** (2.70)

EXAMPLE 1 As a simple case, the sample space could consist of all the 105 items manufactured in a specified day at a plant. The sigma algebra of events consists of all possible subsets. Such events include {no items}, {item 1}, {item 2}, ... {item 105}, {item 1 and item 2}, ..., {all 105 items}. One possible probability measure could be defined as the number of items in an event divided by the total number of manufactured items. For example, the event {item 4, item 27, item 92} has probability measure 3/105.

EXAMPLE 2 As a second example, consider battery lifetimes. If the batteries arrive in the hands of the customer and they have no power, the survival time is 0 h. If the batteries are functional, then their survival times follow some **probability distribution** (2.11), such as an **exponential** (2.58). The collection of survival times is then governed by a distribution that is a mixture between discrete (the proportion of batteries that are not functional to begin with) and continuous (an actual survival time).

bên trái (ví dụ, thời gian hỏng được biết có thể xảy ra ít nhất là 5 h hoặc từ 3 đến 3,5 h), trong đó các ưu điểm của kết cấu này có thể phát huy. Không gian mẫu gồm một nửa đường thẳng thực (các số thực lớn hơn hoặc bằng không). Sigma đại số của biến cố gồm tất cả các khoảng của dạng  $[0, r]$  và tập  $\{0\}$ . Ngoài ra, sigma đại số bao gồm tất cả các hợp đếm được và giao đếm được của các tập này. Độ đo xác suất xác định đối với từng tập hợp, thành phần của nó đại diện cho pin không hoạt động và pin có thời gian tồn tại dương. Chi tiết về việc tính toán thời gian hỏng được đề cập trong toàn bộ điều này ở những chỗ thích hợp.

## 2.69

### Sigma đại số của biến cố

$\sigma$  đại số

sigma trường

$\sigma$  trường

$\mathcal{N}$

tập hợp các biến cố (2.2) có tính chất:

- thuộc  $\mathcal{N}$ ;
- Nếu một biến cố thuộc  $\mathcal{N}$ , thì biến cố bù (2.3) cũng thuộc  $\mathcal{N}$ ;
- Nếu  $\{A_i\}$  là tập hợp biến cố bất kỳ trong  $\mathcal{N}$ , thì hợp  $\bigcup_{i=1}^{\infty} A_i$  và giao  $\bigcap_{i=1}^{\infty} A_i$  của biến cố cũng thuộc  $\mathcal{N}$ .

VÍ DỤ 1: Nếu không gian mẫu là tập hợp các số nguyên thì sigma đại số của biến cố có thể được chọn là tập hợp tất cả các tập con của các số nguyên đó.

VÍ DỤ 2: Nếu không gian mẫu là tập hợp các số thực

For simplicity in this example, it is assumed that the lifetimes of the batteries are relatively short compared to the study time and that all survival times are measured on the continuum. Of course, in practice the possibility of right or left censored survival times (for example, the failure time is known to be at least 5 h or the failure time is between 3 and 3,5 h) could occur, in which case, further advantages of this structure would emerge. The sample space consists of half of the real line (real numbers greater than or equal to zero). The sigma algebra of events includes all intervals of the form  $[0, x]$  and the set  $\{0\}$ . Additionally, the sigma algebra includes all countable unions and intersections of these sets. The probability measure involves determining for each set, its constituents that represent non-functional batteries and those having a positive survival time. Details on the computations associated with the failure times have been given throughout this clause where appropriate.

## 2.69

### sigma algebra of events

$\sigma$ -algebra

sigma field

$\sigma$ -field

$\mathcal{N}$

set of events (2.2) with the properties:

- belongs to  $\mathcal{N}$ ;
- If an event belongs to  $\mathcal{N}$ , then its **complementary event** (2.3) also belongs to  $\mathcal{N}$ ;
- If  $\{A_i\}$  is any set of events in  $\mathcal{N}$ , then the union  $\bigcup_{i=1}^{\infty} A_i$  and the intersection  $\bigcap_{i=1}^{\infty} A_i$  of the events belong to  $\mathcal{N}$ .

EXAMPLE 1 If the sample space is the set of integers, then a sigma algebra of events may be chosen to be the set of all subsets of the integers.

EXAMPLE 2 If the sample space is the set of

thì sigma đại số của biến cố có thể được chọn bao gồm các tập hợp ứng với các khoảng trên đường thẳng thực và tất cả các hợp vô hạn và đếm được cũng như giao của các khoảng này. Có thể mở rộng ví dụ này ra số chiều lớn hơn bằng cách xét "khoảng"  $k$  chiều. Cụ thể, với hai chiều, tập hợp các khoảng có thể bao gồm các vùng xác định bởi  $\{(x,y): x < s, y < t\}$  đối với tất cả các giá trị thực của  $s$  và  $t$ .

CHÚ THÍCH 1: Sigma đại số là tập hợp bao gồm phần tử là các tập hợp. Tập hợp các kết quả  $\Omega$  có thể có là phần tử của sigma đại số của biến cố, như chỉ ra trong tính chất a).

CHÚ THÍCH 2: Tính chất c) liên quan đến tập hợp các phép tính trên các tập con (có thể là hữu hạn đếm được) của sigma đại số của biến cố. Ký hiệu được cho chỉ ra rằng tất cả các hợp đếm được và giao đếm được của các tập hợp này cũng thuộc sigma đại số của biến cố.

CHÚ THÍCH 3: Tính chất c) bao hàm cả tính khép kín (các tập hợp thuộc sigma đại số của biến cố) theo hợp hữu hạn hoặc giao hữu hạn. Hạn định sigma được dùng để nhấn mạnh rằng A khép kín trong các hợp hoặc giao vô hạn đếm được.

## 2.70

### Độ đo xác suất

$\wp$

Hàm không âm xác định trên sigma đại số của biến cố (2.69) sao cho

$$a) \wp(\Omega) = 1,$$

trong đó  $\Omega$  biểu thị không gian mẫu (2.1),

$$b) \wp\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \wp(A_i)$$

trong đó  $\{A_i\}$  là dãy các biến cố (2.2) tách rời nhau đôi một.

VÍ DỤ: Tiếp theo ví dụ về pin ở 2.1, xét biến cố pin hoạt động ít hơn một giờ. Biến cố này gồm cặp các biến cố tách rời {không hoạt động} và {hoạt động ít hơn một giờ nhưng hoạt động từ đầu}. Tương ứng, có thể biểu thị các biến cố  $\{0\}$  và  $(0,1)$ . Độ đo xác suất của  $\{0\}$  là tỷ lệ pin không hoạt động ngay từ lần đầu.

real numbers, then a sigma algebra of events may be chosen to include all sets corresponding to intervals on the real line and all their finite and countable unions and intersections of these intervals. This example can be extended to higher dimensions by considering  $k$ - dimensional "intervals". In particular, in two dimensions, the set of intervals could consist of regions defined by  $\{(x,y): x < s, y < t\}$  for all real values of  $s$  and  $t$ .

NOTE 1 A sigma algebra is a set consisting of sets as its members. The set of all possible outcomes  $\Omega$  is a member of the sigma algebra of events, as indicated in property a).

NOTE 2 Property c) involves set operations on a collection of subsets (possibly countably infinite) of the sigma algebra of events. The notation given indicates that all countable unions and intersections of these sets also belong to the sigma algebra of events.

NOTE 3 Property c) includes closure (the sets belong to the sigma algebra of events) under either finite unions or intersections. The qualifier sigma is used to stress that A is closed even under countably infinite operations on sets.

## 2.70

### probability measure

$\wp$

non-negative function defined on the sigma algebra of events (2.69) such that

$$a) \wp(\Omega) = 1,$$

where  $\Omega$  denotes the sample space (2.1),

$$b) \wp\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \wp(A_i)$$

where  $\{A_i\}$  is a sequence of pair-wise disjoint events (2.2).

EXAMPLE Continuing the battery life example of 2.1, consider the event that the battery survives less than one hour. This event consists of the disjoint pair of events {does not function} and {functions less than one hour but functions initially}. Equivalently, the events can be denoted  $\{0\}$  and



Độ đo xác suất của tập hợp  $(0, 1)$  phụ thuộc vào xác suất phân bố liên tục cụ thể [ví dụ, phân bố mũ (2.58)] chi phối phân bố hỏng.

CHÚ THÍCH 1: Độ đo xác suất ấn định một giá trị từ  $[0, 1]$  đối với mỗi biến cố trong sigma đại số của biến cố. Giá trị 0 ứng với biến cố không thể xảy ra, trong khi giá trị 1 biểu diễn khả năng chắc chắn xảy ra. Cụ thể, độ đo xác suất của tập trống bằng không và độ đo xác suất ấn định cho không gian mẫu là 1.

CHÚ THÍCH 2: Tính chất b) chỉ ra rằng nếu dãy các biến cố không có phần chung khi xét theo cặp thì độ đo xác suất của hợp các biến cố đó là tổng các độ đo xác suất đơn lẻ. Để chỉ rõ thêm cho tính chất b), nếu số biến cố là vô hạn đếm được thì điều này vẫn đúng.

CHÚ THÍCH 3: Ba thành phần của xác suất có liên hệ thông qua các biến ngẫu nhiên. Xác suất (2.5) của các biến cố trong tập ảnh của biến ngẫu nhiên (2.10) tính được từ xác suất của biến cố trong không gian mẫu. Biến cố trong tập ảnh của biến ngẫu nhiên được ấn định xác suất của biến cố trong không gian mẫu chiếu lên nó bằng biến ngẫu nhiên.

CHÚ THÍCH 4: Tập ảnh của biến ngẫu nhiên là tập số thực hoặc bộ  $n$  số thực. (Chú ý tập ảnh là tập hợp mà biến ngẫu nhiên chiếu lên đó.)

$(0,1)$ . The probability measure of  $\{0\}$  is the proportion of batteries that do not function upon the initial attempt. The probability measure of the set  $(0, 1)$  depends on the specific continuous probability distribution [for example, exponential (2.58)] governing the failure distribution.

NOTE 1 A probability measure assigns a value from  $[0, 1]$  for each event in the sigma algebra of events. The value 0 corresponds to an event being impossible, while the value 1 represents certainty of occurrence. In particular, the probability measure associated with the null set is zero and the probability measure assigned to the sample space is 1.

NOTE 2 Property b) indicates that if a sequence of events has no elements in common when considered in pairs, then the probability measure of the union is the sum of the individual probability measures. As further indicated in property b), this holds if the number of events is countably infinite.

NOTE 3 The three components of the probability are effectively linked via random variables. The probabilities (2.5) of the events in the image set of the random variable (2.10) derive from the probabilities of events in the sample space. An event in the image set of the random variable is assigned the probability of the event in the sample space that is mapped onto it by the random variable.

NOTE 4 The image set of the random variable is the set of real numbers or the set of ordered  $n$ -tuples of real numbers. (Note that the image set is the set onto which the random variable maps.)

## Phụ lục A

(tham khảo)

## Ký hiệu

## Annex A

(informative)

## Symbols

Ký hiệu Symbol(s)	Thuật ngữ Tiếng Việt Vietnamese term	Thuật ngữ tiếng Anh English term	Số điều Term No.
$A$	biến cố	event	2.2
$A^C$	biến cố bù	complementary event	2.3
$\mathcal{N}$	sigma đại số của biến cố, $\sigma$ đại số, sigma trường $\sigma$ -trường	sigma algebra of events, $\sigma$ algebra, sigma field $\sigma$ -field	2.69
$\alpha$	mức ý nghĩa	significance level	1.45
$\alpha, \lambda, \mu, \beta, \sigma, \rho, \gamma$ $p, N, M, c, v, a, b, k$	tham số	parameter	
$\beta_2$	hệ số nhọn	coefficient of kurtosis	2.40
$E(X^k)$	mômen mẫu bậc $k$	sample moment of order $k$	1.14
$E[g(X)]$	kỳ vọng của hàm $g$ của biến ngẫu nhiên $X$	expectation of the function $g$ of a random variable $X$	2.12
$F(x)$	hàm phân bố	distribution function	2.7
$f(x)$	hàm mật độ xác suất	probability density function	2.26
$\gamma_1$	hệ số bất đối xứng	coefficient of skewness	2.39
$H$	giả thuyết	hypothesis	1.40
$H_0$	giả thuyết không	null hypothesis	1.41
$H_A, H_1$	đối giả thuyết	alternative hypothesis	1.42
$k$	số chiều	dimension	
$k, r, s$	bậc của mômen	order of a moment	1.14, 2.34, 2.41, 2.42
$\mu$	trung bình	mean	2.35
$\nu$	bậc tự do	degrees of freedom	2.54
$n$	cỡ mẫu	sample size	
$\Omega$	không gian mẫu	sample space	2.1
$(\Omega, \mathcal{N}, \rho)$	không gian xác suất	probability space	2.68
$P(A)$	xác suất của biến cố $A$	probability of an event $A$	2.5
$P(A B)$	xác suất có điều kiện của $A$ cho trước $B$	conditional probability of $A$ given $B$	2.6
$\rho$	độ đo xác suất	probability measure	2.70
$r_{xy}$	hệ số tương quan mẫu	sample correlation coefficient	1.23
$s$	giá trị quan trắc của độ lệch chuẩn mẫu	observed value of a sample standard deviation	
$S$	độ lệch chuẩn mẫu	sample standard deviation	1.17

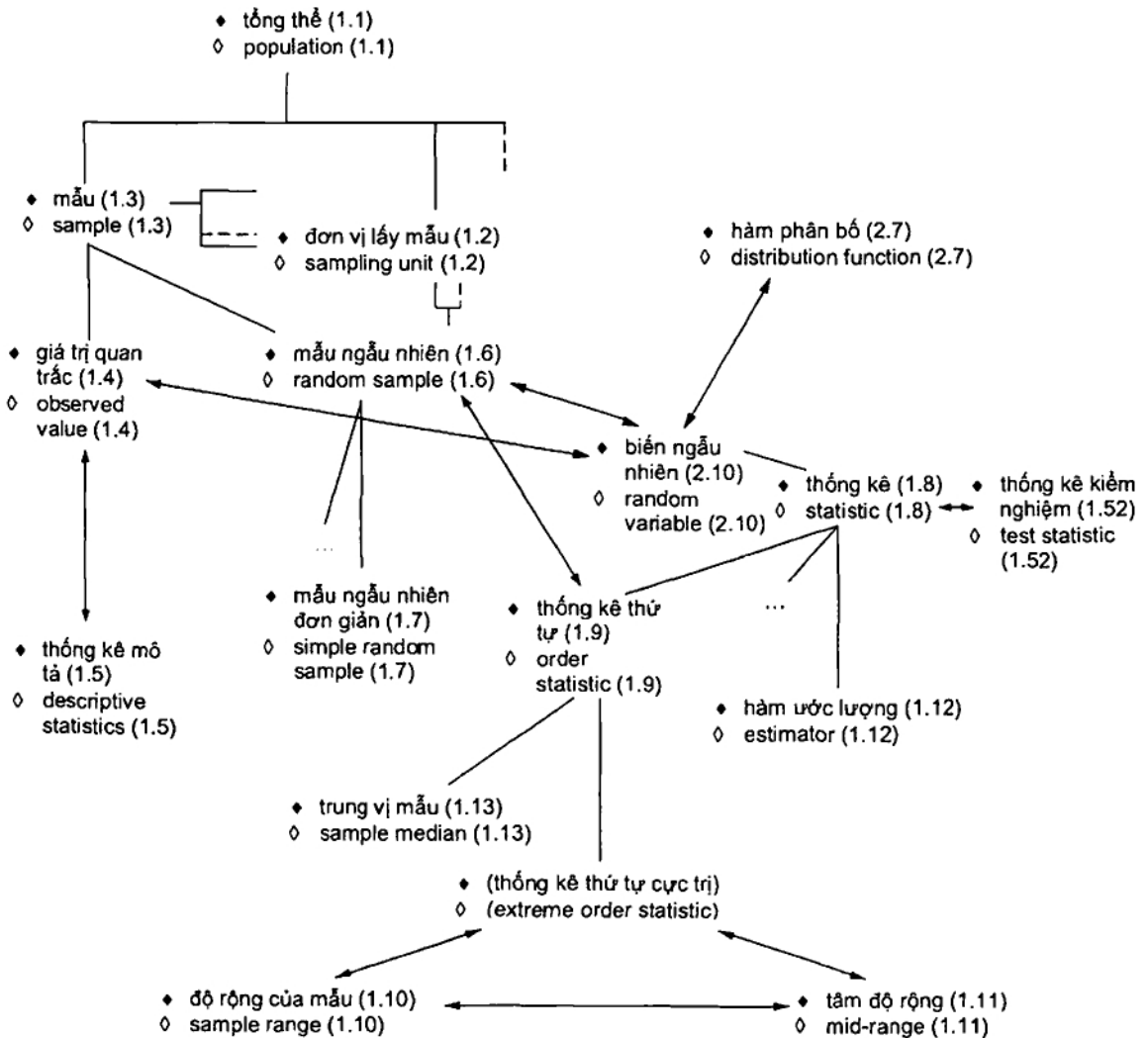
Ký hiệu Symbol(s)	Thuật ngữ Tiếng Việt Vietnamese term	Thuật ngữ tiếng Anh English term	Số điều Term No.
$S^2$	phương sai mẫu	sample variance	1.16
$S_{XY}$	hiệp phương sai mẫu	sample covariance	1.22
$\sigma$	độ lệch chuẩn	standard deviation	2.37
$\sigma^2$	phương sai	variance	2.36
$\sigma_{XY}$	hiệp phương sai	covariance	2.43
$\sigma_{\hat{\theta}}$	sai số tiêu chuẩn	standard error	1.24
$\sigma_{\bar{x}}$	sai số tiêu chuẩn của trung bình mẫu	standard error of the sample mean	
$\theta$	tham số của phân bố	parameter of a distribution	
$\hat{\theta}$	hàm ước lượng	estimator	1.12
$V(X)$	phương sai của biến ngẫu nhiên $X$	variance of a random variable $X$	2.36
$X_{(i)}$	thống kê thứ tự $i$	$i^{\text{th}}$ order statistic	1.9
$x, y, z$	giá trị quan trắc	observed value	1.4
$X, Y, Z, T$	biến ngẫu nhiên	random variable	2.10
$X_p, x_p$	$p$ phân vị	$p$ -quantile	2.13
$\bar{X}, \bar{x}$	trung bình, trung bình mẫu	average, sample mean	1.15

**Phụ lục B**  
(tham khảo)

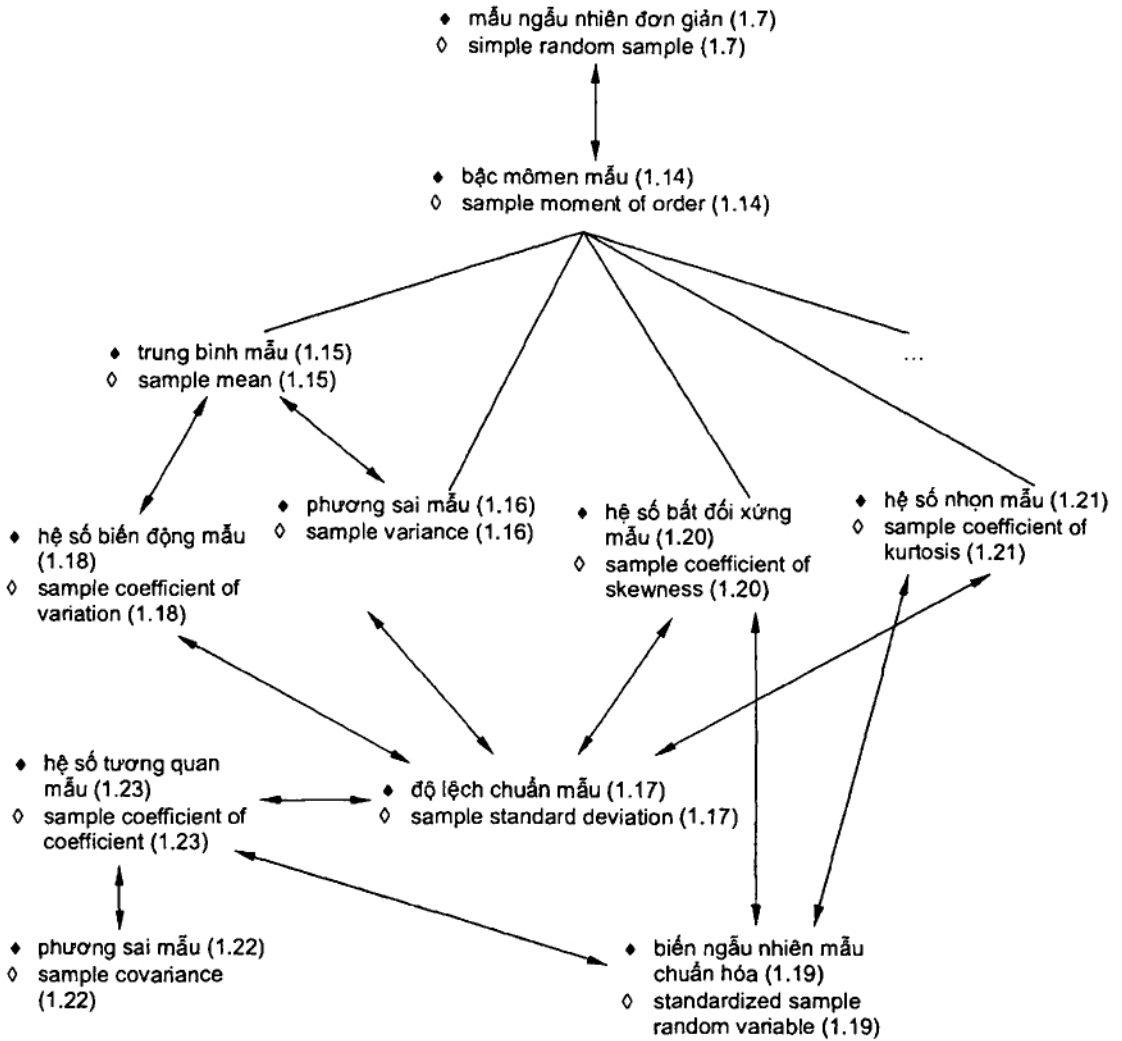
**Sơ đồ khái niệm thống kê**

**Annex B**  
(informative)

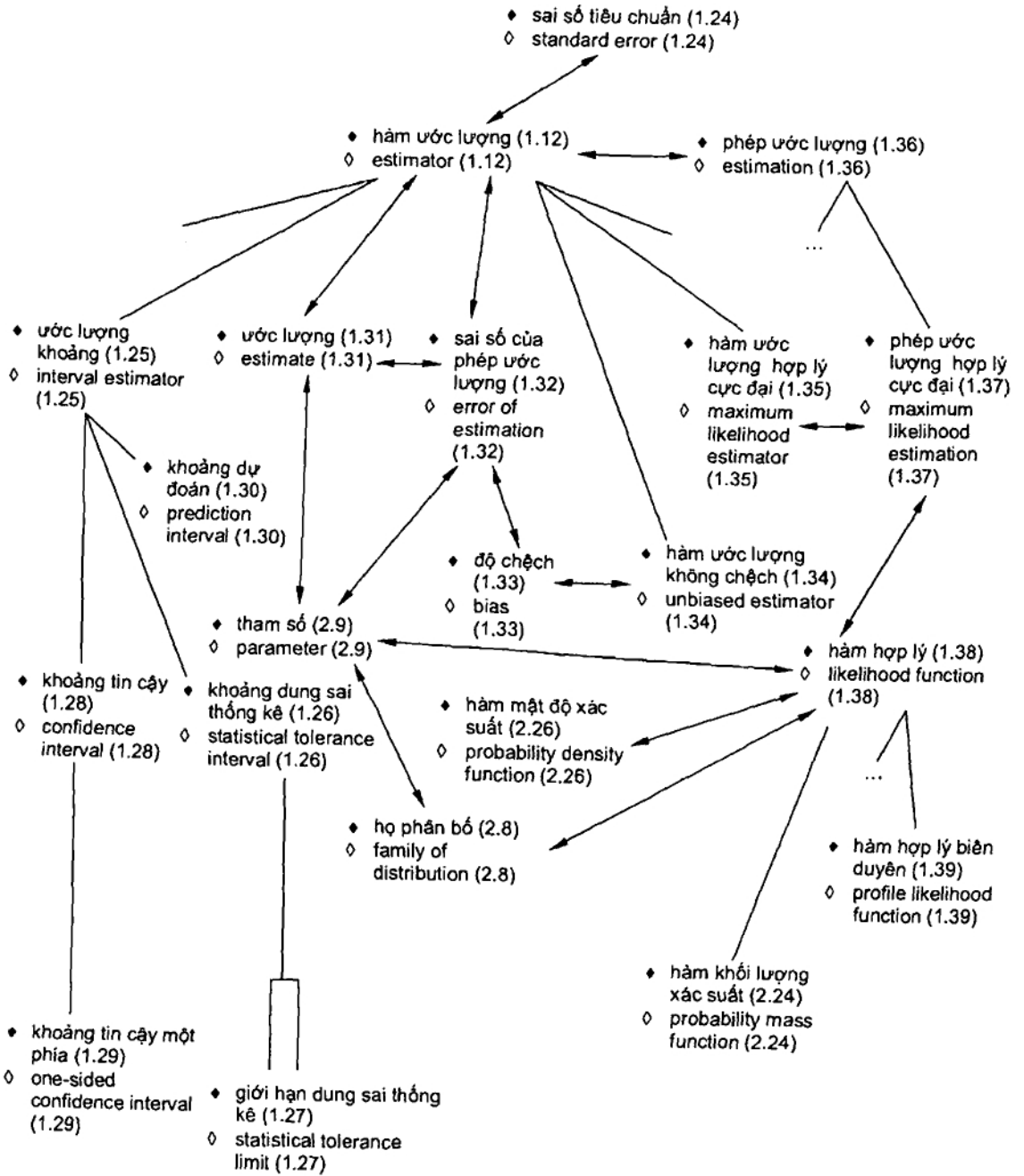
**Statistical concept diagrams**



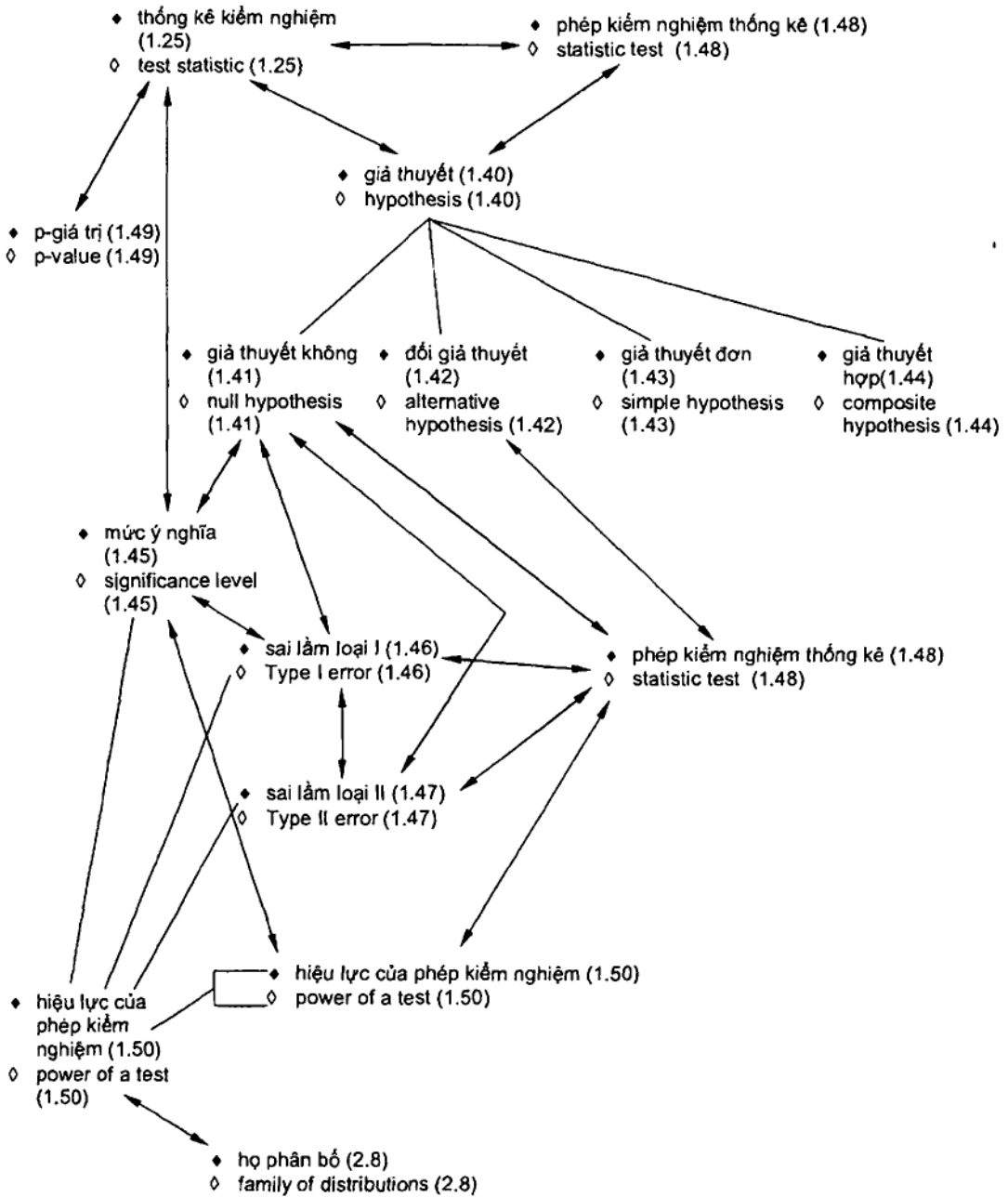
**Hình B.1 – Khái niệm cơ bản về tổng thể và mẫu**  
**Figure B.1 — Basic population and sample concepts**



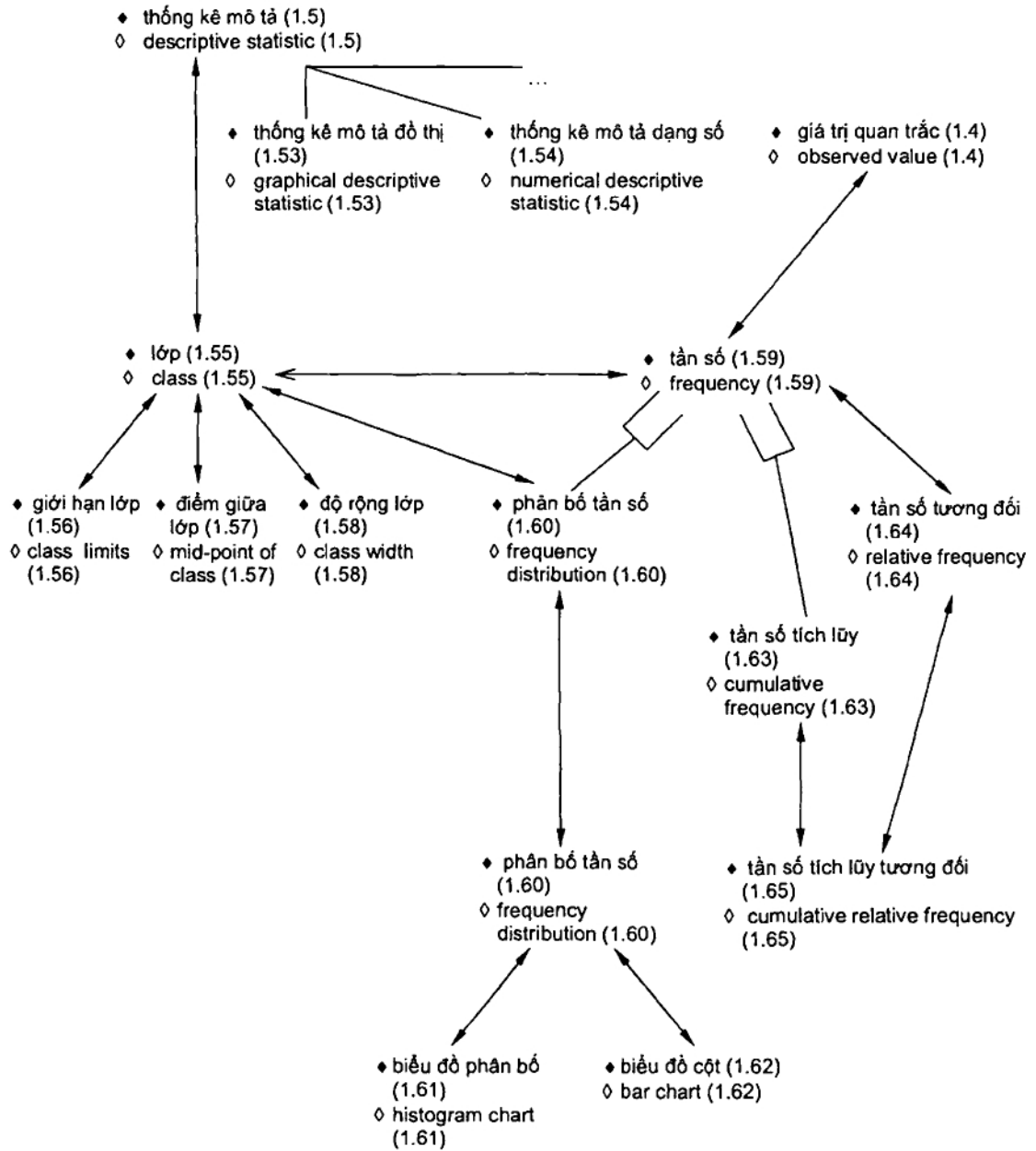
Hình B.2 – Khái niệm về mômen mẫu  
 Figure B.2 – Concepts regarding sample moments



Hình B.3 – Khái niệm ước lượng  
Figure B.3 – Estimation concepts



Hình B.4 – Khái niệm liên quan đến phép kiểm nghiệm thống kê  
 Figure B.4 – Concepts regarding statistical tests



Hình B.5 – Khái niệm liên quan đến lớp và phân bố thực nghiệm  
 Figure B.5 – Concepts regarding classes and empirical distributions



### Phụ lục C

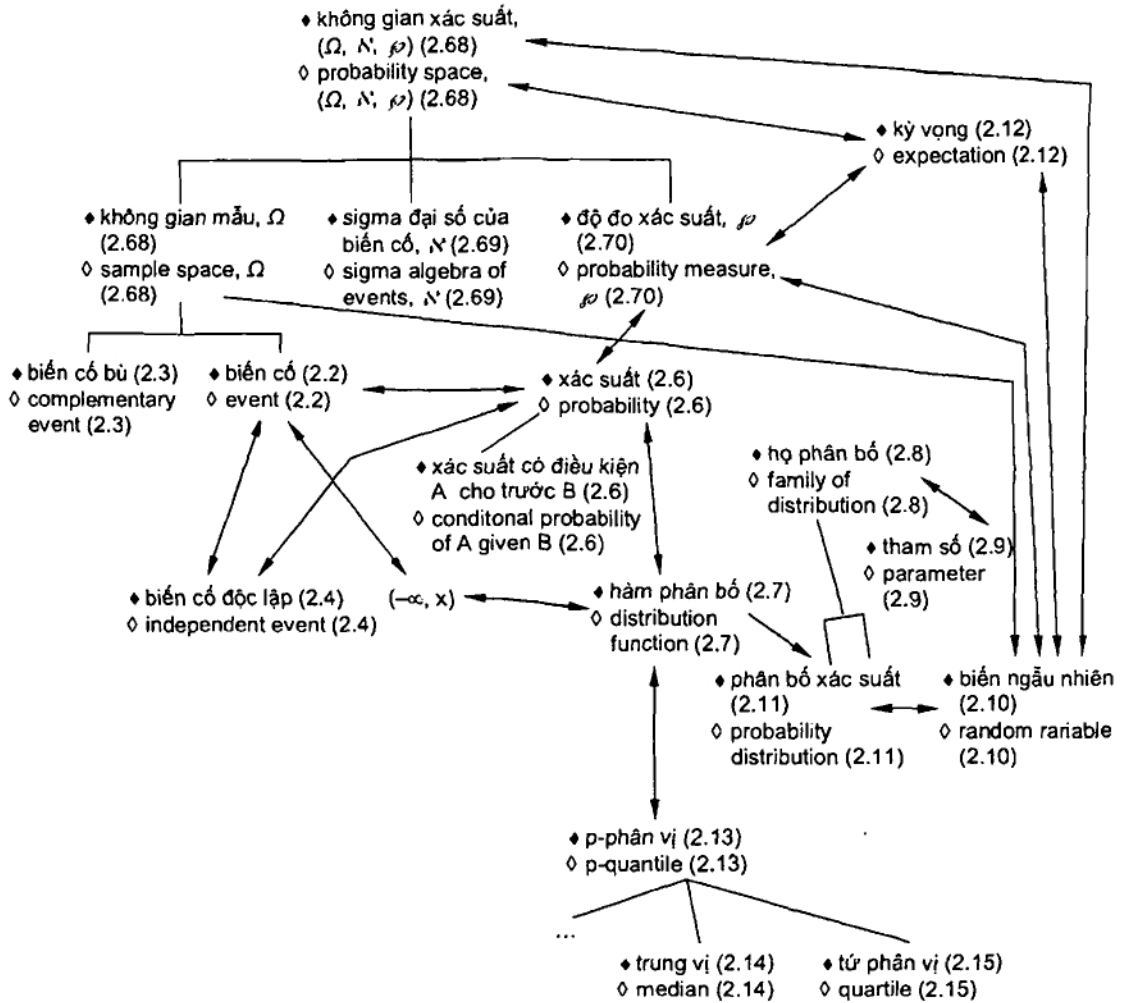
(tham khảo)

### Sơ đồ khái niệm xác suất

### Annex C

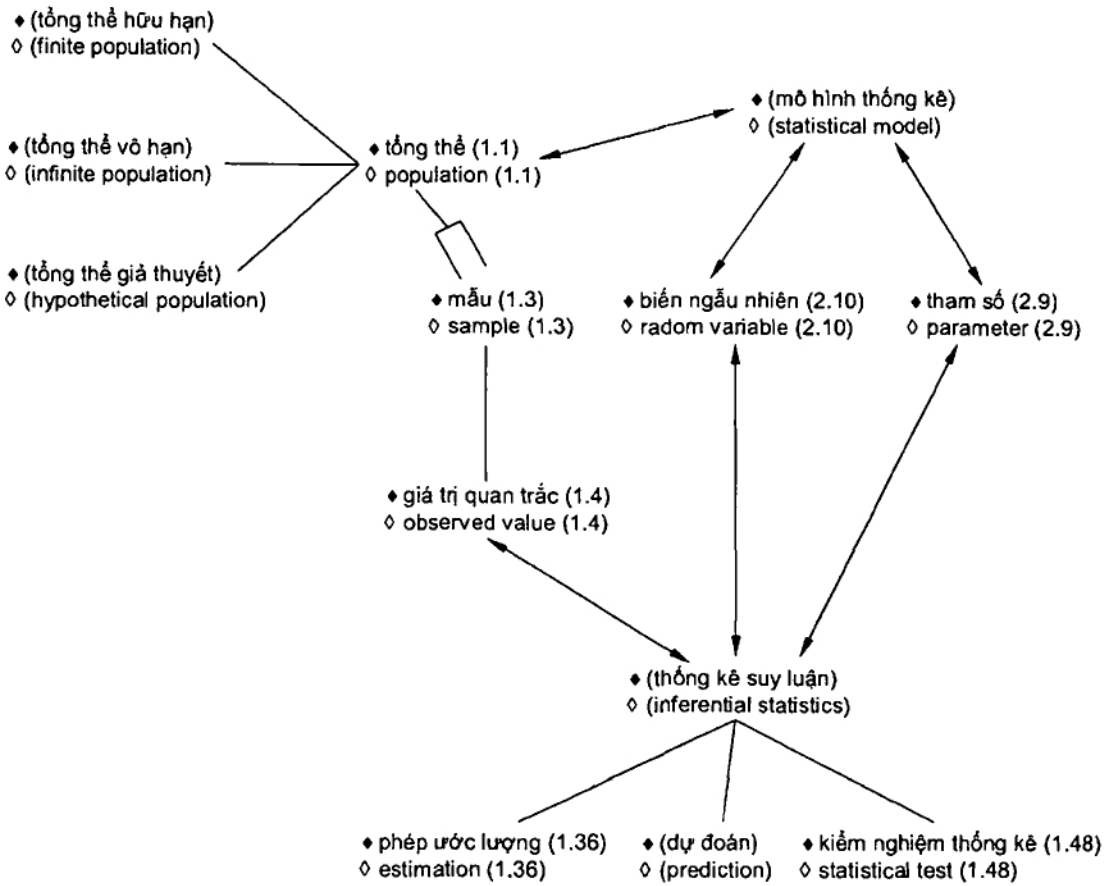
(informative)

### Probability concept diagrams

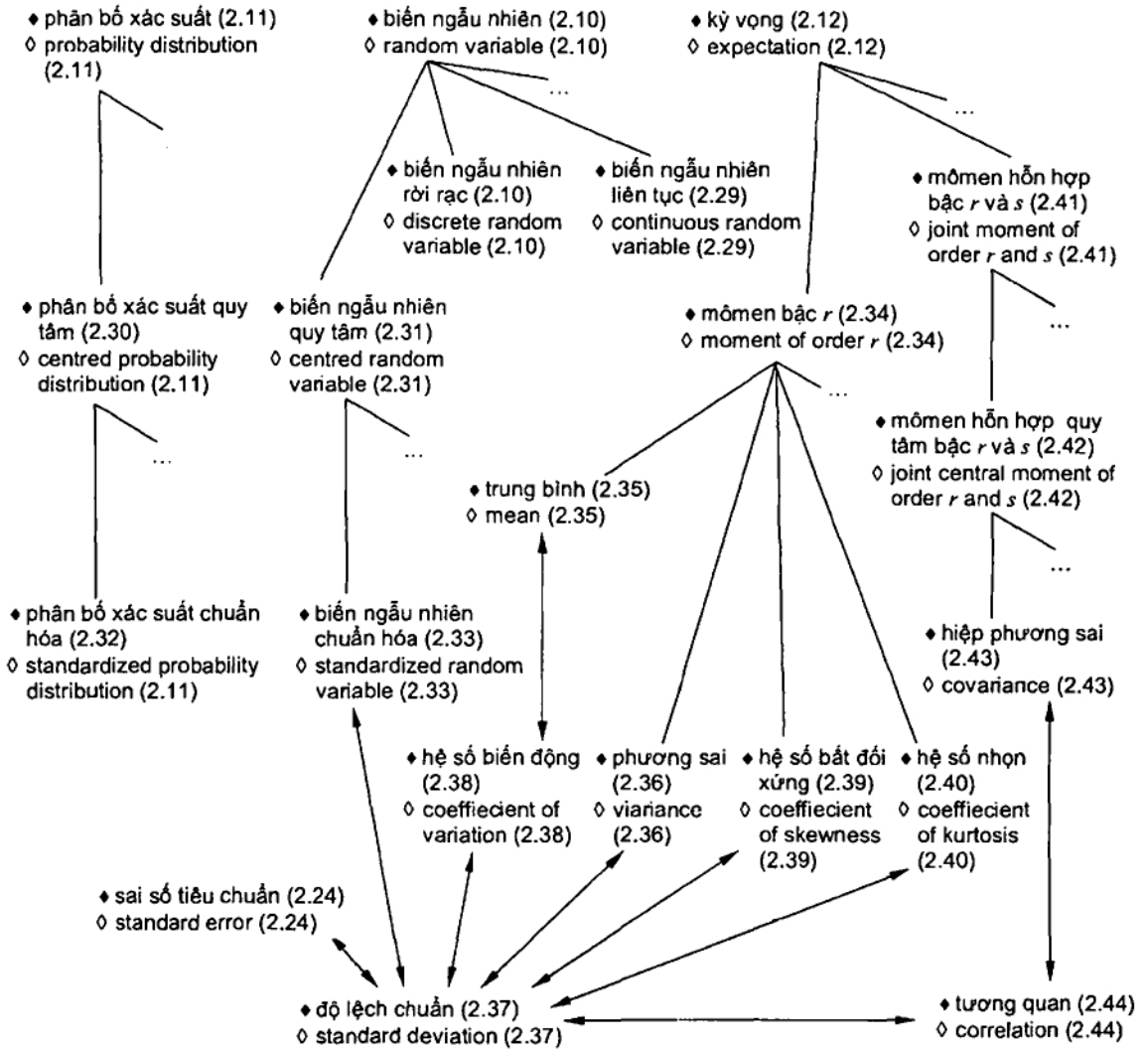


Hình C.1 – Khái niệm cơ bản dùng trong xác suất

Figure C.1 – Fundamental concepts in probability

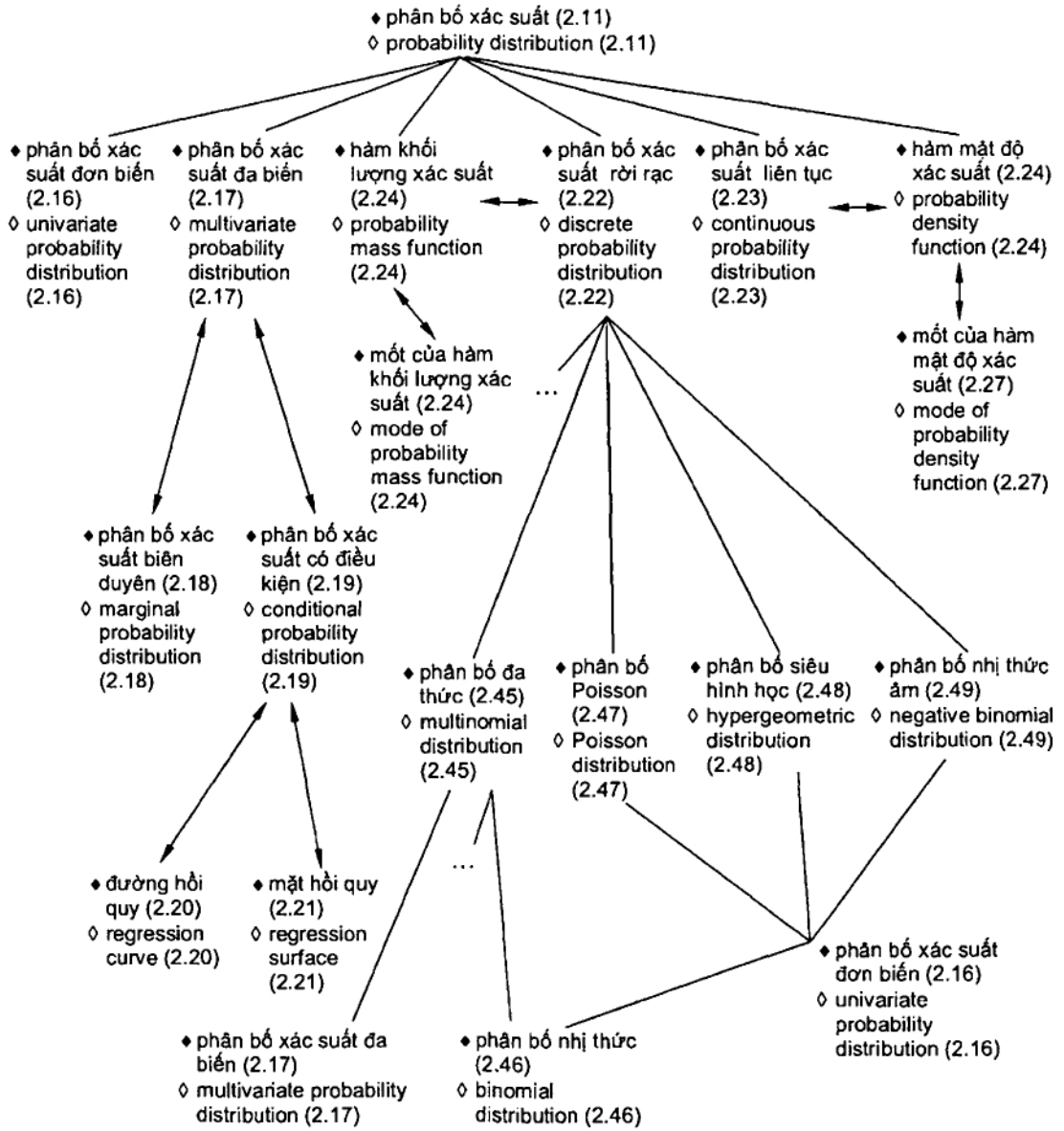


Hình B.6 – Sơ đồ khái niệm suy luận thống kê  
 Figure B.6 – Statistical inference concept diagram

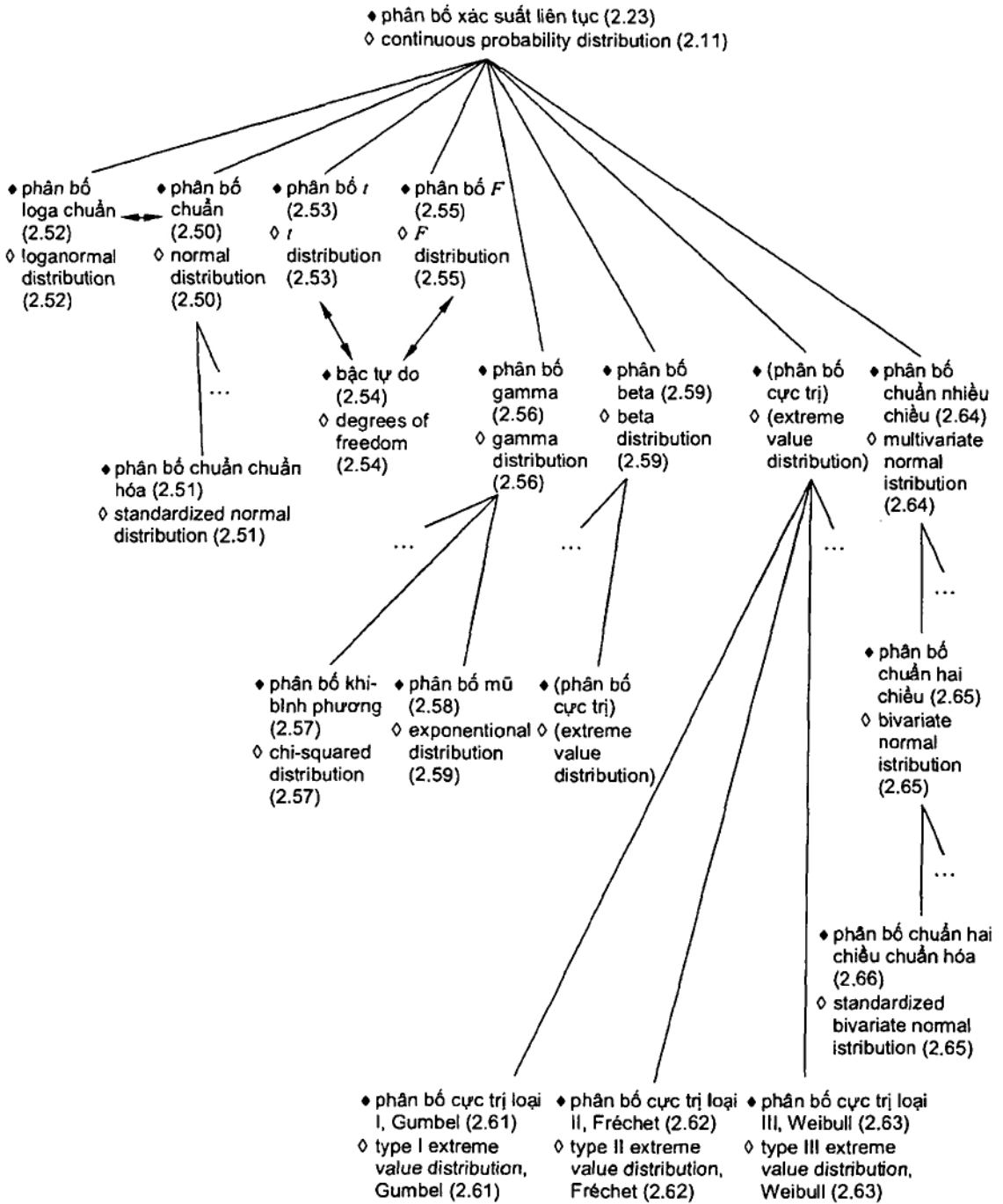


Hình C.2 – Khái niệm liên quan đến mômen

Figure C.2 – Concepts regarding moments



Hình C.3 – Khái niệm liên quan đến phân bố xác suất  
 Figure C.3 – Concepts regarding probability distributions



Hình C.4 – Khái niệm liên quan đến phân bố liên tục  
Figure C.4 – Concepts regarding continuous distributions

**Phụ lục D**

(tham khảo)

**Phương pháp luận sử dụng khi xây dựng từ vựng****D.1 Giới thiệu**

Ứng dụng phổ biến của bộ tiêu chuẩn này đòi hỏi phải sử dụng bộ từ vựng chặt chẽ và hài hòa sao cho những người sử dụng các tiêu chuẩn thống kê ứng dụng hiểu được một cách dễ dàng.

Các khái niệm đều có liên quan đến nhau và việc phân tích mối quan hệ giữa các khái niệm trong lĩnh vực thống kê ứng dụng cũng như sắp xếp chúng theo các sơ đồ khái niệm là điều tiên quyết đối với một bộ từ vựng chặt chẽ. Phân tích này được sử dụng khi xây dựng tiêu chuẩn này. Vì sơ đồ khái niệm sử dụng trong quá trình xây dựng từ vựng có thể hữu ích cho việc tham khảo nên các sơ đồ đã được nhắc lại trong D.4.

**D.2 Nội dung mục từ vựng và quy tắc thay thế**

Khái niệm là đơn vị chuyển đổi giữa các ngôn ngữ (bao gồm cả những khác biệt trong một ngôn ngữ, ví dụ như tiếng Anh Mỹ và tiếng Anh Anh). Đối với mỗi ngôn ngữ, thuật ngữ thích hợp nhất cho tính rõ ràng phổ dụng của khái niệm ở ngôn ngữ đó, nghĩa là không phải một bản dịch, sẽ lựa được chọn.

Định nghĩa hình thành bởi việc mô tả chỉ những đặc trưng thiết yếu để nhận biết khái niệm. Thông tin liên quan đến khái niệm cũng quan trọng nhưng không cần thiết cho việc mô tả khái niệm được đặt trong một hoặc nhiều chú thích kèm theo định nghĩa.

Khi thuật ngữ được thay bằng định nghĩa của nó,

**Annex D**

(informative)

**Methodology used in the development of the vocabulary****D.1 Introduction**

The universal application of the ISO family of standards requires the employment of a coherent and harmonized vocabulary that is easily understandable by potential users of applied statistics standards.

Concepts are interrelated, and an analysis of these relationships among concepts within the field of applied statistics and their arrangement into concept diagrams is a prerequisite of a coherent vocabulary. Such an analysis was used in the development of this part of ISO 3534. Since the concept diagrams employed during the development process may be helpful in an informative sense, they are reproduced in D.4.

**D.2 Content of a vocabulary entry and the substitution rule**

The concept forms the unit of transfer between languages (including variants within one language, e.g. American English and British English). For each language, the most appropriate term for the universal transparency of the concept in that language, i.e. not a literal approach to translation, is chosen.

A definition is formed by describing only those characteristics that are essential to identify the concept. Information concerning the concept which is important but which is not essential to its description is put in one or more notes to the definition.

When a term is substituted by its definition,

có thay đổi nhỏ về cú pháp, sẽ không thay đổi nghĩa văn bản. Sự thay thế như vậy tạo ra một phương pháp đơn giản để kiểm tra tính chính xác của định nghĩa. Tuy nhiên, khi định nghĩa phức tạp theo hướng bao hàm nhiều thuật ngữ, tốt nhất là thực hiện việc thay thế cho một hoặc nhiều nhất là hai định nghĩa. Việc thay thế hoàn toàn tất cả các thuật ngữ sẽ làm cho việc đạt được cú pháp trở nên khó khăn và sẽ không có ích trong việc truyền tải nghĩa.

### **D.3 Mối quan hệ của khái niệm và việc thể hiện chúng bằng sơ đồ**

#### **B.3.1 Khái quát**

Trong thuật ngữ học, ở chừng mực có thể, mối quan hệ giữa các khái niệm được dựa trên thông tin thứ bậc về đặc trưng của loại sao cho mô tả ngắn gọn nhất khái niệm bằng cách gọi tên loại của nó và mô tả đặc trưng phân biệt nó với các khái niệm mẹ hoặc anh em.

Có ba dạng quan hệ khái niệm chính được nêu trong phụ lục này: chung (D.3.2), bộ phận (D.3.3) và liên kết (D.3.4).

#### **D.3.2 Mối quan hệ chung**

Các khái niệm phụ trong hệ thống thứ bậc kế thừa tất cả các đặc trưng của khái niệm chính và bao gồm mô tả về các đặc trưng này để phân biệt chúng với các khái niệm gốc (mẹ) và ngang bằng (anh em), ví dụ quan hệ giữa xuân, hè, thu và đông với mùa.

Mối quan hệ chung được mô tả bằng một sơ đồ quạt hoặc cây không có mũi tên (xem Hình D.1).

subject to minor syntax changes, there should be no change in the meaning of the text. Such a substitution provides a simple method for checking the accuracy of a definition. However, where the definition is complex in the sense that it contains a number of terms, substitution is best carried out taking one or, at most, two definitions at a time. Complete substitution of the totality of the terms will become difficult to achieve syntactically and will be unhelpful in conveying meaning.

### **D.3 Concept relationships and their graphical representation**

#### **D.3.1 General**

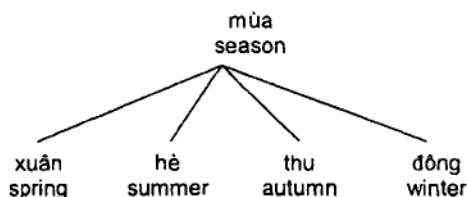
In terminology work, the relationships between concepts are based on the hierarchical formation of the characteristics of a species so that the most economical description of a concept is formed by naming its species and describing the characteristics that distinguish it from its parent or sibling concepts.

There are three primary forms of concept relationships indicated in this annex: generic (D.3.2), partitive (D.3.3) and associative (D.3.4).

#### **D.3.2 Generic relation**

Subordinate concepts within the hierarchy inherit all the characteristics of the superordinate concept and contain descriptions of these characteristics which distinguish them from the superordinate (parent) and coordinate (sibling) concepts, e.g. the relation of spring, summer, autumn and winter to season.

Generic relations are depicted by a fan or tree diagram without arrows (see Figure D.1).



Hình D.1 – Sơ đồ thể hiện mối quan hệ chung  
Figure D.1 – Graphical representation of a generic relation

### B.3.3 Mối quan hệ bộ phận

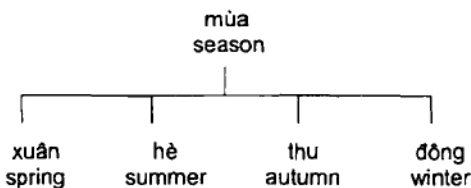
Các khái niệm phụ trong một hệ thống thứ bậc tạo thành các bộ phận cấu thành của khái niệm chính, ví dụ xuân, hạ, thu và đông có thể được xác định là bộ phận của khái niệm năm. Khi so sánh, sẽ không thích hợp nếu định nghĩa thời tiết nắng (một đặc trưng của mùa hè) là bộ phận của năm.

Mối quan hệ thành phần được mô tả bằng hình cái cào, không có mũi tên (xem Hình D.2). Các bộ phận số ít được mô tả bằng một đường thẳng, các bộ phận số nhiều được mô tả bằng hai đường thẳng.

### D.3.3 Partitive relations

Subordinate concepts within the hierarchy from constituent parts of the superordinate concept, e.g. spring, summer, autumn and winter may be defined as parts of the concept year. In comparison, it is inappropriate to define sunny weather (one possible characteristic of summer) as part of a year.

Partitive relations are depicted by a rake, without arrows (see Figure D.2). Singular parts are depicted by one line, multiple parts by double lines.



Hình D.2 – Sơ đồ thể hiện mối quan hệ bộ phận  
Figure D.2 – Graphical representation of a partitive relation

### D.3.4 Mối quan hệ liên kết

Mối quan hệ liên kết không thể đưa ra mô tả ngắn gọn như trong mối quan hệ chung và quan hệ bộ phận nhưng nó rất hữu ích cho việc xác định bản chất mối quan hệ giữa các khái niệm trong một hệ thống khái niệm, ví dụ: nguyên nhân và kết quả, hoạt động và vị trí, hoạt động và kết quả, công cụ và chức năng, vật liệu và sản phẩm.

### D.3.4 Associative relation

Associative relations cannot provide the economies in description that are present in generic and partitive relations but are helpful in identifying the nature of the relationship between one concept and another within a concept system, e.g. cause and effect, activity and location, activity and result, tool and function, material and product.



Mối quan hệ liên kết được mô tả bằng đường thẳng có mũi tên ở hai đầu (xem hình D.3).

Associative relations are depicted by a line with arrowheads at each end (see Figure D.3).



Hình D.3 – Sơ đồ thể hiện mối quan hệ liên kết  
Figure D.3 – Graphical representation of an associative relation

#### D.4 Sơ đồ khái niệm

Các hình từ B.1 đến B.5 thể hiện sơ đồ khái niệm là cơ sở của các định nghĩa trong điều 1 của tiêu chuẩn này. Hình B.6 là sơ đồ khái niệm bổ sung chỉ ra mối quan hệ giữa các thuật ngữ nhất định trước đó được đề cập trong các hình từ B.1 đến B.5. Các hình từ C.1 đến C.4 thể hiện sơ đồ khái niệm là cơ sở của các định nghĩa trong điều 2 của tiêu chuẩn này. Có nhiều thuật ngữ xuất hiện trong nhiều sơ đồ khái niệm được đưa ra, cung cấp mối liên kết giữa các sơ đồ. Điều này được chỉ ra như sau:

#### D.4 Concept diagrams

Figures B.1 to B.5 show the concept diagrams on which the definitions in Clause 1 of this part of ISO 3534 are based. Figure B.6 is an additional concept diagram that indicates the relationship of certain terms appearing previously in Figures B.1 to B.5. Figures C.1 to C.4 show the concept diagrams on which the definitions in Clause 2 of this part of ISO 3534 are based. There are several terms which appear in multiple concept diagrams, thus providing a linkage among the diagrams, are indicated. These are indicated as follows:

Hình B.1 Khái niệm cơ bản về tổng thể và mẫu: thống kê mô tả (1.5) mẫu ngẫu nhiên đơn giản (1.7) hàm ước lượng (1.12) thống kê kiểm nghiệm (1.52) biến ngẫu nhiên (2.10) hàm phân bố (2.7)	Hình B.5 Hình B.2 Hình B.3 Hình B.4 Hình C.1, C.2 Hình C.1
Hình B.2 Khái niệm liên quan đến mômen mẫu: mẫu ngẫu nhiên đơn giản (1.7)	Hình B.1
Hình B.3 Khái niệm ước lượng: hàm ước lượng (1.12) tham số (2.9) họ phân bố (2.8) hàm mật độ xác suất (2.26)	Hình B.1 Hình C.1 Hình B.4, C.1 Hình C.3

Figure B.1 Basic population and sample concepts: descriptive statistics (1.5) simple random sample (1.7) estimator (1.12) test statistic (1.52) random variable (2.10) distribution function (2.7)	Figure B.5 Figure B.2 Figure B.3 Figure B.4 Figure C.1, C.2 Figure C.1
Figure B.2 Concepts regarding sample moments: simple random sample (1.7)	Figure B.1
Figure B.3 Estimation concepts: estimator (1.12) parameter (2.9) family of distributions (2.8)	Figure B.1 Figure C.1 Figure B.4, C.1

hàm khối lượng xác suất (2.24)	Hình C.3	probability density function (2.26)	Figure C.3
Hình B.4 Khái niệm liên quan đến phép kiểm nghiệm thống kê:		probability mass function (2.24)	Figure C.3
thống kê kiểm nghiệm (1.52)	Hình B.1	Figure B.4 Concepts regarding statistical tests:	
hàm mật độ xác suất (2.26)	Hình B.3, C.3	test statistic (1.52)	Figure B.1
hàm khối lượng xác suất (2.24)	Hình B.3, C.3	probability density function (2.26)	Figure B.3, C.3
họ phân bố (2.8)	Hình B.3, C.1	probability mass function (2.24)	Figure B.3, C.3
Hình B.5 Khái niệm liên quan đến lớp và phân bố thực nghiệm:		family of distributions (2.8)	Figure B.3, C.1
thống kê mô tả (1.5)	Hình B.1	Figure B.5 Concepts regarding classes and empirical distributions:	
Hình B.6 Sơ đồ khái niệm suy luận thống kê:		descriptive statistics (1.5)	Figure B.1
tổng thể (1.1)	Hình B.1	Figure B.6 Statistical inference concept diagram:	
mẫu (1.3)	Hình B.1	population (1.1)	Figure B.1
giá trị quan trắc (1.4)	Hình B.1, B.5	sample (1.3)	Figure B.1
phép ước lượng (1.36)	Hình B.3	observed value (1.4)	Figure B.1, B.5
kiểm nghiệm thống kê (1.48)	Hình B.4	estimation (1.36)	Figure B.3
tham số (2.9)	Hình B.3, C.1	statistical test (1.48)	Figure B.4
biến ngẫu nhiên (2.10)	Hình B.1, C.1, C.2	parameter (2.9)	Figure B.3, C.1
Hình C.1 Khái niệm cơ bản dùng trong xác suất:		random variable (2.10)	Figure B.1, C.1, C.2
biến ngẫu nhiên (2.10)	Hình B.1, C.2	Figure C.1 Fundamental concepts in probability:	
phân bố xác suất (2.11)	Hình C.2, C.3	random variable (2.10)	Figure B.1, C.2
họ phân bố (2.8)	Hình B.3, B.4	probability distribution (2.11)	Figure C.2, C.3
hàm phân bố (2.7)	Hình B.1	family of distributions (2.8)	Figure B.3, B.4
tham số (2.9)	Hình B.3	distribution function (2.7)	Figure B.1
Hình C.2 Khái niệm mômen:		parameter (2.9)	Figure B.3
biến ngẫu nhiên (2.10)	Hình B.1, C.1	Figure C.2 Concepts on moments:	
phân bố xác suất (2.11)	Hình C.1, C.3	random variable (2.10)	Figure B.1, C.1
Hình C.3 Khái niệm phân bố xác suất:		probability distribution (2.11)	Figure C.1, C.3
phân bố xác suất (2.11)	Hình C.1, C.2	Figure C.3 Concepts on probability distributions:	
hàm khối lượng xác suất (2.24)	Hình B.3, B.4	probability distribution (2.11)	Figure C.1, C.2
phân bố liên tục (2.23)	Hình C.4	probability mass function (2.24)	Figure B.3, B.4
phân bố một chiều (2.16)	Hình C.4	continuous distribution (2.23)	Figure C.4
phân bố nhiều chiều (2.17)	Hình C.4	univariate distribution (2.16)	Figure C.4
Hình C.4 Khái niệm liên quan đến phân bố liên		multivariate distribution (2.17)	

tục:	
phân bố một chiều (2.16)	Hình C.3
phân bố nhiều chiều (2.17)	Hình C.3
phân bố liên tục (2.23)	Hình C.3

Figure C.4	
Figure C.4 Concepts regarding continuous distributions:	
univariate distribution (2.16)	Figure C.3
multivariate distribution (2.17)	Figure C.3
continuous distribution (2.23)	Figure C.3

Chú thích cuối của Hình C.4, các phân bố dưới đây là ví dụ của phân bố một chiều: phân bố chuẩn, phân bố  $t$ , phân bố  $F$ , phân bố chuẩn chuẩn hóa, phân bố gamma, beta, khi-bình phương, mũ, đều, cực trị loại I, cực trị loại II và cực trị loại III. Các phân bố dưới đây là ví dụ của phân bố nhiều chiều: phân bố chuẩn đa biến, phân bố chuẩn hai biến và phân bố chuẩn chuẩn hóa hai biến. Việc đưa thêm phân bố đơn biến (2.16) và phân bố đa biến (2.17) vào sơ đồ khái niệm có thể làm cho hình trở nên rắc rối quá mức.

As a final note on Figure C.4, the following distributions are examples of univariate distributions: normal,  $t$  distribution,  $F$  distribution, standardized normal, gamma, beta, chi-squared, exponential, uniform, Type I extreme value, Type II extreme value and Type III extreme value. The following distributions are examples of multivariate distributions: multivariate normal, bivariate normal and standardized bivariate normal. To include univariate distribution (2.16) and multivariate distribution (2.17) in the concept diagram would unduly clutter the figure.

**Thư mục tài liệu tham khảo**

**Bibliography**

- |  |  |
|--|--|
| <p>[1] TCVN 6398-11:199 , Đại lượng và đơn vị – Phần 11: Dấu và ký hiệu toán học dùng trong khoa học vật lý và công nghệ</p> <p>[2] TCVN 8244-1, Thống kê học – Từ vựng và ký hiệu – Phần 2: Thống kê ứng dụng</p> <p>[3] TCVN 6910 (tất cả các phần), Độ chính xác (độ đúng và độ chụm) của phương pháp đo và kết quả đo</p> <p>[4] TCVN 6165, Từ vựng quốc tế về đo lường học – Khái niệm, thuật ngữ chung và cơ bản (VIM)</p> | <p>[1] ISO 31-11:1992, Quantities and units – Part 11: Mathematical signs and symbols for use in the physical sciences and technology</p> <p>[2] ISO 3534-2, Statistics – Vocabulary and symbols – Part 2: Applied statistics</p> <p>[3] ISO 5725 (all parts), Accuracy (trueness and precision) of measurement methods and results</p> <p>[4] ISO/IEC Guide 99, International vocabulary of metrology - Basic and general concepts and associated terms (VIM)</p> |
|--|--|

## Chi mục theo bảng chữ cái

phân bố  $\chi^2$  2.57  
 $\sigma$ -đại số 2.69  
 $\sigma$ -trường 2.69

## B

bậc tự do 2.54  
 biến cố bù 2.3  
 biến cố độc lập 2.4  
 biến cố 2.2  
 biên giới lớp 1.56  
 biến ngẫu nhiên chuẩn hóa 2.33  
 biến ngẫu nhiên liên tục 2.29  
 biến ngẫu nhiên mẫu chuẩn hóa 1.19  
 biến ngẫu nhiên quy tâm 2.31  
 biến ngẫu nhiên rời rạc 2.28  
 biến ngẫu nhiên 2.10  
 biểu đồ cột 1.62  
 biểu đồ phân bố 1.61

## Đ

điểm giữa lớp 1.57  
 độ chệch 1.33  
 độ đo xác suất 2.70  
 độ lệch chuẩn 2.37  
 độ lệch chuẩn mẫu 1.17  
 độ rộng lớp 1.58  
 độ rộng mẫu 1.10  
 đối giá thuyết 1.42  
 đơn vị mẫu 1.2  
 đường hiệu lực 1.51  
 đường hồi quy 2.20

## G

giả thuyết đơn 1.43  
 giả thuyết hợp 1.44  
 giả thuyết không 1.41  
 giả thuyết 1.40  
 giá trị quan trắc 1.4  
 giới hạn dung sai thống kê 1.27  
 giới hạn lớp 1.56

## H

hàm hợp lý 1.38  
 hàm hợp lý biên duyên 1.39  
 hàm khối lượng xác suất 2.24  
 hàm mật độ xác suất 2.26  
 hàm phân bố của biến ngẫu nhiên  $X$  2.7

hàm ước lượng 1.12  
 hàm ước lượng hợp lý cực đại 1.35  
 hệ số bất đối xứng 2.39  
 hệ số bất đối xứng mẫu 1.20  
 hệ số biến động 2.38  
 hệ số biến động mẫu 1.18  
 hệ số nhon 2.40  
 hệ số nhon mẫu 1.21  
 hệ số tương quan 2.44  
 hệ số tương quan mẫu 1.23  
 hiệp phương sai 2.43  
 hiệp phương sai mẫu 1.22  
 hiệu lực của phép kiểm nghiệm 1.50  
 họ phân bố 2.8

## K

khoảng dự đoán 1.30  
 khoảng dung sai thống kê 1.26  
 khoảng tin cậy một phía 1.29  
 khoảng tin cậy 1.28  
 không gian mẫu 2.1  
 không gian xác suất 2.68  
 kiểm nghiệm mức ý nghĩa 1.48  
 kiểm nghiệm thống kê 1.48  
 kỳ vọng 2.12

## L

lớp 1.55  
 lớp 1.55.1, 1.55.2, 1.55.3

## M

mật hồi quy 2.21  
 mẫu ngẫu nhiên đơn giản 1.7  
 mẫu ngẫu nhiên 1.6  
 mẫu 1.3  
 mômen bậc  $r$  2.34  
 mômen bậc  $r=1$  2.35.1  
 mômen hỗn hợp bậc  $r$  và  $s$  2.41  
 mômen hỗn hợp quy tâm bậc  $r$  và  $s$  2.42  
 mômen mẫu bậc  $k$  1.14  
 mômen thứ  $r$  2.34  
 một của hàm khối lượng xác suất 2.25  
 một của hàm mật độ xác suất 2.27  
 mức ý nghĩa 1.45

## P

$p$ -giá trị 1.49  
 phân bố 2.11  
 phân bố beta 2.59  
 phân bố biên duyên 2.18  
 phân bố chuẩn 2.50  
 phân bố chuẩn chuẩn hóa hai chiều 2.66  
 phân bố chuẩn chuẩn hóa 2.51  
 phân bố chuẩn đa biến 2.64  
 phân bố chuẩn hai biến 2.65  
 phân bố có điều kiện 2.19  
 phân bố cực trị loại I 2.61  
 phân bố cực trị loại II 2.62  
 phân bố cực trị loại III 2.63  
 phân bố đa thức 2.45  
 phân bố đều 2.60  
 phân bố  $F$  2.55  
 phân bố Fréchet 2.62  
 phân bố gamma 2.56  
 phân bố Gausslan 2.50  
 phân bố Gaussian chuẩn hóa 2.51  
 phân bố Gumbel 2.61  
 phân bố hình chữ nhật 2.60  
 phân bố khí-bình phương 2.57  
 phân bố liên tục 2.23  
 phân bố lôga chuẩn 2.52  
 phân bố mẫu 2.67  
 phân bố một chiều 2.16  
 phân bố mũ 2.58  
 phân bố nhị thức âm 2.49  
 phân bố nhị thức 2.46  
 phân bố nhiều chiều 2.17  
 phân bố Poisson 2.47  
 phân bố rời rạc 2.22  
 phân bố siêu hình học 2.48  
 phân bố Student 2.53  
 phân bố  $t$  2.53  
 phân bố tần số 1.60  
 phân bố Weibull 2.63  
 phân bố xác suất biên duyên 2.18  
 phân bố xác suất chuẩn hóa 2.32  
 phân bố xác suất có điều kiện 2.19  
 phân bố xác suất đa biến 2.17  
 phân bố xác suất đơn biến 2.16  
 phân bố xác suất liên tục 2.23  
 phân bố xác suất quy tâm 2.30  
 phân bố xác suất rời rạc 2.22  
 phân bố xác suất 2.11  
 phân vị 2.15  
 phép ước lượng 1.36  
 phép ước lượng hợp lý cực đại 1.37

# TCVN 8244-1:2010

phương sai mẫu 1.16  
phương sai 2.36  
 $\rho$ -phân vị 2.13

## S

sai lầm loại I 1.46  
sai lầm loại II 1.47  
sai số của phép ước lượng 1.32  
sai số tiêu chuẩn 1.24  
sigma đại số của biến cố 2.69  
sigma trường 2.69

## T

tâm độ rộng 1.11  
tần số 1.59  
tần số tích lũy 1.63  
tần số tích lũy tương đối 1.65  
tần số tương đối 1.64  
tham số 2.9  
thống kê kiểm nghiệm 1.52  
thống kê mô tả 1.5  
thống kê mô tả dạng số 1.54  
thống kê mô tả đồ thị 1.53  
thống kê thứ tự 1.9  
thống kê 1.8  
tổng thể 1.1  
trung bình 1.15, 2.35.1, 2.35.2  
trung bình mẫu 1.15  
trung bình số học 1.15  
trung bình 1.15  
trung vị 2.14  
trung vị mẫu 1.13

## U

ước lượng 1.31  
ước lượng khoảng 1.25  
ước lượng không chệch 1.34

## X

xác suất có điều kiện 2.6  
xác suất của biến cố  $A$  2.5

## Alphabetical index

- $\chi^2$  distribution 2.57  
 $\sigma$ -algebra 2.69  
 $\sigma$ -field 2.69
- A**
- alternative hypothesis 1.42  
arithmetic mean 1.15  
average 1.15
- B**
- bar chart 1.62  
beta distribution 2.59  
bias 1.33  
binomial distribution 2.46  
bivariate normal distribution 2.65
- C**
- centred probability distribution 2.30  
centred random variable 2.31  
chi-squared distribution 2.57  
class 1.55.1, 1.55.2, 1.55.3  
class boundaries 1.56  
class limits 1.56  
class width 1.58  
classes 1.55  
coefficient of kurtosis 2.40  
coefficient of skewness 2.39  
coefficient of variation 2.38  
complementary event 2.3  
composite hypothesis 1.44  
conditional distribution 2.19  
conditional probability 2.6  
conditional probability distribution 2.19  
confidence interval 1.28  
continuous distribution 2.23  
continuous probability distribution 2.23  
continuous random variable 2.29  
correlation coefficient 2.44  
covariance 2.43  
cumulative frequency 1.63  
cumulative relative frequency 1.65
- D**
- degrees of freedom 2.54  
descriptive statistics 1.5  
discrete distribution 2.22  
discrete probability distribution 2.22  
discrete random variable 2.28  
distribution 2.11  
distribution function of a random variable  $X$  2.7
- E**
- error of estimation 1.32  
estimate 1.31  
estimation 1.36  
estimator 1.12  
event 2.2  
expectation 2.12  
exponential distribution 2.58
- F**
- $F$  distribution 2.55  
family of distributions 2.8  
Fréchet distribution 2.62  
frequency 1.59  
frequency distribution 1.60
- G**
- gamma distribution 2.56  
Gaussian distribution 2.50  
graphical descriptive statistics 1.53  
Gumbel distribution 2.61
- H**
- histogram 1.61  
hypergeometric distribution 2.48  
hypothesis 1.40
- I**
- independent events 2.4  
interval estimator 1.25
- J**
- joint central moment of orders  $r$  and  $s$  2.42  
joint moment of orders  $r$  and  $s$  2.41
- L**
- likelihood function 1.38  
lognormal distribution 2.52
- M**
- marginal distribution 2.18  
marginal probability distribution 2.18  
maximum likelihood estimation 1.37  
maximum likelihood estimator 1.35  
mean 1.15, 2.35.1, 2.35.2  
median 2.14  
mid-point of class 1.57  
mid-range 1.11  
mode of probability density function 2.27  
mode of probability mass function 2.25  
moment of order  $r$  2.34  
moment of order  $r = 1$  2.35.1  
multinomial distribution 2.45  
multivariate distribution 2.17  
multivariate normal distribution 2.64  
multivariate probability distribution 2.17
- N**
- negative binomial distribution 2.49  
normal distribution 2.50  
null hypothesis 1.41  
numerical descriptive statistics 1.54
- O**
- observed value 1.4  
one-sided confidence interval 1.29  
order statistic 1.9

<b>P</b>		sigma field 2.69
parameter 2.9		significance level 1.45
<i>p</i> -fractile 2.13		significance test 1.48
Poisson distribution 2.47		simple hypothesis 1.43
population 1.1		simple random sample 1.7
power curve 1.51		standard deviation 2.37
power of a test 1.50		standard error 1.24
<i>p</i> -quantile 2.13		standardized bivariate normal distribution 2.66
prediction interval 1.30		standardized Gaussian distribution 2.51
probability density function 2.26		standardized normal distribution 2.51
probability distribution 2.11		standardized probability distribution 2.32
probability mass function 2.24		standardized random variable 2.33
probability measure 2.70		standardized sample random variable 1.19
probability of an event <i>A</i> 2.5		statistic 1.8
probability space 2.68		statistical test 1.48
profile likelihood function 1.39		statistical tolerance interval 1.26
<i>p</i> -value 1.49		statistical tolerance limit 1.27
<b>Q</b>		Student's distribution 2.53
quartile 2.15		
<b>R</b>		<b>T</b>
random sample 1.6		<i>t</i> distribution 2.53
random variable 2.10		test statistic 1.52
rectangular distribution 2.60		Type I error 1.46
regression curve 2.20		type I extreme value distribution 2.61
regression surface 2.21		Type II error 1.47
relative frequency 1.64		type II extreme value distribution 2.62
<i>r</i> th moment 2.34		type III extreme value distribution 2.63
<b>S</b>		<b>U</b>
sample 1.3		unbiased estimator 1.34
sample coefficient of kurtosis 1.21		uniform distribution 2.60
sample coefficient of skewness 1.20		univariate distribution 2.16
sample coefficient of variation 1.18		univariate probability distribution 2.16
sample correlation coefficient 1.23		
sample covariance 1.22		<b>V</b>
sample mean 1.15		variance 2.36
sample median 1.13		
sample moment of order <i>k</i> 1.14		<b>W</b>
sample range 1.10		Weibull distribution 2.63
sample space 2.1		
sample standard deviation 1.17		
sample variance 1.16		
sampling distribution 2.67		
sampling unit 1.2		
sigma algebra of events 2.69		