



British Standard

Statistical terminology

Part 1. Glossary of terms relating to probability and general terms relating to statistics

Terminologie statistique

Partie 1. Glossaire des termes relatifs à la probabilité et des termes généraux relatifs à la statistique

Begriffe in der Statistik

Teil 1. Begriffe der Wahrscheinlichkeit und allgemeine statistische Begriffe

Contents

	Page		Page
Foreword	Inside front cover	3. General terms relating to methods of sampling	29
Cooperating organizations	Back cover	4. Terms relating to sampling inspection	32
		5. Symbols	39
Scope and field of application	1	English alphabetical index	42
1. Terms used in theory of probability	2	French alphabetical index	45
2. General statistical terms	12		

Foreword

This British Standard is identical with ISO 3534 'Statistics — Vocabulary and symbols'.

The terminology relating to statistical methods has, over the years, presented a barrier to uniformity in the application of principles in this field. A major unification was carried out by M G Kendall and W R Buckland in their 'Dictionary of Statistical Terms', published by the International Statistical Institute, and this is still the authoritative reference over the broad field of statistical terminology.

On a more limited scale, pertaining to standardization, two series of terms were published by the International Organization for Standardization (ISO), under the authority of Technical Committee 69 'Application of Statistical Methods'; the first of these, ISO/R 645, appeared in 1967, and the second, ISO/R 1786, in 1970. In 1975 a Draft International Standard, 3534 'Statistics — Vocabulary', was circulated to all member bodies and at the same time a Draft International Standard 3725 on statistical symbols used in ISO/DIS 3534 was made available. Both these Draft International Standards were also made available for public comment in the UK and were finally combined and published in 1977 as ISO 3534 'Statistics — Vocabulary and symbols'.

Members of the United Kingdom Technical and Advisory Committees felt that a British Standard based on ISO 3534 should be published. It is hoped that Technical Committees having narrower statistical interests will accept, as a basis for statistical methodology, the terms and definitions which follow and will coordinate their more specific language around this work.

It was felt in the UK that this British Standard should be identical to the International Standard (with a dual BS/ISO number) in order to establish international uniformity. It is hoped that the retention of both the English and French texts will help to improve understanding in international standardization.

Attention is drawn to the following.

1. Cross references. The terms are segregated into four categories and where a term might logically be placed in two of these this has been done; the emphasis in each definition depending on the particular category. For example, the term 'median' appears in the section headed 'Terms used in theory of probability' and 'General statistical terms': in the former a probability definition is given whereas a definition applicable to observations or samples is given in the latter. The following terms are those that are defined twice.

Term	Item numbers
Bivariate distribution	1.6 2.22
Multivariate distribution	1.7 2.23
Marginal distribution	1.8 2.26
Conditional distribution	1.9 2.27

Term	Item numbers
Median	1.12 2.30
Mode	1.13 2.31
Moment of order q about the origin	1.20 2.36
Moment of order q about an origin a	1.21 2.37
Centred moment of order q	1.22 2.38
Joint moment of orders q and s about the origin	1.23 2.39
Joint moment of orders q and s about an origin (a, b)	1.24 2.40
Joint centred moment of orders q and s	1.25 2.41
Covariance	1.26 2.42
Coefficient of correlation	1.27 2.43
Regression curve	1.28 2.44
Regression surface	1.29 2.45

2. Acceptable Quality Level (AQL) — Item 4.51. It is strongly recommended that the utilization of an AQL in standards on sampling be restricted to those cases where inspection is of a continuing series of lots. Such a series is one of which the switching rules of BS 6001 have a high probability of coming into play. As such the definition of AQL as given in the second paragraph only of 4.51 should be used, i.e. that relating AQL to a maximum of the process average.

3. Strength of test. Items 2.66 to 2.74 are concerned with various aspects of statistical tests of hypotheses. However, the term 'strength of test' has not been included and as it is felt to be useful the following interpretation of the term is recommended:

'strength of test: the joint relationship between the α and β errors.'

4. Symbols. The use of an upright upper case N is strongly recommended when referring to the Normal distribution. Also, the italic upper case E for expected value is easily confused with the upright upper case E for 'Event'. If there is any possibility of ambiguity, the script $\mathcal{E}$ should be used for the operator 'expected value'.

Terminology and conventions. The text of the international standard has been accepted as suitable for publication, without deviation, as a British Standard. Certain terminology and conventions are used, however, that are not identical with those used in British Standards. Attention is therefore drawn to the following.

Wherever the words 'International Standard' relating to this publication appear, they should be interpreted as 'British Standard'.

The comma has been used as a decimal marker. In British Standards it is current practice to use a full point (a full stop on the baseline) as the decimal marker.

British Standard

Statistical terminology

Part 1. Glossary of terms relating to probability and general terms relating to statistics

SCOPE AND FIELD OF APPLICATION

This International Standard gives definitions, in English and French, of statistical terms 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 under the following main headings :

- 1 Terms used in theory of probability
- 2 General statistical terms
- 3 General terms relating to methods of sampling
- 4 Terms relating to sampling inspection

OBJET ET DOMAINE D'APPLICATION

La présente Norme internationale donne les définitions, en anglais et en français, de termes statistiques susceptibles d'être utilisés dans la rédaction d'autres Normes internationales. En outre, elle définit un ensemble de symboles pour un nombre limité de ces termes.

Les termes sont classés sous les principales rubriques suivantes :

- 1 Termes utilisés en calcul des probabilités
- 2 Termes statistiques généraux
- 3 Termes généraux relatifs aux méthodes d'échantillonnage
- 4 Termes relatifs au contrôle par échantillonnage.

1 TERMS USED IN THEORY OF PROBABILITY

A great number of terms are given both in this clause and in the following clause ("General statistical terms"). Nevertheless, it seems useful to define them separately to draw attention of the reader to the fact that :

- a) the terms in their probabilistic sense apply to principles, independent of any practical application;
- b) the terms in their statistical sense apply to the observations to which they relate : these definitions are of a specifically operational character.

NOTE ON THE NOTION OF PROBABILITY

The concept of probability may be introduced in either of two forms, depending on whether it is intended to designate a degree of belief or whether it is considered as the limit value of a frequency. In both cases, its introduction requires that some precautions be taken which cannot be developed within the context of an International Standard and for which users should refer to specialized literature.

For practical reasons, however, it may be considered that, whenever the conditions of a test can be reproduced, the probability $Pr(E)$ of an event E occurring is the value around which the occurrence frequency of the latter oscillates and towards which it tends when the number of tests is indefinitely increased.

1.1 random variable; variate : A variable which may take any of the values of a specified set of values and with which is associated a probability distribution (see 1.2).

A random variable which may take only isolated values is said to be "discrete". A random variable which may take all the values of a finite or infinite interval is said to be "continuous".

1.2 probability distribution of a random variable : A function which determines the probability that a random variable takes any given value or belongs to a given set of values.

The probability on the whole interval of variation of the variate equals 1.

1.3 distribution function : A function giving, for every value x , the probability that the random variable X be less than or equal to x :

$$F(x) = Pr [X \leq x]$$

1.4 probability density function for a continuous random variable : The derivative (when it exists) of the distribution function.

$$f(x) = \frac{dF(x)}{dx}$$

NOTE — $f(x) dx$ is the "probability element".

$$f(x) dx = Pr [x < X < x + dx]$$

1 TERMES UTILISÉS EN CALCUL DES PROBABILITÉS

De nombreux termes figurent à la fois dans ce chapitre et dans le chapitre suivant («Termes statistiques généraux»). Il a cependant paru utile de les définir séparément pour attirer l'attention du lecteur sur le fait :

- a) que les termes probabilistes s'appliquent à des concepts, indépendamment de toute réalisation;
- b) que les termes statistiques s'appliquent à des observations et aux calculs qui y sont relatifs : ces termes ont un caractère opérationnel que précise la définition.

NOTE SUR LA NOTION DE PROBABILITÉ

La notion de probabilité peut être introduite sous deux formes, soit qu'on veuille s'en servir pour désigner un degré de croyance, soit qu'on la considère comme la valeur limite d'une fréquence. Dans les deux cas, son introduction nécessite certaines précautions qui ne peuvent être développées dans le cadre d'une Norme Internationale et pour lesquelles il convient de se référer aux ouvrages spécialisés.

Pour les besoins de la pratique, on peut toutefois considérer que, dans la mesure où l'on est en état de reproduire les conditions d'un essai, la probabilité $Pr(E)$ d'un événement E susceptible de se produire est la valeur autour de laquelle sa fréquence d'apparition oscille et vers laquelle elle tend, lorsque le nombre d'essais augmente indéfiniment.

1.1 variable aléatoire : Variable pouvant prendre n'importe quelle valeur d'un ensemble déterminé de valeurs, et à laquelle est associée une loi de probabilité (voir 1.2).

Une variable aléatoire qui ne peut prendre que des valeurs isolées est dite «discrète». Une variable aléatoire qui peut prendre toutes les valeurs d'un intervalle fini ou infini est dite «continue».

1.2 loi de probabilité d'une variable aléatoire : Fonction déterminant la probabilité qu'une variable aléatoire prenne une valeur donnée quelconque ou appartienne à un ensemble donné de valeurs.

La probabilité sur tout le domaine de variation de la variable est égale à l'unité.

1.3 fonction de répartition : Fonction donnant, pour toute valeur x , la probabilité que la variable aléatoire X soit inférieure ou égale à x :

$$F(x) = Pr [X \leq x]$$

1.4 densité de probabilité pour une variable aléatoire continue : Dérivée (lorsqu'elle existe) de la fonction de répartition.

$$f(x) = \frac{dF(x)}{dx}$$

NOTE — $f(x) dx$ s'appelle la «probabilité élémentaire».

$$f(x) dx = Pr [x < X < x + dx]$$

1.5 probability for a discrete random variable : x_i being one of the values which can be taken by the discrete random variable X , the probability p_i is :

$$p_i = Pr [X = x_i]$$

1.6 bivariate distribution : A distribution which determines the probability that a pair of variates takes any given values or belongs to a given set of values.

1.7 multivariate distribution : A distribution which determines the probability that several variates considered simultaneously take any given values or belong to a given set of values.

1.8 marginal distribution : In the case of a probability distribution with k variates, a distribution of a subset of p variates, the other $(k-p)$ variates taking any values in their interval of variation.

Example : In a probability distribution with three variates, X, Y, Z , there are :

- three bivariate marginal distributions : distribution of the pairs $(X, Y), (X, Z), (Y, Z)$;
- three univariate marginal distributions : distributions of X, Y and Z .

1.9 conditional distribution : In the case of a probability distribution with k variates, a distribution of a subset of p variates, when the other $(k-p)$ variates have fixed values.

Example : In a probability distribution with two variates X, Y , we have :

- conditional distribution of X . A specific distribution is expressed as "distribution of X for $Y = y$ ";
- conditional distribution of Y . A specific distribution is expressed as "distribution of Y for $X = x$ ".

1.10 correlation : The inter-dependent relationship between two or several variates when such a relationship includes a random part.

NOTE — In this case, it is said that a "stochastic link" exists between the variates.

1.11 fractile (or quantile) of a probability distribution : If p is a number between 0 and 1, the p -fractile is the value of the random variable for which the distribution function equals p or "jumps" from a value less than or equal to p to a value greater than p .

It is possible that the distribution function equals p throughout the interval between consecutive possible values of the variate. In this case, any value in this interval may be considered as the p -fractile.

1.5 probabilité pour une variable aléatoire discrète : x_i étant l'une des valeurs pouvant être prises par la variable aléatoire discrète X , la probabilité p_i est telle que

$$p_i = Pr [X = x_i]$$

1.6 loi de probabilité à deux variables : Loi déterminant la probabilité qu'un couple de variables aléatoires prenne des valeurs données quelconques ou appartienne à un ensemble donné de valeurs.

1.7 loi de probabilité à plusieurs variables : Loi déterminant la probabilité que plusieurs variables aléatoires considérées simultanément prennent des valeurs données quelconques ou appartiennent à un ensemble donné de valeurs.

1.8 loi marginale : Dans une loi de probabilité à k variables, loi d'un sous-ensemble de p variables, les $(k-p)$ autres variables pouvant prendre des valeurs quelconques dans leur domaine de variation.

Exemple : Dans une loi de probabilité à trois variables, X, Y, Z , on distingue

- trois lois marginales à deux variables : loi des couples $(X, Y), (X, Z), (Y, Z)$;
- trois lois marginales à une variable : lois de X , de Y et de Z .

1.9 loi conditionnelle : Dans une loi de probabilité à k variables, loi d'un sous-ensemble de p variables, lorsque les $(k-p)$ autres variables ont des valeurs données.

Exemple : Dans une loi de probabilité à deux variables X, Y , on distingue

- les lois conditionnelles de X . Une loi particulière s'énonce «loi de X pour $Y = y$ »;
- les lois conditionnelles de Y . Une loi particulière s'énonce «loi de Y pour $X = x$ ».

1.10 corrélation : Liaison mutuelle entre deux ou plusieurs variables lorsque cette liaison comporte une partie aléatoire.

NOTE — Dans ce cas, on dit qu'il existe une «liaison stochastique» entre les variables.

1.11 fractile (ou quantile) d'une loi de probabilité : p étant un nombre compris entre 0 et 1, le fractile d'ordre p est la valeur de la variable aléatoire pour laquelle la fonction de répartition prend la valeur p ou «saute» d'une valeur inférieure ou égale à p à une valeur supérieure à p .

Il se peut que la fonction de répartition soit égale à p dans tout l'intervalle entre deux valeurs possibles consécutives de la variable. Dans ce cas, toute valeur de cet intervalle peut être considérée comme fractile d'ordre p .

BS 5532 : Part 1 : 1978

1.12 median: The 0,50 fractile of the probability distribution.

1.13 mode: The value(s) of a random variable such that the probability (discrete variate) or the density (continuous variate) has a maximum for this value (or these values).

If there is one mode, the probability distribution of the variate is said to be "unimodal"; in the other case, it is said to be "multimodal" (bimodal if there are two modes).

1.14 expectation (mean) of a random variable :

a) For a discrete random variable X taking the values x_i with the probabilities p_i , the expectation is defined by

$$E(X) = \sum p_i x_i$$

the sum being extended for all the values x_i which can be taken by X .

b) For a continuous random variable X having the density $f(x)$, the expectation is defined by

$$E(X) = \int x f(x) dx$$

the integral being extended for all values of the interval of variation of X .

No distinction is made between the expectation of a random variable and that of a probability distribution.

NOTES

1 marginal expectation: The expectation of the marginal distribution of a random variable.

2 conditional expectation: The expectation over a conditional distribution of a random variable.

1.15 centred random variable: A random variable the expectation of which equals zero.

NOTE — If the random variable X has an expectation equal to $E(X)$, the corresponding centred variate is $X - E(X)$.

1.16 variance (of a random variable, or of a probability distribution): The expectation of the square of the centred variate. (See 1.22.)

1.17 standard deviation (of a random variable, or of a probability distribution): The positive square root of the variance.

1.18 coefficient of variation (of a random variable, or of a probability distribution): The ratio of the standard deviation to the absolute value of the expectation.

1.12 médiane: Fractile d'ordre $p = 0,50$ de la loi de probabilité.

1.13 mode: Valeur(s) d'une variable aléatoire telle que la probabilité (variable discrète) ou la densité (variable continue) ait un maximum pour cette valeur (ou ces valeurs).

S'il y a un seul mode, la loi de probabilité de la variable est dite «unimodale»; dans le cas contraire, elle est dite «multimodale» (bimodale s'il y a deux modes).

1.14 espérance mathématique (moyenne) d'une variable aléatoire :

a) Pour une variable aléatoire discrète X prenant les valeurs x_i avec les probabilités p_i , l'espérance mathématique est définie par

$$E(X) = \sum p_i x_i$$

la sommation étant étendue à toutes les valeurs x_i susceptibles d'être prises par X .

b) Pour une variable aléatoire continue X de densité $f(x)$, l'espérance mathématique est définie par

$$E(X) = \int x f(x) dx$$

l'intégrale étant étendue à tout le domaine de variation de X .

On dit indifféremment espérance mathématique d'une variable aléatoire ou d'une loi de probabilité.

NOTES

1 espérance mathématique marginale: Espérance mathématique de la loi marginale d'une variable aléatoire.

2 espérance mathématique conditionnelle: Espérance mathématique de la loi conditionnelle d'une variable aléatoire.

1.15 variable aléatoire centrée: Variable aléatoire dont l'espérance mathématique est égale à zéro.

NOTE — Si la variable aléatoire X a pour espérance mathématique $E(X)$, la variable centrée correspondante est $X - E(X)$.

1.16 variance (d'une variable aléatoire ou d'une loi de probabilité): Espérance mathématique du carré de la variable centrée. (Voir 1.22.)

1.17 écart-type (d'une variable aléatoire ou d'une loi de probabilité): Racine carrée positive de la variance.

1.18 coefficient de variation (d'une variable aléatoire ou d'une loi de probabilité): Rapport de l'écart-type à la valeur absolue de l'espérance mathématique.

1.19 standardized variate: A random variable the expectation of which equals zero and the standard deviation of which equals 1.

NOTES

1 If the random variable X has an expectation equal to $E(X)$ and a standard deviation equal to σ , the corresponding standardized variate is the variate

$$\frac{X - E(X)}{\sigma}$$

The distribution of the standardized variate is called "standardized distribution".

2 The concept of standardized variate can be generalized to that of a "reduced variate" which is defined using another location and another scale parameter.

1.20 moment of order q about the origin: * In a univariate distribution, the expectation of the q th power of the variate:

$$E[X^q]$$

NOTE — The moment of order 1 is the expectation of the variate X .

1.21 moment of order q about an origin a : * In a univariate distribution, the expectation of the q th power of the variate $(X - a)$:

$$E[(X - a)^q]$$

1.22 centred moment of order q : * In a univariate distribution, the expectation of the q th power of the central variate $[X - E(X)]$:

$$E\{[X - E(X)]^q\}$$

NOTE — The centred moment of order 2 is the variance of variate X .

1.23 joint moment of orders q and s about the origin: * In a bivariate distribution, the expectation of the product of the q th power of variate X by the s th power of variate Y :

$$E[X^q Y^s]$$

NOTE — The moment of orders 1 and 0 is the expectation of the marginal distribution of X .

The moment of orders 0 and 1 is the expectation of the marginal distribution of Y .

1.24 joint moment of orders q and s about an origin (a, b) : * In a bivariate distribution, the expectation of the product of the q th power of the variate $(X - a)$ by the s th power of the variate $(Y - b)$:

$$E[(X - a)^q (Y - b)^s]$$

* If, in the definition of the moments, the quantities X , $X - a$, Y , $Y - b$, etc., are replaced by their absolute values, i.e. $|X|$, $|X - a|$, $|Y|$, $|Y - b|$, etc., other moments called absolute moments are defined.

1.19 variable aléatoire réduite: Variable aléatoire dont l'espérance mathématique est égale à zéro et dont l'écart-type est égal à l'unité.

NOTES

1 Si la variable aléatoire X a pour espérance mathématique $E(X)$ et pour écart-type σ , la variable réduite correspondante est la variable

$$\frac{X - E(X)}{\sigma}$$

La loi de la variable réduite est appelée «loi réduite».

2 Le terme variable aléatoire réduite peut désigner également une variable transformée de la variable en utilisant une autre valeur centrale et un autre paramètre d'échelle.

1.20 moment d'ordre q par rapport à l'origine: * Dans une loi de probabilité à une variable, espérance mathématique de la $q^{\text{ème}}$ puissance de la variable:

$$E[X^q]$$

NOTE — Le moment d'ordre 1 est l'espérance mathématique de la variable aléatoire X .

1.21 moment d'ordre q par rapport à une origine a : * Dans une loi de probabilité à une variable, espérance mathématique de la $q^{\text{ème}}$ puissance de la variable $(X - a)$:

$$E[(X - a)^q]$$

1.22 moment centré d'ordre q : * Dans une loi de probabilité à une variable, espérance mathématique de la $q^{\text{ème}}$ puissance de la variable centrée $[X - E(X)]$:

$$E\{[X - E(X)]^q\}$$

NOTE — Le moment centré d'ordre 2 est la variance de la variable X .

1.23 moment d'ordres q et s par rapport à l'origine: * Dans une loi de probabilité à deux variables, espérance mathématique du produit de la $q^{\text{ème}}$ puissance de la variable X par la $s^{\text{ème}}$ puissance de la variable Y :

$$E[X^q Y^s]$$

NOTE — Le moment d'ordres 1 et 0 est l'espérance mathématique de la loi marginale de X .

Le moment d'ordres 0 et 1 est l'espérance mathématique de la loi marginale de Y .

1.24 moment d'ordres q et s par rapport à une origine (a, b) : * Dans une loi de probabilité à deux variables, espérance mathématique du produit de la $q^{\text{ème}}$ puissance de la variable $(X - a)$ par la $s^{\text{ème}}$ puissance de la variable $(Y - b)$:

$$E[(X - a)^q (Y - b)^s]$$

* Si, dans la définition des moments, on remplace les quantités X , $X - a$, Y , $Y - b$, etc. par leurs valeurs absolues, c'est-à-dire $|X|$, $|X - a|$, $|Y|$, $|Y - b|$, etc., on définit d'autres moments appelés moments absolus.

1.25 joint centred moment of orders q and s :* In a bivariate distribution, the expectation of the product of the q th power of the centred variate $[X - E(X)]$ by the s th power of the centred variate $[Y - E(Y)]$:

$$E \{ [X - E(X)]^q [Y - E(Y)]^s \}$$

NOTE — The joint centred moment of orders 2 and 0 is the variance of the marginal distribution of X .

The joint centred moment of orders 0 and 2 is the variance of the marginal distribution of Y .

1.26 covariance : The joint centred moment of orders 1 and 1.

$$E \{ [X - E(X)] [Y - E(Y)] \}$$

1.27 coefficient of correlation : The ratio of the covariance of two variates to the product of their standard deviations.

1.28 regression curve : In the case of two variates, the curve defined by the expectation of Y for $X = x$, as x varies, is called the "regression curve of Y on X ".

When the regression curve of Y on X is a straight line, the regression is called "linear".

In this case, the coefficient of linear regression of Y on X is the coefficient of x (slope) in the equation of the regression line.

1.29 regression surface : In the case of three variates, the surface defined by the expectation of Z for $X = x$ and $Y = y$, as x and y vary, is called "the regression surface of Z on X and Y ".

When the regression surface is a plane surface, the regression is called "linear".

In this case, the coefficient of linear regression of Z on X is the coefficient of x in the equation of the regression plane surface.

NOTE — The above definitions may be extended to more than three variates.

1.30 uniform distribution; rectangular distribution : The probability distribution of a continuous variate for which the probability density function is constant within a finite interval and zero outside this interval.

1.25 moment centré d'ordres q et s :* Dans une loi de probabilité à deux variables, espérance mathématique du produit de la $q^{\text{ème}}$ puissance de la variable centrée $[X - E(X)]$ par la $s^{\text{ème}}$ puissance de la variable centrée $[Y - E(Y)]$:

$$E \{ [X - E(X)]^q [Y - E(Y)]^s \}$$

NOTE — Le moment centré d'ordres 2 et 0 est la variance de la loi marginale de X .

Le moment centré d'ordres 0 et 2 est la variance de la loi marginale de Y .

1.26 covariance : Moment centré d'ordres 1 et 1.

$$E \{ [X - E(X)] [Y - E(Y)] \}$$

1.27 coefficient de corrélation : Quotient de la covariance de deux variables par le produit de leurs écarts-types.

1.28 courbe de régression : Dans le cas de deux variables, la courbe définie par l'espérance mathématique de Y pour $X = x$, lorsque x varie, s'appelle «courbe de régression de Y en X ».

Lorsque la courbe de régression de Y en X est une droite, la régression est dite «linéaire».

Dans ce cas, le coefficient de régression linéaire de Y en X est le coefficient de x (pente) dans l'équation de la droite de régression.

1.29 surface de régression : Dans le cas de trois variables, la surface définie par l'espérance mathématique de Z pour $X = x$ et $Y = y$, lorsque x et y varient, s'appelle «surface de régression de Z en X et Y ».

Lorsque la surface de régression est un plan, la régression est dite «linéaire».

Dans ce cas, le coefficient de régression linéaire partielle de Z par rapport à X est le coefficient de x dans l'équation du plan de régression.

NOTE — Les définitions ci-dessus s'étendent à plus de trois variables.

1.30 loi uniforme; loi rectangulaire : Loi de probabilité d'une variable aléatoire continue, telle que la densité de probabilité est constante dans un intervalle fini et nulle hors de cet intervalle.

* See foot-note on page 5.

* Voir note du bas de la page 5.

1.31 normal distribution; Laplace-Gauss distribution : The probability distribution of a continuous random variable X such that, if x is any real number, the probability density is

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{x-m}{\sigma} \right)^2 \right]$$

$$-\infty < x < +\infty$$

NOTE — m is the expectation and σ is the standard deviation of the normal distribution.

1.32 standardized normal distribution : The probability distribution of the standardized normal variate.

For a normal variate X having m and σ as parameters, the standardized random variable is

$$U = \frac{X-m}{\sigma}$$

and the probability density of the U distributed variate is

$$f(u) = \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{u^2}{2} \right)$$

$$-\infty < u < +\infty$$

1.33 chi-squared distribution : The distribution of the sum of the squares of independent standardized normal variates.

The number of these variates is the number ν of degrees of freedom of the χ^2 distributed variate, a parameter of the distribution.

The probability density function of the χ^2 distributed variate is

$$f(\chi^2, \nu) = \frac{(\chi^2)^{(\nu/2)-1}}{2^{\nu/2} \Gamma(\frac{\nu}{2})} \exp \left(-\frac{\chi^2}{2} \right)$$

$$\chi^2 \geq 0$$

NOTE — The probability distribution of the variate $\chi^2/2$ is a gamma distribution of parameter $m = \nu/2$. (See 1.38.)

1.34 t-distribution; Student distribution : The distribution of a quotient of independent random variables, the numerator of which is a standardized normal variate, and the denominator of which is the positive square root of the quotient of a χ^2 distributed variate and its number of degrees of freedom.

The number of degrees of freedom of χ^2 is the number ν of degrees of freedom of the t -distributed variate.

The probability density function of the t -distributed variate is

$$f(t, \nu) = \frac{1}{\sqrt{\pi} \nu} \frac{\Gamma[(\nu+1)/2]}{\Gamma(\nu/2)} \frac{1}{(1+t^2/\nu)^{(\nu+1)/2}}$$

$$-\infty < t < +\infty$$

1.31 loi normale; loi de Laplace-Gauss : Loi de probabilité d'une variable aléatoire continue X , telle que, si x est un nombre réel quelconque, la densité de probabilité est

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{x-m}{\sigma} \right)^2 \right]$$

$$-\infty < x < +\infty$$

NOTE — m est l'espérance mathématique et σ est l'écart-type de la loi normale.

1.32 loi normale réduite : Loi de probabilité de la variable normale réduite.

Pour une variable normale X de paramètres m et σ , la variable aléatoire réduite est

$$U = \frac{X-m}{\sigma}$$

et la densité de probabilité de la variable U est

$$f(u) = \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{u^2}{2} \right)$$

$$-\infty < u < +\infty$$

1.33 loi de χ^2 : Loi de la somme des carrés de variables normales réduites indépendantes.

Le nombre de ces variables est le nombre ν de degrés de liberté de la variable χ^2 , paramètre de cette loi.

La densité de probabilité de la variable χ^2 est

$$f(\chi^2, \nu) = \frac{(\chi^2)^{(\nu/2)-1}}{2^{\nu/2} \Gamma(\frac{\nu}{2})} \exp \left(-\frac{\chi^2}{2} \right)$$

$$\chi^2 \geq 0$$

NOTE — La loi de probabilité de la variable $\chi^2/2$ est une loi gamma de paramètre $m = \nu/2$. (Voir 1.38.)

1.34 loi de t ; loi de Student : Loi du quotient de variables aléatoires indépendantes, le numérateur étant une variable normale réduite et le dénominateur étant la racine carrée positive du quotient d'une variable χ^2 par son nombre de degrés de liberté.

Ce nombre de degrés de liberté de χ^2 est le nombre ν de degrés de liberté de la variable t .

La densité de probabilité de la variable t est

$$f(t, \nu) = \frac{1}{\sqrt{\pi} \nu} \frac{\Gamma[(\nu+1)/2]}{\Gamma(\nu/2)} \frac{1}{(1+t^2/\nu)^{(\nu+1)/2}}$$

$$-\infty < t < +\infty$$

1.35 F-distribution : The distribution of the quotient of two independent χ^2 distributed variates, each one divided by its number of degrees of freedom. The numbers of degrees of freedom of the χ^2 distributed variates of the numerator ν_1 and of the denominator ν_2 are, in this order, the numbers of degrees of freedom of the F -distributed variate.

NOTE — The probability density function of the F -distributed variate is

$$f(F, \nu_1, \nu_2) = \frac{\Gamma[(\nu_1 + \nu_2)/2]}{\Gamma(\nu_1/2) \Gamma(\nu_2/2)} (\nu_1)^{\nu_1/2} (\nu_2)^{\nu_2/2} \frac{F^{(\nu_1/2)-1}}{(\nu_1 F + \nu_2)^{(\nu_1 + \nu_2)/2}}$$

$$F \geq 0$$

1.36 log-normal distribution : The probability distribution of a continuous random variable X which can take any value from a to infinity, with the probability density

$$f(x) = \frac{1}{(x-a) \sigma \sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{\log_e (x-a) - m}{\sigma} \right)^2 \right]$$

$$x > a$$

m and σ being respectively the mean and the standard deviation of $\log_e (X-a)$.

NOTES

1 The probability distribution of the variate $\log_e (X-a)$ is a normal distribution; m and σ are respectively the expectation and the standard deviation of this variate.

2 The $\log_{10}$ is often used instead of the $\log_e$. In this case :

$$f(x) = \frac{0,4343}{(x-a) \sigma \sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{\log_{10} (x-a) - m}{\sigma} \right)^2 \right]$$

m and σ being respectively the mean and the standard deviation of $\log_{10} (X-a)$.

1.37 exponential distribution : The probability distribution of a continuous random variable X which can take any value from 0 to $+\infty$, the distribution function of which is

$$F(x) = 1 - e^{-\lambda x}$$

$$\lambda > 0$$

This probability distribution may be generalized by substituting $x-a$ for x (with $x \geq a$).

1.38 gamma distribution : The probability distribution of a continuous random variable X which can take any value from 0 to $+\infty$ with the probability density

$$f(x) = \frac{e^{-x} x^{m-1}}{\Gamma(m)}$$

with $m > 0, x \geq 0$

and $\Gamma(m) = \int_0^\infty e^{-x} x^{m-1} dx$

m , a positive constant, determines the form of the distribution. When m is an integer, we have

$$\Gamma(m) = (m-1)!$$

1.35 loi de F : Loi du quotient de deux variables χ^2 indépendantes, chacune étant divisée par son nombre de degrés de liberté. Les nombres de degrés de liberté des variables χ^2 du numérateur, ν_1 , et du dénominateur, ν_2 , sont, dans cet ordre, les nombres de degrés de liberté de la variable F .

NOTE — La densité de probabilité de la variable F est

$$f(F, \nu_1, \nu_2) = \frac{\Gamma[(\nu_1 + \nu_2)/2]}{\Gamma(\nu_1/2) \Gamma(\nu_2/2)} (\nu_1)^{\nu_1/2} (\nu_2)^{\nu_2/2} \frac{F^{(\nu_1/2)-1}}{(\nu_1 F + \nu_2)^{(\nu_1 + \nu_2)/2}}$$

$$F \geq 0$$

1.36 loi log-normale : Loi de probabilité d'une variable aléatoire continue X , pouvant prendre toute valeur depuis a jusqu'à l'infini, avec la densité de probabilité

$$f(x) = \frac{1}{(x-a) \sigma \sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{\log_e (x-a) - m}{\sigma} \right)^2 \right]$$

$$x > a$$

m et σ étant respectivement la moyenne et l'écart-type de $\log_e (X-a)$.

NOTES

1 La loi de probabilité de la variable $\log_e (X-a)$ est une loi normale; m et σ sont respectivement l'espérance mathématique et l'écart-type de cette variable.

2 Le $\log_{10}$ est fréquemment employé à la place du $\log_e$. Dans ce cas,

$$f(x) = \frac{0,4343}{(x-a) \sigma \sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{\log_{10} (x-a) - m}{\sigma} \right)^2 \right]$$

m et σ étant respectivement la moyenne et l'écart-type de $\log_{10} (X-a)$.

1.37 loi exponentielle : Loi de probabilité d'une variable aléatoire continue X , pouvant prendre toute valeur de 0 à $+\infty$, dont la fonction de répartition est

$$F(x) = 1 - e^{-\lambda x}$$

$$\lambda > 0$$

Cette loi de probabilité peut être généralisée en remplaçant x par $x-a$ (avec $x \geq a$).

1.38 loi gamma : Loi de probabilité d'une variable aléatoire continue X , pouvant prendre toute valeur de 0 à $+\infty$, avec la densité de probabilité

$$f(x) = \frac{e^{-x} x^{m-1}}{\Gamma(m)}$$

avec $m > 0, x \geq 0$

et $\Gamma(m) = \int_0^\infty e^{-x} x^{m-1} dx$

m , constante positive, détermine la forme de la distribution. Lorsque m est un nombre entier, on a

$$\Gamma(m) = (m-1)!$$

For $m = 1$, the gamma distribution becomes an exponential distribution.

This probability distribution may be generalized by substituting $(x-a)/b$ for x (with $x \geq a$ and $b > 0$).

1.39 Gumbel distribution (type I extreme value distribution) : The probability distribution of a continuous random variable X , the distribution function of which is

$$F(x) = \exp(-e^{-y})$$

with $y = (x-a)/b, b > 0, -\infty < x < +\infty$

1.40 Fréchet distribution (type II extreme value distribution) : The probability distribution of a continuous random variable X , the distribution function of which is

$$F(x) = \exp(-y^k)$$

with $y = (x-a)/b, b > 0, x \geq a$

k , a positive constant, determines the shape of the distribution.

1.41 Weibull distribution (type III extreme value distribution) : The probability distribution of a continuous random variable X , the distribution function of which is

$$F(x) = 1 - \exp(-y^k)$$

with $y = (x-a)/b, b > 0, x \geq a$

k , a positive constant, determines the shape of the distribution.

1.42 binomial distribution : The probability distribution of a discrete random variable such that, if $x \in \{0, 1, 2, \dots, n\}$, then

$$Pr[X=x] = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

$$0 < p < 1$$

1.43 negative binomial distribution : The probability distribution of a discrete random variable X such that, if $x \in \{0, 1, 2, \dots\}$, then

$$Pr[X=x] = \frac{c(c+1)\dots(c+x-1)}{x!} p^c (1-p)^{x-c}$$

with $c > 0$ and $0 < p < 1$

NOTE — The name "negative binomial distribution" is derived from the fact that the successive probabilities for $x = 0, 1, 2, \dots$, are obtained by developing the binomial of negative exponent $-c$:

$$p^c [1 - (1-p)]^{-c}$$

following the positive and integral powers of $1-p$.

Pour $m = 1$, la loi gamma devient une loi exponentielle.

Cette loi de probabilité peut être généralisée en remplaçant x par $(x-a)/b$ (avec $x \geq a$ et $b > 0$).

1.39 loi de Gumbel (loi des valeurs extrêmes du type I) : Loi de probabilité d'une variable aléatoire continue X dont la fonction de répartition est

$$F(x) = \exp(-e^{-y})$$

avec $y = (x-a)/b, b > 0, -\infty < x < +\infty$

1.40 loi de Fréchet (loi des valeurs extrêmes du type II) : Loi de probabilité d'une variable aléatoire continue X dont la fonction de répartition est

$$F(x) = \exp(-y^k)$$

avec $y = (x-a)/b, b > 0, x \geq a$

k , constante positive, détermine la forme de la distribution.

1.41 loi de Weibull (loi des valeurs extrêmes du type III) : Loi de probabilité d'une variable aléatoire continue X dont la fonction de répartition est

$$F(x) = 1 - \exp(-y^k)$$

avec $y = (x-a)/b, b > 0, x \geq a$

k , constante positive, détermine la forme de la distribution.

1.42 loi binomiale : Loi de probabilité d'une variable aléatoire discrète X , telle que, si $x \in \{0, 1, 2, \dots, n\}$, on a

$$Pr[X=x] = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

$$0 < p < 1$$

1.43 loi binomiale négative : Loi de probabilité d'une variable aléatoire discrète X , telle que, si $x \in \{0, 1, 2, \dots\}$, on a

$$Pr[X=x] = \frac{c(c+1)\dots(c+x-1)}{x!} p^c (1-p)^{x-c}$$

avec $c > 0$ et $0 < p < 1$

NOTE — L'appellation de «loi binomiale négative» vient de ce que les probabilités successives, pour $x = 0, 1, 2, \dots$, s'obtiennent en développant le binôme d'exposant négatif $-c$:

$$p^c [1 - (1-p)]^{-c}$$

suivant les puissances entières et positives de $1-p$.

BS 5532 : Part 1 : 1978

1.44 Poisson distribution : The probability distribution of a discrete random variable X such that, if $x \in \{0, 1, 2, \dots\}$, then

$$Pr [X = x] = e^{-m} \frac{m^x}{x!}$$

where m is a positive parameter.

NOTE — The expectation and the variance of the Poisson distribution are both equal to m .

1.45 hypergeometric distribution : If three positive or null integers N , n and d are given such that the numbers in the table below are positive or null integers,

N	d	$N-d$
n	x	$n-x$
$N-n$	$d-x$	$N-n-d+x$

then

$$Pr [X = x] = \frac{n! (N-n)! d! (N-d)!}{N! x! (n-x)! (d-x)! (N-n-d+x)!}$$

1.46 bivariate normal distribution; bivariate Laplace-Gauss distribution : The probability distribution of two continuous variates X and Y , such that the probability density function is

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-m_X}{\sigma_X} \right)^2 - 2\rho \left(\frac{x-m_X}{\sigma_X} \right) \left(\frac{y-m_Y}{\sigma_Y} \right) + \left(\frac{y-m_Y}{\sigma_Y} \right)^2 \right] \right\}$$

with $-\infty < x < +\infty$

and $-\infty < y < +\infty$

NOTES

1 m_X and m_Y are the mathematical expectations and σ_X and σ_Y are the standard deviations of the marginal distributions of X and Y which are normal distributions; ρ is the coefficient of correlation of X and Y .

2 This concept can be extended to a distribution with more than two variates.

1.47 standardized bivariate normal distribution; standardized Laplace-Gauss distribution : The probability distribution of a pair of standardized normal variates. For a pair of normal variates (X, Y) of parameters (m_X, m_Y) and (σ_X, σ_Y) , the corresponding standardized variates are

$$U = \frac{X - m_X}{\sigma_X} \text{ and } V = \frac{Y - m_Y}{\sigma_Y}$$

1.44 loi de Poisson : Loi de probabilité d'une variable aléatoire discrète X , telle que, si $x \in \{0, 1, 2, \dots\}$, on a

$$Pr [X = x] = e^{-m} \frac{m^x}{x!}$$

où m est un paramètre positif.

NOTE — L'espérance mathématique et la variance de la loi de Poisson sont toutes deux égales à la valeur du paramètre m .

1.45 loi hypergéométrique : Étant donné trois entiers positifs ou nuls N , n et d tels que les nombres du tableau suivant soient des entiers positifs ou nuls,

N	d	$N-d$
n	x	$n-x$
$N-n$	$d-x$	$N-n-d+x$

on a

$$Pr [X = x] = \frac{n! (N-n)! d! (N-d)!}{N! x! (n-x)! (d-x)! (N-n-d+x)!}$$

1.46 loi normale à deux variables; loi de Laplace-Gauss à deux variables : Loi de probabilité à deux variables aléatoires continues X et Y , telle que la densité de probabilité soit

avec $-\infty < x < +\infty$

et $-\infty < y < +\infty$

NOTES

1 m_X et m_Y sont les espérances mathématiques et σ_X et σ_Y sont les écarts-types des lois marginales de X et de Y qui sont des lois normales; ρ est le coefficient de corrélation de X et Y .

2 Ce concept s'étend à une distribution à plus de deux variables.

1.47 loi normale réduite à deux variables; loi de Laplace-Gauss réduite à deux variables : Loi de probabilité d'un couple de variables normales réduites. Pour un couple de variables normales (X, Y) de paramètres (m_X, m_Y) et (σ_X, σ_Y) , les variables réduites correspondantes sont

$$U = \frac{X - m_X}{\sigma_X} \text{ et } V = \frac{Y - m_Y}{\sigma_Y}$$

and the probability density function is :

$$f(u, v) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)} (u^2 - 2\rho uv + v^2)\right]$$

with $-\infty < u < +\infty$

and $-\infty < v < +\infty$

(ρ is the coefficient of correlation of X and Y , and also of U and V .)

NOTE — This concept can be extended to a distribution with more than two variates.

1.48 multinomial distribution : The probability distribution of k discrete variates $X_1, X_2, \dots, X_k$, such that, if $x_1, x_2, \dots, x_k$ are integers taken from 0, 1, 2, 3, ..., n , and such that their sum is $x_1 + x_2 + \dots + x_k = n$, then

$$Pr [X_1 = x_1, X_2 = x_2, \dots, X_k = x_k] = \frac{n!}{x_1! x_2! \dots x_k!} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}$$

with $p_i \geq 0$ ($i = 1, 2, \dots, k$)

and

$$\sum_{i=1}^k p_i = 1$$

et la densité de probabilité est

$$f(u, v) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)} (u^2 - 2\rho uv + v^2)\right]$$

avec $-\infty < u < +\infty$

et $-\infty < v < +\infty$

(ρ est le coefficient de corrélation de X et Y aussi bien que de U et V .)

NOTE — Ce concept s'étend à une distribution à plus de deux variables.

1.48 loi multinomiale : Loi de probabilité à k variables discrètes $X_1, X_2, \dots, X_k$, telle que, si $x_1, x_2, \dots, x_k$ sont des entiers pris parmi 0, 1, 2, 3, ..., n et tels que leur somme $x_1 + x_2 + \dots + x_k = n$, on a

avec $p_i \geq 0$ ($i = 1, 2, \dots, k$)

et

$$\sum_{i=1}^k p_i = 1$$

2 GENERAL STATISTICAL TERMS

In order to avoid any confusion with similar terms in clause 1 ("Terms used in theory of probability"), certain definitions in this clause which relate to the concept of sampling may be qualified as empirical. However, this qualifying word is omitted where there is no fear of ambiguity. Furthermore, some terms relating to sampling are used in this clause, which are defined in clause 3 ("General terms relating to methods of sampling").

2.1 item :

- a) An actual or conventional object on which a set of observations may be made, or
- b) A defined quantity of material, on which a set of observations may be made, or
- c) An observed value, either qualitative (attributes) or quantitative (measures).

NOTE — The English terms "individual" and "unit" are sometimes used as synonyms for "item".

2.2 population : The totality of items under consideration.

Every clearly defined part of a population is called a subpopulation.

In the case of a random variable, the probability distribution (1.2) is considered as defining the population of that variable.

2.3 characteristic : A property which helps to differentiate between the items of a given population.

The differentiation may be either quantitative (by variables) or qualitative (by attributes).

2.4 test : An operation made in order to measure or classify a characteristic.

2.5 observed value : The value of a characteristic determined as the result of an observation or test.

2.6 absolute difference : The absolute value of the difference between two values.

2.7 range : The difference between the greatest and the smallest observed values of a quantitative characteristic.

2.8 mid-range : The arithmetic mean of the greatest and the smallest observed values of a quantitative characteristic.

2.9 class : In the case of quantitative characteristics, each of the consecutive intervals into which the total interval of variation is divided.

2 TERMES STATISTIQUES GÉNÉRAUX

Pour éviter toute confusion avec les termes analogues du chapitre 1 («Termes utilisés en calcul des probabilités»), certains termes du présent chapitre relatifs à la notion d'échantillonnage peuvent être qualifiés d'empiriques. Toutefois, ce qualificatif est omis lorsqu'aucune ambiguïté n'est à craindre. D'autre part, il est fait usage dans le présent chapitre, de certains termes relatifs à l'échantillonnage, qui sont définis dans le chapitre 3 («Termes généraux relatifs aux modes d'échantillonnage»).

2.1 individu :

- a) Objet concret ou conventionnel sur lequel un ensemble d'observations peut être fait, ou
- b) Quantité définie de matière, sur laquelle un ensemble d'observations peut être fait, ou
- c) Valeur observée, qu'elle soit qualitative (attributs) ou quantitative (mesures).

NOTE — Les termes anglais «individual» et «unit» sont quelquefois utilisés comme synonymes de «individu».

2.2 population : Totalité des individus pris en considération.

Toute partie nettement définie d'une population est une sous-population.

Dans le cas d'une variable aléatoire, la loi de probabilité (1.2) est considérée comme définissant la population de cette variable.

2.3 caractère : Propriété servant à différencier les individus d'une population donnée.

La différenciation entre individus peut être soit quantitative (par mesures ou par variables), soit qualitative (par attributs).

2.4 essai : Opération faite en vue de mesurer ou de classer un caractère.

2.5 valeur observée : Valeur d'un caractère donnée sous la forme du résultat d'une observation ou d'un essai.

2.6 écart : Valeur absolue de la différence entre deux valeurs.

2.7 étendue : Écart entre la plus grande et la plus petite des valeurs observées d'un caractère quantitatif.

2.8 milieu de l'étendue : Moyenne arithmétique de la plus grande et de la plus petite des valeurs observées d'un caractère quantitatif.

2.9 classes : Dans le cas de caractères quantitatifs, chacun des intervalles contigus en lesquels on a partagé l'intervalle total de variation.

2.10 class limits : The values defining the upper and lower bounds of a class.

NOTE — It should be specified which of the two limits is considered as belonging to the class.

2.11 mid-point of class : The arithmetic mean of the upper and lower limits of a class.

2.12 class interval : The difference between the upper limit and the lower limit of a class.

2.13 absolute frequency; size : The number of items (of a population, lot, sample, class, etc.).

2.14 cumulative absolute frequency : In the case of quantitative characteristics, the number of items the values of which are lower than or equal to a given value, or lower than or equal to the upper limit of a given class.

2.15 relative frequency : The ratio of the number of times a particular value (or a value falling within a given class) is observed to the total number of observations.

2.16 cumulative relative frequency : In the case of quantitative characteristics, the relative frequency of items the values of which are lower than or equal to a given value, or lower than or equal to the upper limit of a given class.

2.17 frequency distribution : The relationship between the values of a characteristic and their absolute or relative frequencies.

The distribution is often presented as a table with special groupings (classes) if the values are measured on a continuous scale.

2.18 univariate distribution : In the case of one variate, the relation between the values, classes or forms of this characteristic on the one hand and their sizes or frequencies on the other hand.

2.19 histogram : A graphical representation of the frequency distribution of a continuous variable.

Classes are represented by contiguous intervals on one of the axes on a linear scale. The absolute (or relative) frequencies of the classes are represented by rectangles having these intervals as bases and areas which are proportional to the absolute (or relative) frequencies of the classes.

2.20 bar diagram : A graphical representation of the frequency distribution of a discrete variate. Straight-line segments (bars), proportional to the absolute (or relative) frequencies of the variates, are drawn perpendicular to an axis on which the values of the variate are represented on a linear scale.

2.21 cumulative absolute (relative) frequency polygon : The polygonal line obtained by joining the points the

2.10 limites de classe : Valeurs qui définissent les bornes supérieure et inférieure d'une classe.

NOTE — On doit préciser laquelle des deux limites est considérée comme appartenant à la classe.

2.11 centre de classe : Moyenne arithmétique des limites supérieure et inférieure d'une classe.

2.12 intervalle de classe : Différence entre la limite supérieure et la limite inférieure d'une classe.

2.13 effectif : Nombre d'individus (d'une population, d'un lot, d'un échantillon, d'une classe, etc.).

2.14 effectif cumulé : Dans le cas de caractères quantitatifs, nombre des individus dont la valeur est inférieure ou égale à une valeur donnée, ou inférieure ou égale à la limite supérieure d'une classe donnée.

2.15 fréquence : Rapport du nombre de fois où une valeur particulière (ou une valeur appartenant à une classe donnée) est observée, au nombre total d'observations.

2.16 fréquence cumulée : Dans le cas de caractères quantitatifs, fréquence des individus dont la valeur est inférieure ou égale à une valeur donnée, ou inférieure ou égale à la limite supérieure d'une classe donnée.

2.17 distribution de fréquence : Relation entre les valeurs d'un caractère et leurs effectifs ou leurs fréquences.

La distribution est souvent présentée sous forme d'un tableau avec des groupements spéciaux (classes) si les valeurs sont mesurées avec une échelle continue.

2.18 distribution à une variable : Dans le cas d'une seule variable, relation entre les valeurs, classes ou modalités de ce caractère d'une part, et leurs effectifs ou leurs fréquences d'autre part.

2.19 histogramme : Représentation graphique d'une distribution de fréquence d'une variable continue.

Les classes sont représentées par des intervalles contigus sur l'un des axes suivant une échelle linéaire. Les effectifs (ou fréquences) des classes sont représentés par des rectangles dont les bases sont les intervalles précédents et dont les aires sont proportionnelles aux effectifs (ou aux fréquences) des classes.

2.20 diagramme en bâtons : Représentation graphique d'une distribution de fréquence d'une variable discrète. Perpendiculairement à un axe sur lequel sont représentées les valeurs de la variable, suivant une échelle linéaire, sont élevés des segments rectilignes (bâtons) proportionnels aux effectifs (ou aux fréquences) de ces valeurs.

2.21 polygone des effectifs (ou des fréquences) cumulés : Ligne polygonale obtenue en joignant les points ayant pour

abscissa of which is the upper limit on each class and the ordinate of which is either the cumulative absolute frequency or the cumulative relative frequency.

2.22 bivariate distribution : In the case of two variates considered simultaneously, the relation between the values, classes or forms of characteristics on the one hand and their sizes or frequencies on the other hand.

2.23 multivariate distribution : In the case of several variates considered simultaneously, the relation between the values, classes or forms of characteristics on the one hand and their sizes or frequencies on the other hand.

2.24 two-way table : A table used to represent a distribution with two characteristics and in which the rows and the columns correspond to the values or classes of the first characteristic and second characteristic respectively.

At the intersection of a row and a column appears the number or the frequency corresponding to the combination of values, classes or forms thus marked.

2.25 contingency table : A two-way table in the case of two qualitative characteristics.

The concept may be generalized to the case of more than two characteristics.

2.26 marginal distribution : In a distribution with two variates, the distribution of one of the variates, the other being ignored. It may be obtained by adding for each value or class of the variate under consideration the corresponding absolute (or relative) frequencies of the other variate.

The notion of marginal distribution may be extended to more than two variates. In this case, a marginal distribution is a univariate or a multivariate distribution, the other variates being integrated out.

2.27 conditional distribution : In a distribution with two variates, the distribution of one of the variates when the other one is fixed.

There are, for each variate, as many conditional distributions as there are values or classes for the other variate.

A conditional distribution is expressed by indicating the variate to which it applies and the value or class fixed for the other variate : for example, conditional distribution of the variate Y given $X = x$.

Conditional distributions of absolute frequencies are read directly in the rows and columns of the two-way table of frequencies. The conditional distributions of relative frequencies are obtained by dividing the numbers in each row (or column) by the total number in the relevant row (or column).

abscisse la limite supérieure de chaque classe et pour ordonnée soit l'effectif cumulé, soit la fréquence cumulée.

2.22 distribution à deux variables : Dans le cas de deux variables considérées simultanément, relation entre les valeurs, classes ou modalités de ces caractères d'une part, et leurs effectifs ou leurs fréquences d'autre part.

2.23 distribution à plusieurs variables : Dans le cas de plusieurs variables considérées simultanément, relation entre les valeurs, classes ou modalités de ces caractères d'une part, et leurs effectifs ou leurs fréquences d'autre part.

2.24 tableau à double entrée : Tableau utilisé pour représenter une distribution à deux caractères et dont les lignes d'une part, les colonnes d'autre part, correspondent respectivement aux valeurs, classes ou modalités de l'un et de l'autre caractère.

À l'intersection d'une ligne et d'une colonne figure l'effectif ou la fréquence correspondant à la combinaison de valeurs, classes ou modalités ainsi repérée.

2.25 tableau de contingence : Tableau à double entrée dans le cas de deux caractères qualitatifs.

La notion s'étend au cas de plus de deux caractères.

2.26 distribution marginale : Dans une distribution à deux variables, distribution de l'une des variables lorsqu'on ne tient pas compte de l'autre. Elle peut s'obtenir en faisant pour chaque valeur ou classe de la variable considérée la somme des effectifs (ou des fréquences) correspondant à l'autre variable.

La notion de distribution marginale s'étend à un nombre de variables supérieur à deux. Une distribution marginale est alors une distribution à une ou plusieurs variables, les autres variables n'étant pas prises en considération.

2.27 distribution conventionnelle : Dans une distribution à deux variables, distribution de l'une des variables lorsque l'autre est fixée.

Il y a, pour chaque variable autant de distributions conditionnelles qu'il y a de valeurs ou classes pour l'autre variable.

Une distribution conditionnelle se désigne en indiquant la variable à laquelle elle s'applique et la valeur ou la classe fixée pour l'autre variable : par exemple, distribution conditionnelle de la variable Y pour $X = x$.

Les distributions conditionnelles des effectifs se lisent directement dans les lignes et les colonnes du tableau à double entrée des effectifs. Les distributions conditionnelles des fréquences s'obtiennent en divisant les effectifs de chaque ligne (ou colonne) par l'effectif total de la ligne (ou colonne).